Essays on forecasting and latent values

ISBN: 978 90 5170 725 0

This book is no. 638 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Essays on forecasting and latent values

Essays over voorspellen en latente waarden

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Thursday, December 17, 2015 at 13:30 hours

by

LAMBERT PIETER DE BRUIJN

born in Schiedam, The Netherlands

Erasmus University Rotterdam

Doctorate Committee

Promotor:	Prof.dr. Ph.H.B.F. Franses
Other members:	Prof.dr. E.A. de Groot
	Prof.dr. M.J.C.M. Verbeek
	Prof.dr. H.P. Boswijk

Acknowledgments

For the last four years I have put time and effort into the hundred-and-some pages in front of you. Doing a PhD in general has put me in situations that I otherwise would not have been in (such as traveling to other continents, presenting at conferences and teaching), but this thesis is the ultimate example of this kind: a book with my name on it, that allows me to physically show other people that I have been doing research in these years. Thanks for taking the time and effort to read (at least this part of) my thesis.

A majority of a PhD is filled with research, and this work does not always result in an immediate payoff. During these past years I and many of my fellow PhD students have experienced the hard reality of the academic world. I haven't tried to count the number of times during lunch someone told about his rejected paper, but if I would have tried, I would have probably lost count. It is a valuable life lesson to know how to keep yourself and others motivated if (or better: when) the revolutionary results that we have expected don't show themselves that easily.

In this endeavor a PhD student needs the supervision and guidance of his promotor, and Philip Hans Franses has fulfilled that part extremely well. Combining teaching, research and supervision while also being the dean of the faculty, it comes to no surprise that he has a busy schedule. Still, he always tried to make time for me if I entered his room, with or without an appointment. During my PhD he recognized when it was best to leave me to my own work during the periods in which I could drive my research myself, but during the other periods he was always willing to provide the feedback or advise needed to get a project back on the road. He has the great gift of turning difficult problems into easy ones.

Next to Philip Hans, I also would like to thank my other co-authors, Rene, Michael and Chia-Lin. During the writing process of each paper on which a chapter of this thesis is based, there has always been a good interaction between me and the corresponding coauthors concerning ideas, modeling, programming, interpreting results and writing the different sections of each paper. I also would like to thank the thesis committee, consisting of Bert de Groot, Marno Verbeek and Peter Boswijk, for their useful comments and quick evaluation of this thesis.

Another part of a PhD is teaching. For a beginning PhD student, teaching can be a very daunting task, but the Assistant and Associate Professors that I have assisted have made it a lot more easy. Christiaan, Andreas and Jan, thanks for the advice and guidance during those intense months in which both you and I had to make long days in front of classes, checking homework or exams or answering long emails containing questions. Despite each of you having a totally different approach to course management, there has never been a problem that we couldn't solve. I also would like to thank support staff like both secretariats, IT, the financial administrators and many more, because without their help, it would be impossible for the academic staff to focus on their core job.

Having a roommate with whom you can discuss research methods, academic bureaucracy and teaching skills is a useful asset, especially in the beginning of the PhD. I am grateful to have had two roommates who both could win a competition on friendliness and positivity. Mathijn has only been my roommate for about the first year of my PhD, but in that year he has been a great help in introducing me to several parts of the institute or research in general. Koen has been my roommate throughout the entirety of my PhD, and next to that he was also my colleague tutorial teacher for most of the tutorials that I have taught. It has been very valuable to have a roommate in exactly the same position for each year. We have had a very good interaction concerning stuff like following and teaching courses, findings answers to difficult problems in our research and the day to day conversation about stuff like the news, the weather or a surprising email that one of us has just received. Also thanks for things like all the reminders for every time I almost forgot an event. Based on these years I can safely say that your new employer could not have recruited a better PhD.

Next to my roommates, I have also had the pleasure to come in contact with many other PhD's throughout the years. Thanks to everyone for the good (and relaxing) conversations during and after lunches, seminars, courses, drinks and outings. Naming all twenty-odd would be a bit too much, but your presence has not gone by unnoticed. The future will show how much such simple conversations can influence the rest of a life. I also would like to thank my friends outside the campus, as well as my parents and family. Eight years have passed since I started my bachelor in Econometrics, but explaining to you what I do has always remained a challenge. Still, I am grateful for the attempts to try to understand it. Thanks for the moments in which the econometrician in me can relax.

Finally, but above all, I would like to thank my wife Liesbeth. There is no one who knows me as well as you do, which makes it a big compliment that you wanted to be my wife. An enormous thanks for all the times in which you have supported me and motivated me to keep going, not only during my PhD but in all parts of life. Thank you for everything that you have taught me. I don't know who I'd be without you, but I know I wouldn't be me.

Bert de Bruijn,
October 2015

Contents

List of Tables

List of Figures

Chapter 1

Introduction and outline

Variables are central in empirical data analysis. Analysts explore and try to explain relationships between variables using various econometric models. One may now think that there is no way to find relationships between variables without having these variables, but actually, though, this is not true. In various situations there are techniques to estimate the values of unobserved variables, also known as latent variables. After discovering these latent variables (and sometimes during their discovery) they can be related to other, observed data. This allows for new insights that otherwise would be unattainable. Unobserved variables are a central theme among in this thesis, which consists of five distinct essays that are partially on latent variables.

One line of research in this area focuses on variables that surely are real, but just unobserved. An example of this is the research on evaluating forecasts created by professional forecasters. These forecasts could be a combination of a model forecast and a judgmental component, often called intuition (see Fildes et al., 2009 for an example and Lawrence et al., 2006 for a review). In some cases only the final forecast is observed by the analyst. The professional forecaster on the other hand knows whether the forecast was purely based on a model or on intuition or a combination, and in the case of the latter, the forecaster also knows what part is model and what part is not. In other words, in this case the unobserved intuition does exist (even though it might be zero if the forecast is a purely model forecast). The analyst might not have access to an explicit expression of the intuition, and thus the analyst may be interested in reconstructing the intuition, in order to evaluate this intuition on several criteria. Other examples of existing but unobserved variables are the set of products a customer considers before buying one (Ben-Akiva and Boccara, 1995) and potential workers' reservation wage (such as in Maloney, 1991).

Sometimes, the unobserved variable does not really exist, but could have existed without a stretch of the imagination. For example, consider a situation in which a researcher collects data for several individuals on different days, but on some days some of the individuals do not turn up. In this case, the researcher could leave out these individuals for the analysis, but this could introduce sampling bias. Indeed, the drop-outs could have characteristics that are related to the variables being researched. Instead, the researcher can also try to reconstruct what the data would have been had the individuals turned up. Examples of this situation can be found in research using questionnaires (Sijtsma and van der Ark, 2003) and experiments (Little, 1995). This kind of what-if research can also be done using different setups, such as in policy research (Cunha et al., 2006) or marketing (Liu, 2010).

Finally, there are also situations in which the unobserved variable most likely does not exist, but is introduced by the analyst to find some kind of structure in an otherwise difficult to analyze dataset. This is often the case when analyzing large datasets. For example, an analyst may want to cluster observations or subjects into several groups to make general conclusions for these clusters. Then, not only the clustering variable itself but also the number of clusters is unobserved. If the true number of clusters is higher than the number chosen by the analyst, then some of the clusters the analyst will find are not really there, but are instead some kind of amalgamation of real clusters. This does not mean that the results are not useful, as it might be the case that there are no large differences between the clusters as compared to the other clusters. Examples of such unobserved clusters and several relevant estimation methods can be found in Wedel and Kamakura (2000). An extreme case of clustering (and implicitly applied very often) concerns the assumption that there is only one cluster, which implies the assumption of homogeneity. This assumption might not always be realistic, but it can be necessary to summarize results, such as when working with large datasets (Evans, 1987). Another situation in which unobserved variables are introduced to provide the researcher with structure in the data concerns the case of multi-level equation models (Fok and Franses, 2007; Fok et al., 2005).

In sum, there are arguments for the value of uncovering unobserved variables in empirical data analysis. It should be noted however that uncovering unobserved variables in an accurate and unbiased manner can be very difficult. For a start, the variables are unobserved, which means that techniques to uncover them cannot be formally tested in these situations. Instead, artificial situations are then created in which researchers act as if they do not have some particular data, which allows them to see how well they have performed in recreating these variables. This provides some idea of how well the reconstruction of data performs for that model, but it cannot control for unknown and unexpected characteristics that make certain variables unobserved. Another difficult part of this type of research is that there is often not much information that guides the reconstruction, and because of this there is often much uncertainty about the accuracy of the estimated values of the unobserved variable. Many researchers rely on properties of large datasets to at least be correct on average, while hoping that opposite deviations from this average mutually cancel out. This might not be true if there are non-linear effects, in which case it might be better to also consider the full distribution of the unknown variable instead of just its mean. But sometimes this is not feasible for computational reasons. A final complicating element is that there are often other confounding factors that blur the sparse information that is available. For example, in the previously discussed situation of trying to split a particular forecast into a model-based part and an intuition part, it might be the case that the model that the forecaster used has changed throughout the sample period. The analyst would be unaware of this change and will probably assume a fixed model, and will thus attribute any change in properties to the intuition component.

An important aspect to keep in mind when doing this type of research is that the estimated values of the unobserved variable are not equal to the true value of the unobserved variable. This may or may not have consequences for the final results. For example, the previously mentioned individuals who do not turn up might have behaved very differently than if they would have turned up. This might or might not be related to the reason for them not to show up. Or in the case of trying to cluster individuals, it might be the case that some individuals of cluster A happen to show properties in the sample that are more comparable to cluster B, and thus they are incorrectly assigned to that other cluster. This shows that is not only important to develop techniques to uncover unobserved variables, but also to find out what are the drawbacks and limitations of those techniques.

As said before, unobserved variables are a common topic in all chapters in this thesis. Another topic that appears in some form in all chapters concerns forecasting. Forecasting as a research topic has many different components. One type of research is focused on producing these forecasts using models, either as the main goal of the paper (such as Batchelor and Dua, 1998; Clements et al., 2004) or as tool to illustrate the usefulness of a certain model or technique (such as Engle and Watson, 1981). Forecasting research can also investigate theoretical forecast properties (Yang, 2004) or actual properties of already existing forecasts (such as Franses et al., 2011; Sheng, 2015, and many others) for various forecasting situations. All of the preceding components of forecasting are featured in at least one of the upcoming chapters of this thesis.

Outline of this thesis

It is common practice to evaluate fixed-event forecast revisions in macroeconomics by regressing current forecast revisions on one-period lagged forecast revisions. Under weak-form (forecast) efficiency, the correlation between the current and one-period lagged revisions should be zero. The empirical findings in the literature suggest that the null hypothesis of zero correlation between the current and one-period lagged revisions is rejected very frequently, where the correlation can be either positive (which is widely interpreted in the literature as "smoothing") or negative (which is widely interpreted as "over-reacting"). Chapter 2 concerns a methodology to be able to interpret such non-zero correlations in a straightforward and clear manner. The approach is based on the assumption that numerical forecasts can be decomposed into both a forecast from an econometric model and random expert intuition. It is shown that the interpretation of the sign of the correlation between the current and one-period lagged revisions depends on the process governing intuition, and the current and lagged correlations between intuition and news (or shocks to the numerical forecasts). It follows that the estimated non-zero correlation cannot be given a direct interpretation in terms of smoothing or over-reaction. It is also shown that smoothing and over-reaction, modelled and interpreted correctly, can change over time. An empirical example is given to highlight the usefulness of the proposed methodology.

Chapter 3 also focuses on expert forecasts. There is ample empirical evidence that expert-adjusted model forecasts can be improved. One way to potential improvement concerns providing various forms of feedback to the sales forecasters. It is also often recognized that the experts (forecasters) might not constitute a homogeneous group. Chapter 3 provides a data-based methodology to discern latent clusters of forecasters, and applies it to a fully new large database with data on expert-adjusted forecasts, model forecasts and realizations. For the data at hand, two clusters can clearly be identified. Next, the consequences for providing subsequent feedback are discussed.

Chapter 4 puts forward a new data collection method to measure daily consumer confidence at the individual level. The data thus obtained allow to statistically analyze the dynamic correlation of such a consumer confidence indicator and to draw inference on transition rates. The latter is not possible for currently available monthly data collected by statistical agencies on the basis of repeated cross-sections. In an application to measuring Dutch consumer confidence, it is shown that the incremental information content in the novel indicator helps to better forecast consumption.

Chapter 5 concerns an analysis of about 300000 earnings forecasts, created by 18000 individual forecasters for earnings of over 300 S&P listed firms. The analysis shows that these forecasts are predictable to a large extent using a statistical model that includes publicly available information. When the focus is on the unpredictable components, which may be viewed as the personal expertise of the earnings forecasters, it can be learned that small adjustments to the model forecasts lead to more forecast accuracy. Based on past track records, it is possible to predict the future track record of individual forecasters.

Finally, in Chapter 6 a new time series model is introduced that can capture the properties of data as is typically exemplified by monthly US unemployment data. These data show the familiar nonlinear features, with steeper increases in unemployment during economic downturns than the decreases during economic prosperity. At the same time, the levels of unemployment in each of the two states do not seem fixed, nor are the transition periods abrupt. Finally, our model should generate out-of-sample forecasts that mimic the in-sample properties. It is demonstrated that the new and flexible model covers all those features, and its illustration to monthly US unemployment data shows its merits, both in and out of sample.

Chapter 2

Analyzing fixed-event forecast revisions

Based on Chang, de Bruijn, Franses, and McAleer (2013). All authors contributed equally.

Abstract

It is common practice to evaluate fixed-event forecast revisions in macroeconomics by regressing current forecast revisions on one-period lagged forecast revisions. Under weak-form (forecast) efficiency, the correlation between the current and one-period lagged revisions should be zero. The empirical findings in the literature suggest that the null hypothesis of zero correlation between the current and one-period lagged revisions is rejected frequently, where the correlation can be either positive (which is widely interpreted in the literature as "smoothing") or negative (which is widely interpreted as "over-reacting"). In this chapter we propose a methodology to be able to interpret such non-zero correlations in a straightforward and clear manner. Our approach is based on the assumption that numerical forecasts can be decomposed into both an econometric model and random expert intuition. We show that the interpretation of the sign of the correlation between the current and one-period lagged revisions depends on the process governing intuition, and the current and lagged correlations between intuition and news (or shocks to the numerical forecasts). It follows that the estimated non-zero correlation cannot be given a direct interpretation in terms of smoothing or over-reaction. It is also shown that smoothing and over-reaction, modelled and interpreted correctly, can change over time. An empirical example is given to highlight the usefulness of the proposed methodology.

2.1 Introduction

There is a substantial body of recent literature on the evaluation of macroeconomic forecasts and, in particular, on forecast revisions. Such revisions involve potential changes in the forecasts for the same fixed event. For example, Consensus Forecasters quote forecasts for the value of an economic variable (such as the inflation rate, unemployment rate, real GDP growth rate) in year T, where the forecast origin starts in January of year $T-1$. When these forecasts continue through to December in year T, there are 24 forecasts for the same fixed event, and hence there are 23 forecast revisions (or updates).

The literature on forecast revisions deals with the empirical merits of these revisions (see, for example, Lawrence and O'Connor, 2000; Cho, 2002) but, for a larger part, it seems to deal with the properties of the updates themselves (see, for example, the recent study of Dovern and Weisser, 2011). The latter seems to be inspired by the recent availability of databases with detailed information of forecasts quoted by a range of professional forecasters.

In this chapter, we contribute to this second stream of literature, that is, an evaluation of the properties of the forecast revisions themselves where, in particular, we show how to interpret a key parameter in an auxiliary testing regression.

In the fixed-event forecast revision literature (see, for example, Chang et al., 2011) numerical forecasts are taken as data. It is not necessarily known how the numerical forecasts were obtained. We denote a forecast given at origin $t-h$, for fixed-event forecast horizon t, as

$$F_{t|t-h} \tag{2.1}$$

where $h = 1, \ldots, H$. Therefore, for each event t, we have H forecasts, ranging from a one-step-ahead forecast to an H-step-ahead forecast. A (first-order) forecast revision is defined by

$$F_{t|t-h} - F_{t|t-(h+1)} \tag{2.2}$$

and it is this type of forecast revision that is the focus of this chapter.

A commonly-used method to examine the potential properties of forecast revisions is to use auxiliary testing regressions of the form:

$$F_{t|t-h} - F_{t|t-(h+1)} = \alpha + \beta(F_{t|t-(h+1)} - F_{t|t-(h+2)}) + \varepsilon_{t,h} \tag{2.3}$$

where the value of β is of key interest.

Nordhaus (1987) introduced the concept of weak-form efficiency, which entails that, under such efficiency the correlation between subsequent forecast revisions is zero. In other words, under weak-form efficiency, it should be the case that $\beta = 0$ in equation (2.3). As Nordhaus (1987) was concerned with forecasts from econometric models, it is appropriate to refer to this concept as "weak-form model forecast efficiency", whereby fixed-event forecasts taken one period apart differ only randomly. Thus, there is no discernible improvement in forecasts as the fixed event becomes less distant.

It should be emphasized that equation (2.3) is solely a testing equation, and is not a model. The sole purpose of equation (2.3) is to test the null hypothesis of weak-form efficiency, that is, $\beta = 0$. It must be emphasized that rejection of $\beta = 0$ is not synonymous with interpreting equation (2.3) as an appropriate specification for modelling forecast revisions. If this were the case, then equation (2.3) would be used to estimate forecast revisions rather than for testing the weak-form efficiency of forecast revisions. Therefore, equation (2.3) can be interpreted only in terms of testing the null hypothesis, $\beta = 0$, so that any other interpretation of β, such as when β is not zero, is intrinsically meaningless. We will return to this key issue below.

A further point to emphasize is that, as an AR(1) process for testing purposes, equation (2.3) exhibits geometric decay, regardless of the sign or magnitude of β. Therefore, the widely-used interpretations of smoothing and over-reaction based on whether β is estimated to be positive or negative, respectively, in equation (2.3), must be taken as inherently flawed.

Interestingly, in various recent studies that have analyzed a range of forecast revisions, it has frequently been found that the null hypothesis $\beta = 0$ is rejected (see Table 2.1). When it is found that $\beta > 0$, the situation is sometimes interpreted as "forecast smoothing" (see, for example, Isengildina et al., 2006). On the other hand, when it is found that $\beta < 0$, it is believed to be a sign of efficient behavior in the event that there is no news in the forecast data (see, for example, Clements, 1997). When there is news, a negative β is interpreted as over-reaction (as will become clear below).

In this chapter, we propose a methodology to provide an interpretation of the alternative sign outcomes of β arising from equation (2.3). The new approach is based on our conjecture that available forecasts are typically the concerted outcome of an econometric model-based forecast, $M_{t|t-h}$, and of the intuition of an expert (such as a professional forecaster), $\nu_{t|t-h}$

(see, for example, Franses et al., 2011 for substantial empirical evidence regarding this conjecture).

There are various reasons why forecasters may deviate from a pure econometric model-based forecast. Examples are that forecasters aim to attract attention (see Laster et al., 1999), or may have alternative loss functions (see, for example, Capistran and Timmermann, 2009).

In what follows, we use the decomposition of an available numerical forecast, which is taken to be the underlying variable of interest, as

$$F_{t|t-h} = M_{t|t-h} + \nu_{t|t-h} \tag{2.4}$$

It will become apparent that changing $M_{t|t-h}$ into $\pi M_{t|t-h}$, with $0 < \pi < 1$, whereby the model forecast may be down-weighted by the expert, does not change the discussion appreciably. Our next step is to propose a model for the intuition $\nu_{t|t-h}$, and to allow for correlation between intuition and the error term $\varepsilon_{t,h}$, in the model. Note that intuition does not need to have mean zero. The interpretation of the sign of the correlation between the current and one-period lagged revisions depends on the process governing intuition, and the correlations between current and one-period lagged intuition and news to the numerical forecast variable. We illustrate our methodology using empirical results that are available in the literature, several of which are presented in Section 2.2, and for some new results based on the well-known Consensus Forecasts. In Section 2.3 we discuss the methodological approach, and in Section 2.4 we relate it to the empirical findings in the literature. Section 2.5 concludes with several further research issues.

2.2 Empirical Findings in the Literature

In this section we review a selection of the empirical results in the forecasting literature, based on the auxiliary testing regression given in equation (2.3). The selection of these papers is primarily intended to show that numerous estimates of β have been presented in the literature. Further references to research on forecast revisions are given in these papers. There are various studies that provide novel estimation tools for variants of (2.3) in the event there are various forecasters who quote forecasts at the same time, or when there is a correlation between the errors of (2.3) for forecast horizon t and the errors in the equation

Table 2.1 Estimation Results for Variants of Equation (2.3)

Source	Estimates of β, with averaging or pooling	
Clements (1997)	-0.414	(average across 5 cases, GDP)
Table 1, p. 233	-0.232	(average across 5 cases, inflation)
Isengildina et al. (2006)	0.396	(average across 5 cases, Corn)
Table 2, p. 1097	0.212	(average across 5 cases, Soybeans)
Dovern and Weisser (2011)	0.089	(average across G7, GDP)
Table 4, p. 463	-0.040	(average across G7, inflation)
	0.001	(average across G7, industrial production)
	-0.021	(average across G7, private consumption)
Ager et al. (2009)	0.309	(average across 12 countries, GDP)
Tables 5 and 6 pp. 178-179	0.163	(average across 12 countries, inflation)
Isiklar et al. (2006) Table II, p. 710	0.330	(pooled estimated across 18 countries, GDP)
Ashiya (2006)	often > 0	(IMF, OECD forecasts, GDP and inflation)
Loungani (2001)	often > 0	(Consensus forecasts, 63 countries, GDP)
Berger and Krane (1985)	often > 0	(DRI, Chase forecast, US, GNP)

for forecast horizon $t+j$. For ease of discussion, these issues are ignored here, and we focus only on the estimates of β in equation (2.3). A summary of the empirical findings is given in Table 2.1.

Clements (1997) analyzes the forecasts for GDP and CPI made by the National Institute of Economics and Social Research in the UK. Using 5 different versions of equation (2.3), Clements (1997) documents an average value of β of -0.414 for GDP forecast revisions and of -0.232 for inflation forecast revisions (see Clements, 1997, Table 1). In 5 of the 10 cases considered, the negative parameter estimate is also significantly less than 0.

Isengildina et al. (2006) examine forecasts for crop production concerning corn and soybeans, where the forecasts are provided by the US Department of Agriculture. The authors also use various versions of (2.3) and obtain an average estimate of 0.396 of β for corn and 0.212 for soybeans, and also show that 8 of the 10 estimates of β are significantly positive.

Dovern and Weisser (2011) analyze the forecasts obtained from the surveys conducted by Consensus Economics. They focus on individual panelist's forecasts for GDP, inflation, industrial production and private consumption for the G7 countries. They conclude that in only a few cases are the estimated values of β significantly different from 0 but, when they are significant, they are predominantly negative. These authors interpret their finding as an indication that forecasters overreact to incoming news.

Ager et al. (2009) also analyze the Consensus Economics forecasts, but they consider pooled forecasts rather than individual forecasts. They analyze the forecast revisions for GDP and inflation for twelve industrial countries for the years 1996 through to 2006. For GDP they report that in all cases the null hypothesis $\beta = 0$ is rejected, with a mean estimate of 0.309 across 24 cases (namely, 12 countries and 2 methods - see their Table 5). In their Table 6, they report a mean estimate of 0.163 across 24 cases for inflation.

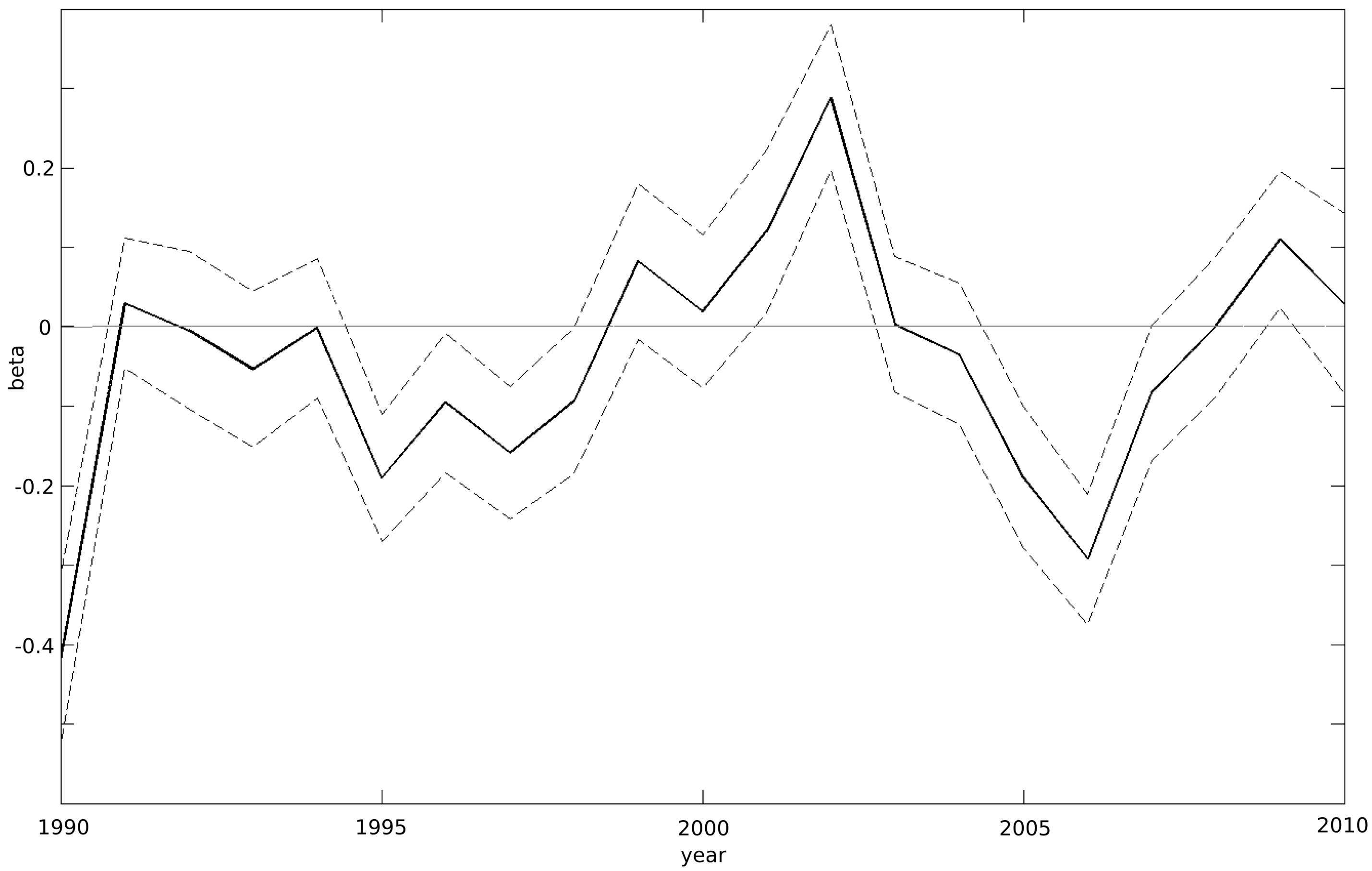
Isiklar et al. (2006) adopt the view that a positive correlation between forecast revisions can occur, and they seek to analyze how long it takes for those correlations to die out. The authors propose using VAR models and impulse response functions, and also use the Consensus Economics forecasts data set, for which they examine 18 industrialized countries and the corresponding GDP growth forecasts. When the authors pool the estimates of β in equation (2.3), they obtain an estimate of 0.330.

Finally, Ashiya (2006), Loungani (2001), and an early study in Berger and Krane (1985), all find small but positive estimates of β in equation (2.3). These results are all interpreted as indications of forecast smoothing, meaning that forecast revisions in one direction are most likely followed by revisions in the same direction.

In summary, we observe from the literature that the estimates of β in equation (2.3) tend to range from -0.5 to 0.5 and, in a significant number of cases the null hypothesis that $\beta = 0$ is rejected. Given the results in Franses et al. (2009) and Chang et al. (2011) regarding the use of biased OLS standard errors in many empirical analyses of forecasts and forecast updates, the frequency of rejecting the null hypothesis is likely to be biased upward.

In Figure 2.1 we report our own estimates of β for the pooled Consensus Forecasts for US real GDP growth in the period 1990 to 2010. We also include the 95% confidence bounds. We see from this graph that around 1995-1998 and 2005-2007, the estimate of β is significantly negative, while in 2001-2002 and 2009 it was significantly positive. We will return to these data below when we illustrate our proposed methodology.

Figure 2.1 Estimates of the autoregression parameter of forecast updates in Concensus forecasts for US GDP growth, 1990-2010.



2.3 Interpreting the Empirical Findings

Despite the wealth of empirical evidence on patterns in forecast revisions, to date there would seem to be no studies that have formally analyzed the meanings of positive and negative estimates of β in equation (2.3). If $\beta > 0$, some kind of smoothing process may exist, but what type of process might this be? Moreover, what does this smoothing process look like? It is the purpose of this section to propose a formal methodology to derive how specific estimates could arise, where we explicitly take into account that a numerical forecast is the concerted effort of an econometric model and an expert individual's intuition.

We begin by introducing some notation, and then derive the first-order autocorrelation of $F_{t|t-h} - F_{t|t-(h+1)}$, which is associated with β in equation (2.3). Finally, we consider several special cases that can be used to explain the observed estimates given in Table 2.1.

Preliminaries

As stated above, the basic assumption for our methodology is that

$$F_{t|t-h} = M_{t|t-h} + \nu_{t|t-h} \tag{2.5}$$

which states that a given numerical forecast is the sum of a model forecast, $M_{t|t-h}$, and of expert intuition, $\nu_{t|t-h}$. For illustrative purposes, we focus on

$$F_{t|t-1} = M_{t|t-1} + \nu_{t|t-1}$$
$$F_{t|t-2} = M_{t|t-2} + \nu_{t|t-2}$$
$$F_{t|t-3} = M_{t|t-3} + \nu_{t|t-3}$$

We use the familiar Wold decomposition of a stationary time series of interest (namely, the numerical forecasts of key economic fundamentals such as real GDP growth, inflation rate, and unemployment rate), y_t, that is:

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3} + \dots \tag{2.6}$$

where $\varepsilon_t \sim (0, \sigma^2)$ is an uncorrelated error process. This error process can be interpreted as a news process (as will be seen below). The parameters, θ_k, $k = 1, 2, 3, \ldots$, are such that the time series is stationary and invertible.

Given (2.6), the econometric time series model forecasts can be written as

$$\begin{aligned} M_{t|t-1} &= \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots \\ M_{t|t-2} &= \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} + \ldots \\ M_{t|t-3} &= \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} + \theta_5 \varepsilon_{t-5} + \ldots \end{aligned}$$

The two subsequent forecast updates are given as

$$\begin{aligned} F_{t|t-1} - F_{t|t-2} &= M_{t|t-1} - M_{t|t-2} + \nu_{t|t-1} - \nu_{t|t-2} \\ &= \theta_1 \varepsilon_{t-1} + \nu_{t|t-1} - \nu_{t|t-2} \end{aligned} \tag{2.7}$$

and

$$\begin{aligned} F_{t|t-2} - F_{t|t-3} &= M_{t|t-2} - M_{t|t-3} + \nu_{t|t-2} - \nu_{t|t-3} \\ &= \theta_2 \varepsilon_{t-2} + \nu_{t|t-2} - \nu_{t|t-3} \end{aligned} \tag{2.8}$$

Note that when

$$F_{t|t-h} = \pi M_{t|t-h} + \nu_{t|t-h} \tag{2.9}$$

with $0 < \pi < 1$, which is the case where the model outcome is only partially taken into account, then similar results will appear as above, as the θ parameters will then be scaled by π.

Correlations

In this subsection we assume that $h = 1$ (and we postpone analysis of other values of h to future research), and that we have data for various numerical forecast events t. In order to derive the correlation between the forecast revision in equation (2.7), that is, the left-hand side variable in equation (2.3), and the variable on the right-hand side in (2.3), as given in equation (2.8), we define the following variances and covariances:

$$
\begin{aligned}
\gamma_0 &= \text{Var}(\nu_{t|t-i}) \\
\gamma_1 &= \text{Covar}(\nu_{t|t-i}, \nu_{t|t-(i+1)}) \\
\gamma_2 &= \text{Covar}(\nu_{t|t-i}, \nu_{t|t-(i+2)}) \\
\omega_0 &= \text{Covar}(\varepsilon_{t-i}, \nu_{t|t-i}) \\
\omega_1 &= \text{Covar}(\varepsilon_{t-(i+1)}, \nu_{t|t-i})
\end{aligned} \tag{2.10}
$$

The first three terms deal with the time series properties of random expert intuition. The last two terms deal with the potential non-zero correlations between current news and current intuition (namely, how intuition might react contemporaneously to news in the numerical forecast), and between one-period lagged news and current intuition (namely, how intuition might react with a one-period lag to news in the numerical forecast). Note that the premise behind forecast smoothing, as it is presented in the literature, is that current news is discarded to some extent, which means that $\omega_0 < 0$.

More precisely, the following definitions will be used to interpret smoothing and over-reaction in a clear and meaningful manner:

- Definition 1a: Contemporaneous Smoothing of intuition to news occurs when $\omega_0 < 0$
- Definition 1b: Dynamic Smoothing of intuition to news occurs when $\omega_1 < 0$
- Definition 2a: Contemporaneous Over-reaction of intuition to news occurs when $\omega_0 > 0$
- Definition 2b: Dynamic Over-reaction of intuition to news occurs when $\omega_1 > 0$

In light of these definitions, it is possible that $\omega_0 < 0$ and $\omega_1 > 0$, or $\omega_0 > 0$ and $\omega_1 < 0$, so that there can be a switch from smoothing to over-reaction, and vice-versa, over time. This possibility does not seem to have been investigated in the literature.

