Advances in Monitoring the Economy

ISBN: 978 90 361 0104 2

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **446** 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.

Advances in Monitoring the Economy

Nieuwe methoden voor de monitoring van de economie

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 29 januari 2009 om 13:30 uur

door

RENGERT SEGERS
geboren te Haarlem.

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor:	Prof.dr. Ph.H.B.F. Franses
Overige leden:	Prof.dr. R. Paap
	Prof.dr. N.R. Swanson
	Prof.dr. D.J.C. van Dijk

To Anne van der Noord

Preface

After a four-year PhD project on monitoring, I guess I should now be able to provide a fully-fledged analysis of all developments that eventually led to this end result. The past few years have been so dynamic, however, that I forgot to collect data on the way. Despite this, I am convinced that I would have never been able to complete my thesis without the support of my close colleagues, friends and family. Let me take this opportunity to thank everyone who contributed to this thesis, either directly or indirectly.

First of all, I am grateful to my promotor Philip Hans Franses. Despite his many duties as the dean of the faculty, he was always more than happy to discuss my latest research ventures. Along the way, he taught me the art of efficient writing, and how to actually get projects done. His enthusiasm is contagious. The second part of this thesis is based on joint work with Richard Paap and Dick van Dijk. Working with Richard and Dick has been a great learning experience, and resulted in my first publication in record time. Besides doing research, I really enjoyed to attend conferences together, or to simply chat about anything we are passionate about. I would also like to thank Norman Swanson, who evaluated my thesis.

The Econometric Institute and the Tinbergen Institute offer a very pleasant working environment. I would like to thank three members personally. First of all, it was Herman van Dijk who aroused my interest in scientific research back when I was a Bachelor student. Since then we have been on a mission to demonstrate that visualization truly matters in Bayesian analysis. In that respect I am very happy with our publication together with Michiel de Pooter and Francesco Ravazzolo. Secondly, I would like to thank Hans Frenk. At some point during my PhD I preferred teaching Simulation above doing research. Doubtless this was partly due to our pleasant cooperation. I enjoyed our numerous discussions on the education program and on the destiny of university professors. Thirdly, at the very heart of the Econometric Institute, there is Elli Hoek van Dijke. Being involved in the organization of the “Rotterdam Econometrics Day” and various activities marking the 50th anniversary of the institute, I found out that Elli is the secret behind all such activities. She truly deserves a medal.

I have been lucky to have many friends around, both on- and off-campus. We had a great laugh, no matter whether we were crossing White Mountain National Forest in our rental "Hummer", making music, enjoying a whisky after a dark movie, fantasizing about the future, or just playing darts. Perhaps I owe most to my roommates Robin and Hendrik, and to Merel, Peter and Stephanie, who had to copy with my mood swings on a daily basis. Thank you all!

Next I would like to thank my family. When I told my parents five years ago that I considered doing a PhD, my father said: "I happen to know it is a terrible job, but *you* will be able to complete it." This trust kept me going, especially when the job indeed seemed terrible from time to time. In Rotterdam, Maaike was always there to help me out, no matter how stupid my concerns were.

I dedicate this thesis to my grandfather, who has always been so proud of me and exactly knew what I was doing. It is very sad that you had to conclude last month that you would not be able to be there at my defence, even though it was just a couple of months away.

Last but not least I would like to thank Qian, who came into my life about two years ago... and stayed! 倩倩心儿,我爱你!

Rene Segers
Rotterdam, november 2008

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction and Motivation

Monitoring involves the collection, analysis and evaluation of information over time. For many professionals, monitoring is a central aspect of their work. For example, policy-makers closely watch the effects of their current policies to set the right course for reform. Likewise, physicians monitor the well-being of their patients to adjust their treatments when necessary. In business, financial investors monitor stock prices and interest rates to optimally time their investments, while marketing managers watch their customers' needs and wants to frame their marketing efforts.

The above examples illustrate that monitoring is crucial in many disciplines to make the right decisions at the right moment. For this reason, there has always been a need for improved monitoring methods. With the advent of increasingly powerful computers and advanced analytical techniques, monitoring systems can nowadays process large amounts of information and have become fully automated where desired. A large body of monitoring methods originate from academics. Especially during the past four decades, many insights from various fields such as economics, statistics, psychometrics and econometrics found their way into everyday monitoring practice. With the overwhelming availability of information in some cases, but also the intrinsic lack of information in other cases, the area is continuously faced with new and highly relevant research challenges.

The aim of this thesis is to contribute to the development of new monitoring methods by offering potential solutions to some of these challenges. The challenges studied in this thesis arise from all three aspects of monitoring, that is from the collection, the analysis as well as from the evaluation of information. We study these challenges within the context of two main applications. The first application concerns monitoring consumer confidence. Here we focus particularly on the way the data is collected. The second application involves a monitor for the state of the economy. Within this context, we develop new methods to analyze data on leading indicator variables and to evaluate the performance of the monitoring system. In the remainder, we discuss some of the main challenges within the areas of data collection, analysis and evaluation in a more general context, providing a first flavor of the solutions to be developed in the thesis.

Data collection

Data collection concerns the continuous process of gathering and administering the most recent information about the subject to be monitored. Ideally, this information should be instantly available at any point in time. This allows the user to monitor continuously and to rapidly act upon the information obtained. In practice, however, it is often very costly or otherwise undesirable to collect data on such a frequent basis. For example, high costs are induced if it is necessary to temporarily close a bridge in order to measure its condition. For this reason, such measurements can only be performed occasionally. Frequent data collection is undesirable in the case of interviewing or surveying individuals. As individuals are surveyed more frequently on the same topic, their (reported) opinions and attitudes regarding the topic may likely be affected once they realize they are closely monitored. One can easily appreciate that this deteriorates the quality of the obtained information.

In sum, the number of data collection occasions is often limited. In these cases it is key to choose the occasions optimally. Generally, this is accomplished by distributing the occasions evenly over time. In this thesis we argue that an even distribution of data collection occasions is *not* desirable in the case of surveying individuals. Instead, we propose to survey individuals at random time intervals. By performing two field studies, we show that this new strategy leads to increased participation in the survey and that individuals' responses are less biased due to being a member of a panel. For the strategy to be successful, it is required that individuals are not informed about the dates they will be surveyed. As a result, individuals are less likely to develop expectations as to when they will be surveyed again. Our findings are useful in situations where it does not suffice to monitor individuals using revealed data alone. Examples of such situations include monitoring patients' mental states in psychology, or monitoring consumers' propensity to churn in marketing.

We show the merits of our new data collection approach by monitoring consumer confidence in The Netherlands at the weekly level. Consumer confidence indicates the extent to which consumers think that the economy is in good shape now, and will be in good shape in the near future. It is based on the sentiments of consumers about the economic climate in general and about their own financial situation. As such, it has proved to be a good predictor of the current and the future state of the economy.

Data analysis

Once the data have been collected, the next challenge is to analyze them. This amounts to translating the data into clear insights about the subject, which guide informed decision-making. Generally, the insights are supported by statistical estimates, such as predictions about the future course of the subject or estimates of the relationship between variables. These estimates are often prepared with the aid of econometric models.

We mention three challenges within this area. The first two challenges stem directly from data limitations. Firstly, it often happens that a substantial part of the data is missing due to the design of the data collection method. An obvious example is the surveying method discussed above, where missing values occur on the occasions when individuals are intentionally not surveyed. On the occasions when individuals *are* surveyed, data may be missing nonetheless due to nonresponse. Generally, it is inappropriate to exclude these cases from the analysis. A better solution is to impute the missing values by simulated model predictions. In our application to monitoring consumer confidence we take this road and we analyze the differences in the results obtained.

A second challenge is that often the collected information is not instantly available. For example, macroeconomic time series are typically released with a publication lag of one or more months. This implies that forecasts are needed to predict not only the future state of the system, but also its current state. Apart from being published with delay, the most recently published data are usually preliminary. As time passes and new information becomes available, the data are constantly revised. As a result, forecasts based on preliminary releases exhibit more uncertainty and may be biased. Monitors on the basis of this type of real-time data should take this aspect into account.

The third and final challenge that we discuss concerns the interrelationship between monitored variables. Changes in one variable are often found to be correlated with, or even to be caused by, changes in other variables. Typically these changes do not occur simultaneously but with a time lag. For example, research has shown that if consumer confidence decreases sharply today, then economic growth is likely to decrease six months later. In other words, consumer confidence tends to lead the business cycle by six months. By monitoring the changes in these so-called leading indicators, we can predict turns in the future course of the economy from expansion to recession or vice versa. It is, however, very unrealistic to assume that the lead times of leading indicators are constant across cycles. In fact, it appears to be the case that many of these indicators have a considerably longer lead time at business cycle peaks than at troughs. In the case of consumer confidence this is likely because of loss aversion, which is the tendency for people to prefer avoiding losses than acquiring gains. Once consumers feel that the economy might slip into a recession, they tend to immediately change their verdict from being optimistic to being pessimistic about the economic outlook, such that when the recession indeed starts this bad news does not come as a surprise. In times of recession, however, consumers tend to be more reserved to change their verdict from being pessimistic to being optimistic, such that if the recession lasts longer than expected, they at least do not suffer unexpected losses due to overoptimism. In this thesis, we develop a formal statistical approach to investigate the possibility of different lead times at peaks and at troughs. In an application to The Conference Board's Composite Coincident Index and its Composite Leading Index we shown that allowing for asymmetric lead times yields improved forecasts of business cycle turning points and economic growth in the US.

Evaluation

Evaluation in the context of monitoring comes down to monitoring the performance of the monitoring system. An examination of the system's historical performance often helps to make better choices for the data collection method and the model used to analyze the data. Generally, the criteria used for evaluation are very specific to the purpose of the monitor. For this reason we necessarily restrict our attention to the evaluation of the two monitoring applications considered in this thesis.

The main purpose of monitoring consumer confidence is to quickly identify changes in consumer confidence over time. To be more precise, given consumers' current verdicts about the economy, we are interested in knowing how many consumers changed their verdicts over a particular time period. In an evaluation of the current consumer confidence monitor for The Netherlands, as produced by Statistics Netherlands, we conclude that this is strictly impossible on the basis of their data. This originates from the way the data are collected. Each month, Statistics Netherlands surveys approximately one thousand consumers and compiles a consumer confidence indicator by subtracting the share of negative answers from the share of positive answers. However, each month this concerns one thousand different individuals. As a consequence, changes in the indicator can either be attributed to changes in consumer confidence, as desired, but also to changes in the composition of the group of respondents. Using our new data collection method, we seek to circumvent this problem.

In our evaluation of monitors for the state of the economy, we consider two performance criteria. First of all, a monitor should detect turning point dates accurately, without providing false signals. Secondly, it should detect turning points quickly. Generally, to detect turning points accurately, it is wise to wait long so that one can make sure that the observed changes in the data are not temporary but persistent. In other words, being very accurate generally implies being slow. The real challenge is thus to find a balance between accuracy and speed of detection. To facilitate the detection of turning points, the models considered in this thesis convert the input data into recession probabilities. We apply visualization techniques to examine at which date a particular model detected a recession and whether or not it generated false warning signals. Our results indicate that the model with asymmetric lead times as discussed earlier leads to more timely and precise identification of peaks and troughs for the two most recent recessions in the US economy, as compared to currently used models.

To summarize, monitoring is crucial to make well-informed decisions at the right moment in various disciplines such as politics, health care, logistics, finance and marketing. This thesis aims to contribute to the methodology and practice of monitoring, focusing on monitors for individuals' opinions and attitudes on the one hand and monitors for the state of the economy on the other hand. To improve monitors for individuals' opinions and attitudes, we propose a novel data collection method where individuals are surveyed at random time intervals. An application to monitoring consumer confidence in the Netherlands demonstrates that our method enables easy interpretation of changes in consumer confidence over time, as opposed to currently used methods. To improve monitors for the state of the economy, we argue that it is crucial to recognize that leading indicators often have different lead times at business cycle peaks and at troughs. Formally allowing for asymmetric lead times yields improved real-time dating of business cycle peaks and troughs and more accurate forecasts of turning points in the US economy.

1.2 Summary and Conclusions

This thesis is divided into two parts, which both consist of two chapters. The chapters are self-contained and can thus be read independently. Part A consists of Chapters 2 and 3. This part focusses on a new data collection method with the purpose to monitor individuals' attitudes and opinions. The method is applied to monitor weekly consumer confidence in the Netherlands. Part B consists of Chapters 4 and 5. The emphasis in this part is on the estimation of the lead times of leading indicator variables at peaks and at troughs with the purpose to monitor the state of the US economy.

Part I: Monitoring Individuals' Attitudes and Opinions

To monitor individuals' opinions and attitudes it would be best to survey the same individuals at multiple points in time. Such surveys are usually referred to as longitudinal or panel surveys. Practice has shown, however, that frequent surveying of the very same individuals likely deteriorates the quality of the information obtained. People get irritated and they disconnect from the panel, thereby making this data collection method less efficient. Or perhaps worse, respondents' (reported) opinions and attitudes may change due to being a member of a panel, which is called panel conditioning.

In Chapter 2 we address whether this problem can be alleviated by designing a panel survey in an alternative way. For this purpose, we perform two field studies where we measure the effects of several panel design characteristics on response rates and response quality. These characteristics include the number of waves and the time between subsequent waves, which may either be fixed or random. Our findings suggest that response rates and response quality can be improved significantly by surveying at random time intervals. It is then crucial that panel members are not informed about the dates they will be surveyed, because in this case respondents are less likely to develop expectations as to when they will be surveyed again. The methodology we put forward can be used to improve the efficiency of a panel study by carefully calibrating the studies' panel design parameters. This chapter is based on Segers and Franses (2007).

The central application in this part is the collection of consumer confidence data. To monitor a countries' consumer confidence level, typically each month statistical agencies survey one thousand consumers. However, each month this concerns one thousand different individuals. A major consequence of this way of collecting data is that changes over time are very difficult to interpret. These changes can either be attributed to developments in consumer confidence, as desired, but also to changes in the composition of the group of respondents. In Chapter 3 we seek to circumvent this problem by applying the alternative method of data collection as introduced in Chapter 2. The data obtained allow us to statistically analyze the dynamic correlation of consumer confidence and to draw inference on transition rates. We argue that this is not possible for currently available monthly data collected by statistical agencies on the basis of repeated cross-sections. An application of the method to various waves of data for the Netherlands shows its advances. Upon temporal aggregation we also show the resemblance of our data with those collected by Statistics Netherlands. This chapter is based on Segers and Franses (2008).

Part II: Monitoring the State of the Economy

Business cycle indicators have proved to be useful tools to monitor the current state of the economy and to predict its state in the near future. Depending on the timing of their movements relative to the business cycle, business cycle indicators are classified as leading, coincident or lagging indicators. Leading indicators are assumed to shift direction in advance of the business cycle. For this reason, they attract most attention when it comes to forecasting the future state of the economy. One can readily imagine that, for an indicator to be useful as a leading indicator, it should have a considerable lead time with respect to the business cycle. Most of the currently popular leading indicator variables are believed to have a lead time between six and eighteen months. At the same time, it appears to be the case that many of these variables have a considerably longer lead time at business cycle peaks than at troughs.

In Chapter 4 we develop a formal statistical approach to investigate the possibility of different lead times at peaks and troughs. For this purpose, we propose a novel Markov switching vector autoregressive model, where economic growth and leading indicators share a common Markov process determining the state, but such that their cycles are non-synchronous with the non-synchronicity varying across the different regimes. Our empirical application involves The Conference Board's monthly Composite Coincident Index (CCI) and its Composite Leading Index (CLI) for the period 1959 - 2007. The CCI is composed of four coincident indicators including employment and industrial production, whereas the CLI is composed of ten leading indicators, including money supply, stock prices and consumer confidence, as studied in Part I of this thesis. The results indicate that on average the CLI leads the CCI by nearly one year at peaks, but only by four months at troughs. In terms of timeliness, the CLI is therefore most useful for signalling oncoming recessions. Furthermore, we find that allowing for asymmetric lead times yields improved real-time dating of business cycle peaks and troughs and more accurate forecasts of turning points and CCI growth rates. This chapter is based on Paap *et al.* (2009).

Apart from having a considerable lead time, leading indicators should lead the business cycle consistently. Consistency refers to the property that a leading indicator should systematically give an accurate indication of the future course of the economy and should not produce false turning point signals too frequently. In practice it has proven to be challenging to find leading indicators that are both timely and consistent. Timely indicator variables are usually not very consistent, and vice versa. A formal assessment of both the timeliness and the consistency of individual leading indicators may shed some light on this. In Chapter 5 we undertake such a formal assessment. For this purpose, we extend the Markov switching model of Chapter 4 such that it allows to simultaneously estimate the individual lead times of a large panel of leading indicator variables. The model relates the turning points of the indicators to the turning points of a reference series, where it is assumed that the cycle of the reference series coincides with the business cycle. Again, the lead times of the indicators are allowed to be different at business cycle peaks and at troughs. Our empirical analysis in this chapter involves the ten separate components included in The Conference Board's CLI along with its CCI, measured over the period 1959 - 2008. Our results suggest that the indicators BUILDING PERMITS, STOCK PRICES, MONEY SUPPLY and CONSUMER EXPECTATIONS are the most timely and consistent indicators among the ten, having average lead times of nine to ten months at peaks and four to five months at troughs with standard deviations of about one month. We apply the model to construct a new, synchronized composite leading index by shifting the ten indicators according to their lead times before aggregation. We show that the synchronized index is more consistent than the CLI and that it yields better in-sample predictions of the business cycle chronology as determined by the NBER. This chapter is based on Segers *et al.* (2008).

Part I

Monitoring Individuals' Attitudes and Opinions

Chapter 2

Panel Design Effects on Response Rates and Response Quality

2.1 Introduction

Understanding changes in individuals' opinions and attitudes is of fundamental interest in the social sciences. To measure such changes, it is desirable to conduct a longitudinal, or panel study, where the same individuals are surveyed at multiple points in time. This allows the researcher to study changes in opinions and attitudes over time at the individual level and to capture the dynamic relationships between events and behavior. Secondly, it is desirable to conduct the survey on a frequent basis. This allows the researcher to establish whether certain changes are permanent or transitory.

One can easily appreciate however that frequent surveying of the very same individuals 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) opinions and attitudes may change due to being a member of a panel, which is called panel conditioning.

In sum, monitoring individuals at a high frequency is useful, but proper data collection is not trivial. It is precisely this topic that we address in this chapter, that is, can we design better ways of data collection? We do believe we can, as we will argue in theory and as we will demonstrate empirically.

The chapter provides two main contributions. Firstly, we propose a general research design to study the effects of panel design characteristics on response rates and response quality. We demonstrate that response rates and response quality can be significantly improved if the design of the panel is carefully calibrated. This makes the survey more efficient. As such, our approach helps to design efficient panels that are cost effective. Secondly, we propose a new panel design where individuals are surveyed at random time intervals. We show that this new design leads to higher response and that the design is less susceptible to panel conditioning bias.

Although nonresponse bias can be handled quite effectively a posteriori using model-based procedures, obviously, preventing nonresponse at the data collection stage is to be preferred as nonresponse due to irritation may harm the relationship with the respondents. This approach has been successful in the field of *questionnaire* design, where the effects of design characteristics such as incentives, the length, and the presentation of the questionnaire have been examined, see for example the work on mail surveys by Adams and Darwin (1982) and Yu and Cooper (1983), and the more recent books by Dillman (2000) and Presser *et al.* (2004), among others. It also relates to the literature on consumer psychology, where the effect has been studied extensively that respondents are often induced by the measurement process to form judgements, see Simmons *et al.* (1993), Dholakia and Morwitz (2002) and Morwitz and Fitzsimons (2004), among others. Other interesting contributions have been made in optimal experimental design theory, where the aim is to design experiments such that statistical inference is most efficient, see for example Atkinson and Donev (1992) for a general exposition, the work in psychological research by Allison *et al.* (1997) and McClelland (1997), and the work in econometrics by De Stavola (1986), Nijman *et al.* (1991) and Ouwens *et al.* (2002). In contrast, to the best of our knowledge, the effects of *panel* design characteristics on response and response quality have not been studied extensively yet.

The chapter is organized as follows. In the next section we discuss various types of response bias relevant to panel surveys, and review several panel designs that have been proposed in the literature to reduce these biases. We then formulate our hypotheses regarding the effects of panel design on response rates and response quality in Section 2.3. Section 2.4 introduces a flexible panel design framework, which embodies the designs discussed in Section 2.2. Within this framework, a panel design is summarized by three design characteristics, namely the time frame of the study, the number of survey requests, and the strategy to determine the timing of these requests. We propose a dynamic panel Tobit-II model to calibrate the design characteristics on experimental data. In Section 2.5 we report the results of two field studies, which are conducted to test our hypotheses as well as to demonstrate the usefulness of the framework. Finally, Section 2.6 concludes.

2.2 Background

In order to provide context for our discussion, this section provides the relevant background on survey nonresponse, response quality, panel survey design and analysis. We are necessarily brief in our explanation. For a more elaborate review, including numerous practical examples, we refer to the excellent paper by Duncan and Kalton (1987).

2.2.1 Nonresponse and Response Quality

Survey nonresponse

Once an individual has agreed to become a member of a panel, three different types of nonresponse may occur, see for example Verbeek (1991) and Schafer and Graham (2002). The weakest form is *item nonresponse*, where panel members do not answer one or more particular questions of the survey. More serious is *wave nonresponse*. In this case, panel members do not participate in the survey during one or more particular waves. Ultimately, *attrition* or *dropout* occurs when panel members disconnect from the survey prematurely.

Response quality

Besides the response rate, also the response quality may be at stake if a panel design demands (too) much from its participants. Response quality is used as an all-embracing term that covers the desire to obtain survey data that is not biased by the survey environment. An often demonstrated threat to response quality is the mere-measurement effect. This is the effect that measuring an individual's preferences may change his or her subsequent behavior, see Dholakia and Morwitz (2002) and Morwitz and Fitzsimons (2004). In this chapter we are particularly concerned with potential biases in response due to panel participation. This form of response bias is usually referred to as panel conditioning or time-in-sample bias. Following Trivellato (1999), we distinguish between two different types of panel conditioning bias. Firstly, as a consequence of being a panel member, respondents may change their *reporting* behavior. For example, because panel members are typically asked to complete similar surveys repeatedly, they tend to get less involved in completing the surveys. For example they may incorrectly report exactly the same now as they did during the previous wave. Alternatively, respondents may give socially desirable responses as they start to be aware of being monitored, an effect usually referred to as the Big Brother effect. Secondly, a respondent may change his or her *actual* behavior due to panel participation. For example, having to report an opinion repeatedly may cause a respondent to think more about his or her opinion, reconsider it, and change it.

As opposed to nonresponse bias, measuring panel conditioning bias through revealed preference data is not trivial. It requires a careful comparison of the responses given in the first wave of data collection, which is free of panel conditioning bias by definition, and in the next waves. We extend the notion of Hansen (1980) who argues that there should not be a difference in the response distribution of different subgroups of panel members who have been exposed to different methods of data collection.[1] In our case, this implies that responses should not depend on the particular panel design chosen, nor on the wave of data collection.

2.2.2 Panel Design

The extremes: complete panels and repeated cross-sections

Longitudinal or panel data sets allow the researcher to capture the dynamic relationships between events and behavior, and to control for time-varying and individual-specific characteristics. Furthermore, by exploiting the time invariance of the unobserved individual characteristics, panels are well-suited to account for unobserved heterogeneity, see for example the excellent books by Arellano (2003), Hsiao (2003) and Baltagi (2008). In terms of data richness, ideally one should be to be able to observe, or in our case to survey, each individual in the panel at each point in time. This survey design is illustrated in Figure 2.1, Panel (a). In this figure individuals are indexed by i, where $i = 1, ..., N$, and time is indexed by t, where $t = 1, ..., T$. A grey square in row i and column t indicates that respondent i is requested to be surveyed at time t. Since complete panels are most vulnerable to nonresponse and to a low quality of response, one can easily appreciate that this design is often not feasible in practice. In this subsection we therefore explore several so-called pseudo-panel designs, mostly designed with the aim to reduce costs and respondent burden. We discuss whether these designs are still attractive from a statistical point of view.

Perhaps the most rigorous way to reduce respondent burden is to collect repeated cross-sectional data instead of complete panel data, as illustrated in Figure 2.1, Panel (b). In this design, at each point in time, a unique group of individuals, to be denoted by g, where $g = 1, ..., G$, is requested to complete a survey. Since in this case individuals are surveyed only once, this ensures that individuals' current opinions are not biased by previous panel participation. As a consequence, there is no panel conditioning bias.

[1] See Deutskens *et al.* (2004) for a recent application of this notion.

Several scholars have argued that the estimation of dynamic models at the individual level is possible on the basis of repeated cross-sections, see Deaton (1985), Verbeek and Nijman (1993), Moffitt (1993) and Collado (1997), among others. However, the identification conditions for these estimators are very strong and potentially unrealistic in many empirical applications (Verbeek and Vella, 2005). Alternatively, one may use matching techniques to match similar individuals and form a pseudo panel, but their assumptions are no less strong.

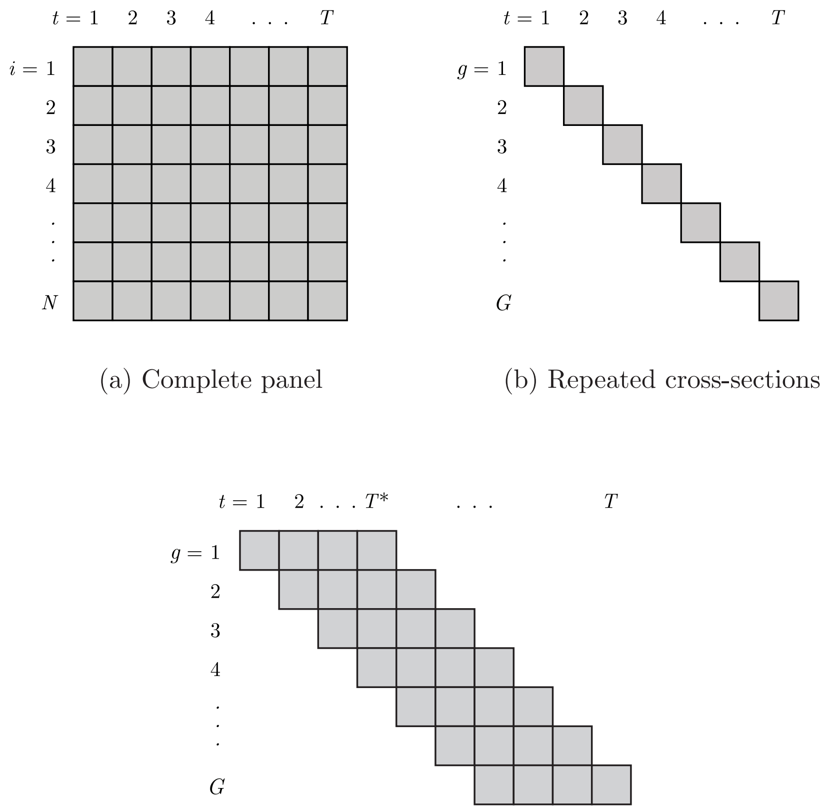
Rotation

In general, one wants to preserve the possibility to study individual dynamics. Still, individuals can only be monitored for a short period of time. A natural compromise is therefore to apply a panel refreshment strategy. This pertains to requesting each panel member to only join the panel for a fixed period of, say T^* time periods. Each wave or block of waves, a new group of individuals is invited to join the panel, with the aim to keep the total number of panel members constant. The latter strategy is also referred to as rotation, see Patterson (1950) and Kish and Hess (1959). An example of a rotating panel is shown in Figure 2.1, Panel (c). For analysis-of-variance models, Nijman *et al.* (1991) demonstrated how to set up a rotating panel to maximize estimation efficiency. They showed that the efficiency gains from using an optimal rotation strategy can be quite substantial, even if the costs of a reinterview equal the costs of acquiring a new panel member.

Continuous sampling and time sampling

Once one has decided upon the number of time periods T^* that an individual will be requested to join the panel and possibly upon a particular rotation strategy, typically the next step is to decide upon the number of survey requests within this period, to be labeled 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 waves. Often the sampling frequency is set equal to the desired data frequency, to be labeled f_{D}. We will refer to this situation as continuous sampling. For example, suppose that the aim is to ask individuals to join a panel for at most two months and the desired sampling frequency f_{D} is weekly, but the maximum accepted number of survey requests n is four. A continuous sampling strategy would then imply that each panel member is interviewed weekly, as desired, but only during the first four weeks. See Table 2.1, Panel (a), for an illustration.

Figure 2.1: Alternative sampling strategies across individuals



(a) Complete panel

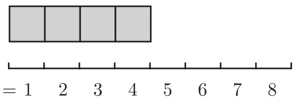
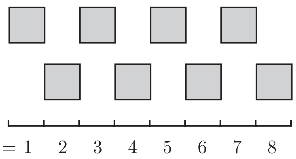
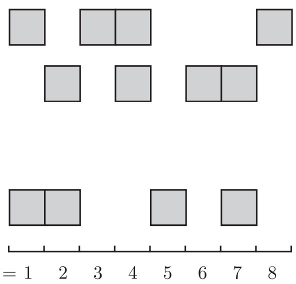
(b) Repeated cross-sections

(c) Rotating panel

Note: In this figure, three sampling strategies across individuals are visualized. In Panel (a), a grey square in row i and column t indicates that respondent i is requested to be surveyed at time t. Likewise, in Panels (b) and (c), a grey square in row g and column t indicates that respondent group g is requested to be surveyed at time t.

In many cases, however, panel members are better observed over the longest horizon possible. This allows the researcher to detect possible changes in opinions in the longer run. Naturally, to extend the panel's horizon, the survey requests should be spaced as much as possible. In our example this would imply to interview biweekly instead of weekly. In order to still be able to monitor on the desired weekly basis, one may now opt to ask one-half of the panel members to complete a survey in the even weeks and the other half to complete a survey in the odd weeks, as illustrated in Table 2.1, Panel (b). This approach, known as time sampling, has gained momentum in recent years. It has been accepted as the natural way to lower the sampling frequency while keeping the data frequency unchanged.

Table 2.1: Alternative sampling strategies over time

Strategy		Group	Sampling scheme
(a)	Continuous sampling	All respondents	
(b)	Time sampling	Respondent group 1	
		Respondent group 2	
(c)	Randomized sampling	Respondent 1	
		Respondent 2	
		⋮	
		Respondent N	

Note: The table presents three alternative strategies to choose the occasions at which respondents are surveyed over time. In the third column these strategies are illustrated for the situation where $T^* = 8$ time periods and $n = 4$ survey requests. The grey squares represent the survey occasions chosen.

Randomized sampling

Recall that we are interested in measuring individuals' attitudes and opinions and their changes over time. In order to be able to identify whether the latter changes are permanent or transitory, we are interested in the autocorrelations of these attitudes and opinions. From this perspective it seems appealing to consider choosing the n survey occasions at random, independently for each panel member, as illustrated in Table 2.1, Panel (c). We argue as follows. To reveal the underlying correlation structure in the data it is important to measure autocorrelations at many different lags. This facilitates the identification of any type of individual dynamics in the data and efficient estimation.

For the three different sampling strategies shown in Figure 2.1 we obtain the following (expected) numbers of observations of the $k-$th autocorrelation that are collected, to be denoted by $A(k)$

- For continuous sampling (CS)

$$A_{\mathrm{CS}}(k) = \begin{cases} n-k & \text{for } k = 1, 2, ..., n-1 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.1}$$

- For time sampling (TS) with frequency $f = n/T^*$

$$A_{\mathrm{TS}}(k) = \begin{cases} n-kf & \text{for } k = 1/f, 2/f, ..., (n-1)/f \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2}$$

- For randomized sampling (RS)[2]

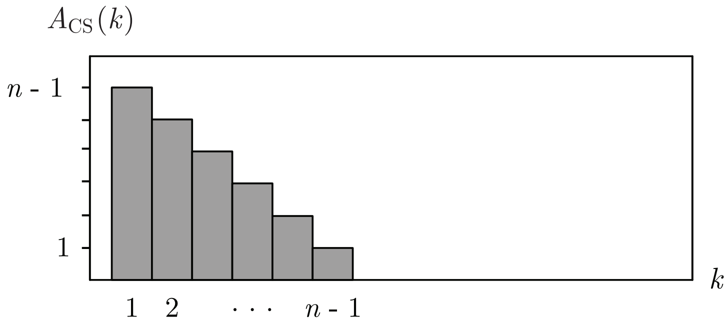
$$E[A_{\mathrm{RS}}(k)] = \begin{cases} \frac{\binom{T^*-2}{n-2}}{\binom{T^*}{n}}(T^*-k) = \frac{\binom{n}{2}}{\binom{T^*}{2}}(T^*-k) & \text{for } k = 1, 2, ..., T^*-1 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.3}$$

The three functions are depicted in Figure 2.2. In the 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. Conceptually, it is convenient to interpret the sampling frequency f in this case as the participation request probability. In the next section we will argue that the randomization approach may not just seem attractive from an estimation point of view but also from a behavioral point of view.

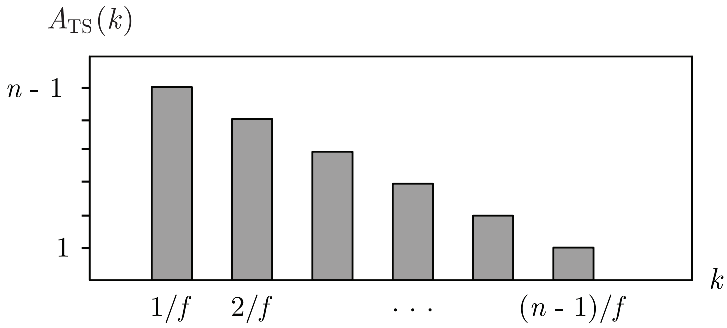
Obviously, different survey designs may be combined or applied to different subsamples of individuals. One may combine repeated cross sections with complete panel data, see for example Nijman and Verbeek (1990) and Hirano *et al.* (2001). This approach facilitates testing for possible sampling biases, as will be discussed later. One may also decide to split the questionnaire and expose different groups of respondents to different subsets of questions in an efficient way. This approach is commonly known as matrix sampling, see for example Shoemaker (1973), Johnson (1992), Raghunathan and Grizzle (1995) and the recent paper by Graham *et al.* (2006). In this chapter, however, we restrict our attention to pure panel design aspects, exposing every panel member to the same questionnaire.

[2]To see that this result is correct, note that we can position a pair of surveys, being k time periods from one another to obtain one autocorrelation of the k-th order, at $(T^* - k)$ different points in time. The remaining $(n-2)$ surveys can be positioned in any of the $(T^* - 2)$ remaining time periods. If we now divide this by the total number of different time-series possible, which is $\binom{T^*}{n}$, we obtain the result as stated.

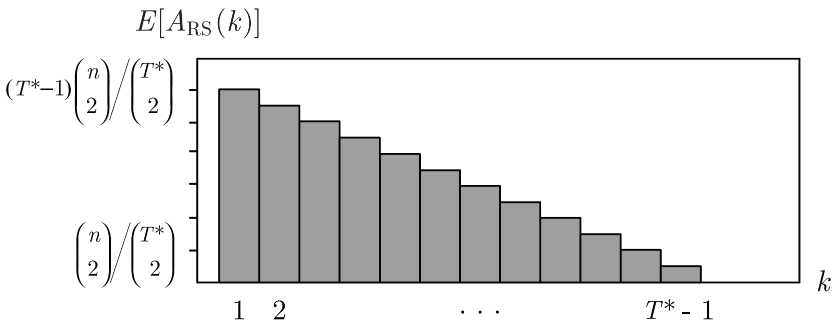
Figure 2.2: Observed autocorrelations using different sampling strategies



(a) Continuous sampling



(b) Time sampling



(c) Randomized sampling

Note: The three histograms depicted above display the number of autocorrelations of order k that are collected when (a) Continuous sampling, (b) Time sampling or (c) Randomized sampling is applied to select n survey occasions among T^* possible occasions in total.

2.2.3 Modelling incomplete panels

As a result of using any of the above advanced panel designs, typically one is confronted with two types of missing data. Firstly, parts of the panel are missing intentionally due to the design of the panel. Secondly, parts of the panel may be missing due to nonresponse.

Missingness by design

Observations that are missing intentionally due to the design of a panel are not related to the variables to be collected in any of the cases mentioned in the previous section. Therefore, following the typology of Rubin (1976), they are *missing completely at random* (MCAR). As a result, missingness by design can be ignored, in the sense that analyzing the incomplete panel will not bias our results. Still, especially in the case of randomization, the panel that is collected may seem rather intractable, as the individual time series are unequally spaced. Modern techniques are readily available however to deal with such data sets.

First of all, if the objective is to estimate first order dynamics, one may consider using a Koyck lag structure, see Ansari *et al.* (2008) and Van Diepen *et al.* (2009) for recent applications. For a panel with elements $y_{i,t}$ this amounts to relating $y_{i,t}$ directly to the observation measured on individual i's previous survey occasion, to be denoted by $y_{i,t-d_{i,t}}$, using

$$y_{i,t} = \rho^{d_{i,t}} y_{i,t-d_{i,t}} + e_{i,t}, \tag{2.4}$$

where $e_{i,t}$ is a white noise process and $d_{i,t}$ denotes the time between $y_{i,t}$ and $y_{i,t-d_{i,t}}$. The factor $\rho^{d_{i,t}}$ is a finite duration adjustment of the geometric lag or Koyck model, which depends on one parameter ρ that measures the effect of the (unobserved) observation $y_{i,t-1}$ on $y_{i,t}$. The advantage of this representation is that unequally spaced time series can be analyzed directly.

To allow for more complex dynamics, one may use multiple imputation or likelihood-based techniques, see the excellent survey of Schafer and Graham (2002), and the books by Schafer (1997) and Little and Rubin (2002). In the first case, each missing value in the panel is substituted by a set of estimated or predicted values based on the available data. The resulting complete panels are then analyzed using conventional complete-data techniques and the results are combined. A widely used likelihood-based approach is maximum likelihood coupled with the EM algorithm of Dempster *et al.* (1977). In this case, the E-Step imputes the best predictors of the missing values, using current estimates of the parameters. It also calculates the adjustments to the estimated covariance matrix needed to allow for imputation of the missing values (Little and Rubin, 2002). A second likelihood-based approach is to reformulate the model as a state space model, where in the observation equation the incomplete panel is transformed into a latent complete panel. In the state equation the complete panel is then regressed on its past values and additional regressors. By alternating between Kalman filtering, smoothing recursions and maximum likelihood estimators, estimation is relatively straightforward, see Palma and Chan (1997), and Shumway and Stoffer (1982, 2006).

Missingness due to nonresponse

In contrast to missingness due to the panel design, missingness as a result of nonresponse can generally not be ignored, as the tendency to drop out is often systematically related to the variables of interest. In some cases, however, the distribution of missingness only depends on observed data and not on missing data. In these cases, the missing data are said to be *missing at random* (MAR), and multiple imputation or likelihood-based techniques can still be applied. If the variables of interest also depend on the unseen responses then the missing data are *missing not at random* (MNAR), see Schafer and Graham (2002) and the references cited there for a more thorough discussion.

In case one does not want to make the MAR assumption, one usually employs either selection models or pattern mixture models. The first amounts to incorporating the response decision explicitly in an econometric model, as a form of non-random selection (Hausman and Wise, 1979). The usual approach is to formulate a two-step model. In the first step, the response decision is modelled. Conditional on response, in the second step, the dependent variables of interest are modelled, see Amemiya (1984) and Heckman (1976, 1979) for details. Pattern mixture models stratify the sample by the pattern of missing data and then model differences in the distribution of the variables of interest over these patterns, see Little (1995), among others.

2.3 The Effects of Panel Design Characteristics

In this section we hypothesize the effects of panel design characteristics on response rates and response quality. In our empirical section below we will collect data to examine the empirical plausibility of these hypotheses.

The sampling frequency and the number of waves

We assume that participating in a survey is a social exchange, similar to responding to a mailing, for example, of a charity organization. Therefore, to hypothesize the effects of the sampling frequency and the number of waves of a panel survey we adopt the Recency, Frequency, Monetary value (RFM) framework, commonly used in direct marketing. The sampling frequency, and thus the (average) time between subsequent waves, is a recency variable. The higher the sampling frequency, the higher the (expected) recency of the last participation request. As in general recency has a negative effect on the probability of (high quality) response, we hypothesize

H_1: The higher the sampling frequency, the lower (a) the response rate and (b) the response quality.

In the same vein, the number of waves can be seen as a frequency effect. The higher the number of waves, the higher the frequency in an RFM context, and hence the lower the probability of (high quality) response. We hypothesize

H_2: As the number of participation requests increases, (a) the response rate and (b) the response quality decrease.

Randomized sampling

To increase response rates it has been argued that potential respondents should not be requested directly to join a panel. Instead, they should only be requested to complete a first survey and to grant permission to be contacted again for follow-up surveys. If respondents are persuaded initially to comply with this smaller request, subsequent requests may less likely be declined (Reingen and Kernan, 1977). This effect is referred to as the foot-in-the-door effect. As a result, overall response may be higher in the first waves. Nevertheless, as the survey evolves, respondents may quickly learn they are members of a panel with a particular sampling frequency and disconnect from the survey after all. In the case of randomized sampling, learning the (average) sampling frequency of the panel may take long, however. As a result, in this case, we expect respondents to participate longer.

