Topics in Applied Macroeconometrics

ISBN 978 90 361 0451 7

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

This book is no. 623 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Topics in Applied Macroeconometrics

Een selectie macro-econometrische toepassingen

Thesis

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

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Thursday September 24, 2015 at 13.30 hours

by

Sjoerd van den Hauwe

born in Oostburg

Erasmus University Rotterdam

Doctoral Committee

Promoters: Prof.dr. D.J.C. van Dijk
Prof.dr. R. Paap

Other members: Dr. C.S. Bos
Prof.dr. D. Fok
Prof.dr. Ph.H.B.F. Franses
Dr. N. Baştürk
Dr. A. Pick

Preface

Writing a thesis is quite an individualistic enterprise mainly consisting of numerous hours spent in front of a computer: looking up papers, composing and decomposing models, preparing data, programming, debugging, generating output, writing the papers, rewriting the papers... and rewriting the papers. It is one of those exercises about which your posterior beliefs substantially deviate from their prior counterparts—and probably for the best; in such a case the learning process has been most informative.

Although the lion's share of this thesis is the result of all these desk-bound hours, I couldn't have finished it without the help of my promoters Richard Paap and Dick van Dijk whom I am grateful for, first, offering me the opportunity to do a PhD and, second, for keeping me (and a few times helping me *back*) on track throughout the years. At the time I embarked the project I wasn't quite Dick's only PhD student, but nevertheless he somehow always managed to find half an hour in his busy daily schedule to talk about the papers and suggest corrections or alternative approaches. Moreover, his academic writing skills helped me a lot in learning how to concisely write a paper. Richard became my second promoter only later during the project (when we needed his Bayesian expertise!), but I already knew him since the start of my studies in Rotterdam, when I was assigned to his mentor class. In all those years I have extensively benefited from his econometric knowledge and tricks. Moreover, I honestly believe it was his Bayesian econometrics course in my Master's that made me decide to continue with econometrics for a few more years.

With the defense of this thesis my doctoral project has come to an end, and I would like to thank Charles Bos, Dennis Fok and Philip Hans Franses for consuming some of their undoubtedly precious time to review the typoscript. In addition to the inner committee members I also thank Nalan Baştürk and Andreas Pick for completing the doctoral committee. Furthermore, I thank both the staffs of the Econometric Institute and the Tinbergen Institute who were always willing to help and support, TI's Carolien and Judith in paricular.

Luckily I also had a bunch of good and kind colleagues who offered the necessary distraction during my stays on campus. I would like to thank Cem, Dennis, Eelco, Hans, Kar Yin, Mathijn, Peter E., Peter Z., Rianne, Sjoerd, Suzanne, and Tim for the entertaining talks, coffee breaks and occasional lunches; Arjan, Yuri and Harwin for sharing an office with me, although sometimes the intersection of our office hours tended to be small. Particular thanks go to Anne and Jorn. Jorn, whom I first met at the TI drinks in september 2007, was an excellent partner during the assignments we jointly worked on in our first year, and an even better colleague and friend in later years—thanks for the agreeable work interruptions and the good laughs we had. One autumn Tuesday night Anne jumped into my office (not improbably literally) asking me something about a matrix decomposition and although I doubt I had an adequate answer to his question, since then we regularly frequented each other's offices or went out for beers or playing pool—thanks for your friendship and never-fading enthusiasm. I strongly appreciate it that both Jorn and Anne have agreed to be my paranymphs during the thesis defense.

Finally, many thanks to the following people were very kind and helpful in making me forget about the thesis every now and then. Kenny, for the many beers, the hours watching football, the serious and less serious talks and reminding me that there is always a way out when you feel like stuck at a "Pampilhosa" train station. Francis, for the Maastricht dinners and drinks and showing that time doesn't really change us. Leontien and Eric, for their love and moral support. Eva, for being the sweetest and smartest little sister I could wish for. Laurens, for taking good care of Eva and tolerating much of my baloney. Mark, for being everything a big brother should be; and you probably were the only one who confidently knew I sooner or later would finish this thesis, despite me telling otherwise—of course you turned out right.

S.v.d.H.
Rotterdam
July 15, 2015.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 General setting

The outcomes of economic decisions depend upon factors yet unknown at the moment when the decisions are made. Therefore, decisionmakers need useful estimates of these factors in order to make a sensible and justified decision which minimizes their expected loss. For example, producers like to know the future state of the economy to adjust their production plans or inventory management; an entrepreneur needs an estimate of any coming recessionary period to decide whether to invest or not; a policymaker is interested in changes in labor market conditions to formulate new policy or suggest labor market interventions. It is one of the econometrician's tasks to deliver these estimates of short- to mid-term future economic conditions by computing quantitative model-based forecasts which serve as inputs to the decisionmaker. Naturally, such forecasts demand insights into the economic dynamics which need to be properly integrated into a model. The topics in this thesis touch upon issues related to the modeling of macroeconomic relations to be applied for forecasting purposes.

Historically, economists have been ambitious in formulating all-encompassing macroeconomic models. These models were supposed to both explain the behavior of economic agents and at the same time to generate practically useful and reliable forecasts. Although tempting and conceptually most satisfying, economic reality turned out to be stubborn, and, in the best case, willing to fit these models only for short periods of time. As a result, different schools of economic thought evolved, each with its own set of basic assumptions. Despite these differing and often conflicting theoretical perspectives, nowadays economists do agree about one thing: an economic model must find substantial support in the data and generate predictions in accordance with what is found empirically.

Diebold (1998) distinguishes two traditions in current macroeconomics. First, the non-structural approach, taking the requirement of empirical evidence for the applied model as starting point, and second, the structural approach which still aspires to both understand and forecast economic behavior. The latter is the more appealing approach from the deductive point of view, and shows promising recent developments with dynamic stochastic general equilibrium modeling. However, this new field still encounters a vast amount of computational challenges to empirically test the theoretical models. The former is the more pragmatic approach to obtain grip on economic reality. Because of its greater empirical effectiveness it is persistently used by policymakers and applied for commercial use, for at least the last four to five decades.

The topics discussed in this thesis are to be placed in the nonstructural tradition as well. The main purpose in each chapter is to establish a statistical relationship between a macroeconomic variable of interest y and predictor variables x. The latter are either past realizations of y or economic variables known at earlier points in time that relate predictively to future y. This approach essentially boils down to the identification of significant cross-variable reduced-form correlations to be exploited for forecasting goals. Boldly stated, the nonstructural approach is a shortcut to circumvent the demanding twofold process of building a realistic structural economic model that at the same time produces useful forecasts. Instead it immediately directs attention to forecasting, mainly motivated by the proven greater empirical effectiveness. The main ingredients of this approach consist of economic intuition, stylized facts based on historical data, and a wide spectrum of statistical techniques.

In the nonstructural tradition, forecasting requires a statistical model of the macroeconomic target variable y. Such a model contains two fundamentals. First, any target variable will always be (at least) partially determined by instrinsically unpredictable components like news, political developments, unforeseen technological advances *et cetera*. On the other hand, the dynamic behavior of y and its relationship to explanatory variables x provide regularities which allow for prediction. The predictable part of y forms the second fundamental of the model. The challenge for the (nonstructural) econometrician is to design statistical models and operationalize methods to identify relevant predictor variables and their predictive relation to the variable of interest such that the unpredictable part of y is as small as possible. Furthermore, his models should provide an adequate description of the statistical characteristics of the unpredictable part in order to properly incorporate this source of uncertainty into predictive statements.

1.2 Motivation

The linear predictive regression model is the workhorse of the econometrician and nearly always serves as the starting point for nonstructural model building. Though powerful, it drastically simplifies reality and as such it will only partially capture the predictable part of y. The chapters in this thesis deal with (substantial) departures from this basic model in various ways, by the relaxation of (parts of) its stringent assumptions. The resulting models better approximate economic reality and produce more accurate and reliable forecasts.

Econometric topics

The first two topics addressed in this thesis include variable selection, that is, the quest for optimal predictors from a given set of variables x, and the optimal selection of the lead time of an x which is *a priori* known for leading y. Both issues feature in, at least conditionally, linear models (Chapters 2 and 3). Two other topics are, first, potentially time-varying effects of predictors on y, which is discussed for both linear and nonlinear models (Chapter 4), and, second, nonstandard relations between x and y, covering nonlinearity and heteroskedasticity, and nonnormality of the target variable (Chapter 5). Why these topics are relevant for macroeconomic forecasting is discussed next.

Not only the computational resources econometricians rely on today have undergone a small revolution, also the amount of (publicly) available economic data has increased tremendously. The basic regression model assumes the set of predictor variables x to be known. Though, with the big set of data on many economic variables this is certainly not straightforward and the econometrician should take one step back and first identify the most relevant predictors for a given forecasting problem. Naively putting all potential predictors into the model is not the best way to proceed. Since many variables will either be irrelevant or strongly related to others, this increases inefficiency and troubles the predictive signal such that the produced forecasts become less precise. On the other hand, omitting relevant predictors from the model leads to biased forecasts. The implementation of model averaging provides an elegant solution, as discussed by Hoeting *et al.* (1999). Each potential predictor gets assigned a posterior probability of being relevant for forecasting y, which is incorporated into the predictive distribution. An alternative is factor modeling, which summarizes the sample information from a large number of economic variables into a small number of factors, which in turn serve as predictors of y (Stock and Watson, 2002b). Or, the partial least squares method which, in an iterative procedure, extracts the info in the data set most strongly related to y (Garthwaite, 1994). Both alternatives are more *ad-hoc* methods to significantly decrease the dimension of the data set, whereas model averaging, as employed in Chapter 2, also allows for economic cross-variable inference about which x's are particulary related to y.