Given the above terms and definitions, we can proceed to show that the variance of $F_{t|t-2} - F_{t|t-3}$ is equal to

$$
\mathbb{E}[(\theta_2\varepsilon_{t-2} + \nu_{t|t-2} - \nu_{t|t-3})(\theta_2\varepsilon_{t-2} + \nu_{t|t-2} - \nu_{t|t-3})] = \theta_2^2\sigma^2 + 2\theta_2\omega_0 + 2\gamma_0 - 2\gamma_1 \tag{2.11}
$$

The covariance between $F_{t|t-1} - F_{t|t-2}$ and $F_{t|t-2} - F_{t|t-3}$ is equal to

$$\mathbb{E}[(\theta_1\varepsilon_{t-1} + \nu_{t|t-1} - \nu_{t|t-2})(\theta_2\varepsilon_{t-2} + \nu_{t|t-2} - \nu_{t|t-3})] = \theta_2^2\omega_1 - \theta_2\omega_0 - \gamma_0 + 2\gamma_1 - \gamma_2 \quad (2.12)$$

Hence, the parameter arising from equation (2.3) for $h = 1$ is given by

$$\beta = \frac{\theta_2\omega_1 - \theta_2\omega_0 - \gamma_0 + 2\gamma_1 - \gamma_2}{\theta_2^2\sigma^2 + 2\theta_2\omega_0 + 2\gamma_0 - 2\gamma_1} \quad (2.13)$$

This expression is the basis for an analysis of alternative special cases below, which serve illustrative purposes.

Special cases

There are several special cases that are worth highlighting, as follows:

Econometric model only

Case (i): $F_{t|t-h} = M_{t|t-h}$

In this case, where the final forecast is just the model forecast with no intuition, such that γ_0, γ_1, γ_2, ω_0 and ω_1 are all equal to 0, it is clear that

$$\mathbb{E}[(\theta_1\varepsilon_{t-1})(\theta_2\varepsilon_{t-2})] = 0 \quad (2.14)$$

so that $\beta = 0$ in (2.3). This is the classic case of weak-form forecast rationality. We will now show that only in this special case of the null hypothesis $\beta = 0$, does a value of β have a straightforward and valid interpretation.

Intuition only

Case (ii): $F_{t|t-h} = \nu_{t|t-h}$

In this case, the final forecast is based only on intuition and no model. Therefore, the forecaster does not consider the use of an econometric model, and also does not have any insights into the news process, ε_t, which means that ω_0 and ω_1 are equal to 0. In this case, the parameter in (2.13) becomes

$$\beta = \frac{-\gamma_0 + 2\gamma_1 - \gamma_2}{2\gamma_0 - 2\gamma_1} \tag{2.15}$$

which, in turn, after dividing the numerator and denominator by γ_0, can be written as

$$\beta = \frac{-1 + 2\rho_1 - \rho_2}{2 - 2\rho_1} \tag{2.16}$$

where the ρ parameters are the one-period and two-period lagged autocorrelations for the intuition process. For illustrative purposes, we consider two alternative processes for intuition, namely an autoregressive (AR) process and a moving average (MA) process:

Process (a): When intuition follows an AR(1) process, with parameter λ, then $\rho_1 = \lambda$ and $\rho_2 = \lambda^2$. Note that intuition can have non-zero mean, which does not matter for these derivations. Substituting these two terms into equation (2.16) gives

$$\beta = \frac{\lambda - 1}{2} \tag{2.17}$$

Clearly, when intuition is a stationary AR(1) process, that is, when $|\lambda| < 1$, then $-1 < \beta < 0$. If intuition is simply a white noise process, then the estimate of β equals -0.5. For an AR(2) process for intuition similar results can be derived.

Process (b): When intuition follows an MA(1) process, with parameter θ, then $\rho_1 = \frac{\theta}{1+\theta^2}$ and $\rho_2 = 0$. Substituting these terms into equation (2.16) gives

$$\beta = \frac{-(\theta - 1)^2}{2(\theta^2 - \theta + 1)} \tag{2.18}$$

In Figure 2.2, we present the parameter, β, as a function of θ. Again, it is clear that β is negative unless $\theta = 1$.

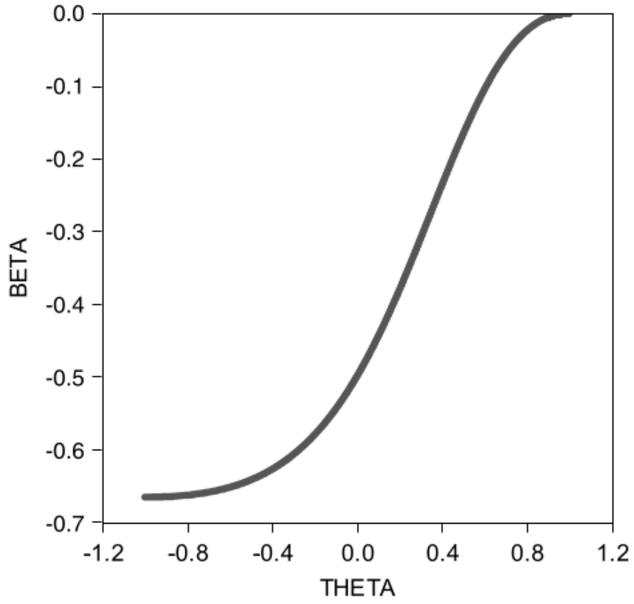
Forecasts based on model and intuition

Case (iii): $F_{t|t-h} = M_{t|t-h} + \nu_{t|t-h}$ with $\omega_0 = 0$ and $\omega_1 = 0$

In this case, where there is no correlation between current and past news and current intuition, the expression for β is

$$\beta = \frac{\gamma_0 + 2\gamma_1 - \gamma_2}{\theta_2^2\sigma^2 + 2\gamma_0 - 2\gamma_1} \tag{2.19}$$

Figure 2.2 Relationship between the parameters θ and β for an MA(1) process underlying intuition.



When a time series process is postulated for intuition, it is easy to see using (2.16) that the value of β is also negative. This is an interesting result as it shows that NO smoothing (Definitions 1a and 1b) and also NO over-reacting (Definitions 2a and 2b) can generate a negative value of β. This would seem to cast serious doubt on the prevailing consensus in the literature regarding the interpretation of β, except when $\beta = 0$.

In the following case where current intuition is correlated with current news, that is

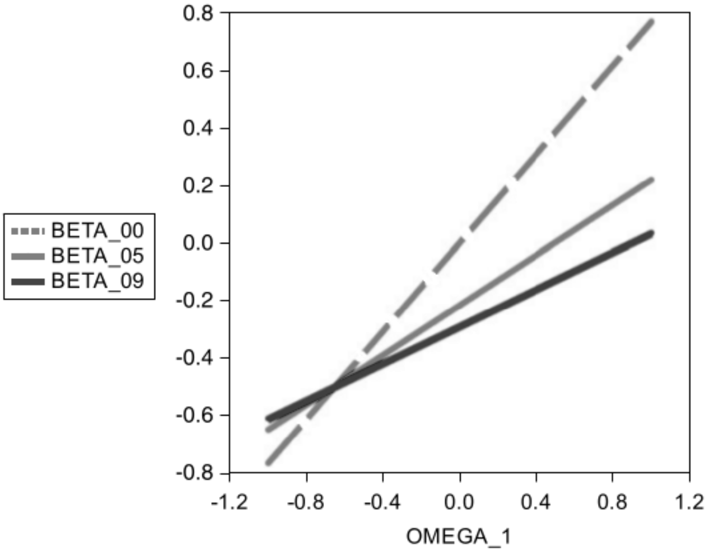
Case (iv): $\quad F_{t|t-h} = M_{t|t-h} + \nu_{t|t-h}$ with $\omega_0 \neq 0$ and $\omega_1 = 0$,

the parameter in (2.13) becomes

$$\beta = \frac{-\theta_2\omega_0 - \gamma_0 + 2\gamma_1 - \gamma_2}{\theta_2^2\sigma^2 + 2\theta_2\omega_0 + 2\gamma_0 - 2\gamma_1} \tag{2.20}$$

A typical macroeconomic variable would show positive autocorrelation, certainly for the first few of these so that, in practice, $\theta_2 > 0$. In this case, for β to become positive, ω_0 should be large and negative. Thus, contemporaneous smoothing of intuition to news (Definition 1a) can lead to a positive value of β, but this interpretation of smoothing is not dependent on the sign of β.

Figure 2.3 Values of β when $\gamma_0 = 1$, $\gamma_1 = -0.8$, $\gamma_2 = 0.6$ (mimicking an AR(2) process), $\sigma^2 = 1$, $\omega_0 = 0$ ($BETA\_00$), 0.5 ($BETA\_05$) or 0.9 ($BETA\_09$), and ω_1 ranges from -1 to 1.



Case (v): $\quad F_{t|t-h} = M_{t|t-h} + \nu_{t|t-h}$ with $\omega_0 \neq 0$ and $\omega_1 \neq 0$

In order to give an impression of which values of β can emerge in practically and relevant cases, consider Figure 2.3. There we depict values of β for the case where $\gamma_0 = 1$, $\gamma_1 = -0.8$, $\gamma_2 = 0.6$ (mimicking an AR(2) process), $\sigma^2 = 1$, $\omega_0 = \{0, 0.5, 0.9\}$, and where ω_1 ranges from -1 to 1. Clearly, when ω_0 and ω_1 are both associated with "smoothing" (whereby they are both negative), the value of β can still be negative. In short, this is a case where there is both contemporaneous and dynamic smoothing, but the literature would typically interpret a negative value of β as over-reaction. Figure 2.3 shows that any value of β is possible for virtually any positive or negative values of ω_0 and ω_1.

In summary, when there is no correlation between current and lagged news and current intuition (Case (iii) above), then $\beta < 0$. When there is a negative correlation between current news and current intuition, and when there is a positive correlation between past news and current intuition, then $\beta > 0$. In the event that $\beta = 0$ this can be associated with the situation (a) where the forecaster relies fully on an econometric model and also (b) where the forecaster relies fully on intuition, and where the time series properties of intuition are a random walk (that is, $\lambda = 1$ in equation (2.17)). In contrast, when only intuition is used and intuition is a white noise process (that is, $\lambda = 0$ in equation (2.17)), then $\beta = -0.5$.

Interestingly, and most importantly, the above derivations and definitions show that the estimated value of β is not directly associated with smoothing or over-reaction, but rather depends heavily on the signs and values of both ω_0 and ω_1.

2.4 Interpreting Table 2.1

Using the results in the previous section, we can now evaluate the empirical results given in Table 2.1. It seems that theoretically the values of β can range from around -1 to $+\infty$, with values in the range -0.5 to slightly greater than zero seem to be the most likely.

A value for β of -0.5 would mean that the forecaster may have discarded the outcome of the model, and has used expert intuition, with the peculiar property that there is zero correlation between $\nu_{t|t-h}$ and $\nu_{t|t-(h+1)}$. This absence of correlation seems quite unusual, as the intuition-based forecasts are concerned with the same fixed event.

Dovern and Weisser (2011, p. 463) interpret a negative value of β as a sign of over-reaction, "i.e., at first, they (forecasters) revise their forecasts too much, then they undo part of this revision during the next forecasting round". Hence, they assume that $\omega_0 > 0$ and $\omega_1 < 0$. The results in the previous section show that there can also be several other situations that lead to negative values of β, specifically the covariance of current and lagged news to the numerical forecasts with current intuition.

A large and positive value of β must mean that forecasters take current and one-period lagged news into account when forming their intuition. A negative correlation between current news and current intuition ($\omega_0 < 0$) means that a forecaster downplays the relevance of current news, that is, there is under-reaction. This could be associated with a forecaster's uncertainty with the most recent releases of data. A positive correlation between one-period lagged news and current intuition ($\omega_1 > 0$) suggests that the forecaster amplifies a recent shock, which might not be there, and hence over-adjusts the model forecast. In the literature, these situations are all presented under the label of "forecast smoothing".

The results in the previous section suggest that, based only on estimates of β, these separate cases cannot be disentangled, which leads to the key issue of identification. Various parameter configurations of γ_0, γ_1, γ_2, ω_0 and ω_1, and especially of ω_0 and ω_1, can lead to various values of positive and negative β.

By far, the optimal value of β is 0. This could mean either that the forecaster has relied fully on an econometric model, or that the forecast is given as

$$F_{t|t-h} = \nu_{t|t-h} \tag{2.21}$$

with $\nu_{t|t-h} = \nu_{t|t-(h+1)} + \zeta_{t,h}$ where $\zeta_{t,h} \sim (0, \eta^2)$ is a white noise process.

What is certain, though, is that, when there is no correlation between news and intuition, it follows that β is negative. For β to be positive a forecaster should under-react to current news and over-react to past news. The latter case seems to occur most frequently in practice (see Table 2.1).

In order to derive what forecasters actually do from the data on numerical values of $F_{t|t-h}$, it is necessary to obtain estimates of the news process and of intuition. This requires fitting an econometric time series model for y_t, the numerical forecast of interest, to obtain estimates of ε_t. Next, this model can be used to create estimates of the model-based forecasts, $M_{t|t-h}$ and, with these, one can estimate a time series model with observations on intuition, $\nu_{t|t-h}$. These two estimated series could then be used to compute the correlations between current intuition and both current and past news. As such, one can obtain estimates of the key parameters, γ_0, γ_1, γ_2, ω_0 and ω_1, and then sensibly interpret the value of the estimated β. As the variables are generated regressors, Franses et al. (2009) recommend using Newey-West HAC standard errors to correct for the measurement errors in the estimated variables.

As an illustration, we return to our own example (Figure 2.1) on the Consensus Forecasts, we do the following. We fit an AR(1) model to the yearly GDP growth figures for the US in the period 1977-2011. With the estimates from this regression we construct one-year and two-year-ahead model forecasts for GDP growth. Also, the residuals of this regression are used as estimates for ε_t. Next to these model forecasts, we also have the Consensus Forecasts for the period 1991-2012. Several forecasters have produced forecasts for current year and next year's GDP on a monthly basis, and we aggregate these into a single forecast. To fairly compare these to the one-year and two-year-ahead forecasts, we only use the January forecasts. Next, intuition is computed as the difference between the model and Consensus forecasts. With these, we compute γ_0 and γ_1. As the forecasters only predict for next year, we do not have γ_2, and hence in what follows next, we set this parameter equal to 0. Finally, we compute ω_0 and ω_1 as the covariance between current and lagged news and intuition. Finally, using the Wold composition we compute the value of θ_2. The numerical results for this case are that γ_0 is 1.544 and γ_1 is 0.312, that ω_0 is 0.161 and ω_1 -0.495 (which is associated with current smoothing and dynamic over-reaction), With a σ^2 estimated as 4.210 and an AR(1) parameter in the model with 0.305, we obtain using (2.13) an estimate of β equal to -0.388, which indeed is approximately the average value as displayed in Figure 2.1.

2.5 Conclusion

This chapter has shown that the interpretation of β in a regression of forecast revisions on previous forecast revisions is not entirely straightforward. Currently, the literature unequivocally assigns meanings such as smoothing, and over-reaction or under-reaction, to positive and negative values of β, but we have shown in this chapter that these are not one-to-one relationships.

The approach developed in the chapter is based on the assumption that numerical forecasts could be decomposed into both an econometric model and random expert intuition. We proposed a methodology to be able to interpret such non-zero correlations in a straightforward and clear manner. In particular, we showed that the interpretation of the sign of the correlation between the current and one-period lagged forecast revisions depends on the process governing intuition, and the current and lagged correlations between intuition and news (or shocks to the numerical forecasts). It follows that the estimated non-zero correlation cannot be given a direct interpretation in terms of smoothing or over-reaction. It was also shown that smoothing and over-reaction, modelled and interpreted correctly, can change over time.

When estimates of γ_0, γ_1, γ_2, ω_0 and ω_1 are available, it also seems possible to examine the validity of other reasons for forecast updates not to be weak-form efficient, or rational. Recent work in Ashiya (2003), Amir and Ganzach (1998), and DellaVigna (2009) sketch various reasons for non-rationality. It would be interesting to examine whether professional forecasters have certain forecasting styles. We postpone such an extensive analysis for future research. Then it would be relevant to compare the behavior with the actual performance of the forecasters. Indeed, as Franses and Legerstee (2010) have shown, in order to evaluate forecast accuracy properly, one needs to know how the forecasts were actually created.

Chapter 3

Heterogeneous forecast adjustment and the provision of feedback

Based on de Bruijn and Franses (2012). Authors contributed following a 75% / 25% split.

Abstract

There is ample empirical evidence that expert-adjusted model forecasts can be improved. One way to potential improvement concerns providing various forms of feedback to the sales forecasters. It is also often recognized that the experts (forecasters) might not constitute a homogeneous group. This chapter provides a data-based methodology to discern latent clusters of forecasters, and applies it to a fully new large database with data on expert-adjusted forecasts, model forecasts and realizations. For the data at hand, two clusters can clearly be identified. Next, the consequences for providing subsequent feedback are discussed.

3.1 Introduction

Sales forecasters often rely on the output of a forecast support system (FSS) when they create their own forecasts. A typical situation is that an FSS delivers forecasts and that an individual forecaster modifies these as he or she sees fit. Oftentimes, an FSS includes forecast algorithms which only include recent past sales data, while the forecaster with domain-specific knowledge may believe that additional information can be useful and thus delivers an adjusted forecast, see Goodwin (2000, 2002). Due to the recent availability of novel databases including FSS based forecasts, managers' forecasts and actual realizations, more insights are gained as to how these adjusted forecasts perform, how they are actually constructed and how they can be improved. Fildes et al. (2009) marked the start of this large databases-based research and Franses (2014) recently summarizes various findings across a range of studies.

The main stylized facts concerning manually adjusted FSS forecasts seem to be that (1) FSS forecasts are almost invariably adjusted by individual forecasters, that (2) expert-adjusted forecasts are often not as accurate as the FSS forecasts and that (3) there are various ways to improve current approaches of adjusting FSS forecasts. One way to achieve improved adjusted forecasts is to provide feedback to the forecasters on their actual behavior and on their past track record. Legerstee and Franses (2014) analyze the behavior of forecasters who deliver forecasts for monthly sales data, where they compare data before and after the moment that experts received different kinds of feedback on their behavior and their task. They conclude that after feedback the adjusted forecasts deviated less from the FSS forecasts and that their accuracy had improved substantially.

At the same time, various studies suggest that individuals who manually adjust model-based forecasts do not constitute a homogeneous group. Boulaksil and Franses (2009) interviewed various forecasters of whom some say that they never look at FSS forecasts when creating their own, and others say that they deviate only a little from the FSS forecasts. In the earnings forecasting literature, there also appear to be one or more different types of behavior of forecasters, depending on whether they want to deliver accurate forecasts or want to stand out with exceptional quotes, see Jegadeesh and Kim (2010) and Clement and Tse (2005). Similarly, in the macroeconomics literature there are also examples of differing behavior across forecasters, see for example Lamont (2002).

When there are different types of forecasters who behave differently when adjusting FSS forecasts, then feedback to these different types of forecasters most likely should also be different, and this is the key premise in this chapter. Hence, to improve the overall quality of adjusted forecasts, one may wish to discern groups of individual forecasters with common behavior within the group and differing behavior across the groups. As records of the adjustment process typically do not exist (Franses, 2014), the division of forecasters into various groups must be done using the available data. In this chapter we therefore propose a method to disentangle groups of forecasters based on their actual behavior. Knowing the clusters can lead to more tailor-made feedback, and we recommend such variants of feedback in our case study. The database in our case study has never been analyzed before and concerns the FSS forecasts, adjusted forecasts and realizations of sales of SKUs (stock keeping units) of a very large Germany-based pharmaceutical company.

The outline of this chapter is as follows. In Section 3.2 we discuss the database. In Section 3.3, we provide our methodology to link the behavior of sales forecasters with their forecast performance, while allowing for latent classes of individual forecasters with common behavior. In Section 3.4 we discuss the main results for our new database. In Section 3.5, we provide some insights on how our results can be used to provide feedback to the forecasters so that they can improve their performance. Finally, Section 3.6 concludes.

3.2 Data

The data set that we use is provided by a globally operating Germany-based pharmaceutical company. Country-specific managers produce sales forecasts for a set of products, and this set is different per country. This means that there is only a single forecaster for each product in a certain country. The dataset also contains the FSS forecasts, the manual adjustments (which are the differences between the managers' and the FSS forecasts) and the actuals. Each pharmaceutical product can be assigned to a specific product category. The FSS forecasts are based on lagged sales data and thus do not account for exceptional events or any other explanatory variables.

The full dataset concerns 11432 products with 29 monthly 3-months ahead forecasts for the period 2009-2012 (May). For many products there are only forecasts for a few of the months in the sample. Next to this, the managers are not always very precise in their reporting behavior. For example, sometimes the actuals and the forecasts are not of the same magnitude, meaning that for example an FSS forecast is reported in thousands of units, while the actual is reported in millions of units. For some cases it is clear how to bring them to the same order of magnitude, for others it is not, and these latter cases are dismissed. We have filtered the products such that only those products for which data in more than half of the sample period is in good condition are selected. All three relevant series (managers' forecasts, FSS forecasts and actuals) were required each to meet this criterion.

Next, we calculate for each product/country combination the median percentage error (MPE) as compared to the median realization of sales of the corresponding product in the corresponding country. We use the median in order to robustify the accuracy measures against unnoticed badly-reported forecasts. We use the percentage error (instead of the error) to make products and countries of different sizes comparable. This is important for our latent class model below. We also calculate the median absolute percentage error, the median percentage adjustment and the median absolute percentage adjustment.

After data cleaning, we have data for 2472 products across 67 forecasters. The average number of products per forecaster is 36.90. The distribution is depicted in Figure 3.1. Clearly, many forecasters deal with less than 40 products, although some are responsible for more than 150 products[1]. The empirical distribution of the median absolute percentage error is heavily skewed to the right. To ensure that a few large observations do not dominate the final results, we will analyze the natural logarithm of the median absolute percentage error, of which the distribution is shown in Figure 3.2. More details on this database are discussed below where we deal with aspects of our methodology.

[1]As a referee rightfully pointed out, it could be that the number of products assigned to a forecaster is not random and related to a forecaster's past performance. Unfortunately, we have no information as to whether such correlation exists. The differences are mainly caused by the size of the country and the related portfolio.

Figure 3.1 The distribution of the
number of products per forecaster. The y-axis gives the frequency and the x-axis gives the number of products.

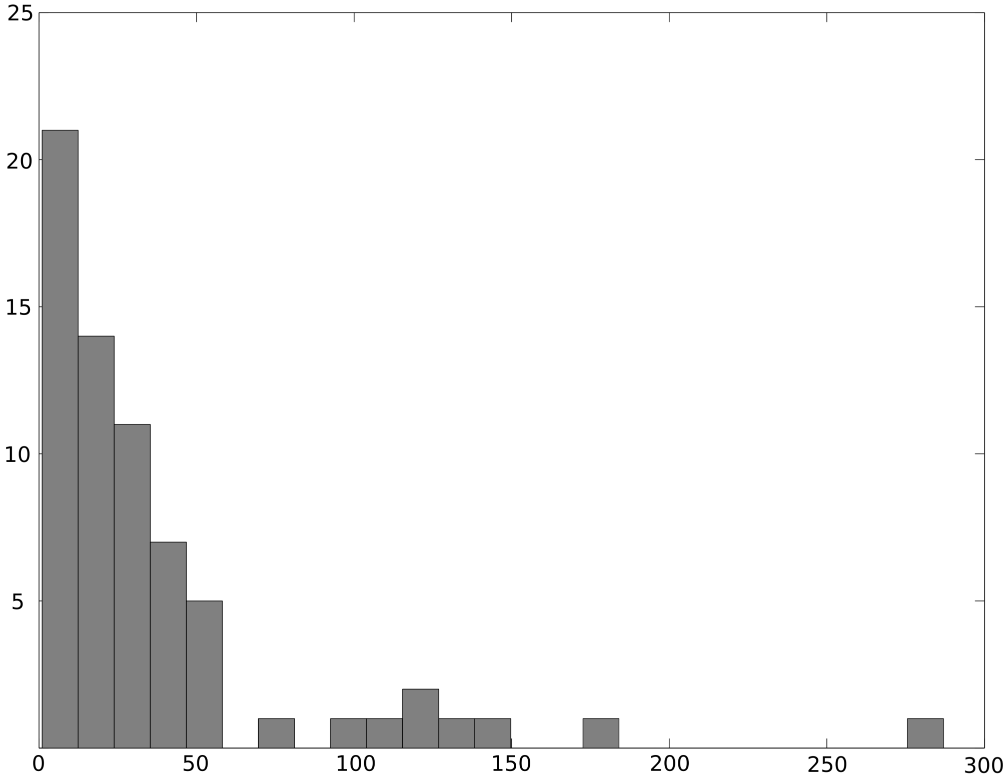
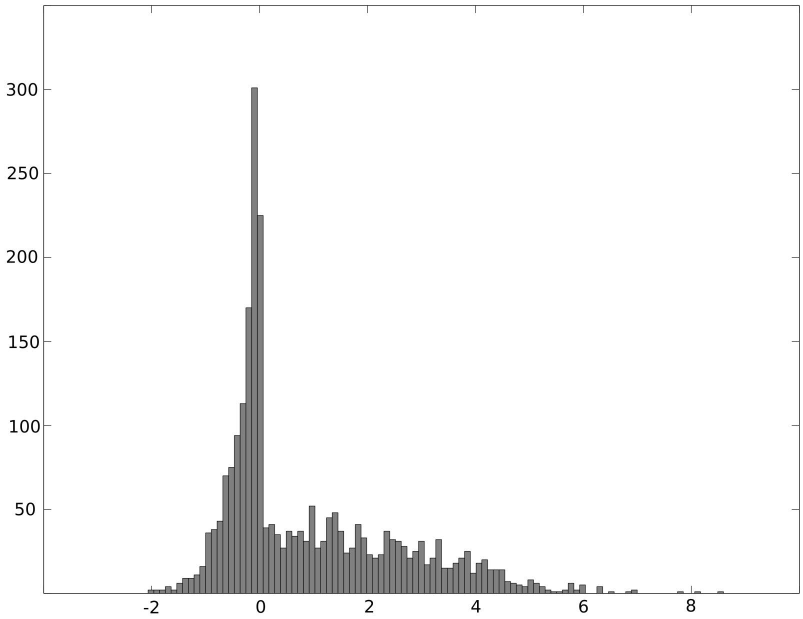


Figure 3.2 The distribution of the logarithm of the median absolute percentage error.



3.3 Methodology

We use a three-level econometric model to discern potential clusters of forecasters with similar behavior. Our model follows the tradition in latent class modeling as outlined in McLachlan and Peel (2000) and Wedel and Kamakura (2000). The variable to be explained (y) is forecast performance. The independent variables for this performance are those that concern the adjustment behavior of the forecasters (collected in X). In a separate level of the model, the parameters that link the performance with X are made a function of moderating variables (Z), like the number of products. Next, we allow that the parameters of the moderating variables associate with S latent classes. For each forecaster we thus estimate a probability of membership of one of the classes.

The model

The equations of our model are

$$y_{p,i} = \beta_{p,i} X_{p,i} + \varepsilon_{p,i} \tag{3.1}$$

$$\beta_{p,i} = \gamma_i Z_{p,i} \tag{3.2}$$

$$\gamma_i = \sum_{s=1}^{S} \psi_s \mathbb{P}[Type_i = s] \tag{3.3}$$

with $p \in [1, ..., P]$ indicating a product, $i \in [1, ..., N]$ indicating a forecaster responsible for in total P_i products and $s \in [1, ..., S]$ indicating potential clusters of forecasters with similar behavior. For our data, the dependent variable in (3.1) is the natural logarithm of the median of the absolute percentage error per product (logMedAbsPercErr). Apart from the intercept, the two variables in X are the natural logarithm of the median of the absolute percentage adjustment + 0.01 (logMedAbsPercAdj) and the median of the percentage adjustment (medPercAdj). Apart from the intercept, the three variables in Z are the natural logarithm of the amount of products the forecaster has been assigned to (logNrProd), the natural logarithm of the number of products in the same category as the respective product and also assigned to the same forecaster (logNrProdCat), and the autocorrelation in the errors of the FSS forecasts (corrErrModel). The latter variable appears prominent in recent studies, see for example Franses and Legerstee (2009). When the parameters in (3.1) are positive, then a larger-sized adjustment occurs simultaneously with larger forecast errors, and even more so for larger upward adjustments. The parameters in (3.2) can amplify or dampen the effects of the variables in X. The third level of the model (3.3) allows for S latent classes for the parameters in (3.2).

Estimation and inference

For the parameter estimation of (3.1)-(3.3), we use an Expectation-Maximization algorithm (EM-algorithm) (Dempster et al., 1977). In such an algorithm there is an Expectation step (E) which concerns the expectation of a set of unobserved variables, given current estimates of the parameters, and a Maximization step (M) in which the likelihood function of the parameters is maximized, given current estimates of the unobserved variables. These steps are repeated until convergence. In our case, the cluster probabilities are the unobserved variables, while the parameters consist of ψ_s together with the unconditional probabilities of belonging to one cluster, for which we use the notation P_s. Using these, estimates for γ_i and $\beta_{p,i}$ can directly be calculated. In the discussion below, y_i is a vector consisting of the various values of $y_{p,i}$ for all p, and similar notation is used for X_i and Z_i.

In the E step the expectation of the individual-group probabilities $\mathbb{P}(Cluster_i = s)$ is taken, given estimates of ψ_s and the unconditional probability P_s for $s = 1, \ldots, S$. This expectation is calculated by comparing for all s the individual densities $f_{i,s}(y_i; X_i, Z_i, \psi_s)$ in case forecaster i would be fully assigned to cluster s, while also incorporating the unconditional probability P_s. Then, the new estimate of $\mathbb{P}(Cluster_i = s)$ is the fraction

$$\frac{P_s f_{i,s}(y_i; X_i, Z_i, \psi_s)}{\sum_{r=1}^{S} P_r f_{i,r}(y_i; X_i, Z_i, \psi_r)}. \tag{3.4}$$

In the M step the likelihood function of the parameters ψ_s and P_s is maximized, given estimates of the individual-group probabilities $\mathbb{P}(Cluster_i = s)$. It can be shown that these two variables can be estimated separately. To estimate P_s one can simply take the average of $\mathbb{P}(Cluster_i = s)$ across all i.

For estimation of the ψ_s we can show that the model reduces to a standard regression. This can be seen by using the multilevel property of the model, that is

$$y_{p,i} = \beta_{p,i} X_{p,i} = \gamma_i Z_{p,i} X_{p,i} = \sum_{r=1}^{S} \psi_r \mathbb{P}(Cluster_i = r) Z_{p,i} X_{p,i}. \tag{3.5}$$

Define $X^*_{p,i,s} = \mathbb{P}(Cluster_i = s) Z_{p,i} X_{p,i}$. Then, the model reduces to the regression of $y_{p,i}$ on all S matrices $X^*_{p,i,s}$. This results in estimates of ψ_s, which can be used to construct estimates of γ_i and $\beta_{p,i}$.

As the EM-algorithm might converge to a local optimum, we use several starting points for each S that we consider. The first starting points are derived from the best likelihood for the case $S - 1$ as follows:

1. Select a cluster s from the previous $S - 1$ clusters.

2. Split this cluster into two clusters by randomly assigning different proportions of $\mathbb{P}(Cluster_i = s)$ to clusters s and S, which is the new cluster. In this step, we use different proportions of the original $\mathbb{P}(Cluster_i = s)$ for each forecaster i.

3. Start the EM-algorithm and run it until convergence.

As there are $S - 1$ clusters to split up in the first step, this results in $S - 1$ outcomes. We also use R other starting points, which are constructed by randomly drawing from a Dirichlet distribution with all S parameters equal to $\frac{1}{S}$ for each forecaster i. We have set R at 2500. Of these total $R + S - 1$ converged estimates, the best one is chosen using the likelihood.

Following the above discussion, we end up with estimates for each S that we consider. Of course, increasing S will increase the likelihood, as in the worst case a cluster can always be cut into two to reduce the ideosyncratic error a bit, even though the forecasters within the clusters actually belong to the same cluster. To choose the final S, one can use the AIC-3 criterion, which has been shown to perform well in a multilayered model (Andrews and Currim, 2003). If a smaller number of types is desired, one can use the BIC criterion. There can also be ad-hoc reasons for the number of clusters.

In the EM-algorithm, one can quickly run into numerical problems. For example, the density when a certain individual is categorized into a certain cluster might be so low that it is almost zero. To avoid numerical problems, natural logarithm formulations of the above are used. If for a forecaster during the process at a certain point all densities are equivalent to 0 (or the log-densities equal to $-\infty$), this forecaster is assigned to the different types using just the unconditional probabilities and then the estimation process is continued.

3.4 Results

We have estimated the model parameters in (3.1)-(3.3) for different values of S. The AIC-3 criterion suggests using 6 clusters, while the BIC recommends 3 clusters. A closer look at the case with 3 clusters indicates that there is one cluster with only a few forecasters (of the 67), and this also holds in the first case where 4 clusters are exceptionally small. Because of this, for this database we therefore decide to limit the number of cluster to $S = 2$.

As the estimated values of $\beta_{p,i}$ and $\gamma_{p,i}$ are directly dependent on the estimated values of ψ_s, we report the estimation results for ψ_s for each of the clusters in Table 3.1. We see that in

Table 3.1 The estimates of ψ_s with standard errors in parentheses. Boldface printed estimates are significant at the 5% significance level.