An often reported source of panel conditioning bias is negativity bias. This is the tendency of individuals to be more negative in their judgements if they expect to be evaluated, see for example Ofir and Simonson (2001). If we survey at random time intervals, panel members are less likely to develop expectations as to when they will be surveyed again. As a result, we expect less panel conditioning bias in this case, which may contribute to a higher quality of response.

Thus we hypothesize the following

H_3: Randomized sampling increases (a) the response rate and (b) the response quality.

2.4 Research Design

Setup

In order to be able to study the effects of panel design characteristics on response rates and response quality, we choose a flexible panel design. Our design is characterized by the collective answers to the following questions

(i) How many times n should we request a panel member to be surveyed at the maximum?

(ii) Within which time span T^* should this occur?

(iii) Given the number of requests and the time span chosen, at which dates should we survey each panel member? In particular, should we apply a time sampling or a randomized sampling approach?

Note that when the sampling frequency $f = n/T^*$ is equal to one, both time sampling and randomized sampling reduce to continuous sampling. We therefore focus on time sampling and randomized sampling only and consider values of f up to one to embed continuous sampling as a limiting case.

We perform two field studies. In the first we follow a foot-in-the-door approach by not informing our subjects about the panel design to be employed. That is, we first request each potential panel member to fill in one questionnaire. We then ask whether they agree to be contacted in the future for further research. To each respondent who reacts positively, we randomly assign a sampling strategy (time sampling or randomized sampling) and a sampling frequency f. We do not fix the total number of survey requests n nor the number of time periods T^*. Obviously these increase as the survey evolves. We expect our panel members to learn about the design through experience, and we assume that they base their further participation decision on this experience.

In the second study, we ask our potential panel members to first complete a typical questionnaire as used in the first study. We then explain again that we would like to contact them repeatedly in the future with the request to complete similar questionnaires. This time, however, we first determine what the respondents find an acceptable panel design for these follow-ups. For this purpose, we present ten randomly generated panel designs to them and request the respondents to indicate which of these designs are acceptable to them, if any. To generate the designs we do not only randomize over the sampling strategy and the sampling frequency f, but also over the number of requests n. Once these data are collected, we randomly select and use one of the accepted designs, separately for each respondent. In personal follow-up mailings the respondents are informed about the design chosen. In this second study, the respondents thus have full information about the panel design used and we assume that possible effects through learning are eliminated.

Model and estimation

To estimate the effects of panel design characteristics on the response decision and subsequent responses separately, it is convenient to use a selection model. We summarize the responses as a result of our participation requests in the response indicator matrix $\boldsymbol{R}$, where its elements $r_{i,t}$, for $i = 1, ..., N$, and $t = 1, ..., T$, register

$$r_{i,t} = \begin{cases} 1 & \text{Member } i \text{ is requested to participate at time } t \text{ and did so} \\ 0 & \text{Member } i \text{ is requested to participate at time } t \text{ but did not so} \\ \cdot & \text{Member } i \text{ is not requested to participate at time } t. \end{cases} \tag{2.5}$$

For positive responses, we additionally observe our questionnaire variables. Typically some of these variables are to be explained, whereas others may serve as explanatory variables. For ease of exposition, we focus in this section on one dependent variable $\boldsymbol{Y}$ with elements $y_{i,t}$ and a m-vector of covariates $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)$ with elements $\boldsymbol{x}_{i,t} = (x_{1,i,t}, \ldots, x_{m,i,t})$. The set of covariates $\mathbf{X}$ may be enriched with additional variables that are available for both respondents and nonrespondents, such as time-invariant demographics collected at earlier occasions. To explain the measurement process $\boldsymbol{R}$ as well as the variable $\boldsymbol{Y}$, our workhorse model will be a panel Tobit-II model, which consists of a Probit model for the elements of $\boldsymbol{R}$ being 0 or 1 and a standard regression model for $\boldsymbol{Y}$. The model can be written as

$$r_{i,t} = \begin{cases} 0 & \text{if } y^*_{i,t} = g(\mathbf{X}, \boldsymbol{\theta}_1, u_{1,i,t}) \leq 0 \\ 1 \quad \text{and} \quad y_{i,t} = g(\mathbf{X}, \boldsymbol{\theta}_2, u_{2,i,t}) & \text{if } y^*_{i,t} = g(\mathbf{X}, \boldsymbol{\theta}_1, u_{1,i,t}) > 0 \end{cases} \tag{2.6}$$

where $u_{1,i,t}$ and $u_{2,i,t}$ are error terms and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are parameter vectors in the Probit part and regression part, respectively. Note that the Tobit-II model allows $\mathbf{X}$ to have a different effect on $\boldsymbol{R}$ and on $\boldsymbol{Y}$ since $\boldsymbol{\theta}_1$ need not be equal to $\boldsymbol{\theta}_2$. To allow for the possibility that respondents resume their participation after one or more waves of nonresponse, the model seeks to explain wave response rather than attrition. The model parameters can be estimated by maximum likelihood.

2.5 Empirical Results

Our two field studies have been conducted among students at Erasmus University Rotterdam over the period September 2004 to March 2007. Our desired frequency of data collection, f_{D}, was weekly, and the sampling frequencies ranged from bimonthly to weekly ($f = 0.125$ to 1). The experiments were conducted online through an interactive website. All correspondence, including personalized participation requests, was generated automatically and sent by e-mail. The general task that we asked the respondents to complete was a test for their knowledge of recent news events. In the test we presented 20 news headlines, among which 10 were literally quoted from newspapers published in the previous week and 10 were slightly altered. The respondents had to indicate which headlines were indeed literally quoted and which were not. The headlines were selected from five different news categories, being DOMESTIC NEWS, FOREIGN NEWS, POLITICS, ECONOMICS, and SPORTS & CULTURE, and were carefully pretested to ascertain that they were equally familiar in each week. On average the respondents completed the test in three minutes.

Additional to the news test, we surveyed the respondents about their attitude towards the experiment, with the aim to measure possible signs of irritation and panel conditioning bias. For this purpose, at each wave we posed three statements per construct, randomly selected out of ten, which had to be rated on a seven-point Likert scale. As a consequence, the overall size of the questionnaire remained constant, but the statements were different. We summarized the scores in two variables, which measure self-reported irritation and self-reported panel conditioning bias.[3] As an incentive to continue participation, we raffled out a \$25 gift voucher at each wave of data collection among the respondents of that particular wave.

To complete the model in (2.6) for our field studies, we have to specify the regressors in $\mathbf{X}$. Presumably the most important regressors in our model are the panel design parameters. Firstly, we assume that respondents learn about the design parameter n through the number of times they have been requested to be surveyed previously, which is $n - 1$. As the survey evolves, n increases, but at a different rate for each individual. We therefore define the variable $n_{i,t}$ as the number of times that individual i has been requested to be surveyed before time t, and include this variable in the model. Secondly, the sampling frequency, to be labeled f_i for individual i, may be learned through the number of weeks

[3]For this purpose, we performed a principal components analysis on the scores obtained on the ten different statement which supposedly measure the same construct, and selected the first component. This component explains over 75% of the variation in the data in both cases.

since the previous request, which is $1/f_i$ in expectation. We denote the number of weeks since the previous request by $w_{i,t}$. Thirdly and finally, we include a time-invariant dummy variable indicating whether time sampling or randomized sampling has been applied, to be labeled $\boldsymbol{S}$ with elements $s_{i,t}$. The three design variables of individual i at time t are summarized in the vector $\boldsymbol{p}_{i,t} = (n_{i,t}, w_{i,t}, s_{i,t})$. Additional to the design variables, we include lagged response using a Koyck lag structure, as discussed in Section 2.2.3, to allow for first order dynamics. Finally, we include k individual-specific regressors such as demographics, self-reported irritation and panel conditioning, to be denoted by the vector $\mathbf{Q} = (Q_1, \ldots, Q_k)$ with elements $\boldsymbol{q}_{i,t}$. The regression equations can now be written as

$$g(\boldsymbol{x}_{i,t}, \boldsymbol{\theta}_s, u_{s,i,t}) = \rho^{d_{i,t}} y_{i,t-d_{i,t}} + \boldsymbol{p}_{i,t}\boldsymbol{\beta} + \boldsymbol{q}_{i,t}\boldsymbol{\gamma} + u_{s,i,t} \quad \text{for } s = 1, 2, \tag{2.7}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are parameter vectors. Finally, we specify the error terms $u_{s,i,t}$ in (2.6) as

$$u_{s,i,t} = (\tau_{s,t} + a_{s,i} + e_{s,i,t}), \quad \text{for } s = 1, 2. \tag{2.8}$$

The parameters $\tau_{s,t}$ are time-specific fixed effects to account for possible variation in response due to the particular week of data collection and the particular news test. For example, these parameters may capture lower response rates during examination periods or higher news scores due to a relatively easy test. The variables $\alpha_{s,i}$ denote individual-specific random effects to account for unobserved heterogeneity among respondents. We approximate the distribution of the random effects conditional on the initial conditions $y_{i,j,1}$, as suggested by Wooldridge (2005)

$$\alpha_{s,i} | y_{i,1}, \boldsymbol{x}_i \sim N(\alpha_{s,0} + \alpha_{s,1} y_{i,1} + \boldsymbol{x}_i \boldsymbol{\alpha}_2, \sigma^2_{\alpha_s}), \quad \text{for } s = 1, 2, \tag{2.9}$$

where $\boldsymbol{x}_i$ is the subset of all nonredundant explanatory variables in $\boldsymbol{x}_{i,t}$ in all time periods. The variables $e_{1,i,t}$ and $e_{2,i,t}$ are uncorrelated idiosyncratic shocks, which are assumed to satisfy

$$e_{s,i,t} | y_{i,t-d_{i,t}}, \ldots, y_{i,1}, \boldsymbol{x}_i \sim N(0, \sigma^2_{e_s}), \quad \text{for } s = 1, 2. \tag{2.10}$$

The variance $\sigma^2_{e_1}$ is set equal to 1 for identification purposes. We assume that the dynamics are correctly specified, which means that only $y_{i,t-d_{i,t}}$ appears in the distribution given outcomes back to the initial time period. The variables $\boldsymbol{x}_i$ are assumed to be strictly exogenous conditional on a_i, see Wooldridge (2005) for details.

Study 1: No information provided about the design

Among the 623 students we contacted, 290 agreed to participate in the first study. The total sample period of the study was 26 weeks. In Table 2.2 we summarize the response rates, where we classify the respondents to four groups according to the sampling

Table 2.2: Response rates

Frequency	Wave				
	2	3	4	6	12
Three times per month - weekly ($0.75 < f \leq 1.00$)	0.67	0.69	0.59	0.45	0.06
Biweekly - three times per month ($0.50 < f \leq 0.75$)	0.80	0.75	0.65	0.59	0.22
Monthly - biweekly ($0.25 < f \leq 0.50$)	0.83	0.82	0.76	0.68	–
Once every eight weeks - monthly ($0.125 \leq f \leq 0.25$)	0.88	0.84	0.77	–	–

Note: This table presents the response rates of Study 1 after $2, 3, 4, 6$ and 12 waves of data collection, where the respondents are classified to four groups according to the sampling frequency f assigned to them. The response rates for the bottom right cases are missing, due to the studies' time window of 26 weeks.

frequency assigned to them. It is already apparent from this table that the sampling frequency negatively influences response. For example, after six waves of data collection, the response rate among those respondents who have been surveyed monthly to biweekly is 0.68, which is over 50% higher than the response rate of 0.45 measured for the group of respondents who have been surveyed three times per month to weekly.

Table 2.3 shows the estimation results of the dynamic panel Tobit-II model. In the first panel the results of the Probit part of the model are presented. The second panel presents the results of the regression part. In this part we seek to explain four different variables $\boldsymbol{Y}$, and as a consequence we estimate our model four times. The estimates within the Probit part are the same across all variables $\boldsymbol{Y}$.

Inspecting the results for the Probit part, we find that the number of weeks between subsequent participation requests has a positive effect on the probability of response. Hence the higher the sampling frequency f, the lower the response rate, as hypothesized in H_{1a}. The number of waves n has the expected significant negative effect. Perhaps more surprising, the dummy variable that distinguishes between randomized sampling (1) and time sampling (0) is significantly positive. This indicates that it is beneficial to request panel members to be surveyed at random points in time rather than at fixed time intervals. We thus find support for both hypotheses H_{2a} and H_{3a}. As can be seen from the hitrate, the model predicts 70% of the responses correctly. Moreover, it does not seem to seriously overpredict either response or nonresponse.

Next, we inspect the results of the regression part with the purpose to test our hypotheses regarding the effects of panel design characteristics on response quality. The first two variables we consider here are the individuals' scores on the news test and the time needed to complete this test. There is a positive effect of the number of weeks since the previous request both on the score and on the time needed to complete the test. This implies that as the time between subsequent waves gets shorter, respondents tend to score lower on the test and spend less time to complete it. Since the score on the news test should obviously not depend on the design of the panel, we interpret this result as a signal of panel conditioning bias. The fact that the time needed to complete the test tends to be shorter as the number of weeks since the previous request gets shorter can have two explanations. Respondents complete the task faster either because they are familiar with it since they they have recently completed it already, or because they devote less effort to the task as a result of the more demanding panel design. The second two variables are our measures for the stated level of irritation of an individual and the extent to which he or she feels his or her (reporting) behavior has changed due to participation in the panel study. As these variables are self-reported, they have to be interpreted with a bit more care. First of all, although there is a weak effect on irritation, the sampling frequency f does seem to influence respondents' stated level of panel conditioning bias. We do find an effect on the news test score however. This suggests that, even though respondents are biased in their (reporting) behavior due to the sampling frequency, they do not perceive this effect. In contrast, the number of waves n does seem to drive stated panel conditioning bias. Randomization seems to lower both irritation and panel conditioning bias. These two variables do not significantly influence the test performance however. In sum, we find support for hypothesis H_{1b} in the revealed data, but support for H_{2b} and H_{3b} only in the stated data.

Finally, we discuss the results for the additional regressors and demographics. The highly significant estimates of ρ suggest that there is a strong relationship between current and past responses, which cannot be ignored. The effects of respondents' news consumption levels on their performance on the news test are positive, as expected. Looking at the effects of demographics, we observe that women tend to be slightly more loyal to the experiment, as can be seen from the negative effect of gender on response and irritation. Also they tend to score slightly higher on the news test.

Study 2: Full information provided about the design

For the second study, we found 148 students willing to participate out of a sample of 292. In the first part of the experiment, we presented ten panel designs to each respondent and

Table 2.3: Results of the panel Tobit-II model for Study 1

Variable	Probit part ($\boldsymbol{R}$)		Regression parts ($\boldsymbol{Y}$)							
			News score (0 - 20)		Time needed (in mins.)		Irritation (stated; − 1.0 - 1.0)		Panel conditioning (stated; − 1.0 - 1.0)	
Design parameters										
No. of weeks since previous request	0.322**	(0.143)	0.249**	(0.095)	0.670**	(0.305)	0.017*	(0.010)	0.012	(0.014)
No. of times requested before	−0.728***	(0.054)	−0.014	(0.044)	0.001	(0.207)	0.001	(0.003)	0.011**	(0.005)
Randomized sampling	0.585***	(0.201)	0.567	(0.409)	1.070	(0.978)	−0.102***	(0.038)	−0.098*	(0.051)
Additional regressors										
Dynamics	0.939***	(0.290)	0.397***	(0.036)	0.465***	(0.048)	0.547***	(0.041)	0.682***	(0.040)
No. of hours spent consuming news			0.139**	(0.065)	−0.091	(0.062)	−0.003	(0.002)	−0.004	(0.003)
Newspaper subscriber			0.459*	(0.259)	−0.524	(0.616)	−0.044	(0.044)	−0.052	(0.032)
Demographics										
Age	0.046	(0.034)	−0.003	(0.037)	−0.074	(0.085)	−0.001	(0.003)	0.008*	(0.004)
Gender (0: female; 1: male)	−0.389*	(0.237)	0.688*	(0.411)	−0.154	(0.961)	−0.120**	(0.059)	−0.110**	(0.051)
Additional model parameters										
Intercept	−5.557***	(0.908)	6.308***	(1.115)	5.190*	(3.155)	−0.232**	(0.099)	−0.014	(0.130)
Cross-section random st. dev.	0.362		2.306		5.385		0.163		0.233	
Idiosyncratic random st. dev.	1.000[1]		1.466		9.594		0.161		0.190	
Mean of the dependent	0.683		10.223		3.255		−0.646		−0.126	
McFadden / Adjusted R^2	0.555		0.583		0.379		0.756		0.768	
Hitrate	0.702									
Prop. of correctly predicted response	0.730									
Prop. of correctly predicted nonresponse	0.642									

Note: The table presents the estimation results of the panel Tobit-II model for Study 1, given by (2.6) - (2.7). White heteroskedasticity-consistent standard errors are in parentheses. The superscripts ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The estimates of α_1, $\boldsymbol{\alpha}_2$ and the time-specific fixed effects are not displayed for ease of presentation. [1] Standardized for identification purposes.

Table 2.4: Results of the panel Probit model for Study 2

Variable		
Design parameters		
No. of weeks since previous request	0.135***	(0.016)
No. of times requested before	−0.161***	(0.011)
Randomized sampling	0.162	(0.117)
Demographics		
Age	−0.024	(0.035)
Gender (0: female; 1: male)	0.131	(0.500)
Additional model parameters		
Intercept	0.777	(0.784)
Cross-section random st. dev.	0.750	
Idiosyncratic random st. dev.	1.000[1]	
Mean of the dependent	0.362	
McFadden / Adjusted R^2	0.569	
Hitrate	0.728	
Prop. of correctly predicted acceptance	0.626	
Prop. of correctly predicted rejection	0.785	

Note: The table presents the estimation results of the panel Probit model for the individuals' acceptance decisions in Study 2. White heteroskedasticity-consistent standard errors are in parentheses. The superscripts ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The estimates of α_1, $\boldsymbol{\alpha}_2$ and the time-specific fixed effects are not displayed for ease of presentation. [1] Standardized for identification purposes.

requested them to state which of these designs would be acceptable to them. Among the accepted designs, we found the median of the maximum number of waves to be 5.5, and the median of the maximum frequency to be biweekly, which is high considered the actual participation to Study 1. There were 22 respondents who did not accept any design. To study the effects of panel design parameters on stated acceptance, we first estimate the acceptance decision using the Probit part of our Tobit-II model. The estimation results are shown in Table 2.4. Again, the effects of the sampling frequency and the number of waves are prominent. It is interesting to see, however, that a respondent's probability of acceptance is not influenced by the choice for a randomized or a time sampling strategy. Respondents state to be indifferent between the two strategies.

Next, we check whether our respondents' stated participation matches their revealed participation, by indeed surveying them according to one of their accepted designs. The results are shown in Table 2.5. Clearly, there is a big distinction between respondents'

Table 2.5: Results of the panel Tobit-II model for Study 2

Variable	Probit part ($\boldsymbol{R}$)		Regression parts ($\boldsymbol{Y}$)							
			News score (0 - 20)		Time needed (in mins.)		Irritation (stated; − 1.0 - 1.0)		Panel conditioning (stated; − 1.0 - 1.0)	
Design parameters										
No. of weeks since previous request	0.282***	(0.072)	0.329**	(0.167)	0.524**	(0.272)	0.020**	(0.011)	0.008	(0.031)
No. of times requested before	−0.583***	(0.101)	−0.071	(0.062)	−0.167	(0.304)	0.003	(0.005)	0.013**	(0.007)
Randomized sampling	0.193	(0.197)	−0.777	(0.735)	−0.698	(1.910)	−0.022	(0.068)	−0.092	(0.097)
Additional regressors										
Dynamics	0.931***	(0.296)	0.471***	(0.055)	0.698***	(0.063)	0.484***	(0.064)	0.575***	(0.063)
No. of hours spent consuming news			0.143***	(0.042)	−0.106	(0.095)	−0.006	(0.036)	−0.003	(0.005)
Newspaper subscriber			0.300	(0.639)	−0.102	(1.764)	0.172	(0.193)	0.173	(0.182)
Demographics										
Age	0.016	(0.029)	−0.101	(0.073)	−0.243	(0.197)	−0.018*	(0.015)	0.004	(0.010)
Gender (0: female; 1: male)	−0.287	(0.199)	1.118*	(0.599)	1.881	(1.651)	−0.113*	(0.060)	−0.136*	(0.085)
Additional model parameters										
Intercept	−4.712***	(1.066)	8.955***	(1.811)	8.353	(5.551)	−0.095	(0.168)	−0.154	(0.237)
Cross-section random st. dev.	0.276		2.214		4.502		0.166		0.233	
Idiosyncratic random st. dev.	1.000[1]		1.401		5.621		0.184		0.264	
Mean of the dependent	0.755		9.746		2.879		−0.611		−0.143	
McFadden / Adjusted R^2	0.639		0.594		0.555		0.741		0.780	
Hitrate	0.690									
Prop. of correctly predicted response	0.720									
Prop. of correctly predicted nonresponse	0.598									

Note: The table presents the estimation results of the panel Tobit-II model for Study 2, given by (2.6) - (2.7). White heteroskedasticity-consistent standard errors are in parentheses. The superscripts ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The estimates of α_1, $\boldsymbol{\alpha}_2$ and the time-specific fixed effects are not displayed for ease of presentation. [1] Standardized for identification purposes.

Table 2.6: Summary of hypothesis-testing results

Hypothesis		Information about the panel design	
		Not provided (Study 1)	Provided (Study 2)
H_{1a}:	The sampling frequency on response rates	Supported	Supported
H_{1b}:	The sampling frequency on response quality	Supported	Supported
H_{2a}:	The number of participation requests on response rates	Supported	Supported
H_{2b}:	The number of participation requests on response quality	Weak support[1]	Weak support[1]
H_{3a}:	Randomized sampling on response rates	Supported	Not supported
H_{3b}:	Randomized sampling on response quality	Weak support[1]	Not supported

Note: The table presents a summary of the testing results for the hypotheses in Section 2.3.
[1] Only supported by stated preference data.

promised participation and actual participation as there still is considerable nonresponse in this experiment ($1-0.755 = 24.5\%$). Finally, we examine whether there are differences in behavior between the informed respondents of Study 2 and the uninformed respondents of Study 1. The informed respondents' probability of response, performance on the news test, stated level of irritation and panel conditioning bias still seem to be influenced by the design parameters n and f in the same manner as they did in Study 1. This indicates that providing information about the design of the experiment does not alleviate response or panel conditioning bias. In contrast, the influence of randomized sampling on any of our dependent variables disappeared in the informed case. This supports our conjecture that randomization has an effect on response rates and response quality as a result of learning about the design. Response rates are improved as panel members actually have low awareness of being such a member when they are uninformed about the panel design. Response quality gets improved as respondents are less likely develop expectations as to when they will be surveyed again. As a result, they are more surprised by the survey request, and have not been able to prepare for the survey.

We summarize the testing results for our hypotheses in Table 2.6, where we distinguish between the informed and uninformed case.

Panel calibration

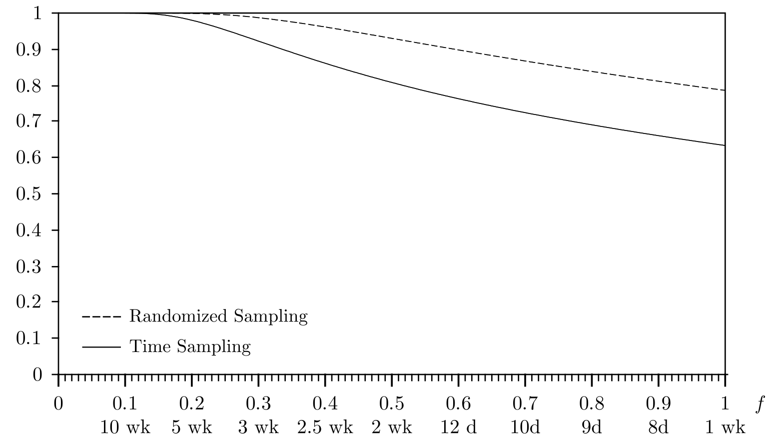
One of the advantages of our approach is that it can be used to calibrate the design of a new panel survey. To illustrate this, suppose that we conducted Study 1 as a pilot study with the aim to calibrate a continuous monitor for students' knowledge of recent news events. Using our parameter estimates, we can now simulate data from the model for various values of the design parameters **P**. These data can be used to compute the expected response rates under different scenario's. In Figure 2.3 we plot the response curves, which are the expected response rates as a function of the sampling frequency f, at various waves of data collection. We distinguish between the expected response rates obtained using a time sampling and a randomized sampling strategy.

Generally, since the sampling frequency has an effect on a student's news score, which is undesirable, but the sampling strategy and the number of waves do not have an effect, it is advisable to adopt randomized sampling, to choose a high number of waves, but with a low sampling frequency. Suppose now that our budget restricts us to have at least 50% expected response in the last wave. As can be read from the graph, in this case, if we want to collect as much as 12 waves of data, we should set the sampling frequency f equal to 0.47, which is close to biweekly, or lower. In case time sampling is to be preferred, then the frequency should be equal to 0.35 or lower.

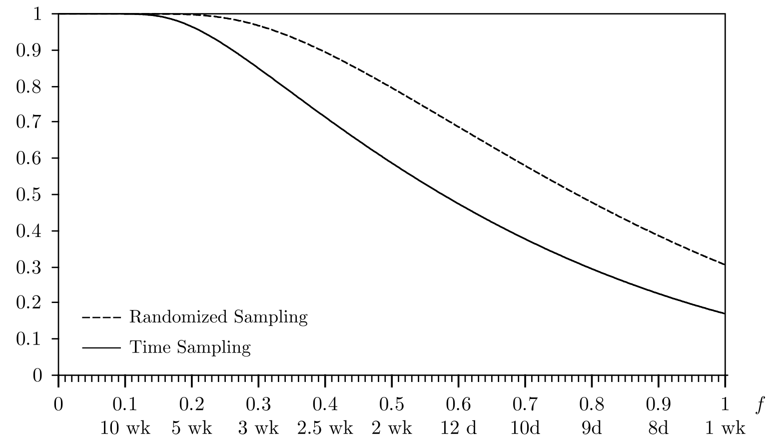
2.6 Conclusions

The response rate and the response quality of a panel survey critically depend on the design of the panel. It is therefore important to choose this design carefully. To facilitate this choice, in this chapter we have studied the effects of typical panel design characteristics on response rates and response quality. We hypothesized that response rates and response quality decrease as the sampling frequency and the number of waves of a panel increase. Further, we proposed a new sampling strategy, labeled randomized sampling, where individuals are surveyed at random time intervals. We hypothesized that, as compared to the often applied time sampling strategy, randomized sampling improves response rates and response quality. To test our hypotheses, we proposed a two-step selection model, where in the first step we explained the response decision and in the second step we explained subsequent responses, conditional on response in the first step. Two empirical studies indeed confirm the above hypotheses. We do find, however, that the effect of randomized sampling is only significant in a setting where panel members are not informed about the dates they will be surveyed, because in this case respondents are less likely to develop expectations as to when they will be surveyed again. Finally, we have demonstrated that our research design is well-suited to calibrate new panels. Careful calibration should lead to an efficient panel design where expected response is maximized and threats to response quality due to the design are suppressed. This in turn reduces survey costs, since fewer respondents need to be acquired and higher quality data is collected. The methodology

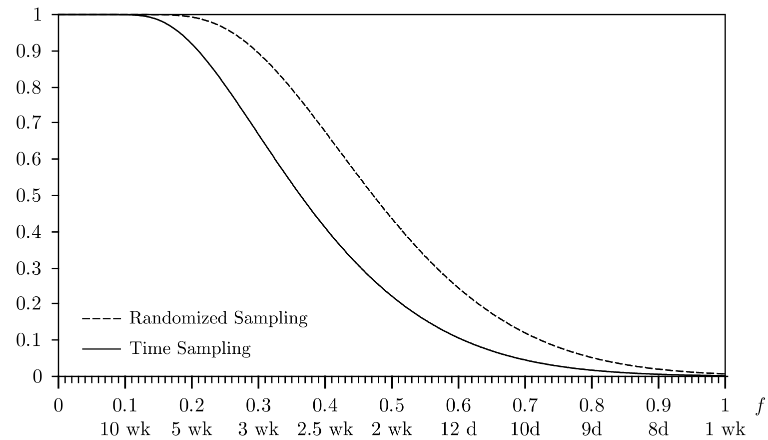
Figure 2.3: Response curves



(a) Expected response rates of the 2nd wave



(b) Expected response rates of the 6th wave



(c) Expected response rates of the 12th wave

Note: The graphs present the expected response rates as a function of the sampling frequency f, based on simulated data from the panel Tobit-II model for Study 1, given by (2.6) - (2.7). Panels (a) - (c) show the curves for the 2nd, 6th and 12th wave, respectively, where we distinguish between the expected response rates obtained using a time sampling and a randomized sampling strategy.

may also be used to capture residual response bias and panel conditioning bias, to the extent that these biases are explained by the model.

A promising avenue for further research would be to investigate whether our results also apply in extreme cases where the incentives to participate in the panel are either very weak or very strong. Examples of panels with weak participation incentives are large internet panels, where in general the owners of the panel do not know their panel members and no monetary incentives are offered. In these cases it may be helpful to concentrate on minimizing attrition rather than on minimizing wave nonresponse. Examples of panels with strong incentives can be found in health care, where patients join a panel to allow their physicians to monitor the effects of their treatments. Other interesting examples are economic expert panels, such as the Survey of Professional Forecasters conducted by the Federal Reserve Bank of Philadelphia, where panel membership is considered very prestigious. In the latter two cases the focus should be directed more towards the panel's response quality, than towards response rates. A second avenue would be to conduct more field experiments, with the aim to better understand the learning process of panel members over time.

Chapter 3

Monitoring Weekly Consumer Confidence

3.1 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 a recent 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, focussing 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 final 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 just the financial position or just employment. 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 very 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 socioeconomic 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 two research angles to be discussed below seem more promising.

A third angle concerns the fact that the data are only available at the monthly level. Indeed, businesses tend to operate in terms of weeks, and also many other economic indicators, like stock market returns, interest rates, and industry-specific figures like the number of temporary employees, are available at the weekly level. In fact, it seems that a weekly figure of consumer confidence, reported at the beginning of a new week, would be a helpful indicator for many people in business and industry. In this chapter we therefore aim to collect such data.

A fourth angle, which is also addressed in the present chapter, 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 -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. 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 the present chapter we also seek to do that, but now by applying the alternative method of data collection as introduced in Chapter 2.

In sum, in this chapter we put forward a method to collect weekly consumer confidence data at the individual level. We keep the Katona-type questions intact, but we merely focus on the collection of the data, which should be comparable from week to week, that is, we try to prevent facing ecological fallacy problems. To that end, we need to collect data such that we have the same (though not all) individuals being interviewed from one week to another, without them being annoyed or becoming uninterested.

The outline of the chapter is as follows. In Section 3.2 we present our method of data collection, and we argue that it has various convenient properties for the purpose at stake. Next, we introduce the model that will be used to describe longitudinal developments in consumer confidence at the individual level. In Section 3.3 we illustrate the usefulness of our method by surveying individuals during periods of at most three months. We show that we obtain weekly confidence data that can be compared across the weeks. We also show how to compute confidence bounds around these numbers, so that one can infer whether this week's figure is significantly different from the previous week's figure. When we compare our figures with the actually published figures by Statistics Netherlands (SN), we observe a remarkable resemblance. Hence, when we temporally aggregate our weekly figures, we get the actually published monthly data. In Section 3.4 we conclude with an outline of various areas for further research.

3.2 Methodology

In this section we present our method to collect consumer confidence data at the individual level. Additionally, we introduce the model that will be used in our empirical work.

3.2.1 Data Collection Method

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 the researcher to study developments in confidence at the individual level and to capture the dynamic relationships between confidence and events. However, surveying the very same individuals frequently, such as weekly, 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.

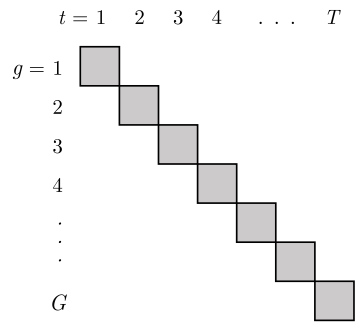
For this reason, most statistical agencies decide to collect repeated cross-sections instead of panel data. This amounts to surveying a new group of individuals at each survey occasion, which implies that individuals are surveyed only once. The design is illustrated in Figure 3.1, Panel (a). Here we index time by t, where $t = 1, \dots, T$, individuals by i, where $i = 1, \dots, N$ and groups of individuals by g, where $g = 1, \dots, G$. A grey square in row g and column t indicates that group g is requested to be surveyed at time t. Although clearly this design reduces respondent burden and eliminates potential panel conditioning biases, the design does not have the advantages of a panel. Therefore it seems promising to collect longitudinal data nevertheless, but to calibrate the design of the panel carefully, such that the above adverse effects are negligible or, at least, manageable.

In panel design, three key decisions have to be made. Firstly, as individuals obviously cannot be surveyed continuously, one has to decide on the total time-span a panel member is requested to join the panel, to be denoted by T^*. As panel members leave the panel, one may decide to invite new individuals to join the panel in such a way that the total number of panel members remains constant. This strategy is referred to as rotation, see Patterson (1950) and Kish and Hess (1959). Naturally, the next step would be to decide upon the number of survey requests within this period, to be labeled 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 subperiod. 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 data are collected continuously. The above strategy is mostly referred to as time sampling.

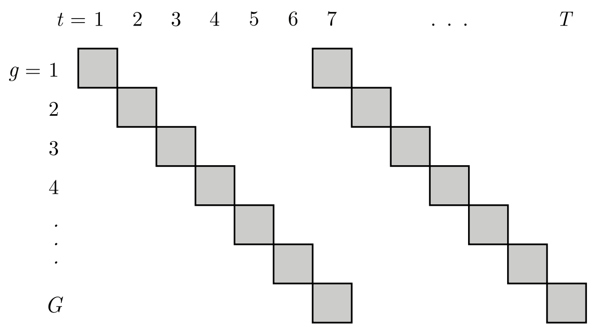
To the best of our knowledge, the only consumer confidence indicator that is not measured through repeated cross-sections is the Index of Consumer Sentiment of the University of Michigan. Michigan adopts a rotating panel design in which the respondents are requested to be re-interviewed six months after the first interview, see Curtin (1982) for details. This design is illustrated in Figure 3.1, Panel (b). In our terminology, we would characterize the Michigan panel as a rotating panel where $T^* = 12$ months, $n = 2$ survey occasions per individual and time sampling is applied.

As an alternative, the new data collection method approach which we developed in Chapter 2 seems very useful here. Recall that in this chapter, we choose the n survey occasions at random, independently for each panel member. We showed 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 calibrate the actual panel design to be used in this chapter, we can use the results as obtained in Section 2.5, as we will argue in the next section. An example of a randomized rotating panel, where two new individuals are invited to join the panel in

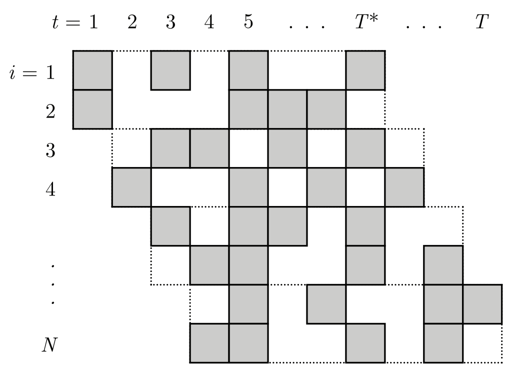
Figure 3.1: Possible panel designs to measure consumer confidence



(a) Repeated cross-sections



(b) The Michigan panel



(c) An example of a randomized rotating panel

Note: In this figure, three panel designs are visualized that can be employed to measure consumer confidence over time. In Panels (a) and (b), a grey square in row g and column t indicates that respondent group g is requested to be surveyed at time t. Likewise, in Panels (c), a grey square in row i and column t indicates that individual i is requested to be surveyed at time t. Each dotted area in this panel encloses all survey requests assigned to one particular cohort of individuals in the randomized rotating panel.

each time period is shown in Figure 3.1, Panel (c). In this example, we set the maximum time-span 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.

3.2.2 Modelling Consumer Confidence

To model respondents' answers to the five questions which together summarize their consumer confidence level, we employ a dynamic panel version of the Ordered Probit model, as originally developed by McKelvey and Zavoina (1975). The model is relevant in applications such as surveys in which respondents express their preferences on an ordinal scale. In our case, respondents are requested to indicate whether their economic situation has been or will be better (1), the same (0) or worse (-1), see Appendix 3.A for details. We index respondents by i, where $i = 1, \dots, N$, questions by j, where $j = 1, \dots, J$, and time by t, where $t = 1, \dots, T$. A respondent's unobserved assessment of the change in his or her economic situation is denoted by $y^*_{i,j,t}$ while we only observe the multinomial variable $y_{i,j,t} \in \{-1, 0, 1\}$. We assume that the latent variable $y^*_{i,j,t}$ can be explained by a set of explanatory variables $\boldsymbol{z}_{i,j,t}$ and the previous assessment through

$$y^*_{i,j,t} = \boldsymbol{z}_{i,j,t}\boldsymbol{\beta} + \rho_j^{d_{i,t}} y_{i,j,t-d_{i,t}} + a_{i,j} + e_{i,j,t} \tag{3.1}$$

where $(\boldsymbol{\beta}', \rho_1, \dots, \rho_J)$ are unknown parameters. The variable $y_{i,j,t-d_{i,t}}$ denotes the previous observation, which is measured $d_{i,t}$ time periods before $y_{i,j,t}$. The question-specific factor $\rho_j^{d_{i,t}}$ is a finite duration adjustment of the geometric lag or Koyck model, see Ansari *et al.* (2008) and Van Diepen *et al.* (2009) for recent applications. We assume that $\rho_j < 1$ for $j = 1, \dots, J$. This implies that the effect of an individual's previous opinion on his or her present opinion decreases, as the time between the present and the previous survey gets longer. The above representation allows us to analyze any incomplete panel directly, even if the observations are unequally spaced. Finally, $a_{i,j}$ denotes an individual- and question-specific random effect and $e_{i,j,t}$ denotes an idiosyncratic error term. It is assumed that $a_{i,j}$ and $e_{i,j,t}$ are mutually independent and independent of the regressors $\boldsymbol{z}$.

The latent variable $y^*_{i,j,t}$ gets mapped onto $y_{i,j,t}$ by the rule

$$y_{i,j,t} = \begin{cases} 1 & \text{if } \gamma_0 < y^*_{i,j,t} \leq \gamma_1 \\ 0 & \text{if } \gamma_{-1} < y^*_{i,j,t} \leq \gamma_0 \\ -1 & \text{if } \gamma_{-2} < y^*_{i,j,t} \leq \gamma_{-1}, \end{cases} \tag{3.2}$$

where the parameters γ_{-2} to γ_1 are unobserved thresholds which must satisfy $\gamma_{c-1} < \gamma_c$ for $c = -1, 0, 1$. Because the boundary values of our latent variable $y^*_{i,j,t}$ are unknown, we set γ_{-2} and γ_1 equal to $-\infty$ and $+\infty$, respectively. We normalize γ_{-1} to 0 in order to be able to include an intercept in the model. The threshold γ_0 will be estimated from the data.

In dynamic nonlinear panel data models with unobserved heterogeneity, such as the model specified above, the treatment of the initial observations is an important issue. An incorrect treatment of the initial observations may lead to a bias in the parameter estimates which only gets reduced when T is large, see Heckman (1981) for details. To deal with this problem, we apply the Wooldridge (2005) approach. This amounts to approximating the distribution of the random effects $a_{i,j}$ conditional on the initial conditions $y_{i,j,1}$ rather than the distribution of $y_{i,j,1}$ conditional on $a_{i,j}$, as suggested by Heckman (1981). Specifically, we assume that

$$a_{i,j}|y_{i,j,1}, \boldsymbol{z}_{i,j} \sim N(\alpha_0 + \alpha_1 y_{i,j,1} + \boldsymbol{z}_{i,j}\boldsymbol{\alpha}_2, \sigma_a^2), \tag{3.3}$$

where $\boldsymbol{z}_{i,j}$ is the subset of all nonredundant explanatory variables in $\boldsymbol{z}_{i,j,t}$ in all time periods and $(\alpha_0, \alpha_1, \boldsymbol{\alpha}_2', \sigma_a^2)$ are unknown parameters. The idiosyncratic error is assumed to satisfy

$$e_{i,j,t}|y_{i,j,t-d_{i,t}}, \ldots, y_{i,j,1}, \boldsymbol{z}_{i,j}, a_{i,j} \sim N(0, 1). \tag{3.4}$$

The variance of $e_{i,j,t}$ is set equal to 1 as no scaling of the underlying utility model can be deduced from the observed data. We assume that the dynamics are correctly specified, which means that only $y_{i,j,t-d_{i,t}}$ appears in the distribution given outcomes back to the initial time period. The variables $\boldsymbol{z}_{i,j}$ are assumed to be strictly exogenous conditional on $a_{i,j}$, see Wooldridge (2005) for details.