In the 1920s, empirical business cycle modeling experienced a growing interest (see Morgan, 1990, Ch. 2), and instead of tracing the source of business cycle fluctuations—although that goal remained active in the background—attention shifted to finding a “barometer” anticipating any changes in business activity. One of the main findings was the identifcation of three types of macroeconomic variables, which is maintained up to this day (The Conference Board, 2001). That is, once the business cycle is approximated (*e.g.* by the cyclical movements in gross domestic product, of which the estimation is a demanding exercise itself), a series coincides with the cycle, or either leads or lags it. Obviously, the set of leading variables forms the input to the construction of the business barometer. Thus, the problem of variable selection as encounterd in Chapter 2 is already taken care of. Given the set consisting of a moderate amount of leading variables, the econometrician's task lies in properly relating them to business fluctuations y. More in particular, the number of months the x's are leading and the relative strength of their relation to y need to be identified. These issues are addressed in Chapter 3, and though their nature has not changed in nearly a hundred years, the methods to analyze them have. Persons (1919) analyzed similar issues in his seminal work in the field. He used slides with the plotted time series of the macroeconomic variables and, using an overhead projector, he shifted these horizontally over the business cycle series such that a panel of “observers” could visually decide upon each leading indicator's lead time and correlation with the business cycle (Morgan, 1990, provides an historical discussion of Persons' methods). In Chapter 3 this procedure is “automated” using modern computational techniques, replacing the subjective judgement by quantitative statistical inference.

The previous two topics are examined with models that presume relationships between x and y to be stable over time. When assuming the parameters of these functional relations to be constant, all available data can be used in the estimation routine to identify their unknown values. In macroeconomic forecasting, (dynamic) models of which the parameters are estimated using time-series data, *i.e.*, a set of successive observations of the economic variables, various theoretical reasons exist that make it plausible that the model parameters are time-dependent instead. Moreover, substantial empirical evidence for this assumption is found as well. Stock and Watson (1996) report evidence for permanent changes in time-series models' parameters by applying statistical testing procedures. In contrast to these structural changes, Hamilton (1989) introduces models that formulate switching between different parameter regimes that determine the way x and y are related, and shows their empirical importance. Ignoring the possibly time-varying character of model parameters, either permanent or transitory, has obvious consequences for identifying cross-variable relations and macroeconomic forecasting. In Chapter 4, a method is discussed to deal with parameter changes and their implications for empirical modeling.

In the linear model the effect of a change in x on y is always the same. In a model with time-varying parameters the effect does change, but it depends on time and not on the value of x. Assuming a nonlinear relationship between predictor and predicted variable allows for testing whether the *strength* of the relation changes with the value of x. For instance, a leading indicator can contain more predictive power for negative growth rates of the economy than for predicting positive growth. Such nonlinearities directly affect the predicted y. Threshold regressions, models in which the parameter values change once x drops below a threshold value (and their smooth-transition generalizations), can be applied to incorporate nonlinearities (Teräsvirta, 2006). If the decisionmaker has a nonstandard loss function (see Geweke, 2005, Section 2.4, for examples), and requires forecasting statements of the value-at-risk type like "with a 5% probability y will decline with at least 8% next quarter," modeling the (possibly nonlinear) conditional mean of the target variable may not suffice. If a decision heavily depends upon the likelihood of atypical events, for example a severe recessionary period, the entire conditional distribution is important, with special interest turned to the left tail of y. A model is required that adequately describes the statistical properties of the unpredictable part and prevents underestimation of the likelihood of large negative shocks. This substantial set of considerations demands the econometrician to make a whole series of model assumptions. Instead, a nonparametric approach eliminates most of such possibly restrictive assumptions and it lets the data to be primarily leading in pointing out what form of distribution y fits best. In Chapter 5, such an approach is implemented for forecasting changes in economic activity by exploiting the flexibility of mixture modeling (Richardson and Green, 1997).

Bayesian analysis

During the process of finding an econometric solution to the initial problem—how to construct useful macroeconomic forecasts—many subproblems are encountered. In this thesis, such subproblems as which predictors to use (Chapter 2), what lead times of business-cycle-leading indicators to set (Chapter 3), whether to consider model parameters constant over time (Chapter 4), or what functional form between x and y to specify (Chapter 5). To arrive at a practically workable model, these questions need to be answered. However, each answer and resulting choice of model entails the exclusion of all other models. Since there generally is no guarentee that the right choices are made, ideally econometricians incorporate this type of uncertainty into their outcomes.

The Bayesian framework for statistical analysis accounts for the different sources of uncertainty in an unambiguous and formal way, according to the rules of probability. It produces the so-called posterior predictive distribution which formalizes the beliefs about future y given the predictor x, after having observed a set of predictor-target pairs from the past.

First, the econometrician puts forward his prior belief about the target variable by suggesting a (moderate) range of specifications that statistically relate x to y. These specifications contain parameters of which the values are unknown, and a belief about what values are *a priori* considered likely is specified as well. The prior is formulated relying on experience and (expert) economic knowledge. Second, using a set of observed (x, y)-pairs the prior belief is updated to obtain its posterior counterpart. That is, the prior belief and the sampled data combined lead to a new ranking of models and their posterior likely parameter values. Third and finally, this posterior belief is used to construct the posterior *predictive* belief about a future y, which is a “weighted average” of the statistical specifications according to their respective posterior model probabilities (model uncertainty) and posterior likely parameter values (parameter uncertainty). All applications in this thesis are analyzed following this principle such that the various types of uncertainty are made an inherent part of any predictive statement.

Beside being conceptually appealing, the Bayesian approach offers some far-reaching computational advantages as well. The fact that the linear predictive regression model is easy to estimate made it the fundamental empirical model, and econometricians of course admitted it to be too simplistic in many cases. However, computational resources at the time inhibited the implementation of more sophisticated and realistic models. Nowadays, the development of modern simulation techniques (see, for example, Robert and Casella, 2004, for an overview) combined with exponentially grown computing power and storage capacity have made it possible to estimate, and forecast with, models substantially deviating from this basic statistical model. Applying the Bayesian principle, to implement these more involved models they are “broken down” into computational subproblems which are often solved by applying simulation methods. The topics and empirical applications in this thesis would have been impossible to examine and operationalize were it not for modern computing power and techniques like latent-variable modeling (see, *e.g.*, Smith, 1973) which are employed in all four chapters. Geweke (1999, 2001) provides a discussion of the Bayesian framework, both from the perspective of decisionmaking, dealing with the different sources of uncertainty, and with regard to the extensive computational advantages it provides.

1.3 Outline

The four chapters after this introduction deal with empirical applications of the previously mentioned topics and can be read independently. Appendices A and B contain definitions of probability distributions and results for Bayesian regression settings, respectively, which are frequently applied throughout this thesis. The following outline provides per chapter the macroeconomic application on hand, a short description of the econometric methodology and brief empirical results.

Chapter 2 is build on the question which macroeconomic and financial variables possess the most predictive ability for the (qualitative) federal funds target rate decisions as made by the Federal Open Market Committee (FOMC). The analysis is conducted for the 157 FOMC decisions during the period January 1990–June 2008, using dynamic ordered probit models with a Bayesian endogenous variable selection methodology to implement the model averaging principle. Real-time data on a set of 33 candidate predictor variables are used in order to produce practically useful forecasts. Indicators of economic activity and forward-looking term structure variables, as well as survey measures turn out to be most informative from a forecasting perspective. For the full sample period, in-sample probability forecasts achieve a hit rate of 90 percent. Based on out-of-sample forecasts for the period January 2001–June 2008, 82 percent of the decisions are predicted correctly. This chapter is based on Van den Hauwe *et al.* (2013).

Macroeconomic indicators leading the business cycle form the core of Chapter 3. Potential heterogeneity in both the lead times and the relative importance of the ten leading indicators which jointly (and contemporaneously) build The Conference Board's (TCB) composite leading index, is examined. To this end, both business-cycle information contained in continuous coincident series like industrial production and employment, as well as the zero-one recession indicator are linked to the ten individual leading indicators with unknown lead times and possibly heterogenous weights. In the Bayesian framework, the resulting mixed discrete-continuous parameter space is straightforwardly dealt with and, additionally, it provides the context to generate real-time recession-probability forecasts. Posterior results show that leading indicators can be grouped according to their lead time, being either one to three months, or longer horizons up to one year. Financial variables turn out to get relatively the largest weights. We exploit these results to form an alternative composite leading economic index with heterogeneity in lead times as well as in weights which is better capable of forecasting recessions both in- and out-of-sample. Chapter 3 is based on Van den Hauwe *et al.* (2014).

In Chapter 4, an approach based on Dirichlet process mixtures models is proposed to deal with parameter changes in empirical macroeconomic models. The key element of the approach consists of a dynamic latent-variable specification for the model parameters which supports new-born parameter values at each point in time. New parameter values are generated from a so-called base prior distribution. The modeling procedure consists of the choice of a parametric likelihood specification and a base prior with the proper support for the parameters liable to change. The flexibility in the combinations of these two inputs shows advantageous compared to existing methods. Moreover, the approach accounts for both an unknown number of in-sample and potential out-of-sample parameter changes, since these numbers are specified as stochastic variables. The computational procedure involves one-observation likelihood evaluations and simulation from mixture distributions. The flexibility of the approach is illustrated on nonlinear and discrete time series models, and models with restrictions on the parameter space. Applications cover forecasting employment and gross domestic product and provide decisive empirical evidence for changes in parameter values. The contents of this chapter are based on Van den Hauwe *et al.* (2015).