		Intercept	logMedAbsPercAdj	medPercAdj
Type 1	Intercept	**1.881** (0.165)	**1.108** (0.070)	**-0.145** (0.050)
	logNrProd	-0.083 (0.054)	**-0.179** (0.024)	**0.054** (0.016)
	logNrProdCat	-0.001 (0.054)	**0.087** (0.025)	-0.015 (0.009)
	corrErrModel	-0.178 (0.110)	**0.173** (0.050)	-0.048 (0.026)
Type 2	Intercept	**0.877** (0.133)	**0.504** (0.059)	**0.049** (0.024)
	logNrProd	**-0.097** (0.046)	-0.022 (0.021)	0.003 (0.009)
	logNrProdCat	**0.172** (0.049)	0.004 (0.022)	-0.003 (0.010)
	corrErrModel	0.178 (0.102)	0.082 (0.045)	0.034 (0.026)

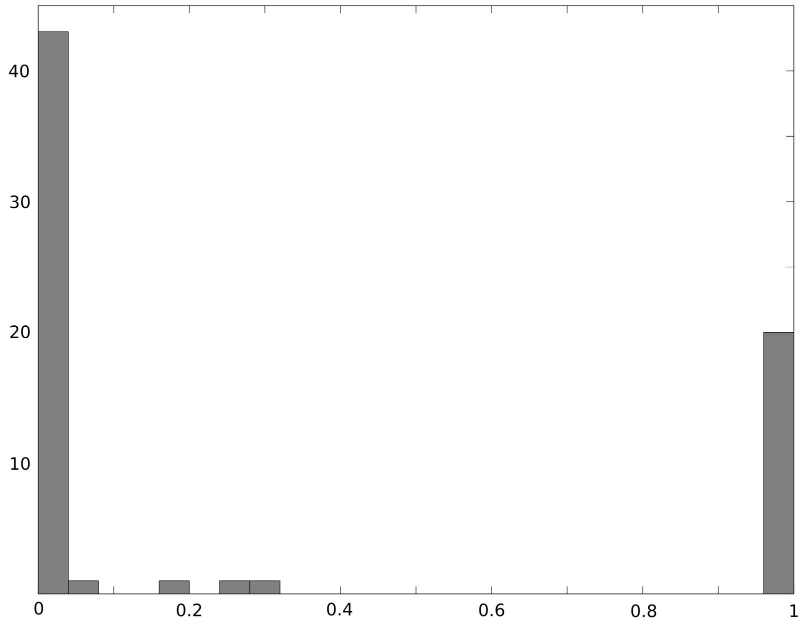
Table 3.2 The estimates of ψ_s while assuming that there is only one cluster (homogeneity). Standard errors in parentheses. Boldface printed estimates are significant at the 5% significance level

	Intercept	logMedAbsPercAdj	medPercAdj
Intercept	**1.218** (0.104)	**0.727** (0.045)	0.007 (0.021)
logNrProd	-0.051 (0.036)	**-0.043** (0.016)	-0.001 (0.007)
logNrProdCat	**0.091** (0.037)	-0.005 (0.016)	0.012 (0.006)
corrErrModel	0.108 (0.077)	**0.139** (0.034)	**-0.066** (0.017)

both clusters the size of the forecast error increases if the size of the adjustment increases. For forecasters in cluster 1 this increase is much larger (1.11 versus 0.50). Additionally, the effect of this variable for cluster 1 forecasters increases when the product is part of a larger category of products (0.09) and when the autocorrelation in the FSS forecasts increases (0.17), but it decreases if there is an increase in the total number of products assigned to the forecaster (-0.18). For the forecasters in cluster 2 these effects are not significant.

The forecast error increases for negative adjustments (compared to positive adjustments of the same size) for forecasters in cluster 1 (-0.15), while cluster 2 forecasters have a larger error in the case of positive adjustments (0.05). Also, the forecast error decreases for cluster 2 forecasters in the countries with more products (-0.10), but it increases if the product is part of a larger product category (0.17). In the next section, we will discuss in more detail what the managerial implications are of these estimates, but for now we can conclude that for our data there are two types of forecasters with clearly distinct behavioral characteristics. This can also be learned from the estimation results in Table 3.2, where we present the parameter

Figure 3.3 The distribution of the estimated probabilities of being a member of type 1 forecasters.



estimates in case we assume that all forecasters are homogenous. Due to averaging various significant effects seem to disappear.

Figure 3.3 shows the distribution of the estimated probabilities of a forecaster being a member of cluster 1. As can be seen, our model allows us to clearly categorize most of the forecasters into separate clusters. Note that such a distinction would be impossible by just looking at the graphs in Figure 3.1 and 3.2. There are only a few forecasters who have properties of both clusters. Figure 3.3 clearly shows that there is a distinction between the two types of forecasters, and also that these classes are substantially large. We have tried to explain the categorization of our multi-level approach using available explanatory variables (for example, using a binomial probit), and we have found no significant parameters. This again indicates that one needs a multi-level or mixture model such as ours to disentangle different classes of forecast behavior.

To examine if the forecasters get assigned different time series with different properties, we compute the first order autocorrelation and the standard deviation of all the time series. The average autocorrelations are 0.64 and 0.64 for the series that are dealt with by the fore-

Table 3.3 Several characteristics of both forecasters clusters. Standard deviations in parentheses.

	Cluster 1	Cluster 2
P_s, unconditional probability	0.312	0.688
Average number of products	52.23 (14.27)	29.95 (5.44)
Average of logMedAbsPercErr	1.162 (0.054)	0.736 (0.040)
Average of logMedAbsPercErr for FSS	0.969 (0.049)	0.630 (0.037)
Average absolute adjustment	22.50 (1.088)	31.74 (4.462)
Percentage upward adjustments	63.3 (0.003)	61.4 (0.002)

casters in clusters 1 and 2, respectively. The respective standard deviations are 36978 and 39166, and these numbers also do not differ much.

Table 3.3 shows several characteristics of the forecasters per cluster, which are weighted averages with as weights the cluster probabilities. Notice that both types of forecasters perform worse than the FSS forecasts, on average (which is consistent with earlier findings in the literature). There are twice as many forecasters in cluster 2, but those in cluster 1 are concerned with about twice as many products. Cluster 1 forecasters are more active forecasters, and may have more experience. Cluster 1 forecasters adjust more upwards (63.3 versus 61.4), which at first glance seems beneficial for them due to their negative sign of β_2 in Table 3.1 (-0.18). This effect may be countered by their number of products. Indeed, $29.95 \times 0.054 = 1.62 > 0.15$, their average parameter of X_2 is positive and higher than that of cluster 2. For X_1 the reverse holds true as the larger number of products makes the coefficient to decrease. Taking all factors into account simultaneously, the percentage error for cluster 1 forecasters is larger than the percentage error of cluster 2, and this may be due to their actions or due to a more difficult forecasting task.

3.5 Implication of the estimation results for feedback

The three types of feedback that are commonly provided to forecasters who can rely on FSS forecasts and may quote modified forecasts are outcome feedback, performance feedback and cognitive process feedback. Outcome feedback provides the forecaster with the realizations of the relevant variables, and it seems that this type of feedback is the least effective, see Goodwin and Fildes (1999) and Lawrence et al. (2006). Performance feedback provides the forecaster with information on past forecast accuracy. Remus et al. (1996) did

Table 3.4 The effect of one standard deviation (calculated per cluster) change in some characteristics on logMedAbsPercErr for the average forecaster of both types.

Variable	Change	Cluster 1 Outcome	% Effect	Change	Cluster 2 Outcome	% Effect
logMedAbsPercAdj	0.067	1.113	4.4 %	0.058	0.748	2.7 %
medPercAdj	0.731	1.092	2.3 %	0.129	0.730	0.9 %
logNrProd	0.029	1.071	0.2 %	0.025	0.719	-0.2 %
logNrProdCat	0.027	1.067	-0.2 %	0.025	0.725	0.4 %
corrErrModel	0.009	1.066	-0.3 %	0.008	0.722	0.1 %
Benchmark forecaster		1.069			0.721	

not find evidence of successful performance feedback, but Bolger and Önkal Atay (2004), Stone and Opel (2000) and Athanasopoulos and Hyndman (2011) did. Note that all these studies assumed homogeneous behavior of the forecasters, and this may perhaps explain the mixed results. Third, cognitive process feedback gives the forecaster information on what the forecaster does and how he or she reacts to changes in the context of in the information. At present the successfulness of this type of feedback is doubtful, see Remus et al. (1996), Balzer et al. (1992) and Lim et al. (2005). Finally, task properties feedback provides the forecaster with statistical information on the variable of interest, and it seems the most successful type of feedback, see Remus et al. (1996), Sanders (1992), Welch et al. (1998), Goodwin and Fildes (1999) and Lawrence et al. (2006).

In order to assign appropriate feedback to the forecasters in each of the clusters, the parameter estimates in Table 3.1 are not directly helpful. Therefore, we now address the issue of which changes in behavior and context of the forecasters could potentially improve their forecast track record. Table 3.4 contains some highlights of potential results.

For an average forecaster in cluster 1, the value for logMedAbsPercErr is 1.07, while for cluster 2 it is 0.72, see the bottom line of Table 3.4. This table also shows how much this criterion would change if certain characteristics would be adjusted by one standard deviation while the values of other characteristics are kept constant. As this effect is the effect on a log-measure, this can be interpreted as a percentage change. For example, a change of one standard deviation in the median absolute percentage adjustment (first line of Table 3.4), which means more deviation from the FSS forecast, increases the median absolute percentage error of the average cluster 1 forecaster 4.12% and for cluster 2 it is 3.74%. The first two lines of Table 3.4 show that the effects of the two $X_{p,i}$ variables have positive effects, although the effects can differ in size. Increasing the size of the adjustment increases the size of the error, and making more positive adjustments also increases the size of the error

In terms of feedback, both sets of forecasters should receive performance feedback that they adjust too much, and should do less. Also, task properties feedback indicating that there is no need to most often adjust upwards as the FSS generally would provide unbiased forecasts, can also help. Given that the clusters look similar in this dimension here there seems no need to diversify the types of feedback.

When we consider the first two lines of the second panel in Table 3.4, we see that the amount of products and categories do deliver opposite effects across clusters. For example, increasing the number of products for a forecaster with one standard deviation (while keeping the number of products per category equal, which effectively means introducing new categories to this forecaster and increasing the task) increases the absolute percentage error with 0.19% for cluster 1 forecasters, while cluster 2 forecasters will see a decrease of -0.28

As a fictitious case, if we were to assign (on average) 3.78 products less to the cluster 1 (more experienced) forecasts and 1.21 more product categories, then their performance improves with 0.3%. In contrast, if we assign 1.94 products more to cluster 2 forecasters, and 0.73 categories less, then performance for this group increases with 0.8%.

Most evidence of the relevance of diversifying feedback (here: task properties) is given in the last line of the second panel of Table 3.4. When the persistence in forecast errors of the FSS increases with one standard deviation, the forecasters in cluster 1 better do not change their behavior (as the forecast error decreases with 0.28%), while forecasters in cluster 2 should be recommended to act.

3.6 Conclusion

In this chapter we have proposed a methodology to discern clusters of forecasters who behave differently with varying performance success, which in turn can be used to appropriately provide feedback to the forecasters. Upon applying our new methodology to a new and large database with expert-adjusted forecasts, FSS forecasts and realizations for a pharmaceutical company, we could discern two clear disjoint clusters of forecasters, and we could also suggest more precise types of feedback for these clusters.

The downside of our study is that unfortunately we cannot test our recommended feedback on real people. We have no access to the individual forecasters of the company, and we also will not gain access in the future due to many job moves since we collected the data. Hence, in a sense our results could be interpreted as merely speculative. In our future work, we plan to run behavioral experiments in a laboratory.

Even though our study analyzed only a single large database, we believe that the main take away is that the assumption of a single homogeneous set of forecasters may be too strong and that this assumption also can obstruct the proper implementation of feedback. As the heterogeneity can only be estimated using the available data, our methodology seems to be practically relevant. It may also be that various findings in the literature of ineffective feedback may perhaps be due to the homogeneity assumption. Further empirical evidence is however needed.

Chapter 4

A novel approach to measuring consumer confidence

Based on de Bruijn, Segers, and Franses (2014). All authors contributed equally.

Abstract

This chapter puts forward a new data collection method to measure daily consumer confidence at the individual level. The data thus obtained allow to statistically analyze the dynamic correlation of such a consumer confidence indicator and to draw inference on transition rates. The latter is not possible for currently available monthly data collected by statistical agencies on the basis of repeated cross-sections. In an application to measuring Dutch consumer confidence, we show that the incremental information content in the novel indicator helps to better forecast consumption.

4.1 Introduction

Consumer Confidence Indicators (CCIs) are often regarded as useful variables to measure the current state of the economy as well as to forecast its future states at reasonably short horizons, see Ludvigson (2004) for an assessment. Most industrialized countries report such indicators at a monthly level. Typically consumer confidence is measured by surveying one thousand or more individuals each month. The individuals are asked whether they believe that their situation has improved in the previous period or will improve in the next period, while focusing on their financial situation, employment, and, for example, their purchases of durable and more expensive products in particular. The answer categories are (very) positive, neutral, and (very) negative, and their origin goes back to Katona (1951). The total indicator is constructed by subtracting the percentage of negative answers from the percentage of positive answers. Many countries also report more specific indicators, which are confined to the financial position or employment only. Publicly available data are published in original format as well as after seasonal adjustment.

Despite their widespread use and interpretation, it can be of interest to investigate if the way consumer confidence is measured can be improved. One research angle can concern the questions asked and the way indicators are constructed from these questions. One may for example consider replacing the traditional qualitative questions by probabilistic questions inquiring about more well-defined events, as suggested in Dominitz and Manski (2004). Also the fact that consumer confidence data show signs of seasonality can be viewed as inconvenient, and perhaps a rephrasing of the questions can overcome this potential drawback.

A second angle for potential improvement of consumer confidence indicators would be to better understand how consumer confidence varies across individuals with different socio-economic and demographic characteristics. These insights could be exploited to reduce sampling error due to the use of small and possibly unrepresentative samples in the data collection stage, which improves the reliability of the indicators. We believe that improvement in these two directions can be relevant, but the third research angle to be discussed next seems more promising.

A third angle is that consumer confidence data are usually so-called repeated cross-sectional data. That is, each month approximately one thousand individuals are interviewed, but each month this concerns one thousand *different* individuals. A major consequence of this way of collecting data is that developments over time are difficult to interpret. Basically, when an indicator is, say, -18 in December, while it was -21 in November, we must conclude that the average fraction of more negative answers in December was smaller than in November. We could even say that in December consumer confidence has increased with 3 points, but we must be aware that this does not concern the same individuals. Hence, an interpretation of a sequence of monthly consumer confidence levels is prone to the so-called ecological fallacy. This fallacy concerns the situation where we seek to derive micro behavior from aggregated data, and here it would mean that we think that the same individuals have changed their confidence. In the literature there are various suggestions to circumvent or solve this problem, see King (1997), Moffitt (1993), Sigelman (1991), and the collection of papers in King et al. (2004), among many others. In this chapter we seek to do that, but now by applying an alternative method of data collection.

In this chapter we put forward a method to collect high-frequency consumer confidence data at the individual level. We keep the Katona-type questions intact, but we merely focus on the collection and analysis of the data, trying to prevent ecological fallacy[1]. To that end, we collect data such that we have the same (though not all) individuals being interviewed from one week to another, to alleviate the problem that respondents from becoming annoyed or uninterested. In order to statistically analyze the dynamic correlation of our CCI and to draw inference on transition rates, we develop a Markov transition model. The model describes the persistence in an individual's confidence level. We exploit the Markov transition properties of the model to estimate an expected response for each individual on the days on which the individual did not complete the questionnaire. This enables us to compute a daily consumer confidence indicator as if the entire panel was surveyed every day. To illustrate the usefulness of our approach, we employ the indicator to forecast Dutch consumption. We show that the incremental information content in the novel indicator helps to improve forecasting accuracy.

The outline of the chapter is as follows. In Section 4.2 we present our method of data collection, and we argue that it has various convenient properties for the purpose of measuring consumer confidence. Next, in Section 4.3 we introduce the Markov transition model that will be used to describe longitudinal developments in consumer confidence at the individual level. In Section 4.4 we illustrate the usefulness of our method by forecasting Dutch

[1]The University of Michigan also uses a rotating design, and we will explain its differences from our method below.

household consumption using the novel consumer confidence indicator. In Section 4.5 we conclude with an outline of various areas for further research.

4.2 Methodology

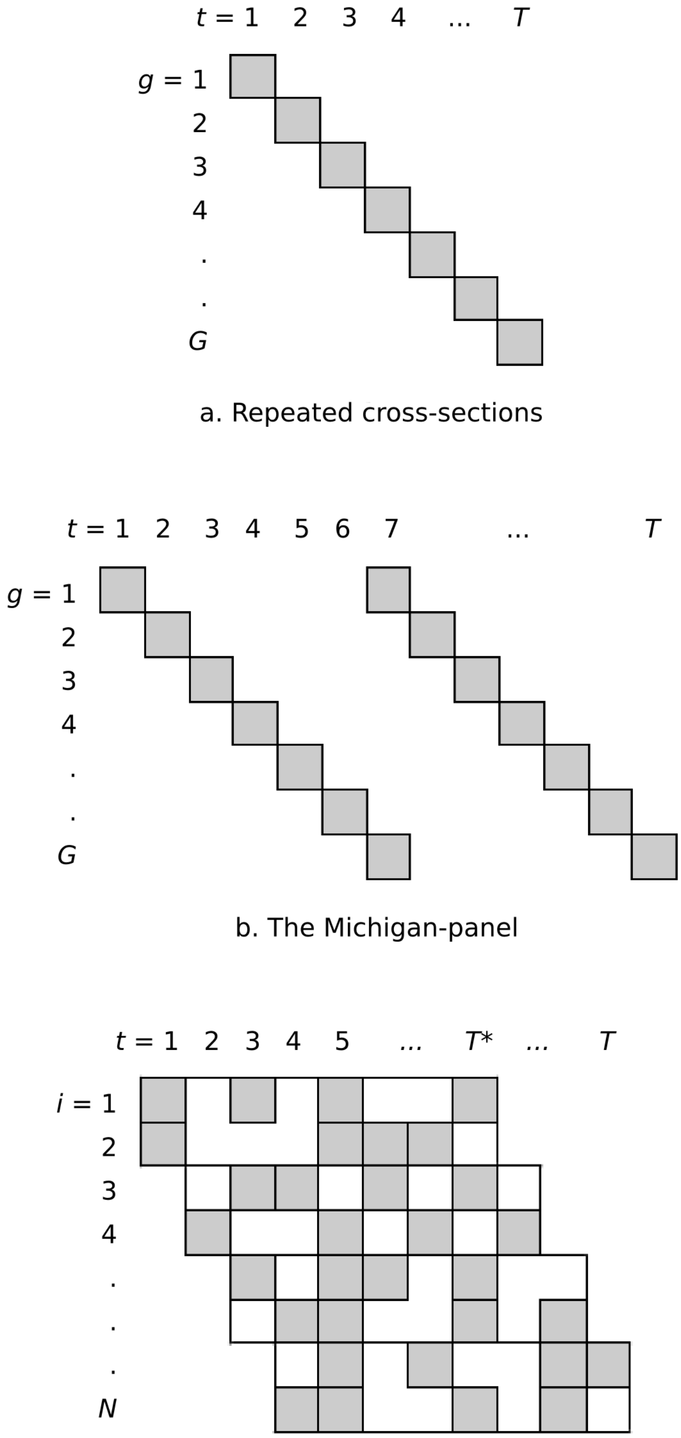
To measure developments in consumer confidence over time it is desirable to conduct a longitudinal or panel study where the same individuals are surveyed at multiple points in time. This allows us to study developments in confidence at the individual level. However, surveying the very same individuals frequently likely deteriorates the quality of the survey. People get irritated and they disconnect from the panel, thereby making the panel less efficient. Or perhaps worse, respondents' (reported) confidence levels may change due to being a member of a panel, which is called panel conditioning. Most statistical agencies therefore collect repeated cross-sections instead of panel data. This amounts to surveying a new group of individuals at each survey occasion, which implies that individuals are surveyed only once. The design is illustrated in Figure 4.1a. Here we index time by t, where $t = 1, \ldots, T$, individuals by i, where $i = 1, \ldots, N$ and groups of individuals by g, where $g = 1, \ldots, G$. A grey square in row g and column t indicates that group g is requested to be surveyed at time t. While repeated cross-sections reduce respondent burden and eliminate potential panel conditioning bias, developments at the individual level cannot be derived without making many assumptions, see the excellent treatment in Moffitt (1993). Therefore it seems promising to collect longitudinal data nevertheless, but to choose the design of the panel carefully such that the adverse effects of repeated interviewing are negligible or, at least, manageable.

To design a panel for a specific purpose, three decisions have to be made. Firstly, as individuals cannot be surveyed continuously, one has to decide on the total time-span that a panel member is requested to join the panel, to be denoted by T^*. To keep the total number of panel members constant, one may decide to invite new individuals to join the panel when existing panel members disconnect from the panel. This strategy is referred to as rotation, see Patterson (1950) and Kish and Hess (1959). Naturally, the next step is to decide upon the number of survey requests within this period, to be labelled n. Note that T^* and n together constitute the sampling frequency $f = n/T^*$ of the survey, which is equal to the reciprocal of the time between subsequent survey occasions, or waves. Thirdly and finally, one needs to decide when to conduct the n surveys within the time-span T^*. We will refer to this aspect as date selection. A natural way is to divide the time-span T^* into n equally long time periods, and to survey around the beginning of each sub period. Typically in this case the implied sampling frequency f is lower than the desired data frequency. Again one may therefore apply rotation, such that at each point in time t a new group of panel members is surveyed and the data are collected continuously. The above strategy is mostly referred to as time sampling. As an alternative, Segers and Franses (2007) proposed to choose the n survey occasions at random, independently for each panel member. They show that in this case of randomized sampling, data is collected to measure every possible autocorrelation up to $T^* - 1$ lags, where the lower lag orders are sampled most frequently. This facilitates the identification of any type of individual dynamics in the data and it allows for efficient estimation.

To the best of our knowledge, the only consumer confidence indicator that is not obtained from repeated cross-sections is the Index of Consumer Sentiment of the University of Michigan. Michigan adopts a rotating panel design in which the respondents are requested to be re-interviewed six months after the first interview, see Curtin (1982) for details. This design is illustrated in Figure 4.1b. In our terminology, we would characterize the Michigan panel as a rotating panel where $T^* = 12$ months and $n = 2$ survey occasions per individual and time sampling is applied.

An example of our preferred randomized rotating panel, where two new individuals are invited to join the panel in each time period, is shown in Figure 4.1c In this example, we set the maximum time-span that a panel member is requested to join the panel, T^*, equal to 8 and the number of survey requests, n, equal to 4. As a consequence, the sampling frequency f is 0.5. Each dotted area encloses all survey requests assigned to one particular cohort of individuals.

Figure 4.1 Panel designs to measure consumer confidence.



a. Repeated cross-sections

b. The Michigan-panel

c. A randomized rotating panel

4.3 Modeling Consumer Confidence

Notation and motivation

Typically, consumer confidence surveys are composed of 3 to 10 questions. Consumers are asked whether they believe that the economic conditions in their country have improved in the previous period or will improve in the next period. Often, similar questions are then posed related to the respondents' believes about their private financial situation. In practice, the number of answer categories varies from 3 (negative, neutral or positive) to 201 (-100, 0, +100). To obtain one score per individual, the negative answers are then deducted from the positive answers, or the answers are simply averaged, depending on the design of the survey.

To convert the survey data to a daily measure of consumer confidence, we will need to use an econometric model, as we will argue below. We aim to keep the model specification flexible, such that the model can be easily adopted to match a wide range of consumer confidence surveys. First, let us denote the response of individual i on the j-th survey question on day t by $z_{i,j,t}$, with i and t as before, and $j = 1, \dots, J$, where J denotes the total number of questions. Second, let $z_{i,t}$ be the overall confidence score of an individual, which is generally obtained by simply taking the sum or the average of the J answers provided by the individual i at time t. We assume that $z_{i,t}$ is scaled such that all scores are in the closed interval $[-1, \dots, 1]$, with -1 being the most negative response and 1 being the most positive response possible. Finally, but not necessarily, the data is then further reduced by classifying an individual as being in a negative, neutral or positive state of confidence, using

$$q_{i,t} = \begin{cases} 1 & \text{if } z_{i,t} > \tau \\ 0 & \text{if } |z_{i,t}| \leq \tau \\ -1 & \text{if } z_{i,t} < -\tau \end{cases} \tag{4.1}$$

where $q_{i,t}$ is the final classification of the individual and τ is a threshold parameter. We collect the values of $q_{i,t}$ in the $I \times T$ matrix Q, defined as

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,t} & \cdots & q_{1,T} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,t} & \cdots & q_{2,T} \\ \vdots & \vdots & & \vdots & & \vdots \\ q_{i,1} & q_{i,2} & \cdots & q_{i,t} & \cdots & q_{i,T} \\ \vdots & \vdots & & \vdots & & \vdots \\ q_{I,1} & q_{I,2} & \cdots & q_{I,t} & \cdots & q_{I,T} \end{bmatrix} \tag{4.2}$$

To obtain a daily consumer confidence index c_t, one may calculate the average of the collected $q_{i,t}$'s. This would be equal to the proportion of positive responses minus the proportion of negative responses. The matrix of responses Q is sparse because of two reasons. First, by design, in each wave only a limited number of respondents is requested to be surveyed. Second, respondents may decide not to participate in the survey. This implies that taking the average of $q_{i,t}$ would yield estimates with large measurement error. We can produce more accurate daily averages if we impute the missing values in Q, before taking the averages over all respondents. To impute the missing values in Q, we employ a Markov transition model.

The Markov Transition model

As we want to focus on the number of positive and negative responses while also incorporating dynamics, we use a Markov Transition model, see Ross (2007) among others for a detailed assessment. This model uses transition probabilities to reflect the probability of transferring from one state of confidence to the next. We jointly denote these probabilities by the transition matrix P. For the situation with a positive, neutral and negative state, we define P as

$$P = \begin{bmatrix} p_{1,1} & p_{1,0} & p_{1,-1} \\ p_{0,1} & p_{0,0} & p_{0,-1} \\ p_{-1,1} & p_{-1,0} & p_{-1,-1} \end{bmatrix} \tag{4.3}$$

The elements $p_{l,m}$ correspond to the probability of transitioning in one day from state l to state m: $\mathbb{P}(q_{i,t} = m | q_{i,t-1} = l)$, with $l, m \in \{-1, 0, 1\}$. The k-day transitioning matrix P_k can now simply be obtained by taking the k-th power of P: $P_k = P^k$. We denote its elements by $p_{k,l,m}$. Note that it is straight-forward in this setup to include more states. Using extra states allows to also capture slightly positive and negative states of confidence.

Given the elements of the matrix P, we can estimate the probability that individual i is in state l at time t, to be denoted by $\pi_{i,t,l}$. For this purpose, we need to administer two additional variables, for each $q_{i,t}$. The first is the number of days that have passed since the last survey response of individual i at time t, to be denoted by Δt. Further, denote the day of the last response by $s = t - \Delta t$. The second is the value of the state $q_{i,s}$. Then

$$\mathbb{P}(q_{i,t} = l) = \pi_{i,t,l} = p_{\Delta t, q_{i,s}, l} = [P^{\Delta t}]_{q_{i,s}, l} \tag{4.4}$$

In order to estimate the elements of P, we use a maximum likelihood approach (see Cameron and Trivedi, 2005, among others), which we set up as follows. First, we construct the individual likelihood contribution $\mathcal{L}_{i,t}$. Given an estimate $\hat{P}$, we can calculate an estimate of $\pi_{i,t,l}$, using (4.4). But for some combinations of i and t, $q_{i,t}$ is known. The likelihood of observing such a single data point is captured by $\mathcal{L}_{i,t} = \prod_{l=-1}^{1} \pi_{i,t,l}^{I[q_{i,t}=l]}$, which is equal to 1 if there is no response for individual i at day t or if the response $q_{i,t}$ is not state l. The latter is equal to the estimated state probability $\pi_{i,t,l}$ if $q_{i,t} = l$. Next, we find the total likelihood by multiplication of the individual contributions, for all time periods and for all individuals, that is,

$$\mathcal{L} = \prod_{i=1}^{I} \prod_{t=1}^{T} \mathcal{L}_{i,t} = \prod_{i=1}^{I} \prod_{t=1}^{T} \prod_{l=-1}^{1} \pi_{i,t,l}^{I[q_{i,t}=l]} \tag{4.5}$$

For reasons of numerical stability and feasibility, we maximize the logarithm of $\mathcal{L}$, which is given by

$$log\mathcal{L} = \sum_{i=1}^{I} \sum_{t=1}^{T} log\mathcal{L}_{i,t} = \sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{l=-1}^{1} I[q_{i,t} = l] \; log \, \pi_{i,t,l} \tag{4.6}$$

We need to maximize $log\mathcal{L}$ over the elements of matrix P. For each row of P it is the case that if two of the elements of that row are known, the third element can be calculated because the elements must sum to 1. The number of parameters over which we must maximize this function is thus equal to six.

Once consistent estimates of the model parameters are obtained, we impute the missing values in the matrix Q by the difference between the probability of being in the positive and the probability of being in the negative state, at each wave, that is,

$$\hat{q}_{i,t} = \mathbb{P}(q_{i,t} = 1) - \mathbb{P}(q_{i,t} = -1) = \hat{\pi}_{i,t,1} - \hat{\pi}_{i,t,-1} \tag{4.7}$$

The value of the aggregated consumer confidence index, c_t, is then equal to the average of these state estimates:

$$c_t = \frac{1}{I}\sum_{i=1}^{I} \hat{q}_{i,t} \tag{4.8}$$

4.4 Empirical Illustration

We illustrate the usefulness of our method of data collection and the Markov transition model by showing how the data collection method and model can be employed to produce forecasts of household consumption.

Data collection

To collect consumer confidence data at the individual level using our data collection method, we invited all alumni who graduated as a BSc or MSc at our institute to join an expert panel specifically set up for the occasion. A first group of alumni was invited to join the panel in January 2010, and a second group was invited to join the panel in September 2011. The alumni who accepted the invitation were requested to complete a short questionnaire at most four times per year. The invitations were sent out every week on Monday at 9 am. In order to be able to compare our indicator to the official CCI of Statistics Netherlands (SN), we used the same questionnaire as SN. The survey consists of five questions regarding the economic climate in The Netherlands and the respondents' willingness to buy durable goods, see Appendix 4.A for details. The number of answer categories (7 or 21) as well as the sampling strategy (time-sampling or randomized) was varied across the respondents. Over the period January 4, 2010 up to October 13, 2014, 250 alumni participated in the experiment. We sent out 3625 survey requests. The response rate was 33.9%. We fixed the

Table 4.1 Estimated transition matrix of the Markov switching model with its implied long-term properties

Transition from state	Transition to state (day-to-day)			Transition to state (90 days)			Steady state
	Optimistic	Neutral	Pessimistic	Optimistic	Neutral	Pessimistic	
Optimistic	0.996	0.000	0.004	0.732	0.128	0.141	0.418
Neutral	0.006	0.206	0.788	0.194	0.387	0.420	0.279
Pessimistic	0.000	0.730	0.270	0.191	0.388	0.421	0.304

threshold parameter τ in Equation (1) at 0.1, which resulted in three categories of comparable size.

In-sample results

We have estimated the transition probabilities $p_{l,m}$ by maximizing the likelihood function as discussed in Section 4.3. The results are shown in Table 4.1. The day-to-day transition matrix is quite similar to the identity matrix, which suggests that a respondent's state of confidence typically does not change on a daily basis. Therefore, in the second panel of Table 4.1, we also report the transitions over a period of 90 days. 73.2% of the respondents who are optimistic about the economy remain positive after 90 days, while 12.8% becomes neutral and 14.1% becomes pessimistic. Staying in the neutral or the pessimistic state after 90 days is more unlikely than staying in the optimistic state with probabilities of 38.7% and 42.1%, respectively. Finally, in the third panel of Table 4.1 we report the steady state distribution. In the long-run, 41.8% of our respondents is optimistic, 27.9% is optimistic and 30.4% is pessimistic about the economy. These implications on the longer horizon seem realistic in the context of our data, as the number of responses is approximately the same for each state of confidence.

Construction of the index

Next, we discuss how we can use our estimates of c_t to forecast monthly consumption in the Netherlands. We aim to compare our forecasts to the forecasts one would obtain using the official CCI of SN, to be denoted by CCI_{SN}. SN only collects their data during the first ten working days of each month. To allow for a fair forecast comparison, we construct our monthly CCI from the daily CCI by taking the value obtained on the tenth working day of each month only. We denote our monthly CCI by CCI_{MS}.