To derive the likelihood function of the model, it is convenient to replace $a_{i,j}$ by

$$a_{i,j} = \alpha_0 + \alpha_1 y_{i,j,1} + \boldsymbol{z}_{i,j}\boldsymbol{\alpha}_2 + a_{i,j}^*, \tag{3.5}$$

so that

$$a_{i,j}^*|y_{i,j,1}, \boldsymbol{z}_{i,j} \sim N(0, \sigma_a^2). \tag{3.6}$$

The model specified in (3.1) now reads as

$$\begin{aligned} y_{i,j,t}^* &= \boldsymbol{z}_{i,j,t}\boldsymbol{\beta} + \rho_j^{d_{i,t}} y_{i,j,t-d_{i,t}} + \alpha_0 + \alpha_1 y_{i,j,1} + \boldsymbol{z}_{i,j}\boldsymbol{\alpha}_2 + a_{i,j}^* + e_{i,j,t} \\ &= f(\boldsymbol{x}_{i,j,t}, \boldsymbol{\theta}) + a_{i,j}^* + e_{i,j,t}, \end{aligned} \tag{3.7}$$

where $\boldsymbol{x}_{i,j,t}$ summarizes the explanatory variables, $\boldsymbol{x}_{i,j,t} = (\boldsymbol{z}_{i,j,t}', y_{i,j,t-d_{i,t}}, y_{i,j,1}, \boldsymbol{z}_{i,j}')$ and $\boldsymbol{\theta}$ summarizes the parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\rho}', \alpha_0, \alpha_1, \boldsymbol{\alpha}_2', \sigma_a^2)$ with $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_J)$. For notational convenience we write $\boldsymbol{z}_{i,j} = \boldsymbol{z}$ and $a_{i,j}^* = a^*$ in the remainder. The model's conditional density of $(y_{i,j,2}, \ldots, y_{i,j,T})$ given $(y_{i,j,1}, \boldsymbol{z}', a^*, \boldsymbol{\theta}')$ is given by

$$g(y_{i,j,2}, \ldots, y_{i,j,T}|y_{i,j,1}, \boldsymbol{z}, a^*, \boldsymbol{\theta}) = \prod_{i,j,t,c} \Pr[y_{i,j,t} = c|\boldsymbol{x}_{i,j,t}]^{\mathbb{I}[y_{i,j,t}=c]} \tag{3.8}$$
$$= \prod_{i,j,t,c} \left(\Phi\big(\gamma_c - f(\boldsymbol{x}_{i,j,t}, \boldsymbol{\theta}) - a^*\big) - \Phi\big(\gamma_{c-1} - f(\boldsymbol{x}_{i,j,t}, \boldsymbol{\theta}) - a^*\big) \right)^{\mathbb{I}[y_{i,j,t}=c]},$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the Normal distribution, and $\mathbb{I}[\cdot]$ the indicator function, which takes the value 1 if the argument is true and zero otherwise. Integrating out a^*, we obtain the likelihood of $(y_{i,j,2}, \ldots, y_{i,j,T})$ conditional on $(y_{i,j,1}, \boldsymbol{z}', \boldsymbol{\theta}')$

$$\mathcal{L}(y_{i,j,2}, \ldots, y_{i,j,T}|y_{i,j,1}, \boldsymbol{z}, \boldsymbol{\theta}) = \int_{\mathbb{R}} g(y_{i,j,2}, \ldots, y_{i,j,T}|y_{i,j,1}, \boldsymbol{z}, a^*, \boldsymbol{\theta}) \times (1/\sigma_a)\phi(a^*/\sigma_a)\mathrm{d}a^*, \tag{3.9}$$

where $\phi(\cdot)$ denotes the probability density function of the Normal distribution. The model parameters can be estimated by maximizing the log conditional likelihood, $\log \mathcal{L}(y_{i,j,2}, \ldots, y_{i,j,T}|y_{i,j,1}, \boldsymbol{z}, \boldsymbol{\theta})$.

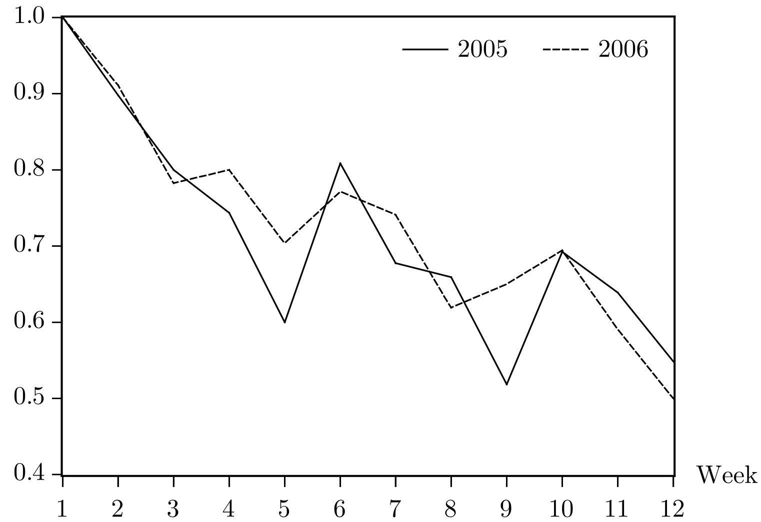
3.3 Empirical Illustration

In this section we apply our data collection method to measure consumer confidence as defined by SN among students at Erasmus University Rotterdam over two periods of three months. Using the model presented in Section 3.2 we illustrate the advances of collecting weekly panel data rather than monthly repeated cross-sectional data.

3.3.1 Panel Calibration and Data Collection

As the target population of our survey coincides with the target population in Chapter 2, we use the results as presented in Section 2.5 to calibrate the design of our panel. Particularly, we inspect the response curves as shown in Figure 2.3 to determine the maximum sampling frequency f to be used for the present study, provided that we want to survey our respondents over a period as long as 12 weeks and apply a randomized sampling strategy. The curves indicate that in order to have at least 50% expected response in the twelfth week, we need to set the sampling frequency f equal to 0.47 or lower. In order to collect as much data as possible, we decide to set f equal to this maximum value, which implies that we will survey each student close to biweekly on average, or 5.6 times within three months. Once the data are collected we verify whether indeed the response rates are

Figure 3.2: Weekly response rates



Note: The graph shows the weekly fractions of panel members who completed the consumer confidence survey in Appendix 3.A on our request.

as high as expected, and, perhaps more importantly, we use our model to check whether no signs of panel conditioning are apparent from the data.

We apply the design to two different cohorts of students, surveying the first cohort from October 2nd, 2005, to January 7th, 2006, and the second cohort from October 1st, 2006, to January 6th, 2007. We allow the students to join the panel in the first, second and third week of data collection and monitor each student up to 12 weeks, which implies that we collect 14 waves of data in total. Note that we do not apply a rotation strategy. We measure the respondents' confidence levels using the questionnaire as developed by SN, see Appendix 3.A for details. The survey is conducted online through an interactive website. All correspondence, including the participation requests, is generated automatically and sent by e-mail. No survey follow-up is performed to increase response. In total, 78 students agreed to participate in 2005 and 52 students in 2006. The response rates of the survey are shown in Figure 3.2, separately for each cohort. During the first 5 weeks, typically the response rate decreased by about 8% per week, whereas in the later weeks it decreased by about 4%. The response rate of the 12th week was still just above 50% for both cohorts, as desired.

3.3.2 Comparison with Officially Published Statistics Netherlands Data

In Table 3.1 we compare our *weekly* consumer confidence indicator to *monthly* consumer confidence as measured by SN. SN report their indicator for the Netherlands around the 22nd of each month, and the surveys are conducted during the first ten working days of that particular month. Therefore we compare the average of our weekly indicator over the first two weeks, which is in fact a biweekly indicator, to the monthly indicator of SN. As students, especially those close to graduation, are generally more positive about their future financial situation as compared to the overall Dutch population, we anticipate our levels of CCI to be higher as compared to the levels as recorded by SN, who use a representative sample of the population. The longitudinal developments in our indicator, however, should be roughly similar. Especially in 2005 this is indeed the case. SN measured an increase of 4% in November, where we measured an increase of 5%. Also the changes in December (3% versus 2%) and January (6% versus 4%) compare well. There is slightly more variation in the 2006 estimates, possible due to the smaller sample of only 52 respondents. Overall, however, we believe our longitudinal changes resemble those of SN quite closely.

For ease of comparison, we also plotted the weekly CCI levels as collected using our randomized panel against the monthly CCI levels of SN in Figure 3.3. It is interesting to see that changes in CCI from week to week can be substantial. This supports our conjecture that it is useful to measure CCI at a higher frequency. The figure also nicely illustrates that the developments in weekly and monthly CCI are equal in the longer run. Both indicators show a steady upward trend in 2005, a slight decrease in confidence from October to November 2006, and another increase in confidence towards the end of 2006.

3.3.3 The Dynamics of Consumer Confidence

The second angle for potential improvement of consumer confidence indicators which we pursue with our data collection method is the fact that we collect observations of the same individuals at multiple moments in time. This allows us to study developments in consumer confidence at the individual level, correcting for heterogeneity among individuals. In particular, we may assess to what extent developments in consumer confidence are driven on the one hand by the observed and unobserved characteristics of respondents and on the other hand by state dependence. For this purpose, we analyze our consumer confidence panel data set using the dynamic Ordered Probit model with random effects, as put forward in Section 3.2.

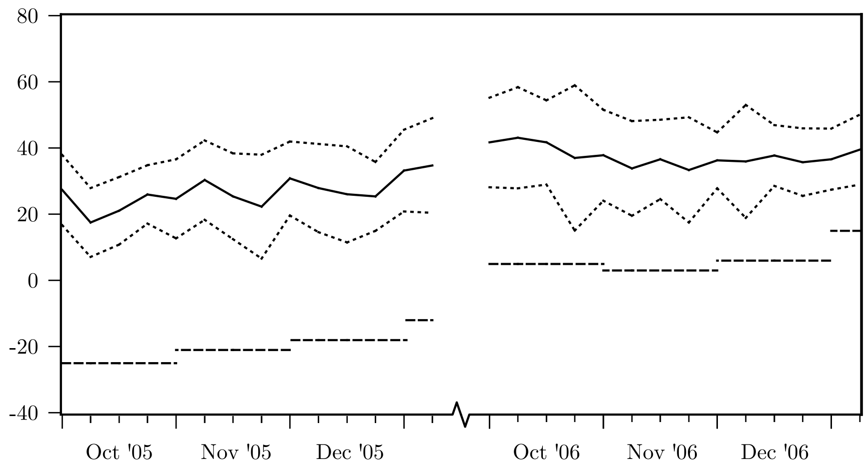
Table 3.1: A comparison of weekly CCI to monthly CCI

Month (Approx.)	Week	(a) Randomized panel data					(b) Repeated cross-sections	
		Weekly CCI			Biweekly CCI		Monthly CCI	
		Average Level		Weekly change	Average Level	Monthly change	Average Level	Monthly change
Oct '05	41	27	(5.4)	–	23 (3.8)	–	-25	–
	42	19	(5.3)	-8				
	43	21	(5.2)	2	24 (3.6)	–		
	44	28	(4.5)	7				
Nov '05	45	26	(6.1)	-3	28 (4.4)	5	-21	4
	46	32	(6.1)	6				
	47	27	(6.6)	-6	26 (5.1)	2		
	48	25	(8.0)	-1				
Dec '05	49	32	(5.7)	6	30 (4.4)	2	-18	3
	50	28	(6.8)	-3				
	51	27	(7.4)	-1	28 (5.0)	2		
	52	29	(5.3)	2				
Jan '06	53	33	(6.3)	4	34 (5.5)	4	-12	6
	1	35	(7.3)	1				
Oct '06	40	42	(6.9)	–	42 (5.6)	–	5	–
	41	43	(7.8)	1				
	42	42	(6.5)	-1	40 (5.8)	–		
	43	37	(11.2)	-5				
Nov '06	44	38	(7.0)	1	36 (5.1)	-6	3	-2
	45	34	(7.3)	-4				
	46	37	(6.1)	3	36 (4.9)	-4		
	47	33	(8.1)	-3				
Dec '06	48	36	(4.3)	3	36 (4.1)	0	6	3
	49	36	(8.7)	0				
	50	38	(4.7)	2	37 (3.5)	1		
	51	36	(5.2)	-2				
Jan '07	52	37	(4.7)	1	38 (3.5)	4	15	9
	1	40	(5.4)	3				

Note: The table compares the weekly consumer confidence indicator as measured in this chapter to the monthly seasonally unadjusted consumer confidence indicator as measured by Statistics Netherlands. The average levels and changes of the weekly indicator are shown in Panel (a), whereas the average levels and changes of the monthly indicator are shown in Panel (b). Since the monthly indicator of Statistics Netherlands is in fact a biweekly indicator, measured over the first two weeks of each month, Panel (a) also shows the average biweekly levels and monthly changes in these biweekly levels, based on the weekly data. The monthly changes are comparable across the two data collection methods. Standard errors are in parentheses.

We include the following explanatory variables. Firstly, to explicitly control for some observed heterogeneity, we include the demographics age and gender as $\boldsymbol{z}_{i,j}$ variables. Secondly, individual-specific dummies, with coefficients denoted by κ_i, are included to account for unobserved heterogeneity among respondents. Likewise, we aim to capture

Figure 3.3: A comparison of weekly CCI to monthly CCI



Note: The graph compares the weekly consumer confidence indicator as measured in this chapter (solid line) to the monthly indicator as measured by SN (dashed line). The dotted lines represent the upper and lower 95% confidence bounds of the weekly indicator.

possible heterogeneity among specific questions and weeks by question- and week-specific dummies with coefficients λ_j and μ_t, respectively. Finally, we want to verify whether there are no signs of panel conditioning bias apparent from the data. This requires a careful comparison of the responses given in the first wave of data collection, which is free of panel conditioning bias by definition, and in the next waves. We extend the notion of Hansen (1980) who argues that there should not be a difference in the response distribution of different subgroups of panel members who have been exposed to different methods of data collection. This implies that responses should not depend on the particular panel design chosen, nor on the wave of data collection. To verify this, we include the panel design parameters as explanatory variables, and test for their (joint) significance. We include n directly through a variable that indicates the number of times respondents have been requested to be surveyed previously. As the sampling frequency f is set equal to 0.47 for all panel members, we cannot include this variable directly. However, we can include a variable that indicates the number of weeks since the previous participation request, which is $1/f$ in expectation. Note that by construction the above two variables change over time. For this reason, we include the values of the two variables in all time periods as $\boldsymbol{z}_{i,j}$ variables, and their values in week t as $\boldsymbol{z}_{i,j,t}$ variables.

Estimation Results

We estimate the parameters of the above model on the 2005 data, and use the 2006 data for an out-of-sample forecasting evaluation. The estimation results are shown in Table 3.2. In order to detect possibly redundant explanatory variables, we consider in each column a different subset of these variables. First of all, we note that there are no signs of panel conditioning in the collected data, as none of the variables that are based on the panel design parameters have an effect on $y_{i,j,t}$ or on unobserved heterogeneity as measured by $a_{i,j}$ at any reasonable level of significance, irrespective of the model specification chosen. This indicates that the design is calibrated properly. A second general observation is that we do not find evidence for differences in consumer confidence based on age or gender. The absence of an age effect, however, is most likely simply due to the fact there is not enough variation in age among the students interviewed.

For the other variables the results differ among the different specification. The first column displays the most general version of the model, where we included the question-, time-, as well as individual-specific dummy variables. For this specification, there seems to be state dependence only in the answers given to Question 4 and to a lesser extent in the answers given to Question 5, as only the coefficients $\hat{\rho}_4$ and $\hat{\rho}_5$ are statistically different from zero. The bottom panel displays the in- and out-of-sample hit rates, as defined in Franses and Paap (2001, Section 5.3). The in-sample hit rate indicates the fraction of correctly explained answers to the five CCI questions as given by the 2005 cohort, whereas the out-of-sample hit rate indicates the fraction of correctly predicted answers as given by the 2006 cohort. Model 1 explains 70% of the answers correctly, which is promising. As the data for 2006 are collected using a new cohort of students, we substitute the average estimate of the individual- and time-specific coefficients for every individual in every week of the out-of-sample period. Even though this greatly simplifies the model, it still predicts 55% of the answers as given by this cohort correctly.

The second column presents the results for the model without individual-specific dummy variables. Perhaps not surprising, this specification allows more room for state dependence, as unobserved heterogeneity is only captured by the model's random effects, see Keane (1997) and the discussion in Erdem and Sun (2001) in the context of the Wooldridge (2005) approach. This also holds for Model 3, where also the time-specifics are excluded, and Model 4, where even the question-specifics are excluded. Looking across the results for Models 2 to 4, we conclude that there is more state dependence in the answers given to Questions 3, 4 and 5, which together constitute the Willingness To Buy indicator, than in the answers given to Questions 1 and 2, which together constitute the Economic Climate indicator, see Appendix 3.A. Despite this result, in the rightmost col-

Table 3.2: Estimation results of the dynamic Ordered Probit model with random effects

Variable		Model 1.	Model 2.	Model 3.	Model 4.	Model 5.
<u>Panel design parameters</u>						
No. of weeks since prev. request	β_1	0.011	−0.023	0.019	0.020	0.012
		(0.504)	(0.082)	(0.032)	(0.031)	(0.029)
No. of times requested before	β_2	0.003	−0.021	0.005	0.002	0.006
		(0.186)	(0.046)	(0.029)	(0.029)	(0.026)
<u>Demographics</u>						
Age	$\alpha_{2,1}$	−0.028	−0.012	−0.013	—	−0.014
		(0.124)	(0.022)	(0.014)		(0.011)
Gender	$\alpha_{2,2}$	0.175	0.240	0.248	—	0.214
		(0.850)	(0.199)	(0.164)		(0.138)
<u>State dependence</u>						
Question 1	ρ_1	0.814	0.857***	0.854***	0.858***	—
		(1.223)	(0.133)	(0.123)	(0.096)	
Question 2	ρ_2	0.715	0.764***	0.764***	0.762***	—
		(0.692)	(0.144)	(0.126)	(0.095)	
Question 3	ρ_3	0.847	0.875***	0.867***	0.868***	—
		(0.689)	(0.107)	(0.097)	(0.069)	
Question 4	ρ_4	0.900**	0.923***	0.917***	0.921***	—
		(0.357)	(0.106)	(0.096)	(0.077)	
Question 5	ρ_5	0.908*	0.902***	0.891***	0.898***	—
		(0.550)	(0.113)	(0.099)	(0.080)	
All questions	ρ	—	—	—	—	0.898***
						(0.043)
<u>Additional model parameters</u>						
Intercept	α_0	0.904	0.725	0.248	0.143	0.092
		(6.620)	(0.928)	(0.455)	(0.260)	(0.315)
Initial condition	α_1	0.281	0.282***	0.276***	0.280***	0.302***
		(0.770)	(0.069)	(0.057)	(0.056)	(0.054)
Threshold	γ_0	2.205***	2.169***	2.149***	2.149***	2.078***
		(0.288)	(0.058)	(0.053)	(0.053)	(0.042)
Variance of the random effects	σ_a^2	0.237***	0.272***	0.265***	0.284***	0.207***
		(0.532)	(0.074)	(0.054)	(0.056)	(0.053)
Question-specifics	λ_j	Included	Included	Included	—	—
Time-specifics	μ_t	Included	Included	—	—	—
Individual-specifics	κ_i	Included	—	—	—	—
Hit rate in-sample		0.712	0.716	0.727	0.722	0.713
Hit rate out-of-sample		0.546	0.556	0.607	0.538	0.548
Max. log-likelihood value		-963.1	-1007.8	-1014.1	-1016.8	-1041.4

Note: The table presents the estimation results of the dynamic Ordered Probit model with random effects given by (3.1) - (3.4) for five different subsets of explanatory variables. The superscripts ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The estimates of the question-, time- and individual-specific parameters, as well as the $\boldsymbol{\alpha}_2$ parameters that measure the effects of the panel design parameters in all time periods on $y_{i,j,t}$, are not displayed for ease of presentation. Standard errors are in parentheses.

Table 3.3: Prediction-realization table

Observed	Cohort 2005: In-sample predictions				Cohort 2006: Out-of-sample predictions			
	Negative (−1)	Neutral (0)	Positive (1)	Total	Negative (−1)	Neutral (0)	Positive (1)	Total
Negative	0.045	0.047	0.007	0.099	0.007	0.036	0.005	0.048
Neutral	0.028	0.377	0.090	0.494	0.020	0.325	0.169	0.513
Positive	0.007	0.094	0.305	0.406	0.002	0.162	0.274	0.439
Total	0.080	0.519	0.401	**0.727**	0.030	0.522	0.448	**0.607**

Note: The table presents the prediction-realization tables for both the in-sample estimates and the out-of-sample predictions obtained using Model 3. The bold-faced numbers are the hit rates of the model, which are computed as the sum of the diagonal elements. Possible inconsistencies are due to rounding.

umn we also present the result of a specification in which all questions share the same dynamic parameter ρ. This simplification, however, leads to a substantial decrease in likelihood. The initial values of $y_{i,j,t}$ is highly significant across all but the first specification, which indicates that in general there is substantial correlation between the initial condition and unobserved heterogeneity.

Finally, inspecting the hit rates across the different specifications, we observe that the in-sample hit rate differs only very marginally. This indicates that there is not much difference in terms of in-sample fit. In terms of out-of-sample forecasting performance, however, Model 3 is clearly to be preferred above the other model specifications, with an out-of-sample hit rate of 61%. This suggests that the time- and individual-specific variables have no added value for out-of-sample classification. For this reason we use Model 3 in the remainder of our analysis.

The prediction-realization tables for both the in-sample estimates and the out-of-sample predictions obtained using Model 3 are shown in Table 3.3. The table allows us to check whether the model tends to predict well in all answer categories, or whether it owes its hit rate solely to good predictions in a subset of the categories. We find that our model indeed predicts neutral and positive answers best, but that these two answer categories account for over 90% of the data. The model tends to have more difficulties predicting negative answers correctly, with a failure rate slight over 50%. However, as the share of negative answers is small, this does not affect the model's overall performance substantially.

3.3.4 Correcting for Changes in the Sample Composition

Consumer confidence indicators may change due to a change in the average level of consumer confidence among the population, as desired, but also simply due to fact that at each survey occasion different individuals are surveyed. Obviously data collection agencies seek to ensure that each time a representative sample is drawn from the population. Nevertheless, this causes additional variation in the indicator, which is undesirable. Our data collection method reduces the uncertainty due to changes in the sample composition as respondents are not replaced at each wave of data collection but join the panel for a prolonged period of time. However, since at each time period we only request a subsample of our panel members to be surveyed rather than all members, and some do not respond, our index is still prone to this type of uncertainty. We correct for this by imputing the missing values in our panel by simulated model predictions. In particular, we apply a multiple imputation approach, as in Schafer (1997) and Little and Rubin (2002). This implies that we replace all missing values by simulated model predictions not just once, but S times, where S is the number of simulation runs. The S complete panels are then used to compute an average consumer confidence index, which accounts for imputation uncertainty and sampling uncertainty. It proves to be advantageous to impute forward in time, so that a realization of the past observations, $\hat{y}_{i,j,t-1}$, is always available by the time $y_{i,j,t}$ has to be imputed. This allows us to substitute the dynamic component of the model in (3.1), which is $\rho_j^{d_{i,t}} y_{i,j,t-d_{i,t}}$, by $\hat{\rho}_j \hat{y}_{i,j,t-1}$.

In order to be able to compare the model imputations with the observed data, we first plot the consumer confidence indicator solely based on the imputations in Figure 3.4, Panel (a). The imputed indicator is based on $S = 10,000$ imputed panels. A comparison between the imputed indicator and the observed indicator as displayed in Figure 3.3 shows that the imputed indicator resembles the observed indicator rather closely, also during the out-of-sample period. This reinforces our confidence in the model. In Figure 3.4, Panel (b) we show the indicator based on the complete panel, that is, on the observed panel where all missing values are substituted by model imputations. Note that missing values are either due to nonresponse or due to the design of the panel. As nonresponse increases over time, typically during the first weeks a smaller portion of the data is missing as compared to the final weeks. This explains why the confidence bounds around the observed indicator are tighter during the first weeks and less tight during the final weeks, whereas for the imputed indicator the opposite holds. As a consequence, the combined indicator relies more heavily on model imputations during the final weeks. This potential adverse effect can be avoided by applying a rotation strategy. Obviously, for the complete index to be correct, we do have to assume that our model is correctly specified. Whether it is to be preferred to use the complete index instead of the observed index, depends on the particular application of the indicator and one's belief in the imputation model.

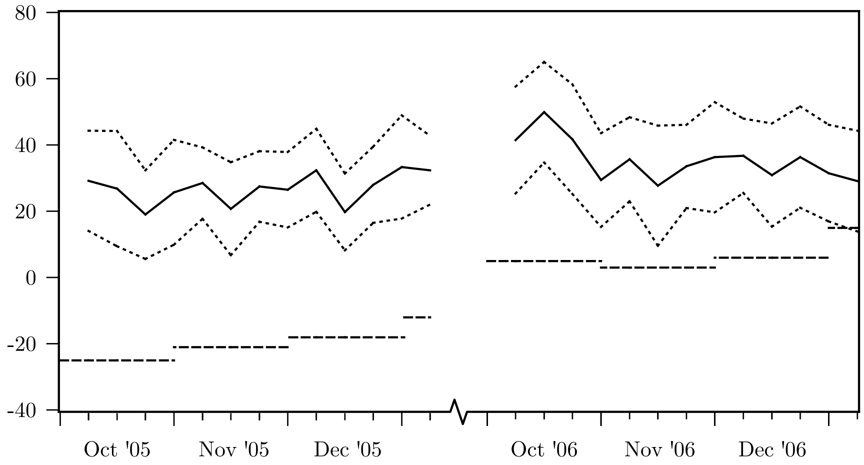
3.3.5 Weekly Changes in Consumer Confidence

We conclude our analysis with an assessment of changes in consumer confidence. Recall that one of the advances of collecting panel data rather than repeated cross-section is that panel data allow us to exactly track changes in respondents' opinions over time at the individual level. This allows us, for example, to identify whether an increase in the index is due to respondents changing their opinions from negative to neutral, or from neutral to positive. On the other hand, suppose that the index does not change significantly, we may assess whether this is because respondents' opinions have not changed, or whether their opinions have changed, but in such a way that the share of respondents who became more negative is equal to the share of respondents who became more positive. This type of polarization in opinions may have important implications to policy makers, yet it cannot be detected on the basis repeated cross-sections.

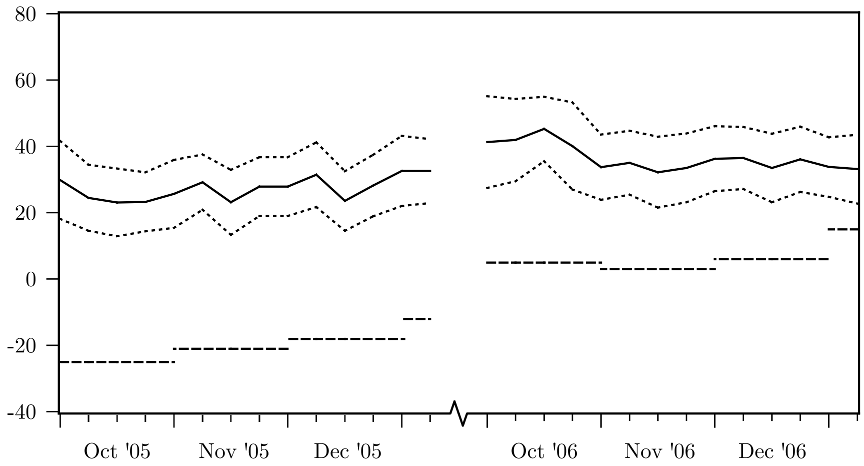
In Table 3.4, we tabulate the transition rates from any previous state $y_{i,j,t-d_{i,t}}$ to the next state $y_{i,j,t}$, separately for our two cohorts. This table reveals that respondents changed their answers frequently. In fact, in 2005, 28% $(1 - 0.718)$ of the answers given to the five questions which together comprise a respondent's confidence level changed as compared to the previous answers given by the same respondent. Strikingly, in 2006, even 44% $(1 - 0.564)$ of the answers changed. Note that, based on the overall index in Figure 3.3, it would have been tempting to conclude that there were no significant changes in opinions during this period, which clearly is totally wrong.

The changes in opinions as tabulated in Table 3.4 occurred between time $t - d_{i,t}$ and time t, that is, between the previous and the present survey occasion. As in our panel design the time between subsequent waves is random, this complicates the interpretation of these changes. It would be more insightful to study weekly changes instead. For this purpose, we may again repeatedly impute the missing values in our panel data set, and calculate the transition rates separately for all imputed panels. The variation in these transition rates across the different imputed panels allows us to derive confidence bounds around the point estimates of the transition rates. Both the point estimates as well as the 95% confidence bounds are shown in Table 3.5. Clearly, the percentage of unchanged opinions in this figure is lower as compared to Table 3.4. Note that, effectively, in the latter table we showed the transition rates based on an average time between the past and present state of $1/f = 2.13$ weeks. As the percentages of unchanged answers is

Figure 3.4: A comparison of weekly CCI using model imputations to monthly CCI



(a) Model imputations



(b) Combination of observed data and model imputations

Note: In Panel (a) of this figure the model imputations of the weekly consumer confidence indicator (solid line) are compared to the monthly indicator as measured by SN (dashed line). Similarly, in Panel (b) the composite weekly consumer confidence indicator, which is composed of the observed data as depicted in Figure 3.3 and the imputed data as depicted in Panel (a), is compared to the monthly indicator as measured by SN. The dotted lines represent the upper and lower 95% confidence bounds of the weekly indicators. The results are based on 10,000 imputed panels.

Table 3.4: Transition rates in the incomplete panel data set

From ($y_{i,j,t-d_{i,t}}$)	Cohort 2005: To ($y_{i,j,t}$) Negative (−1)	Neutral (0)	Positive (1)	Total	Cohort 2006: To ($y_{i,j,t}$) Negative (−1)	Neutral (0)	Positive (1)	Total
Negative	0.060	0.043	0.013	0.116	0.010	0.048	0.006	0.064
Neutral	0.032	0.353	0.088	0.474	0.031	0.279	0.158	0.468
Positive	0.008	0.098	0.305	0.410	0.003	0.189	0.276	0.469
Total	0.099	0.494	0.406	**0.718**	0.045	0.516	0.440	**0.564**

Note: The table presents the transition rates in the incomplete panel data set. The bold-faced numbers are the fractions of unchanged answers, which are computed as the sum of the diagonal elements. Possible inconsistencies are due to rounding.

significantly lower at the weekly level, we conclude that individuals are more likely to change their opinions in the longer run.

Finally, we look at the developments over time in the transition rates. To visualize these it is convenient to classify the transitions into three categories, which are transitions towards a more positive state, transitions to the same state, and transitions towards a more negative state. In Figure 3.5, we indicate the shares of the transitions within each of these three categories by the white, light and dark grey area's, respectively. The dotted lines indicate the confidence bounds around the shares of positive and negative transitions. This figure illustrates perhaps most convincingly that, while the CCI was stable over time in 2006, there was a lot of variability in individuals' answers over this period. Especially in October and December 2006 we observe large up- and downswings. This could not have been noticed on the basis of repeated cross-sections.

3.4 Conclusions

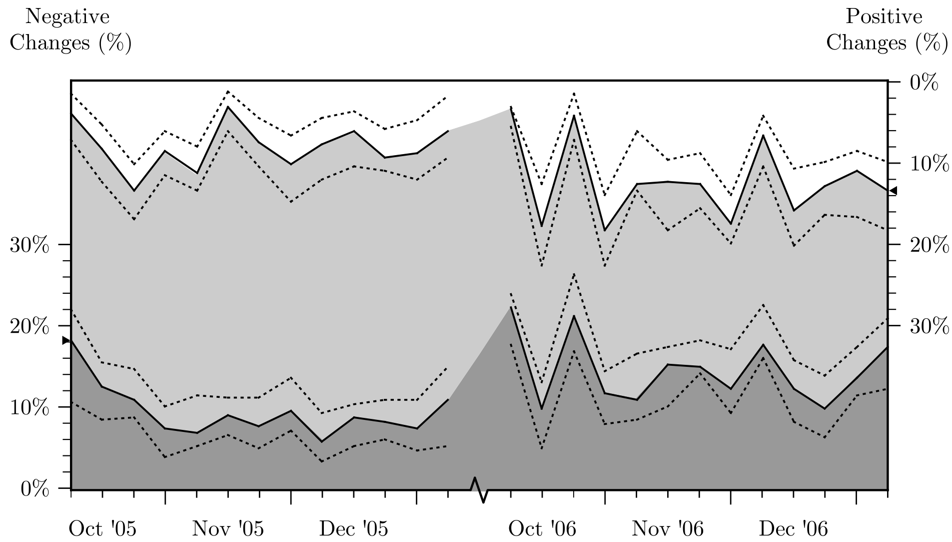
In this chapter we considered two angles for potential improvement of monitors for consumer confidence. Firstly, we considered measuring consumer confidence at the weekly instead of the monthly level. Secondly, we considered collecting panel data rather than repeated cross-sections. This allowed us to measure and statistically test longitudinal changes in weekly consumer confidence. We verified that upon temporal aggregation these changes matched with the officially published ones, and obtained evidence that reliable and more insightful indicators may be constructed on the basis of relatively small panels rather than on larger repeated cross-sections.

Table 3.5: Transition rates in the imputed panel data set

From	Cohort 2005: To ($y_{i,j,t}$)				Cohort 2006: To ($y_{i,j,t}$)			
($y_{i,j,t-1}$)	Negative (−1)	Neutral (0)	Positive (1)	Total	Negative (−1)	Neutral (0)	Positive (1)	Total
Negative	0.081 (0.070, 0.093)	0.027 (0.016, 0.038)	0.009 (0.004, 0.015)	0.117 (0.090, 0.146)	0.039 (0.031, 0.048)	0.032 (0.021, 0.0440)	0.007 (0.002, 0.013)	0.078 (0.054, 0.105)
Neutral	0.023 (0.013, 0.034)	0.430 (0.400, 0.459)	0.059 (0.043, 0.074)	0.512 (0.456, 0.568)	0.023 (0.014, 0.033)	0.367 (0.334, 0.400)	0.107 (0.089, 0.126)	0.498 (0.437, 0.559)
Positive	0.008 (0.003, 0.014)	0.050 (0.034, 0.066)	0.314 (0.290, 0.338)	0.372 (0.326, 0.418)	0.009 (0.003, 0.017)	0.091 (0.071, 0.110)	0.324 (0.294, 0.354)	0.425 (0.369, 0.481)
Total	0.112 (0.086, 0.141)	0.506 (0.450, 0.564)	0.381 (0.336, 0.427)	**0.825** **(0.760, 0.891)**	0.071 (0.048, 0.098)	0.490 (0.426, 0.554)	0.439 (0.386, 0.493)	**0.731** **(0.659, 0.802)**

Note: The table presents the transition rates in the imputed panel data set. The lower and upper 95% confidence bounds of these rates are shown in parentheses. The bold-faced numbers are the fractions of unchanged answers, which are computed as the sum of the diagonal elements. Possible inconsistencies are due to rounding. The results are based on 10,000 imputed panels.

Figure 3.5: Developments in weekly changes in consumer confidence over time



Note: This figure displays the share of respondents over time who changed their answer in the positive direction (white area, values on the right axis) and in the negative direction (dark grey area, values on the left axis), as compared to their (imputed) answer in the previous week. The share of respondents who did not change their answer is indicated by the light grey area. The dotted lines represent the upper and lower 95% confidence bounds of the shares of positive and negative answers. The results are based on 10,000 imputed panels.

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. Secondly, we can use our data collection methods in other application areas such as customer monitoring in marketing. An application in finance would be to monitor the perceived conditions of financial markets with the aim to construct a financial barometer. Finally, it may be important to study whether the parameters of our model vary over time. This is particularly relevant if the model is used for imputation. Possibly it is necessary to regularly update the parameter estimates or to develop a time-varying parameter model for our purpose. We may also consider enriching the model with various other explanatory variables to obtain better insights into the determinants of consumer confidence. For example, it has been hypothesized that factors such as mood, temperature 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.

3.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 over time.[1]

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)

[1] 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 percentage of positive answers minus that of negative answers, averaged over Questions 1 and 2. Similarly, the willingness to buy indicator is computed as the percentage of positive minus negative answers, averaged over Questions 3 to 5. Finally, the consumer confidence indicator is defined as the average of the economic climate indicator and the willingness to buy indicator.

Part II

Monitoring the State of the Economy

Chapter 4

Do Leading Indicators Lead Peaks More Than Troughs?

4.1 Introduction

Reliable leading indicators of the business cycle are of great importance for policy-makers, firms, and investors. It is therefore not surprising that economists set out on an intensive quest for such leading indicators, ever since the initial attempts of Mitchell and Burns (1938) for the US economy. This research has provided much insight into the construction, use, and evaluation of leading indicators, see Marcellino (2006) for a recent survey.

Reliability of a leading indicator variable includes aspects such as consistency and timeliness. Consistency refers to the property that a leading indicator should systematically give an accurate indication of the future course of the economy and should not produce false turning point signals too frequently, for example. Timeliness means that in order to be useful, a leading indicator variable should have a considerable lead time with respect to business cycle turning points. Most of the currently popular leading indicator variables are believed to have a lead time between six and eighteen months. At the same time, it appears to be the case that many of these variables have a considerably longer lead time at business cycle peaks than at troughs. For example, the Composite Leading Index (CLI) currently published by The Conference Board has led cyclical downturns in the economy by eight to twenty months, and upturns by one to ten months during the post-World War II period (The Conference Board, 2001).

In this chapter we develop a formal approach to investigate whether leading indicator variables have different lead times at peaks and at troughs. For this purpose, we propose a novel Markov switching vector autoregressive model, where economic growth and leading indicator variables share a common nonlinear cycle determined by a single Markov process, but such that their regime-switching is not exactly synchronous with the length of the displacement. Instead, the lead/lag times are allowed to vary across the different regimes. We follow a Bayesian approach for estimation of the model parameters, with posterior results being obtained through flexible Markov Chain Monte Carlo techniques. The advantage of Bayesian analysis of the model is that it allows us to treat the lead/lag times as unknown parameters. We can use their posterior distributions to conduct statistical inference on the asymmetry of the lead/lag structure at peaks and at troughs.

We provide an empirical application involving The Conference Board's monthly Composite Coincident Index (CCI)[1] and Composite Leading Index (CLI) over the period January 1959 - June 2007. We find that on average the CLI leads CCI by nearly one year at peaks, but only by one quarter at troughs. This suggests that, in terms of timeliness, the CLI is most useful for signaling oncoming recessions. The posterior results provide convincing evidence in favor of the presence of a non-synchronous common cycle with asymmetric lead times. The Bayes factor relative to an alternative specification with equal lead times at cyclical downturns and upturns, is very large. The same applies to models with synchronous cycles and with independent cycles in the different variables. In addition, the CLI is more consistent and more timely in terms of signaling oncoming recessions when embedded in the general model specification. In order to examine the practical usefulness of allowing for asymmetric lead times we conduct a business cycle dating and forecasting exercise for the period from December 1988 - July 2007, using real-time data for both the CLI and CCI. We find that allowing for asymmetric lead times leads to more timely and precise identification of peaks and troughs for the 1990/1991 and 2001 recessions, as well as more accurate out-of-sample forecasts of turning points and CCI growth rates.

The chapter is organized as follows. In Section 4.2, we introduce our novel Markov switching vector autoregressive model. In addition, we describe (nested) alternative specifications, which allow for a non-synchronous common cycle but with identical lead times at all possible regime switches, for a synchronous common cycle, and for independent cycles. To facilitate the interpretation of the models, we focus on the bivariate case, where both economic growth (or the coincident indicators) and the leading indicators are represented by a single variable. We provide details of the Bayesian approach for parameter estimation and inference in Section 4.3. In Section 4.4 we discuss the empirical results

[1]Note that in Part I the abbreviation CCI stands for Consumer Confidence Indicator, whereas CCI refers to the Composite Coincident Index in this part.

based on estimating the different model specifications over the complete sample period. In Section 4.5, we consider the real-time performance of the alternative cycle representations in terms of identifying peaks and troughs, and with respect to forecasting turning points and CCI growth. Finally, we conclude in Section 4.6.