The joint distribution of macroeconomic activity and a leading economic variable is estimated using a Bayesian nonparametric approach in Chapter 5. The data indicate the *ex ante* unknown form of cross-sectional dependence structure, which is operationalized by the flexibility of multivariate mixture modeling. The implied conditional distribution allows for an examination of any nonlinearities between the predictor and the economic activity measure. Further nonstandard statistical characteristics as heteroskedasticity and nonnormality of economic shocks are checked for by comparing the model to parametric threshold-regression alternatives. Applying both TCB's fully revised and real-time composite leading indexes as predictors for economic activity during the period January 1965–December 2013, decisive empirical evidence is found in favor of the nonparametric model. Not merely nonlinearity and heteroskedasticity find support in the data, but nonnormality, caused by a heavier left tail which assigns more weight to extreme negative economic shocks, as well. Chapter 5 is based on Van den Hauwe (2015).

CHAPTER 2

Federal Funds Target Rate Decisions and Model Averaging

2.1 Introduction

The federal funds target rate is one of the key monetary policy instruments of the Federal Reserve. As it signals the current stance of monetary policy, the federal funds rate is commonly considered to be an important indicator of the state of the U.S. economy. Not surprisingly then, decisions concerning the target rate as made by the Federal Open Market Committee (FOMC) are closely watched by investors, firms, and other economic agents. Likewise, speeches, interviews, and other types of communication by FOMC members are routinely scrutinized for information about future target rate decisions. Surprises in target rate decisions have been documented to have a pronounced impact on financial markets, see Bernanke and Kuttner (2005), Faust *et al.* (2007), Andersen *et al.* (2007), and Chulia-Soler *et al.* (2010), among many others.

Federal funds target rate decisions are made by the FOMC during their meetings held approximately every six weeks, and are the outcome of a complicated decision-making process. The target rate is set as a guideline for the Federal Reserve's open market operations, that is, purchases and sales of U.S. Treasury and federal agency securities, which is one of the Fed's principal tools for implementing its monetary policy (in addition to the discount window and reserve requirements). Numerous economic indicators are closely monitored by the FOMC, in order to determine the most appropriate course of action. Most attention is believed to be paid to inflation (in deviation from a target) and the output gap, in accordance with the main goals of the Federal Reserve's monetary policy to promote price stability and maximum sustainable output growth and employment (Federal Reserve Board of Governors, 2005), as formalized in the Taylor rule, see Taylor (1993). The minutes of FOMC meetings indicate, however, that a large number of other economic variables, reflecting developments in the labor market, housing market, and financial markets, also play a substantial role in the considerations.

The aim of this chapter is to assess which macroeconomic and financial variables are most informative for the FOMC's federal funds target rate decisions from a forecasting perspective. We analyze the 157 target rate decisions made during the period January 1990–June 2008, and consider a set of 33 possible predictors. The variables in this set are classified into three categories. First, we include recent realizations of inflation, output and unemployment, based on the fact that these macroeconomic variables are most directly related to the Federal Reserve's monetary policy goals. Second, we examine the information embedded in several other macroeconomic and financial variables. Most of the macro variables in this group are established leading indicators, providing signals about future economic developments that are potentially useful for predicting FOMC decisions. Similarly, the forward-looking nature of asset prices such as stock prices and interest rates has been shown to result in predictive ability for macro variables like output and inflation, see Stock and Watson (2003), among others. It seems natural to examine whether this also holds for FOMC target rate decisions. Third, we include survey measures of consumer confidence and expectations, as well as professional forecasts of inflation, output and interest rates. This is motivated by the results in Ang *et al.* (2007), who demonstrate that survey-based measures and forecasts outperform macro variables and asset prices in forecasting inflation.

We should emphasize at the outset that we do not aim to determine which variables are leading in the actual FOMC deliberations; we do not intend to get into the heads and minds of the FOMC members. Rather, we take the perspective of an "outsider" who is merely interested in forecasting upcoming FOMC decisions. For this reason, for example, we do not include the Greenbook forecasts as these would not be available in real-time to outsiders.

We employ ordered probit models for our analysis, to take into account that actual target rate decisions are discrete, in the sense that, with few exceptions, they occur in multiples of 25 basis points. We limit ourselves to modeling and forecasting the sign of the target rate decisions (or, in economic terms, the direction of monetary policy), making no distinction between changes of different magnitudes.

Federal funds target rate decisions of a given type come in clusters. On the one hand, this feature often is interpreted as a sign of interest rate smoothing by the FOMC. Several reasons for this "inertia" in monetary policy decisions have been put forward. These include uncertainty about the true structure of the economy and uncertainty about the accuracy of initial data releases of important macro variables, in particular output and employment. A third reason is the possibility to influence market expectations if monetary policymakers demonstrate that they (are willing to) implement a consistent, long-run interest rate policy, see Dueker and Rasche (2004) for a discussion. On the other hand, the clustering of target rate changes may also reflect persistence in shocks to the macroeconomic variables that drive monetary policy in general and the FOMC decisions in particular. Rudebusch (2002) suggests that much of the evidence for interest rate smoothing may in fact be the result of omitting some relevant determinants of target rate decisions from the model. As it is not really possible to include all relevant variables, empirically it is difficult to distinguish between policy inertia and persistence of macroeconomic shocks. While this distinction is important for the economic interpretation of the clustering of similar target rate decisions, from a forecasting perspective the true explanation is less crucial. Following Rudebusch (2002), we simply allow for autocorrelated errors to capture the temporal dependence in the FOMC decisions. In the empirical analysis we compare this dynamic ordered probit model to a static version, in order to assess the importance of explicitly accounting for the observed clustering.

FOMC target rate decisions, and possible determinants thereof, have been studied previously by means of (dynamic) ordered probit and logit models by Dueker (1999), Vanderhart (2000), Hamilton and Jordà (2002), Dueker and Rasche (2004), Hu and Phillips (2004), Piazzesi (2005), Grammig and Kehrle (2008), Kim *et al.* (2009), Hayo and Neuenkirch (2010), Monokroussos (2011), Scotti (2011) and Kauppi (2012), among others. We make three contributions to the existing literature. First, previous studies consider a predetermined set of explanatory variables. We, however, explicitly address the question which macroeconomic and financial variables bear most predictive content for target rate decisions. This is made possible by adopting a Bayesian approach for inference and forecasting. In particular, we employ the endogenous variable selection approach developed by Kuo and Mallick (1998), which for each candidate predictor renders a probability that it should be included in the forecasting model. Alternatively, we can interpret this procedure as a form of Bayesian model averaging. In terms of forecasting, model uncertainty is accounted for by averaging across different models, based on the posterior probabilities of inclusion of the different predictors. The Bayesian approach also facilitates to specify the dynamic probit model for FOMC decisions in calendar time. Months without a decision can be considered as missing observations which can easily be dealt with in a Bayesian setting.

Second, while most of the previous literature on target rate decisions only considers the in-sample fit of ordered probit (or other) models, we explicitly focus on out-of-sample forecasting. For this purpose, we update the parameter beliefs each time the outcome of a new FOMC meeting is observed by employing a recursive importance sampling scheme. As we integrate with respect to these updated posterior beliefs, our real-time forecasts do not only account for model uncertainty but also for parameter uncertainty.

Third, we take care to construct our probability forecasts of the target rate decisions in real time, in order to mimic the FOMC decisionmaking as realistically as possible. In addition to the recursive updating of the parameter beliefs mentioned above, this means that we account for the fact that many macroeconomic variables are revised after their initial release. Hence, the values of variables like output, employment and inflation as they are available to us now are not necessarily the same as those that were available to the FOMC members in the past at the time of their target rate decisions, nor to outside forecasters at the time of their predictions of the FOMC decisions. Neglecting this aspect may crucially affect the results of historical forecasting exercises as conducted here, see Diebold and Rudebusch (1991a), Rudebusch (2001), Stark and Croushore (2002), and Croushore (2006, 2011), among others. In order to address this issue we put together a real-time data set for the macroeconomic variables that we consider as possible predictors of the target rate decisions. This data set enables us to use the values of these variables as they actually were available to the forecaster (and the FOMC) at each point in time.

Our main empirical results are as follows. First, we find that measures of recent (changes in) economic activity like the output gap and industrial production, which are closely linked to the Federal Reserve's monetary policy goals, indeed have substantial predictive content for the FOMC target rate decisions. Perhaps surprisingly, the predictive ability of past inflation is much more limited. The most important individual predictor is the spread between the six-month T-bill rate and the effective federal funds rate, reflecting the forward-looking nature of the term structure. In agreement with Ang *et al.* (2007), we find that survey-based measures and forecasts contribute to forecasting the FOMC decisions. This holds in particular for consumer confidence, expectations about the labor market and term structure forecasts. Survey forecasts of inflation, however, do not bear useful predictive information. Second, the dynamic ordered probit model with endogenous variable selection correctly predicts 90 percent of the target rate decisions in-sample (during the complete sample period January 1990–June 2008) and 82 percent during the out-of-sample period January 2001–June 2008. It is crucially important to incorporate the clustering of similar target rate decisions in the model. Compared to the hit rates of 90 and 82 percent for the in-sample and out-of-sample forecasts achieved by the dynamic model, the static model predicts 74 and 77 percent of the target rate decisions correctly. Finally, using real-time data instead of fully revised data (as available at the time of the analysis) does not lead to any reduction whatsoever in forecasting performance. Using fully revised data, the dynamic probit model produces forecasts that result in hit rates of 89 and 81 percent.