Besides our two main indicators, CCI_{SN} and CCI_{MS}, we consider two additional indicators, which cannot be observed if the data are collected using repeated cross-sections. The first aims capture the spread in confidence across individuals, and is defined as the time-varying standard deviation of $\hat{q}_{i,t}$:

$$S_{MS,t} = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I}(\hat{q}_{i,t} - c_t)^2} \tag{4.9}$$

The second indicator measure the rate at which respondents have recently changed their state of confidence. We define this rate of change as the aggregated absolute 30 day change in $\hat{q}_{i,t}$:

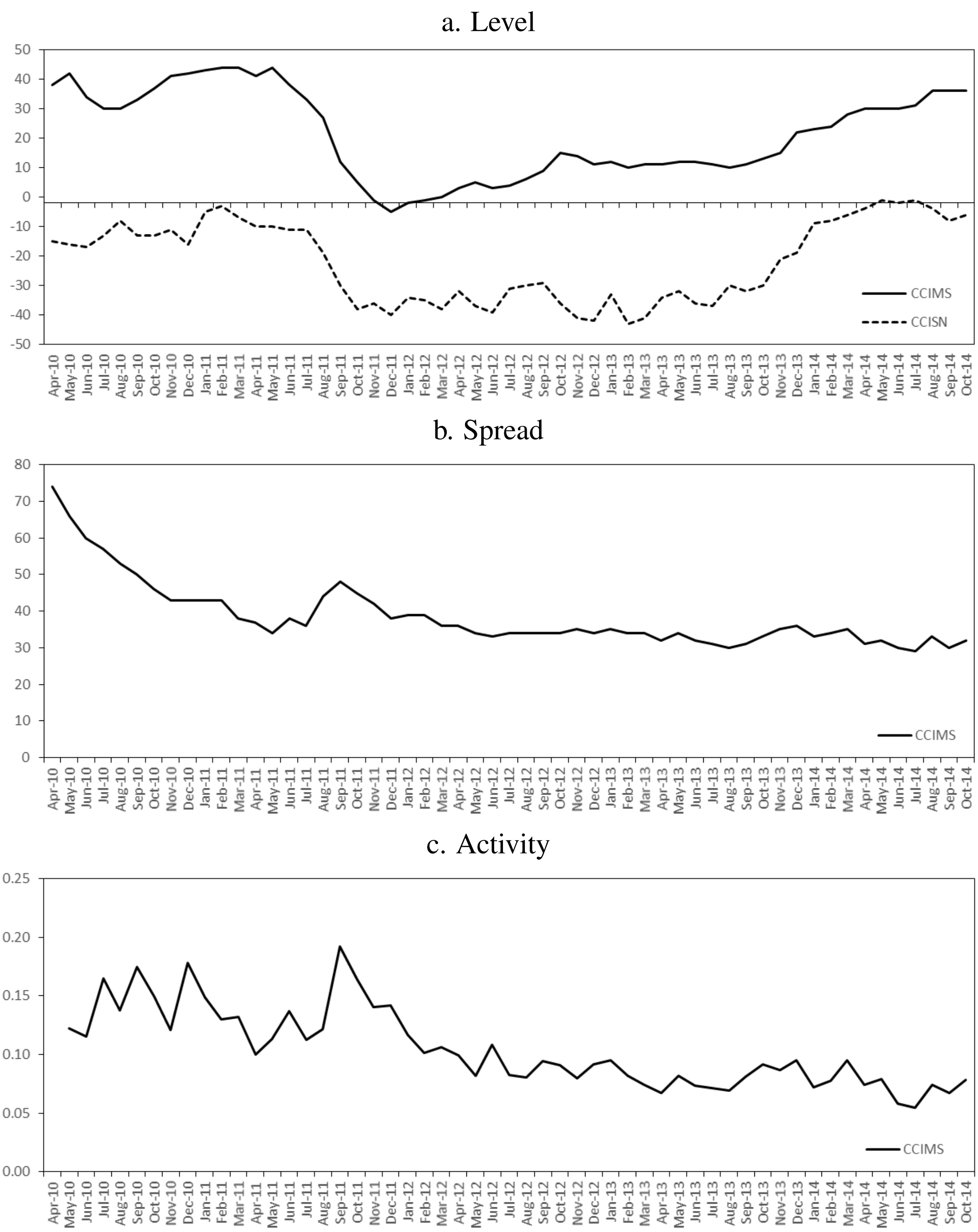
$$A_{MS,t} = \sum_{i=1}^{I} |\hat{q}_{i,t} - \hat{q}_{i,t-30}| \tag{4.10}$$

Figure 4.2a shows the developments in both the CCI_{MS} and the CCI_{SN} over the period April 2010 to October 2014. First of all, note that the average level of the novel indicator is about 43 points higher than that of SN. This is most probably due to the fact that our panel consists of university alumni only, who are generally positive about their financial situation, while SN invites aims to select a representative sample of the population. However, the cycles of the two indicators are very similar, with the most prominent turning point being the moment when the impact of the European debt crisis became apparent. Finally, we note that there tends to me more short-term variation in the indicator of Statistics Netherlands, compared to the model based indicator. Figure 4.2b and 4.2c display the spread, $S_{MS,t}$, and the activity, $A_{MS,t}$, of the novel indicator. Both indicators show an increase during the freefall of the index in the summer of 2011 and peak in September 2011, but also distinct fluctuations that do not necessary coincide with developments in the index.

Forecasting comparison

The target variables in our forecast comparison are Dutch national household consumption and its four subcomponents. Household consumption is measured by SN and is subdivided into SERVICES AND GOODS, where GOODS are further subdivided into FOOD, DRINK AND TOBACCO, DURABLES and OTHER GOODS. DURABLES are defined as goods that in principle last more than one year, such as clothes and textiles, shoes, furniture, consumer elec-

Figure 4.2 Level, spread and activity of our consumer confidence index.



tronics and cars. OTHER GOODS include energy, motor fuels and all other consumer goods that cannot be classified as FOOD, DRINK AND TOBACCO, or DURABLES.

To forecast consumption one month ahead, we consider three basic forecasting models. In the first we only include the level of the consumer confidence indicator of SN, CCI_{SN}, and its first difference, alongside with a constant term and AR(1) dynamics:

$$y_t = \alpha + \beta CCI_{SN,t-1} + \gamma \Delta CCI_{SN,t-1} + \phi y_{t-1} + \varepsilon_t \tag{4.11}$$

In the second forecasting model, we replace CCI_{SN} by our index CCI_{MS}:

$$y_t = \alpha + \beta CCI_{MS,t-1} + \gamma \Delta CCI_{MS,t-1} + \phi y_{t-1} + \varepsilon_t \tag{4.12}$$

Finally, the third forecasting model is equal to (4.12) but now also includes the spread and activity indicators, S_t and A_t:

$$\begin{aligned} y_t = \alpha + \beta_1 CCI_{MS,t-1} + \gamma_1 \Delta CCI_{MS,t-1} + \beta_2 S_{MS,t-1} + \gamma_2 \Delta S_{MS,t-1} + \\ \beta_3 A_{MS,t-1} + \gamma_3 \Delta A_{MS,t-1} + \phi y_{t-1} + \varepsilon_t \end{aligned} \tag{4.13}$$

The three forecasting models are all estimated using Ordinary Least Squares.

In the first set of columns in Table 4.2, we display the R^2 values of models (4.11), (4.12) and (4.13). All models for the components FOOD, DRINK AND TOBACCO and TOTAL CONSUMPTION explain less than 10% of the variation in the data. This suggests that confidence indicators are less useful for explaining the latter two components. In contract, SERVICES is explained best by the models, with a R^2 of 0.93 for models (4.11) and (4.12) and 0.94 for model (4.13). This series is relatively persistent, as demonstrated by the estimated value of the AR parameter ϕ (0.89, p-value: 0.08), see columns 4 and 5. We then test whether the additional four parameters of model (4.13) relative to model (4.12) are jointly significant using an F-test. The corresponding p-values are shown in column 6. For the components DURABLE GOODS, OTHER GOODS and SERVICES, (4.12) is rejected in favor of (4.13) at the 10% significance level. This suggests that the additional components S_t and A_t have indeed explanatory power.

Table 4.2 Diagnostics of the estimated forecasting models (4.11), (4.12) and (4.13)

Target variable	R^2 (4.11)	(4.12)	(4.13)	ϕ in (4.13)	$SE_{\phi,(4.13)}$	p-value of (4.13) vs (4.12)
Food, drinks and tobacco	0.05	0.06	0.17	-0.15	0.17	0.26
Durable goods	0.37	0.25	0.38	-0.07	0.17	0.06
Other goods	0.59	0.56	0.63	0.67	0.10	0.09
Services	0.93	0.93	0.94	0.89	0.08	0.03
Total consumption	0.06	0.06	0.10	0.00	0.17	0.76

Table 4.3 Relative forecast errors of the forecasting equations (4.12) and (4.13), compared to (4.11)

Target variable[2]	(4.12): CCI_{SN}, level only	(4.13): CCI_{SN}, level, spread, activity
Food, drinks and tobacco	0.998	0.901
Durable goods	1.127	0.877
Other goods	1.127	0.804
Services	0.878	0.474
Total consumption	0.989	0.957
Average	1.024	0.802

Finally, we compare the forecasting performance of models (4.11), (4.12) and (4.13) in Table 4.3. The first column displays the ratio of the forecast error of (4.12) to the forecast error of (4.11). As a result, values below 1 indicate that (4.12) outperforms (4.11). Since on average, however, the ratio is close to 1, we conclude that the level of the novel indicator CCI_{MS} performs approximately as good as model (4.11), which contains the indicator CCI_{SN}. As both models contain the same number of parameters, this indicates that no clear decision between both models can be made. The next column displays the ratio of the forecast error of (4.13) to the forecast error of (4.11). Model (4.13) has relative forecast error variances that are 20% smaller than those of (4.11), on average. Especially the forecasts for components DURABLE GOODS, OTHER GOODS and SERVICES benefit from the use of the novel indicator and its additional components S_t and A_t.

4.5 Conclusions

In this chapter we proposed to collect randomized panel data rather than repeated cross-sections to measure consumer confidence. Randomized panel data allow us to not only observe longitudinal changes in confidence across our respondents, but also to observe changes

[2] Measured in terms of the total value, where the average value over 2000 is normalized to be 100. Source: http://www.cbs.nl/en-GB/menu/themas/dossiers/conjunctuur/publicaties/conjunctuurbericht/inhoud/conjunctuurklok/toelichtingen/ck-06.htm

in the spread in confidence across individuals, as well as changes in rate at which respondents change their state of confidence over time.

We demonstrated the usefulness our approach in an application to measuring consumer confidence in The Netherlands over the period April 2010 to October 2014 with the purpose to produce one-month ahead forecasts of Dutch consumption. We showed that the incremental information content in the novel panel data indicator improves the forecasting accuracy by 20% on average.

We mention various directions for further research. The first is that we can now correlate significant weekly changes with weekly observed macroeconomic variables, in order to study whether consumer confidence has predictive value. Indeed, currently most such studies concern monthly observed cross-sectional data, and it may well be that substantial information is lost. We may also consider enriching the model with various explanatory variables that do correlate with consumer confidence but do not necessarily correlate with the target variable. For example, it has been hypothesized that factors such as the weather and specific events such as terrorist attacks have an impact on consumer confidence. An indicator that is corrected for one or more of these factors might be a better predictor of the future course of the economy than the uncorrected indicators that are currently in use. Finally, it might be a worthwhile endeavor to extend the Markov transition model. For example, we have assumed constant transition probabilities, while in practice transition probabilities might be time-varying. This might be particularly relevant when the economy shifts between expansions and recessions.

4.A The Consumer Confidence Survey of Statistics Netherlands

As opposed to the consumer confidence indicator measured by the European Commission, the indicator measured by Statistics Netherlands not only concerns consumers' opinions on their financial situation, the economy in general, willingness to save and unemployment in the next twelve months, but also consumers' present situations and their opinions on the previous twelve months. The two indicators show roughly the same developments overtime.[3]

Consumer confidence is based on five questions from a more elaborate consumer survey. These questions are subdivided into a section on the economic climate and a section on the respondent's willingness to buy. The questions are formulated as follows:

Economic Climate

1. How do you think the general economic situation in this country has changed over the last twelve months?

   Possible answers: At present, it is better (1) / the same (0) / worse (-1)

2. How do you think the general economic situation in this country will develop over the next twelve months?

   Possible answers: It will be better (1) / the same (0) / worse (-1)

Willingness To Buy

3. How does the financial situation of your household now compare to what it was twelve months ago?

   Possible answers: At present, it is better (1) / the same (0) / worse (-1)

4. How do you think the financial situation of your household will change over the next twelve months?

   Possible answers: It will be better (1) / the same (0) / worse (-1)

[3] See http://www.cbs.nl/en-GB for details.

5. Do you think that at present there is an advantage for people to make major purchases, such as furniture, washing machines, TV sets, or other durable goods?

   Possible answers: Yes, now it is the right time (1) / It is neither the right nor the wrong time (0) / No, it is the wrong time (-1)

The economic climate indicator is computed as the summed score of questions 1 and 2 averaged across all individuals. Similarly, the willingness to buy indicator is computed as the summed score of questions 3 to 5 averaged across all individuals. Finally, consumer confidence is defined as the average of all five scores.

Chapter 5

How informative are the unpredictable components of earnings forecasts?

Based on de Bruijn and Franses (2013). Authors contributed following a 90% / 10% split.

Abstract

An analysis of about 300000 earnings forecasts, created by 18000 individual forecasters for earnings of over 300 S&P listed firms, shows that these forecasts are predictable to a large extent using a statistical model that includes publicly available information. When we focus on the unpredictable components, which may be viewed as the personal expertise of the earnings forecasters, we see that small adjustments to the model forecasts lead to more forecast accuracy. Based on past track records, it is possible to predict the future track record of individual forecasters.

Acknowledgements: We are grateful to Roy Batchelor, Scott Stickel and workshop participants of the International Symposium on Forecasting in Boston (2012) and Rotterdam (2014) for their helpful comments.

5.1 Introduction

Earnings forecasts can provide useful information for investors. They have been the focus in many studies, where typically their accuracy and their managerial usefulness are studied, see for example Abarbanell et al. (1995), Abarbanell and Lehavy (2000, 2003), Bradshaw et al. (2013), Bradshaw et al. (2012), Givoly et al. (2009), and Mikhail et al. (1999, 2004). In contrast, in our present study we focus on the predictability of earnings forecasts, and in particular we aim to learn more about the non-predictable part.

When investors rely on these forecasts, it is important to have insights into how earnings forecasters create their forecasts. Knowledge about the key drivers of the earnings forecasts is relevant as it allows for the analysis of the added value of earnings forecasters. More precise, one may want to disentangle the part of earnings forecasts that can be predicted using publicly available information from the part involving private knowledge that the earnings forecasters themselves have. In the present chapter we will first estimate the predictable part and then focus on the usefulness of the unpredictable part. That is, to what extent does the genuine contribution of an earnings forecaster, beyond that part that can be predicted using publicly available information, lead to more forecast accuracy?

Our present study extends the important study of Stickel (1990) in various dimensions. He investigated whether the change in the forecast by an individual forecaster could be predicted by a change in the average forecast of other forecasters, the deviation of the forecaster's previous forecast relative to that total average and the cumulative stock returns since a previous forecast. In Stickel (1990) it is concluded that virtually all explanatory power is associated with the first variable, that is, the change in the average forecast.

We update and extend in several ways. First, we use recent data for the period 1995 to 2011. Second, instead of looking at changes in forecasts we consider the actual levels, in particular as we want to zoom in on the unpredictable part of the forecasts. Third, we allow for the inclusion of more potential explanatory variables in our statistical model. And last, but most importantly, we focus on the added value of the earnings forecasts and seek to derive informative rules to discern the better forecasters from the lesser performing forecasters.

A summary of our findings is the following. A key predictor of the earnings forecasts appears to be the average of all available earnings forecasts concerning the same forecast event. A second predictor is the most recent difference between the individual forecaster's forecast and the average of the currently available forecasts. As the sign is positive, this means that a forecaster who previously was more optimistic about the earnings of a particular firm can be expected to persist in quoting above-average values. Other variables do have some predictive value in individual cases, but we do not find consistent effects. When we focus on the unpredictable components then one of our key findings is that a larger unpredictable component associates with less forecast accuracy. We also document that alternative weights to these unpredictable components can lead to more accuracy. Separating the data in an estimation sample and evaluation sample allows us to draw our final conclusion which is that past track records of forecasters have predictive value for future track records.

The outline of this chapter is as follows. In Section 5.2 we provide a concise summary of the empirical evidence in the literature. In Section 5.3 we discuss the data and in Sections 5.4 and 5.5 we present our results. Section 5.6 summarizes our findings.

5.2 Literature review

Earnings forecasts have been the topic of interest for many academic studies. For an extensive discussion of research on earnings forecasts in the period 1992-2007, see Ramnath et al. (2008). For earlier overviews we refer to Schipper (1991) and Brown (1993).

One stream of earnings forecasts research has focused on relationships between forecast performance and forecaster characteristics. Performance can be measured by forecast accuracy and by forecast impact on stock market fluctuations. The characteristics of these performance measurements have been related to timeliness (Cooper et al., 2001; Kim et al., 2011), the number of firms that the forecaster follows (Kim et al., 2011; Bolliger, 2004), the firm-specific experience of the forecaster (Bolliger, 2004), age (Bolliger, 2004), the size of the firm for which the forecasts are created and the size of the company where the forecaster works (Kim et al., 2011; Bolliger, 2004), and whether the forecaster works individually or in a team (Brown and Hugon, 2009).

A second stream of research concerns the behaviour of an earnings forecaster and how it is related to what other forecasters do. In particular, herding behaviour is considered, which occurs when forecasters produce forecasts that converge towards the averages of those of the other forecasters. There have been efforts to categorize earnings forecasters into two groups, corresponding to leaders and followers or to innovators and herders (Jegadeesh and

Kim, 2010; Clement and Tse, 2005). This is relevant for many reasons as such different forecasters may consult different sources of information, which in turn can be useful for investors to incorporate this information into their investment decisions. Indeed, a leading or innovative forecaster is perhaps more useful than a herding forecaster. This does not directly imply that leading forecasts are also more accurate, as it is documented that accuracy and the type of forecaster are not necessarily related. In fact, it has been documented that the aggregation of leading forecasts is a fruitful tactic to produce accurate forecasts (Kim et al., 2011).

Recently, Clement et al. (2011) have studied the effect of stock returns and other forecasters' forecasts on what forecasters do. In contrast to Jegadeesh and Kim (2010) and Clement and Tse (2005), Clement et al. (2011) do not consider categorizing the forecasters into different clusters. Instead, they consider how the first forecast revision after a forecast announcement is affected by how the stock market and other forecasters have reacted to that forecast announcement. Landsman et al. (2012) also look at how earnings announcements affect the stock market, where these authors focus on how mandatory IFRS adoption has moderated this effect. Sheng and Thevenot (2012) propose a new earnings forecast uncertainty measure, which they use to demonstrate that forecasters focus more on the information in the earnings announcement if there is more dispersion in the available earnings forecasts.

In sum, earnings forecasts have been studied concerning their performance and a few of their potential drivers. In this chapter we extend the knowledge base by considering many more drivers of earnings forecasts, while we pay specific attention to the relevance of the unpredictable component of earnings forecasts.

For our study we go back to Stickel (1990) and seek to extend this important study in various dimensions. In that paper it is concluded that in a statistical model for predicting changes in earnings forecasts the key explanatory variable is the change in the average of all other forecasts. We extend this study by considering more and more recent data and also by including more variables in a model for the levels (and not the changes) of earnings forecasts. A key extension however is that we use the statistical model to disentangle the predictable component from the unpredictable component, and then we zoom in on the latter component. We do so to see to what extent the unobservable knowledge of forecasters contributes to the quality of the earnings forecasts. Also, we aim to examine if forecasters who successfully rely on their knowledge do so persistently. That is, is a past successful track record an indicator for future success?

5.3 Data and sample selection

Our data have been collected from WRDS[1], using the I/B/E/S database for the analyst forecasts and the CRSP data for the stock prices and stock returns.

Concerning the earnings forecasts, we have collected data for all firms which have been part of the S&P500 during the period 1995 to 2011. Sometimes the sample size was too small, and in other cases we could not properly link the forecasts with the firms, so in the end we have useful information concerning 316 firms with some 270000 earnings forecasts. We focus on the within-year annual earnings forecasts, that is, the forecasts that are produced to forecast the earnings of the current year.

The structure of the data is illustrated in Table 5.1. This figure shows an "x" for the moment when a forecaster makes a forecast available, which is not necessarily the same moment that other forecasters give their quotes as not all forecasters have the same frequency of quotes. This figure also shows the variables which we measure at the highest frequency and these are the daily observed stock returns. Finally, this figure shows double vertical lines depicting the moments of the earnings announcements, at which point the realization occurs of the variable that is to be forecasted. We only use the within-year earnings forecasts, which means that we only include forecasts for the next upcoming annual earnings announcement, and hence we abstain from forecasts for year T made in year $T-1$.

Some descriptives of the data are shown in Table 5.2. The data until and including 2005 cover the estimation sample, and the data from 2006 onwards constitute the evaluation sample. We make this distinction in order to examine if past track records have predictive value for future track records. And, we also want to see if estimated parameters in the estimation sample provide reasonably constant inference in a post-estimation period.

All data are used to create and evaluate the statistical models for the earnings forecasts. For that purpose we have data on 18338 forecasters and more than 270000 forecasts, with the latter about equally spread over the estimation and evaluation samples. When it comes

[1] Wharton Research Data Services (WRDS) was used in preparing this chapter. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers. `http://wrds-web.wharton.upenn.edu/wrds/`

Table 5.1 An example of the data format, with an x indicating an earnings forecast and the double vertical lines indicating when a new yearly earnings announcement takes place. This figure shows for five forecasters for two years a variety of hypothetical patterns of forecasts, including analysts that follow a very regular forecasting pattern, or the opposite, and including forecasters that quit producing forecasts or that join during a later year. In contrast, (daily) returns are shown as an example of the explanatory variables, which are observed at every measurement point.

Analyst 1	x			x			x			x			x				x	
Analyst 2		x				x						x			x			
Analyst 3				x									x			x		
Analyst 4											x			x			x	
Analyst 5	x		x	x	x		x	x	x		x	x						
Returns	r	r	r	r	r	r	r	r	r	r	r	r	r	r	r	r	r	r

Table 5.2 The number of firms, forecasters and forecasts for each upcoming section and subsection. The number of forecasts is shown separately for the estimation sample, which is up until 2005, and the evaluation sample, which is from 2006 onwards.

	Number of firms	Number of forecasters	Number of forecasts	
			Estimation sample	Evaluation sample
Section 5.4 and 5.5.1	316	18338	146319	126651
Section 5.5.2	316	1835	52236	36403
Section 5.5.3	316	4541	90190	28000

to forecaster-specific regressions and correlations, we need enough data points to run these computations, and then our sample size drops to about one-third of the forecasts. Still, this is a large database and therefore we are confident that our results below are informative.

5.4 Predicting earnings forecasts

To create the unpredictable components of earnings forecasts, we first have to create the predictable components. For this we put forward a statistical model to predict earnings forecasts using information publicly available up until the day before the publication of the earnings forecast. In this section we first introduce the statistical model that we use to make predictions of the earnings forecasts. We present the explanatory variables and the relevant estimation results. Next, we apply a correction method to account for the firms for which we have a small number of forecasts.

5.4.1 The statistical model

For predicting the earnings forecasts we use a linear regression model. The list of explanatory variables is presented in Table 5.3.

Table 5.3 The variables that are used to forecast the earnings forecasts. These variables enter a linear model. They all use one-day lagged information. Several variables are based on historic forecasters' behaviour, others are based on stock market data.

Variable	Description
Intercept	
Average Forecast	The average of all most recent forecasts of every forecaster, until the previous day
Average Forecast is	*also included in an interaction term with two indicator variables:*
	1. Whether the number of forecasters is lower than 10 or not, $I[nrF < 10]$
	2. Whether the time until the announcement of the earnings is more than two weeks or not, $I[TUA > 10]$
Δ Average Forecast	First difference in Average Forecast
Δ Previous Forecast	The difference between the previous forecast of the forecaster, and the average forecast at that time
Previous Forecast	The previous forecast of the individual forecaster
Stock Price Firm	The daily stock price of the firm for which the earnings are forecasted
Stock Returns Firm	The daily returns of the firm for which the earnings are forecasted
Cumulative Stock Returns Firm	Stock returns of the firm since the day of the previous forecast by this forecaster
Stock Index S&P500	The daily stock market index of the S&P500
Stock Returns S&P500	The daily stock market returns of the S&P500
Cumulative Stock Returns S&P500	Stock market returns of the S&P500 since the day of the previous forecast by this forecaster

Following Stickel (1990), we expect earnings forecasters to look at the recent forecasts of competing forecasters. We thus include in our model the average of all most recent forecasts across individual forecasters. Note that we only include forecasts that have been made within the same year for the same forecast event.

We also include in our model several variables that are related to this average forecast. First, the average forecast may contain more useful information when it is based on a larger number of forecasters. To see whether this holds true, we include an interaction term of the average forecast with an indicator function that is 1 if the number of forecasters is below 10 and that is 0 otherwise. The contribution to the fit of average forecast may also increase when the moment of the actual announcement of the true value of the earnings comes closer. We therefore include an interaction term with an indicator function which is 1 when the data concern the last two weeks before the announcement, and 0 otherwise. The final explanatory variable related to the average forecast is the day-to-day change in this average forecast. Indeed, when the average forecast has increased on one day, then individual forecasters could be tempted to extrapolate this growth to the next day.

The second set of explanatory variables concerns the own previous forecasts of a forecaster. We include the most recent forecast and the difference between this previous forecast and the average forecast at that particular moment in time. These two variables can allow for persistence in the opinions of a forecaster, implying that forecasters can be more optimistic or pessimistic for some period of time.

Finally, the third set of explanatory variables concerns the stock market. We include the most recent stock price of the firm for which its earnings are predicted. Also, recent changes in the stock price can be relevant, and for that purpose we include the daily returns and the returns relative to the most recent moment when an individual forecaster produced a forecast. Next to these three firm-related stock prices, we include similar variables for the entire S&P500 stock exchange.

We estimate the parameters in the regression models using Ordinary Least Squares (OLS) for data for each of the 316 firms, and summaries of the estimation results across these 316 firms are presented in Table 5.4. The first five columns show results on the OLS based estimates, including the mean of the estimated parameters, the median and their standard deviation and also the 5% and a 95% percentiles of these estimates. The next two columns concern a summary of the standardized estimates, which are the estimates that are found if

Table 5.4 A summary of estimation results of forecasting earnings forecasts. Results are for the estimation sample, which amounts to 316 firms, 18338 forecasters and 146319 forecasts. The variable to be explained concerns the earnings forecasts. As explanatory variables we include the variables mentioned in Table 2. The regression is run individually for each firm, and the table shows statistics which summarize these results. The first five columns contain summary results on the regular parameter estimates (average, median, standard deviation and bounds of a 90% interval). The last three columns show results for the standardized estimate, which is included to compare contributions to the fit. The standardized estimate is defined as the estimate that would have been obtained had the regressor been standardized beforehand (which is a transformation to having an average of 0 and a standard deviation of 1).

	Estimate					Standardized estimate		
	Average	Median	Standard Deviation	Bounds of 90% interval		Median	Median of absolute	Contribution to total fit
Intercept	-0,028	-0,026	0,278	-0,395	0,229			
Average Forecast	1,050	1,077	0,384	0,557	1,483	0,490	0,491	96,9%
Average Forecast x $I[nrF < 10]$	0,018	0,005	0,172	-0,053	0,078	0,001	0,004	0,0%
Average Forecast x $I[TUA > 14]$	-0,038	-0,034	0,292	-0,212	0,058	-0,017	0,025	0,3%
Δ Average Forecast	0,765	0,607	1,053	-0,526	2,438	0,005	0,006	0,0%
Δ Previous Forecast	0,498	0,548	0,359	-0,151	1,051	0,029	0,030	0,4%
Previous Forecast	-0,046	-0,078	0,273	-0,388	0,375	-0,028	0,076	2,3%
Stock Price Firm	0,002	0,001	0,004	-0,001	0,007	0,009	0,013	0,1%
Stock Returns Firm	0,302	0,150	1,039	-0,264	1,291	0,005	0,007	0,0%
Cumulative Stock Returns Firm	0,054	0,018	0,206	-0,092	0,285	0,003	0,005	0,0%
Stock Index S&P500	0,000	0,000	0,002	-0,001	0,002	0,001	0,009	0,0%
Stock Returns S&P500	-0,214	-0,096	1,226	-2,155	1,421	-0,001	0,004	0,0%
Cumulative Stock Returns S&P500	-0,032	-0,006	0,238	-0,408	0,256	0,000	0,005	0,0%

the variables are first all standardized by subtracting the mean and scaling the variance to 1. Such standardized estimates can be helpful when comparing the contribution of each of the variables to the overall fit. In the last column of Table 5.4 we present these contributions as percentages.

The results in Table 5.4 show that, on average, the coefficient of the recent average forecast is about 1. The distribution of this effect across firms, indicated by standard deviation and percentiles) indicates that the sign of this effect is consistently positive. When we scroll down the table, we see that none of the other variables have this property. Also, looking at the contribution to the fit, it is clear that the average forecast is most important, and that the previous forecast and its difference to the average forecast are a distant second and third useful explanatory variable.

To continue with these regression models, Table 5.5 shows summary statistics on the t-statistic values for each of the explanatory variables. The last column of this table shows that all variables are significant for at least 20% of the firms, but it also repeats the finding that

Table 5.5 A summary of t-statistics when forecasting earnings forecasts. Results are for the estimation sample, which amounts to 316 firms, 18338 forecasters and 146319 forecasts. The variable to be explained concerns the earnings forecasts. As explanatory variables we include the variables mentioned in Table 2. The regression is run individually for each firm, and the table shows statistics which summarize these results.

	Median t-statistic	Median absolute of t-statistic	Percentage significant at 5% level
Intercept	-0,986	1,920	48,4%
Average Forecast	9,865	9,865	96,4%
Average Forecast x $I[nrF < 10]$	0,530	1,118	27,9%
Average Forecast x $I[TUA > 14]$	-1,662	2,212	51,6%
Δ Average Forecast	1,680	1,766	47,2%
Δ Previous Forecast	4,794	4,794	78,0%
Previous Forecast	-0,804	1,599	38,6%
Stock Price Firm	1,928	2,402	55,8%
Stock Returns Firm	1,378	1,653	40,1%
Cumulative Stock Returns Firm	0,730	1,329	32,9%
Stock Index S&P500	0,151	1,928	49,3%
Stock Returns S&P500	-0,395	1,192	24,9%
Cumulative Stock Returns S&P500	-0,110	1,196	26,7%

most of the variables are not consistent in the sign of their effect (and thus, the sign of their t-statistic). Again, the average forecast is seen to be most relevant as it is associated with the largest percentage of significant cases. Additionally, the difference of the previous forecast to the average stands out with a higher percentage significant and a high median value of the t-statistic (78% of the cases).

5.4.2 Correcting for small sample sizes

The results in Tables 5.4 and 5.5 show that various explanatory variables do have a statistically significant effect, but at the same time this effect does not have a consistent sign. The latter causes the finding in Table 5.4 that on average these effects are equal to 0. Now it could be that this finding is a small-sample effect, as for some firms we only have a small number of earnings forecasts.

To correct for these small sample sizes, we employ the following method that is detailed in Appendix 5.A. This method amounts to an assumption that the collection of firm-specific (population) parameters for one of the variables corresponds to a normal distribution. Suppose that the parameters of this distribution are known. As a consequence, there are two sources of information for the value of each individual estimated parameter, and these are the estimated OLS coefficient and the parameters of this common distribution. The optimal choice is a weighted average of these two values, with weights determined by the standard error of the estimated coefficient and the standard deviation of the underlying distribution. For firms with only a few observations, the weight for the estimated coefficient most likely will be low, and the best estimate will thus be relatively close to the mean of the common distribution. On the other hand, for firms with many observations the weight of the estimated coefficient will be high and the best estimate will not deviate much from the OLS estimated parameters.

In our application, we of course do not know the values of the common distribution in advance. We therefore apply an iterative process. First, the two parameter values are initialized on the sample mean and standard deviation of all OLS estimates. Then, we adjust the estimates using the weights. After adjustment, we use the weighted mean and weighted standard deviation to construct a new value of the two parameters, with weights equal to the reciprocal of the estimated standard error. This is again followed by a new adjustment of the estimated parameters, and then again the calculation of a new set of parameters. We do this until convergence.

When we apply this method we obtain the summarized results in Table 5.6. Comparing the numbers in this table with those in Table 5.4, we can see that the average and median values have not changed much. In contrast, and as expected, the standard deviation and the width of the 90% interval have clearly decreased. There are now more variables that are (almost) consistent in their estimated sign, and among them are the parameters for firm-specific stock price and the S&P500 stock market index. At the same time, however, the contribution to the fit as reported in the last column has stayed about the same.

To conclude, whether we employ a small-sample correction or not, the key result is that the recent past average forecast is the main explanatory variable for current earnings forecasts. At the same time, for many individual cases (out of the 316 cases) we find various other variables to be relevant, and we will use these variables in our analyses below. Like

Table 5.6 A summary of estimation results of forecasting earnings forecasts, after using the correction method to account for small-sample error. Results are for the estimation sample, which amounts to 316 firms, 18338 forecasters and 146319 forecasts. The variable to be explained concerns the earnings forecasts. As explanatory variables we include the variables mentioned in Table 2. The regression is run individually for each firm, and the table shows statistics which summarize these results. The first five columns contain summary results on the regular parameter estimates (average, median, standard deviation and bounds of a 90% interval). The last three columns show results for the standardized estimate, which is included to compare contributions to the fit. The standardized estimate is defined as the estimate that would have been obtained had the regressor been standardized beforehand (which is a transformation to having an average of 0 and a standard deviation of 1). The correction method is based on the assumption of an underlying normal distribution out of which each parameter (for the different firms) is drawn. This provides additional information on the firm-specific estimate especially in the case when the firm has only a few observations.

	Estimate					Standardized estimate		
	Average	Median	Standard Deviation	Bounds of 90% interval		Median	Median of absolute	Contribution to total fit
Intercept	-0,019	-0,021	0,051	-0,113	0,068			
Average Forecast	1,097	1,094	0,134	0,880	1,287	0,479	0,479	98,5%
Average Forecast x $I[nrF < 10]$	0,003	0,004	0,011	-0,012	0,019	0,001	0,002	0,0%
Average Forecast x $I[TUA > 14]$	-0,048	-0,040	0,063	-0,161	0,018	-0,019	0,022	0,2%
Δ Average Forecast	0,741	0,687	0,409	0,148	1,444	0,005	0,005	0,0%
Δ Previous Forecast	0,553	0,564	0,192	0,206	0,865	0,029	0,029	0,4%
Previous Forecast	-0,081	-0,087	0,120	-0,253	0,117	-0,031	0,044	0,8%
Stock Index Firm	0,001	0,001	0,001	0,000	0,002	0,009	0,009	0,0%
Stock Returns Firm	0,160	0,120	0,175	-0,043	0,508	0,005	0,005	0,0%
Cumulative Stock Returns Firm	0,016	0,013	0,022	-0,013	0,055	0,002	0,003	0,0%
Stock Index S&P500	0,000	0,000	0,000	0,000	0,000	0,000	0,005	0,0%
Stock Returns S&P500	-0,087	-0,081	0,214	-0,410	0,251	-0,001	0,001	0,0%
Cumulative Stock Returns S&P500	-0,002	-0,002	0,036	-0,065	0,060	0,000	0,002	0,0%

Stickel (1990) we find that earnings forecasts can be predicted, and as such we substantiate these earlier findings.