4.2 Model Specification

Our point of departure is the assumption that the cycles in both the coincident indicator of aggregate economic activity (which can be an index but also an individual measure such as industrial production or GDP) and the leading indicator consist of two regimes (although extensions to multiple regimes are possible, as discussed below) labeled 'recession' and 'expansion', which are characterized by different mean growth rates of these variables. To make this precise, let $y_{1,t}$ and $y_{2,t}$ denote the growth rates of the coincident and leading indicators, respectively, in period t. Consider the unobserved binary random variables $s_{1,t}$ and $s_{2,t}$, where $s_{j,t}$ takes the value 0 in case $y_{j,t}$ is in expansion and 1 in case $y_{j,t}$ is in the recession regime. The mean growth rate conditional on the state $s_{j,t}$ is denoted as $\mu_{j,s_{j,t}} \equiv \mathsf{E}[y_{j,t}|s_{j,t}]$, for $j = 1, 2$, where typically $\mu_{j,1} < 0 < \mu_{j,0}$ such that recessions and expansions correspond to periods with negative and positive average growth, respectively. The properties of $s_{1,t}$ and $s_{2,t}$ determine the relationship between the cyclical behavior of the coincident and leading indicators, and we return to these in detail below. For the moment it is sufficient to say that they are assumed to be homogeneous first-order Markov processes. Finally, assuming first-order autoregressive dynamics in the demeaned growth rates, we arrive at the specification

$$\begin{aligned} y_{1,t} - \mu_{1,s_{1,t}} &= \phi_{1,1}(y_{1,t-1} - \mu_{1,s_{1,t-1}}) + \phi_{1,2}(y_{2,t-1} - \mu_{2,s_{2,t-1}}) + \varepsilon_{1,t} \\ y_{2,t} - \mu_{2,s_{2,t}} &= \phi_{2,1}(y_{1,t-1} - \mu_{1,s_{1,t-1}}) + \phi_{2,2}(y_{2,t-1} - \mu_{2,s_{2,t-1}}) + \varepsilon_{2,t}, \end{aligned} \tag{4.1}$$

where

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim \text{i.i.d.} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{pmatrix} \right), \tag{4.2}$$

and ε_{1,t_1} and ε_{2,t_1} are independent of s_{1,t_2} and s_{2,t_2} for all t_1 and t_2. We can write this model in vector notation as

$$(\boldsymbol{Y}_t - \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t}) = \Phi(\boldsymbol{Y}_{t-1} - \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_{t-1}}) + \boldsymbol{\mathcal{E}}_t \text{ with } \boldsymbol{\mathcal{E}}_t \sim N(0, \boldsymbol{\Sigma}), \tag{4.3}$$

where

$$\boldsymbol{Y}_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}, \quad \boldsymbol{\mathcal{S}}_t = \begin{pmatrix} s_{1,t} \\ s_{2,t} \end{pmatrix}, \quad \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t} = \begin{pmatrix} \mu_{s_{1,t}} \\ \mu_{s_{2,t}} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\mathcal{E}}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}. \tag{4.4}$$

The Markov switching vector autoregressive (MS-VAR) model in (4.3) obviously needs to be completed by specifying the exact dynamic properties of $s_{1,t}$ and $s_{2,t}$. We consider four different specifications, which allow for varying degrees of interrelation between the cycles in the coincident and leading indicators. Firstly, an extreme standpoint would be to assume that these cycles are completely independent. In this case, the state vectors $s_{1,t}$ and $s_{2,t}$ can be defined as two independent first-order two-state homogeneous Markov processes with transition probabilities

$$\Pr[s_{j,t} = 0|s_{j,t-1} = 0] = p_j \quad \text{and} \quad \Pr[s_{j,t} = 1|s_{j,t-1} = 1] = q_j, \qquad j = 1, 2. \tag{4.5}$$

Secondly, the other extreme would be to assume that the variables $y_{1,t}$ and $y_{2,t}$ share a common business cycle, which is obtained by imposing

$$s_{2,t} = s_{1,t} \qquad \text{for all } t. \tag{4.6}$$

As a consequence, a single underlying Markov process with transition probabilities p and q can be used to model the business cycle. We refer to Krolzig (1997), Paap and Van Dijk (2003), and Chauvet and Hamilton (2006) for extensive treatments of this model specification.

Thirdly, a more subtle approach, as proposed in Hamilton and Perez-Quiros (1996), is to assume that, although the coincident and leading indicators share the same business cycle, the cycle of the leading indicators leads or lags the cycle of the coincident variable by κ periods, that is

$$s_{2,t} = s_{1,t+\kappa} \qquad \text{for } \kappa \in \mathbb{Z}. \tag{4.7}$$

Note that positive values of κ correspond to the situation that the cycle of $y_{2,t}$ leads the cycle of $y_{1,t}$ by κ periods, whereas negative values correspond to a lag of $|\kappa|$ periods. We may treat the lead time κ as an unknown parameter to be estimated.

As discussed in the introduction, stylized facts show that on average leading indicators have a longer lead time when entering a recession than when entering an expansion. To capture this phenomenon, we consider a new specification of the state parameters accompanying the MS-VAR model (4.3) such that $s_{2,t}$ leads $s_{1,t}$ by κ_1 periods at peaks and by κ_2 periods at troughs. This may be formalized by defining $s_{2,t}$ as

$$s_{2,t} = \begin{cases} \prod_{i=\kappa_1}^{\kappa_2} s_{1,t+i} & \text{if } \kappa_1 \leq \kappa_2 \\ 1 - \prod_{i=\kappa_2}^{\kappa_1} (1 - s_{1,t+i}) & \text{if } \kappa_1 > \kappa_2. \end{cases} \tag{4.8}$$

To understand that this specification indeed gives rise to the desired asymmetric lead times, consider the case where $\kappa_1 \leq \kappa_2$. Defining $s_{2,t}$ as the product from $s_{1,t+\kappa_1}$ to

$s_{1,t+\kappa_2}$, essentially implies that recessions in $y_{2,t}$ start κ_1 periods before recessions in $y_{1,t}$, while they end κ_2 periods earlier. Note that, consequently, recessions in $y_{2,t}$ are $|\kappa_2 - \kappa_1|$ periods shorter than recessions in $y_{1,t}$. On the other hand, if $\kappa_1 > \kappa_2$, recessions in $y_{2,t}$ are $|\kappa_1 - \kappa_2|$ periods longer than recessions in $y_{1,t}$. Obviously, lengthening the recessions is equivalent to shortening the expansions. For that reason we define $s_{2,t}$ in (4.8) in terms of the product over $(1 - s_{1,t})$ in this case.

Note that the specification of $s_{2,t}$ in (4.8) embeds the specifications with a synchronous common cycle ($\kappa_1 = \kappa_2 = 0$), and with a non-synchronous common cycle with symmetric lead/lag times at peaks and at troughs ($\kappa_1 = \kappa_2$) as special cases. This facilitates testing for the degree of interrelation between the two cycles. The four specifications of the state vectors $s_{j,t}$, for $j = 1, 2$, discussed above are illustrated in Table 4.1.

The bivariate MS-VAR(1) model in (4.3) may be extended in several directions to make it more realistic and useful in empirical practice. Firstly, we may want to consider multiple coincident indicator variables, based on the original idea of Burns and Mitchell (1946, p.3) that the business "cycle consist of expansions (and recessions) occurring at about the same time in many economic activities." One may also use the information content in the recession indicator of the National Bureau of Economic Research (NBER) as in Issler and Vahid (2006), although the relatively large publication lag in this indicator makes it less useful for forecasting. Similarly, it may be beneficial to include multiple leading indicator variables, as different recessions have different sources and characteristics and thus may be signaled by different leading indicators, see Stock and Watson (2003), among others. This may be accommodated by taking $y_{j,t}$ to be a $(m_j \times 1)$-vector, for $j = 1, 2$, such that the model includes m_1 coincident indicators and m_2 leading indicators. In case both $m_1 > 1$ and $m_2 > 1$, it may be cumbersome to clearly define the relationships between the states $s_{j,t}$, $j = 1, \ldots, m_1 + m_2$ directly as in (4.8), as now there are $m_1 \times m_2$ different lead/lag times to consider. A possible solution then is to employ a dynamic factor structure as in Chauvet (1998), where all coincident and leading indicators are related with a certain lead/lag time to a latent common factor that exhibits regime-switching behavior.

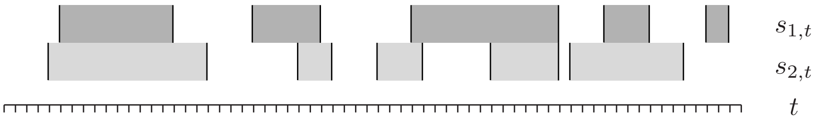
Secondly, the model may be extended to incorporate higher-order dynamics in the coincident and leading indicators. For any lag order $k \geq 0$, the general MS-VAR model reads

$$\begin{aligned}(\boldsymbol{Y}_t - \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t}) &= \boldsymbol{\Phi}_1(\boldsymbol{Y}_{t-1} - \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_{t-1}}) + \cdots + \boldsymbol{\Phi}_k(\boldsymbol{Y}_{t-k} - \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_{t-k}}) + \boldsymbol{\mathcal{E}}_t \qquad (4.9)\\ &= \sum_{i=1}^{k} \boldsymbol{\Phi}_i(\boldsymbol{Y}_{t-i} - \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_{t-i}}) + \boldsymbol{\mathcal{E}}_t,\end{aligned}$$

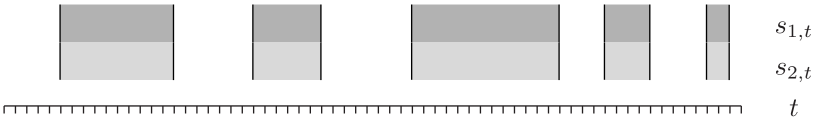
Table 4.1: Possible cycle interrelationships

Specification	Graphical example

(a) Independent cycles

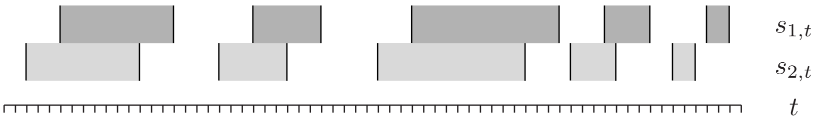


(b) Synchronous common cycle



(c) Non-synchronous common cycle with lead time κ

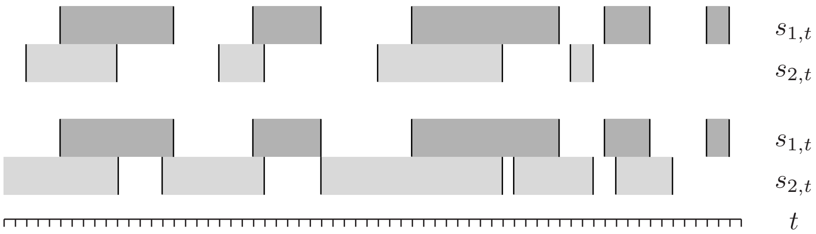
$\kappa = 3$



(d) Non-synchronous common cycle with asymmetric lead times κ_1 for peaks and κ_2 for troughs

$\kappa_1 = 3$
$\kappa_2 = 5$

$\kappa_1 = 8$
$\kappa_2 = 5$



Note: The table shows four possible types of specifications for the processes $s_{1,t}$ and $s_{2,t}$ in the bivariate Markov Switching model (4.3), with different types of relationship between the cycles in $y_{1,t}$ and $y_{2,t}$: (a) Independent cycles as implied by (4.5), (b) A synchronous common cycle as in (4.6), (c) A non-synchronous common cycle with identical lead/lag time κ at peaks and at troughs as in (4.7), and (d) A non-synchronous common cycle with different lead/lag times κ_1 at peaks and κ_2 at troughs as in (4.8). The dark and light grey shaded areas correspond to recession periods in $y_{1,t}$ and $y_{2,t}$, respectively.

or, using lag polynomial notation

$$\Phi(\boldsymbol{L})\boldsymbol{Z}_t = (\boldsymbol{I} - \boldsymbol{\Phi}_1\boldsymbol{L} - \cdots - \boldsymbol{\Phi}_k\boldsymbol{L}^k)\boldsymbol{Z}_t = \boldsymbol{\mathcal{E}}_t, \tag{4.10}$$

where $\boldsymbol{Z}_t = \boldsymbol{Y}_t - \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t}$.

A third possible extension of the model concerns the possibility of multiple regimes. Several applications of Markov switching models to US GDP, for example, have found that allowing for a third regime to capture the so-called 'bounce-back effect', that is, a short period of rapid recovery following recessions, considerably improves the description of the cyclical dynamics of output, see Sichel (1994), Boldin (1996), and Clements and Krolzig (2003), among others. In the case of multiple regimes, specifying the lead/lag structure of the regime-switches for the different variables in the model may be complicated and has to be done with care.

Fourthly, a model specification in which the transition probabilities of the Markov processes $s_{j,t}$, for $j = 1, 2$, depend on observed explanatory variables may be considered, see Diebold *et al.* (1994), Filardo (1994), and Diebold and Rudebusch (1996), among others.

Fifthly and finally, regime-dependent heteroskedasticity and correlations among the shocks $\boldsymbol{\mathcal{E}}_t$ may be captured by replacing the assumption $\boldsymbol{\mathcal{E}}_t \sim N(0, \boldsymbol{\Sigma})$ in (4.3) by $\boldsymbol{\mathcal{E}}_t|\boldsymbol{\mathcal{S}}_t \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}_t})$. In the empirical application below we consider another type of dynamics in the error (co-)variances to accommodate the effects of the 'Great Moderation', that is the large and persistent decline in volatility of US macro-economic time series since the mid-1980s, see McConnell and Perez-Quiros (2000), Sensier and Van Dijk (2004), and Herrera and Pesavento (2005), among others. Specifically, we allow for a single structural break in the covariance matrix of $\boldsymbol{\mathcal{E}}_t$

$$\begin{aligned}\boldsymbol{\Sigma}_t &= \begin{cases}\boldsymbol{\Omega}_0 & \text{if } t < \tau \\ \boldsymbol{\Omega}_1 & \text{if } t \geq \tau\end{cases} \\ &= \boldsymbol{\Omega}_0\mathbb{I}[t < \tau] + \boldsymbol{\Omega}_1\mathbb{I}[t \geq \tau],\end{aligned} \tag{4.11}$$

where $\mathbb{I}[\cdot]$ denotes the indicator function, taking the value one if the argument is true and zero otherwise, and $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ are (2×2) covariance matrices. We treat the break point τ as an unknown parameter to be estimated.

4.3 Estimation and Inference

Parameter estimation and inference on the regimes in MS-(V)AR models is commonly done using maximum likelihood coupled with the EM-algorithm, see Hamilton (1989,

1994) for details. However, as we want to conduct inference on the discrete lead/lag time parameters κ_1 and κ_2 in (4.8), a frequentist approach is not feasible. We therefore adopt a Bayesian approach. In Section 4.3.1 we derive the likelihood function of the model. Sections 4.3.2 and 4.3.3 discuss prior specification and posterior simulation.

4.3.1 The likelihood function

We first derive the complete data likelihood function. We focus on the derivation for the bivariate MS-VAR model (4.9) with asymmetric lead/lag structure as given in (4.8). The likelihood of the other specifications can be derived in a similar way.

Following Hamilton (1989) and Paap and Van Dijk (2003) we replace $\boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t}$ by

$$\boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t} = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_t, \tag{4.12}$$

where $\odot$ denotes the Hadamard or element-by-element product and where $\boldsymbol{\mathcal{S}}_t$ is given in (4.4) with (4.8). Hence, $\boldsymbol{\Gamma}_0 = (\mu_{1,0}, \mu_{2,0})'$ and $\boldsymbol{\Gamma}_1 = (\mu_{1,1} - \mu_{1,0}, \mu_{2,1} - \mu_{2,0})'$. Model (4.9) then reads

$$(\boldsymbol{Y}_t - \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_t) = \sum_{i=1}^{k} \boldsymbol{\Phi}_i(\boldsymbol{Y}_{t-i} - \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_{t-i}) + \boldsymbol{\mathcal{E}}_t, \qquad \boldsymbol{\mathcal{E}}_t \sim N(0, \boldsymbol{\Sigma}_t) \tag{4.13}$$

for $t = k+1, \ldots, T$, where T denotes the sample size and $\boldsymbol{\Sigma}_t$ is specified in (4.11).

The conditional density of $\boldsymbol{Y}_t$ for this model given the past and current states $\boldsymbol{\mathcal{S}}^t = \{\boldsymbol{\mathcal{S}}_1, \ldots, \boldsymbol{\mathcal{S}}_t\}$ and given the past observations $\boldsymbol{Y}^{t-1} = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}\}$ is given by

$$f(\boldsymbol{Y}_t|\boldsymbol{Y}^{t-1}, \boldsymbol{\mathcal{S}}^t, \boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \boldsymbol{\Phi}, \kappa_1, \kappa_2, \tau) = \frac{1}{(\sqrt{2\pi})^2}|\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\boldsymbol{\mathcal{E}}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mathcal{E}}_t\Big), \tag{4.14}$$

where $\boldsymbol{\mathcal{E}}_t$ follows from (4.13). Hence the complete data likelihood function for model (4.13) conditional on the first k observations $\boldsymbol{Y}^k$ equals

$$\begin{aligned}\mathcal{L}(\boldsymbol{Y}^T, \boldsymbol{\mathcal{S}}^T|\boldsymbol{Y}^k, \boldsymbol{\theta}) = {} & p^{\boldsymbol{\mathcal{N}}_{0,0}}\,(1-p)^{\boldsymbol{\mathcal{N}}_{0,1}}\,q^{\boldsymbol{\mathcal{N}}_{1,1}}\,(1-q)^{\boldsymbol{\mathcal{N}}_{1,0}} \\ & \times \prod_{t=k+1}^{T} f(\boldsymbol{Y}_t|\boldsymbol{Y}^{t-1}, \boldsymbol{\mathcal{S}}^t, \boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \boldsymbol{\Phi}, \kappa_1, \kappa_2, \tau),\end{aligned} \tag{4.15}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \boldsymbol{\Phi}, \kappa_1, \kappa_2, \tau, p, q\}$, and where $\boldsymbol{\mathcal{N}}_{s_1^*, s_2^*}$ denotes the number of transitions from state s_1^* to state s_2^*. The unconditional likelihood function $\mathcal{L}(\boldsymbol{Y}^T|\boldsymbol{Y}^k, \boldsymbol{\theta})$ can be obtained by summing over all possible realizations of $\boldsymbol{s}_1$

$$\mathcal{L}(\boldsymbol{Y}^T|\boldsymbol{Y}^k, \boldsymbol{\theta}) = \sum_{s_{1,1}=0}^{1} \sum_{s_{1,2}=0}^{1} \cdots \sum_{s_{1,T}=0}^{1} \mathcal{L}(\boldsymbol{Y}^T, \boldsymbol{\mathcal{S}}^T|\boldsymbol{Y}^k, \boldsymbol{\theta}). \tag{4.16}$$

The complete data likelihood in case $\kappa = \kappa_1 = \kappa_2$ follows directly from (4.15). In case we have separate cycles for the two series in $\boldsymbol{Y}_t$ we have to extend (4.15) with the likelihood contribution of the second cycle in a straightforward manner.

4.3.2 Prior specification

We opt for a prior specification that is relatively uninformative compared to the information in the likelihood. For the transition probabilities p and q, we take independent and uniformly distributed priors on the unit interval $(0, 1)$

$$f(p) = \mathbb{I}[0 < p < 1] \qquad \text{and} \qquad f(q) = \mathbb{I}[0 < q < 1]. \tag{4.17}$$

Under flat priors for p and q special attention must be paid to the priors for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$. It is easy to show that the likelihood has the same value if we switch the role of the states, switch κ_1 with κ_2, and change the values of $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_1$, p and q into $\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1$, $-\boldsymbol{\Gamma}_1$, q and p respectively, see Frühwirth-Schnatter (2001). This complicates proper posterior analysis if we are interested in the values of $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$, see also Geweke (2007) for a discussion. Following Paap and Van Dijk (2003) we take priors for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$ on subspaces which identify the regimes, that is,

$$\begin{aligned}
f(\boldsymbol{\Gamma}_0) &\propto \begin{cases} 1 & \text{if } \boldsymbol{\Gamma}_0 \in \{\boldsymbol{\Gamma}_0 \in \mathbb{R}^2 | \boldsymbol{\Gamma}_{0,1} > 0\} \\ 0 & \text{elsewhere,} \end{cases} \\
f(\boldsymbol{\Gamma}_1 | \boldsymbol{\Gamma}_0) &\propto \begin{cases} 1 & \text{if } \boldsymbol{\Gamma}_1 \in \{\boldsymbol{\Gamma}_1 \in \mathbb{R}^2 | \boldsymbol{\Gamma}_{0,1} + \boldsymbol{\Gamma}_{1,1} \leq 0\} \\ 0 & \text{elsewhere.} \end{cases}
\end{aligned} \tag{4.18}$$

Hence, we impose that the growth rate μ for the first series is positive if $s_{1,t} = 0$ and negative if $s_{1,t} = 1$, see also Smith and Summers (2005) for a discussion on this issue. For the model specification with two independent cycles, or, put differently, two independent Markov processes $s_{1,t}$ and $s_{2,t}$, we take the priors given in (4.17) for both sets of transition probabilities. In that case, the prior for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$ as given in (4.18) is augmented with the additional restrictions $\boldsymbol{\Gamma}_{0,2} > 0$ and $\boldsymbol{\Gamma}_{0,2} + \boldsymbol{\Gamma}_{1,2} < 0$ for identification of the regimes of $y_{2,t}$. Note that Smith and Summers (2005) are less restrictive and only impose that the diagonal elements of $\boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_0$ are smaller or equal to 0 and have no further restrictions on the diagonal elements of $\boldsymbol{\Gamma}_0$.

For the shift parameters κ_j we take a discrete uniform prior

$$f(\kappa_j) = \begin{cases} \frac{1}{2c_j+1} & \text{if } \kappa_j \in \{-c_j, \ldots, 0, \ldots, c_j\} \\ 0 & \text{elsewhere} \end{cases} \tag{4.19}$$

for $j = 1, 2$. Hence, we allow for a maximum lead/lag time of c_j periods. The same prior is used for the lead time κ in the model specification with a non-synchronous common cycle but equal lead times at peaks and at troughs based on (4.7).

For the autoregressive parameters we use flat priors

$$f(\boldsymbol{\Phi}_i) \propto 1 \qquad \text{for } i = 1, \ldots, k-1, \tag{4.20}$$

and for $\boldsymbol{\Omega}_j$ we take the uninformative prior

$$f(\boldsymbol{\Omega}_j) \propto |\boldsymbol{\Omega}_j|^{-3/2} \tag{4.21}$$

for $j = 0, 1$. This prior results from a standard Wishart prior by letting the degrees of freedom approach zero, see Geisser (1965).

Finally, for the break parameter τ we take a discrete uniform prior

$$f(\tau) = \begin{cases} \frac{1}{T-k-2b} & \text{if } \tau \in \{k+b+1, \ldots, T-b\} \\ 0 & \text{elsewhere,} \end{cases} \tag{4.22}$$

hence not allowing for a break in the first and last b observations of the sample period.

The joint prior for the model parameters $f(\boldsymbol{\theta})$ is given by the product of (4.17)–(4.22).

4.3.3 Posterior distributions

The posterior distribution for the model parameters of the Markov switching vector autoregressive model is proportional to the product of the joint prior $f(\boldsymbol{\theta})$ and the unconditional likelihood function $\mathcal{L}(\boldsymbol{Y}^T|\boldsymbol{Y}^k, \boldsymbol{\theta})$. To obtain posterior results we use the Gibbs sampling algorithm of Geman and Geman (1984) together with the data augmentation method of Tanner and Wong (1987). The unobserved state variables $\{\boldsymbol{\mathcal{S}}_t\}_{t=1}^T$ are simulated alongside the model parameters $\boldsymbol{\theta}$, see Albert and Chib (1993), McCulloch and Tsay (1994), Chib (1996), and Kim and Nelson (1999), among others.

The Gibbs sampler is an iterative algorithm, where one consecutively samples from the full conditional posterior distributions of the model parameters. This produces a Markov chain, which converges under mild conditions. The resulting draws can be considered as a sample from the posterior distribution, see De Pooter *et al.* (2008) for a lucid survey and Smith and Roberts (1993) and Tierney (1994) for technical details. In Appendix 4.A we derive the full conditional posterior distributions resulting from $f(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{Y}^T, \boldsymbol{\mathcal{S}}^T|\boldsymbol{Y}^k, \boldsymbol{\theta})$ for the model specification with asymmetric lead/lag structure.

4.4 Empirical Results

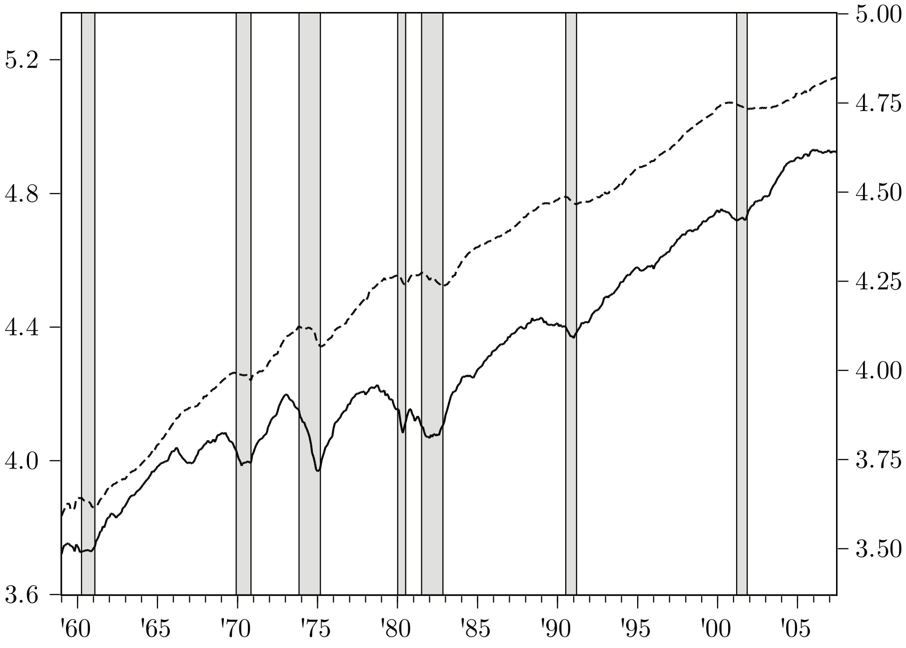
We apply the Markov switching VAR models proposed in Section 4.2 to examine the lead times at business cycle peaks and troughs of the Composite Leading Index as issued by The Conference Board. As a measure of economic activity we consider The Conference Board's Composite Coincident Index, which is comprised of employment, industrial production, manufacturing and trade sales, and personal income less transfer payments. Both time series are transformed to monthly growth rates. The sample period runs from January 1959 - June 2007. The estimation results reported in this section use the revised data as available in July 2007. The next section considers out-of-sample forecasting of turning points and CCI growth rates based on real-time data.

Figure 4.1 displays a time series plot of the log levels and monthly percentage growth rates of both series, together with the recession periods as determined by the NBER. In general, the leading index seems to have a similar cyclical pattern as the coincident index, but with turning points clearly occurring earlier. In addition, the visual evidence in Figure 4.1 already suggests that the CLI turning points have a longer lead time for business cycle peaks than for troughs. We apply the four different specifications of the Markov switching model discussed in Section 4.2 to investigate formally whether this indeed is the case, and to examine by how many periods the leading indicator actually is leading at peaks and at troughs. In addition, for comparison we include a linear vector autoregressive model, which can be obtained from (4.9) by setting $\boldsymbol{\mathcal{M}}_{\mathcal{S}_t} = \boldsymbol{\mathcal{M}}$ for all t.

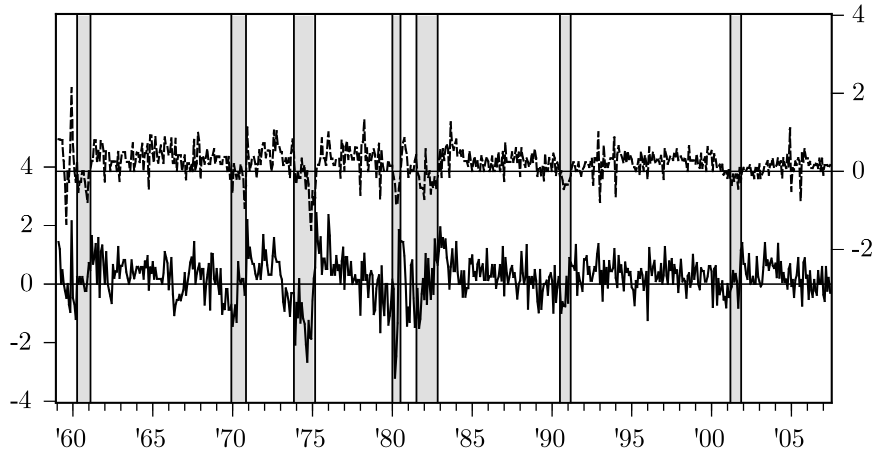
All five models are estimated without and with a single structural break in volatility as in (4.11). We find that the marginal likelihoods of the models with a structural break are clearly smaller than for the models without, and hence there is strong posterior evidence for a structural break in volatility. In other respects, the models with and without volatility break are quite similar. For those reasons and to save space we only report results for the models that incorporate a volatility break. Detailed results for the models that do not allow for a structural break in volatility are available upon request.

To perform inference we use the Bayesian approach as discussed in Section 4.3 with the prior specifications given in Section 4.3.2. We set the parameters c_1 and c_2 in the priors for the lead/lag times κ_1 and κ_2 equal to 18, which implies that we allow for a maximum non-synchronicity of one and a half year in the cycles of both series. The parameter b in the prior for the break date is set equal to 6, so that we do not allow for a break in the (co-)variances in the first and last six observations. We consider several specifications for the autoregressive dynamics in (4.9). Unreported Bayes factors based on moderately informative priors on $\boldsymbol{\Phi}$ indicate that a lag order $k = 1$ with additional

Figure 4.1: Time series plots of The Conference Board's Composite Coincident Index and the Composite Leading Index



(a) Log levels



(b) Percentage growth rates

Note: The graph presents the logarithmic levels in Panel (a) and the month-to-month growth rates in Panel (b) of The Conference Board's Composite Coincident Index (dashed lines, left axes) and the Composite Leading Index (solid lines, right axes) for the period January 1959 - June 2007.

restrictions $\phi_{1,1} = \phi_{2,1} = 0$ is most appropriate, see Hamilton and Perez-Quiros (1996) for a similar specification. Hence, only lagged CLI growth rates enter the equations for both CCI and CLI.

Posterior results of the five estimated models are shown in Table 4.2, based on 100,000 simulations following a burn-in period of 10,000 simulations. The first panel of the table shows that in the linear VAR model the posterior mean of the average monthly growth rate of CLI and CCI is roughly equal over the sample period, at 0.21% and 0.20%, respectively. The posterior standard deviation of the growth rate of the leading index is, however, substantially higher. In the MS-VAR models we observe clear differences in the average growth rates during recession and expansion periods. Depending on the model specification, the posterior means of the average growth rates during expansions are between 0.26% and 0.30% for the CCI series. For recessions the posterior means are between –0.26% and –0.12%. The posterior mean of the probability of staying in an expansion regime is about 0.97, while the probability of staying in a recession regime is considerably lower at about 0.87 on average. This obviously reflects the fact that recessions typically last shorter than recessions. Based on these average transition probability estimates, the expected duration of a recession is between seven and eight months, compared to 33 months for expansions. Note that there is much more variability in the posterior means of the probability of staying in a recession than for the probability of staying in an expansion across the different model specifications. In particular, the probability of staying in recession varies between 0.83 for the asymmetric non-synchronous cycle specification and 0.90 for the specification with independent cycles.

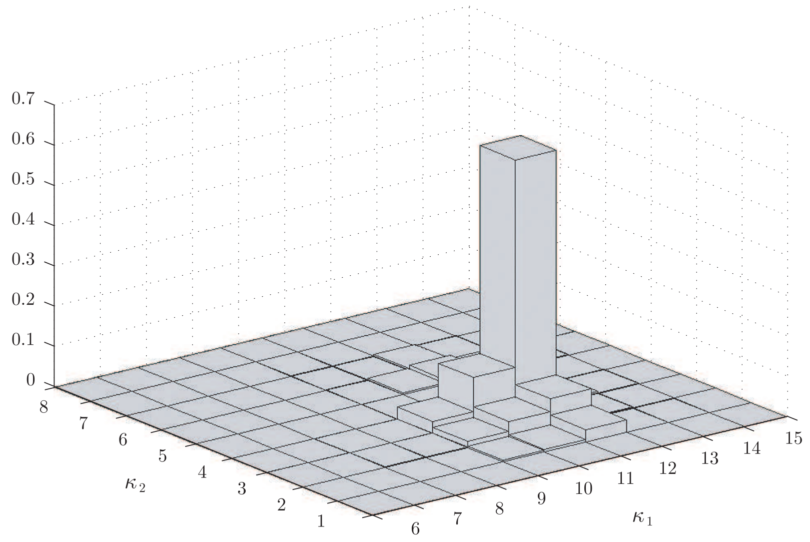
The fourth panel of Table 4.2 shows that in the model with a non-synchronous common cycle (4.7) the posterior mean of the lead time of the CLI is around nine months. This is considerably longer than the lead time of one quarter as found in Hamilton and Perez-Quiros (1996). This large discrepancy can be attributed to substantial differences in methodology. In particular, Hamilton and Perez-Quiros (1996) obtain their results for quarterly GDP and CLI data on a shorter sample period using frequentist inference methods. In the novel model specification allowing for asymmetric lead/lag times κ_1 and κ_2, the posterior mean of the lead time at peaks is about 11.7 months compared to about 3.8 months at troughs. This confirms the informal visual evidence in Figure 4.1 that the CLI signals oncoming recessions earlier than expansions. Figure 4.2 displays the posterior distribution of the κ_j parameters, for $j = 1, 2$ in this model specification, showing that the posterior mode is $\kappa_1 = 12$ and $\kappa_2 = 4$. The posterior probability that $\kappa_1 = \kappa_2$ is only 0.02, providing complementary evidence that the lead times at the start of recessions and expansions really are different.

Table 4.2: Posterior means and standard deviations (in parentheses) of parameters in linear and MS-VAR models with volatility break

		Linear VAR		MS-VAR							
				(a) Independent cycles		(b) Synchronous common cycle		(c) Non-synchronous common cycle		(d) Asymmetric Non-synchronous common cycle	
Growth rates	$\mu_{1,0}$	0.197	(0.016)	0.304	(0.019)	0.274	(0.018)	0.281	(0.018)	0.262	(0.014)
	$\mu_{1,1}$			−0.118	(0.042)	−0.208	(0.055)	−0.138	(0.063)	−0.257	(0.043)
	$\mu_{2,0}$	0.214	(0.040)	0.334	(0.037)	0.216	(0.049)	0.332	(0.051)	0.373	(0.036)
	$\mu_{2,1}$			−0.839	(0.118)	−0.039	(0.044)	−0.295	(0.129)	−0.296	(0.061)
Transition Probabilities	p_1			0.972	(0.011)	0.966	(0.011)	0.964	(0.012)	0.973	(0.008)
	q_1			0.904	(0.038)	0.841	(0.053)	0.870	(0.044)	0.833	(0.046)
	p_2			0.974	(0.011)						
	q_2			0.797	(0.080)						
Dynamics	$\phi_{1,2}$	0.105	(0.022)	0.121	(0.026)	0.096	(0.024)	0.068	(0.026)	0.047	(0.024)
	$\phi_{2,2}$	0.376	(0.041)	0.297	(0.053)	0.394	(0.041)	0.279	(0.058)	0.217	(0.043)
Lead/lag times	κ_1							9.228	(2.712)	11.710	(0.637)
	κ_2									3.794	(0.553)
Covariance matrix	$\omega^1_{1,1}$	0.180	(0.015)	0.935	(0.576)	0.132	(0.012)	0.140	(0.014)	0.136	(0.012)
	$\omega^1_{2,2}$	0.548	(0.046)	1.172	(0.791)	0.543	(0.046)	0.482	(0.050)	0.464	(0.040)
	ρ^1	0.413	(0.049)	0.711	(0.182)	0.478	(0.052)	0.520	(0.051)	0.547	(0.046)
	$\omega^2_{1,1}$	0.069	(0.006)	0.087	(0.006)	0.051	(0.005)	0.055	(0.006)	0.060	(0.006)
	$\omega^2_{2,2}$	0.255	(0.022)	0.310	(0.022)	0.261	(0.024)	0.229	(0.024)	0.220	(0.020)
	ρ^2	0.407	(0.051)	0.535	(0.037)	0.460	(0.057)	0.423	(0.060)	0.403	(0.054)
Most likely break date	τ	1984:02		1984:02		1984:02		1984:02		1984:02	
Log marg. likelihood		–671.6		–656.6		–656.7		–635.8		-622.8	

Note: The table presents posterior means and standard deviations (in parentheses) of parameters in linear and MS-VAR models with a single structural change in the covariance matrix as in (4.11), estimated for monthly growth rates of The Conference Board's CCI and CLI over the period January 1959 - June 2007. The parameters ρ^1 and ρ^2 denote the correlation between the CCI and CLI growth rates before and after break, respectively. The most likely break date is defined as the mode of the posterior distribution of τ. The four specifications for the processes $s_{1,t}$ and $s_{2,t}$ in the bivariate MS-VAR model are such that CCI and the CLI have (a) Independent cycles as implied by (4.5), (b) A synchronous common cycle as in (4.6), (c) A non-synchronous common cycle with identical lead/lag time κ at peaks and at troughs as in (4.7), and (d) A non-synchronous common cycle with different lead/lag times κ_1 at peaks and κ_2 at troughs as in (4.8). Posterior results are based on 100,000 simulations. Number of burn-in simulations is 10,000.

Figure 4.2: Joint posterior density of the lead/lag parameters κ_1 and κ_2

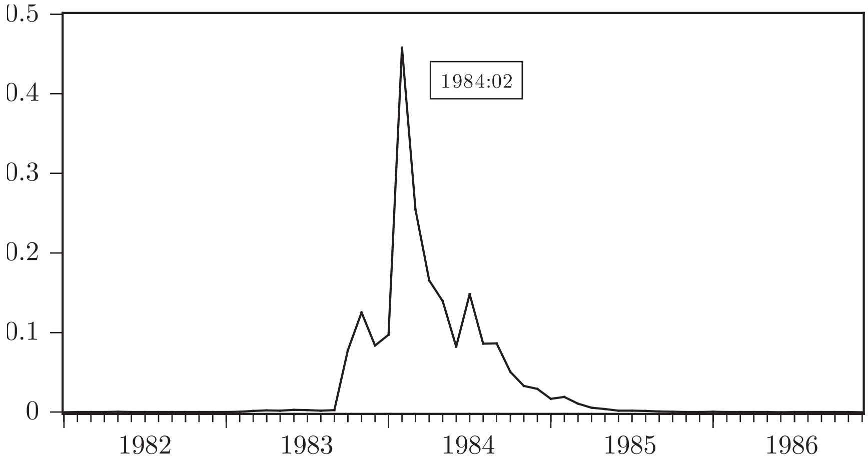


Note: The graph presents the joint posterior density of the lead/lag parameters κ_1 and κ_2 in the MS-VAR model with a non-synchronous common cycle with asymmetric lead/lag times κ_1 at peaks and κ_2 at troughs as in (4.8) and a single structural change in the covariance matrix as in (4.11), estimated for monthly growth rates of The Conference Board's CCI and CLI over the period January 1959 - June 2007.

The bottom panel of the table shows that the posterior mode of the break point parameter τ is 1984:02 in all model specifications, which corresponds with the break point estimate for GDP volatility as reported by McConnell and Perez-Quiros (2000), among others. Figure 4.3 displays the posterior density of the break parameter for the model with an asymmetric lead/lag structure for the period January 1982 - December 1986. Almost all posterior mass is located in the years 1983 and 1984. In the MS-VAR model with asymmetric lead/lag times, the posterior means of the variances are such that for CCI the variance after the break is about 43% of the variance before. For the leading indicator the reduction in variance also is very large at 47%. The posterior means of the correlation between the CCI and CLI growth rates are 0.55 and 0.40 before and after the break, respectively. This suggests that the strength of the co-movement between the series also declined.