The outline of the chapter is as follows. In Section 2.2 we describe the dynamic ordered probit model with endogenous variable selection for the FOMC target rate decisions. Subsequently we discuss the Bayesian procedure for inference and real-time forecasting. In Section 2.3 we summarize the main features of the federal funds target rate during the sample period January 1990–June 2008, and introduce the data set of candidate predictors and its real-time properties. We discuss the posterior and forecasting results in Section 2.4, and conclude in Section 2.5.

2.2 Methodology

In this section we discuss the dynamic probit model that we use to describe the discrete federal funds target rate changes. Section 2.2.1 deals with model specification. We opt for a Bayesian approach for inference and forecasting. We consider the prior for the model parameters in Section 2.2.2 and discuss the blocks of our posterior simulation scheme in Section 2.2.3. Finally, in Section 2.2.4, we outline our procedure for obtaining real-time Bayesian forecasts. Full details of the computational procedures for posterior simulation and forecasting are provided in Appendix 2.A.

2.2.1 Model specification

Federal funds target rate decisions are (mostly) made during the scheduled meetings of the FOMC. These are held approximately every six weeks, in total eight times per year. Our dynamic probit modeling framework is not specified in "meeting time" (as in Hamilton and Jordà, 2002 and Hayo and Neuenkirch, 2010) though, but in calendar time with a monthly frequency, as in Hu and Phillips (2004) and Kim *et al.* (2009). A crucial difference with these studies, which rely upon static models, discarding months without an FOMC meeting is not the natural way to proceed in our set-up with dynamic probit models. We discuss this issue in detail below.

We define r_t as the prevailing target rate at the end of month t for $t = 1, \dots, T$. As we aim to model the direction of FOMC target rate decisions we take the sign of $\Delta r_t = r_t - r_{t-1}$ as our variable of interest. We construct the discrete dependent variable y_t according to the classification

$$y_t = \begin{cases} 1 & \text{if } \Delta r_t < 0 \text{ (target rate decrease)}, \\ 2 & \text{if } \Delta r_t = 0 \text{ (no change)}, \\ 3 & \text{if } \Delta r_t > 0 \text{ (target rate increase)}. \end{cases} \tag{2.1}$$

This variable summarizes the target rate decisions made by the FOMC during its meetings and, therefore, is necessarily missing if there is no meeting in month t. To describe the discrete and ordinal nature of y_t, we introduce a latent continuous variable r_t^* that drives the classification, see Albert and Chib (1993) and Geweke (2005), for example. The introduction of the latent variable has two justifications. First, it establishes the link of y_t with potential explanatory variables. Second, we can conveniently interpret r_t^* as the publicly unobserved target rate that is central to the FOMC in choosing its policy. With our application in mind, if in month t the (absolute) deviation between the previously announced target rate r_{t-1} and the latent "desired" target r_t^* becomes too large, the target rate is adjusted. Formally this decision rule, providing the link between r_t, y_t and r_t^*, becomes

$$y_t = j \quad \Leftrightarrow \quad r_t^* - r_{t-1} \in (\alpha_{j-1}, \alpha_j], \qquad (j = 1, 2, 3), \tag{2.2}$$

in which the α_j's are threshold parameters satisfying the restriction $-\infty = \alpha_0 < \alpha_1 < \alpha_2 < \alpha_3 = \infty$. If, for example, r_t^* is higher than r_{t-1} by an amount that exceeds α_2, the target rate is increased in month t such that $r_t > r_{t-1}$.

For predicting future FOMC decisions we specify the way r_t^* relates to macroeconomic and financial information available at the moment of constructing the forecast. We assume a linear relation between the unobserved target rate r_t^* and covariates summarized in the vector $\mathbf{x}_t = (x_{1t}, \ldots, x_{Kt})'$, such that

$$r_t^* = \boldsymbol{\beta}'\mathbf{x}_t + u_t, \tag{2.3}$$

in which $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$ and $\{u_t\}$ is an unobserved random process. We stress that in the implementation we require that information included in $\mathbf{x}_t$ is predetermined at the end of month $t-1$ (including publication lags), such that the model can indeed be used for real-time prediction of the target rate decision in month t. We note that our approach is conservative, in the sense that the FOMC may have had more recent information at its disposal when making the target rate decision in month t. We, however, can produce the forecast earlier in time.

The FOMC monitors a large number of macroeconomic and financial variables to guide its target rate decisions, in addition to inflation and output measures that are directly related to monetary policy objectives. This raises the question which economic indicators are the most useful for predicting the FOMC decisions. We incorporate this uncertainty with respect to the exact content of the vector $\mathbf{x}_t$ in our analysis of the ordered probit model. This approach has two advantages. First, it allows us to do inference on which economic variables are important predictors of the Fed's monetary policy decisions. Second, when constructing forecasts of future target rate decisions, we can account for this model uncertainty by averaging over different models containing different combinations of predictor variables.

We follow the approach of Kuo and Mallick (1998) because it is computationally easy to implement and does not require extensive tuning as, for example, the method proposed by George and McCulloch (1993). We have K potential predictors x_{kt}, $(k = 1, \ldots, K)$. For describing the selection of these covariates in the model we introduce K additional binary parameters γ_k, indicating whether the k-th variable is included in the model $(\gamma_k = 1)$ or not $(\gamma_k = 0)$. Effectively, we decompose the regression parameters β_k in (2.3) as $\beta_k = \psi_k \cdot \gamma_k$ with ψ_k denoting the effect of x_{kt} on the latent target rate when it is included in the model. Modeling the uncertainty regarding relevant predictors this way boils down to determining which of the 2^K different predictor combinations are most likely to have predictive power for the target rate decisions.

The final part of our model specification concerns the distributional assumptions for u_t in (2.3).[1] We allow for temporal dependence in u_t in order to capture the persistence in FOMC decisionmaking. In particular, we assume the Gaussian first-order autoregressive structure

$$u_t = \varphi u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1). \tag{2.4}$$

A similar type of persistence is considered plausible by Rudebusch (2002). This structure allows for temporal dependence in the latent target rate r_t^* in (2.3), in addition to the temporal dependence due to autocorrelation in the predictor variables $\mathbf{x}_t$. We assume that the shocks ε_t are normally distributed and independent of the predictors $\mathbf{x}_s$, $(s \leq t)$. Their effect on the latent target rate dies out exponentially over time, which is in contrast with the dynamic ordered probit models of Eichengreen *et al.* (1985) and Dueker (1999), because they impose the latent interest rate to be integrated. In the empirical application, we compare the dynamic ordered probit model with a static model, obtained by setting $\varphi = 0$ in (2.4).

Finally, two remarks are in order. First, our dynamic probit model is specified in calendar time. About one third of our sample period concerns months without an FOMC meeting. In these months, we do not observe an outcome y_t as defined in (2.2). However, we do know that the unobserved target rate behaves like $r_t^* \,|\, \{\boldsymbol{\theta}, r_{t-1}^*\} \sim \mathcal{N}(\boldsymbol{\beta}'\mathbf{x}_t + \varphi(r_{t-1}^* - \boldsymbol{\beta}'\mathbf{x}_{t-1}), 1)$, according to (2.3)–(2.4). For example, if in month $t-1$ the FOMC decides to increase the target rate, in t there is no meeting and in $t+1$ they keep the target rate constant, the persistence in the latent target rate suggests an "interpolated" value for r_t^*. However, it is not restricted to be in one of the three bandwidths. The availability of both the interpolated latent target rate of month t and predictors $\mathbf{x}_t$, make these months informative about $\boldsymbol{\beta}$ (if and only if $\varphi \neq 0$). We note that this situation is different from the one in which month t does have an FOMC meeting but it is decided to keep the target unchanged. According to the model specification in (2.2), in such a case the difference between the latent target rate r_t^* and the prevailing actual target rate r_{t-1} must have realized in the range $(\alpha_1, \alpha_2]$ and this truncates r_t^*.

At this point it is also useful to note that another possibility to incorporate dynamics in the probit model is to add y_{t-1} to the regressors $\mathbf{x}_t$ in (2.3). We note that this variable, representing the actual target rate decision in month $t-1$, is not observed if the FOMC has no meeting in the previous month. Dealing with these missing values in a proper way turns out to be more complicated with respect to simulating r_t^* in our posterior sampler, especially if there are no FOMC meetings several months in a row. Hence, we opt for the more elegant AR(1) specification for u_t in (2.4).

Second, because we do not exclude months without meetings, neither in the model estimation, nor in our forecasting experiment discussed later, our model provides monthly predictions. Hence, due to our sample selection, every forecast we construct is necessarily conditioned on the event of a meeting and each forecasting statement should start with saying: "If the FOMC organized a meeting next month, the forecast for its outcome would be ...."

[1] If we consider r_t^* as the latent target rate which directs monetary policy, we can adopt the economic interpretation of Bernanke and Blinder (1992) and label u_t as an unanticipated monetary policy shock as opposed to the anticipated part captured by $\boldsymbol{\beta}'\mathbf{x}_t$.

2.2.2 Prior distribution

We adopt a Bayesian approach for inference and forecasting in the dynamic probit model with endogenous variable selection. Therefore we have to specify prior beliefs about the parameter values in (2.2)–(2.4).

For the threshold parameters α_j we opt for a flat (improper) prior, not imposing any kind of asymmetry that would *a priori* favor a particular target rate adjustment category. However, due to the ordering of the categories of y_t, we do restrict this prior to the region $\{(\alpha_1, \alpha_2)' \in \mathbb{R}^2 \;:\; \alpha_1 < \alpha_2\}$.

Our prior for the autoregressive parameter φ is such that stationarity of the u_t process is guaranteed. We take the truncated normal distribution $\varphi \sim \mathcal{N}(b, B) \times \mathbb{I}_{\{\varphi \in S\}}$ with $S = (-1, 1)$. Obviously, for large B the prior for φ becomes a uniform distribution on the region defined by S. We take $b = 0$ and $B = 10$. If r_t^* would be integrated (as is imposed in Eichengreen *et al.*, 1985 and Dueker, 1999), we should find the posterior mode of φ at one.