The estimation results for the statistical models so far are informative in their own right, but for our present study they mainly serve to each time disentangle a predictable component from an unpredictable component. This last component will become the focus of our interest in the rest of this chapter.

5.5 How useful are the predictable and unpredictable components?

Now we have seen that earnings forecasts can be predicted to quite some extent, we will now analyse to what extent earnings forecasters add some value to the statistical model that can be constructed using publicly available data. We adopt three focus points. The first concerns all forecasts, then we see if we can evaluate individual forecasters against each other, and finally we consider the forecasts from the same forecaster and compare these with his or her own other forecasts.

5.5.1 All forecasts

We start with an examination of the predictive accuracy and compare the performance of the forecasts of the earnings forecasters (which are of course equal to the sum of the predictable and unpredictable components) with the statistical model forecasts (which are just the predictable components). Next, we zoom in on the size of the unpredictable components and examine if larger deviations from the statistical model forecasts are better or not. Finally, we look at whether we can use the unpredictable component in an alternative and perhaps better way by using different weights.

Do earnings forecasts improve on statistical model forecasts?

Table 5.7 shows some statistics on a newly created variable that seeks to highlight the differences across the two sets of forecasts. This variable is the median value (across 316 firms) of the ratio of squared earnings forecast errors over squared model forecast errors. The difference between these two sets of forecasts is the unpredictable component, so if this

Table 5.7 A summary of results on median $\frac{EFE^2}{PCE^2}$, the median ratio of squared earnings forecast error over squared predictable component error. The predictable component error is the error made if we use the part of the earnings forecast that we can predict using a statistical model. This ratio shows whether the inclusion of the unpredictable component results in an improvement or not. The last two rows show the average and median for the percentage of the forecasts for which the ratio is smaller than 1. We show results for 18338 forecasters across 316 firms, separated for the estimation (146319 forecasts) and evaluation (126651 forecasts) samples.

Period	Estimation sample	Evaluation sample
Average	0,609	0,638
Median	0,655	0,631
Standard Deviation	0,304	0,571
5% percentile	0,071	0,061
95% percentile	1,031	1,207
Percentage < 1, average	65,1%	66,4%
Percentage < 1, median	63,7%	64,4%

performance ratio is different from 1 in either direction then that must be due to this unpredictable component. The table presents this median ratio for both the estimation sample and the evaluation sample. In the evaluation sample we use the model parameters as they have been estimated using the estimation sample.

The bottom panel of Table 5.7 shows that the performance ratio is below 1 in about 65% of the cases. Hence, for 65% of the 316 firms, the earnings forecasts created by the earnings forecasters provide more accuracy than the predictable component from the statistical model. The results across the estimation sample and the evaluation sample are similar. Note that this thus means that for 35% of the firms one could easily rely on the statistical model forecasts.

Table 5.8 concerns the outcomes for the same ratio, but now for different parts of the year. These parts correspond with the four periods between the quarterly announcements and the three periods around the quarterly announcements (except for the quarterly announcement that coincides with the yearly announcement). The results in this table show that the performance ratio increases throughout the year, meaning that the unpredictable component leads to more accuracy in the beginning than towards the end. This might be due to an increase in the accuracy of the statistical model simply because the predictable components are then based on more observations and thus the sample period which may require added expertise from the earnings forecaster becomes smaller. All in all, we can conclude that the earnings forecasters can substantially contribute to the quality of the final forecasts, which is most obvious from the two bottom rows in Table 5.8.

Table 5.8 A summary of results on median $\frac{EFE^2}{PCE^2}$, the median ratio of squared earnings forecast error over squared predictable component error. The predictable component error is the error made if we use the part of the earnings forecast that we can predict using a statistical model. This ratio shows whether the inclusion of the unpredictable component results in an improvement or not. The last two rows show the average and median for the percentage of the forecasts for which the ratio is smaller than 1. We show results for 18338 forecasters across 316 firms, for a total number of 272970 observations spread over seven periods in the year leading up to the earnings announcement. The seven periods roughly correspond to the periods around the quarterly earnings announcement (excluding the fourth quarter, which coincides with the announcement of the earnings of interest) and the four periods in-between.

	During Q1	Announcement Q1	During Q2	Announcement Q2	During Q3	Announcement Q3	During Q4
Average	0,455	0,414	0,632	0,613	0,797	0,789	0,921
Median	0,335	0,350	0,640	0,625	0,808	0,750	0,887
Standard Deviation	0,464	0,329	0,351	0,379	0,411	0,683	0,552
5% percentile	0,020	0,025	0,077	0,088	0,166	0,180	0,266
95% percentile	1,170	1,032	1,136	1,209	1,463	1,378	1,581
Percentage $<$ 1, average	69,7%	72,3%	64,4%	66,6%	60,2%	62,1%	58,1%
Percentage $<$ 1, median	69,0%	71,8%	62,8%	66,2%	60,0%	62,1%	57,1%

Is there an optimal size of the unpredictable component?

As the unpredictable component does seem to help for improved forecast accuracy we now want to know what kind of added value of an earnings forecaster makes the difference. For this, we regress the squared earnings forecast error on a constant, on the unpredictable component and on the squared unpredictable component. We do this three times, once using the raw data and twice using two standardization approaches. Standardization might be relevant as it may occur that earnings are difficult to forecast as the data may be unstable, and this could then have an effect on both the squared earnings forecast error and the unpredictable components. The first standardization employs the variance of the predictable component for a firm, whereas the second considers the variance of the unpredictable component. The results are presented in Table 5.9.

The left-hand columns of Table 5.9 show the unstandardized results, while the other columns concern the standardized results. In all cases, even across the estimation and evaluation samples, the result for the squared unpredictable component is the same. That is, the larger is the squared unpredictable component, the larger is the squared forecast error of the

Table 5.9 Regression of squared forecast error of the earnings forecasts on the unpredictable component and its square: $FE^2 = \beta_0 + \beta_1 UC + \beta_2 UC^2$. We do this for all 316 firms and 18338 forecasters simultaneously in two regressions, one for the estimation sample (n=146319) and one for the evaluation sample (n=126651). Next to OLS estimation of the above linear model, we also use two standardization methods to account for firm differences in the size of earnings and the uncertainty of earnings. Standardization 1 uses the variance of the predictable component per firm. Standardization 2 uses the variance of the unpredictable component per firm. Standard errors in parentheses.

		No standardization		Standardization 1		Standardization 2	
Estimation sample	intercept	0,116	(0,003)	0,079	(0,002)	2,427	(0,034)
	UC	0,539	(0,020)	0,257	(0,012)	0,151	(0,044)
	UC^2	0,891	(0,011)	0,940	(0,008)	0,934	(0,012)
	R^2	0,044		0,085		0,040	
Evaluation sample	intercept	0,777	(0,024)	0,163	(0,004)	5,955	(0,061)
	UC	0,292	(0,054)	0,250	(0,015)	-1,182	(0,042)
	UC^2	0,980	(0,002)	1,006	(0,001)	0,879	(0,007)
	R^2	0,625		0,810		0,107	

earnings forecast. So, in general, earnings forecasts that are close to what a statistical model would predict are most accurate.

The story for whether an earnings forecast is better off by being higher or lower than the predictable component is not so clear, as can be seen from the right-hand side columns of Table 5.9. Using no standardization or the first standardization suggests that negative unpredictable components perform better (see the positive parameters on UC). However, the second type of standardization gives no relationship (in the estimation sample) or the opposite relationship (in the evaluation sample). In all cases, however, the parameter for UC^2 stays close to 1. So, we do not find evidence that systematically adjusting upwards or downwards leads to more accuracy.

How useful is the unpredictable component?

Table 5.10 presents our OLS-based estimation results of the regression of the actual (the true earnings observations) on various functions of the predictable and unpredictable components. We include interaction terms with the number of forecasts, as the predictable component might be more accurate when it is based on more forecasts. We also include interaction terms with the time until the announcement, as forecasts just before the announcement might have already incorporated all information into the predictable component, and as such leaving not much room for extra expertise of the earnings forecaster.

Table 5.10 A summary of results of the regression of the actual earnings on predictable and unpredictable component variables: $Actual = \alpha + \beta PCV + \gamma UCV$. PCV does not include only the predictable component itself, but also interaction terms of the predictable component with logNF, the logarithm of the number of forecasts on which Average Forecast is based at that moment, and with logTUA, the logarithm of the number of days until the announcement. In a similar way UCV is based on the unpredictable component and interaction terms of the unpredictable component with logNF and logTUA. We run these regressions for each firm separately (of the 316 firms) but pool the results of all 18338 forecasters. The total number of observations in the regressions across all firms is 146319 in the estimation sample and 126651 in the evaluation sample. We present several statistics (average, median, standard deviation, 90% interval) on the estimated parameters and also the average and median of the standard error of the parameters.

		Estimated coefficient					Standard error	
		Average	Median	Standard deviation	Bounds of 90% interval		Average	Median
Estimation sample	intercept	0,067	0,021	0,348	-0,243	0,509	0,034	0,020
	PC	1,060	1,099	3,815	-3,274	4,489	0,663	0,444
	PC x logNF	-0,018	-0,019	1,176	-0,980	1,240	0,233	0,152
	PC x logTUA	-0,020	-0,016	0,704	-0,582	0,834	0,122	0,081
	PC x logNF x logTUA	0,004	0,004	0,215	-0,249	0,182	0,043	0,027
	UC	2,400	1,506	24,167	-34,384	40,469	11,212	10,087
	UC x logNF	-0,663	-0,456	8,279	-13,757	12,491	4,005	3,460
	UC x logTUA	-0,198	0,028	4,458	-7,161	6,155	2,068	1,836
	UC x logNF x logTUA	0,082	0,029	1,543	-2,411	2,530	0,744	0,646
Evaluation sample	intercept	0,394	0,278	0,972	-0,498	1,704	0,067	0,046
	PC	0,054	0,711	5,900	-8,086	5,740	0,950	0,506
	PC x logNF	0,242	0,068	1,888	-1,590	2,451	0,326	0,174
	PC x logTUA	0,084	0,027	0,988	-0,959	1,272	0,172	0,095
	PC x logNF x logTUA	-0,024	-0,006	0,321	-0,367	0,343	0,059	0,033
	UC	-2,370	-2,213	43,991	-52,879	50,791	13,371	9,822
	UC x logNF	0,712	0,793	14,559	-18,921	16,258	4,661	3,402
	UC x logTUA	0,811	0,655	7,789	-9,058	9,518	2,435	1,832
	UC x logNF x logTUA	-0,224	-0,229	2,584	-3,003	3,002	0,853	0,631

Several results in this table are interesting. The estimated parameters for the predictable and unpredictable components in the estimation sample seem to suggest that they need to be made more important than what they are in the actual forecast. In the latter their weights are equal to 1 by construction, but the table suggests that alternative weights could be beneficial. Note that these larger weights are downplayed by the interaction terms with the number of forecasts and the time until announcement, which are two variables that are both strictly positive and have an associated negative parameter estimate. To visualize this findings, consider Figure 5.1 which shows the effective parameters for both components throughout the year, both in the estimation and in the evaluation sample. This figure demonstrates that the optimal weight for each of the components is always below 1. For the predictable component, the contribution is relatively stable throughout the year, whereas for the unpredictable component the contribution is highest at the beginning of the year.

Another result from Table 5.10 is that the predictable component parameters are all estimated more accurately than their unpredictable component counterparts. Also, and not unexpectedly, the estimation results are more reliable in the estimation sample than in the evaluation sample.

These results altogether suggest that the optimal contribution of the unpredictable component can be less than 1. Hence, in other words, perhaps the earnings forecasters are adding too much of their unobservable expertise on top of what a statistical model already could achieve. This is not to say that the contribution of this expertise should be set at 0, as the results in Table 5.11 clearly indicate that this unpredictable component matters. This table presents the results on the F-test for the joint statistical relevance of the four variables that are associated with the unpredictable component. In both samples, the median F-statistic is larger than 20, and the 5% based F-test rejects no significant effect in more than 90% of the cases. Hence, there are clear signs that the unpredictable component does add useful information.

The next step is to examine how much the unpredictable component actually contributes to forecast accuracy. Table 5.11 also shows the R^2 values when using the predictable component variables and also when additionally including the unpredictable component variables. The increase in the median R^2 is about 2 to 3 percent, which is not that much. On the other hand, the median R^2 using only the predictable variables is already around 90% so there is not much left to be explained.

Figure 5.1

The effective parameter of the predictable and unpredictable component in forecasting the actual earnings throughout the year (with earnings announcement at t=0), after filling in average actual values for the number of forecasts and time until announcement across all firms and all years in the estimation or evaluation sample.

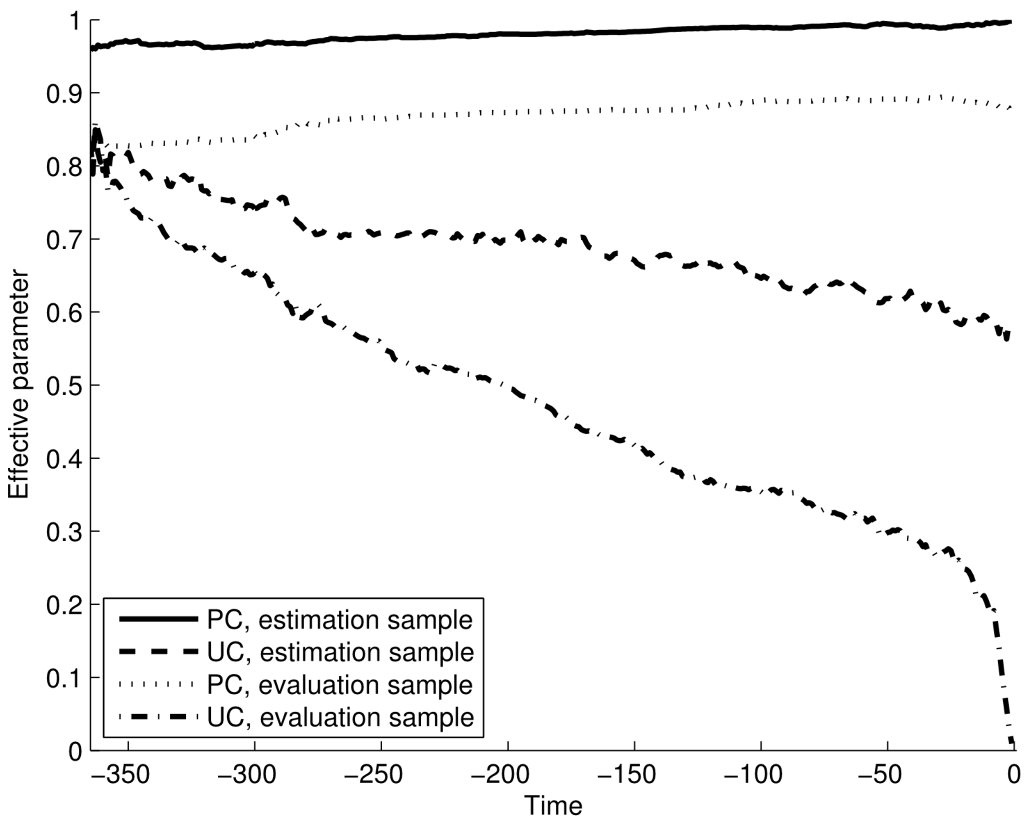


Table 5.11 A summary of results on the comparison between the regressions of (1) $Actual = \alpha + \beta PCV$, the actual earnings on only predictable component variables and (2) $Actual = \alpha + \beta PCV + \gamma UCV$, the actual earnings on both the predictable and unpredictable component variables. We run these regressions for each firm separately (of the 316 firms) but pool the results of all 18338 forecasters. The total number of observations in the regressions across all firms is 146319 in the estimation sample and 126651 in the evaluation sample. The F-statistic is based on the test for the joint significance of γ, the parameters of the unpredictable component variables, and the results for the associated p-value are shown in the column labelled p-value. Also shown are results for the R^2 values for both the restricted and the unrestricted model.

	Estimation sample				Evaluation sample			
	F-statistic	p-value	R^2 without UC	R^2 with UC	F-statistic	p-value	R^2 without UC	R^2 with UC
Average	33,776	0,015	0,868	0,893	29,292	0,039	0,817	0,850
Median	21,554	0,000	0,918	0,935	20,794	0,000	0,878	0,901
Standard Deviation	44,568	0,103	0,155	0,129	30,484	0,164	0,180	0,155
5% percentile	3,066	0,000	0,554	0,610	1,442	0,000	0,422	0,508
95% percentile	101,095	0,032	0,997	0,997	85,891	0,233	0,988	0,993
Significant at 5% level	96,2%				91,8%			

To complete our story on weights of the two components that could constitute an accurate forecast, we look at the comparison of the accuracy of optimally weighted forecasts to its constituent earnings forecasts and statistical model forecasts, and we report the results in Table 5.12.

From this table we see that the ratios that include the errors of the optimal forecast (OFE) are smaller than 1 for the samples for which the optimality is based on the parameters in the estimation sample for the estimation sample data or on the parameters found in the evaluation sample for the evaluation sample data. In contrast, when we use the estimation sample parameters for the evaluation sample data, the relevant ratio is larger than 1 compared to both the model forecast and the earnings forecast. This suggests that the optimal weights do not yield a stable performance over time and that they apparently need to be re-estimated on a regular basis.

5.5.2 Comparison across forecasters

The final part of our empirical analysis concerns an examination into which properties of individual earnings forecasters make them to display superior forecast performance. We first

Table 5.12 A summary of results on the median ratios between two squared errors. We use combinations of the following: FE^2, the squared earnings forecast error, PCE^2, the squared error of using the predictable component as forecast, and OE^2, the squared error of the optimal combination of the predictable component and unpredictable component variables. We calculate these median ratios for each firm separately (of the 316 firms) but pool the ratios of all 18338 forecasters. The total number of observations across all firms is 146319 in the estimation sample and 126651 in the evaluation sample. We calculate some ratios in the evaluation sample twice: once with the weights (used in the construction of the optimal forecast) as estimated in the estimation sample, and once using weights based on the evaluation sample itself.

	Estimation sample with estimation sample weights			Evaluation sample with estimation sample weights			Evaluation sample with evaluation sample weights	
	$\frac{EFE^2}{PCE^2}$	$\frac{OFE^2}{PCE^2}$	$\frac{OFE^2}{EFE^2}$	$\frac{EFE^2}{PCE^2}$	$\frac{OFE^2}{PCE^2}$	$\frac{OFE^2}{EFE^2}$	$\frac{OFE^2}{PCE^2}$	$\frac{OFE^2}{EFE^2}$
Average	0,609	0,669	1,499	0,638	2,878	7,463	0,570	1,123
Median	0,655	0,532	0,877	0,631	1,107	2,165	0,311	0,591
Standard Deviation	0,304	1,206	4,425	0,571	5,703	18,399	3,097	5,330
5% percentile	0,071	0,059	0,279	0,061	0,142	0,642	0,031	0,164
95% percentile	1,031	1,371	3,437	1,207	10,102	33,621	0,943	1,757

look at the key aspects that make earnings forecasters outperform a statistical model, and next we zoom in on optimal properties of the added expertise of the forecaster.

How to find outperforming earnings forecasters?

To investigate which earnings forecasters do best, we introduce a new measure, which is the balanced relative difference defined by $BRD(E,P) = \frac{EFE^2 - PCE^2}{EFE^2 + PCE^2}$, where EFE refers to the forecast error of the earnings forecaster and PCE refers to the forecast error of the predictable component. The variable to be explained concerns the data in the evaluation sample. Table 5.13 presents the results of a regression of the balanced relative difference between the earnings forecasts and the predictable component thereof (in the evaluation sample) on an intercept, the ratio of the squared unpredictable component to the squared predictable component and three balanced relative differences. These latter three variables are the BRD(E,P) itself and the $BRD(U,P) = \frac{UC^2 - PCE^2}{UC^2 + PCE^2}$ and the $BRD(O,P) = \frac{OFE^2 - PCE^2}{OFE^2 + PCE^2}$, where UC denotes the unpredictable component and OFE refers to the optimal forecast. Three variables show significant results. First, the BRD(E,P) in the evaluation sample is significantly related to its previous value in the estimation sample. So, the past track record seems to have predictive value for the future track record. Next to this, the previous value of the relative size of the unpredictable component to the predictable component and to the previous value of BRD(E,P) also contain predictive information.

Table 5.13 The results for the regressions to predict better analysts in the evaluation sample using variables in the estimation sample. This is based on 1835 forecasters (since we only include forecasters with a minimum of 10 observations in both sample periods) with a total of 52236 forecasts in the estimation sample and 36403 forecasts in the evaluation sample. We put the data across all firms in one regression. We use two interpretations for what a better earnings forecaster is: a forecaster who has a smaller forecast error compared to the predicted component ("better performing") and a forecaster whose associated optimally constructed forecasts have smaller forecast errors compared to the predicted component error ("having more information"). These might overlap if the forecasters with more information also use them well (so if the optimal forecast is similar to the earnings forecast), but there could also be forecasters that do not use their information well, which is why we separate these measures. In these regressions we use the balanced relative difference: $BRD(x,y) = \frac{x-y}{x+y}$ with x and y being combinations of E (for the earnings forecast error, EFE^2), P (for the predictable component error, PCE^2), O (for the optimal forecast error, OFE^2) and U (for the squared unpredictable component, UC^2). As performance variable, we use $BRD(E,P)$, while we use $BRD(O,P)$ as information variable. The variables to be explained are measured in the evaluation sample, while the regressors are measured in the estimation sample. Standard errors are shown in parentheses.

	Variable to explain			
	$BRD(E,P)$		$BRD(O,P)$	
intercept	-0,174	(0,020)	-0,003	(0,022)
$\frac{UC^2}{PC^2}$	1,152	(0,398)	0,244	(0,440)
BRD(U,P)	-0,098	(0,029)	-0,148	(0,032)
BRD(E,P)	0,407	(0,035)	0,302	(0,038)
BRD(O,P)	0,042	(0,026)	0,266	(0,029)

These results suggest that the forecasters who will predict best in the evaluation sample are those that have predicted best in the estimation sample (autoregressive), who have a small unpredictable component relative to the predictable component and who have a small unpredictable component relative to the error of the predictable component. Of these three, the autoregressive type variable has the highest statistical significance.

We can now use the above regressions to produce forecasts for the median balanced relative difference of each forecaster. Next, we can then compare the actual errors of the 50% forecasters that who we predict to have the best performance to the 50% forecasters who we predict to perform the worst. The ratio of the median squared error of the best 50% to the median squared error of the worst 50% turns out to be 0.600. Also, the predicted probabilities of having a negative balanced relative difference, which are the probabilities of outperforming the statistical model, are on average 80.8% and 61.9% for the best and worst half, respectively. This indicates that it is indeed possible to select a subset of all forecasters who will perform better in future.

Which forecasters have most expertise?

We use a similar approach to investigate whether it is possible to select forecasters who have more useful information in their unpredictable component, where we define this situation as where the optimal forecast performs best. We again use balanced relative differences. The variable to be explained now is BRD(O,P) in the evaluation sample. The results are presented in Table 5.13 in the right-hand side panel.

Again, the autoregressive type variable is statistically most significant, whereas the other two significant regressors are the other two balanced relative differences, that is, BRD(U,P) and BRD(E,P). Hence, the forecasters with the most useful information (meaning low values of BRD(O,P)) in the evaluation sample are those with the most useful information in the estimation sample, who are most accurate in the estimation sample and, surprisingly, who have a large unpredictable component.

When we compare the actual optimal forecast errors of two groups that are predicted to have the most and the least information, we get that the relative median squared optimal forecast error is 0.566. Hence, it is indeed possible to select a subset of the forecasters that contains those who have more informative unpredictable components.

Do the best performing forecasters have most expertise?

One may now wonder if there is an overlap between the best-performing forecasters and those who have most (unobservable) expertise. To investigate this, we calculate the hit rate, which is the percentage of cases in which a forecaster is categorized in the same cluster for both measures. It so turns out that this hit rate is 85.4%, which to us indicates that the question in the title can be answered affirmatively.

5.5.3 Comparison within forecasters

In this last subsection, we take a look at individual forecasts and compare their properties to other forecasts by the same earnings forecaster. Indeed, a large unpredictable component might be much more surprising if produced by a forecaster who usually has small unpredictable components than if produced by someone else who usually has large unpredictable components. In the first case, this single forecast may be based on unique and important information, but it might also mean that the forecaster quoted at random.

Table 5.14 A summary of results on the correlation between three balanced relative difference variables and two unpredictable component variables, calculated per individual forecaster. This is based on 4541 forecasters, with 90190 forecasts in the estimation sample and 28000 in the evaluation sample. We calculate the correlation of the UC variables with three balanced relative difference variables, with the definition $BRD(x,y) = \frac{x-y}{x+y}$ with x and y being combinations of E (for the earnings forecast error, EFE^2), P (for the predictable component error, PCE^2) and O (for the optimal forecast error, OFE^2).

		Correlation with $\lvert UC \rvert$			Correlation with UC^2		
		BRD(E,P)	BRD(O,P)	BRD(O,E)	BRD(E,P)	BRD(O,P)	BRD(O,E)
Estimation sample	Average	-0,096	-0,185	-0,116	-0,069	-0,166	-0,121
	Median	-0,125	-0,214	-0,133	-0,124	-0,210	-0,146
	Standard Deviation	0,322	0,279	0,273	0,335	0,278	0,272
	5% percentile	-0,582	-0,598	-0,535	-0,559	-0,556	-0,526
	95% percentile	0,454	0,324	0,359	0,506	0,354	0,358
Evaluation sample	Average	-0,146	-0,122	0,033	-0,125	-0,105	0,030
	Median	-0,173	-0,129	0,049	-0,177	-0,116	0,046
	Standard Deviation	0,477	0,490	0,472	0,481	0,487	0,468
	5% percentile	-0,953	-0,967	-0,877	-0,954	-0,963	-0,863
	95% percentile	0,790	0,852	0,909	0,816	0,835	0,892

We compute for each forecaster the correlation between the size of the unpredictable component and the three balanced relative difference variables, which are BRD(E,P), BRD(O,P) and BRD(O,E), of which the latter is defined as $BRD(O,E) = \frac{OFE^2 - EFE^2}{OFE^2 + EFE^2}$. This last measure can be interpreted as how much the earnings forecaster could improve his forecast if he would optimally use his available information. As measures for the size of the unpredictable component we use both $|UC|$ and UC^2. A summary of the results across all forecasters is presented in Table 5.14.

Table 5.14 shows that only negative correlations are found. The negative correlations between the size variables of UC and BRD(E,P) indicate that large unpredictable components for a particular forecaster are associated with a better performance compared to the statistical model. Similarly, the negative correlations with BRD(O,M) show that large unpredictable components are associated with more information in that unpredictable component. Finally, the negative correlations with BRD(O,E) show that large unpredictable components are associated with a better optimal forecast than the actual earnings forecasts, and thus with less optimal use of the unpredictable component by the earnings forecaster.

Table 5.14 also covers the evaluation sample. In this case, not all correlations are negative. The correlations with BRD(E,P) and BRD(O,P) result in the same qualitative conclusion as before, that is, large unpredictable components are associated with a better performance and more information than smaller unpredictable components produced by the same forecaster. The positive correlation of BRD(O,E) with the size of the unpredictable component indicates that in this case, on average, large unpredictable components coincide with less opportunity to set optimal weights in the combination of the unpredictable component with the model forecast. This finding may be due to unstable weights over time.

Overall, we find that in general small-sized added expertise of an earnings forecaster to a statistical model forecast is beneficial. At the same time, when an individual forecaster with a track record of small-sized added expertise suddenly makes large adjustments, then this usually leads to an increased accuracy of the earnings forecasts.

5.6 Conclusion

Earnings forecasts are an important factor in the decision making process of investors. In this chapter we have shown that earnings forecasts can be predicted, which allows investors to already incorporate the predictable part in their investment decision. Furthermore, we also show that the unpredictable part of an earnings forecast can be used. One way to use it, is to improve the forecast based on just the predictable part. This is especially beneficial in the beginning of the year. Another use of the predictable and unpredictable components concerns the selection of earnings forecasters, which can be relevant if an investor wants to ignore the forecasters with a poor track record. We have shown that there is persistence in the performance of forecasters compared to the predictable component, that is, earnings forecasters who perform better in our estimation sample, also perform better, on average, in the evaluation sample. Similarly, the information in the unpredictable component, that can be used to improve the optimal forecast, is also persistent, that is, earnings forecasters whose unpredictable components are more useful in the estimation sample also have this property in the evaluation sample.

In general, large unpredictable components seem to be a bad sign, as they are associated with large relative forecast errors. This is not the case if the earnings forecaster normally produces small unpredictable components. In that case, a large unpredictable component is a sign of both good performance and more useful information in this unpredictable component.

5.A Small-sample error correction method

In Section 5.4 we use a model to predict earnings forecasts, including a correction to account for small-sample error. In this appendix, we present the mathematical definition of the model.

We will describe the regression by using the familiar notation

$$y_{i,j,t} = X_{i,j,t}\beta_j + \varepsilon_{i,j,t}, \tag{5.1}$$

with subscript i denoting the individual forecaster, j the firm for which the earnings are forecasted and t the day on which the forecast is produced. The parameter coefficients are denoted by β_j, which is a vector consisting of $\beta_{j,k}$ for $k = 1, .., K$, one parameter for each variable in $X_{i,j,t}$. We will let the vector of coefficients differ per firm, but not per individual nor for different time periods. Also, the error variance $\sigma^2_{\varepsilon,j}$ differs per firm. This is the model without the small-sample error correction.

Now we introduce the small-sample error correction, for which we use a latent variable model for β_j. We can use this latent variable model to correct estimates that have been estimated with a small number of data points and which are thus less accurate and more prone to outliers. These estimates can be adjusted towards the overall mean of that respective parameter, and we do that in such a way that estimates based on more than thousand observations are hardly affected. As a necessary assumption for this model we use

$$\beta_j \sim N(\beta^*, \Sigma_\beta) \tag{5.2}$$

which means that the latent parameter vector β_j (the estimated parameters for firm j) is related to the overall mean parameter vector β^*. For simplicity, we will assume the covariance matrix Σ_β to be diagonal. Then we employ the following steps:

1. The elements of β^* and Σ_β are estimated by taking the weighted average and weighted variance of all individual estimates.

2. We update each individual estimate by taking a weighted average:

$$\beta_{j,k}^{(u)} = w_{j,k}\beta_k^* + (1 - w_{j,k})\beta_{j,k} \tag{5.3}$$

$$w_{j,k} = \frac{\frac{1}{\sigma_{\beta,k}}}{\frac{1}{\sigma_{\beta,k}} + \frac{n_k}{\sigma_{\varepsilon,j}}} \tag{5.4}$$

   The weights are calculated using the inverses of the latent variable standard deviation and the standard error of the regression, as these determine how accurate both sources of information on the $\beta_{j,k}$ estimate are.

We will repeat (5.3) and (5.4) until convergence.

Chapter 6

Stochastic levels and duration dependence in US unemployment

Based on de Bruijn and Franses (2015). Authors contributed according to a 95% / 5% split.

Abstract

We introduce a new time series model that can capture the properties of data as is typically exemplified by monthly US unemployment data. These data show the familiar nonlinear features, with steeper increases in unemployment during economic downswings than the decreases during economic prosperity. At the same time, the levels of unemployment in each of the two states do not seem fixed, nor are the transition periods abrupt. Finally, our model should generate out-of-sample forecasts that mimic the in-sample properties. We demonstrate that our new and flexible model covers all those features, and our illustration to monthly US unemployment data shows its merits, both in and out of sample.

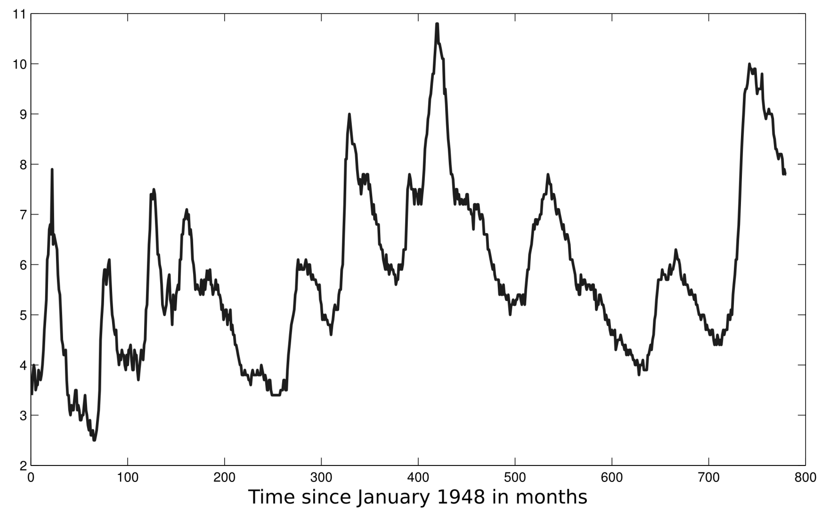


Figure 6.1 Monthly unemployment in the United States in the period 1948 to 2012.