The bottom panel of the table also shows the log marginal likelihoods of the five models. As we have proper priors on the transitions probabilities we can compare the log marginal likelihoods of the four Markov switching models to assess the appropriateness of the different cycle specifications. The log marginal likelihood of the model with asymmetric lead/lag structure is clearly bigger than for the other models. The log Bayes

Figure 4.3: Posterior density of the variance breakpoint parameter τ

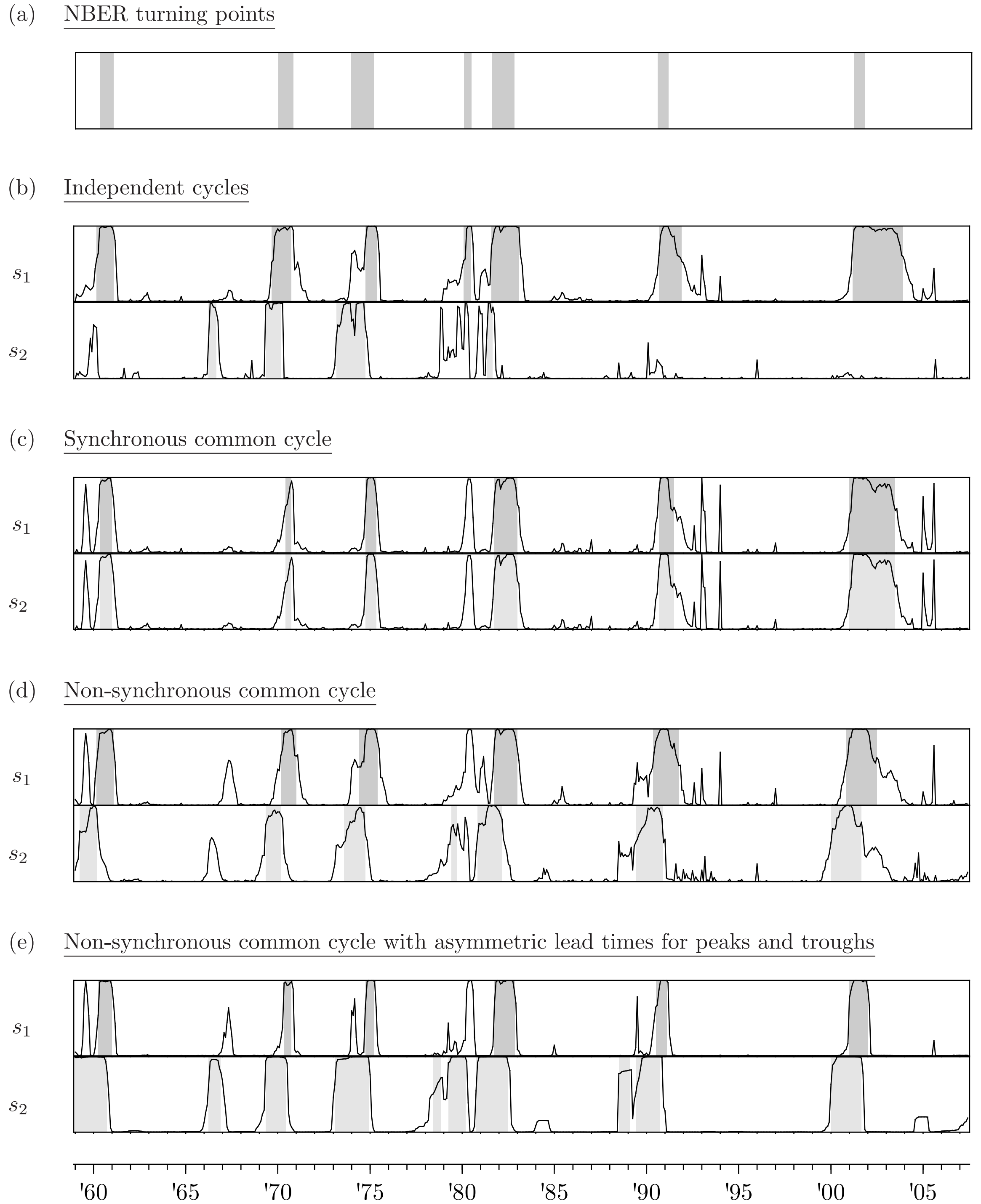


Note: The graph presents the posterior density of the variance break date τ (for the period January 1982 - December 1986) in the MS-VAR model with a non-synchronous common cycle with different lead/lag times κ_1 at peaks and κ_2 at troughs as in (4.8) and a single structural change in the covariance matrix as in (4.11), estimated for monthly growth rates of The Conference Board's CCI and CLI over the period January 1959 - June 2007.

factor compared with the non-synchronous common cycle specification is equal to 13 and, hence, there is strong posterior evidence for the more general specification with different lead/lag times at peaks and at troughs.

We proceed with judging the different model specifications on their ability to signal turning points and to identify recession periods. Figure 4.4 shows the posterior means of the state variables $s_{j,t}$, $j = 1, 2$ for the four Markov switching models. The shaded areas indicate 'recession periods' defined as six consecutive months where the posterior mean of $s_{j,t}$ is larger than 0.5. This corresponds with the popular rule of thumb saying that the economy is in recession whenever economic growth is negative during two consecutive quarters. We emphasize that we do not propose this censoring rule as a formal means to identify business cycle regimes. For that purpose, it is better to consider the posterior probabilities of the state variable $s_{1,t}$ for the coincident indicator directly. We compare the posterior means of $s_{1,t}$ with the business cycle expansions and recessions as implied by the NBER turning points, thus assuming that the latter are correct. It is useful to note that the CCI consists of the same four monthly series closely monitored by the NBER business cycle dating committee. At the same time, of course the NBER also considers other indicators of economic activity such as real GDP.

Figure 4.4: Posterior recession probabilities in MS-VAR models with volatility break



Note: The graphs present posterior recession probabilities in MS-VAR models with a single structural change in the covariance matrix as in (4.11), estimated for monthly growth rates of The Conference Board's CCI and CLI over the period January 1959 - June 2007, with different types of relationships between their cycles. See Table 4.2 for definitions of the specifications for the processes $s_{1,t}$ and $s_{2,t}$. The shaded areas in Panel (a) are recession periods based upon the NBER turning points. The dark and light grey shaded areas in Panels (b) - (e) correspond to periods of (at least) six consecutive months where the posterior mean of $s_{j,t}$ is larger than 0.5.

The bottom graph in Figure 4.4 reveals that the model with an asymmetric lead/lag time at peaks and at troughs as in (4.8) does an excellent job identifying the regimes of the business cycle. For all the official NBER-dated recessions that occurred during the sample period, the posterior mean of $s_{1,t}$ is very close to one. In addition, no false recession signals are given except around 1967, which corresponds with a growth rate cycle recession, according to the Economic Cycle Research Institute.[2] A comparison with the graphs in Panels (b) - (d) demonstrates the advantage of allowing for different lead times at peaks and at troughs in two different ways. Firstly, the posterior probabilities obtained from the asymmetric lead time specification allow for much sharper inference concerning the business cycle regimes. Although the other specifications also identify all recessions that occurred during the sample period, their signal is more noisy in the sense that the posterior mean of $s_{1,t}$ often lingers at values between zero and one. This is especially noticeable for the two most recent recessions in 1990/1991 and 2001, where the posterior means of $s_{1,t}$ in Panels (b) - (d) return to zero much later than the official troughs in March 1991 and November 2001. By contrast, in Panel (e) we find a clear and timely signal that these recessions have ended as the posterior mean of $s_{1,t}$ drops to zero close to these trough dates. Secondly, the model with a non-synchronous common cycle with an asymmetric lead/lag structure is more timely, in the sense that it signals oncoming recessions (by showing an increase in the probability that $s_{2,t} = 1$) quite a bit earlier than the other specifications. This is most obvious for the model with a synchronous common cycle, which is almost by definition not able to identify a recession before it actually occurs. In the model with independent cycles in Panel (b), we do find a positive lead time of $s_{2,t}$ but only for the recessions in the first part of the sample period, before the break in volatility. The 1990/1991 and 2001 recessions are completely missed by the CLI in this specification. Comparing the two non-synchronous common cycle specifications in Panels (d) and (e), the difference in the lead time at peaks is not very large, as suggested already by the difference in posterior means of just 2.5 months, see Table 4.2. The signal provided by the state variable of the CLI ($s_{2,t}$) is, however, again much more convincing in the asymmetric lead time specification.

4.5 Real-time Business Cycle Dating and Forecasting

The full-sample estimation results discussed in the previous section demonstrate that using the CLI within a MS-VAR model delivers an accurate description of US business cycle dynamics *ex post*. The practical usefulness of leading indicator variables, however,

[2]See `http://www.businesscycle.com`.

crucially hinges upon their ability to signal changes in the business cycle *ex ante*. Furthermore, as both the CCI and CLI are subject to substantial revisions after their initial release, a realistic assessment of this issue requires the use of real-time data that was actually available when the forecasts were supposed to be made. Two related aspects of real-time performance are of interest. Firstly, we consider real-time business cycle *dating*, as in Chauvet and Piger (2008), and examine how quickly the different models provide a reliable signal that the business cycle regime has changed. Secondly, we consider genuine out-of-sample *forecasting* of both turning points and output growth in real time. A number of previous studies examining the real-time predictive ability of the CLI have rendered mixed results, depending on the choice of time series model as well as the coincident indicator variable(s), see Diebold and Rudebusch (1991), Hamilton and Perez-Quiros (1996), Camacho and Perez-Quiros (2002), Filardo (1999, 2004), and McGuckin *et al.* (2007), among others. Here we examine whether there is any added value of allowing for different lead times at peaks and at troughs for predictive accuracy.

Our real-time data set for the CCI and CLI consists of 223 releases, or vintages, released from January 1989 until July 2007. Each vintage contains a complete time series of monthly observations from January 1959 until one month prior to the release date. Except for McGuckin *et al.* (2007), previous studies examining the real-time predictive ability of the CLI use revised data as available at present for the coincident indicator or output measure, based on the idea that the revised data is closer to the truth that we (should) aim to forecast. However, this comes at the cost of making the real-time experiment less realistic as revisions in output measures are substantial, see Swanson and Van Dijk (2006) for a recent assessment. As we would like to approximate the actual possibilities of a business cycle analyst as closely as possible, we make use of real-time CCI data instead.

We construct real-time estimates and forecasts of the business cycle indicator $s_{1,t}$ and the monthly CCI growth rate $y_{1,t}$ for each vintage in the period January 1989 until July 2007, as follows. Using the data release of month T, which contains observations until month $T-1$, we first obtain the posterior distribution of the model parameters. Of particular interest is the posterior distribution of the state variable $f(s_{1,t}|\boldsymbol{Y}^{T-1})$ for $t = 1, 2, \ldots, T-1$, as this can be used for real-time dating of business cycle peaks and troughs. Next, we determine the predictive densities $f(s_{1,T-1+h}|\boldsymbol{Y}^{T-1})$ and $f(y_{1,T-1+h}|\boldsymbol{Y}^{T-1})$ for $h = 1, 2, \ldots$. Draws from these predictive densities can easily be obtained from the Gibbs output of the posterior distribution. Given a draw from the posterior of the parameters and states $\boldsymbol{\mathcal{S}}_1, \ldots, \boldsymbol{\mathcal{S}}_{T-1}$, we simply simulate future observations taking the model as data generating process. We use the means of the predictive distributions as point forecasts.

This implies that the forecast for $s_{1,T-1+h}$ is the predictive probability that $s_{1,T-1+h}$ equals 1 or, put differently, the probability that month $T-1+h$ is part of a recession. We consider forecast horizons up to $h = 12$ steps ahead, where it should be noted that in fact the one-step ahead prediction is a nowcast as it is made at the end of month T. Finally, we remark that the out-of-sample forecasting exercise reported in this section is conducted using the models with a volatility break as in (4.11). The implicit assumption is that at the start of the forecasting period, that is January 1989, the business cycle analyst is aware of the volatility break and incorporates this into the model.

4.5.1 Business cycle dating results

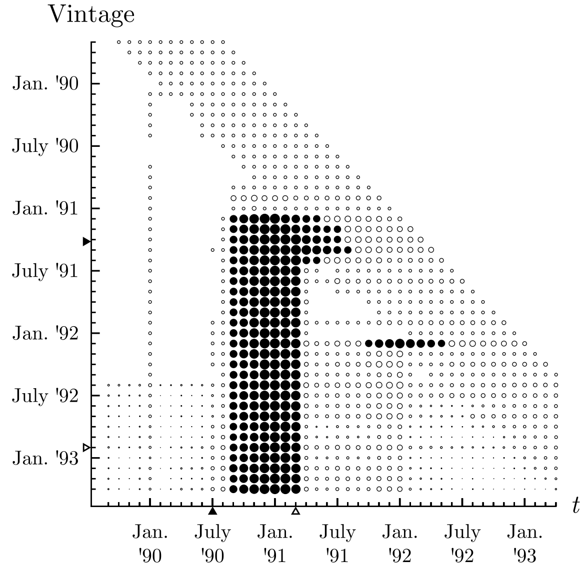
For a business cycle dating procedure to be useful in real-time, it should strike a balance between the speed at which regime shifts are detected and the accuracy of the estimated turning point dates. The posterior distribution $f(s_{1,t}|\boldsymbol{Y}^{T-1})$ constructed at the end of month T delivers posterior probability estimates $\Pr[s_{1,t} = 1|\boldsymbol{Y}^{T-1}]$ for $t = 1, 2, \ldots, T-1$. To convert these recession probabilities into turning point estimates, we may again use a specific dating rule as we did in Section 4.4, where we considered the rule of thumb defining a recession as a period of six consecutive months where the recession probability exceeds 0.5. Some business cycle analysts, however, may be more inclined to accept also weaker signals, if the speed of detection is of utmost importance. Others may prefer to wait longer, in order to gain accuracy and certainty about the dates obtained. For this reason, instead, we visualize the posterior probabilities $\Pr[s_{1,t} = 1|\boldsymbol{Y}^{T-1}]$ and leave the exact dating rule to the reader.

Figures 4.5 and 4.6 display the real-time posterior probabilities $\Pr[s_{1,t} = 1|\boldsymbol{Y}^{T-1}]$ obtained using the vintages before, during and after the recessions in 1990/1991 and 2001, respectively. On the horizontal axis of the graphs we display time t and on the vertical axis the release date of the vintage $T-1$. Hence, each row in these graphs displays the values of the posterior probabilities of a recession over time based on the vintage released at the end of the month as indicated on the vertical axis. If the value of the recession probability $\Pr[s_{1,t} = 1|\boldsymbol{Y}^{T-1}]$ is 0.5 or above we report a black dot. If the probability is below 0.5 we report a white dot. Moreover, the dots increase in size with the value of the recession probability. If for one particular vintage the dots change from white to black at a certain calender month and remain black consistently thereafter, we gain confidence that this month should be marked as a business cycle peak. A change from black to white similarly indicates a trough. Conversely, looking across rows reveals how this assessment changes across data releases. If the posterior probabilities based on some vintage show a particular recession for the first time and this recession persists in the results thereafter, we mark the release date of the first vintage as the detection date of the recession.

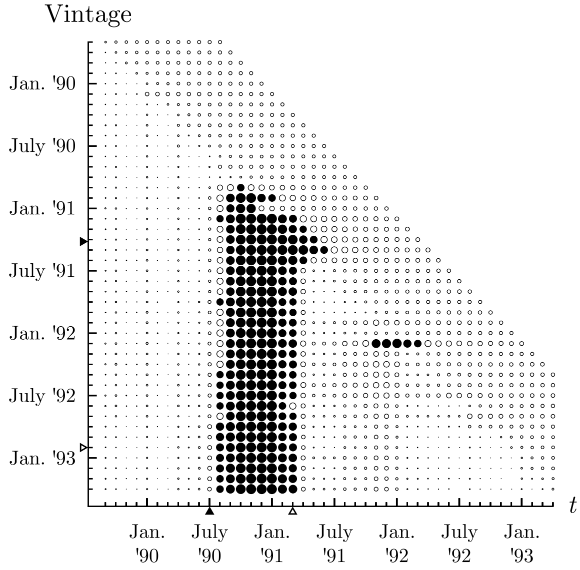
For the period before, during and after the July 1990 - March 1991 recession, all models produce a fairly stable pattern of recession probabilities from the July 1991 vintage onwards. Consistent with the results for the July 2007 vintage in Figure 4.4, we do find that the signals of the synchronous and the non-synchronous common cycle specifications are less clear as compared to the other two specifications, in the sense that we do not find sharp regime switches. Particularly, it would be difficult to date the trough using the synchronous cycle specification. Using the non-synchronous cycle specification we would be unsure about the date of the peak. Based on the asymmetric non-synchronous common cycle specification, we would date the peak at September 1990, two months later than the peak date of the NBER. Our estimate of the trough, however, does exactly match the NBER's date. As regards the speed of detection of the business cycle peak, we observe a first string of recession probabilities higher than 0.5 for the data release of December 1990, four months before the NBER's announcement of the peak. The independent cycles and the non-synchronous common cycle specifications detect the recession two months later, in February 1991. As early as June 1991 our model indicates that a trough had occurred in March. The NBER announced the end of this recession in December 1992, nearly one and a half year after the model's time of detection.

The results for the March - November 2001 recession in Figure 4.6 demonstrate the advantage of the asymmetric non-synchronous common cycle specification more convincingly. The most prominent feature of the recession probabilities in these graphs is the false signal given by the other three cycle specifications for the period September 2002 - September 2003. The increase in recession probabilities during this period is much less for the asymmetric non-synchronous common cycle specification. For the 2001 recession itself, the two non-synchronous common cycle models provide a much earlier and clearer signal of the peak than the independent cycles and synchronous common cycle specifications. For both non-synchronous common cycle specifications we observe that the recession probabilities increase for the first time with the July 2001 vintage. Again the models signal this recession considerable faster than the NBER, whose announcement of the peak came only four months later in November 2001. The timeliness of the models is even more pronounced for the subsequent trough, which we first detect when using the March 2002 data vintage, whereas the corresponding NBER announcement came only in July 2003. Concerning the dating of the turning points, based on the asymmetric lead/lag specification we would have concluded that the recession started in January 2001,

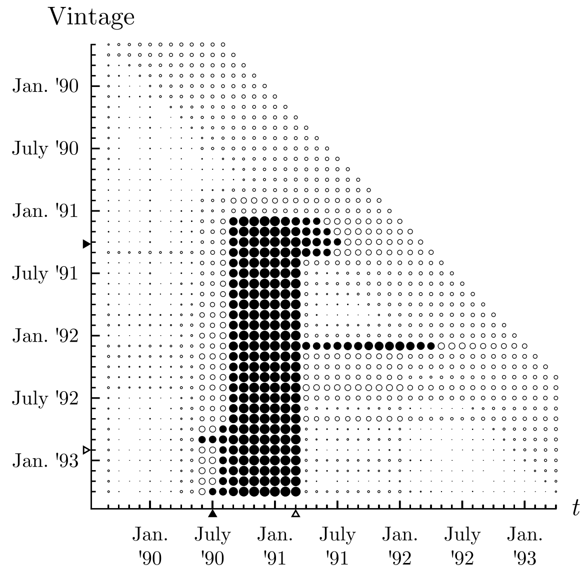
Figure 4.5: In-sample estimates and out-of-sample predictions of recession probabilities in a rolling horizon: The July 1990 - March 1991 recession



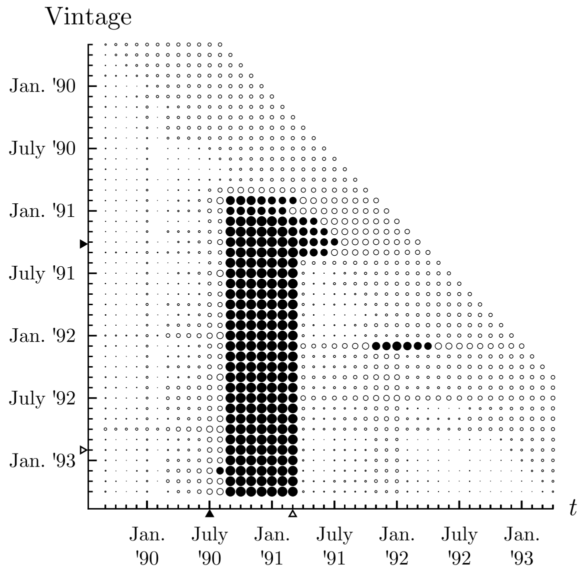
(a) Independent cycles



(b) Synchronous common cycle



(c) Non-synchronous common cycle

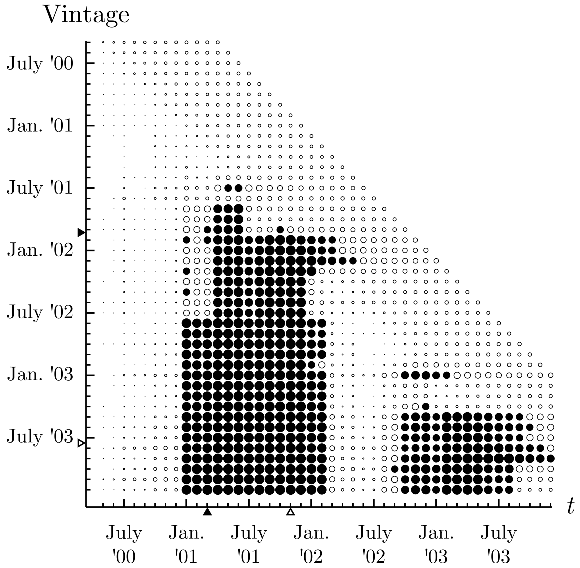


(d) Asymmetric non-synchronous common cycle

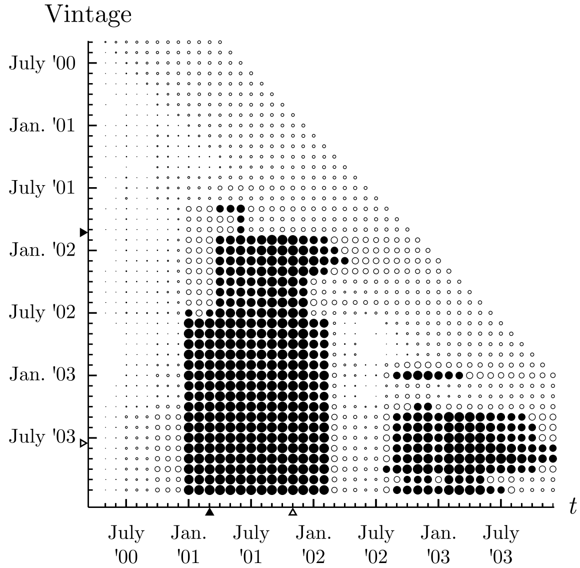
Probability scale: 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Note: The graphs present the estimated and predicted recession probabilities in a rolling horizon, where at every point in time the latest data vintage is used to compute in-sample estimates for the past and out-of-sample predictions for the next 12 months ahead. On the vertical axes, the announcement date of the July 1990 business cycle peak (April 25, 1991) and the March 1991 business cycle trough (December 22, 1992) are marked by the black and white pointers, respectively. Likewise, on the horizontal axes, the pointers mark the dates of the peak and trough itself.

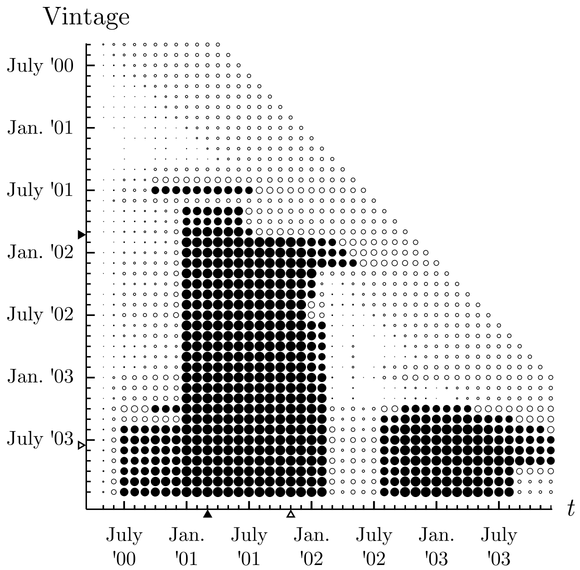
Figure 4.6: In-sample estimates and out-of-sample predictions of recession probabilities in a rolling horizon: The March - November 2001 recession



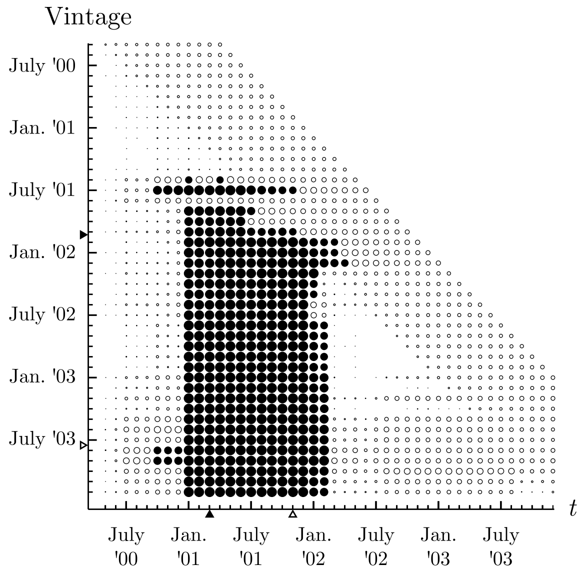
(a) Independent cycles



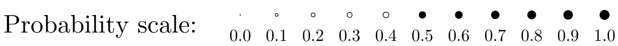
(b) Synchronous common cycle



(c) Non-synchronous common cycle



(d) Asymmetric non-synchronous common cycle



Note: The graphs present the estimated and predicted recession probabilities in a rolling horizon, where at every point in time the latest data vintage is used to compute in-sample estimates for the past and out-of-sample predictions for the next 12 months ahead. On the vertical axes, the announcement date of the March 2001 business cycle peak (November 26, 2001) and the November 2001 business cycle trough (July 17, 2003) are marked by the black and white pointers, respectively. Likewise, on the horizontal axes, the pointers mark the dates of the peak and trough itself.

two months earlier than it started according to the NBER. The end of the recession also occurs a few months after the official trough in November 2001.

4.5.2 Forecasting results

We conclude our analysis by evaluating the (relative) accuracy of real-time h-month ahead forecasts of the business cycle regime and of CCI growth rates, for $h = 1, 2, \ldots, 12$. To evaluate the latter forecasts we use the mean squared forecast error (MSFE)

$$\text{MSFE}(h) = \frac{1}{T_2 - h - T_1} \sum_{t=T_1-1}^{T_2-h} \left(y_{1,t+h} - \hat{y}_{1,t|t+h}\right)^2, \tag{4.23}$$

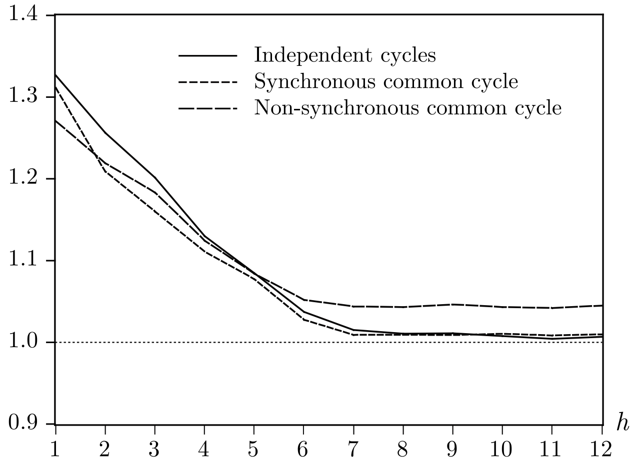
where T_1 and T_2 correspond to the first and last data vintages considered (January 1989 and July 2007), where $\hat{y}_{1,t|t+h}$ denotes the h-step ahead forecast made at time $t+1$ and where $y_{1,t+h}$ is the monthly CCI growth rate resulting from the release of the series in July 2007. To measure the deviation between the regime variable $s_{1,t}$ and the recessions according to the NBER turning points, we use the turning point forecast error (TPFE)

$$\text{TPFE}(h) = \frac{1}{T_2 - h - T_1} \sum_{t=T_1-1}^{T_2-h} \left(\text{NBER}_{t+h} - \hat{s}_{1,t|t+h}\right)^2, \quad \text{for } h = 1, \ldots, 12, \tag{4.24}$$

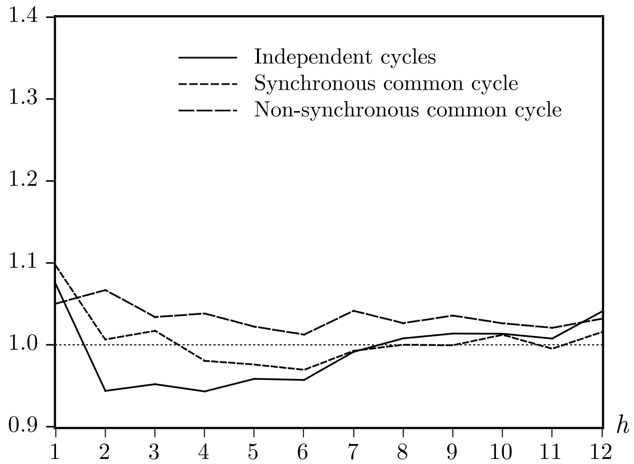
where $\hat{s}_{1,t|t+h}$ denotes the h-step ahead forecast of the state variable made at time $t+1$, and where NBER_{t+h} is a binary variable which equals 1 if, according to the NBER turning points, the economy is in recession at time $t+h$. An alternative approach to forecasting turning points based directly on the NBER turning points or a different sequence of peaks and troughs is the Qual VAR model developed in Dueker (2005). Extending this model to explicitly incorporate asymmetric lead times at peaks and at troughs is an interesting challenge, but beyond the scope of this chapter.

To facilitate the forecast comparison, we take the most general model specification, that is, the asymmetric non-synchronous common cycle model as reference point. The first panel of Figure 4.7 displays ratios of the TPFE(h) for h-step ahead forecasts of the probability of recession obtained from the MS-VAR models with independent cycles, with a synchronous common cycle and with a symmetric non-synchronous common cycle relative to the asymmetric non-synchronous common cycle model. We observe that the asymmetric model outperforms the other models in forecasting turning points as all ratios exceed unity. The improvement in forecast accuracy is especially large for short horizons, up to 30 percent. For horizons longer than 6 months the differences are relatively small, especially with the models with independent cycles and with a synchronous common cycle.

Figure 4.7: Regime and CCI growth rate prediction error ratios



(a) Regime



(b) CCI growth rate

Note: The graphs present forecasting error ratios obtained from the MS-VAR models with independent cycles, with a synchronous common cycle and with a symmetric non-synchronous common cycle relative to the asymmetric non-synchronous common cycle model. In Panel (a), the TPFE(h) for forecasts of the recession probabilities are shown, where $h = 1, 2, \ldots, 12$. Panel (b) shows the MSFE(h) for forecasts of CCI growth.

We test whether the differences in TPFE's are statistically significant by means of the Diebold and Mariano (1995) test of equal predictive accuracy. The results in the left panel of Table 4.3 for forecast horizons $h = 1, 2, \ldots, 6, 9$ and 12 months confirm the graphical evidence in Figure 4.7. In particular, the asymmetric non-synchronous common cycle model provides significantly more accurate forecasts than the other three variants for horizons up to five months. For longer horizons, only the symmetric non-synchronous cycle specification is still significantly worse than our asymmetric lead time specification.

The second panel of Figure 4.7 displays the ratios of the MSFE(h) for $h = 1, \ldots, 12$ for CCI growth forecasts for the same three models relative to the asymmetric non-synchronous cycle model. Here the results are more mixed. For forecast horizons of one month and longer than seven months, the novel cycle specification outperforms the other models. The Diebold and Mariano (1995) test results displayed in the right panel of Table 4.3 indicate that the difference in MSFE for one-month ahead forecasts is significant. The same holds for 12-month ahead forecasts for the models with independent cycles and with the symmetric non-synchronous common cycle. At intermediate horizons, the independent cycle specification provides the most accurate forecasts, although the gain relative to the asymmetric non-synchronous common cycle specification is not statistically significant, see Table 4.3.

In sum, our model produces sharper and more accurate turning point estimates, in particular for business cycle peaks. Concerning the speed of detection, our model proves to be advantageous especially to detect business cycle troughs. For the last two recessions, the troughs where detected over a year ahead of the NBER's announcements. The asymmetric non-synchronous common cycle model also provides more accurate forecasts than the other models, especially for turning points.

4.6 Conclusions

In this chapter we have developed a formal statistical approach to investigate whether the lead time of leading indicator variables is different at business cycle peaks and troughs. A novel Markov switching vector autoregressive model, where economic growth and leading indicators share a common Markov process determining the state but with different lead times at switches between the different regimes, was proposed for this purpose. The empirical application involving The Conference Board's monthly Composite Coincident Index and Composite Leading Index demonstrates the usefulness of the new model specification. For the period January 1959 - June 2007, we found that on average the CLI led CCI by nearly one year at peaks, but only by one quarter at troughs. Therefore, in

Table 4.3: Testing equal predictive accuracy

Horizon	Predicting s_1			Predicting y_1 (CCI)		
h	Independent cycles	Synchronous common cycle	Non-synchronous common cycle	Independent cycles	Synchronous common cycle	Non-synchronous common cycle
1	2.26***	2.16**	1.88**	0.63**	0.83***	0.43***
	(0.83)	(0.88)	(0.80)	(0.30)	(0.30)	(0.14)
2	1.92***	1.57**	1.65**	−0.46	0.05	0.54***
	(0.73)	(0.70)	(0.69)	(0.33)	(0.20)	(0.16)
3	1.59**	1.26**	1.45**	−0.41	0.15	0.29
	(0.62)	(0.60)	(0.58)	(0.41)	(0.39)	(0.20)
4	1.06***	0.91*	1.01***	−0.47	−0.16	0.31*
	(0.39)	(0.50)	(0.37)	(0.37)	(0.27)	(0.17)
5	0.71**	0.65*	0.71***	−0.33	−0.19	0.17
	(0.31)	(0.36)	(0.26)	(0.34)	(0.23)	(0.19)
6	0.32	0.24	0.45*	−0.35	−0.25	0.10
	(0.20)	(0.25)	(0.23)	(0.28)	(0.21)	(0.20)
9	0.09	0.07	0.39**	0.10	0.00	0.26***
	(0.16)	(0.20)	(0.15)	(0.14)	(0.12)	(0.10)
12	0.05	0.08	0.38***	0.29**	0.11	0.23***
	(0.16)	(0.19)	(0.11)	(0.14)	(0.14)	(0.08)

Note: The left panel of the table shows the difference between the TPFE(h) defined in (4.24) of the indicated cycle specification and the TPFE(h) of the asymmetric non-synchronous common cycle specification. The right panel shows corresponding differences for the MSFE(h) defined in (4.23). Heteroskedasticity and autocorrelation consistent (HAC) standard errors in parentheses. All numbers are ×100. The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively, of the Diebold-Mariano (1995) statistic for testing the null hypothesis that the difference in TPFE(h) or MSFE(h) is equal to zero.

terms of timeliness, the CLI is most useful for signaling oncoming recessions. In addition, in a real-time business cycle dating and forecasting exercise for the vintages in the period January 1989 until July 2007, we found that allowing for asymmetric lead times leads to more timely and precise identification of peaks and troughs for the 1990/1991 and 2001 recessions, as well as more accurate out-of-sample turning point forecasts.

4.A Full Conditional Posterior Distributions

4.A.1 Sampling of $\mathbf{\Gamma}_0$

To sample $\mathbf{\Gamma}_0$ we rewrite (4.13) as

$$\begin{aligned}\mathbf{\Sigma}_t^{-\frac{1}{2}} \Big((\boldsymbol{Y}_t - \mathbf{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_t) - \textstyle\sum_{i=1}^k \mathbf{\Phi}_i (\boldsymbol{Y}_{t-i} - \mathbf{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_{t-i}) \Big) \\ = \mathbf{\Sigma}_t^{-\frac{1}{2}} \Big(\boldsymbol{I}_m - \textstyle\sum_{i=1}^k \mathbf{\Phi}_i \Big) \mathbf{\Gamma}_0 + \mathbf{\Sigma}_t^{-\frac{1}{2}} \boldsymbol{\mathcal{E}}_t,\end{aligned} \tag{4.A.1}$$

where $\boldsymbol{I}_m$ denotes the $(m \times m)$ identity matrix. This is a regression model with parameter $\mathbf{\Gamma}_0$ and an error term with unit variance. Hence, the full conditional posterior distribution of $\mathbf{\Gamma}_0$ is normal with mean $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$ and covariance matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ where $\boldsymbol{Z} = (\boldsymbol{Z}'_{k+1}, \ldots, \boldsymbol{Z}'_T)'$ with $\boldsymbol{Z}_t = \mathbf{\Sigma}_t^{-\frac{1}{2}} \big((\boldsymbol{Y}_t - \mathbf{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_t) - \sum_{i=1}^k \mathbf{\Phi}_i (\boldsymbol{Y}_{t-i} - \mathbf{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_{t-i}) \big)$ and $\boldsymbol{X} = (\boldsymbol{X}'_{k+1}, \ldots, \boldsymbol{X}'_T)'$ with $\boldsymbol{X}_t = \mathbf{\Sigma}_t^{-\frac{1}{2}} (\boldsymbol{I}_m - \sum_{i=1}^k \mathbf{\Phi}_i)$, see for example Zellner (1971, Chapter III). The prior restriction for identification can easily be incorporated by sampling from truncated normal distributions.

4.A.2 Sampling of $\mathbf{\Gamma}_1$

To sample $\mathbf{\Gamma}_1$ we rewrite (4.13) as

$$\begin{aligned}\mathbf{\Sigma}_t^{-\frac{1}{2}} \Big((\boldsymbol{Y}_t - \mathbf{\Gamma}_0) - \textstyle\sum_{i=1}^k \mathbf{\Phi}_i (\boldsymbol{Y}_{t-i} - \mathbf{\Gamma}_0) \Big) \\ = \mathbf{\Sigma}_t^{-\frac{1}{2}} \Big(\boldsymbol{I}_m \odot (\boldsymbol{\mathcal{S}}_t - \textstyle\sum_{i=1}^k \mathbf{\Phi}_i \boldsymbol{\mathcal{S}}_{t-i}) \Big) \mathbf{\Gamma}_1 + \mathbf{\Sigma}_t^{-\frac{1}{2}} \boldsymbol{\mathcal{E}}_t.\end{aligned} \tag{4.A.2}$$

This is again a regression model with parameter $\mathbf{\Gamma}_1$ and an error term with unit variance. The full conditional posterior distribution of $\mathbf{\Gamma}_1$ is normal with mean $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$ and covariance matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ where $\boldsymbol{Z} = (\boldsymbol{Z}'_{k+1}, \ldots, \boldsymbol{Z}'_T)'$ with $\boldsymbol{Z}_t = \mathbf{\Sigma}_t^{-\frac{1}{2}} \big((\boldsymbol{Y}_t - \mathbf{\Gamma}_0) - \sum_{i=1}^k \mathbf{\Phi}_i (\boldsymbol{Y}_{t-i} - \mathbf{\Gamma}_0) \big)$ and $\boldsymbol{X} = (\boldsymbol{X}'_{k+1}, \ldots, \boldsymbol{X}'_T)'$ with $\boldsymbol{X}_t = \mathbf{\Sigma}_t^{-\frac{1}{2}} \times \big(\boldsymbol{I}_m \odot (\boldsymbol{\mathcal{S}}_t - \sum_{i=1}^k \mathbf{\Phi}_i \boldsymbol{\mathcal{S}}_{t-i}) \big)$. Again, the prior restriction for identification can easily be incorporated by sampling from truncated normal distributions.

4.A.3 Sampling of $\mathbf{\Phi}$

To sample $\mathbf{\Phi}$ we note that (4.13) is a multivariate regression model with regression parameters $\mathbf{\Phi}_i$ for $i = 1, \ldots, k$. Define $\boldsymbol{Z}_t = (\boldsymbol{Y}_t - \mathbf{\Gamma}_0 - \mathbf{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_t)$ and $\boldsymbol{Z}^j = (\boldsymbol{Z}_{k-j+1}, \ldots, \boldsymbol{Z}_{T-j})'$. This multivariate regression model can be written as

$$\boldsymbol{Z} = \boldsymbol{X}\mathbf{\Phi} + \boldsymbol{e} \tag{4.A.3}$$

where $\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{Z}^0$, $\boldsymbol{X} = (\boldsymbol{D}\boldsymbol{Z}^1, \boldsymbol{D}\boldsymbol{Z}^2, \ldots, \boldsymbol{D}\boldsymbol{Z}^k)$, $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{\Sigma}_{k+1}^{-\frac{1}{2}}, \ldots, \boldsymbol{\Sigma}_T^{-\frac{1}{2}})$, $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_k)'$ and $\boldsymbol{e} = (\boldsymbol{\Sigma}_{k+1}^{-\frac{1}{2}}\boldsymbol{\mathcal{E}}_{k+1}, \ldots, \boldsymbol{\Sigma}_T^{-\frac{1}{2}}\boldsymbol{\mathcal{E}}_T)'$. Hence, the full conditional posterior distribution of $\boldsymbol{\Phi}$ is a matric variate normal distribution with mean $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$ and covariance matrix $\boldsymbol{I}_m \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}$, see Zellner (1971, Chapter VIII). Here $\otimes$ denotes the Kronecker product. To sample in case we have zero restrictions on the elements of $\boldsymbol{\Phi}$ we rewrite (4.A.3) in a univariate linear regression model with regression parameter $\mathrm{vec}(\boldsymbol{\Phi})$ using the vec operator, that is,

$$\mathrm{vec}(\boldsymbol{Z}) = \mathrm{vec}(\boldsymbol{I} \otimes \boldsymbol{X})\mathrm{vec}(\boldsymbol{\Phi}) + \mathrm{vec}(\boldsymbol{e}) \tag{4.A.4}$$

and hence one can sample the nonzero elements of $\mathrm{vec}(\boldsymbol{\Phi})$ from a normal distribution with mean $(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}(\tilde{\boldsymbol{X}}'\mathrm{vec}(\boldsymbol{Z}))$ and covariance matrix $(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}$ where $\tilde{\boldsymbol{X}}$ contains the columns of $\mathrm{vec}(\boldsymbol{I} \otimes \boldsymbol{X})$ corresponding to the nonzero elements of $\mathrm{vec}(\boldsymbol{\Phi})$.

4.A.4 Sampling of $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$

It is easy to see from the conditional likelihood (4.15) and the prior specification (4.21) that the full conditional posterior of density $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ is proportional to

$$\begin{aligned} f(\boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1 | \boldsymbol{s}_1^T, \boldsymbol{\theta}\backslash\{\boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1\}, \boldsymbol{Y}^T) &\propto |\boldsymbol{\Omega}_0|^{-\frac{1}{2}(\tau-k)} |\boldsymbol{\Omega}_1|^{-\frac{1}{2}(T-\tau+2)} \\ &\times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_0^{-1}\left(\sum_{t=k+1}^{\tau-1}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t'\right) + \boldsymbol{\Omega}_1^{-1}\left(\sum_{t=\tau}^{T}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t'\right)\right)\right), \end{aligned} \tag{4.A.5}$$

where $A\backslash B$ denotes the set A excluding the elements in set B. Hence, the covariance matrices $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ can be sampled from inverted Wishart distributions with scale parameters $(\sum_{t=k+1}^{\tau-1}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t')$ and $(\sum_{t=\tau}^{T}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t')$ and degrees of freedom $\tau - k - 1$ and $T - \tau + 1$, respectively, see Zellner (1971, p. 395).