The regression parameters in (2.3) deserve extra attention due to the endogenous variable selection procedure. As discussed before, we decompose $\boldsymbol{\beta}$ into the product of inclusion indicators $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)'$ and regression effects $\boldsymbol{\psi} = (\psi_1, \dots, \psi_K)'$. We take the inclusion of the k-th predictor to be *a priori* independent of the inclusion of the other $K-1$ variables. Consequently, the prior for each γ_k is a Bernoulli distribution with probability of success π_k, *i.e.*, $\gamma_k \,|\, \pi_k \sim \mathcal{B}er(\pi_k)$. Alternatively we could follow Chipman (1996) to find a prior distribution that incorporates a dependence structure for variable inclusion. However, such an alternative would force us to think *a priori* about the predictors in terms of complements and substitutes. We prefer a jointness analysis *a posteriori*.

A prior on the π_k's implies a prior on the model size. The latter is defined as the number of included covariates given by $N(\boldsymbol{\gamma}) = \sum_{k=1}^{K} \gamma_k$. We restrict the prior inclusion probabilities to be equal, that is, we set $\pi_k = \pi$ for $k = 1, \dots, K$. We opt for the conditionally conjugate Beta prior $\pi \sim \mathcal{B}e(c_1, c_2)$. In this setting the prior expected model size is equal to $Kc_1/(c_1 + c_2)$.[2] We make the prior on the inclusion probability uninformative in the sense that we consider all values $\pi \in [0, 1]$ to be equally likely (which is achieved by setting $c_1 = c_2 = 1$). As a result, the prior distribution of the model size is, with a variance of 96.25, quite dispersed as well.[3] Some researchers prefer smaller models, therefore we experiment with more stringent values of the c_i. In our case it turns out that sample information strongly dominates the inclusion prior though.

One of the advantages of Kuo and Mallick's (1998) approach is the specification of the prior for $\boldsymbol{\psi}$. It is specified independent of the inclusion indicators $\boldsymbol{\gamma}$, in contrast to the mixture of normals idea of George and McCulloch (1993). This facilitates the posterior simulation considerably and has no undesirable model restrictions. We use a conditionally conjugate Gaussian prior for $\boldsymbol{\psi}$ with mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$, that is, $\boldsymbol{\psi} \sim \mathcal{N}(\mathbf{a}, \mathbf{A})$. The conditional conjugacy of this prior allows us to integrate the regression effects analytically. As we demonstrate below, in our simulation scheme to obtain posterior results (based on the Gibbs sampler), $\boldsymbol{\gamma}$ is sampled after integrating with respect to $\boldsymbol{\psi}$. For the regression effects it

[2] If we notice that $N(\boldsymbol{\gamma}) \,|\, \pi \sim \mathcal{B}in(K, \pi)$ and apply the law of iterated expectations, we compute the prior expected model size as $\mathsf{E}[N(\boldsymbol{\gamma})] = \mathsf{E}_\pi \mathsf{E}_{\boldsymbol{\gamma}\,|\,\pi}[N(\boldsymbol{\gamma})] = \mathsf{E}_\pi[K\pi] = K\bar{\pi}$, with $\bar{\pi} = \mathsf{E}[\pi] = c_1/(c_1 + c_2)$.

[3] In a fashion similar to deriving the first moment, we compute the second moment of $N(\boldsymbol{\gamma})$, and after some manipulation we obtain the prior variance of model size $\mathsf{Var}[N(\boldsymbol{\gamma})] = \mathsf{E}\left[N(\gamma)^2\right] - \mathsf{E}[N(\gamma)]^2 = K\{\bar{\pi}(1-\bar{\pi}) + (K-1)\mathsf{Var}[\pi]\}$. With the variance of the Beta distribution being $\mathsf{Var}[\pi] = c_1c_2/[(c_1 + c_2 + 1)(c_1 + c_2)^2]$. To provide a benchmark value for the variance in our case: the most informative Beta prior with mean 0.5 gives the lower bound of the variance of 8.25.

holds that if $\gamma_k = 0$, the corresponding ψ_k is not identified by the data and is approximately (exactly if $\mathbf{A}$ is diagonal) sampled from its prior distribution. Kuo and Mallick (1998) recommend to make the prior relatively uninformative and choose a diagonal prior covariance matrix $\mathbf{A}$ with elements equal to 16 if the explanatory variables are standardized. We adopt their recommendation and set $\mathbf{a} = \mathbf{0}_K$ and $\mathbf{A} = 16\mathbf{I}_K$, though we have experimented with different values. Provided that these prior variances do not take extreme values, we find that posterior results are robust.

2.2.3 Posterior simulation

We obtain posterior results by using Markov chain Monte Carlo (MCMC) methods, see Tierney (1994) and Robert and Casella (2004). The latent variables $\mathbf{r}^* = (r_1^*, \ldots, r_T^*)'$ are simulated alongside the model parameters $\boldsymbol{\theta} = \{\alpha_1, \alpha_2, \pi, \varphi, \boldsymbol{\psi}, \boldsymbol{\gamma}\}$, see Tanner and Wong (1987) and Albert and Chib (1993). The posterior density of the parameters and latent variables after having observed the sample $\mathbf{y} = (y_1, \ldots, y_T)'$ is given by

$$p(\boldsymbol{\theta}, \mathbf{r}^* \,|\, \mathbf{y}) \propto p(\boldsymbol{\theta}) p(\mathbf{y}, \mathbf{r}^* \,|\, \boldsymbol{\theta}),$$

in which $p(\boldsymbol{\theta})$ and $p(\mathbf{y}, \mathbf{r}^* \,|\, \boldsymbol{\theta})$ are the prior density of the model parameters and the complete data likelihood function of the model, respectively.

It is well known that Gibbs sampling in ordered probit models may suffer from mixing difficulties, especially if some form of temporal dependence in the latent variables is allowed as in (2.4), see Liu and Sabatti (2000), for example. The main reason for this is that the sampled threshold parameters display high correlation. We solve this issue partly by fixing the lower threshold in (2.2) at zero and at the same time including an intercept in (2.3). Removing one of the threshold parameters from the specification clearly improves the mixing of the sampler. We only use this reparameterization for posterior simulation. For inference, results are easily converted back into the familiar, interpretable model specification. The MCMC simulation scheme to sample from the posterior consists of the following six steps.

Step 1. Sample the threshold α_2 from its full conditional posterior (uniform distribution);[4]

Step 2. Sample φ given the other parameters and $\mathbf{r}^*$ with a Metropolis–Hastings sampler as suggested by Chib and Greenberg (1994) (proposal values from truncated normal distribution);

Step 3. Sample π from its full conditional posterior (Beta distribution);

Step 4. Sample the latent target rate r_t^* from its full conditional posterior for $t = 1, \ldots, T$, as in Girard and Parent (2001) (truncated normal distributions);

Step 5. Sample γ_k for $k = 1, \ldots, K$ in a random order, from its conditional posterior after marginalization with respect to $\boldsymbol{\psi}$, see Kuo and Mallick (1998) (Bernoulli distributions);

Step 6. Sample $\boldsymbol{\psi}$ from its full conditional posterior, see also Kuo and Mallick (1998) (multivariate normal distribution).[5]

[4] We note that we sample the transformed upper threshold, but given the intercept we have a draw of the original upper threshold as well.

[5] Due to the reparameterization, in this step we simulate the regression parameters *and* the intercept in one block from their joint multivariate normal distribution.

The first step proceeds in a similar fashion as for a standard (static) ordered probit model, see *e.g.* Albert and Chib (1993). Since we do not condition on the first observation, we cannot sample the autoregressive parameter from a truncated normal distribution. Therefore, in Step 2 we implement the Metropolis–Hastings sampler to simulate φ as suggested by Chib and Greenberg (1994). In Step 3, the full conditional posterior of π reduces to $\pi \,|\, \boldsymbol{\gamma} \sim \mathcal{B}e(c_1{+}N(\boldsymbol{\gamma}),\, c_2{+}K{-}N(\boldsymbol{\gamma}))$. Finally, to take care of the dynamics in the latent variables and the treatment of the first observation, we rewrite the conditional model in linear regression form and execute the next three steps of the simulation scheme. We provide details of our posterior simulation procedure in Appendix 2.A.2.

2.2.4 Forecasting

To construct real-time Bayesian forecasts of the FOMC target rate decisions that account for both model and parameter uncertainty, we derive the posterior predictive distributions. The one-step-ahead predictive distribution of y_s, derived at the end of month $s-1$, is given by its probability density function

$$p(y_s \,|\, \mathbf{y}^{1,s-1}) = \int p(y_s \,|\, \mathbf{y}^{1,s-1}, \mathbf{r}^{*;1,s-1}, \boldsymbol{\theta}) p(\mathbf{r}^{*;1,s-1}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1}) \, \mathsf{d}\{\mathbf{r}^{*;1,s-1}, \boldsymbol{\theta}\}, \tag{2.5}$$

in which $\mathbf{v}^{n,m} = (v_n, \ldots, v_m)'$, $(1 \leq n < m \leq T)$. Given draws from the posterior distribution with pdf $p(\mathbf{r}^{*;1,s-1}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1})$, we use Monte Carlo integration to solve the integral in (2.5). We note that we can analytically compute $p(y_s \,|\, \mathbf{y}^{1,s-1}, \mathbf{r}^{*;1,s-1}, \boldsymbol{\theta}) = p(y_s \,|\, r^*_{s-1}, \boldsymbol{\theta})$ by evaluations of the normal cumulative distribution function. Because $\boldsymbol{\theta}$ contains the predictor inclusion parameters $\boldsymbol{\gamma}$, each draw from the posterior also indicates a particular selected combination of predictors. According to this principle, forecasts by models which are *a posteriori* considered more likely get a larger weight in the “model-free” predictive distribution.