6.1 Introduction

In this chapter we introduce a new time series model that can capture the properties of data as exemplified by monthly US unemployment data as depicted in Figure 6.1. Clearly the data show nonlinear features, as the increases in unemployment during economic downswings are much steeper than the decreases during economic prosperity. At the same time, the levels of unemployment in each of the two states do not seem fixed, nor are the transition periods abrupt. Finally, one would want a time series model that can generate out-of-sample forecasts that mimic the in-sample properties. Our new and flexible model will be shown to cover just those aspects, and our illustration to the very same US unemployment data shows its merits.

To analyze time series data with regime switching features, a natural starting point is the familiar Markov Switching model. Markov Switching (MS) models (Hamilton, 1989) are suitable for data fluctuating around two levels, where these levels associate with each of the two states. In the initially proposed MS models, the occurrence of a state at time t is only dependent on the state at time $t-1$ and it is governed by transition probabilities p_{ij}, the probability of switching from state i to state j. In these initial models, the probabilities are fixed across the sample. Figure 6.2a shows a simulated example of a MS model with 2 states. Recent applications of such basic MS models include Kim (2009), Bauwens et al. (2010),

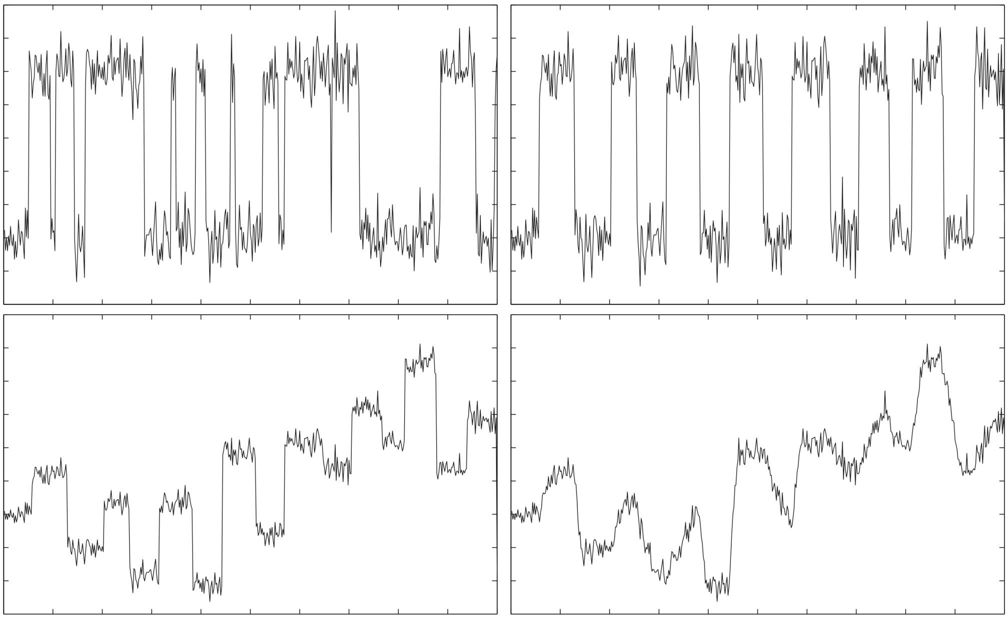


Figure 6.2 Four stylized series to characterize the gradual steps from a standard MS model to our model.

Nalewaik (2011), Cunningham and Kolet (2011), Guérin and Marcellino (2013) and Chen and Schorfheide (2013).

One of the features of such a basic MS model is that it is not capable of dealing with cyclicality, which entails for example that the forecasts produced by a basic two-state MS model are monotonically convergent. One modification to account for cyclical behavior is to introduce duration dependence in the transition probabilities (Diebold and Rudebusch, 1990; Durland and McCurdy, 1994). For example, one could let the probabilities be $p_{ij} = F(\beta_0 + \beta_1 d_t)$ with F any CDF and d_t the duration of the current spell at time t, which is the period since the last state switch. This means that the probability of switching now has become dependent on the duration of the spell. Figure 6.2b shows an example of a duration dependent MS model with 2 states, in which the duration dependence is positive, that is, the probability of switching out of a state increases the longer the time series has been within that state. Such MS models are implemented in, among others, Sichel (1991), Lunde and Timmermann (2004), Lam (2004), Layton and Smith (2007), Castro (2010) and Cunningham and Kolet (2011).

A common feature of the two MS models so far is that the mean in each of the states is fixed. For example, if one were to use a two-state MS model to model unemployment from 1980 to now, one assumes that the states of high unemployment and low unemployment imply the same mean for both the eighties and the current decade. This assumption might not be considered as realistic, which is also clear from Figure 6.1. Hence, one may wish to allow the means to be stochastic. In our model we alleviate the restriction by allowing the means to alternate in such a way that the difference is different each time. An example of the kind of data that can be generated by such a model can be found in Figure 6.2c.

Finally, as already indicated, and is visible from Figure 6.1, the transitions from one state to the other may not be immediate, as there might be a gradual transition from the previous state mean to the new state mean. At the same time, the time it takes to switch from one regime to the other may also not be the same across the entire sample, and hence we wish to allow the transition process to be a stochastic process too. Figure 6.2d shows a simulated time series with these properties, and it is clear that the pattern starts to come close to the unemployment data graphed in Figure 6.1.

To wrap up, in the present chapter we propose a Markov Switching model with duration dependence, and with stochastic processes for the levels in each of the states and for the transitions from one regime to the other. We will illustrate our new model for monthly US unemployment from 1948 to 2012 (see Figure 6.1). As this new model is computationally demanding and also requires the data to be informative, we run various simulations to see how well parameters can be estimated.

The outline of the remainder of this chapter is as follows. In Section 6.2 we will formally introduce our new MS model. Using simulations, we will highlight some of the data characteristics that align with this model. Section 6.3 discusses estimation of the parameters and inference of the latent variables. We will also show how one can produce forecasts using this model, and we will demonstrate that these forecasts continue the in-sample data features into the future. In Section 6.4 we illustrate our model and the associated estimation procedure for US unemployment. We also outline how our new model can be used for real-time monitoring of the data. In Section 6.5 we simulate from the DGP using the estimates from Section 6.4 to investigate how accurate the parameters can be estimated. Finally, we conclude in Section 6.6 with some final remarks and thoughts for further research.

6.2 Modeling and simulations

We first reintroduce the Markov Switching model. We choose for a notation that will make it easier to describe extensions. Denote the time series of interest as y_t with $t = 1, \ldots, T$. We relate y_t to the two state means μ_0 and μ_1, and the differences between the data and the state means are contained in the error term $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. In the basic MS model the probability of being in one of the states $s_t \in 0, 1$ depends only on the state in the previous time period, and in the basic MS model these probabilities are assumed as fixed. The so far discussed properties of the model can be captured by the following expressions:

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{6.1}$$

$$\mu_t = \mu_{s_t} \tag{6.2}$$

$$P(s_t = j | s_{t-1} = i) = p_{ij}, \quad i, j \in 0, 1 \tag{6.3}$$

The time period of the κ^{th} switch is described by the variable τ_κ, so $s_{\tau_\kappa} \neq s_{\tau_\kappa - 1}$ $\forall \kappa$ and $s_t = s_{t-1}$ for all other t.

A first extension of this model that is often considered in practice concerns allowing for transition probabilities that are duration dependent. One way to do this is to use a link function to transform a linear function of the state duration to a variable between 0 and 1. One possible link function that is commonly used in various applications is the standard normal CDF $\Phi(.)$. If the duration of the relevant state at time t is captured by the variable d_t, then a first extension amounts to replacing (6.3) by $P(s_t = j | s_{t-1} = i) = \Phi(\beta_0 + \beta_1 d_t)$. This makes the transition probability dependent on the duration d_t, but note that it also assumes that switches from one state to the other occur in a similar way. One way to incorporate a possible difference in such switching behavior is to consider

$$\begin{aligned} P(s_t = j | s_{t-1} = i) &= \Phi(\beta_0 + \beta_1 I[s_{t-1} = 1] + \beta_2 d_t + \beta_3 d_t I[s_{t-1} = 1]) \\ &= \Phi(\beta' D_t) \end{aligned} \tag{6.4}$$

with $I[.]$ the indicator function and $D_t' = [1 \quad I[s_{t-1} = 1] \quad d_t \quad d_t I[s_{t-1} = 1]]$. If both β_1 and β_3 are zero, then there is no difference in the switching behavior across the two states. If both β_2 and β_3 are zero, there is no duration dependence, and the model reduces to the basic MS model as in equations (6.1)-(6.3).

In practice we need to estimate the value of d_t when $t = 1$ as we do not know whether a switch has occurred just before the start of the sample or whether it has occurred a long time before that. For this, we introduce the variable d_1^*, and we will set d_1 equal to that, and calculate the other d_t's by either adding 1 to the previous value, or by resetting it to 1.

Next, we propose a second extension by allowing for stochastic state means, instead of fixing these to two values μ_0 and μ_1. To allow for this in the notation, we introduce the difference between the type of state s_t, which is either H (high) or L (low), and the sequential number of the state at time t, for which we extend the variable κ to have an index $\kappa_t = 0, 1, \ldots$. The state type will switch around each time the data enter a new state. We assume the following relation between two subsequent states means, that is

$$\mu_{\kappa_t} \sim N(\mu_{\kappa_t - 1} + \Delta\mu^* \times (-1)^{I[s_t = L]}, \sigma^2_{\Delta\mu}) \tag{6.5}$$

This relation assumes that the new state mean on average differs $\Delta\mu^*$ from the previous state mean. This difference is however not fixed, so it is not exactly the same each time. Also, whether the change in the state mean is upwards or downwards depends on what type of state s_t will be associated with the new state mean μ_{κ_t}. We do not want new state means to be on the wrong side of the previous state mean (for example, having a state mean of type $s_t = H$ being lower than the directly preceding state mean of the low type). Therefore, we adjust the preceding relation to a Truncated Normal distribution with parameters for the bounds denoted by lb_{κ_t} and ub_{κ_t} as

$$\mu_{\kappa_t} \sim TN(\mu_{\kappa_t - 1} + \Delta\mu^* \times (-1)^{I[s_t = L]}, \sigma^2_{\Delta\mu}, lb_{\kappa_t}, ub_{\kappa_t}) \tag{6.6}$$

$$lb_{\kappa_t} = \begin{cases} -\infty & \text{if } s_t = L \\ \mu_{\kappa_t - 1} & \text{if } s_t = H \end{cases} \qquad ub_{\kappa_t} = \begin{cases} \mu_{\kappa_t - 1} & \text{if } s_t = L \\ \infty & \text{if } s_t = H \end{cases} \tag{6.7}$$

The observation mean μ_t can now be generalized from (6.2) to

$$\mu_t = \mu_{\kappa_t} \tag{6.8}$$

Our third and final extension concerns the stochastic linear transitions. We introduce the notation λ_{κ_t} to denote the time taken by the transition of state mean $\mu_{\kappa_t - 1}$ to μ_{κ_t}. The linear transition property indicates that between the start of the transition τ_{κ_t} and the end of the transition at $\tau_{\kappa_t} + \lambda_{\kappa_t}$, the state mean is not equal to either μ_{κ_t-1} or μ_{κ_t}, but a weighted sum of these, with weights dependent on the length of the transition period. To calculate weights, we can use the duration variable d_t. This results in partly replacing (6.8) by

$$\mu_t = \frac{d_t}{\lambda_{\kappa_t}}\mu_{\kappa_t} + (1 - \frac{d_t}{\lambda_{\kappa_t}})\mu_{\kappa_t-1} \tag{6.9}$$

We write "partly" because this replacement is only relevant for the cases in which $t \in [\tau_{\kappa_t}; \tau_{\kappa_t} + \lambda_{\kappa_t}]$. For the other cases, (6.8) remains valid, that is, $\mu_t = \mu_{\kappa_t}$ if $t \in [\tau_{\kappa_t} + \lambda_{\kappa_t}; \tau_{\kappa_t+1}]$. We impose a distribution on λ_{κ_t} that needs to be positive, which is why we use the lognormal distribution. We allow the precise distribution to be dependent on whether the switch is upwards or downwards, as there might be a difference in transition speed. For the upwards switch, we propose $\lambda^u_{\kappa_t} \sim LN(\lambda^*_u, \sigma^2_{\lambda,u})$. Similarly for the downwards switch, we assume $\lambda^d_{\kappa_t} \sim LN(\lambda^*_d, \sigma^2_{\lambda,d})$. Finally, we impose that the transition periods have come to an end before the next one starts. The latter amounts to the restriction

$$\tau_{\kappa_t+1} - \tau_{\kappa_t} \geq \lambda_{\kappa_t} \tag{6.10}$$

for all κ_t.

To wrap up, our new model reads as

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2_\varepsilon) \tag{6.11}$$

$$\mu_t = \begin{cases} \mu_{s_t} & \text{if } t \in [\tau_{\kappa_t} + \lambda_{\kappa_t}; \tau_{\kappa_t+1}] \\ \frac{d_t}{\lambda_{\kappa_t}}\mu_{\kappa_t} + (1 - \frac{d_t}{\lambda_{\kappa_t}})\mu_{\kappa_t-1} & \text{if } t \in [\tau_{\kappa_t}; \tau_{\kappa_t} + \lambda_{\kappa_t}] \end{cases} \tag{6.12}$$

$$P(s_t = j | s_{t-1} = i) = \Phi(\beta D_t) \tag{6.13}$$

$$D'_t = [1 \quad I[s_{t-1} = L] \quad d_t \quad d_t I[s_{t-1} = L]] \tag{6.14}$$

$$\mu_{\kappa_t} \sim TN(\mu_{\kappa_t-1} + \Delta\mu^* \times (-1)^{I[s_t=L]}, \sigma^2_{\Delta\mu}, lb_{\kappa_t}, ub_{\kappa_t}) \tag{6.15}$$

$$lb_{\kappa_t} = \begin{cases} -\infty & \text{if } s_t = L \\ \mu_{\kappa_t-1} & \text{if } s_t = H \end{cases} \qquad ub_{\kappa_t} = \begin{cases} \mu_{\kappa_t-1} & \text{if } s_t = L \\ \infty & \text{if } s_t = H \end{cases} \tag{6.16}$$

$$\lambda^u_{\kappa_t} \sim LN(\lambda^*_u, \sigma^2_{\lambda,u}) \qquad \lambda^d_{\kappa_t} \sim LN(\lambda^*_d, \sigma^2_{\lambda,d}) \tag{6.17}$$

$$\tau_{\kappa_t+1} - \tau_{\kappa_t} \geq \lambda_{\kappa_t} \tag{6.18}$$

Our new model includes 12 parameters that need to be estimated from the data, and these are σ_ε, $\Delta\mu^*$, $\sigma_{\Delta\mu}$, λ_u^*, $\sigma_{\lambda,u}$, λ_d^*, $\sigma_{\lambda,d}$, β_0, β_1, β_2, β_3, d_1^*. Additionally, we need to estimate $3\kappa_T + 1$ latent variables, that is $\mu_{\kappa_t}, \tau_{\kappa_t}, \lambda_{\kappa_t}$ $\forall\kappa_t$, plus the start state mean μ_0. The number of latent variables depends on the size of the sample and on the frequency of state switches. The other variables such as the observation mean μ_t or the state duration d_t can be directly calculated from the estimates of the latent variables.

Hypothetical data

To examine how time series data can look like if they are generated from the new model, we run a few simulations. We generate data from four data generating processes (DGPs). The reference DGP, for which the associated hypothetical data are plotted in the top left part of each upcoming graph, is based on the following parameter configuration, that is, $\sigma_\varepsilon = 1$, $\mu_0 = 0$, $\Delta\mu^* = 6$, $\sigma_{\Delta\mu} = 2$, $\lambda_u^* = \lambda_d^* = 2.5$, $\sigma_{\lambda,u} = \sigma_{\lambda,d} = 0.5$, $\beta = [-4\ 0\ 0.1\ 0]'$ and $d_1^* = 0$, where we set the sample size at $T = 500$. This corresponds with a duration-dependent model in which the transition behavior is the same for upward and downward switches. The other DGPs differ from the benchmark DGP each time for only a few parameters, that is, we consider (i) $\sigma_{\Delta\mu} = 1$, which amounts to a process with more similar-sized jumps between the state means, and thus this process is closer to a model without stochastic means. Next, we consider (ii) $\beta = [-3\ -3\ 0.12; 0]'$, for which the most important difference is $\beta_1 = -3$ instead of $\beta_1 = 0$. The intercept for downward switches is $\beta_0 + \beta_1$, thus a negative value for β_1 makes downward transitions take more time to initialize than upward transitions, which have just β_0 as intercept. Finally, we consider (iii) $\lambda_d^* = 4$, which means that the downward transitions take more time to complete than the upward transitions.

Figure 6.3 shows several simulated series using each of the four configurations of the parameters. The top-left graph shows three series generated using the reference DGP, with two series having a different μ_0 which ensures that these lines do not overlap. The state durations are relatively stable, which illustrates the duration dependence. Also, the stochastic means are evident from the fact that the level is not the same each time the data switch between states. Especially the top line shows a quickly changing mean.

The other graphs in Figure 6.3 show comparable but slightly different behavior. The top-right panel shows alternative (i), which incorporates a lower $\sigma_{\Delta\mu}$. This is visible as this

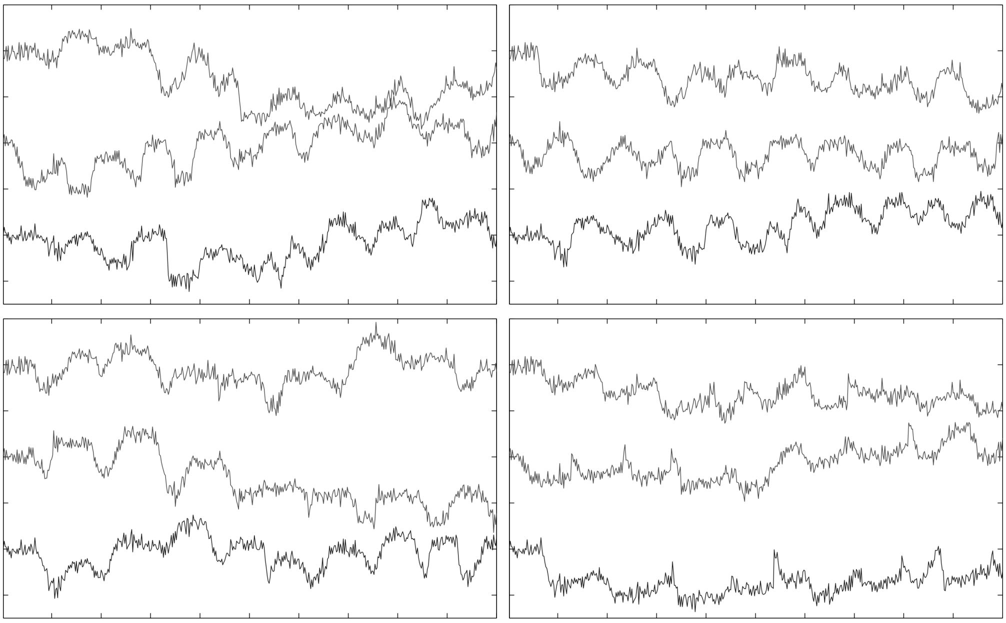


Figure 6.3 Three sample series of each of a reference DGP and three alternatives.

series shows less drifting behavior, and stays closer to its starting value. Alternative (ii) in the lower-left graph depicts the case in which the duration dependence is different per state. The graph clearly shows that more time is spent in the high states than in the low states. The lower-right graph represents alternative (iii), in which the transition time is different across the types of switches. The series in this graph clearly have a longer upwards transition time than downwards.

Figure 6.4 shows the first half of one of the series for each type, and then uses 10000 simulations to construct a prediction interval for the remainder of the series. The top-left graph shows how our new general model's short-term forecasts can capture the cyclical behavior rather well. Of course, for the longer term one eventually becomes less certain about whether there will be an upward or a downward state. Also, the confidence intervals increase as the aggregated effect of the unknown future information increases. The smaller $\sigma_{\Delta\mu}$ in alternative (i) is clearly visible in the top-right graph of Figure 6.4, as the intervals are smaller, especially for the longer term where the aggregated effect of $\sigma_{\Delta\mu}$ could have much of an effect. The intervals for both alternatives (ii) and (iii) seem to be comparable in size for all horizons relative to the reference DGP. This shows that the estimation of the stochastic mean shall be the most important part of the estimation process as this mean dominates the uncertainty in long-term forecasts.

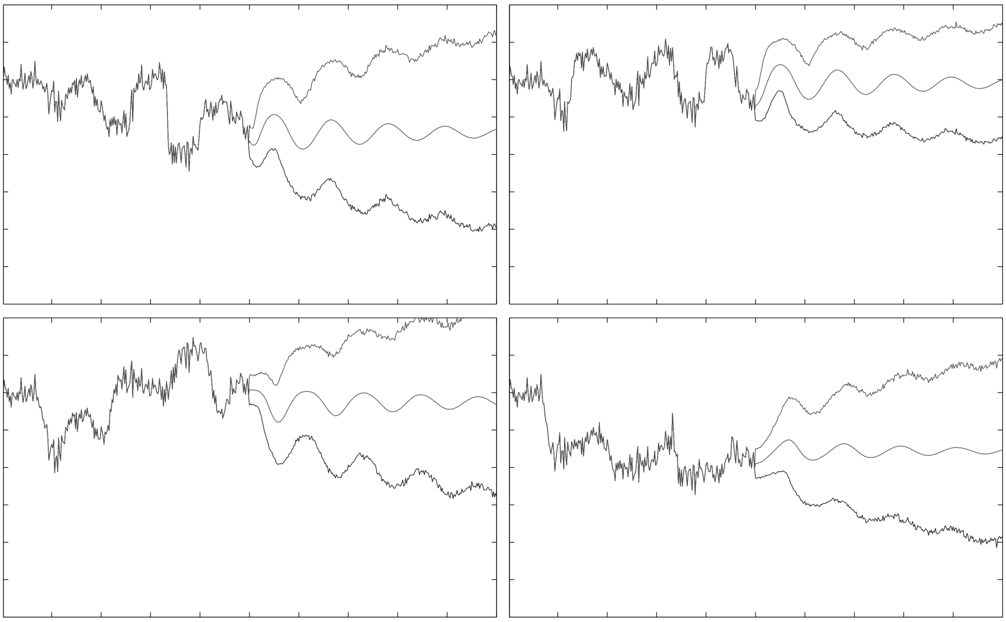


Figure 6.4 Simulated confidence intervals for the second half of one series for each of a reference DGP and three alternatives.

Finally, Figure 6.5 shows a simulated histogram of the length of a full cycle (switching up and down) for each parameter configuration, based on 50000 replications. While the first three histograms (with symmetric switching behavior between both types of states) have an approximately symmetrically distributed cycle length, the last histogram shows a much more asymmetric distribution. This shows that an asymmetric switching occurrence (alternative (ii)) and an asymmetric transition length (alternative (iii)) can have different effects on the cycle length.

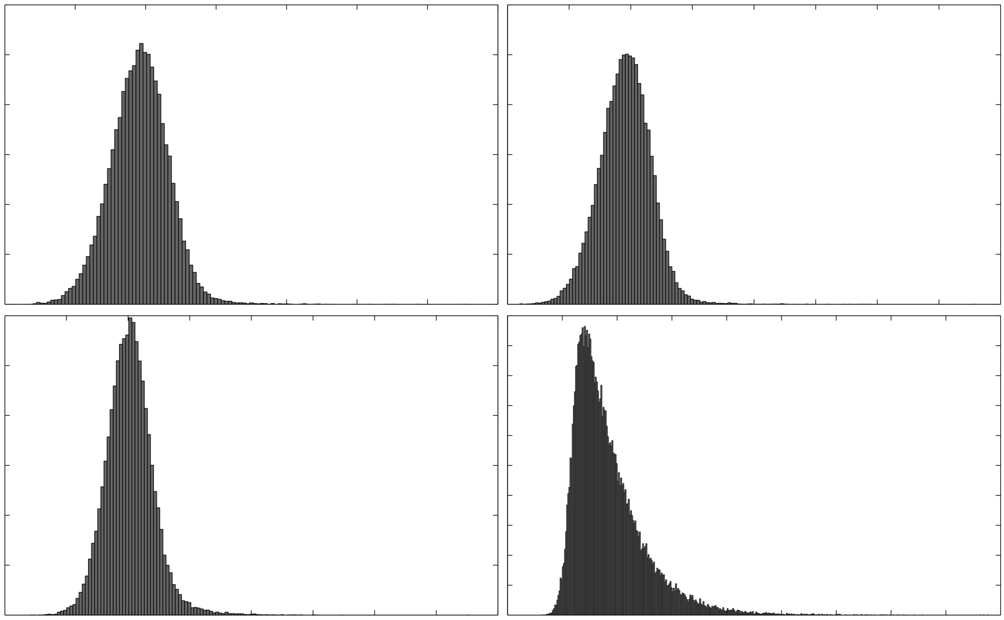


Figure 6.5 Simulated histograms for the length of a full cycle (consisting of an upward and a downward switch).

6.3 Estimation and inference

In this section we present the estimation routine for the estimation of the parameters and latent variables in our new general model. We also show how these estimates can be used for forecasting purposes.

Parameter estimation

We start with assuming that the number of state switches in the sample is known and is equal to K. To estimate the parameters we will make use of Gibbs Sampling with Data Augmentation. This method uses conditional distributions of parameters and latent variables given other parameters and latent variables to draw parameter values in an iterative manner. If chosen starting values of the parameters and latent variables are reasonably close to their posterior distribution, then after convergence the draws will be draws from the posterior distribution of the parameters. From these, one can take for example the mean to obtain a point estimate. We denote the draw in iteration m with the superscript (m). For example, $\mu_0^{(251)}$ denotes the value of the latent variable μ_0 in iteration round 251.

The conditional distributions need to be constructed for different sets of variables. Per set, one needs to be able to draw all parameters and latent variables within that set simultaneously (given the other parameters and latent values), so we need to group them accordingly. The sets that we create are as follows: $[\sigma_\varepsilon, \sigma_{\Delta\mu}, \sigma_{\lambda,u}, \sigma_{\lambda,d}, d_1^*]$, $[\Delta\mu^*, \lambda_u^*, \lambda_d^*, \beta]$, $[\lambda_1, \lambda_2, ..., \lambda_K]$, $[\mu_0]$, $[\mu_1]$, ..., $[\mu_K]$, $[\tau_1]$, $[\tau_2]$, ..., $[\tau_K]$. This amounts to $2K+4$ sets. We denote the sets using the notation $B_1, B_2, ..., B_{2K+4}$. To denote all sets except B_i, we use the notation B_{-i}. To denote the sets with a lower of higher index, we use $B_{<i}$ and $B_{>i}$.

The conditional distributions

We now present and discuss the conditional distributions of each individual set.

First we discuss the conditional distribution $B_1^{(m)}|B_{-1}^{(m-1)}$. This set consists of all the σ type parameters. The conditional distribution can be derived for each parameter separately, as these parameters do not directly affect each others' contribution to the likelihood function. To draw $\sigma_\varepsilon^{(m)}$, we calculate the residuals of (6.11) and denote these residuals as $\hat{\varepsilon}_t^{(m)}$. Then we have

$$\sigma_\varepsilon^{2(m)}|B_{-1}^{(m-1)} \sim IG(\sum_{t=1}^{T} \hat{\varepsilon}_t^2, T) \tag{6.19}$$

with IG denoting the Inverted Gamma distribution. Similarly, the other σ type variables can be drawn by rewriting their defining equations in a residual form, that is,

$$\sigma_{\Delta\mu}^2{}^{(m)}|B_{-1}^{(m-1)} \sim IG(\sum_{i=1}^{K}(|\mu_i - \mu_{i-1}| - \Delta\mu^*)^2, K-1) \tag{6.20}$$

$$\sigma_{\lambda,u}^2{}^{(m)}|B_{-1}^{(m-1)} \sim IG(\sum_{s_\kappa=H}(\lambda_\kappa^u - \lambda_u^*)^2, \sum_{\kappa=1}^{K} I[s_\kappa = H]) \tag{6.21}$$

$$\sigma_{\lambda,d}^2{}^{(m)}|B_{-1}^{(m-1)} \sim IG(\sum_{s_\kappa=L}(\lambda_\kappa^d - \lambda_d^*)^2, \sum_{\kappa=1}^{K} I[s_\kappa = L]) \tag{6.22}$$

Finally, for the draw of d_1^* we only need to observe the moment of the first switch $\tau_1^{(m-1)}$ and the duration dependence parameters $\beta^{(m-1)}$. The contribution to the likelihood of d_1^* is then the probability of switching at $t = \tau_1$ times the probability of not switching earlier, like

$$L(d_1^*) \propto \Phi(\beta^{(m-1)} D_{d_1^*+\tau_1-1}^{(m-1)}) \prod_{t=1}^{d_1^*+\tau_1-1} (1 - \Phi(\beta^{(m-1)} D_t^{(m-1)})) \tag{6.23}$$

We draw the new value for $d_1^{*(m)}$ from 0, 1, ... using the probabilities $p(j) = \frac{L(j)}{\sum_{i=1}^{\infty} L(i)}$.

Next, for the draws of $B_2^{(m)}|B_1^{(m)}, B_{>2}^{(m-1)}$, we again can split the set into parts that have no influence on each others' likelihood contribution. For $\Delta\mu^*|B_1^{(m)}, B_{>2}^{(m-1)}$ we can rewrite (6.15) to a normal distribution with mean equal to the average difference between subsequent state means and the variance equal to the sample mean variance, that is,

$$\Delta\mu^*|B_1^{(m)}, B_{>2}^{(m-1)} = N(\frac{1}{K}\sum_{\kappa=1}^{K}|\mu_\kappa - \mu_{\kappa-1}|, \frac{\sigma^2_{\Delta\mu}}{K}) \tag{6.24}$$

After applying a logarithmic transformation to (6.17), we can apply the same method to find the conditionals of λ_u^* and λ_d^*:

$$\lambda_u^*|B_1^{(m)}, B_{>2}^{(m-1)} = N(\frac{1}{K}\sum_{s_\kappa=H}\lambda_\kappa^u, \frac{\sigma^2_{\lambda,u}}{\sum_{\kappa=1}^{K} I[s_\kappa = H]}) \tag{6.25}$$

$$\lambda_d^*|B_1^{(m)}, B_{>2}^{(m-1)} = N(\frac{1}{K}\sum_{s_\kappa=L}\lambda_\kappa^d, \frac{\sigma^2_{\lambda,d}}{\sum_{\kappa=1}^{K} I[s_\kappa = L]}) \tag{6.26}$$

For the simulation of β, we rewrite (6.13) by introducing the latent variable z_t:

$$z_t = \beta D_t + \eta_t, \qquad \eta_t \sim N(0,1) \tag{6.27}$$

$$\begin{aligned} s_t \neq s_{t-1} \quad &\text{if } z_t \geq 0 \\ s_t = s_{t-1} \quad &\text{if } z_t < 0 \end{aligned} \tag{6.28}$$

Then, we simulate z_t from a truncated normal using the observation that there is a switch or not at time t, that is,

$$z_t^{(m)} \sim \begin{cases} TN(\beta^{(m-1)}D_t^{(m-1)}, 1, 0, \infty) & \text{if } s_t^{(m-1)} \neq s_{t-1}^{(m-1)} \\ TN(\beta^{(m-1)}D_t^{(m-1)}, 1, -\infty, 0) & \text{if } s_t^{(m-1)} = s_{t-1}^{(m-1)} \end{cases} \tag{6.29}$$

After that, we can simulate $\beta^{(m)}$ using a normal distribution based on the OLS regression of $z_t^{(m)}$ on $D_t^{(m-1)}$, like

$$\beta^{(m)} \sim N(\hat{\beta}_{OLS}^{(m)}, ({D_t}^{(m-1)\prime}{D_t}^{(m-1)})^{-1}) \tag{6.30}$$

For $B_3^{(m)}|B_{<3}^{(m)}, B_{>3}^{(m-1)}$, we make use of a Metropolis-Hastings sampler (MH; see Chib and Greenberg, 1995) for each individual λ_κ. For the MH sampler we need a candidate-generating function and a likelihood function for evaluation. For the candidate-generating function, we make use of (6.17) and (6.18) to draw from $g(\lambda_\kappa)$, which is a truncated log-normal distribution with parameters λ_u^* and $\sigma_{\lambda,u}$ for an upwards transition (λ_d^* and $\sigma_{\lambda,d}$ for downwards) and an upperbound equal to $\tau_{\kappa+1} - \tau_\kappa$. For the likelihood function we use the contribution of λ_κ to the likelihood, $f(\lambda_\kappa)$, which is based on its effect on μ_t via (6.12) and on its own likelihood via (6.17). The probability of accepting the candidate is

$$\alpha = \min(\frac{f(\lambda_\kappa^*)g(\lambda_\kappa^{(m-1)})}{f(\lambda_\kappa^{(m-1)})g(\lambda_\kappa^*)}, 1), \tag{6.31}$$

otherwise, $\lambda_\kappa^{(m)} = \lambda_\kappa^{(m-1)}$. As the definition $g(\lambda_\kappa)$ is part of the definition $f(\lambda_\kappa)$, this drops out of the fraction and thus we can also define $h(\lambda_\kappa) = \frac{f(\lambda_\kappa)}{g(\lambda_\kappa)}$, which only looks at the contribution to the likelihood based on its effect on μ_t via (6.12), and then use

$$\alpha = \min(\frac{h(\lambda_\kappa^*)}{h(\lambda_\kappa^{(m-1)})}, 1) \tag{6.32}$$

We use this approach for each individual λ_κ. The drawn value of one λ_κ will not affect the distribution of the other λ_κ's, which is the reason we can include them all in one set B_3.