4.A.5 Sampling of $\boldsymbol{p}$ and $\boldsymbol{q}$

From the conditional likelihood function (4.15) it follows that the full conditional posterior densities of the transition parameters are given by

$$f(p|\boldsymbol{s}_1^T, \boldsymbol{\theta}\backslash\{p\}, \boldsymbol{Y}^T) \propto p^{\boldsymbol{\mathcal{N}}_{0,0}}(1-p)^{\boldsymbol{\mathcal{N}}_{0,1}} \tag{4.A.6}$$

$$f(q|\boldsymbol{s}_1^T, \boldsymbol{\theta}\backslash\{q\}, \boldsymbol{Y}^T) \propto q^{\boldsymbol{\mathcal{N}}_{1,1}}(1-q)^{\boldsymbol{\mathcal{N}}_{1,0}}, \tag{4.A.7}$$

where $\boldsymbol{\mathcal{N}}_{i,j}$ again denotes the number of transitions from state i to state j. This implies that the transition probabilities can be sampled from Beta distributions with parameters $\boldsymbol{\mathcal{N}}_{0,0} + 1$ & $\boldsymbol{\mathcal{N}}_{0,1} + 1$, and $\boldsymbol{\mathcal{N}}_{1,1} + 1$ & $\boldsymbol{\mathcal{N}}_{1,0} + 1$, respectively. In case we have separate state variables for the two series we can sample both transition probabilities separately.

4.A.6 Sampling of $\boldsymbol{\tau}$

The full conditional posterior density of τ is given by

$$f(\tau|\boldsymbol{s}_1^T, \boldsymbol{\theta}\backslash\{\tau\}, \boldsymbol{Y}^T) \propto \mathbb{I}[b+k+1 < \tau \leq T-b] \times |\boldsymbol{\Omega}_0|^{-\frac{1}{2}(\tau-k)} |\boldsymbol{\Omega}_1|^{-\frac{1}{2}(T-\tau+2)} \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_0^{-1}\left(\sum_{t=k+1}^{\tau-1}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t'\right) + \boldsymbol{\Omega}_1^{-1}\left(\sum_{t=\tau}^{T}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t'\right)\right)\right). \quad (4.A.8)$$

As τ has can only take discrete values on the range $(\nu+k+1, T-\nu]$ we can easily sample from its full conditional posterior distribution.

4.A.7 Sampling of $\boldsymbol{\kappa}$

If our model contains non-synchronous cycles in both series we have to sample one or two κ parameters. As the κ parameters are discrete we can compute the value of the posterior distribution for $\kappa_j \in \{-c_j, \ldots, c_j\}$ and scale these values such that they add up to one. We can now easily sample a value for κ. Note that we can sample κ_1 and κ_2 at once from their joint full conditional distribution.

4.A.8 Sampling of the states

To sample the states, we need the full conditional posterior density of $s_{1,t}$, denoted by $f(s_{1,t}|\boldsymbol{s}_1^{-t}, \boldsymbol{\theta}, \boldsymbol{Y}^T)$, $t = 1, \ldots, T$, where $\boldsymbol{s}_1^{-t} = \boldsymbol{s}_1^T\backslash\{s_{1,t}\}$. Since $s_{1,t}$ follows a first-order Markov process, it is easily seen that

$$f(s_{1,t}|\boldsymbol{s}_1^{-t}) \propto f(s_{1,t}|s_{1,t-1})\, f(s_{1,t+1}|s_{1,t}) \quad (4.A.9)$$

due to the Markov property. Hence, the full conditional distribution of $s_{1,t}$ is given by

$$f(s_{1,t}|\boldsymbol{s}_1^{-t}, \boldsymbol{\theta}, \boldsymbol{Y}^T) \propto f(s_{1,t}|s_{1,t-1}, \boldsymbol{\theta})\, f(s_{1,t+1}|s_{1,t}, \boldsymbol{\theta}) \prod_{i=t-\kappa_{\min}}^{t+k+\kappa_{\max}} f(\boldsymbol{Y}_i|\boldsymbol{Y}^{i-1}, \boldsymbol{\mathcal{S}}^i, \boldsymbol{\theta}), \quad (4.A.10)$$

where $f(\boldsymbol{Y}_t|\boldsymbol{Y}^{t-1}, \boldsymbol{\mathcal{S}}^t, \boldsymbol{\theta})$ is defined in (4.14), $\kappa_{\max} = \max(\kappa_1, \kappa_2)$, $\kappa_{\min} = \min(\kappa_1, \kappa_2)$ and the constant of proportionality can be obtained by summing over the two possible values of $s_{1,t}$. At time $t = T$ the term $f(s_{1,T+1}|s_{1,T}, \boldsymbol{\theta})$ drops out. The first k states can be sampled from the full conditional distribution

$$f(s_{1,t}|\boldsymbol{s}_1^{-t}, \boldsymbol{\theta}, \boldsymbol{Y}^T) \propto f(s_{1,t}|s_{1,t-1}, \boldsymbol{\theta})\, f(s_{1,t+1}|s_{1,t}, \boldsymbol{\theta}) \prod_{i=k+1}^{t+k+\kappa_{\max}} f(\boldsymbol{Y}_i|\boldsymbol{Y}^{i-1}, \boldsymbol{\mathcal{S}}^i, \boldsymbol{\theta}) \quad (4.A.11)$$

for $t = 1, \ldots, k$, where at time $t = 1$ the term $f(s_{1,t}|s_{1,t-1}, \boldsymbol{\theta})$ is replaced by the unconditional density $f(s_{1,1}|\boldsymbol{\theta})$, which is a binomial density with probability $(1-p)/(2-p-q)$.

Sampling of the state variables can be done as follows. Take the most recent value of $\boldsymbol{s}_1^T$ and sample the states backward in time, one after another, starting with $s_{1,T}$. After each step, the t-th element of $\boldsymbol{s}_1^T$ is replaced by its most recent draw.

Chapter 5

Evaluating the Consistency and Timeliness of Leading Indicators

5.1 Introduction

There has been considerable debate about the recent performance of the leading indicators selected by The Conference Board for inclusion in its Composite Leading Index (CLI), see Stock and Watson (2003). Apart from being statistically reliable and promptly available, the primary criteria for evaluating the quality of a leading indicator are its timeliness and consistency, see Marcellino (2006) for an excellent discussion. Timeliness means that in order to be useful, a leading indicator variable should have a considerable lead time with respect to business cycle turning points. Consistency refers to the property that the lead timing of an indicator should ideally be constant across cycles. In practice it has proven to be challenging to find leading indicators that are both timely and consistent. Indicator variables that tend to have a considerable lead time are usually not very consistent, and vice versa. For example, the INTEREST RATE SPREAD has been lauded for its timeliness and its ability to predict turning points (see e.g. Stock and Watson, 1989, Estrella and Hardouvelis, 1991, Dueker, 1991, and Estrella and Mishkin, 1998), but questioned for its inconsistency in recent decades (see e.g. Heubrich and Dombrosky, 1996, and Dotsey, 1998). In contrast, while the indicator CHANGE IN CONSUMER INSTALLMENT DEBT was considered sufficiently consistent, in 1975 it was dropped from the list of leading indicators because it lacked timeliness, in particular at troughs (Zarnowitz and Boschan, 1975). In 1989 it was restored in the form of CHANGE IN CREDIT OUTSTANDING, but it was dropped again in 1996, see Klein (2001) for an historical perspective on the shortlist of indicators.

In our view, there are two aspects which complicate the discussion on timeliness and consistency of leading indicator variables. Firstly, it has long been a challenge to statistically estimate the lead time of an indicator and to test for its significance, see Hamilton and Perez-Quiros (1996) and Camacho and Perez-Quiros (2002) for successful work in this area. Indeed, based on the limited number of observed turning points during the post-war period alone it is hard to render clear-cut lead time estimates, and models have to be developed to estimate the lead times using the complete time series. Secondly, the apparent inconsistency of leading indicators may merely be due to the fact that their lead time tends to vary with the phase of the business cycle. Indeed, leading indicators tend to have a longer lead time when entering a recession than when entering an expansion, see The Conference Board (2001). An assessment of the true consistency of leading indicators should take this feature into account.

The goal of the present chapter is to develop a formal approach to assess the timeliness and consistency of business cycle indicators. For this purpose, we develop a Markov Switching panel data model to relate the cycles of the indicators to the cycle of a reference series, where the cycle of the reference series is assumed to coincide with the business cycle. We employ the asymmetric non-synchronous common cycle specification of Chapter 4, which assumes that the latent underlying cycle of the reference series is shared by the indicators, but with a certain shift that can be different for expansions and recessions. Consequently, this specification allows us to estimate the lead times of the indicators separately at business cycle peaks and at troughs. As such, we aim to account for fluctuations in the lead/lag times purely due to the phase of the business cycle.

For the lead time estimates to be reliable, it is crucial that the model captures the business cycle as good as possible. To facilitate this, we incorporate the information in a large number of business cycle indicators. This draws upon the notion of Burns and Mitchell (1946, p. 3) that the business "cycle consist of expansions (and recessions) occurring at about the same time in many economic activities." As a consequence, we choose to formulate our model for panels of time series rather than as a vector autoregression (VAR). Although VARs are frequently applied in the business cycle literature, see for example the work by Hamilton and Perez-Quiros (1996), Camacho and Perez-Quiros (2002), Paap *et al.* (2009), they are typically not tractable for moderately large numbers of variables, as relevant here.

To perform inference on the model parameters we employ a Bayesian approach, with posterior results being obtained through flexible Markov Chain Monte Carlo techniques. This approach allows us to treat the lead times at peaks and at troughs as unknown parameters and to obtain the posterior distributions of these lead times directly. The posterior lead time distributions both provide estimates of the average timeliness of the indicators, but their shape also reveals what is the uncertainty around this average, or in other words, whether the indicators have consistent lead timing.

Our empirical analysis involves The Conference Board's Composite Coincident Index (CCI) and the ten components included in its Composite Leading Index (CLI), measured over the period January 1959 - February 2008. We find that the indicators BUILDING PERMITS, STOCK PRICES, MONEY SUPPLY, and CONSUMER EXPECTATIONS are the most timely and consistent indicators among the ten, having posterior mean lead times of nine to ten months at peaks and four to five months at troughs with standard deviations of around one month. In spite of this, none of the individual indicators seems to be as timely and consistent as the CLI is, according to the results in Chapter 4. This validates the use of composite leading indices and the ongoing search for better indices. To contribute to this search, we assess whether we can construct a model based composite leading index that is more reliable than the currently popular CLI. This assessment builds upon the seminal work of Stock and Watson (1989). Recent contributions in this area include Kim and Nelson (1998), McGuckin *et al.* (2007) and Serati and Amisano (2008). We construct a separate leading index at peaks and at troughs by synchronizing the ten indicators according to their respective posterior lead time distributions before aggregation. The synchronized indices thus obtained seem to indeed outperform the CLI in the sense that they yield better in-sample predictions of the business cycle chronology as determined by the NBER.

The chapter is organized as follows. In Section 5.2, we introduce the novel Markov switching panel data model. We discuss our empirical application to The Conference Board's CCI and the ten components included in its CLI in Section 5.3. Along the way, we also provide the most important features of the Bayesian approach for estimation and inference in the model, with full details being presented in Appendix 5.A. The estimation results are used in Section 5.4 to construct our synchronized composite leading index. Finally, we conclude in Section 5.5 and discuss several directions for further research.

5.2 Model Specification

Consider a panel of growth rates of economic indicators $y_{i,t}$, where the indicators are indexed by i, for $i = 1, \ldots, N$, and time in months by t, for $t = 1, \ldots, T$. We assume that the indicators exhibit cyclical behavior, and that their cycles consist of two regimes, to be labeled 'recession' (1) and 'expansion' (0), which are characterized by different mean growth rates. More specifically, we assume that all indicators share the same cycle, but possibly with a time shift. Hence, the individual variables may be leading, coincident, or lagging indicators.

The most parsimonious model which satisfies the above reads

$$y_{i,t} = \mu_{i,s_{i,t}} + \varepsilon_{i,t}, \tag{5.1}$$

where the latent indicator $s_{i,t} \in \{0,1\}$ denotes the state of the cycle of each series, and where $\varepsilon_{i,t}$ is a white noise process with zero mean and (co-)variances denoted as $\mathsf{E}[\varepsilon_{i_1,t}\varepsilon_{i_2,t}] = \sigma_{i_1,i_2}$, where $i_1, i_2 = 1, \ldots, N$. In addition, we assume that ε_{i_1,t_1} is independent of s_{i_2,t_2} for all $t_1, t_2 = 1, \ldots, T$ and all $i_1, i_2 = 1, \ldots, N$. We impose $\mu_{i,1} < 0 < \mu_{i,0}$ for $i = 1, \ldots, N$, such that recessions and expansions correspond to periods with negative and positive average growth, respectively.

The model is completed by specifying the properties of the regime indicators $s_{i,t}$. For identification purposes we label the first series in the panel $y_{1,t}$ as the reference series, for which the turning points coincide with the business cycle. As a result, the regime indicator $s_{1,t}$ of this series will be referred to as the business cycle. We assume that the business cycle can be described as a homogenous first-order Markov process with transition probabilities

$$\Pr[s_{1,t} = 0 | s_{1,t-1} = 0] = p \quad \text{and} \quad \Pr[s_{1,t} = 1 | s_{1,t-1} = 1] = q. \tag{5.2}$$

The cycles of the other $N-1$ indicators $y_{2,t}, \ldots, y_{N,t}$ are assumed to be the same as the cycle of the references series, but are allowed to be non-synchronous such that they lead, coincide with, or lag the cycle of $y_{1,t}$ by $\kappa_{i,1}$ months at peaks and $\kappa_{i,2}$ months at troughs. We can formalize this interrelationship by defining the regime indicators $s_{i,t}$ of the other series as

$$s_{i,t} = \begin{cases} \prod\limits_{k=\kappa_{i,1}}^{\kappa_{i,2}} s_{1,t+k} & \text{if } \kappa_{i,1} \leq \kappa_{i,2} \\ 1 - \prod\limits_{k=\kappa_{i,2}}^{\kappa_{i,1}} (1 - s_{1,t+k}) & \text{if } \kappa_{i,1} > \kappa_{i,2} \end{cases} \quad \text{for } i = 2, \ldots, N, \tag{5.3}$$

as in Chapter 4. Their corresponding lead times $\kappa_{i,1}$ and $\kappa_{i,2}$ are treated as parameters to be estimated.

An advantage of our modeling approach is that we do not require the indicators $y_{2,t}, \ldots, y_{N,t}$ to be leading indicators. These may also be coincident or even lagging indicators, or a mixture of different types of indicators. Positive values of $\kappa_{i,1}$ and $\kappa_{i,2}$ indicate that the indicator i leads $\kappa_{i,1}$ periods at business cycle peaks and $\kappa_{i,2}$ periods at troughs. Similarly, negative values of $\kappa_{i,1}$ and $\kappa_{i,2}$ correspond to lags of the indicator at peaks and at troughs, respectively. Finally, our specification allows $\kappa_{i,1}$ to be positive while $\kappa_{i,2}$ is negative, or vice versa. The first case, for example, corresponds to the situation that an indicators leads the business cycle at peaks but lags the business cycle at troughs. Although such indicators are typically not selected for inclusion in leading indices, it may be useful to include them in the model in order to improve identification of the common business cycle.

It is also useful to note that our model includes the special case of synchronicity between the cycle of indicator i and the business cycle by setting $\kappa_{i,1} = \kappa_{i,2} = 0$ as in Krolzig (1997), Paap and Van Dijk (2003), and Chauvet and Hamilton (2006). When $\kappa_{i,1} = \kappa_{i,2} \neq 0$, the two cycles are non-synchronous but such that the lead/lag timing of the indicator is the same at peaks and at troughs, as in Hamilton and Perez-Quiros (1996).

5.3 Empirical Results

Our empirical analysis involves The Conference Board's Composite Coincident Index (CCI) and Composite Leading Index (CLI), and the four and ten individual components included therein, respectively. An overview of these series is provided in Table 5.1. The estimation results reported in this section use the revised monthly data as published in March 2008. The sample period starts in January 1959. Most of the series, including the two indices, have a publication lag of one month, which implies that the sample period runs until February 2008. We transform all time series to month-to-month symmetric percentage changes, with some exceptions, as listed in Table 5.1. These transformations are equal to those used by The Conference Board in the construction of their CCI and CLI. Figures 5.3, 5.4 and 5.5 in Appendix 5.B display the time series of the logarithmic levels and growth rates of the CCI, CLI and their individual components, together with the recession periods as determined by the NBER. Generally, the turning points in the coincident indicators seem to coincide with the turning points as implied by the NBER recession periods. The turning points in the leading indicators, however, clearly occur earlier. This is perhaps most easily seen by looking at the leading index. According to preliminary turning point estimates in The Conference Board (2001), only three recessions were missed by the leading indicators during the period 1959 - 1999. The indicator STOCK PRICES missed the recession in 1980 whereas the indicator MONEY SUPPLY missed both the recession in 1960-1961 and the recession in 1981-1982. Inspecting the growth rates, we conclude that the coincident indicators tend to evolve more smoothly than the leading indicators components, an exception being MONEY SUPPLY.

Table 5.1: Dataset and mnemonics

Mnemonic	Variable	Growth rate conversion	Publication lag (in months)
Coincident indicators			
CCI	Composite Coincident Index	SP	1
EM	Employment	SP	1
IP	Industrial Production	SP	1
PI	Personal income less transfer payments	SP	2
MTS	Manufacturing and trade sales	SP	3
Leading indicators			
CLI	Composite Leading Index	SP	1
AWH	Average weekly hours	SP	1
AIC	Average initial claims[1]	−SP	1
NCO	New orders, consumer goods	SP	1
VP	Vendor performance	SP	1
NCA	New orders, capital goods	SP	1
BP	Building permits	SP	1
SP	Stock prices	SP	1
M2	Money supply	SP	1
IRS	Interest rate spread	AC	1
CE	Consumer expectations	SP	1

Note: SP denotes symmetric percentage changes, computed as $\Delta y_{i,t} = 200(y_{i,t} - y_{i,t-1})/(y_{i,t} + y_{i,t-1})$, whereas AC denotes arithmetic changes. [1] This series has been inverted, since typically initial claims increase when employment conditions worsen, see The Conference Board (2001) for details.

In line with McConnell and Perez-Quiros (2000), Sensier and Van Dijk (2004), and Herrera and Pesavento (2005), among others, we observe that for most of our time series the volatility of their growth rates seems to have declined persistently since the mid-1980s.[1] For this reason we extend our Markov switching panel data model by allowing for a single structural break in the (co-)variances of $\varepsilon_{i,t}$, that is

$$\sigma_{i_1,i_2,t} = \begin{cases} \sigma_{i_1,i_2,0} & \text{if } t < \tau \\ \sigma_{i_1,i_2,1} & \text{if } t \geq \tau \end{cases} \tag{5.4}$$

The break point τ is treated as an unknown parameter to be estimated. Detailed results for the model that does not allow for a structural break in (co-)variances are available upon request.

[1] One exception is the indicator CONSUMER EXPECTATIONS. However, this is purely due to the fact that CONSUMER EXPECTATIONS data were originally collected on a quarterly basis. Beginning with 1978, the data are collected monthly. To match the sampling frequency over the sample from 1959 to 2008, the data from 1959 to 1977 are linearly interpolated.

Inference on the parameters and on the regimes of our Markov Switching panel data model is performed by adopting a Bayesian approach. For the purpose of our analysis it is very convenient to perform Bayesian inference, as it provides us with an estimate of the conditional posterior distributions of $\boldsymbol{\kappa}_1 = (\kappa_{2,1}, \ldots, \kappa_{N,1})$ and $\boldsymbol{\kappa}_2 = (\kappa_{2,2}, \ldots, \kappa_{N,2})$ in (5.3), rather than just with point estimates. This helps us to examine the consistency of indicators without having to make distributional assumptions. In Appendix 5.A we derive the likelihood function of the model and discuss prior specification and posterior simulation. We set the indicator-specific maximum and minimum lead/lag times in the priors for $\kappa_{i,j}$, denoted by $l_{i,j}$ and $u_{i,j}$, respectively, equal to -24 and 60 for all series i and for $j = 1, 2$. Hence, we allow for a maximum lead time of five years and a maximum lag time of two years, both at peaks and at troughs.

We estimate the Markov Switching panel data model in (5.1)–(5.3) for the ten individual leading indicators, using The Conference Board's Composite Coincident Index as the reference series $y_{1,t}$. We obtain posterior distributions of the model parameters based on 100,000 simulations following a burn-in period of 10,000 simulations.

We start by examining the properties of the estimated cycle $s_{1,t}$ of the reference series, CCI. This allows us to judge the model on its ability to signal turning points and to identify recession periods. The posterior distribution $f(s_{1,t}|Y^T)$ constructed at the end of month T delivers posterior probability estimates $r_t = \Pr[s_{1,t} = 1|Y^T]$ for $t = 1, 2, \ldots, T$, which we will refer to as recession probabilities. Figure 5.1 shows the developments in these recession probabilities over time. The shaded areas indicate 'recession periods' defined as six consecutive months where the posterior mean of $s_{1,t}$ is larger than 0.5. This corresponds with the popular rule of thumb saying that the economy is in recession whenever economic growth is negative during two consecutive quarters. For all the 'official' recessions that occurred during the sample period according to the NBER, the posterior mean of $s_{1,t}$ is very close to one. In addition, no false recession signals are given except around 1967, which corresponds with a growth rate cycle recession, according to the Economic Cycle Research Institute, see `http://www.businesscycle.com`. The estimation results also suggest that the US economy slipped into a recession around the end of 2007.

Posterior estimates of the transition probabilities p and q are shown at the bottom of Table 5.2. The posterior mean of the probability of staying in the expansion regime is about 0.95. This corresponds to an expected duration of expansions of 20 months. The probability of staying in the recession regime is considerably lower at about 0.82 on average, which corresponds to an expected duration of recessions of between five and six months.

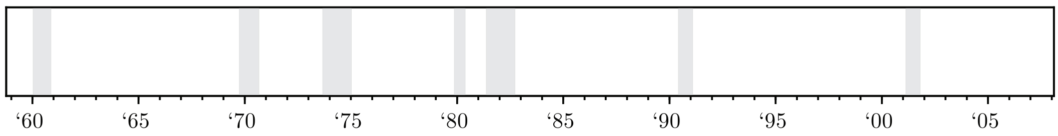
Table 5.2: Posterior means and standard deviations (in parentheses) of the parameters in the Markov Switching panel data model with a structural break in the covariance matrix

Variable	Growth rates		Lead/lag times		Variances	
	$\mu_{i,0}$	$\mu_{i,1}$	$\kappa_{i,1}$	$\kappa_{i,2}$	$\sigma^2_{i,0}$	$\sigma^2_{i,1}$
CCI	0.273***	−0.330***			0.120***	0.060***
	(0.014)	(0.028)			(0.010)	(0.005)
AWH	0.038***	−0.143***	4.956*	1.961	0.351***	0.201***
	(0.027)	(0.070)	(2.624)	(1.124)	(0.029)	(0.018)
AIC	0.438***	−1.756***	6.128***	3.136***	30.384***	17.583***
	(0.234)	(0.594)	(1.570)	(0.765)	(2.562)	(1.561)
NCO	0.317***	−0.576***	8.480***	2.389**	4.855***	3.019***
	(0.162)	(0.329)	(1.777)	(0.956)	(0.404)	(0.269)
VP	0.364***	−1.472***	5.772***	2.316**	55.452***	13.077***
	(0.243)	(0.611)	(2.632)	(0.873)	(4.663)	(1.212)
NCA	0.748***	−1.678***	3.531	−1.036	54.864***	52.053***
	(0.366)	(0.769)	(3.531)	(3.868)	(4.472)	(4.659)
BP	0.596***	−1.917***	8.977***	4.081***	39.926***	20.669***
	(0.280)	(0.611)	(1.038)	(0.495)	(3.335)	(1.841)
SP	1.079***	−1.643***	8.752***	5.210***	10.723***	10.008***
	(0.248)	(0.644)	(1.120)	(0.666)	(0.893)	(0.910)
M2	0.462***	−0.492***	10.051***	4.212***	0.102***	0.204***
	(0.022)	(0.028)	(0.474)	(0.520)	(0.009)	(0.021)
IRS	0.060***	−0.095***	34.643***	17.456***	0.266***	0.087***
	(0.055)	(0.062)	(10.134)	(3.778)	(0.022)	(0.011)
CE	0.468***	−1.256***	8.369***	5.250***	20.073***	31.497***
	(0.272)	(0.713)	(1.438)	(1.096)	(1.673)	(2.760)
Transition probabilities			p	0.951***	q	0.818***
				(0.011)		(0.040)
Most likely break date			τ	1984:03		

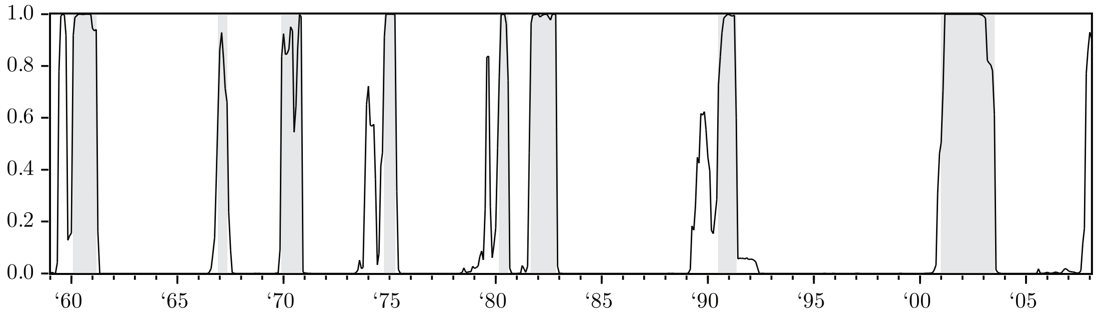
Note: The table presents posterior means and standard deviations (in parentheses) of parameters in the Markov Switching panel data model with a single structural change in the covariance matrix as in (5.4), estimated for monthly growth rates of The Conference Board's CCI and the ten components of its CLI over the period January 1959 - February 2008. The most likely break date is defined as the mode of the posterior distribution of τ. Posterior results are based on 100,000 simulations. Number of burn-in simulations is 10,000. The superscripts *, **, *** indicate that the value zero is not contained in the 90%, 95%, 99% HPD-region, respectively.

Figure 5.1: Posterior recession probabilities in the Markov Switching panel data model

(a) NBER turning points



(b) Posterior recession probabilities



Note: The graph presents posterior recession probabilities in the Markov Switching panel data model with a single structural change in the covariance matrix as in (5.4), estimated for monthly growth rates of The Conference Board's CCI and the ten components of its CLI over the period January 1959 - February 2008. The shaded areas in Panel (a) are recession periods based upon the NBER turning points. The shaded areas in Panel (b) correspond to periods of (at least) six consecutive months where the posterior mean of $s_{1,t}$ is larger than 0.5.

Columns 2 and 3 of Table 5.2 show the posterior means and standard deviations of the growth rates in expansion and recession regimes. Both $\mu_{i,0}$ and $\mu_{i,1}$ vary considerably across series, with growth rates ranging from 0.1% to 1.1% during expansions and -1.9% and -0.1% during recessions. It holds that for all series growth is higher in absolute sense in recessions than in expansions, reflecting the fact that recessions are shorter than expansion but more severe in terms of growth.

Columns 4 and 5 of Table 5.2 show the estimates of the lead/lag times at peaks and at troughs, as measured by the $\kappa_{i,1}$ and $\kappa_{i,2}$ parameters, respectively. In Figure 5.6 in Appendix 5.B, we depict the marginal posterior distributions of $\kappa_{i,1}$ and $\kappa_{i,2}$, as well as their joint posterior distribution. Firstly, we focus on the posterior means of the lead time parameters, which provide an indication of the timeliness of the corresponding indicators. Most leading indicators tend to lead the business cycle by about five to ten months at business cycle peaks and two to five months at troughs. Two interesting exceptions are NEW ORDERS, CAPITAL GOODS and the INTEREST RATE SPREAD. The first indicator does not seem to significantly lead or lag the cycle of the reference series at all, based on the fact that the value zero is contained in the 90% highest posterior density region of its posterior lead time density both at peaks and at troughs. In contrast, the INTEREST RATE SPREAD is the most prominent 'leader' with a lead time of about three years at peaks and one and a half years at troughs. Secondly, we inspect the posterior standard deviations of the lead times, and more in general, the shapes of the lead time distributions. A low posterior standard deviation provides some indication that the corresponding indicator is more consistent across cycles. Based on this criterium, we conclude that MONEY SUPPLY seems to be the most favorable leading indicator at the peaks, whereas BUILDING PERMITS is to be preferred for the troughs. In these two cases, over 80% of the marginal probability mass is located at the mode, which suggests that these indicators are very useful for signalling peaks and troughs, respectively. Regrettably, the often-lauded INTEREST RATE SPREAD, which is the most timely indicator according to the model, seems to have very inconsistent lead times. Its lead time distribution at troughs is more or less peaked, but its lead time distribution at peaks is close to being uniform between one and six years. The joint lead time distributions for AVERAGE WEEKLY HOURS and VENDOR PERFORMANCE seem to be bimodal, which suggests that these indicators are less useful as leading indicators as well. Judging all indicators both on their timeliness and consistency, the indicators MONEY SUPPLY is to be preferred overall, followed by BUILDING PERMITS, STOCK PRICES and CONSUMER EXPECTATIONS.

To judge the overall performance of the model, we proceed by comparing the model based lead time estimates to the lead time estimates in The Conference Board (2001, p. 35) which are based on the turning points of the indicators as determined by its dating committee. We note that The Conference Board's estimates are based on the turning points of the indicators as tabulated for the six official recessions during the period 1959 to 1999, whereas our estimates are based on data for the period January 1959 - February 2008, which includes the 2001 recession. Even though this somewhat complicates a one-to-one comparison, we believe it is worth studying the results. The posterior means and modes of the lead time distributions as well as the average lead times found by The Conference Board (2001) are presented in Table 5.3. We also include the posterior standard deviations of the lead times and the sample standard deviations of The Conference Board's lead times. Both measures provide an indication of the consistency of indicators relative to their competitors. Again, we distinguish between the lead times at peaks and at troughs. Comparing the lead time estimates at peaks, we conclude that on average The Conference Board's estimates are similar to the model based estimates, but that for the more timely indicators some large discrepancies are found. In particular, The

Conference Board estimates the average lead time of BUILDING PERMITS to be nearly 21 months but with a large standard deviation of more than 18 months, whereas we obtain an estimate of the lead time of 9 months with standard deviation of only one month. Hence, our model estimates suggest that BUILDING PERMITS is less timely but more consistent. Similarly, the INTEREST RATE SPREAD has a lead time of 26 months according to The Conference Board versus 35 months according to the model. However, both approaches suggest that this indicator is inconsistent. We do not observe any large deviations in the estimates for the lead times at troughs. It is worth noticing that according to both the model and The Conference Board, NEW ORDERS, CAPITAL GOODS does not lead the business cycle at troughs. We further mention that MONEY SUPPLY, which is the best performing indicator overall according to the model, is indeed timely according to The Conference Board, but less consistent in comparison to its close competitor STOCK PRICES.

The posterior density of τ is displayed in Figure 5.2 for the period January 1982 - December 1986. Almost all posterior mass is located in the years 1983 and 1984, and the most likely date of the structural break in the (co-)variances is March 1984. This confirms the earlier findings by McConnell and Perez-Quiros (2000), among others. The two rightmost columns of Table 5.2 show the posterior means of the variances of the indicators before and after the break point τ. We find that the volatility of all series has decreased, except for MONEY SUPPLY and CONSUMER EXPECTATIONS.

In Table 5.4 we report the posterior means of the correlations, where the upper- and lower-triangular elements refer to the posterior means of the correlations before and after the break point τ, respectively. Interestingly, before the break point, 26 pairs of indicators were correlated, as indicated by the fact that the value zero is not contained in their estimated 90% highest posterior density regions. After the break, this number is reduced to only 16 pairs. This suggests that the break does not only mark a decline in volatility, but also a decline in the strength of the co-movement between the indicators.

Finally, as a robustness check we estimate the Markov Switching panel data model including the four coincident indicator variables separately, instead of the coincident index. Given that EMPLOYMENT is often considered to be the most reliable component among the four, see Issler and Vahid (2006) among others, we choose this series as the reference series. Interestingly, the results obtained for the leading indicators and for the estimated recession probabilities are very similar. The lead times at peaks and at troughs of the three coincident indicators relative to EMPLOYMENT are all estimated to be zero with posterior probabilities larger than 99%. This reassures that the four coincident indicators indeed share the exact same cycle. As we are primarily interested in the results for the leading

Table 5.3: Comparison of the lead times as determined by The Conference Board (2001) to the model based lead time estimates

Variable	The Conference Board (2001)				Markov Switching panel data model						Difference			
	Peaks		Troughs		Peaks			Troughs			Peaks		Troughs	
	Mean	St.dev.	Mean	St.dev.	Mean	Mode	St.dev.	Mean	Mode	St.dev.	Mean	Mode	Mean	Mode
AWH	8.2	(4.8)	0.8	(1.0)	5.0	6	(2.6)	2.0	2	(1.1)	3.21	2.17	-1.1	-1.2
AIC	11.8	(6.6)	1.2	(1.0)	6.1	6	(1.6)	3.1	3	(0.8)	5.71	5.83	-2.0	-1.8
NCO	6.3	(4.7)	0.7	(0.8)	8.5	9	(1.8)	2.4	2	(1.0)	-2.15	-2.67	-1.7	-1.3
VP	4.8	(5.7)	3.5	(4.8)	5.8	2	(2.6)	2.3	2	(0.9)	-0.94	2.83	1.2	1.5
NCA	5.5	(7.4)	0.2	(2.3)	3.5	3	(3.5)	-1.0	0	(3.9)	1.97	2.50	1.2	0.2
BP	20.8	(18.6)	5.0	(5.2)	9.0	9	(1.0)	4.1	4	(0.5)	11.86	11.83	0.9	1.0
SP	8.0	(4.2)	4.2	(0.8)	8.8	8	(1.1)	5.2	5	(0.7)	-0.75	0.00	-1.0	-0.8
M2	12.5	(7.8)	1.5	(6.0)	10.1	10	(0.5)	4.2	4	(0.5)	2.45	2.50	-2.7	-2.5
IRS	26.3	(12.4)	13.2	(7.5)	34.6	28	(10.1)	17.5	17	(3.8)	-8.31	-1.67	-4.3	-3.8
CE	10.3	(6.9)	4.5	(2.4)	8.4	8	(1.4)	5.3	5	(1.1)	1.96	2.33	-0.8	-0.5
Average difference											1.50	2.57	-1.03	-0.93
Average absolute difference											3.93	3.43	1.69	1.46

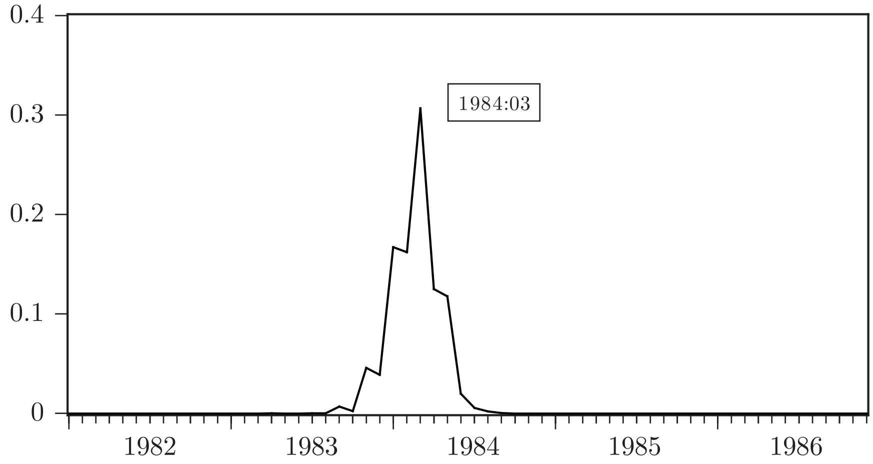
Note: The table compares the lead times of the leading indicators as estimated by The Conference Board (2001, p. 35), to the model based lead time estimates. The Conference Board's lead time estimates are based on the lead times as tabulated for the six official recessions during the period 1959 to 1999. The model based lead times are estimated over the period January 1959 - February 2008.

Table 5.4: Posterior means and standard deviations (in parentheses) of the correlations among the indicators

Variable	CCI	AWH	AIC	NCO	VP	NCA	BP	SP	M2	IRS	CE
CCI	1	0.429***	0.441***	0.576***	0.134**	0.127**	0.183***	0.026	0.095	−0.207***	−0.023
		(0.049)	(0.049)	(0.042)	(0.059)	(0.059)	(0.058)	(0.061)	(0.062)	(0.057)	(0.060)
AWH	0.201***	1	0.300***	0.247***	0.070	−0.051	0.232***	−0.008	−0.028	−0.092	0.011
	(0.062)		(0.053)	(0.055)	(0.059)	(0.058)	(0.055)	(0.059)	(0.060)	(0.058)	(0.058)
AIC	0.266***	0.095	1	0.376***	0.098*	0.172***	0.257***	0.083	0.092	−0.037	0.128**
	(0.059)	(0.062)		(0.051)	(0.058)	(0.057)	(0.054)	(0.059)	(0.059)	(0.059)	(0.057)
NCO	0.361***	−0.090	0.287***	1	0.276***	0.082	0.222***	0.082	0.121**	−0.016	0.022
	(0.057)	(0.062)	(0.058)		(0.055)	(0.058)	(0.057)	(0.059)	(0.060)	(0.058)	(0.059)
VP	0.066	0.038	−0.047	0.036	1	0.103*	0.088	0.057	0.143**	−0.002	0.124**
	(0.063)	(0.062)	(0.062)	(0.062)		(0.058)	(0.058)	(0.058)	(0.060)	(0.058)	(0.058)
NCA	0.136**	−0.009	0.096	0.191***	−0.014	1	−0.011	−0.005	−0.021	−0.110*	0.052
	(0.064)	(0.062)	(0.061)	(0.059)	(0.062)		(0.058)	(0.058)	(0.059)	(0.058)	(0.058)
BP	0.237***	0.122**	0.148**	0.133**	0.138**	−0.031	1	0.103*	0.137**	0.095*	0.084
	(0.063)	(0.062)	(0.060)	(0.061)	(0.061)	(0.062)		(0.059)	(0.058)	(0.057)	(0.058)
SP	−0.031	−0.040	0.038	0.014	−0.053	0.038	0.012	1	0.170***	0.008	0.253***
	(0.064)	(0.062)	(0.063)	(0.062)	(0.063)	(0.062)	(0.063)		(0.058)	(0.058)	(0.055)
M2	0.067	−0.076	0.022	−0.029	−0.077	0.030	0.149**	−0.024	1	0.097	0.100*
	(0.069)	(0.071)	(0.069)	(0.080)	(0.072)	(0.067)	(0.070)	(0.082)		(0.061)	(0.059)
IRS	0.058	0.075	0.041	0.048	0.020	−0.090	−0.013	−0.024	0.169**	1	0.026
	(0.065)	(0.063)	(0.064)	(0.064)	(0.064)	(0.063)	(0.064)	(0.064)	(0.075)		(0.057)
CE	0.126**	0.038	0.157***	−0.018	−0.066	0.045	0.023	0.306***	0.074	0.080	1
	(0.063)	(0.062)	(0.060)	(0.063)	(0.062)	(0.061)	(0.061)	(0.056)	(0.068)	(0.062)	

Note: The table presents posterior means and standard deviations (in parentheses) of the correlations $\sigma_{i_1,i_2}/\sqrt{\sigma_{i_1,i_1}\sigma_{i_2,i_2}}$ for any pair of indicators i and j in the Markov Switching panel data model with a single structural change in the covariance matrix as in (5.4), estimated for monthly growth rates of The Conference Board's CCI and the ten components of its CLI over the period January 1959 - February 2008. The upper- and lower-triangular elements are the correlations before and after the break point τ, respectively. The superscripts *, **, *** indicate that the value zero is not contained in the 90%, 95%, 99% HPD-region, respectively.

Figure 5.2: Posterior density of the variance break point parameter τ



Note: The graph presents the posterior density of the variance break date τ (for the period January 1982 - December 1986) in the Markov Switching panel data model with a single structural change in the covariance matrix as in (5.4), estimated for monthly growth rates of The Conference Board's CCI and the ten components of its CLI over the period January 1959 - February 2008.

indicators we only report the results of the model for the CCI and the ten components included in the CLI in the remainder of the chapter. Detailed results for the model with the four components of the CCI are available upon request.