For model evaluation we gauge the predictive power by considering one-step-ahead forecasts. We compute these forecasts for the last $T - \tau$ observations, that is, for months $s = \tau + 1, \ldots, T$. For every prediction we incorporate all sample information revealed by that date. Therefore we update our posterior distribution for every forecast we make.

We limit the computational burden by using importance sampling methods. The forecast for y_s is based on all information available up to and including month $s-1$. Hence, based on draws from $p(\mathbf{r}^{*;1,s-1}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1})$ we construct the forecast for month s according to (2.5). Before computing the next one-step-ahead forecast, for y_{s+1}, we incorporate the additional information revealed by the latest target rate decision y_s to update the posterior beliefs. We avoid rerunning the entire MCMC sampler by using importance sampling with the posterior from the previous period, $p(\mathbf{r}^{*;1,s-1}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1})$, as importance function. To construct the forecast for y_{s+1} we actually need a sample from $p(\mathbf{r}^{*;1,s}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s})$, but our importance function does not provide a draw for r^*_s. We solve this issue by extending the posterior from the previous period with $p(r^*_s \,|\, \mathbf{y}^{1,s}, \mathbf{r}^{*;1,s-1}, \boldsymbol{\theta})$ and, hence, we obtain r^*_s by simulating from a truncated normal distribution. In Appendix 2.A.3 we describe this forecasting procedure in full detail and show how to derive the importance weights.

2.3 Data and target rate characteristics

We investigate the federal funds target rate at a monthly frequency for the period January 1990–June 2008. It can be argued that during this sample period, covering most of

Greenspan's term as chairman of the Board of Governors plus the start of Bernanke's reign, the Federal Reserve's monetary policy objectives have been kept constant.[6] We build on this by assuming that the set of macroeconomic and financial variables that are most informative for predicting target rate changes (possibly because they are most closely monitored by the FOMC) have been the same throughout this period. Besides, a structural break in the parameters would be hard to identify given the limited time span of the (clustered and discrete-valued) data.

The FOMC meets eight times per year at previously set dates. Our sample period, which covers 222 months in total, contains 148 months with such a scheduled meeting. *Un*scheduled meetings are held occasionally (sometimes by conference call), if required by sudden unexpected economic developments or other major events affecting the economy. In addition to the 148 months with scheduled meetings, we observe nine months with an FOMC target rate decision not made during a scheduled meeting.[7] We retain these months in our sample and in fact consider them identical to months with regular meetings.[8]

The announced target rates are displayed in Figure 2.1, with summary statistics being provided in Table 2.1. During the sample period the target rate varied considerably, between a minimum of 1.00 percent in 2003–04 and a maximum of 8.25 percent during the first half of 1990. Figure 2.1 clearly shows that decisions of the same type appear in clusters. For example, periods of sustained declines of the target rate occurred in the early nineties, in 2001, and during the final year of the sample period. To a large extent these target rate declines coincide with U.S. recessions as declared by the National Bureau of Economic Research (NBER) and shown in Figure 2.1 by the shaded areas. Similarly, multiple consecutive decisions to increase the target were made during 1994, 1999, and the period from mid-2004 till mid-2006. These clusters can be captured by our model via two mechanisms. First, they may be due to temporal dependence in the predictor variables and, second, conditional on the predictors interest rate smoothing allows for additional persistence in the decisions.

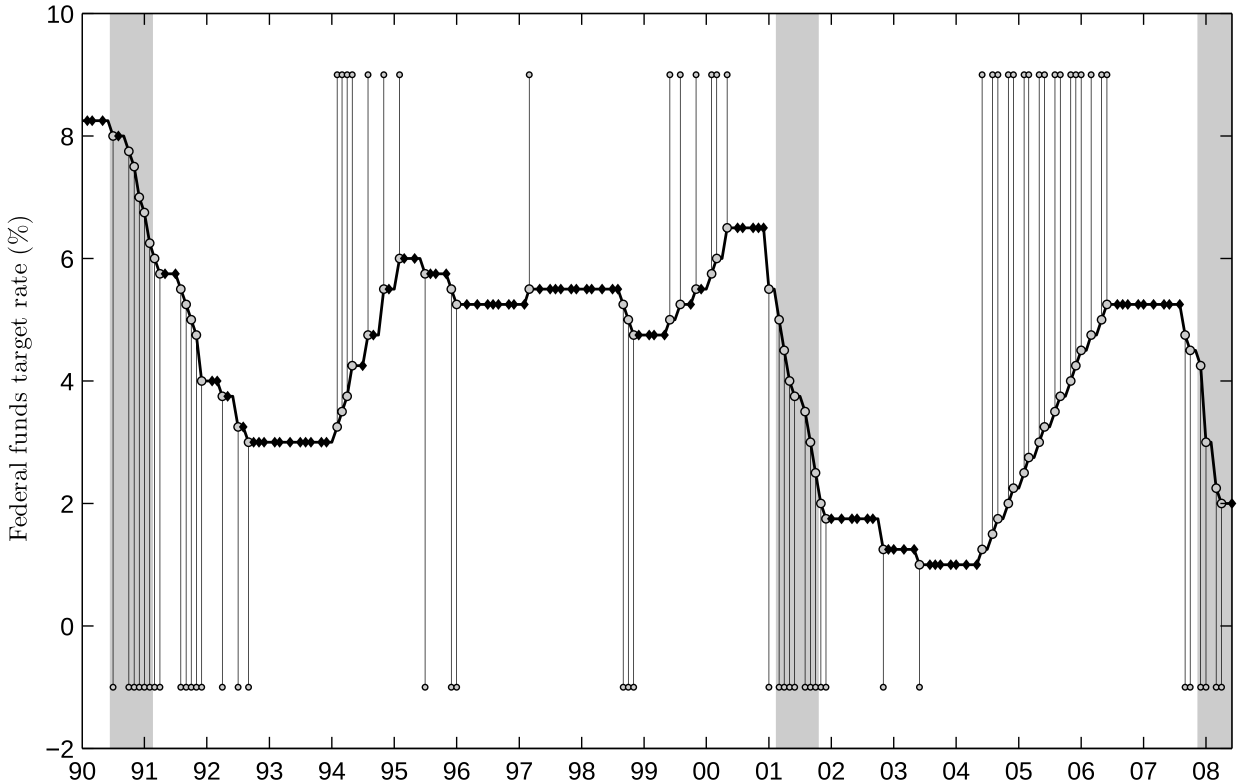
Our data set further consists of a set of macroeconomic and financial variables that are considered potential predictors for the FOMC decisions. These variables can be categorized in three groups. The first group comprises measures related to inflation, output and (un)employment. These variables are most closely related to the monetary policy objectives of the Federal Reserve and for this reason might be expected to play a key role in the FOMC decision-making process. The second group of variables consists of recent realizations of several other macro and financial variables that provide information on economic developments. Most of the variables in this group, such as new orders and building permits, have a forward-looking character. As discussed in the introduction, the FOMC considers multiple

[6]During our sample period, substantial changes did occur in Federal Reserve's operating procedures and communication policy (see Poole, 2005, Ehrmann and Fratzscher, 2007, among others).

[7]Specifically, these months cover decisions made on January 9, 1991; April 30, 1991; 13 September, 1991; April 9, 1992; September 4, 1992; October 15, 1998; April 18, 2001; and September 17, 2001 (target rate decreases) and on April 18, 1994 (target rate increase). There are also months with both unscheduled and scheduled meetings. For example, an unscheduled target rate decrease was announced on January 22, 2008, but this was followed later that month by another target rate cut following the scheduled meeting on January 29–30. We include such months among the 148 scheduled months.

[8]We have examined an alternative definition of the dependent variable y_t in (2.1), which has observations if and only if there is a *scheduled* meeting in month t. Months with an unscheduled meeting are not omitted from the sample, but y_t is again treated as missing in such a case. Estimation results for this specification are not much different from those reported in Section 2.4. In particular, similar predictor variables are considered relevant in forecasting the target rate decisions and estimated thresholds do not indicate substantial differences.

Figure 2.1 *Federal funds target rate and FOMC decisions, January 1990–June 2008*



Notes: The graph shows monthly observations of the federal funds target rate (black graph), with circles indicating the FOMC decisions to change the target rate and diamonds representing no-change decisions. The shaded areas correspond with U.S. recessions, according to the NBER business cycle turning points.