For the sets $B_4^{(m)}|B_{<4}^{(m)}, B_{>4}^{(m-1)}$; ... ; $B_{K+4}^{(m)}|B_{<K+4}^{(m)}, B_{>K+4}^{(m-1)}$ we can use the same approach. For each B_i with $i = 4, \ldots, K+4$, we only draw μ_κ with $\kappa = i - 4$. For this, we at first will neglect the restriction that subsequent state means must alternately be higher and lower. We can then rewrite (6.11), (6.12) and (6.15) to a regression of y_t and $\Delta\mu^*$ on transformations of d_t and λ_{κ_t} and on 1 and -1, that is,

$$y_t = \begin{cases} \mu_{\kappa_t} + \varepsilon_t & \text{if } t \in [\tau_{\kappa_t} + \lambda_{\kappa_t}; \tau_{\kappa_t+1}] \\ \mu_{\kappa_t}\frac{d_t}{\lambda_{\kappa_t}} + \mu_{\kappa_t-1}(1 - \frac{d_t}{\lambda_{\kappa_t}}) + \varepsilon_t & \text{if } t \in (\tau_{\kappa_t}; \tau_{\kappa_t} + \lambda_{\kappa_t}) \end{cases} \tag{6.33}$$

$$\Delta\mu^* = \begin{cases} \mu_\kappa - \mu_{\kappa-1} + \zeta_\kappa & \text{if } s_\kappa = H \\ \mu_{\kappa-1} - \mu_\kappa + \zeta_\kappa & \text{if } s_\kappa = L \end{cases} \tag{6.34}$$

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \qquad \zeta_\kappa \sim N(0, \sigma_{\Delta\mu}^2) \tag{6.35}$$

We standardize all equations by dividing each term by the associated standard deviation and collect the variables on the right hand side (except for μ_κ itself) in the matrix X. We collect the μ_κ variables in the vector μ. Without the alternating increase or decrease in state mean, we could now sample all μ_κ using $\mu \sim N(\hat{\mu}_{OLS}, (X'X)^{-1})$. As we do not want to interfere in the alternating state mean restriction, we can sample the μ_κ one by one, conditional on all the others, using the standard formula for conditional normal distributions[1]. We restrict these to be either higher or lower than both the newly drawn previous state mean $\mu_{\kappa-1}^{(m)}$ and the next state mean of the previous iteration $\mu_{\kappa+1}^{(m-1)}$.

For the sets $B_{K+5}^{(m)}|B_{<K+5}^{(m)}, B_{>K+5}^{(m-1)}$; ... ; $B_{2K+4}^{(m)}|B_{<2K+4}^{(m)}, B_{>2K+4}^{(m-1)}$ we can use the same approach for each individual set. Each one of these sets consists of only one latent variable, that is, τ_{κ_t}. For this, we again make use of a MH-sampling method, for which we again need a likelihood-evaluating function f and a candidate generating function g. As candidate-generating function, we consider a discrete uniform distribution between the end of the previous transition ($\tau_{\kappa-1}^{(m)} + \lambda_{\kappa-1}^{(m)}$) and the end of the current transition ($\tau_\kappa^{(m-1)} + \lambda_\kappa^{(m)}$). As we want to let τ_κ only influence the start of the transition and not also the end, we adjust the transition length: $\lambda_\kappa^* = \lambda_\kappa^{(m-1)} + (\tau_\kappa^{(m-1)} - \tau_\kappa^*)$. For the evaluation of the likelihood, we need to observe that τ_κ influences the likelihood in two ways, that is, (i) changing the observation mean μ_t via (6.12), and (ii) the time periods during which the probit in (6.13) equals 1 (and thus also when it is 0). The adjusted λ_κ also affects (6.12), and along with that it contributes to the likelihood via (6.17). As our candidate-generating function is a uniform distribution, its pdf has the same value for each input in its support and it disappears from the equation to calculate the acceptance probability. This means that in this case the probability of accepting the candidate $\tau_\kappa^{(m)} = \tau_\kappa^*$ equals

$$\alpha = \min(\frac{f(\tau_\kappa^*, \lambda_\kappa^*)}{f(\tau_\kappa^{(m-1)}, \lambda_\kappa^{(m-1)})}, 1), \tag{6.36}$$

otherwise, $\tau_\kappa^{(m)} = \tau_\kappa^{(m-1)}$, and similarly for $\lambda_\kappa^{(m)}$.

Finally, we relax the assumption that the number of switches is known to be K. Instead, assume that the number of state switches in the sample is an element of {K, K+1, K+2}. That is, we allow for two states for which it is open whether they are part of the sample or not. For this, we will allow for two additional sets of state parameters, μ_{K+1}, μ_{K+2}, λ_{K+1}, λ_{K+2}, τ_{K+1} and τ_{K+2}. The state means and transition speeds might be partly simulated using the time series, if the associated switches occur before $t = T$. Otherwise, the simulated values will be entirely due to the distribution assumptions (6.13)-(6.18). Similarly, the simulation of

[1]If X_1 and X_2 are both multivariate normally distributed vectors with means μ_1 and μ_2, covariance matrices Σ_1 and Σ_2 and cross-covariance matrix Σ_{12}, then $X_1|X_2 \sim N(\mu_1 + \Sigma_{12}'\Sigma_2^{-1}(X_2 - \mu_2), \Sigma_1 - \Sigma_{12}'\Sigma_2^{-1}\Sigma_{12})$

the parameters β, $\Delta\mu^*$, $\sigma_{\Delta\mu}$, λ^*_u, λ^*_d, $\sigma_{\lambda,u}$ and $\sigma_{\lambda,d}$ now incorporate the additional two states and its latent values.

In practice, we would advise to set K such that K+1 equals the suspected number of switches in the sample. This way one can account for the situation that the expected last switch might not have happened, or otherwise, that an unexpected switch did occur. To evaluate the number of switches outside this interval, one could compare the estimated average likelihoods for different choices of K, possibly including a penalizing term for higher values of K.

Forecasting and real-time monitoring

We can of course also construct forecasts of y_t, $t > T$, and also for the associated latent values. For this one can fix the parameters to the mean or median of the draws obtained using the Gibbs sampling method. To account for parameter uncertainty, it is however better to draw the parameters used in forecasting from the entire posterior distribution. This can be easily done in practice by constructing a forecast in each iteration of the estimation process using the values of the parameters in that iteration. More elaborate sampling methods need to be used if one wants to forecast from a different starting point than at the end of the estimation sample. In that case, we would also need to re-estimate all the latent variables using only information up until that starting point.

We can forecast the observation mean μ_t by simulating from (6.13)-(6.18). To account for the restriction in (6.18), we first simulate the transition length λ, and then simulate the next state switch moment τ so that the λ is smaller than the difference between the two subsequent state switches. To obtain a full forecasting distribution, we also simulate the observation error ε_t using (6.11). If necessary, point and interval forecasts can be obtained using expected quantiles of this distribution.

Related to forecasting is the concept of real-time monitoring, in which the estimates of now relevant latent variables and short-horizon forecasts are updated each time a new data point becomes available. The approach for this is comparable to the one we use for forecasting, with the main difference that for real-time monitoring one needs to reapply the sampling procedure each time period. Due to the newly available sample point, the latent variables associated with the final states in the sample will change, and the best way to account for this is by re-estimating the previous latent variables and the parameters. This estimation process may take some time (depending on processor speed, programming efficiency and software), which can be in contrast to the goals of real-time monitoring, for which in fact one needs the updated estimates as quickly as possible after obtaining the new data point. For a quicker updating of the latent variables of the final states, one could fix the parameters and the latent variables of the previous states. That way, re-estimating will be done using less sets of parameters in the Gibbs sampling, which leads to less autocorrelation in the draws and thus to a smaller simulation sample that is necessary to obtain an accurate distribution.

6.4 Application to US unemployment

In this section we illustrate our model and estimation process on monthly unemployment in the United States for the period 1948 to 2012. Our estimation sample runs until 1992 (covering 540 months), which leaves 240 months for the forecast evaluation. For the estimation sample we restrict our K to be an element of $\{15, 16, 17\}$, based on visual inspection of the data.

Estimation results

Our estimation results are shown in Table 6.1 for which we have used in total 110000 iterations in the sampling process. After accounting for the burn-in period of 10000 iterations and a thinning factor of 45, this results in 2000 as-good-as-independent draws from the posterior distribution of the parameters and the latent variables.

The estimates of β_1 and β_3 show that there is asymmetry concerning the switching behavior, although it is not statistically significant. The estimate of β_2 shows that the upwards switch is not duration dependent, as zero is approximately in the middle of the HPD interval. In contrast, for the downwards switch the results show that there might be duration dependence as zero is on the border of the interval for $\beta_2 + \beta_3$. In fact, there are many ways in which we could have applied our thinning differently and that could have resulted in an interval that would not include zero.

Next, the estimate of $\Delta\mu^*$ shows that on average the high states and low states in unemployment differ about 2.8 percentage points. The ratio of $\frac{\Delta\mu^*}{\sigma_{\Delta\mu}}$ suggests that switches in

Table 6.1 Results on the posterior density of the main parameters of applying our model to US unemployment.

Parameter	Average	Standard Error	95% HPD interval	
σ_ε	0,313	0,010	0,292	0,332
β_0	-1,967	0,342	-2,688	-1,347
β_1	-0,594	0,476	-1,471	0,415
β_2	0,021	0,019	-0,016	0,055
β_3	-0,011	0,019	-0,051	0,023
$\Delta\mu^*$	2,806	0,280	2,244	3,333
$\sigma_{\Delta\mu}$	1,040	0,222	0,673	1,469
λ_u^*	2,375	0,265	1,857	2,909
λ_d^*	3,184	0,326	2,554	3,792
$\sigma_{\lambda,u}$	0,713	0,125	0,519	0,953
$\sigma_{\lambda,d}$	0,816	0,164	0,566	1,144
d_1^*	46,455	29,592	0	98
$\beta_0 + \beta_1$	-2,561	0,329	-3,218	-1,940
$\beta_2 + \beta_3$	0,011	0,006	-0,001	0,022
$\frac{\Delta\mu^*}{\sigma_{\Delta\mu}}$	2,810	0,605	1,663	3,982
$\Phi(\frac{0-\Delta\mu^*}{\sigma_{\Delta\mu}})$	0,008	0,015	0,000	0,033
$e^{\lambda_u^*+\frac{1}{2}\sigma^2_{\lambda,u}}$	14,558	4,612	7,857	23,568
$e^{\lambda_d^*+\frac{1}{2}\sigma^2_{\lambda,d}}$	36,669	17,654	15,177	63,317

the wrong direction are not likely even if we would remove the truncation, as the average of this ratio is 2.810, of which the negative (-2.810) is the 0.25-th percentile of the normal distribution. In fact, if we would calculate for each iteration the probability that one individual state mean change is in the wrong direction ($\Phi(\frac{0-\Delta\mu^*}{\sigma_{\Delta\mu}})$), then the HPD-region of these probabilities only runs to 0.033 and it has a mean of 0.008. The median is even lower, that is, 0.003. This shows that the truncation restriction is not of much influence on our estimation results.

Finally, we discuss the estimates of the transition speed parameters. The estimates of λ_u^*, λ_d^*, $\sigma_{\lambda,u}$ and $\sigma_{\lambda,d}$ seem to be quite accurate, considering their relatively small HPD intervals. The average transition length however is an exponential function of both of them, which can result in blowing up small differences to large effects. The average upwards transition length, $e^{\lambda_u^*+\frac{1}{2}\sigma_{\lambda,u}^2}$, is 14.6, which means that on average it takes almost five quarters for unemployment to transit from a local minimum to a local maximum, what would be called a recession. The opposite movement takes much more time, as the average for $e^{\lambda_d^*+\frac{1}{2}\sigma_{\lambda,d}^2}$ amounts to more than three years. This shows the familiar property for unemployment that an increase in unemployment is much quicker than a decrease.

Table 6.2 shows the average results for the latent variables. The first two columns present results on the timing of the start of the switch (τ_κ). Next, the results on the length of the transitions (λ_κ) follow, and the final two columns contain results on the state means (μ_κ).

The first fifteen states all have a state switch that falls entirely inside the sample that runs to $t = 540$. These state switches can thus be estimated quite accurately in most cases. Only the 10th state switch, which occurred around December 1971 ($t = 287$), has a standard deviation that is relatively high (4.150). For this state, the observations gradually start to decline and so there is no clear visual starting point of the transition. The model arrives at the same conclusion. The last two states do not fall entirely inside the sample. For $\kappa = 17$, the state switch is definitely after $t = 540$, but for $\kappa = 16$, this is not so clear. On average, the switch occurs after $t = 540$, but in fact in 50.5% of the iterations the state switch τ_{16} occurs on or before $t = 540$. This shows why it is important to account for multiple possible numbers of switches as it might be unclear whether a switch has occurred or not, and then one can account for both situations. As can be expected, both τ's that fall (partially) outside the sample are estimated less accurately than those inside the sample.

Table 6.2 Average results on the latent variables. For example, the sixth state switch is estimated to start around $t = 127$, the transition length for this switch is about 9 and the level of the sixth state is about 5.7.

	τ_κ		λ_κ		μ_κ	
κ	Mean	StDev	Mean	StDev	Mean	StDev
0					7,273	0,097
1	11,386	0,734	8,634	1,358	4,019	0,185
2	22,685	0,979	17,582	1,402	8,024	0,059
3	69,541	0,593	5,732	1,104	5,093	0,155
4	79,874	1,515	10,316	2,370	6,885	0,062
5	116,981	0,655	7,565	1,110	3,555	0,176
6	127,021	0,999	9,207	1,523	5,746	0,089
7	150,353	1,573	6,686	2,133	4,342	0,090
8	159,719	1,667	69,800	3,184	7,373	0,060
9	263,398	0,985	12,534	1,750	5,013	0,107
10	287,225	4,150	17,099	5,828	6,031	0,090
11	318,829	0,854	8,314	1,293	2,337	0,120
12	329,791	1,674	40,408	3,041	5,156	0,113
13	380,478	1,521	36,920	1,817	1,394	0,074
14	418,860	0,816	65,760	1,992	5,672	0,069
15	507,878	1,605	21,436	3,643	3,474	0,133
16	541,674	6,440	18,445	10,247	6,070	1,105
17	578,684	17,344	9,725	6,210	3,264	1,461

For the values of λ_κ in Table 6.2, we can observe an alternating pattern of high and low values. This is due to the different transition behavior for upwards and downwards transitions, which was also evident from the final two rows of Table 6.1. Again, of all in-sample states, state 10 has the most uncertain estimate of the latent variable. This makes sense, because if the start of the transition is unclear, then the length of the transition is also most likely unclear, as that depends on the start. Also, the two latent variables that fall outside the sample ($\lambda_1 6$ and $\lambda_1 7$) are both less accurate than all in-sample estimates, which is similar to the situation for the corresponding τ variables.

For the in-sample state means (μ_κ in Table 6.2), we see again an alternating pattern of high and low values, which follows directly from the relation between two subsequent state means. The two state means that are most close to each other on average are $\mu_9 \approx 5.013$ and $\mu_{10} \approx 6.031$, which still differ more than 1 percentage point. Also, there is no single value that is clearly less accurate than the others. The highest in-sample posterior standard deviation is 0.185 for μ_1, while the lowest is 0.060 for μ_8. The numbers in between seem to be about evenly spread out. Even the two state means around the tenth switch, μ_9 and μ_{10}, are both estimated quite accurately, in contrast to the situation for τ_{10} and λ_{10}. Again, both state means that are (partially) outside of the sample are estimated less accurately.

Based on our estimates of the parameters and the latent variables, some interesting graphs can be constructed. Figure 6.6 shows the original data along with the estimated mean and HPD intervals for the state means μ_t until the end of the estimation sample $t = 540$. We also present 7.5 years of forecasts, all constructed using information until $t = 540$. It can be seen that the in-sample intervals are much smaller than the out-of-sample intervals, as could well be expected. Quite noticeable is that the forecasts exhibit the same cyclical property of the model, and this is a feature that does not follow from a standard MS model with two states.

Figure 6.7 shows the probability of switching out of the state for both the upward and the downward switch, with on the horizontal axis the time already spent in that state. These probabilities are calculated for the entire posterior distribution of β and they incorporate the parameter uncertainty. For both state types, the probability of switching away increases the longer the duration of the state. For the downward switch this increase is obtained earlier than for the upward switch, as its line is mostly above the line of the upward switch.

Figure 6.8 shows the histogram of the durations of both states, accounting for both the uncertainty in β and for the completed transition restriction in (6.18). The histograms clearly

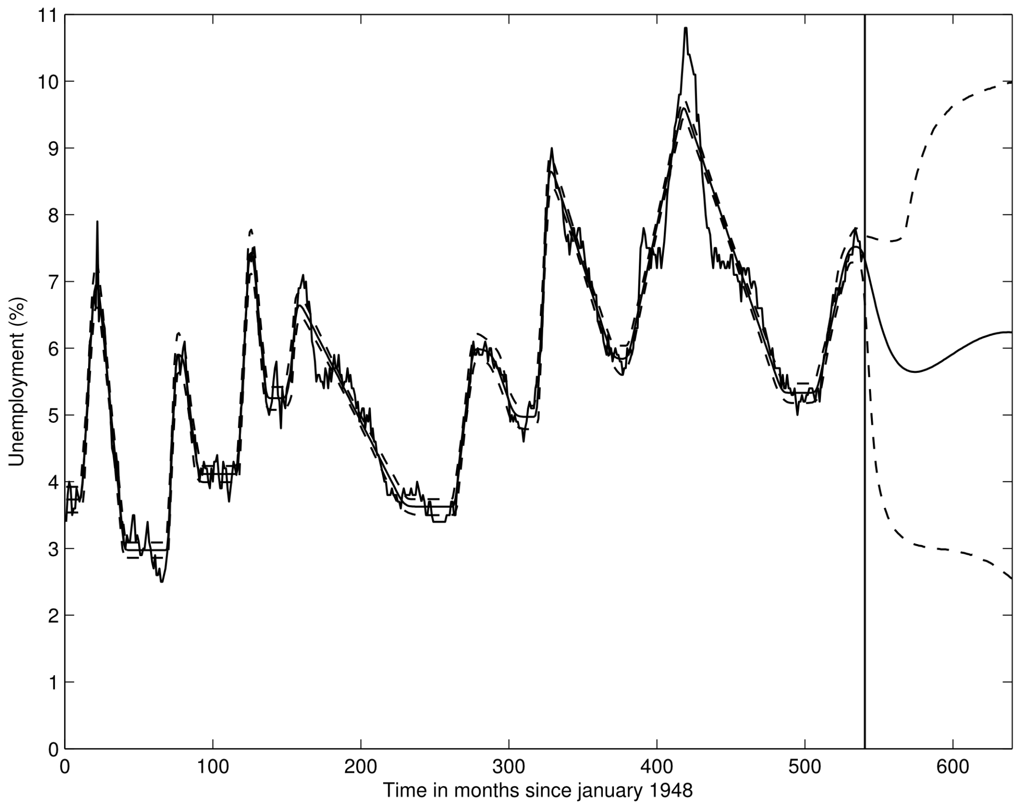


Figure 6.6 The mean and 95 % bounds for the estimation and forecasted observation means. Up until the vertical line at $t = 540$ the mean and bounds are in-sample and they are shown together with the original unemployment series for the period 1948 to 1992. For $t > 540$, mean and bounds of out-of-sample forecasts are shown, all constructed with the information set at $t = 540$.

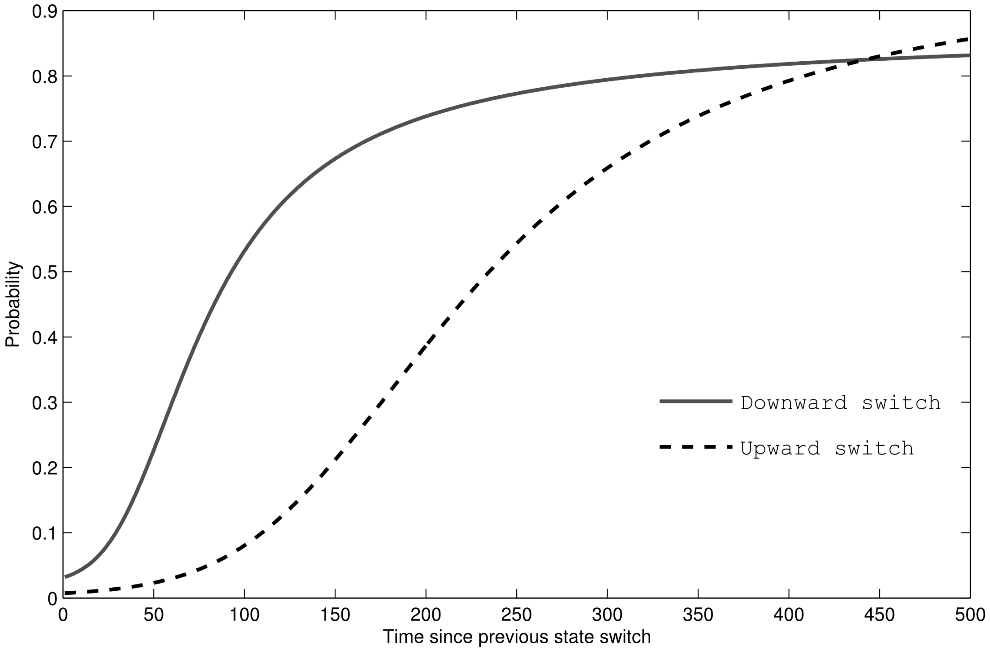


Figure 6.7 Estimated posterior probability of switching out of a state, accounting for the uncertainty in β by using the entire posterior distribution.

show that the downward switch happens faster than the upward switch. In fact, the averages amount to 2.3 years and 5.6 years, so upward switches occur after more than twice the time of downward switches, which means that high states last twice as short as the lower states. For both states an immediate switch is not impossible, although it is unlikely. As a comparison, for a standard MS model this histogram would be monotonically declining, and an immediate switch is actually the most likely.

Finally, Figure 6.9 shows the histogram of the length of a total cycle, thus incorporating both an upward and a downward switch using the values of β in the same iteration of the Gibbs sampler for both cycles. The pattern is comparable to the patterns in Figure 6.8, which was to be expected as it is the sum of both histograms in that figure. The average cycle length amounts to 7.8 years, which corresponds well with the common socio-economic cycle periods mentioned in de Groot and Franses (2012).

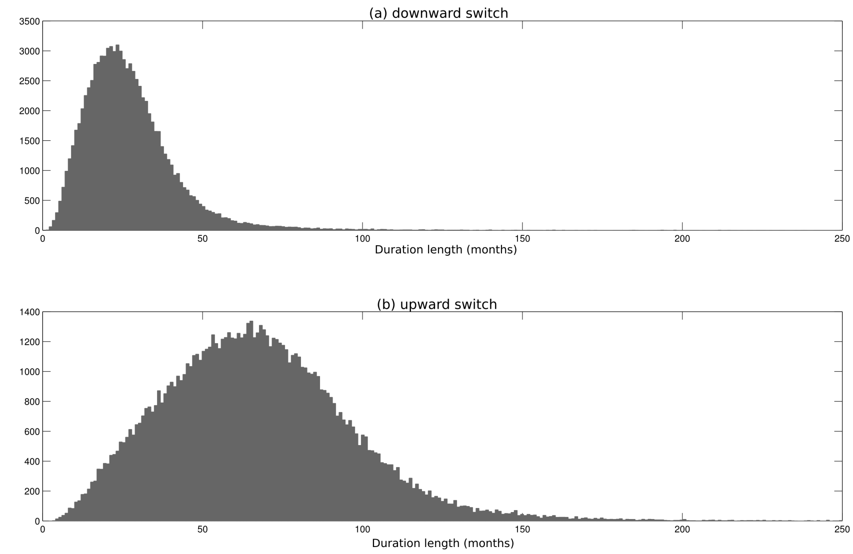


Figure 6.8 Estimated histograms
for the duration of one state, accounting for the uncertainty in β by using the entire posterior distribution.

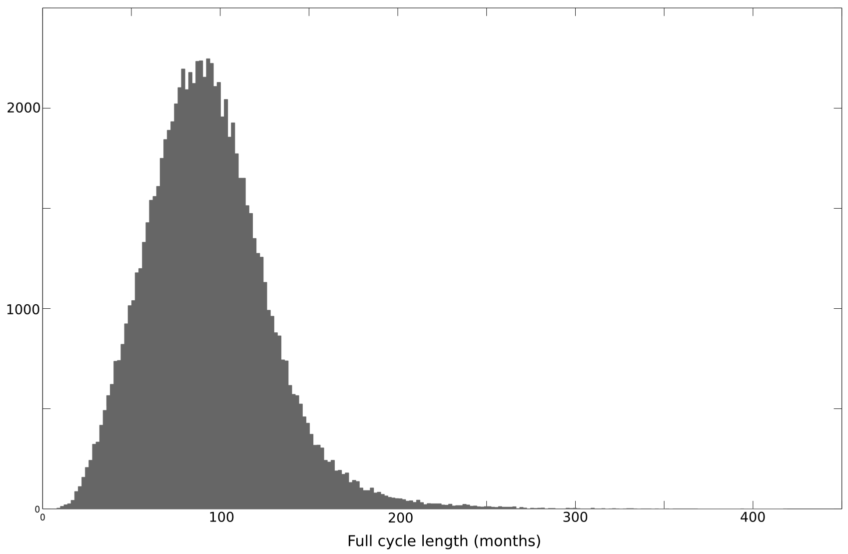


Figure 6.9 Estimated histograms for the length of one full cycle,
which incorporates both one upwards switch and one downwards switch together with their transition periods.

Real-time monitoring

An interesting application of our model is real-time monitoring, in which one investigates how much the estimates of latent variables get updated when new data points become available. We now illustrate how real-time monitoring can be applied using our model. For this we first re-estimate all latent variables using the data only until $t = 1$, only until $t = 2$, up to until $t = 540$, while fixing the parameters to the posterior distribution that has been estimated previously. For this part of the estimation process, we can use less draws as we can randomly draw the parameters from the posterior distribution, thereby decreasing the autocorrelation of the draws of the latent variables. That is why we decided to use 10000 draws with a burn-in sample of 1000 for each individual monitoring process.

This results in 540 estimated distributions for each individual latent variable. Using this we can calculate all sorts of statistics to see how the estimates evolve over time. For example, we can calculate the estimated mean of state 5, μ_5, using any possible set of information, to see from which point on the estimate of μ_5 does not vary any more. Or, we can find the width of the 95% HPD interval of μ_5, to see when the estimate of μ_5 first meets a certain accuracy requirement.

Figure 6.10 shows the average width (calculated over the entire sample period) of the 95% HPD interval for μ_t. This width has been calculated using information starting from $t - 100$ up until $t + 100$, thus using a total time period of over 16 years around each observation. This shows how the accuracy in estimating the observations' mean evolves when more information about that mean becomes available. It can be seen that the width decreases the fastest just before and just after time t. There is a slower decrease for the time periods that are well before t, while after a few months after t there is no information gain left anymore.

Next, Figure 6.11 shows the in-sample Root Mean Squared Error for each observation y_t, calculated using the in-sample forecast $\hat{y}_{t|t+h}$. As expected, for low h this value approaches σ_ε, while for high values of h the forecast error is larger. Even though this graph shows a different characteristic than Figure 6.10, they both show a similar pattern.

Forecasting

We have constructed forecasts for the last 240 hold-out observations, allowing for varying forecast horizons. We produce these forecasts starting at different starting points, for which

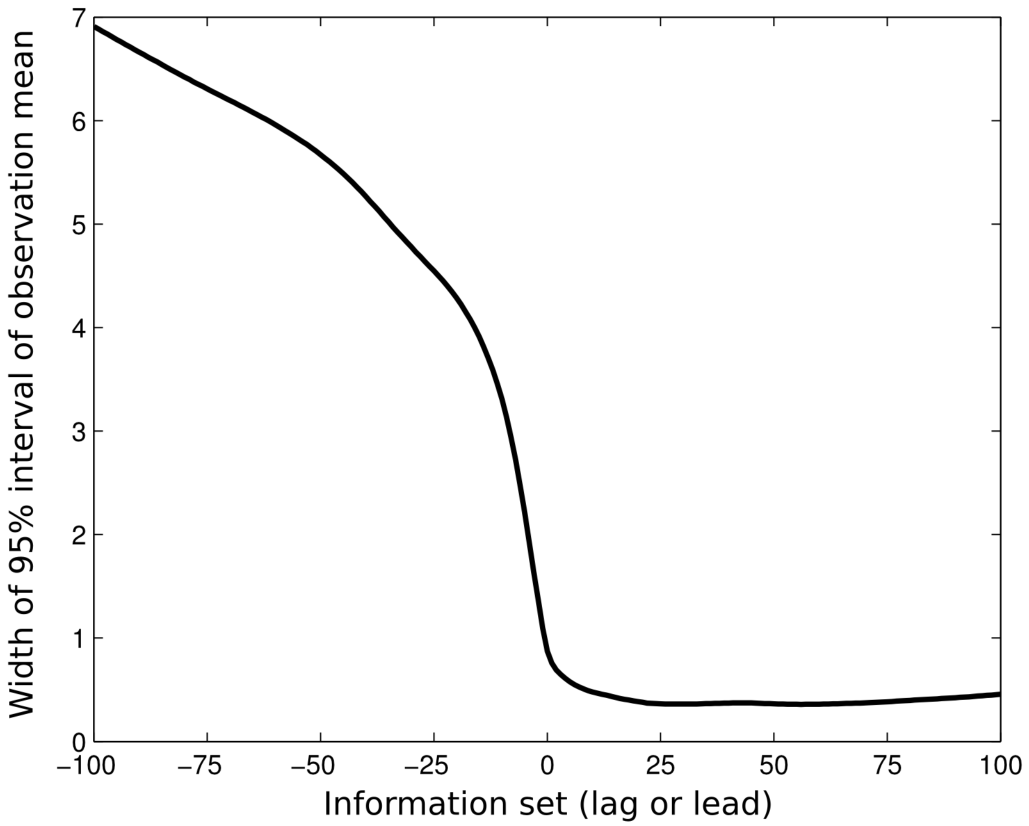


Figure 6.10 The average width of the 95 % interval of the observation mean for information sets that lead or lag with horizons up to $h = 100$.

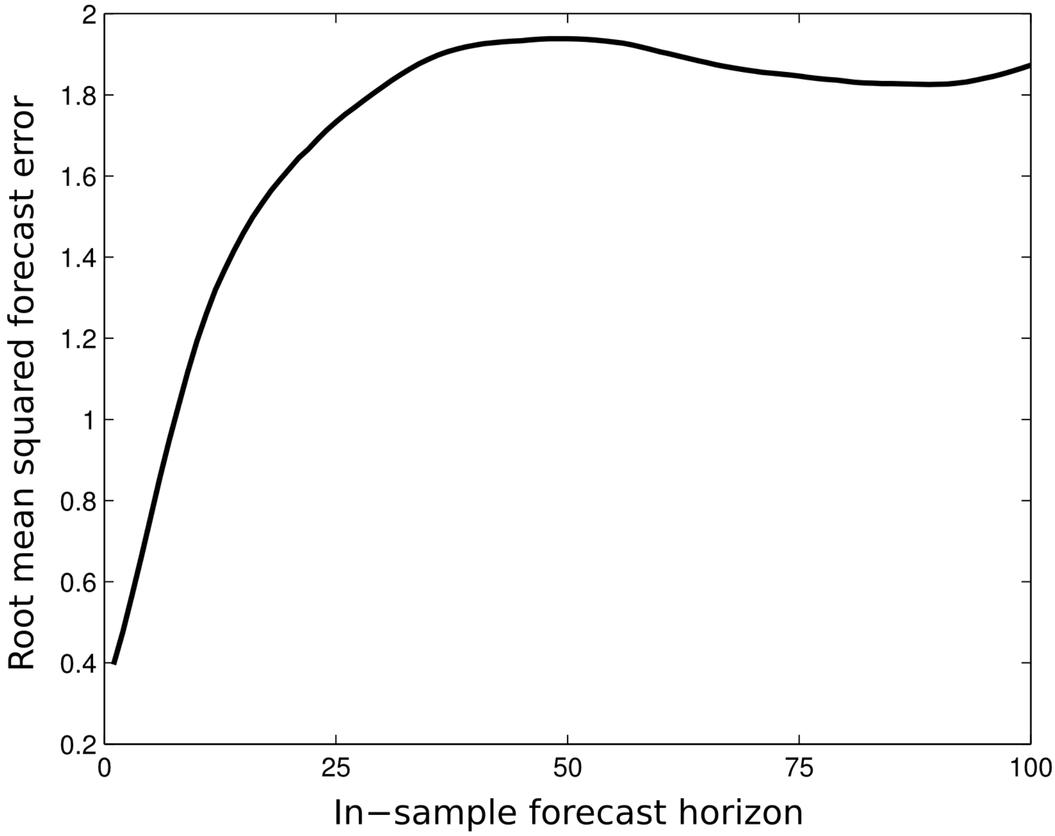


Figure 6.11 The in-sample root mean squared forecast error for horizons up to $h = 100$.

we each time need to re-estimate the latent variables to account for the new information. We will not update the parameter estimates, however, as these are based on the first 540 observations. For simulating the latent variables we have used the same approach as discussed for real-time monitoring, with again 10000 iterations per information set.

We will compare the forecasting performance of our model with several others. We use two simple forecasting models as a baseline comparison. The first of these is just taking the average of y_t for $t = 1, \ldots, 540$, so this means that we forecast the future using the sample average of the estimation sample. The second simple model is the Random Walk model, in which each forecast is just the most recently observed value at that time. In other words, $\hat{y}_{t+h|t} = y_t$. We also compare our model with two other Markov Switching models, namely the two-state and the three-state models. These models do not incorporate the duration dependence or stochastic means of our model, but instead they use fixed transition probabilities and fixed state means. To use an approach that comes close to our approach, we have estimated these models using Gibbs sampling on the same estimation sample and we forecast using draws from the entire posterior distribution of the parameters. For these simpler models, we have used 50000 iterations to estimate the parameters (after 100 burn-in iterations), and 5000 iterations for updating the estimates as the forecasting windows moves.