5.4 Constructing a Synchronized Leading Index

In this section we consider different methods to construct synchronized leading indices based on the lead time distributions of the included leading indicators. We then assess whether these synchronized indices yield better in-sample predictions of the business cycle chronology as determined by the NBER as compared to The Conference Board's CLI.

Methodology

Let us first briefly describe The Conference Board's procedure to construct its CLI. Firstly, for each of the ten leading indicators a volatility measure v_i is computed as the inverse of the standard deviation of its month-to-month changes. Each component is then adjusted to equalize the volatility of the components using the so-called standardization factor $w_i = v_i / \sum_{i=1}^{N} v_i$. Thirdly, the standardized month-to-month-changes are aggregated across the components for each month. This results in the sum of the adjusted contributions, $m_t = \sum_{i=1}^{N} w_i y_{i,t}$. The levels of the index are then calculated recursively, starting from the initial value $\text{CLI}_1 = (200 + m_1)/(200 - m_1)$ for the start of the sample period by

$$\text{CLI}_t = \text{CLI}_{t-1}\frac{200 + m_t}{200 - m_t} \qquad \text{for } t = 2, \ldots, T. \tag{5.5}$$

Finally, CLI_t is rebased to average 100 in the current base month, which is May 2004.

In order to identify the changes in performance between synchronized leading indices and the non-synchronized CLI, we will use the exact same procedure as described above to weight and combine the components in this section. We thus choose to refrain from weighting procedures based on, for example, the estimated variances in the model or on model based scoring systems as in Moore and Shiskin (1967) and Diebold and Rudebusch (1989).

We consider three different ways to synchronize leading indicators according to their estimated lead time distributions. Either way, we will synchronize the indicators separately for peaks and for troughs. Firstly, it seems perhaps most natural to synchronize the peaks (troughs) of all components to the peaks (troughs) of the reference component (CCI) using the posterior distribution of $\kappa_{i,1}$ ($\kappa_{i,2}$), for $i = 2, \ldots, N$. This can easily be achieved by shifting the components back or forth in time in each simulation run z according to the obtained draws $\kappa_{i,1}^{(z)}$ and $\kappa_{i,2}^{(z)}$. As a result, we obtain the posterior distribution of synchronized growth rates for each component at each point in time. We then compute the means of these distributions to obtain the posterior synchronized indicators $\text{SLI}_{i,t,j}$ at peaks and at troughs

$$\text{SLI}_{i,t,j} = \frac{1}{Z}\sum_{z=1}^{Z} y_{i,t-\kappa_{i,j}^{(z)}} \tag{5.6}$$

where Z denotes the total number of simulations. While conceptually straightforward, a limitation of this method is that it may require values of $y_{i,t}$ outside the sample to construct indicator values at the beginning or near the end of the sample. Since the very beginning of the sample is usually not of interest but the end certainly is, we restrict our attention to the latter. Near the end of the sample, the values $y_{i,T-\kappa^*}$ are not available for negative κ^*'s, corresponding to lags. We may either predict these future values, as in McGuckin *et al.* (2007), or impose additional restrictions on the distributions of $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$. Here we consider the latter. In particular, we truncate the lead time distributions from below at zero.[2] We may then synchronize according to the truncated posterior distributions using (5.6). This implies that we assume our leading indicators to either lead or coincide with the business cycle. Thirdly and finally, we consider only using the modes of the posterior

[2] Strictly speaking, even when we truncate at zero, we may still encounter missing observations if one of the publication lags of the indicators exceed the publication lag of the reference series. In this case, one may decide to further enlarge the truncation bound.

distributions of $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$ and synchronizing the components accordingly just once. Since the modes of our lead time distributions all correspond to leads, we do not encounter the above problem in our application.

After having combined the synchronized indicators using The Conference Board's procedure, we obtain one leading index for business cycle peaks and one index for troughs, to be denoted by $\text{SCLI}_{t,1}$ and $\text{SCLI}_{t,2}$, respectively. We now consider different ways to recombine the two, so that we obtain one overall index that is easier to interpret. Again, we consider three alternatives. Firstly, we consider weighting $\text{SCLI}_{t,1}$ and $\text{SCLI}_{t,2}$, for $t = 1, \ldots, T$, using the series of posterior recession probabilities r_t

$$\text{SCLI}_t^{r_t} = r_t \text{SCLI}_{t,1} + (1 - r_t)\text{SCLI}_{t,2}. \tag{5.7}$$

Alternatively, to make a clear-cut decision at each time point between either $\text{SCLI}_{t,1}$ and $\text{SCLI}_{t,2}$, one may decide to use $\text{SCLI}_{t,1}$ if the recession probability is above 0.5 at time t or $\text{SCLI}_{t,2}$ otherwise

$$\text{SCLI}_t^{\mathbb{I}[r_t]} = \mathbb{I}\left[r_t > \frac{1}{2}\right]\text{SCLI}_{t,1} + \mathbb{I}\left[r_t \leq \frac{1}{2}\right]\text{SCLI}_{t,2}, \tag{5.8}$$

where $\mathbb{I}[\cdot]$ denotes the indicator function, taking the value one if the argument is true and zero otherwise. A disadvantage of the above two approaches is that they require an estimate of r_t to be available for all t. This is especially inconvenient for forecasting. We therefore also consider simply using the average recession probability $\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$

$$\text{SCLI}_t^{\bar{r}} = \bar{r}\text{SCLI}_{t,1} + (1 - \bar{r})\text{SCLI}_{t,2}. \tag{5.9}$$

Application

We apply our method to construct a SCLI based on the lead time distributions obtained in Section 5.3 for the model of the March 2008 vintages of The Conference Board's CCI and the ten components of its CLI. We restrict the evaluation-sample to March 1961 - July 2004 so that we do not encounter missing observations using any of the approaches outlined above. To select the best performing approach to construct the index, we evaluate the performance of the indices as regressors in a logit model. With the logit model we aim to explain the binary recession indicator as implied by the NBER turning points, thus assuming that the latter are correct. This approach is along the same lines as Issler and Vahid (2006). The model is formulated as

$$\Pr[\text{NBER}_t = 1|\Delta\text{CLI}_t^*] = F(\alpha + \beta\Delta\text{CLI}_t^*) = \frac{F(\alpha + \beta\Delta\text{CLI}_t^*)}{1 + F(\alpha + \beta\Delta\text{CLI}_t^*)}, \tag{5.10}$$

where ΔCLI_t^* denote the growth rates of a particular (synchronized) leading index CLI_t^*.

We select the best performing Logit model and its corresponding leading index by its McFadden R^2, the percentage of correctly predicted recessionary months and the overall percentage of correctly predicted states. That is, the hit rate. We also compare the performance of our index to that of The Conference Board. For this purpose, we synchronize The Conference Board's CLI in the same way as we synchronized our components. That is, we re-estimated the Markov Switching panel data model only on the CCI and CLI, and use the resultant posterior distribution of $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$. The differences in performance between the SCLI and the CLI are thus solely due to the fact that the SCLI is constructed using the individually synchronized components included in the CLI, whereas the CLI is only synchronized itself.

Our results are summarized in Table 5.5. Firstly, inspecting the differences among the three classes of indices which make use of the posterior distribution of $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$, we conclude that it is not advisable to only use the modes of the lead time distributions. The predictive performance of these indicators is much weaker. This supports the Bayesian philosophy that full parameter distributions are more informative than point estimates. Secondly, truncating the posterior lead time distributions from below at zero does tend to seriously affect the results which seems promising for applications where this is necessary. Comparing the results obtained for the different methods to combine $\text{SCLI}_{t,1}$ and $\text{SCLI}_{t,2}$, we conclude that an optimal time-invariant weight performs best, with a R^2 of 0.49 and a hit rate of 92%. The indicator $\text{SCLI}_t^{\bar{r}}$, which is based on the average recession probability, underperforms slightly.

We proceed by inspecting the results for the bivariate model of the CCI and CLI. We find that negative values of $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$ have no posterior mass. As a result, the truncated and non-truncated case provides the same result. Concerning the choice for the posterior modes of $\boldsymbol{\kappa}$ versus the complete posterior distribution, we reach the same conclusions as for the indices based on individual synchronization. This also holds for the choice of the explanatory variables. Comparing the performance of the SCLI's to the performance of the CLI, we conclude that the SCLI's outperform the CLI in all relevant cases. This suggests that it is beneficial to synchronize leading indicators before aggregation.

5.5 Conclusions

In this chapter we have evaluated the timeliness and consistency of the ten leading indicators included in The Conference Board's CLI. For this purpose, we developed a Markov switching model which allows to simultaneously estimate the individual lead times of a

Table 5.5: Performance of the Synchronized Leading Index

$\boldsymbol{\kappa}$'s used	Explanatory variable(s)	Individual Leading Indicators			Composite Leading Index		
		R^2	% 1's corr.	% corr.	R^2	% 1's corr.	% corr.
Posterior distributions	$\text{SCLI}_{t,1}$, $\text{SCLI}_{t,2}$	**0.49**	**57.75**	**92.32**	**0.44**	**50.00**	**89.44**
	$\text{SCLI}_{t,1}$	0.36	39.74	88.95	0.27	35.37	87.91
	$\text{SCLI}_{t,2}$	0.24	22.54	87.01	0.29	28.05	85.41
	$\text{SCLI}_t^{r_t}$	0.30	23.94	88.19	0.26	34.15	87.72
	$\text{SCLI}_t^{\mathbb{I}[r_t]}$	0.23	21.13	88.02	0.14	8.54	83.69
	$\text{SCLI}_t^{\bar{r}}$	0.46	53.52	91.14	0.40	43.90	88.48
Posterior modes	$\text{SCLI}_{t,1}$, $\text{SCLI}_{t,2}$	*0.22*	*21.95*	*85.80*	*0.25*	*23.17*	85.41
	$\text{SCLI}_{t,1}$	0.15	13.41	85.03	0.10	6.10	84.84
	$\text{SCLI}_{t,2}$	0.12	9.76	84.45	0.18	13.41	83.69
	$\text{SCLI}_t^{r_t}$	0.12	10.98	85.22	0.14	15.85	*85.99*
	$\text{SCLI}_t^{\mathbb{I}[r_t]}$	0.07	0.00	83.11	0.08	6.10	84.07
	$\text{SCLI}_t^{\bar{r}}$	0.20	20.73	85.41	0.19	19.51	85.03
Truncated posterior distributions	$\text{SCLI}_{t,1}$, $\text{SCLI}_{t,2}$	*0.47*	*54.88*	*90.60*	**0.44**	**50.00**	**89.44**
	$\text{SCLI}_{t,1}$	0.36	40.24	88.48	0.27	35.37	87.91
	$\text{SCLI}_{t,2}$	0.24	23.17	85.60	0.29	28.05	85.41
	$\text{SCLI}_t^{r_t}$	0.20	18.29	85.80	0.26	34.15	87.72
	$\text{SCLI}_t^{\mathbb{I}[r_t]}$	0.09	0.00	82.73	0.14	8.54	83.69
	$\text{SCLI}_t^{\bar{r}}$	0.45	52.44	90.40	0.40	43.90	88.48

Note: The numbers in italics indicate the best performing methods among the set of methods with the same usage of the lead time distributions. The bold-faced numbers indicate the best performing methods overall.

large panel of leading indicator variables. The model relates the turning points of the indicators to the turning points of a reference series, where it is assumed that the cycle of the reference series coincides with the business cycle. As in Chapter 4, we allowed the indicators to have different lead times at peaks and at troughs.

Our empirical results for the period January 1959 - February 2008 suggest that the indicators BUILDING PERMITS, STOCK PRICES, MONEY SUPPLY, and CONSUMER EXPECTATIONS are the most consistent indicators among the ten, having lead times of nine to ten months at peaks and four to five months at troughs with standard deviations of about one month. Interestingly, these indicators also rank among the most timely indicators in the panel. This contradicts the often-heard statement that timely leading indicators are usually not very consistent and vice versa. One indicator that does strongly confirm this statement, however, is the INTEREST RATE SPREAD. On average, the lead time of this indicator is nearly three years at peaks and one and a half year at peaks, but its lead time distributions have exceptionally fat tails.

The fact that none of the individual indicators seem to be as timely and consistent as the CLI validates the use of composite leading indices and the ongoing search for better indices. To contribute to this search we examined whether it is useful to synchronize leading indicators according to their estimated lead times before aggregation. In an application to the ten components of the CLI, we found that synchronized indices indeed outperform the CLI in the sense that they yield better in-sample predictions of the business cycle chronology as determined by the NBER.

Further research

We conclude by offering several avenues for further research. The first is to consider more flexible model specifications. We mention four possibilities here. Firstly, as advocated in Diebold *et al.* (1994), Filardo (1994), and Diebold and Rudebusch (1996) it might be interesting to allow the transition probabilities p and q of the Markov process to vary over time. For example, one may wish to relate p and q to observed explanatory variables. Secondly, since growth volatility is often found to be high during recessions, it would be interesting to examine whether this can be exploited to identify the regimes. One may, for example, allow the variances in the model to be regime-dependent alongside the growth rates. Thirdly, to capture the persistence in the demeaned growth, the model can be extended by allowing for autoregressive dynamics. Finally, the lead time estimates of individual indicators may benefit from an indicators specific break point parameter τ in the model.

A second avenue would be to use the model for indicator selection. For example, indicators can be selected by means of a formal scoring system as in Moore and Shiskin (1967) and Diebold and Rudebusch (1989), where the scores of the indicators are based on their estimated risk (inconsistency) versus return (timeliness). The scores could also serve as indicator weights in the construction of new composite indices. In the same vein it seems promising to develop a formal test for business cycle independence of leading indicators based on their estimated consistency.

Finally, it would be interesting to employ the model in other situations where variables are found to lead or to lag a common cycle. For example, the model can be used to estimate the lead/lag times of the cycles of the smaller economies within the euro area relative to the German cycle. Another interesting opportunity would be to assess whether switches between bull and bear regimes in different financial markets can be predicted on the basis of their interrelationship.

5.A Estimation and Inference

In Section 5.A.1 we specify the likelihood function of the model. Sections 5.A.2 and 5.A.3 discuss the differences in prior specification and posterior simulation as compared to the approach in Chapter 4.

5.A.1 The likelihood function

For ease of exposition, we write the model in matrix format as

$$\boldsymbol{Y}_t = \boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t} + \boldsymbol{\mathcal{E}}_t, \qquad \boldsymbol{\mathcal{E}}_t \sim N(0, \boldsymbol{\Sigma}_t), \tag{5.A.1}$$

where $\boldsymbol{Y}_t = (y_{1,t}, \ldots, y_{N,t})'$ is the vector of growth rates of all N variables in month t, $\boldsymbol{\mathcal{E}}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})'$, $\boldsymbol{\mathcal{S}}_t = (s_{1,t}, \ldots, s_{N,t})'$, $\boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t} = (\mu_{s_{1,t}}, \ldots, \mu_{s_{N,t}})'$. The $(N \times N)$ covariance matrix $\boldsymbol{\Sigma}_t$ is specified as $\boldsymbol{\Sigma}_t = \boldsymbol{\Omega}_0 \mathbb{I}[t < \tau] + \boldsymbol{\Omega}_1 \mathbb{I}[t \geq \tau]$, where $\boldsymbol{\Omega}_0 = \{\sigma_{i_1,i_2,0}\}_{i_1,i_2=1}^N$ and $\boldsymbol{\Omega}_1 = \{\sigma_{i_1,i_2,1}\}_{i_1,i_2=1}^N$ contain the (co-)variances before and after the structural break at $t = \tau$, respectively. Finally, for notational convenience, we define $\boldsymbol{\mathcal{M}}_{\boldsymbol{\mathcal{S}}_t} = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_t$, such that the model can be written as

$$\boldsymbol{Y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \odot \boldsymbol{\mathcal{S}}_t + \boldsymbol{\mathcal{E}}_t, \qquad \boldsymbol{\mathcal{E}}_t \sim N(0, \boldsymbol{\Sigma}_t). \tag{5.A.2}$$

The conditional density of $\boldsymbol{Y}_t$ for the model given the current state $\boldsymbol{\mathcal{S}}_t$ is given by

$$f(\boldsymbol{Y}_t | \boldsymbol{\mathcal{S}}_t, \boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \tau) = \frac{1}{(\sqrt{2\pi})^N} |\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2} \boldsymbol{\mathcal{E}}_t' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mathcal{E}}_t \Big), \tag{5.A.3}$$

where $\boldsymbol{\mathcal{E}}_t$ follows from (5.A.1). Hence the complete data likelihood function for model (5.A.1) equals

$$\begin{aligned} \mathcal{L}(\boldsymbol{Y}^T, \boldsymbol{\mathcal{S}}^T | \boldsymbol{\theta}) = {} & p^{\boldsymbol{\mathcal{N}}_{0,0}} (1-p)^{\boldsymbol{\mathcal{N}}_{0,1}} q^{\boldsymbol{\mathcal{N}}_{1,1}} (1-q)^{\boldsymbol{\mathcal{N}}_{1,0}} \\ & \times \prod_{t=1}^{T} f(\boldsymbol{Y}_t | \boldsymbol{\mathcal{S}}_t, \boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \tau), \end{aligned} \tag{5.A.4}$$

with $\boldsymbol{\mathcal{S}}^T = \{\boldsymbol{\mathcal{S}}_1, \ldots, \boldsymbol{\mathcal{S}}_T\}$, $\boldsymbol{Y}^T = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T\}$, $\boldsymbol{\theta} = \{\boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \tau, p, q\}$, and where $\boldsymbol{\mathcal{N}}_{s_1^*, s_2^*}$ denotes the number of transitions from state s_1^* to state s_2^*. The unconditional likelihood function $\mathcal{L}(\boldsymbol{Y}^T | \boldsymbol{\theta})$ can be obtained by summing over all possible realizations of $\boldsymbol{s}_1$

$$\mathcal{L}(\boldsymbol{Y}^T | \boldsymbol{\theta}) = \sum_{s_{1,1}=0}^{1} \sum_{s_{1,2}=0}^{1} \cdots \sum_{s_{1,T}=0}^{1} \mathcal{L}(\boldsymbol{Y}^T, \boldsymbol{\mathcal{S}}^T | \boldsymbol{\theta}). \tag{5.A.5}$$

5.A.2 Prior specification

We adopt the same prior specification as in Section 4.3.2 for the growth parameters $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$, the transition parameters p and q, the covariance matrices $\mathbf{\Omega}_0$ and $\mathbf{\Omega}_1$, and the breakpoint parameter τ. For the two vectors of lead time parameters $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$, however, our prior specification is necessarily slightly more restrictive. We take the joint discrete uniform priors given by

$$f(\kappa_{i,1}, \kappa_{i,2}|s_{1,t}) \propto \begin{cases} 1 & \text{if } (\kappa_{i,1}, \kappa_{i,2}) \in K \\ 0 & \text{elsewhere} \end{cases} \tag{5.A.6}$$

with

$$K = \{(\kappa_{i,1}, \kappa_{i,2}) \in \mathbb{Z} | l_{i,1} \leq \kappa_{i,1} \leq u_{i,1}, \quad l_{i,2} \leq \kappa_{i,2} \leq u_{i,2}, \quad \kappa_{i,2} \leq \kappa_{i,1} + \max L[\boldsymbol{s}_1]\} \tag{5.A.7}$$

for $i = 2, \ldots, N$, where the function $\max L[\cdot]$ returns the length in months of the longest recession (or string of ones) in the given state vector. Hence, we define indicator-specific maxima and minima on the lead/lag time at peaks and at troughs, denoted by $l_{i,j}$ and $u_{i,j}$, respectively. To understand that the third condition on the set K is necessary, consider the boundary case where $\kappa_{i,2} = \kappa_{i,1} + \max L[\boldsymbol{s}_1]$ for any i. This case implies that the cycle of the reference series $\boldsymbol{s}_1$ converts into the cycle of series i, $\boldsymbol{s}_i$, by shortening all recessions in the reference series by the length of its longest recession. The resultant state vector $\boldsymbol{s}_i$ would then be zero, for $t = 1, \ldots, T$, which is a feasible model outcome. As we cross the boundary, however, any $\kappa_{i,2} < \kappa_{i,1} - \max L[\boldsymbol{s}_1]$ yields the same state vector $\boldsymbol{s}_i$ and the same likelihood. This induces an identification problem.

5.A.3 Posterior simulation

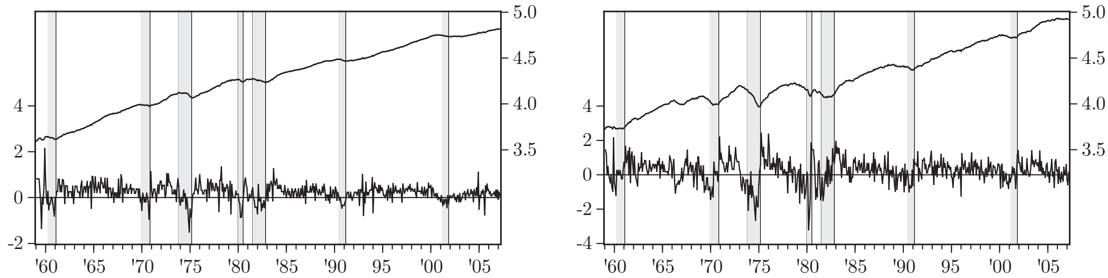
To obtain posterior results for the model parameters of the Markov Switching panel data model (5.1)–(5.3), we use the Gibbs sampling algorithm of Geman and Geman (1984) together with the data augmentation method of Tanner and Wong (1987). The unobserved state variables $\{\boldsymbol{\mathcal{S}}_t\}_{t=1}^T$ are simulated alongside the model parameters $\boldsymbol{\theta}$, see Albert and Chib (1993), McCulloch and Tsay (1994), Chib (1996), and Kim and Nelson (1999), among others.

The full conditional posterior distributions of the state variables $\{\boldsymbol{\mathcal{S}}_t\}_{t=1}^T$, the growth parameters $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$, the transition parameters p and q, the covariance matrices $\mathbf{\Omega}_0$ and $\mathbf{\Omega}_1$, and the breakpoint parameter τ are provided in Section 4.A.

The lead time parameters $\kappa_{i,1}$ and $\kappa_{i,2}$ are sampled for each series i in the panel $\boldsymbol{Y}$ separately. Hence, we sample the pair $\kappa_{i,1}$ and $\kappa_{i,2}$ from its joint full conditional distribution for $i = 2, \ldots, N$.

5.B Additional Figures

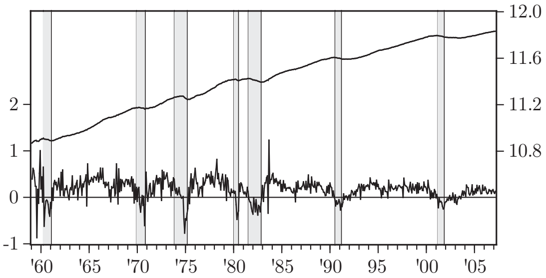
Figure 5.3: Time series plots of The Conference Board's Composite Coincident Index and the Composite Leading Index



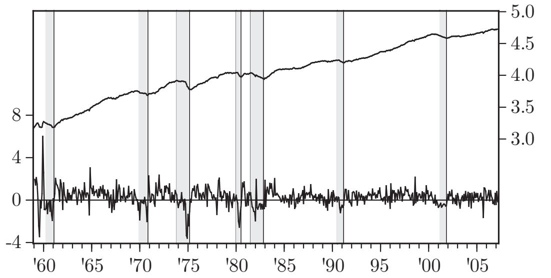
(a) Composite Coincident Index (b) Composite Leading Index

Note: The graphs present the logarithmic levels on the top (right axis) and the month-to-month growth rates on the bottom (left axis) of the series as indicated, measured over the period January 1959 - February 2008. The shaded areas are recession periods based upon the NBER turning points.

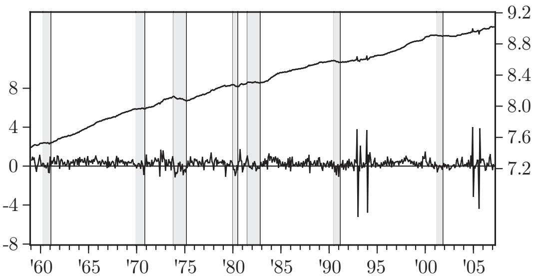
Figure 5.4: Time series plots of the components of The Conference Board's Composite Coincident Index



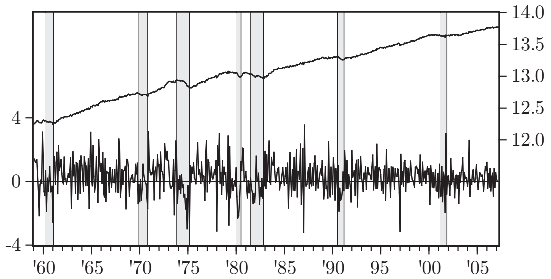
(a) Employment



(b) Industrial Production



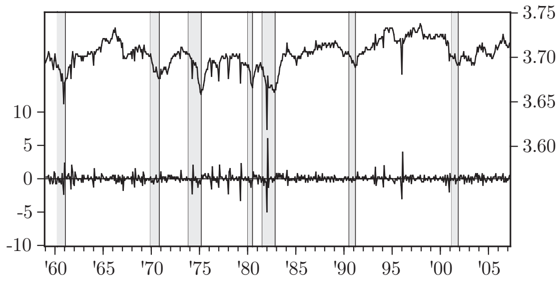
(c) Personal income less transfer payments



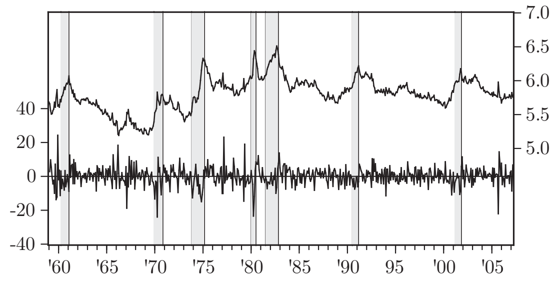
(d) Manufacturing and trade sales

Note: The graphs present the logarithmic levels on the top (right axis) and the month-to-month growth rates on the bottom (left axis) of the window, measured over the period January 1959 - February 2008. The shaded areas are recession periods based upon the NBER turning points.

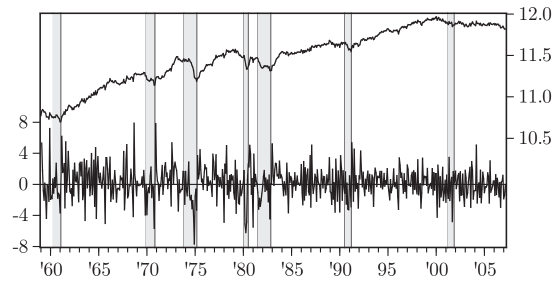
Figure 5.5: Time series plots of the components of The Conference Board's Composite Leading Index



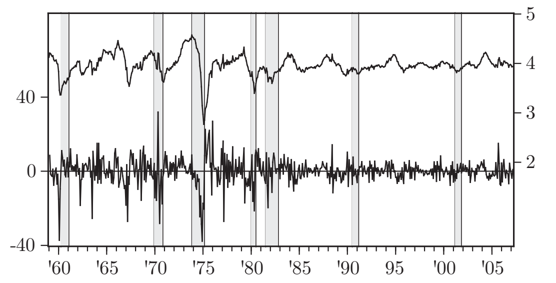
(a) Average weekly hours



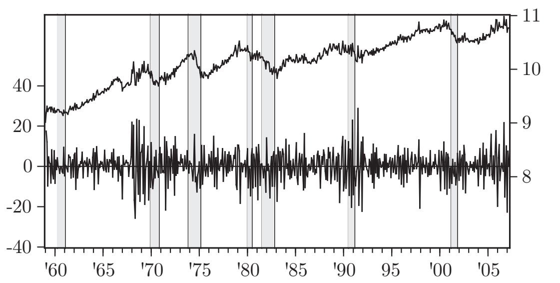
(b) Average initial claims



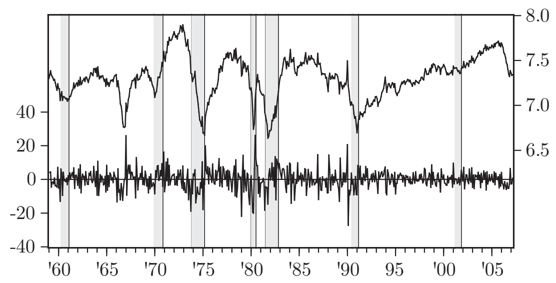
(c) New orders, consumer goods



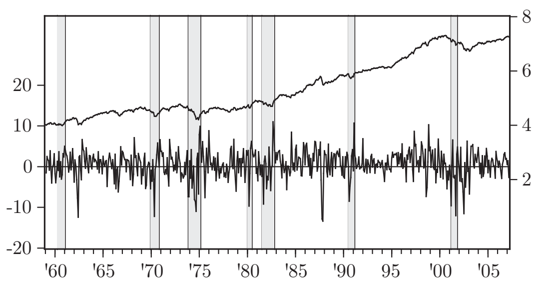
(d) Vendor performance



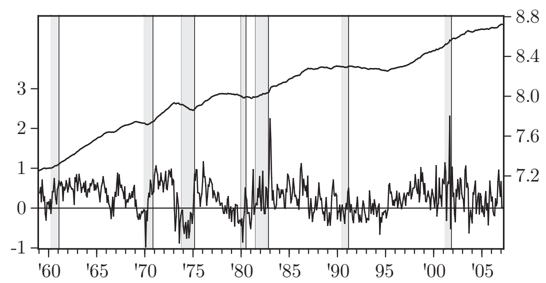
(e) New orders, capital goods



(f) Building permits

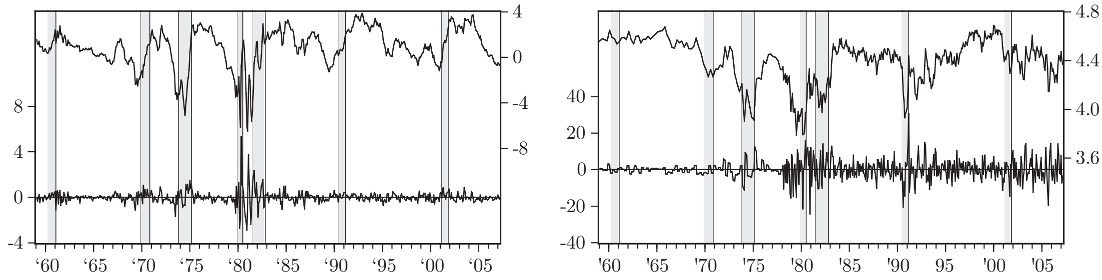


(g) Stock prices



(h) Money supply

Figure 5.5 (*continued*)

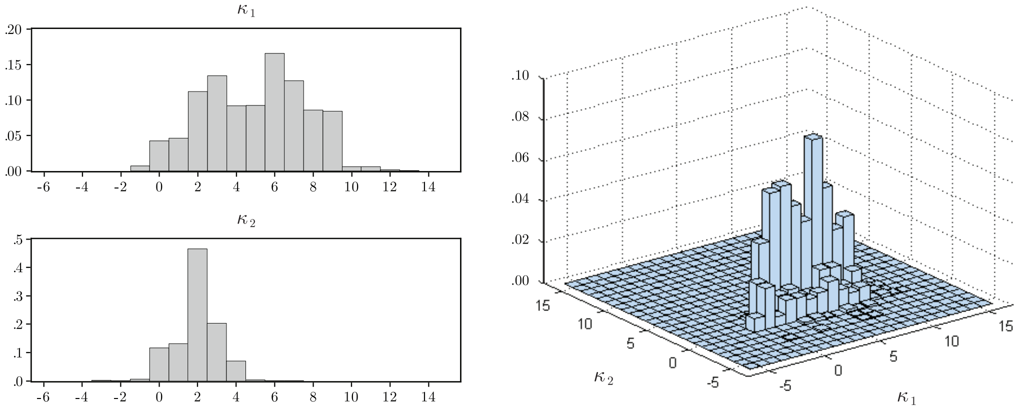


(i) Interest rate spread[1]

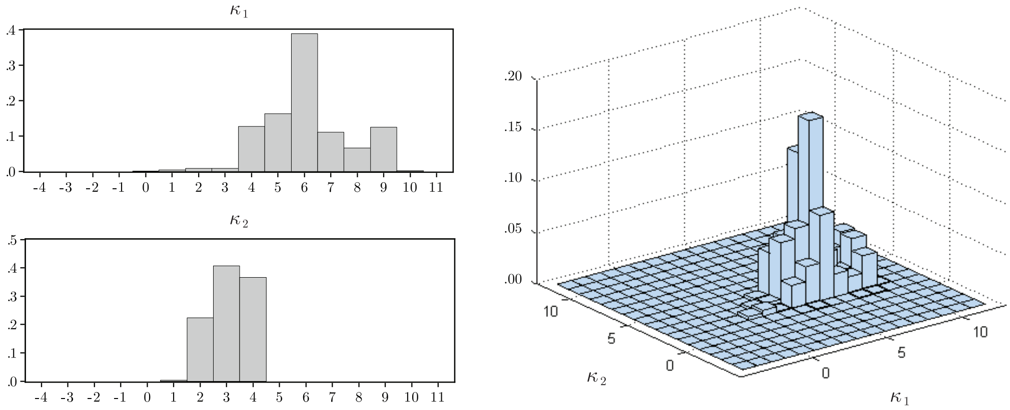
(j) Consumer expectations

Note: The graphs present the logarithmic levels on the top (right axis) and the month-to-month growth rates on the bottom (left axis) of the window, measured over the period January 1959 - February 2008. The shaded areas are recession periods based upon the NBER turning points. [1] For this series, we show the levels instead of the logarithmic levels.

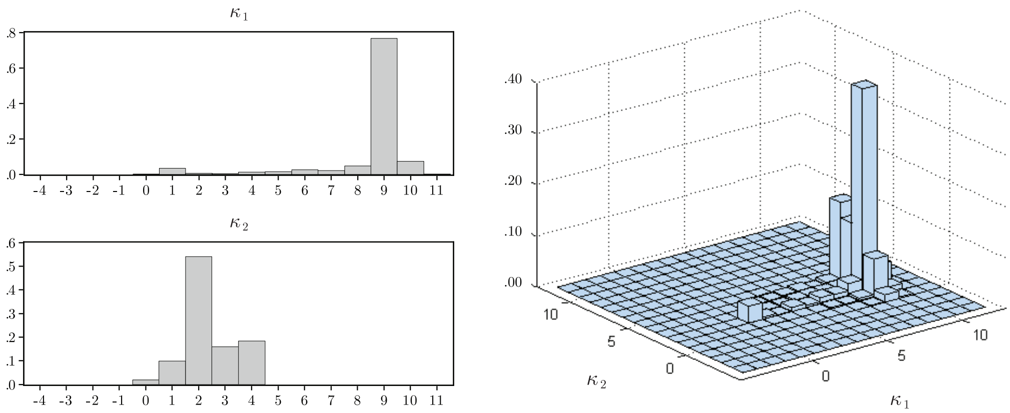
Figure 5.6: Posterior densities of the lead times $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$



(a) Average weekly hours

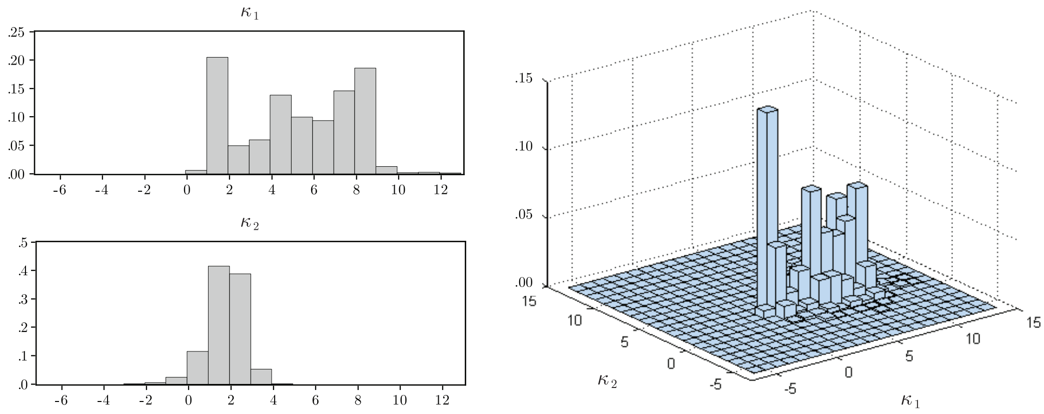


(b) Average initial claims

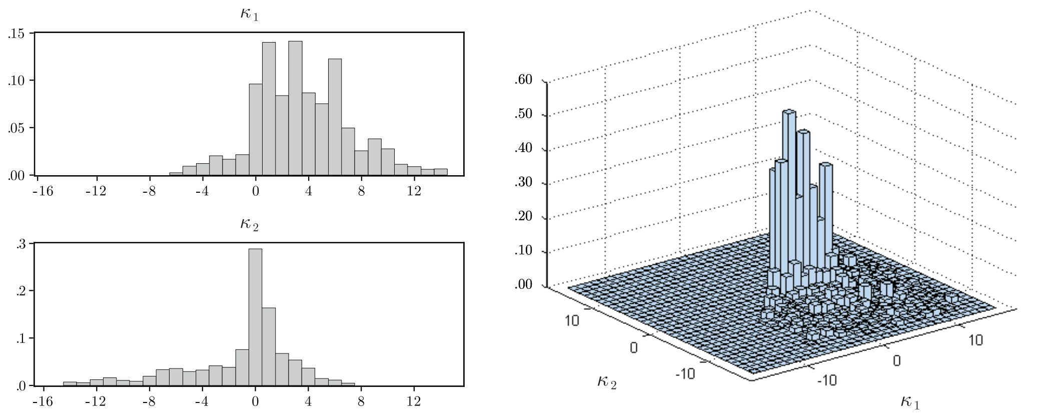


(c) New orders, consumer goods

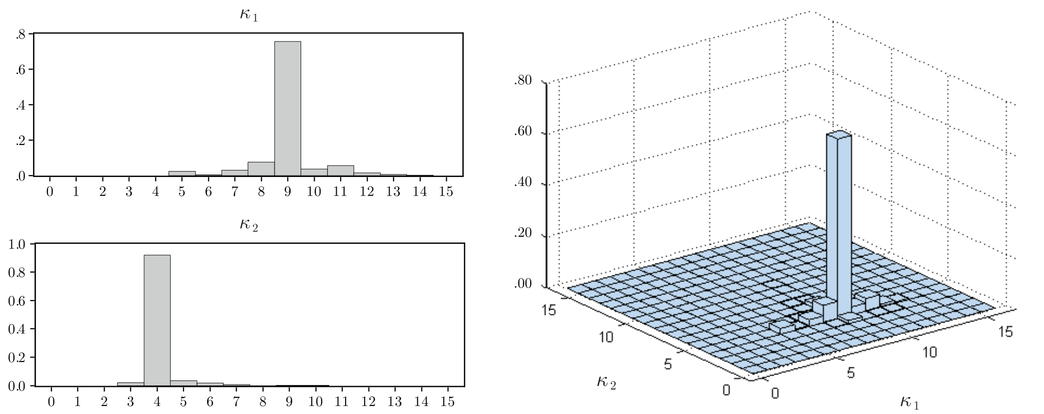
Figure 5.6 (*continued*)



(d) Vendor performance

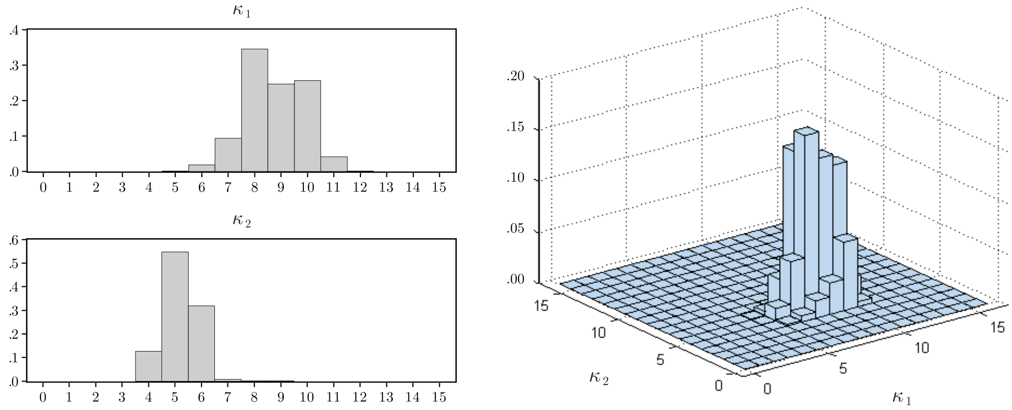


(e) New orders, capital goods

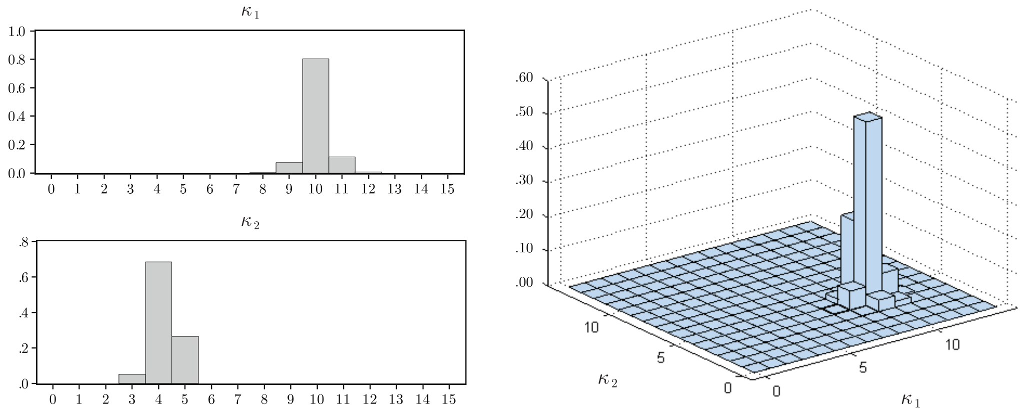


(f) Building permits

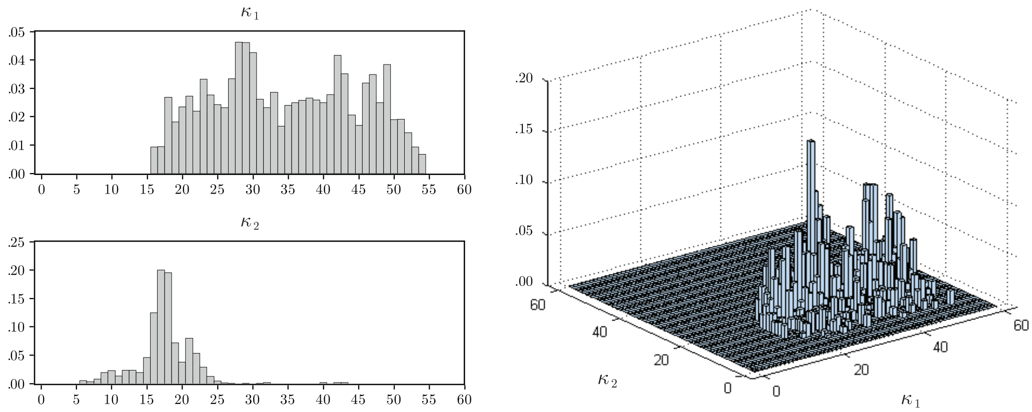
Figure 5.6 (*continued*)



(g) Stock prices

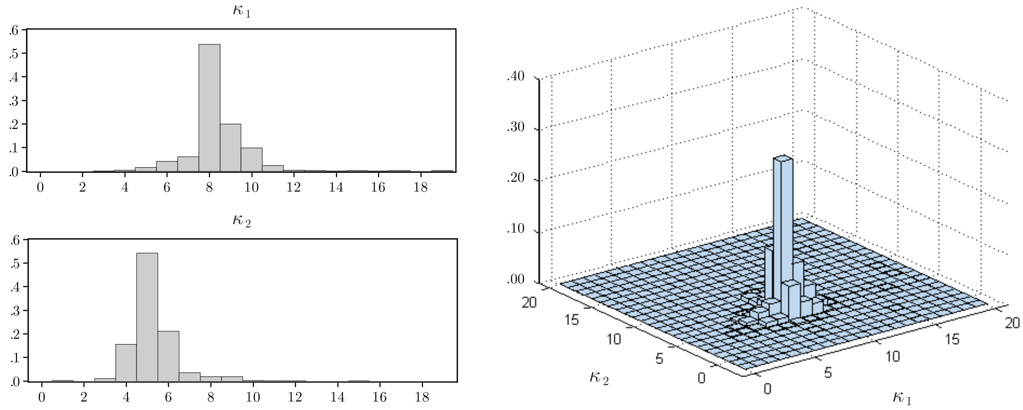


(h) Money supply



(i) Interest rate spread

Figure 5.6 (*continued*)



(j) Consumer expectations

Note: The graphs present the individual and joint posterior densities of the lead/lag parameters κ_1 and κ_2 in the Markov Switching panel data model with a single structural change in the covariance matrix as in (5.4), estimated for monthly growth rates of The Conference Board's CCI and the ten components of its CLI over the period January 1959 - February 2008.