Table 2.1 *Summary statistics*

FOMC decisions		Federal funds target rate	
# Decreases	40	Mean	4.35
# No-changes	86	Minimum (June 2003–May 2004)	1.00
# Increases	31	Maximum (January 1990–June 1990)	8.25
		Standard deviation	1.85

Notes: The table presents summary statistics for the federal funds target rate and the 157 FOMC decisions during the period January 1990–June 2008 (222 months).

economic indicators in its deliberations, reflecting developments in financial markets, the labor market, and the housing market, among others. Including this second group of variables allows us to examine whether these indicators provide any supplemental information for predicting target rate decisions, in addition to inflation and output. The third group of variables consists of survey-based measures of consumer/purchasing managers' confidence and forecasts of inflation, output and interest rates. The latter are taken from the Survey of Professional Forecasters conducted by the Federal Reserve Bank of Philadelphia, see Croushore (1993). Including this group of variables is motivated, among others, by Ang *et al.* (2007) and Campbell and Diebold (2009), who demonstrate that survey-based measures provide more accurate out-of-sample forecasts for inflation and stock returns than historical macroeconomic and financial variables. Koenig (2002) establishes an empirical relation between the purchasing managers' index and monetary policy.[9]

An important feature of the macroeconomic variables in the first and second group is that they are subject to revisions after their initial release. As a consequence, the currently available time series is different from the one that was at the FOMC members' disposal at the time they met. These revisions can be substantial, in particular for output- and employment-related variables. Diebold and Rudebusch (1991a) and Stark and Croushore (2002), among others, analyze the consequences and potential pitfalls in case a forecasting experiment is conducted with latest-available data instead of variables measured in real time, see also the recent surveys of Croushore (2006, 2011). We note that the survey-based measures in the third group of predictors are not subject to revisions, which can be considered an additional advantage of this type of variables. In order to make our empirical analysis as realistic as possible, we employ data as available on a real-time basis. For this purpose, we combine data taken from the ALFRED database of the Federal Reserve Bank of St. Louis, the Federal Reserve Bank of Philadelphia's real-time database and real-time data for the components of The Conference Board's leading economic index.[10] Our resulting "real-time vintage data" (Koenig *et al.*, 2003) is set up such that $\mathbf{x}_t$ contains the observations on the covariates as they were historically available at the end of month $t-1$, when the forecast for the target rate decision in month t was made.

Table 2.2 lists the complete set of 33 potential predictors we use in our analysis. Some variables are transformed, mostly by converting them to growth rates (see the final column of Table 2.2). Two further issues are worth mentioning. First, variables measured only at a quarterly frequency, such as the SPF forecasts, are transformed to monthly observations by keeping the value constant for the three months within the quarter. We justify this method by observing that during the three-month period no new information about the variable is revealed. Second, we allow for a so-called "averaging" period. That is, we take moving averages or growth rates over the m most recent available observations of the predictors. This reflects the idea that possibly the FOMC does not only focus on what has happened to a certain economic indicator in the most recent month, which can be a noisy signal, but instead also considers developments over a longer period of a few months. We experiment with different values of m to analyze the robustness of our results.

2.4 Results

This section is divided into three parts. First, we report posterior results using the full sample period January 1990–June 2008, focusing on the question which variables appear

[9]A referee pointed us at the non-manufacturing constituent of the purchasing managers' index, but since we can only reconstruct it back to July 1997, we examine the all-industries index instead.

[10]The real-time database of the Federal Reserve Bank of Philadelphia is constructed and maintained as described in Croushore and Stark (2001). We use their (monthly) vintages of GDP, but due to redefinitons of real potential GDP, we encounter problems in forming a real-time output gap. Therefore we construct our own series of real potential GDP in real time by applying a Hodrick–Prescott filter per vintage to real-time GDP.

Table 2.2 *Set of potential predictors*

Predictor's description	Ab.	Post.In.Pr. Dy.	Post.In.Pr. St.	Tr.
Panel A: Monetary policy variables				
1. Inflation, CPI: U.S. city average: All items: Seasonally adjusted	INF	0.02	0.19	*gr*
2. Output gap, Real GDP less its HP-filtered trend	OUT	0.47	0.12	*av*
3. Unemployment gap, Unemployment rate less CBO NAIRU	UG	0.05	0.04	*av*
Panel B: Other macroecnomic and financial variables				
4. The Conference Board's leading economic index	LEI	0.02	0.01	*gr*
5. Average weekly hours, manufacturing	WHM	0.05	0.02	*gr*
6. Average weekly initial claims for unemployment insurance	CUI	0.01	0.01	*gr*
7. Manufacturers' new orders, consumer goods and materials	NOC	0.02	0.01	*gr*
8. Vendor performance, slower deliveries diffusion index	VPI	0.02	0.02	*gr*
9. Manufacturers' new orders, nondefense capital goods	NOK	0.01	0.01	*gr*
10. Building permits, new private housing units	NBP	0.04	0.01	*gr*
11. Stock prices, 500 common stocks	SP5	0.01	0.01	*gr*
12. Money supply, M2	M2	0.01	0.01	*gr*
13. Interest rate spread, 10-y Treasury bonds less Fed funds	10TFF	0.05	0.97	*av*
14. The Conference Board's coincident economic index	CEI	0.34	0.02	*gr*
15. Employees on nonagricultural payrolls	ENP	0.04	0.01	*gr*
16. Personal income less transfer payments	PI	0.07	0.02	*gr*
17. Index of industrial production	IP	0.48	0.41	*gr*
18. Manufacturing and trade sales	MTS	0.01	0.02	*gr*
19. Interest rate spread, 6-month T-bill less Fed funds	6TFF	1.00	1.00	*av*
20. Capacity utilization: Manufacturing	CUM	0.07	0.04	*av*
21. Household sector: Household credit market debt outstanding	HCM	0.11	0.06	*gr*
22. Total consumer credit outstanding	TCC	0.02	0.01	*gr*
Panel C: Survey measures and professional forecasts				
23. TCB Consumer confidence	CC	0.08	0.14	*av*
24. TCB Consumer confidence: Present situation	CCP	0.52	0.13	*av*
25. TCB Consumer confidence: Expectations	CCE	0.04	0.03	*av*
26. Un. of Michigan consumer sentiment index	CS	0.03	0.09	*av*
27. Un. of Michigan consumer sentiment index: Current conditions	CSC	0.04	0.59	*av*
28. Un. of Michigan consumer sentiment index: Expectations	CSE	0.03	0.05	*av*
29. Average (mean) duration of unemployment	MDU	0.28	0.17	*av*
30. ISM Purchasing managers' index	ISM	0.02	0.10	*av*
31. Anxious index: Probability decline real GDP in following quarter	AI	0.02	0.02	*av*
32. SPF: Mean expected 3-month T-bill rate in the following 4 quarters	3TE	0.14	0.97	*av*
33. SPF: Mean expected CPI inflation rate in the following 4 quarters	INE	0.05	0.03	*av*

Notes: "Ab." reports the variable's abbreviation; "Post.In.Pr." contains the predictor's posterior inclusion probabilities for the dynamic ("Dy.") and static ("St.") model, respectively; "Tr." indicates the applied transformation, which either is the annualized growth rate $1200/m(x_t - x_{t-m})/x_{t-m}$ (*gr*), or the moving average $(x_t + x_{t-1} + \ldots + x_{t-m+1})/m$ (*av*) over the m most recent months.

to be most informative for FOMC target rate decisions. Next, we present real-time out-of-sample forecasts of the target rate decisions during the period January 2001–June 2008. Third and finally, with a number of robustness checks we examine the sensitivity of the results to the choice of priors and the transformations applied to the predictor variables.

Table 2.3 *Properties of marginal posterior distributions*

Dynamic ordered probit						Static ordered probit					
			Percentiles						Percentiles		
Param.	Mean	St.D.	5th	95th	In.Pr.	Param.	Mean	St.D.	5th	95th	In.Pr.
β_{6TFF}	2.97	0.58	2.05	3.95	1.00	β_{6TFF}	2.77	0.41	2.12	3.47	1.00
β_{CCP}	2.39	0.93	0.88	3.89	0.52	β_{10TFF}	−1.40	0.35	−1.98	−0.85	0.97
β_{IP}	0.67	0.22	0.32	1.03	0.48	β_{3TE}	1.39	0.29	0.88	1.79	0.97
β_{OUT}	1.83	0.67	0.70	2.91	0.47	β_{CSC}	0.80	0.23	0.44	1.14	0.59
β_{CEI}	0.55	0.20	0.24	0.89	0.34	β_{IP}	0.47	0.18	0.20	0.77	0.41
β_{MDU}	−1.24	0.50	−2.05	−0.41	0.27	β_{INF}	0.36	0.14	0.13	0.60	0.19
β_{3TE}	1.41	0.86	0.07	2.90	0.14	β_{MDU}	−0.74	0.36	−1.40	−0.18	0.17
β_{HCM}	−0.49	0.23	−0.86	−0.13	0.11	β_{OUT}	0.61	0.27	0.17	1.05	0.17
β_{CC}	0.98	1.64	−1.54	4.17	0.08	β_{CC}	0.72	0.49	0.06	1.26	0.14
β_{PI}	0.46	0.27	0.01	0.89	0.07	β_{CCP}	0.77	0.51	−0.01	1.41	0.13
π	0.15	0.07	0.05	0.28	–	π	0.18	0.07	0.08	0.31	–
α_1	−3.29	1.25	−5.33	−1.30	–	α_1	−1.76	0.27	−2.24	−1.36	–
α_2	3.75	1.34	1.72	6.05	–	α_2	2.06	0.30	1.63	2.60	–
φ	0.93	0.04	0.85	0.97	–						

Notes: The table shows properties of the marginal posterior distributions of the model parameters in the dynamic and static ordered probit models, obtained with the full-sample data. Properties of the posteriors of the parameters associated with the ten most relevant predictors (that is, with the highest posterior inclusion probabilities $\mathsf{Pr}[\gamma_k = 1 \,|\, \mathbf{y}]$ ["In.Pr."]) are shown conditional on the variable's inclusion: Posterior expectation $\mathsf{E}[\beta_k \,|\, \mathbf{y}, \gamma_k = 1]$ ("Mean"); posterior standard deviation $\mathsf{Var}[\beta_k \,|\, \mathbf{y}, \gamma_k = 1]^{1/2}$ ("St.D."); the 5th and 95th percentiles.