Table 6.3 shows the results of this forecast comparison. For each competing model, the RMSE (Root Mean Squared Error) has been calculated and then divided by the RMSE of our model. Values above 1 indicate that the alternative model performs worse, and values below 1 indicate the opposite. As can be seen, our model is the best model for the short-to-mid-term: for forecasting 6 months to 2 years ahead, our model beats the Random Walk model and a simple first order autoregression, and it is much better than both other MS models. For the other forecast horizons, our model is beaten by other models. On the very short horizon of 1 month, this defeat is no surprise, as our model takes no short-term information into account and the two models that beat our model here do. On the other hand, we easily outperform both other MS models again for 1 month ahead. On the longer term (more than 3 years), our model loses from all alternatives. For 3 to 5 years ahead the forecasting performance is not too bad as our model still beats the Random Walk, and no other model outperforms by more than 8%. For the two longest horizons however, our model performs very poorly. This might be because in the long-run our simulated state cycles are often out-of-sync with reality. This latter feature is studied next, using simulations.

Table 6.3 RMSE of other models
relative to our model. A value larger than 1 indicates that the corresponding model performs less than our model for the corresponding horizon, while a value smaller than 1 indicates that the model outperforms our model.

Forecast horizon	Fixed Mean	Random Walk	AR1	MS2	MS3
1 month	9,599	0,762	0,766	5,290	4,615
6 months	3,314	1,095	1,083	2,256	2,016
1 year	1,838	1,096	1,055	1,496	1,426
1.5 years	1,366	1,106	1,038	1,228	1,233
2 years	1,142	1,131	1,033	1,083	1,163
3 years	0,978	1,184	1,035	0,972	1,036
4 years	0,951	1,218	1,031	0,955	0,973
5 years	0,947	1,175	0,968	0,941	0,921
7.5 years	0,769	0,786	0,680	0,751	0,726
10 years	0,791	1,062	0,783	0,770	0,787

6.5 Simulations

In this section we investigate the accuracy of our estimation method in practically realistic situations. We do this by simulating multiple time series from a Data Generating Process (DGP) and by applying our estimation method to these time series.

As DGP we use the model and its parameters as presented in Section 6.4. We set the sample size in our simulations at 540, which is the length of the time series used in Section 6.4, and at 2160. The number of simulated time series in both cases is 400, while we use 10000 iterations in the estimation process after a burn-in of 1000 iterations.

Various summary statistics of the simulation results are shown in Table 6.4. As these results are calculated across 400 time series with 10000 iterations each in the estimation process, we report summary scores of various statistics, like for example the standard deviation of the mean. In that case, the standard deviation is calculated based on 400 values of the mean, of which each individual value is based on 10000 iterations.

Table 6.4 also shows the values of the parameters in the DGP. For most parameters, the mean of the mean and median of the median are quite close to the true DGP values. This holds true for $T = 540$ and even more so if $T = 2160$. Moreover, in all cases the spread in the point estimate (*StDev of mean*) is smaller when using more observations, which was to be expected. For some variables the *StDev of mean* for $T = 2160$ is about 10% of the same statistic for $T = 540$, while for most it is about a half. The least improvement is made

Table 6.4 Summary results of multiple parameter distributions estimated using simulated data from the same Data Generating Process for time series length $T = 540$ and $T = 2160$, with 400 time series for each case. Parameters have been estimated using Gibbs sampling with 10000 iterations (after a burn-in of 100 iterations). The table presents the mean and standard deviation of the mean of the posterior distribution, the median of the median of the posterior distribution, the mean of the standard deviation of the posterior distribution, and the mean of the quantile of the DGP parameter in the posterior distribution.

		$T = 540$					$T = 2160$				
	Data Generating Process	Mean of mean	StDev of mean	Median of median	Mean of StDev	Mean of Quantile	Mean of mean	StDev of mean	Median of median	Mean of StDev	Mean of Quantile
σ_ε	0,313	0,315	0,011	0,315	0,012	0,544	0,315	0,007	0,314	0,005	0,565
β_0	-1,967	-3,813	3,225	-2,919	1,105	0,072	-2,597	0,304	-2,568	0,255	0,030
β_1	-0,594	0,116	3,404	-0,354	1,462	0,599	-0,319	0,421	-0,352	0,338	0,693
β_2	0,021	0,094	0,162	0,053	0,047	0,835	0,037	0,016	0,036	0,010	0,821
β_3	-0,011	-0,068	0,163	-0,030	0,050	0,315	-0,025	0,017	-0,023	0,010	0,251
$\Delta\mu^*$	2,806	2,742	0,309	2,743	0,344	0,445	2,775	0,169	2,781	0,163	0,447
$\sigma_{\Delta\mu}$	1,040	1,080	0,247	1,023	0,279	0,482	1,045	0,111	1,036	0,119	0,482
λ_u^*	2,375	1,914	5,935	2,350	1,356	0,495	2,344	0,163	2,342	0,156	0,442
λ_d^*	3,184	3,151	0,364	3,155	0,358	0,469	3,161	0,187	3,176	0,165	0,467
$\sigma_{\lambda,u}$	0,713	0,726	0,130	0,675	0,202	0,416	0,683	0,055	0,672	0,063	0,321
$\sigma_{\lambda,d}$	0,816	0,764	0,139	0,709	0,200	0,302	0,738	0,062	0,728	0,070	0,185
d_1^*	46,455	49,858	18,771	49	24,947	0,497	51,329	15,241	50	28,022	0,516

for d_1^* (only a 16% drop to 84% of previous value), which might be explained by the fact that this value is for a large part affected by the initial observations only, and these cannot be influenced by the choice of T.

The β parameters are the only ones that are not always accurately estimated, especially if $T = 540$. The reason for the large differences in the mean of the posterior distribution (column *StDev of mean*) is the apparently small number of state switches in a sample of this length. On average the number of state switches is about equal to 15, as in the estimation sample of Section 6.4, which is already quite small, and for some of the simulated series this number dropped to as low as 10. Naturally, estimating the parameters in a probit model with four explanatory variables using only 10 observations with a '1' results in substantial uncertainty around the estimates. Having a longer time series obviously will make this situation less likely, and this is evident from the much lower values of *StDev of mean* for β when $T = 2160$. Also, for larger T the point estimates are on average much closer to the true values, even though there is still room for improvement.

The final two columns for both values of T provide an indication of how the DGP configurations are located as compared to the posterior. The *mean of StDev* shows how narrow (or wide) the estimated posterior is. Most posteriors are much more narrow for a higher T, with d_1^* as only exception as that parameter is not really affected by the value of T. The *mean of quantile* shows the quantiles where the true parameters are located. Both sample sizes show a similar pattern, so the sample size does not seem to matter much.

6.6 Conclusion

In this chapter we have introduced a new model that can deal with changing levels and cyclicality in time series. We have proposed a Markov switching model with two states that each have a stochastic mean, where the transition behavior of these states is governed by duration dependence and stochastic linear transition periods. We have shown with artificial data that data from this model have characteristics comparable with actual data. We have presented an estimation method that uses Gibbs Sampling with Data Augmentation, which also generates a density forecast. We have applied this estimation method to postwar monthly US unemployment and we have found that for two to three years ahead forecasts, our model has superior forecasting performance compared to a set of benchmark models. We have also shown, using a set of simulations, that the parameters of our model can be estimated quite accurately, granted that there is a sufficient number of state switches.

We envisage various potential extensions to our model and analysis. The major drawback of our model, as we have seen in the simulation exercise, is the potential difficulty in estimating parameters that fully depend on state switches. For many currently available samples of macroeconomic data, one typically encounters a limited number of state switches. One way to alleviate this is to jointly model several time series for which a common parameter can be assumed. Alternatively, the parameters of the different time series can be linked using an underlying joint distribution.

Applications to other than macroeconomic series can be particularly interesting. We then think of high-frequency financial time series data or data in marketing contexts, where different regimes may occur much more frequently. Other extensions could include implementing an autoregressive model to the stochastic-mean part of the model (6.15) or to the shocks in (6.11). An alternative distribution like a log-normal distribution instead of the truncated normal in (6.15) could also be considered.

Nederlandse samenvatting (Summary in Dutch)

Het is gebruikelijk om herzieningen in voorspellingen voor een bepaalde macro-economische variabele te evalueren door die herzieningen te regresseren op de herzieningen van een periode ervoor. Onder de veronderstelling van voorspelefficiëntie in zwakke vorm moet de correlatie tussen de huidige herziening en die van een periode terug gelijk aan nul zijn. De empirische bevindingen in de literatuur suggereren dat de stelling dat deze correlatie nul is zeer vaak verworpen wordt, waarbij men een positieve correlatie vindt (met als interpretatie een zogeheten onderreactie), maar ook een negatieve correlatie (die als overreactie geïnterpreteerd wordt). Hoofdstuk 2 betreft een methode om dergelijke correlaties op een eenvoudige en duidelijke manier te kunnen interpreteren, ook als ze niet nul zijn. De aanpak is gebaseerd op de veronderstelling dat de numerieke voorspellingen kunnen worden opgesplitst in een voorspelling vanuit een econometrisch model en de intuïtie van een expert. Het blijkt dat de interpretatie van het teken van de correlatie tussen de huidige herziening en die van één periode eerder afhankelijk is van het proces dat intuïtie stuurt, en daarnaast van de relaties tussen intuïtie en nieuws (ook wel: schokken in de numerieke voorspellingen). Hieruit volgt dat de geschatte correlaties niet rechtstreeks kunnen worden geïnterpreteerd als onderreactie of overreactie. Daarnaast is het aangetoond dat de onderreactie en overreactie kan veranderen over de tijd, mits de voorspellingen goed worden gemodelleerd en geïnterpreteerd. Het hoofdstuk bevat ook een empirisch voorbeeld om de bruikbaarheid van de voorgestelde werkwijze te laten zien.

Hoofdstuk 3 richt zich ook op voorspellingen van experts. Er is voldoende empirisch bewijs dat de wijze waarop aanpassingen door experts van modelvoorspellingen worden gedaan verbeterd kan worden. Eén manier om mogelijke verbetering te krijgen betreft het verstrekken van verschillende vormen van feedback aan de experts. Ook wordt vaak gevonden dat de experts geen homogene groep vormen. Hoofdstuk 3 introduceert een op data gebaseerde methodologie om latente clusters van voorspellers te onderscheiden, en past dit toe op een volledig nieuwe, grote dataset met daarin modelvoorspellingen, zowel inclusief als exclusief aanpassingen door experts, en realisaties van de betreffende variabelen. Met behulp van deze data kunnen twee clusters worden geïdentificeerd. Vervolgens worden de gevolgen voor het verstrekken van feedback aan deze clusters besproken.

Hoofdstuk 4 introduceert een nieuwe methode van dataverzameling om consumentenvertrouwen per dag per individu te meten. De aldus verkregen data maakt het mogelijk om de dynamische correlatie van deze index van consumentenvertrouwen statistisch te analyseren en conclusies te trekken over de overgangskansen. Dat laatste is niet mogelijk voor de maandelijkse index zoals die verzameld wordt door statistische bureaus op basis van herhaalde steekproeven. In een toepassing op het meten van het Nederlandse consumentenvertrouwen wordt aangetoond dat de additionele informatie in de nieuwe indicator helpt om consumentengedrag beter te voorspellen.

Hoofdstuk 5 betreft een analyse van ongeveer 300.000 winstverwachtingen, gecreëerd door 18.000 individuele voorspellers voor de winst van meer dan 300 S&P beursgenoteerde bedrijven. De analyse toont aan dat deze winstverwachtingen voor een groot deel te voorspellen zijn met behulp van een statistisch model dat openbaar beschikbare informatie bevat. Daarna wordt de focus verlegd op de onvoorspelbare componenten, die kunnen worden gezien als de eigen expertise van de voorspellers. Hieruit kan worden geleerd dat kleine afwijkingen ten opzichte van de voorspelde winstverwachtingen leiden tot een betere voorspelnauwkeurigheid. Op basis van prestaties in het verleden is het mogelijk om de toekomstige prestaties van de individuele voorspellers in zekere mate te voorspellen.

Tenslotte wordt in Hoofdstuk 6 een nieuw tijdreeksmodel geïntroduceerd dat patronen beschrijft die overeenkomen met wat de maandelijkse werkloosheidscijfers van de Verenigde Staten laat zien. Dit betreft bekende niet-lineaire eigenschappen, zoals steilere toenames in werkloosheid tijdens recessies dan de dalingen in expansies. Daarnaast bevat de data ook andere eigenschappen die minder vaak meegenomen worden, zoals twee stochastische werkloosheidsniveaus, met transitieperiodes tussen elke keer dat de werkloosheid omslaat van laag naar hoog. Daarnaast moet een geschikt model in staat zijn om voorspellingen buiten de steekproef te genereren met eigenschappen die lijken op de eigenschappen in de steekproef. Het nieuwe en flexibele model heeft al deze eigenschappen. Een toepassing op de maandelijkse werkloosheid in de Verenigde Staten laat zien dat dit model een toegevoegde waarde heeft, zowel om patronen in de steekproef te verklaren als om werkloosheid buiten de steekproef te voorspellen.

Bibliography

Abarbanell, J. S., Lanen, W. N., Verrecchia, R. E., 1995. Analysts forecasts as proxies for inventor beliefs in empirical research. Journal of Accounting and Economics 20, 31–60.

Abarbanell, J. S., Lehavy, R., 2000. Can stock recommendations predict earnings management and analysts' earnings forecast errors? Journal of Accounting Research 41, 1–31.

Abarbanell, J. S., Lehavy, R., 2003. Biased forecasts or biased earnings? the role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts. Journal of Accounting and Economics 36, 105–146.

Ager, P., Kappler, M., Osterloh, S., 2009. The accuracy and efficiency of the Consensus forecasts: A further application and extension of the pooled approach. International Journal of Forecasting 25, 167–181.

Amir, E., Ganzach, Y., 1998. Overreaction and underreaction in analysts' forecasts. Journal of Economic Behavior & Organization 37, 333–347.

Andrews, R. L., Currim, I. S., 2003. A comparison of segment retention criteria for finite mixture logit models. J. of Mark. Res. 40 (2), 235–243.

Ashiya, M., 2003. Testing the rationality of Japanese GDP forecasts: The sign of forecast revision matters. Journal of Economic Behavior & Organization 50, 263–269.

Ashiya, M., 2006. Testing the rationality of forecast revisions made by the IMF and the OECD. Journal of Forecasting 25, 25–36.

Athanasopoulos, G., Hyndman, R. J., 2011. The value of feedback in forecasting competitions. International Journal of Forecasting 27, 845–849.

Balzer, W. K., Sulsky, L. M., Hammer, L. B., Sumner, K. E., 1992. Task information, cognitive information, or functional validity information: Which components of cognitive feedback affect performance? Organizational Behavior and Human Decision 53, 35–54.

Batchelor, R., Dua, P., 1998. Improving macro-economic forecasts: The role of consumer confidence. International Journal of Forecasting 14 (1), 71 – 81.

Bauwens, L., Preminger, A., Rombouts, J., 2010. Theory and inference for a Markov switching GARCH model. The Econometrics Journal 13, 218–244.

Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12 (1), 9 – 24.

Berger, A. N., Krane, S. D., 1985. The informational efficiency of econometric model forecasts. Review of Economics and Statistics 67, 128–134.

Bolger, F., Önkal Atay, D., 2004. The effects of feedback on judgmental interval predictions. International Journal of Forecasting 20, 29–39.

Bolliger, G., 2004. The characteristics of individual analysts' forecasts in Europe. Journal of Banking & Finance 28 (9), 2283–2309.

Boulaksil, Y., Franses, P. H., mar 2009. Experts' stated behavior. Interfaces 39 (2), 168–171.

Bradshaw, M. T., Brown, L. D., Huang, K., 2013. Do sell-side analysts exhibit differential target price forecasting ability? Review of Accounting Studies 18, 930–955.

Bradshaw, M. T., Drake, M. S., Myers, J., Myers, L., 2012. A re-examination of analysts' superiority over time series forecasts of annual earnings. Review of Accounting Studies 17, 944–968.

Brown, L. D., 1993. Earnings forecasting research: its implications for capital markets research. International Journal of Forecasting 9 (3), 295–320.

Brown, L. D., Hugon, A., 2009. Team earnings forecasting. Review of Accounting Studies 14, 587–607.

Cameron, A. C., Trivedi, P. K., 2005. Maximum Likelihood. Cambridge University Press, Ch. 5.6, pp. 139–146.

Capistran, C., Timmermann, A., 2009. Disagreement and biases in inflation expectations. Journal of Money, Credit and Banking 41, 365–396.

Castro, V., March 2010. The duration of economic expansions and recessions: More than duration dependence. Journal of Macroeconomics 32 (1), 347–365.

Chang, C.-L., de Bruijn, B., Franses, P. H., McAleer, M., 2013. Analyzing fixed-event forecast revisions. International Journal of Forecasting 29 (4), 622–627.

Chang, C.-L., Franses, P. H., McAleer, M., 2011. How accurate are government forecasts of economic fundamentals? the case of Taiwan. International Journal of Forecasting 27 (4), 1066–1075.

Chen, F., F. D., Schorfheide, F., December 2013. A Markov-switching multifractal inter-trade duration model, with application to US equities. Journal of Econometrics 177 (2), 320–342.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings Algorithm. The American Statistician 49 (4), 327–335.

Cho, D. W., 2002. Do revisions improve forecasts? International Journal of Forecasting 18, 107–115.

Clement, M. B., Hales, J., Xue, Y., 2011. Understanding analysts' use of stock returns and other analysts' revisions when forecasting earnings. Journal of Accounting and Economics 51 (3), 279–299.

Clement, M. B., Tse, S. Y., 2005. Financial analyst characteristics and herding behavior in forecasting. The Journal of Finance 60 (1), 307–341.

Clements, M. P., 1997. Evaluating the rationality of fixed-event forecasts. Journal of Forecasting 16, 225–239.

Clements, M. P., Franses, P. H., Swanson, N. R., 2004. Forecasting economic and financial time-series with non-linear models. International Journal of Forecasting 20 (2), 169 – 183.

Cooper, R. A., Day, T. E., Lewis, C. M., 2001. Following the leader: a study of individual analysts' earnings forecasts. Journal of Financial Economics 61 (3), 383–416.

Cunha, F., Heckman, J., Navarro, S., 2006. Counterfactual Analysis of Inequality and Social Mobility. Stanford University Press, Ch. 11, pp. 290–348.

Cunningham, R., Kolet, I., 2011. Housing market cycles and duration dependence in the United States and Canada. Applied Economics 43 (5), 569–586.

Curtin, R., 1982. Indicators of Consumer Behavior: The University of Michigan Survey of Consumers. Public Opinion Quarterly 46, 340–352.

de Bruijn, B., Franses, P. H., 2012. Managing sales forecasters. Tinbergen Institute Discussion Paper Series 12-131/III.

de Bruijn, B., Franses, P. H., 2013. Forecasting earnings forecasts. Tinbergen Institute Discussion Paper Series 13-121/III.

de Bruijn, B., Franses, P. H., 2015. Stochastic levels and duration dependence in us unemployment. Erasmus University Rotterdam: Econometric Institute Report Series EI 2015-20.

de Bruijn, B., Segers, R., Franses, P. H., 2014. A novel approach to measuring consumer confidence. Erasmus University Rotterdam: Econometric Institute Report Series EI 2014-30.

de Groot, B., Franses, P. H., January 2012. Common socio-economic cycle periods. Technological Forecasting and Social Change 79 (1), 59–68.

DellaVigna, S., 2009. Psychology and economics: Evidence from the field. Journal of Economic Literature 47, 315–372.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum Likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39 (1), 1–38.

Diebold, F., Rudebusch, G., June 1990. A nonparametric investigation of duration dependence in the American business cycle. Journal of Political Economy 98 (3), 596–616.

Dominitz, J., Manski, C. F., 2004. How should we measure consumer confidence? Journal of Economic Perspectives 18, 51–66.

Dovern, J., Weisser, J., 2011. Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: An empirical comparison for the G7. International Journal of Forecasting 27, 452–465.

Durland, J., McCurdy, T., 1994. Duration-dependent transitions in a Markov model of U.S. GNP growth. Journal of Business & Economic Statistics 12 (3), 279–288.

Engle, R., Watson, M., 1981. A one-factor multivariate time series model of metropolitan wage rates. Journal of the American Statistical Association 76 (376), pp. 774–781.

Evans, D. S., 1987. The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. The Journal of Industrial Economics 35 (4), pp. 567–581.

Fildes, R., Goodwin, P., Lawrence, M., Nikolopoulos, K., 2009. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. International Journal of Forecasting 25 (1), 3–23.

Fok, D., Dijk, D. V., Franses, P. H., 2005. A multi-level panel star model for us manufacturing sectors. Journal of Applied Econometrics 20 (6), pp. 811–827.

Fok, D., Franses, P. H., 2007. Modeling the diffusion of scientific publications. Journal of Econometrics 139 (2), 376 – 390, the Econometrics of Intellectual Property.

Franses, P. H., 2014. Expert Adjustment of Model Forecasts: Theory, Practice and Strategies for Improvement. Cambridge University Press.

Franses, P. H., Kranendonk, H. C., Lanser, D., 2011. One model and various experts: Evaluating Dutch macroeconomic forecasts. International Journal of Forecasting 27, 482–495.

Franses, P. H., Legerstee, R., 2009. Properties of expert adjustments on model-based SKU-level forecasts. International Journal of Forecasting 25 (1), 35–47.

Franses, P. H., Legerstee, R., 2010. Do experts adjustments on model-based SKU-level forecasts improve forecast quality? Journal of Forecasting 29, 331–340.

Franses, P. H., McAleer, M., Legerstee, R., 2009. Expert opinion versus expertise in forecasting. Statistica Neerlandica 63, 334–346.

Givoly, D., Hahn, C., Lehavy, R., 2009. The quality of analysts' cash flow forecasts. The Accounting Review 84, 1877–1911.

Goodwin, P., 2000. Improving the voluntary integration of statistical forecasts and judgment. International Journal of Forecasting 16 (1), 85–99.

Goodwin, P., 2002. Integrating management judgment and statistical methods to improve short-term forecasts. Omega 30 (2), 127–135.

Goodwin, P., Fildes, R., 1999. Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? Journal of Behavioral Decision Making 12, 37–53.

Guérin, P., Marcellino, M., 2013. Markov-Switching MIDAS models. Journal of Business & Economic Statistics 31 (1), 45–56.

Hamilton, J., March 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.

Isengildina, O., Irwin, S. H., Good, D., 2006. Are revisions to USDA crop production forecasts smoothed? American Journal of Agricultural Economics 88, 1091–1104.

Isiklar, G., Lahiri, K., Loungani, P., 2006. How quickly do forecasters incorporate news? evidence from cross-country surveys. Journal of Applied Econometrics 21, 703–725.

Jegadeesh, N., Kim, W., 2010. Do analysts herd? An analysis of recommendations and market reactions. Review of Financial Studies 23 (2), 901–937.

Katona, G., 1951. Psychological Analysis of Economic Behavior. McGraw Hill.

Kim, C., January 2009. Markov-switching models with endogenous explanatory variables II: A two-step MLE procedure. Journal of Econometrics 148 (1), 46–55.

Kim, Y., Lobo, G. J., Song, M., 2011. Analyst characteristics, timing of forecast revisions, and analyst forecasting ability. Journal of Banking & Finance 35 (8), 2158–2168.

King, G., 1997. A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data. Princeton University Press.

King, G., Rosen, O., Tanner, M. A., 2004. Ecological Inference: New Methodological Strategies. Cambridge University Press.

Kish, L., Hess, I., 1959. A ”replacement” procedure for reducing the bias of nonresponse. The American Statistician 13 (4), 17–19.

Lam, P., February 2004. A Markov-Switching model of GNP growth with duration dependence. International Economic Review 45 (1), 175–204.

Lamont, O. A., 2002. Macroeconomic forecasts and microeconomic forecasters. Journal of Economic Behavior & Organization 48, 265–280.

Landsman, W. R., Maydew, E. L., Thornock, J. R., 2012. The information content of annual earnings announcements and mandatory adoption of IFRS. Journal of Accounting and Economics 53 (1–2), 34–54.

Laster, D., Bennett, P., Geoum, I. S., 1999. Rational bias in macroeconomic forecasts. Quarterly Journal of Economics 114 (1), 293–318.

Lawrence, M., Goodwin, P., O’Connor, M., Onkal, D., 2006. Judgmental forecasting: A review of progress over the last 25 years. International Journal of Forecasting 22 (3), 493–518.

Lawrence, M., O’Connor, M., 2000. Sales forecasting updates: How good are they in practice? International Journal of Forecasting 16, 369–382.

Layton, A., Smith, D., December 2007. Business cycle dynamics with duration dependence and leading indicators. Journal of Macroeconomics 29 (4), 855–875.

Legerstee, R., Franses, P. H., 2014. Do experts’ sku forecasts improve after feedback? Journal of Forecasting, 69–79.

Lim, K. H., O’Connor, M. J., Remus, W. E., 2005. The impact of presentation media on decision making: Does multimedia improve the effectiveness of feedback? Information and Management 42, 305–316.

Little, R. J. A., 1995. Modeling the drop-out mechanism in repeated-measures studies. Journal of the American Statistical Association 90, 1112–1121.

Liu, H., 2010. Dynamics of pricing in the video game console market: Skimming or penetration? Journal of Marketing Research 47 (3), 428–443.

Loungani, P., 2001. How accurate are private sector forecasts? cross-country evidence from consensus forecasts of output growth. International Journal of Forecasting 17, 419–432.

Ludvigson, S. C., 2004. Consumer confidence and consumer spending. Journal of Economic Perspectives 18, 29–50.

Lunde, A., Timmermann, A., 2004. Duration dependence in stock prices. Journal of Business & Economic Statistics 22 (3), 253–273.

Maloney, T., 1991. Unobserved variables and the elusive added worker effect. Economica 58 (230), pp. 173–187.

McLachlan, G. J., Peel, D., 2000. Finite Mixture Models. Wiley.

Mikhail, M. B., Walther, B. R., Willis, R. H., 1999. Does forecast accuracy matter to security analysis? The Accounting Review 74, 185–200.

Mikhail, M. B., Walther, B. R., Willis, R. H., 2004. Do security analysts exhibit persistent differences in stock picking ability? Journal of Financial Economics 74, 67–91.

Moffitt, R., 1993. Identification and estimation of dynamic models with a time series of repeated cross-sections. Journal of Econometrics 59, 99–123.

Nalewaik, J., April-June 2011. Incorporating vintage differences and forecasts into Markov switching models. International Journal of Forecasting 27 (2), 281–307.

Nordhaus, W. D., 1987. Forecasting efficiency: concepts and applications. Review of Economics and Statistics 69, 667–674.

Patterson, H. D., 1950. Sampling on successive occasions with partial replacements of units. Journal of the Royal Statistical Society Series B 49, 241–255.

Ramnath, S., Rock, S., Shane, P., 2008. The financial analyst forecasting literature: A taxonomy with suggestions for further research. International Journal of Forecasting 24 (1), 34–75.

Remus, W. E., O'Connor, M. J., Griggs, K., 1996. Does feedback improve the accuracy of recurrent judgmental forecasts? Organizational Behavior and Human Decision Processes 66, 22–30.

Ross, S. M., 2007. Markov Chains, 9th Edition. Elsevier, Ch. 4, pp. 185–280.

Sanders, N. R., 1992. Accuracy of judgmental forecasts: a comparison. Omega 20, 353–364.

Schipper, K., 1991. Analysts ' forecasts. Accounting Horizons 5 (4), 105–121.

Segers, R., Franses, P. H., 2007. Panel design effects on response rates and response quality. Erasmus University Rotterdam: Econometric Institute Report Series EI 2007-29.

Sheng, X., Thevenot, M., 2012. A new measure of earnings forecast uncertainty. Journal of Accounting and Economics 53 (1–2), 21–33.

Sheng, X. S., 2015. Evaluating the economic forecasts of {FOMC} members. International Journal of Forecasting 31 (1), 165 – 175.

Sichel, D., May 1991. Business cycle duration dependence: A parametric approach. The Review of Economics and Statistics 73 (2), 254–260.

Sigelman, L., 1991. Turning cross sections into a panel: A simple procedure of ecological inference. Social Science Research 20, 150–170.

Sijtsma, K., van der Ark, L. A., 2003. Investigation and treatment of missing item scores in test and questionnaire data. Multivariate Behavioral Research 38, 505–528.

Stickel, S. E., 1990. Predicting individual analyst earnings forecasts. Journal of Accounting Research 28 (2), 409–417.

Stone, E. R., Opel, R. B., 2000. Training to improve calibration and discrimination: The effects of performance and environmental feedback. Organizational Behavior and Human Decision Processes 83, 282–309.

Wedel, M., Kamakura, W. A., 2000. Market Segmentation: Conceptual and Methodological Foundations, 2nd Edition. Kluwer Academic Publishers.

Welch, E., Bretschneider, S., Rohrbaugh, J., 1998. Accuracy of judgmental extrapolation of time series data. characteristics, causes, and remediation strategies for forecasting. International Journal of Forecasting 14, 95–110.

Yang, Y., 2004. Combining forecasting procedures: Some theoretical results. Econometric Theory 20, 176–222.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

588. B.Z. YUESHEN *Frictions in Modern Financial Markets and the Implications for Market Quality*

589. D. VAN DOLDER *Game Shows, Gambles, and Economic Behavior*

590. S.P. CEYHAN *Essays on Bayesian Analysis of Time Varying Economic Patterns*

591. S. RENES *Never the Single Measure*

592. D.L. IN 'T VELD *Complex Systems in Financial Economics: Applications to Interbank and Stock Markets*

593. Y. YANG *Laboratory Tests of Theories of Strategic Interaction*

594. M.P. WOJTOWICZ *Pricing Credits Derivatives and Credit Securitization*

595. R.S. SAYAG *Communication and Learning in Decision Making*

596. S.L. BLAUW *Well-to-do or doing well? Empirical studies of wellbeing and development*

597. T.A. MAKAREWICZ *Learning to Forecast: Genetic Algorithms and Experiments*

598. P. ROBALO *Understanding Political Behavior: Essays in Experimental Political Economy*

599. R. ZOUTENBIER *Work Motivation and Incentives in the Public Sector*

600. M.B.W. KOBUS *Economic Studies on Public Facility use*

601. R.J.D. POTTER VAN LOON *Modeling non-standard financial decision making*

602. G. MESTERS *Essays on Nonlinear Panel Time Series Models*

603. S. GUBINS *Information Technologies and Travel*

604. D. KOPÁNYI *Bounded Rationality and Learning in Market Competition*

605. N. MARTYNOVA *Incentives and Regulation in Banking*

606. D. KARSTANJE *Unraveling Dimensions: Commodity Futures Curves and Equity Liquidity*

607. T.C.A.P. GOSENS *The Value of Recreational Areas in Urban Regions*

608. Ł.M. MARĆ *The Impact of Aid on Total Government Expenditures*

609. C. LI *Hitchhiking on the Road of Decision Making under Uncertainty*

610. L. ROSENDAHL HUBER *Entrepreneurship, Teams and Sustainability: a Series of Field Experiments*

611. X. YANG *Essays on High Frequency Financial Econometrics*

612. A.H. VAN DER WEIJDE *The Industrial Organization of Transport Markets: Modeling pricing, Investment and Regulation in Rail and Road Networks*

613. H.E. SILVA MONTALVA *Airport Pricing Policies: Airline Conduct, Price Discrimination, Dynamic Congestion and Network Effects.*

614. C. DIETZ *Hierarchies, Communication and Restricted Cooperation in Cooperative Games*

615. M.A. ZOICAN *Financial System Architecture and Intermediation Quality*

616. G. ZHU *Three Essays in Empirical Corporate Finance*

617. M. PLEUS *Implementations of Tests on the Exogeneity of Selected Variables and their Performance in Practice*

618. B. VAN LEEUWEN *Cooperation, Networks and Emotions: Three Essays in Behavioral Economics*

619. A.G. KOPÁNYI-PEUKER *Endogeneity Matters: Essays on Cooperation and Coordination*

620. X. WANG *Time Varying Risk Premium and Limited Participation in Financial Markets*

621. L.A. GORNICKA *Regulating Financial Markets: Costs and Trade-offs*

622. A. KAMM *Political Actors playing games: Theory and Experiments*

623. S. VAN DEN HAUWE *Topics in Applied Macroeconometrics*

624. F.U. BRÄUNING *Interbank Lending Relationships, Financial Crises and Monetary Policy*

625. J.J. DE VRIES *Estimation of Alonso's Theory of Movements for Commuting*

626. M. POPŁAWSKA *Essays on Insurance and Health Economics*

627. X. CAI *Essays in Labor and Product Market Search*

628. L. ZHAO *Making Real Options Credible: Incomplete Markets, Dynamics, and Model Ambiguity*

629. K. BEL *Multivariate Extensions to Discrete Choice Modeling*

630. Y. ZENG *Topics in Trans-boundary River sharing Problems and Economic Theory*

631. M.G. WEBER *Behavioral Economics and the Public Sector*

632. E. CZIBOR *Heterogeneity in Response to Incentives: Evidence from Field Data*

633. A. JUODIS *Essays in Panel Data Modelling*

634. F. ZHOU *Essays on Mismeasurement and Misallocation on Transition Economies*

635. P. MULLER *Labor Market Policies and Job Search*

636. N. KETEL *Empirical Studies in Labor and Education Economics*

637. T.E. YENILMEZ *Three Essays in International Trade and Development*