Nederlandse Samenvatting (Summary in Dutch)

Monitoren is het systematisch volgen van een proces, door regelmatig informatie over de toestand van het proces te verzamelen, deze informatie te analyseren en te evalueren. Wellicht onbewust monitort ieder mens meerdere processen gelijktijdig, zij het niet beroepsmatig, dan wel privé. Zo monitoren ouders de ontwikkelingen van hun kinderen, doktoren de toestand van hun patiënten, investeerders het rendement op hun investeringen en marketeers de wensen van consumenten. Monitoring is van doorslaggevend belang om de juiste beslissingen te nemen op het juiste moment. Vooral vanuit de beroepspraktijk is er daarom altijd behoefte aan verbeterde monitoringsmethoden. Door gebruik te maken van moderne computersystemen en geavanceerde analytische technieken, kunnen monitoringssystemen tegenwoordig grote hoeveelheden informatie verwerken en zijn ze, indien nodig, volledig geautomatiseerd. Veel van deze technieken vinden hun oorsprong in de wetenschap. In de afgelopen vier decennia omarmde de beroepspraktijk inzichten uit diverse vakgebieden zoals de statistiek, de psychometrie en de econometrie.

Wetenschappers die zich met de ontwikkeling van monitoringsmethoden bezighouden staan constant voor nieuwe uitdagingen. Dit proefschrift levert een bijdrage aan de ontwikkeling van nieuwe monitoringsmethoden door oplossingen te bieden voor enkele van deze uitdagingen. De bestudeerde uitdagingen hebben zowel betrekking op dataverzameling, data-analyse, als op systeemevaluatie. Om de uitdagingen en de oplossingen te illustreren maken we gebruik van twee toepassingen. De eerste toepassing is de monitoring van consumentenvertrouwen. Bij deze toepassing blijkt met name de dataverzameling uitdagend te zijn. De tweede toepassing is de monitoring van de conjuncturele toestand van de economie. Bij deze toepassing richten we ons op nieuwe data-analysemethoden en evaluatietechnieken. In het onderstaande bespreken we de deelgebieden dataverzameling, data-analyse en systeemevaluatie eerst in algemene zin, waarna we ingaan op de in het proefschrift bestudeerde uitdagingen.

Dataverzameling

In de dataverzamelingsfase worden de meest recente gegevens, ofwel data, over de toestand van het te monitoren proces waargenomen en geregistreerd. Idealiter zijn deze data op ieder moment direct beschikbaar. In dit geval kan de gebruiker het onderwerp continu monitoren en onmiddellijk reageren wanneer dat nodig is. In de praktijk is het echter vaak onmogelijk, dan wel zeer kostbaar of onwenselijk, om continu data te verzamelen. Zo is het bijvoorbeeld zeer kostbaar om een brug voor het verkeer af te sluiten zodat de conditie van de brug gemeten kan worden. Als gevolg van deze belemmering worden conditiemetingen slechts zelden uitgevoerd. Het frequent verzamelen van data is onwenselijk als het gaat om het ondervragen van individuen. Naarmate een individu vaker ondervraagd wordt over hetzelfde onderwerp, is de kans groot dat de verzamelde data beïnvloed worden doordat de ondervraagde het gevoel krijgt nauwlettend gevolgd te worden. Men kan zich goed voorstellen dat dit gevoel de kwaliteit van de verzamelde data niet ten goede komt.

We kunnen dus stellen dat het aantal momenten waarop data verzameld kunnen worden vaak beperkt is. In deze gevallen is het belangrijk dat deze momenten zorgvuldig geselecteerd worden. In veel gevallen wordt een optimale selectie bereikt wanneer de momenten gelijkmatig over de tijd verspreid worden. In dit proefschrift beargumenteren we dat een gelijkmatige verdeling van dataverzamelingsmomenten echter niet altijd de voorkeur heeft. We denken hierbij aan de zojuist geschetste situatie waarbij individuen herhaaldelijk ondervraagd worden. In dit geval blijkt dat het beter is om de momenten willekeurig over de tijdsperiode van het onderzoek te verspreiden. Aan de hand van twee experimenten laten we zien dat deze dataverzamelingsmethode resulteert in een hogere participatie in het onderzoek. Bovendien blijkt dat de op deze manier verzamelde data in mindere mate beïnvloed zijn door de opzet van het onderzoek. Een belangrijke randvoorwaarde daarbij is dat de respondenten niet vooraf geïnformeerd worden over de momenten waarop ze ondervraagd worden. Deze voorwaarde waarborgt dat respondenten geen uitgesproken verwachtingen zullen hebben over het tijdstip van het volgende ondervragingsmoment. Onze resultaten zijn zeer bruikbaar in situaties waar het onmogelijk is om individuen te monitoren op basis van feitelijke data over hun gedrag. Voorbeelden van dergelijke situaties zijn de monitoring van de gemoedstoestand van psychiatrische patiënten en de monitoring van de mate waarin consumenten bereid zijn over te stappen naar concurrerende producten of diensten.

We demonstreren de voordelen van onze dataverzamelingsmethode door wekelijks het consumentenvertrouwen in Nederland te monitoren. Het consumentenvertrouwen geeft het oordeel weer van consumenten over de wijze waarop de economie als geheel en de eigen financiële situatie zich ontwikkelt. Uit studies is gebleken dat deze indicator bruikbaar is voor het voorspellen van de huidige en toekomstige conjuncturele toestand van de economie.

Data-analyse

Nadat de dataverzamelingsfase is afgerond, is de volgende uitdaging om deze data te analyseren. Dit komt neer op het vertalen van de data in heldere inzichten. Over het algemeen worden deze inzichten onderbouwd met statistische schattingen, zoals voorspellingen van het toekomstig verloop van een proces, of schattingen van de samenhang tussen verschillende variabelen. Om deze schattingen te verkrijgen worden vaak econometrische modellen gebruikt.

Binnen het gebied van data-analyse bespreken we drie onderzoeksuitdagingen. De eerste twee uitdagingen komen direct voort uit databeperkingen. Ten eerste gebeurt het vaak dat een substantieel deel van de data als gevolg van de gebruikte dataverzamelingsmethode ontbreekt. Denk hierbij bijvoorbeeld aan de dataverzamelingsmethode die zojuist besproken is. Op de momenten dat een individu uit voorzorg niet geïnterviewd worden, is het vanzelfsprekend dat alle data over hem of haar ontbreekt. Op de momenten dat een individu wel geïnterviewd worden, ontbreken er ook data wanneer hij of zij weigert om sommige vragen te beantwoorden. Over het algemeen is het niet verstandig om deze ontbrekende data te negeren in de analysefase. Het is beter om de ontbrekende data te vervangen door voorspelde data, en daarbij rekening te houden met de onzekerheid in deze voorspellingen. Wanneer we onze dataverzamelingsmethode toepassen op de monitoring van het consumentenvertrouwen, maken we van dergelijke technieken gebruik en bestuderen we wat daarvan de meerwaarde is.

Een tweede uitdaging is dat de verzamelde data vaak niet direct beschikbaar zijn voor analyse. Macro-economische tijdreeksen worden bijvoorbeeld gepubliceerd met een vertraging van een of meerdere maanden. Dit impliceert dat niet alleen de toekomstige, maar ook de huidige toestand van het proces voorspeld moet worden. Afgezien hiervan zijn de gepubliceerde reeksen vaak nog niet definitief, maar zijn het eerste inschattingen. Naarmate de tijd vordert en nieuwe data beschikbaar komt, worden de reeksen herhaaldelijk gereviseerd. Het gevolg hiervan is dat de voorspellingen die op eerste inschattingen gebaseerd zijn, onnauwkeurig of zelfs onzuiver kunnen zijn. Wanneer men dergelijke reeksen wil monitoren is het belangrijk om met dit aspect rekening te houden.

De derde en laatste uitdaging die we bespreken betreft de samenhang tussen te monitoren variabelen. Veranderingen in een variabele blijken vaak gecorreleerd te zijn met, of zelfs veroorzaakt te worden door, veranderingen in andere variabelen. Over het algemeen veranderen variabelen echter niet gelijktijdig, maar loopt de ene variabele voor op de andere. Uit onderzoek blijkt bijvoorbeeld dat een scherpe daling in het consumentenvertrouwen vaak een signaal is dat de economie zes maanden later zal dalen. In andere woorden, de indicator consumentenvertrouwen loopt zes maanden voor op de economie. Door de veranderingen in dergelijke voorlopende indicatoren nauwlettend in de gaten te houden, kunnen we de toekomstige conjuncturele toestand van de economie voorspellen. Het is echter zeer onrealistisch om aan te nemen dat de voorlooptijden van deze indicatoren niet variëren al naar gelang de toestand van de economie. Vaak blijkt het dat voorlopende indicatoren verder voorlopen op de economie ten tijden van pieken in de economische cyclus dan ten tijde van dalen. In het geval van consumentenvertrouwen kan dit verklaard worden door de theorie dat mensen het voorkomen van verliezen verkiezen boven het behalen van winsten. Wanneer consumenten het voorgevoel hebben dat de economie in een recessie zou kunnen raken, slaat optimisme over de economie meestal onmiddellijk om in pessimisme. Wanneer de economie dan inderdaad in een recessie raakt komt dit slechte nieuws in ieder geval niet onverwacht. Ten tijde van een recessie echter heerst er niet snel weer optimisme, ook al zien de vooruitzichten er goed uit. De reden hiervoor is dat men niet graag het risico wil nemen dat de recessie toch nog langer duurt en men meer verliezen moet accepteren. In dit proefschrift ontwikkelen we een wiskundig model waarmee onderzocht kan worden of er inderdaad een verschil bestaat in de voorlooptijd van indicatoren ten tijde van pieken en dalen. Met behulp van data van de Amerikaanse economie laten we zien dat de voorspellingen van zowel de conjuncturele toestand als van de economische groei verbeterd kunnen worden door dit model toe te passen.

Systeemevaluatie

De evaluatiefase komt neer op het monitoren van de prestaties van het monitorsysteem. Hierbij kijkt men kritisch of de prestaties van het systeem betrouwbaar zijn gebleken over een langer tijdbestek en of de verzamelde data daadwerkelijk voldoende informatie bieden voor de toepassing die het systeem heeft. Vaak komt men tot de conclusie dat de dataverzamelingsmethode en de analysetechnieken nog voor verbetering vatbaar zijn. Over het algemeen zijn de kwaliteitscriteria van een monitorsysteem zeer toepassingsgebonden. In deze sectie richten we ons daarom strikt op de twee toepassingen die in dit proefschrift aan de orde komen.

Het voornaamste doel van het monitoren van consumentenvertrouwen is het tijdig identificeren van veranderingen in dit vertrouwen. Maandelijks stelt men zich de vraag: “Hoeveel mensen hebben in de afgelopen maand hun oordeel over de economie bijgesteld, in positieve dan wel negatieve richting?” In onze evaluatie van de methoden die gebruikt worden om het consumentenvertrouwen in Nederland te monitoren, komen we tot de conclusie dat deze vraag simpelweg niet te beantwoorden is op basis van de door het Centraal Bureau voor de Statistiek (CBS) verzamelde data. Iedere maand ondervraagt het CBS ongeveer duizend Nederlandse consumenten en berekent zij een indicator voor het consumentenvertrouwen op basis van de enquêteresultaten. Dit betreft echter telkens duizend andere consumenten. Als gevolg hiervan kunnen veranderingen in de indicator enerzijds het gevolg zijn van veranderingen in consumentenvertrouwen, zoals men in de media graag stelt, maar ook het gevolg zijn van veranderingen in de steekproefsamenstelling. Dit is zeer verwarrend. Door onze nieuwe dataverzamelingsmethode toe te passen, waarbij dezelfde consumenten herhaaldelijk ondervraagd worden, kan deze verwarring voorkomen worden.

Bij de evaluatie van monitorsystemen voor de conjuncturele toestand van de economie beschouwen we twee evaluatiecriteria. Ten eerste moet een monitor de omslagpunten in de conjunctuur nauwkeurig kunnen dateren. Ten tweede moet het systeem omslagpunten zo snel mogelijk detecteren, zonder valse signalen af te geven. Om omslagpunten zo nauwkeurig mogelijk te dateren blijkt het verstandig te zijn om lang te wachten met het geven van het oordeel, zodat men zeker weet dat men de omslagpunten niet baseert op schommelingen in de economische groei op de zeer korte termijn. Het liefste wacht men zelfs tot ongeveer één à twee jaar na het omslagpunt alvorens dit te dateren. Anders gezegd wordt precisie bereikt door zeer geduldig te zijn. Beleidsmakers en ondernemers verwachten echter al veel eerder een oordeel over de conjuncturele toestand. De uitdaging is dan ook om een balans te vinden tussen precisie en detectiesnelheid. Om economische groeicijfers te vertalen in voorspellingen van omslagpunten, gebruiken we in dit proefschrift een systeem dat op ieder tijdstip de kans op een recessie bepaalt. Door deze kansen te visualiseren kan snel bekeken worden op welke momenten het systeem een recessie voor het eerst aan zag komen en of dit oordeel altijd correct en nauwkeurig is geweest. Uit deze resultaten blijkt dat het systeem dat rekening houdt met de asymmetrie in de voorlooptijden van conjunctuurindicatoren beter presteert dan concurrerende systemen.

Conclusies

Monitoring is van cruciaal belang om onderbouwde beslissingen te nemen op de juiste momenten in verscheidene vakgebieden zoals politiek, gezondheidszorg, logistiek, financiering en marketing. Dit proefschrift draagt bij aan de ontwikkeling van monitoringsmethoden, waarbij er in het bijzonder gekeken wordt naar monitoringssystemen voor opinies en voor de conjuncturele toestand van economieën. Ter verbetering van monitorsystemen voor opinies ontwikkelen we een nieuwe dataverzamelingsmethode waarbij de ondervragingsmomenten willekeurig over de onderzoeksperiode verspreid worden. Door onze methode toe te passen op het meten van het consumentenvertrouwen in Nederland demonstreren we dat de veranderingen in deze indicator eenvoudig te interpreteren zijn. Dit in tegenstelling tot de veranderingen in de momenteel gehanteerde indicator van het CBS. Monitoringssystemen voor de conjuncturele toestand van de economie kunnen sterk verbeterd worden door in te zien dat voorlopende indicatoren een andere voorlooptijd hebben ten tijde van pieken dan ten tijde van dalen. Door deze asymmetrie in voorlooptijd te modelleren kunnen omslagpunten in de economie nauwkeuriger en eerder voorspeld worden.

Bibliography

Adams, L. M. and G. Darwin (1982), Solving the Quandary Between Questionnaire Length and Response Rate in Educational Research, *Research in Higher Education*, 17, 231–240.

Albert, J. H. and S. Chib (1993), Bayes Inference via Gibbs Sampling of Autoregressive Time Series subject to Markov Mean and Variance Shifts, *Journal of Business & Economic Statistics*, 11, 1–15.

Allison, D. B., R. L. Allison, M. S. Faith, F. Paultre, and F. X. Pi-Sunyer (1997), Power and Money: Designing Statistically Powerful Studies While Minimizing Financial Costs, *Psychological Methods*, 2, 20–33.

Amemiya, T. (1984), Tobit Models: A Survey, *Journal of Econometrics*, 24, 3–61.

Ansari, A., C. F. Mela, and S. A. Neslin (2008), Customer Channel Migration, *Journal of Marketing Research*, 45, 60–76.

Arellano, M. (2003), *Panel Data Econometrics*, Oxford University Press, New York.

Atkinson, A. C. and A. N. Donev (1992), *Optimum Experimental Designs*, Clarendon Press, Oxford Science Publications, Oxford, England.

Baltagi, B. H. (2008), *Econometric Analysis of Panel Data*, fourth edn., John Wiley & Sons, Chichester.

Boldin, M. D. (1996), A Check on the Robustness of Hamilton's Markov Switching Approach to the Economic Analysis of Business Cycles, *Studies in Nonlinear Dynamics and Econometrics*, 1, 35–46.

Burns, A. F. and W. C. Mitchell (1946), *Measuring Business Cycles*, National Bureau of Economic Research, New York.

Camacho, M. and G. Perez-Quiros (2002), This Is What the Leading Indicators Lead, *Journal of Applied Econometrics*, 17, 61–80.

Chauvet, M. (1998), An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching, *International Economic Review*, 39, 969–996.

Chauvet, M. and J. D. Hamilton (2006), Dating Business Cycle Turning Points, in C. Milas, P. Rothman, and D. Van Dijk (eds.), *Nonlinear Time Series Analysis of Business Cycles*, Elsevier, North-Holland, pp. 1–54.

Chauvet, M. and J. Piger (2008), A Comparison of the Real-Time Performance of Business Cycle Dating Methods, *Journal of Business & Economic Statistics*, 26, 42–49.

Chib, S. (1996), Calculating Posterior Distributions and Model Estimates in Markov Mixture Models, *Journal of Econometrics*, 75, 79–97.

Clements, M. P. and H.-M. Krolzig (2003), Business Cycle Asymmetries: Characterization and Testing Based on Markov-Switching Autoregressions, *Journal of Business & Economic Statistics*, 21, 196–211.

Collado, M. D. (1997), Estimating Dynamic Models from Time Series of Independent Cross-sections, *Journal of Econometrics*, 82, 37–62.

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

De Pooter, M., F. Ravazzolo, R. Segers, and H. K. Van Dijk (2008), *Bayesian Near-Boundary Analysis in Basic Macroeconomic Time Series Models*, vol. 23, Bayesian Econometrics of *Advances in Econometrics*, Elsevier Science, forthcoming.

De Stavola, B. (1986), Sampling Designs for Short Panel Data, *Econometrica*, 54, 415–424.

Deaton, A. (1985), Panel Data from Time Series of Cross-sections, *Journal of Econometrics*, 30, 109–126.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), ML Estimation from Incomplete Data via the EM algorithm, *Journal of the Royal Statistical Society*, B39, 1–38.

Deutskens, E., K. De Ruyter, M. Wetzels, and P. Oosterveld (2004), Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study, *Marketing Letters*, 15, 21–36.

Dholakia, U. M. and V. G. Morwitz (2002), The Scope and Persistence of Mere-measurement Effects: Evidence from a Field-study of Customer Satisfaction, *Journal of Consumer Research*, 29, 159–167.

Diebold, F. X., J. Lee, and G. C. Weinbach (1994), Regime Switching and Endogenous Transition Probablities, in C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*, chap. 10, Oxford University Press, Oxford, pp. 283–302.

Diebold, F. X. and R. S. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business & Economic Statistics*, 13, 253–263.

Diebold, F. X. and G. D. Rudebusch (1989), Scoring the Leading Indicators, *Journal of Business*, 62, 369–391.

Diebold, F. X. and G. D. Rudebusch (1991), Forecasting Output with the Composite Leading Index: An Ex Ante Analysis, *Journal of the American Statistical Association*, 86, 603–610.

Diebold, F. X. and G. D. Rudebusch (1996), Measuring Business Cycles: A Modern Perspective, *Review of Economics and Statistics*, 78, 67–77.

Dillman, D. A. (2000), *Mail and Internet Surveys: The Tailored Design Method*, second edition edn., John Wiley & Sons, New York.

Dominitz, J. and C. F. Manski (2004), How Should We Measure Consumer Confidence?, *Journal of Economic Perspectives*, 18, 51–66.

Dotsey, M. (1998), The Predictive Content of the Interest Rate Term Spread for Future Economic Growth, *Federal Reserve Bank of Richmond Economic Quarterly*, 84-3, 31–51.

Dueker, M. J. (1991), Strengthening the Case for the Yield Curve as a Predictor of U.S. Recessions, *Federal Reserve Bank of St. Louis Review*, 79, 41–51.

Dueker, M. J. (2005), Dynamic Forecasts of Qualitative Variables: A Qual VAR Model of U.S. Recessions, *Journal of Business & Economic Statistics*, 23, 96–104.

Duncan, G. J. and G. Kalton (1987), Issues of Design and Analysis of Surveys Across Time, *International Statistical Review*, 55, 97–117.

Erdem, T. and B. Sun (2001), Testing for Choice Dynamics in Panel Data, *Journal of Business & Economic Statistics*, 19, 142–152.

Estrella, A. and G. A. Hardouvelis (1991), The Term Structure as a Predictor of Real Economic Activity, *Journal of Finance*, 46, 555–576.

Estrella, A. and F. S. Mishkin (1998), Predicting U.S. Recessions: Financial Variables as Leading Indicators, *Review of Economic Studies*, 80, 45–61.

Filardo, A. J. (1994), Business-Cycle Phases and Their Transitional Dynamics, *Journal of Business & Economic Statistics*, 12, 299–308.

Filardo, A. J. (1999), How Reliable Are Recession Prediction Models?, *Federal Reserve Bank of Kansas City Economic Review*, 84, 35–55.

Filardo, A. J. (2004), The 2001 US Recession: What Did Recession Prediction Models Tell Us?, Bank of International Settlements Working Paper No. 148.

Franses, P. H. B. F. and R. Paap (2001), *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge, UK.

Frühwirth-Schnatter, S. (2001), Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models, *Journal of American Statistical Association*, 96, 194–209.

Geisser, S. (1965), A Bayes Approach for Combining Correlated Estimates, *Journal of the American Statistical Association*, 60, 602–607.

Geman, S. and D. Geman (1984), Stochastic Relaxations, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Geweke, J. F. (2007), Interpretation and Inference in Mixture Models: Simple MCMC Works, *Computational Statistics and Data Analysis*, 51, 3529–3550.

Graham, J. W., B. J. Taylor, A. E. Olchowski, and P. E. Cumsille (2006), Planned Missing Data Designs in Psychological Research, *Psychological Methods*, 11, 323–343.

Hamilton, J. D. (1989), A New Approach to the Econometric Analysis of Nonstationary Time Series and Business Cycles, *Econometrica*, 57, 357–384.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, New Jersey.

Hamilton, J. D. and G. Perez-Quiros (1996), What Do the Leading Indicators Lead?, *Journal of Business*, 69, 27–49.

Hansen, R. A. (1980), A Self-Perception Interpretation of the Effect of Monetary and Nonmonetary Incentives on Mail Survey Respondent Behavior, *Journal of Marketing Research*, 17, 77–83.

Hausman, J. A. and D. A. Wise (1979), Attrition Bias in Experimental and Panel Data: The Gary Income Maintainance Experience, *Econometrica*, 47, 455–473.

Heckman, J. J. (1976), The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement*, 5, 475–492.

Heckman, J. J. (1979), Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153–161.

Heckman, J. J. (1981), The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time - Discrete Data Stochastic Process, in C. F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, pp. 179–195.

Herrera, A.-M. and E. Pesavento (2005), The Decline in U.S. Output Volatility: Structural Changes and Inventory Investment, *Journal of Business & Economic Statistics*, 23, 462–472.

Heubrich, J. G. and A. M. Dombrosky (1996), Predicting Real Growth Using the Yield Cruve, *Federal Reserve Bank of Cleveland Economic Review*, 32, 26–35.

Hirano, K., G. Imbens, G. Ridder, and D. B. Rubin (2001), Combining Panel Data Sets with Attrition and Refreshment Samples, *Econometrica*, 69, 1645–1659.

Hsiao, C. (2003), *Analysis of panel data*, second edn., Cambridge University Press.

Issler, J. and F. Vahid (2006), The missing link: using the NBER recession indicator to construct coincident and leading indices of economic activity, *Journal of Econometrics*, 132, 281–303.

Johnson, E. G. (1992), The Design of the National Assessment of Educational Progress, *Journal of Educational Measurement*, 29, 95–110.

Katona, G. (1951), *Psychological Analysis of Economic Behavior*, McGraw Hill, New York.

Keane, M. P. (1997), Modeling Heterogeneity and State Dependence, *Journal of Business & Economic Statistics*, 310–327.

Kim, C.-J. and C. R. Nelson (1998), Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching, *The Review of Economic and Statistics*, 80, 188–201.

Kim, C.-J. and C. R. Nelson (1999), *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, MIT Press, Cambridge, MA.

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

King, G., O. Rosen, and M. A. Tanner (2004), *Ecological Inference: New Methodological Strategies*, Cambridge University Press, New York.

Kish, L. and I. Hess (1959), A "Replacement" Procedure for Reducing the Bias of Non-response, *The American Statistician*, 13-4, 17–19.

Klein, P. A. (2001), The Leading Indicators in Historical Perspective, in *Business Cycle Indicators Handbook*, pp. 23–28.

Krolzig, H.-M. (1997), *Markov Switching Vector Autoregressions: Modelling, Statistical Inference, and an Application to Business Cycle Inference*, Springer, Berlin.

Little, R. J. A. (1995), Modeling the Drop-out Mechanism in Repeated-Measures Studies, *Journal of the American Statistical Association*, 90, 1112–1121.

Little, R. J. A. and R. B. Rubin (2002), *Statistical Analysis with Missing Data*, second edn., John Wiley & Sons.

Ludvigson, S. C. (2004), Consumer Confidence and Consumer Spending, *Journal of Economic Perspectives*, 18, 29–50.

Marcellino, M. (2006), Leading Indicators: What Have We Learned?, in G. Elliott, C. W. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Elsevier, North-Holland, pp. 879–960.

McClelland, G. H. (1997), Optimal Design in Psycological Research, *Psychological Methods*, 2, 3–19.

McConnell, M. and G. Perez-Quiros (2000), Output Fluctuations in the United States: What Has Changed since the 1980s?, *American Economic Review*, 90, 1464–1476.

McCulloch, R. E. and R. S. Tsay (1994), Statistical Analysis of Economic Time Series via Markov Switching Models, *Journal of Time Series Analysis*, 15, 523–540.

McGuckin, R. H., A. Ozyildirim, and V. Zarnowitz (2007), A More Timely and Useful Index of Leading Indicators, *Journal of Business & Economic Statistics*, 25, 110–120.

McKelvey, R. D. and W. Zavoina (1975), A Statistical Model for the Analysis of Ordinal Level Dependent Variables, *Journal of Mathematical Sociology*, 4, 103–120.

Mitchell, W. C. and A. F. Burns (1938), *Statistical Indicators of Cyclical Revivals*, National Bureau of Economic Research, New York.

Moffitt, R. (1993), Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-sections, *Journal of Econometrics*, 59, 99–123.

Moore, G. H. and J. Shiskin (1967), Indicators of Business Expansions and Contractions, *NBER Occasional Paper*, 103.

Morwitz, V. G. and G. J. Fitzsimons (2004), The Mere-Measurement Effect: Why Does Measuring Intentions Change Actual Behavior?, *Journal of Consumer Psychology*, 14, 64–74.

Nijman, T. and M. Verbeek (1990), Estimation of Time Dependent Parameters in Linear Models using Cross-sections, Panels or Both, *Journal of Econometrics*, 46, 333–346.

Nijman, T., M. Verbeek, and A. Van Soest (1991), The Efficiency of Rotating-panel Designs in an Analysis-of-variance Model, *Journal of Econometrics*, 49, 373–399.

Ofir, C. and I. Simonson (2001), In Search of Negative Customer Feedback: The Effect of Expecting to Evaluate on Satisfaction Evaluations, *Journal of Marketing Research*, 38, 170–182.

Ouwens, M. J. N. M., F. E. S. Tan, and M. P. F. Berger (2002), Maximin D-Optimal Designs for Longitudinal Mixed Effects Models, *Biometrics*, 58, 735–741.

Paap, R., R. Segers, and D. Van Dijk (2009), Do Leading Indicators Leads Peaks More Than Troughs?, *Journal of Business & Economic Statistics*, forthcoming.

Paap, R. and H. K. Van Dijk (2003), Bayes Estimates of Markov Trends in Possibly Cointegrated Series: An Application to US Consumption and Income, *Journal of Business & Economic Statistics*, 21, 547–563.

Palma, W. and N. H. Chan (1997), Estimation and Forecasting of Long-memory Time Series with Missing Values, *Journal of Forecasting*, 16, 395–410.

Patterson, H. D. (1950), Sampling on Successive Occasions with Partial Replacement of Units, *Journal of the Royal Statistical Society Series B*, 49, 241–255.

Presser, S., J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, and E. Singer (2004), *Methods for Testing and Evaluating Survey Questionnaires*, Wiley, New York.

Raghunathan, T. and J. Grizzle (1995), A Split Questionnaire Survey Design, *Journal of the American Statistical Association*, 90, 54–63.

Reingen, P. H. and J. B. Kernan (1977), Compliance with an Interview Request: A Foot-in-the-Door, Self-Perception Interpretation, *Journal of Marketing Research*, 14, 365–369.

Rubin, D. B. (1976), Inference and Missing Data, *Biometrika*, 63, 581–592.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, Chapman and Hall, New York.

Schafer, J. L. and J. W. Graham (2002), Missing Data: Our view of the State of the Art, *Psychological Methods*, 7, 147–177.

Segers, R. and P. H. B. F. Franses (2007), Panel Design Effects on Response Rates and Response Quality, *Econometric Institute Report Series*, EI 2007-29, Erasmus University Rotterdam.

Segers, R. and P. H. B. F. Franses (2008), Measuring Weekly Consumer Confidence, *Econometric Institute Report Series*, EI 2008-01, Erasmus University Rotterdam.

Segers, R., R. Paap, and D. Van Dijk (2008), Evaluating the Consistency and Timeliness of Business Cycle Indicators, *Econometric Institute Report Series*, forthcoming, Erasmus University Rotterdam.

Sensier, M. and D. Van Dijk (2004), Testing for Volatility Changes in US Macroeconomic Time Series, *Review of Economics and Statistics*, 86, 833–839.

Serati, M. and G. Amisano (2008), Building Composite Leading Indexes in a Dynamic Factor Model Framework: A New Proposal, *Liuc Papers n. 212, Serie Economia e Impresa*, 58.

Shoemaker, D. M. (1973), *Principles and Procedures of Multiple Matrix Sampling*, Ballinger Publishing, Cambridge.

Shumway, R. H. and D. S. Stoffer (1982), An Approach to Time Series Smoothing and Forecasting using the EM Algorithm, *Journal of Time Series Analysis*, 3, 253–264.

Shumway, R. H. and D. S. Stoffer (2006), *Time Series Analysis and Its Applications*, second edn., Springer, New York.

Sichel, D. E. (1994), Inventories and the Three Phases of the Business Cycle, *Journal of Business & Economic Statistics*, 12, 269–277.

Sigelman, L. (1991), Turning Cross Sections into a Panel: A Simple Procedure of Ecological Inference, *Social Science Research*, 20, 150–170.

Simmons, C. J., B. A. Bickart, and J. G. Lynch, Jr. (1993), Capturing and Creating Public Opinion in Survey Research, *Journal of Consumer Research*, 20, 316–329.

Smith, A. F. M. and G. O. Roberts (1993), Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society B*, 55, 3–23.

Smith, P. A. and P. M. Summers (2005), How Well do Markov Switching Models describe Actual Business Cycles? The Case of Synchronization, *Journal of Applied Econometrics*, 20, 253–274.

Stock, J. H. and M. W. Watson (1989), New Indexes of Coincident and Leading Economic Indicators, in O. J. Blanchard and S. Fischers (eds.), *NBER Macroeconomics Annual 1989*, MIT Press, pp. 352–394.

Stock, J. H. and M. W. Watson (2003), How Did Leading Indicator Forecasts Perform During the 2001 Recession?, *Federal Reserve Bank of Richmond Economic Quarterly*, 89, 71–90.

Swanson, N. R. and D. Van Dijk (2006), Are Statistical Reporting Agencies Getting it Right? Data Rationality and Business Cycle Asymmetry, *Journal of Business & Economic Statistics*, 24, 24–42.

Tanner, M. A. and W. H. Wong (1987), The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, 82, 528–550.

The Conference Board (2001), *Business Cycle Indicators Handbook*, The Conference Board, New York.

Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions, *The Annals of Statistics*, 22, 1701–1762.

Trivellato, U. (1999), Issues in the Design and Analysis of Panel Studies: A Cursory Review, *Quality & Quantity*, 33, 339–352.

Van Diepen, M., B. Donkers, and P. H. B. F. Franses (2009), Dynamic and Competitive Effects of Direct Mailings, *Journal of Marketing Research*, forthcoming.

Verbeek, M. (1991), *The Design of Panel Surveys and the Treatment of Missing Observations*, Ph.D. thesis, Tilburg Univesity.

Verbeek, M. and T. Nijman (1993), Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross-sections, *Journal of Econometrics*, 59, 125–136.

Verbeek, M. and F. Vella (2005), Estimating Dynamic Models from Repeated Cross-sections, *Journal of Econometrics*, 127, 83–102.

Wooldridge, J. M. (2005), Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity, *Journal of Applied Econometrics*, 20, 39–54.

Yu, J. and H. Cooper (1983), A Quantitative Review of Research Design Effects on Response Rates to Questionaires, *Journal of Marketing Research*, 20, 36–44.

Zarnowitz, V. and C. Boschan (1975), New Composite Indexes of Coincident and Lagging Indicators, *Business Condition Digest*, November, Reprinted in *Handbook of Cyclical Indicators*, 1944, 170–183.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Author Index

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit 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:

396. P.D.E. DINDO, *Bounded rationality and heterogeneity in economic dynamic models.*
397. D.F. SCHRAGER, *Essays on asset liability modeling.*
398. R. HUANG, *Three essays on the effects of banking regulations.*
399. C.M. VAN MOURIK, *Globalisation and the role of financial accounting information in Japan.*
400. S.M.S.N. MAXIMIANO, *Essays in organizational economics.*
401. W. JANSSENS, *Social capital and cooperation: An impact evaluation of a women's empowerment programme in rural India.*
402. J. VAN DER SLUIS, *Successful entrepreneurship and human capital.*
403. S. DOMINGUEZ MARTINEZ, *Decision making with asymmetric information.*
404. H. SUNARTO, *Understanding the role of bank relationships, relationship marketing, and organizational learning in the performance of people's credit bank.*
405. M.Á. DOS REIS PORTELA, *Four essays on education, growth and labour economics.*
406. S.S. FICCO, *Essays on imperfect information-processing in economics.*
407. P.J.P.M. VERSIJP, *Advances in the use of stochastic dominance in asset pricing.*
408. M.R. WILDENBEEST, *Consumer search and oligopolistic pricing: A theoretical and empirical inquiry.*
409. E. GUSTAFSSON-WRIGHT, *Baring the threads: Social capital, vulnerability and the well-being of children in Guatemala.*
410. S. YERGOU-WORKU, *Marriage markets and fertility in South Africa with comparisons to Britain and Sweden.*
411. J.F. SLIJKERMAN, *Financial stability in the EU.*
412. W.A. VAN DEN BERG, *Private equity acquisitions.*
413. Y. CHENG, *Selected topics on nonparametric conditional quantiles and risk theory.*
414. M.D. DE POOTER, *Modeling and forecasting stock return volatility and the term structure of interest rates.*
415. F. RAVAZZOLO, *Forecasting financial time series using model averaging.*
416. M.J.E. KABKI, *Transnationalism, local development and social security: the functioning of support networks in rural Ghana.*
417. M. POPLAWSKI RIBEIRO, *Fiscal policy under rules and restrictions.*
418. S.W. BISSESSUR, *Earnings, quality and earnings management: the role of accounting accruals.*
419. L. RATNOVSKI, *A Random Walk Down the Lombard Street: Essays on Banking.*

420. R.P. NICOLAI, *Maintenance models for systems subject to measurable deterioration.*
421. R.K. ANDADARI, *Local clusters in global value chains, a case study of wood furniture clusters in Central Java (Indonesia).*
422. V. KARTSEVA, *Designing Controls for Network Organizations: A Value-Based Approach.*
423. J. ARTS, *Essays on New Product Adoption and Diffusion.*
424. A. BABUS, *Essays on Networks: Theory and Applications.*
425. M. VAN DER VOORT, *Modelling Credit Derivatives.*
426. G. GARITA, *Financial Market Liberalization and Economic Growth.*
427. E. BEKKERS, *Essays on Firm Heterogeneity and Quality in International Trade.*
428. H. LEAHU, *Measure-Valued Differentiation for Finite Products of Measures: Theory and Applications.*
429. G. BALTUSSEN, *New Insights into Behavioral Finance.*
430. W. VERMEULEN, *Essays on Housing Supply, Land Use Regulation and Regional Labour Markets.*
431. I.S. BUHAI, *Essays on Labour Markets: Worker-Firm Dynamics, Occupational Segregation and Workplace Conditions.*
432. C. ZHOU, *On Extreme Value Statistics.*
433. M. VAN DER WEL, *Riskfree Rate Dynamics: Information, Trading, and State Space Modeling.*
434. S.M.W. PHLIPPEN, *Come Close and Co-Create: Proximities in pharmaceutical innovation networks.*
435. A.V.P.B. MONTEIRO, *The Dynamics of Corporate Credit Risk: An Intensity-based Econometric Analysis.*
436. S.T. TRAUTMANN, *Uncertainty in Individual and Social Decisions: Theory and Experiments.*
437. R. LORD, *Efficient pricing algorithms for exotic derivatives.*
438. R.P. WOLTHOFF, *Essays on Simultaneous Search Equilibrium.*
439. Y.-Y. TSENG, *Valuation of travel time reliability in passenger transport.*
440. M.C. NON, *Essays on Consumer Search and Interlocking Directorates.*
441. M. DE HAAN, *Family Background and Children's Schooling Outcomes.*
442. T. ZAVADIL, *Dynamic Econometric Analysis of Insurance Markets with Imperfect Information*
443. I.A. MAZZA, *Essays on endogenous economic policy*
444. R. HAIJEMA, *Solving large structured Markov Decision Problems for perishable-inventory management and traffic control*
445. A.S.K. WONG, *Derivatives in Dynamic Markets*