2.4.1 Posterior results

We present the key properties of the marginal posterior distributions of the dynamic ordered probit model's parameters in the left panel of Table 2.3. We obtain these results by applying the MCMC simulation scheme of Section 2.2.3 using the full sample period January 1990–June 2008 and an averaging period equal to one month ($m = 1$). For comparison, we present the posterior results for the nested static ordered probit model, that is, with the restriction $\varphi = 0$ in (2.4), in the right panel of this table. In Appendix 2.B we provide convergence diagnostics for the MCMC procedure.

Several conclusions emerge from these posterior results. First, we find strong evidence for temporal dependence in the target rate decisions that goes beyond the dependence caused by the inclusion of autocorrelated predictor variables. The posterior mean of the autoregressive parameter in the dynamic probit model is equal to 0.93. Since the posterior mass is tightly concentrated around 0.9, the effect of a shock ε_t remains noticeable for a substantial period of time. The posterior mean of 0.93, for example, corresponds with a half-life of shocks of nine months. Despite this high persistence we do not find evidence for r_t^* to be integrated. The posterior mode of φ equals 0.94 and $p(\varphi = 0.94 \,|\, \mathbf{y}) \approx 2 \cdot p(\varphi = 0.99 \,|\, \mathbf{y})$.

Second, relying on the posterior inclusion probabilities $\mathsf{Pr}[\gamma_k = 1 \,|\, \mathbf{y}]$, $(k = 1, \ldots, K)$, only a limited number of predictor variables appear to be informative for the target rate decisions. We report these probabilities for the ten most frequently sampled covariates in the rightmost columns of the panels in Table 2.3 and compare these to the posterior average inclusion probability π.[11] For the dynamic model we find that six variables have a conditional posterior inclusion probability that exceeds the posterior mean $\mathsf{E}[\pi \mid \mathbf{y}]$, which equals 0.15. These variables are the spread between the six-month T-bill rate and the effective federal funds rate (6TFF), The Conference Board's consumer confidence index: present situation (CCP), industrial production growth (IP), the output gap (OUT), The Conference Board's coincident economic index (CEI) and the expected mean duration of unemployment (MDU). 6TFF is by far the most important predictor, in agreement with Hamilton and Jordà (2002). This spread is included in the model with probability equal to one, which is almost twice as high as the inclusion probability of the second-most frequently sampled variable, The Conference Board's consumer confidence measure of present economic conditions. Although its inclusion probability of 0.14 is slightly smaller than the posterior of π, the SPF forecast of the three-month T-bill rate (3TE) also contains predictive content. Interestingly, inflation only has a posterior inclusion probability of 0.02, indicating that lagged inflation is not important for predicting FOMC decisions. Similarly, the SPF inflation expectation (INE) has a small posterior inclusion probability of 0.05.

For the static model we also find six variables for which $\mathsf{Pr}[\gamma_k = 1 \mid \mathbf{y}] > \mathsf{E}[\pi \mid \mathbf{y}] = 0.18$. These include the spread between the six-month T-bill and the effective federal funds rate, growth in industrial production and the professional forecast of the three-month T-bill rate, which also are found to be important in the dynamic model. In addition, the spread between the ten-year T-bond rate and the effective federal funds rate (10TFF), the University of Michigan's consumer sentiment index: current conditions (CSC), and the inflation rate (INF) satisfy this condition. Although their inclusion probabilities are below the posterior mean of π, the output gap, expected mean duration of unemployment, The Conference Board's consumer confidence measures (CC, CCP) and the purchasing managers' index (ISM) (see Table 2.2) contribute substantially to describing FOMC decisions.[12]

A priori we were almost uninformative about the size of the model. The posterior mean of the inclusion probability π indicates that the data are in favor of small models with about five predictors included. Furthermore, the posterior inclusion probabilities are lower for the dynamic probit model, which is due to the strong explanatory power of the autoregressive dynamics in this model. In Section 2.4.3 we discuss other Beta priors for π to check the robustness of our results.

If we exclusively focus on the marginal inclusion properties of the predictors, the interaction between covariates remains hidden, as pointed out by Doppelhofer and Weeks (2009a). In order to investigate these interactions we compute Doppelhofer and Weeks' (2009a) jointness measure for all pairs of explanatory variables $\{x_{kt}, x_{lt}\}$, $(k \neq l = 1, \ldots, K)$.[13] This measure, denoted $J_{k,l}$, accounts for both joint inclusion and exclusion of the two variables over the total model space to assess their mutual dependence. Negative values indicate that the two variables are substitutes, which means that they contain approximately the same information with respect to the target rate decisions. Positive values suggest a complemen-

[11] We display the marginal posterior inclusion probabilities for all examined predictors in Table 2.2.

[12] We also have estimated the model with the set of predictors augmented with the non-manufacturing part (NMI) of the purchasing managers' index. We can only do this for the period August 1997–June 2008, but we find results in agreement with Koenig (2002). That is, NMI is included as predictor with probability 0.27. Moreover, for this sample period the "overall" index ISM seems more important as well, with an inclusion probability of 0.14. We stress that we obtain these results with a limited time span of data.

[13] There are alternatives to this measure, see the comments in Doppelhofer and Weeks (2009b).

Table 2.4 *Posterior jointness measures for predictor pairs*

Dynamic ordered probit				Static ordered probit			
Complements		Substitutes		Complements		Substitutes	
Pair	$J_{k,l}$	Pair	$J_{k,l}$	Pair	$J_{k,l}$	Pair	$J_{k,l}$
CC CCE	2.6	CEI IP	−3.2	CSC 3TE	3.0	10TFF OUT	−3.4
MDU OUT	1.9	CCP OUT	−2.7	10TFF CSC	2.6	CC CSC	−2.7
CS CSE	1.6	CEI PI	−1.9	CS CSE	2.6	CSC OUT	−2.5
INF PI	1.6	CCP MDU	−1.6	MDU OUT	2.3	CCP CSC	−2.4
3TE INE	1.5	WHM PI	−1.3	CUM MDU	1.9	CS CSC	−2.1
INF CSC	1.4	CEI HCM	−1.2	IP MDU	1.7	CCE 3TE	−1.9
CEI ENP	1.4	NOC IP	−1.0	CCE CUM	1.6	10TFF INE	−1.8
NOK HCM	1.4	CC CCP	−0.9	INE OUT	1.5	10TFF CCP	−1.7
WHM OUT	1.3	WHM IP	−0.9	CCP 3TE	1.4	CSC CSE	−1.6
PI MTS	1.2	CCP HCM	−0.8	IP CCP	1.3	CSC MDU	−1.6

Notes: The table shows Doppelhofer and Weeks' (2009a) jointness measure for predictor pairs in the dynamic and static ordered probit models obtained with the full-sample data. The jointness measure for predictors x_{kt} and x_{lt} is defined as $J_{k,l} = \log_{10}\left[q(k,l)q(\bar{k},\bar{l})\right] - \log_{10}\left[q(k,\bar{l})q(\bar{k},l)\right]$ in which $q(k,l) = \mathsf{Pr}\left[\gamma_k = 1, \gamma_l = 1 \,|\, \mathbf{y}\right]$ is the posterior joint inclusion probability, $q(\bar{k},\bar{l}) = \mathsf{Pr}\left[\gamma_k = 0, \gamma_l = 0 \,|\, \mathbf{y}\right]$ the posterior joint exclusion probability, and $q(\bar{k},l) = \mathsf{Pr}\left[\gamma_k = 0, \gamma_l = 1 \,|\, \mathbf{y}\right]$ and $q(k,\bar{l}) = \mathsf{Pr}\left[\gamma_k = 1, \gamma_l = 0 \,|\, \mathbf{y}\right]$ are the two posterior mutual exclusion probabilities. Jeffreys' classification (Greenberg, 2008) is used to assess the strength of the relation: $|J_{k,l}| > 2$ indicates a decisively strong relation; $1 < |J_{k,l}| < 2$ a (very) strong relation, and $|J_{k,l}| < 1$ corresponds with a (weak) substantial relation. Negative values indicate substitutes, positive values signal a complementary relation.

tary pair such that the two variables are jointly in/excluded most of the time and together they are more informative than either one in isolation. If $J_{k,l} = 0$, inclusion of variable k is *a posteriori* statistically independent of variable l's inclusion.

Table 2.4 shows the ten pairs of strongest complements and substitutes for both the dynamic (left panel) and static (right panel) ordered probit models. TCB's consumer confidence index and its expectations component are strong complements. We find a similar result for the consumer sentiment indexes of the University of Michigan (CS and CSE); measures of the current economic situation need the forward-looking character of consumers' future expectations to jointly predict FOMC decisions. Many of the complementary pairs contain a current economic activity measure and a type of forward-looking variable. For example, the output gap and the expected duration of unemployment, which both are important predictors when considered in isolation as shown in Table 2.3, are often jointly in/excluded. Growth in TCB's coincident index is a strong substitute for both industrial production and personal income. This may not be surprising, given that these variables are two of the four constituents of the CEI (together with employment and manufacturing sales). The pairs of substitutes mostly consist of current economic activity measures, for example, present consumer confidence and the output gap or hours of manufacturing and personal income.

For the static probit model we find more predictor pairs that form strong substitutes. Although in this specification the model size is generally larger, many of the different included variables contain the same information for the target rate decisions. Especially pairs of economic activity variables and different present consumer confidence measures contain similar predictive content. We observe that the SPF forecast of the three-month T-bill rate is a strong complement for current consumer sentiment ($J_{\text{CSC,3TE}} \approx 3$). The latter, in turn, forms a strong complementary pair with the spread between the ten-year T-bond and the federal funds rate; the pair has a jointness measure of 2.6. For complementary pairs we find a similar result as in the dynamic model: activity variables and forward-looking measures mutually complement each other.

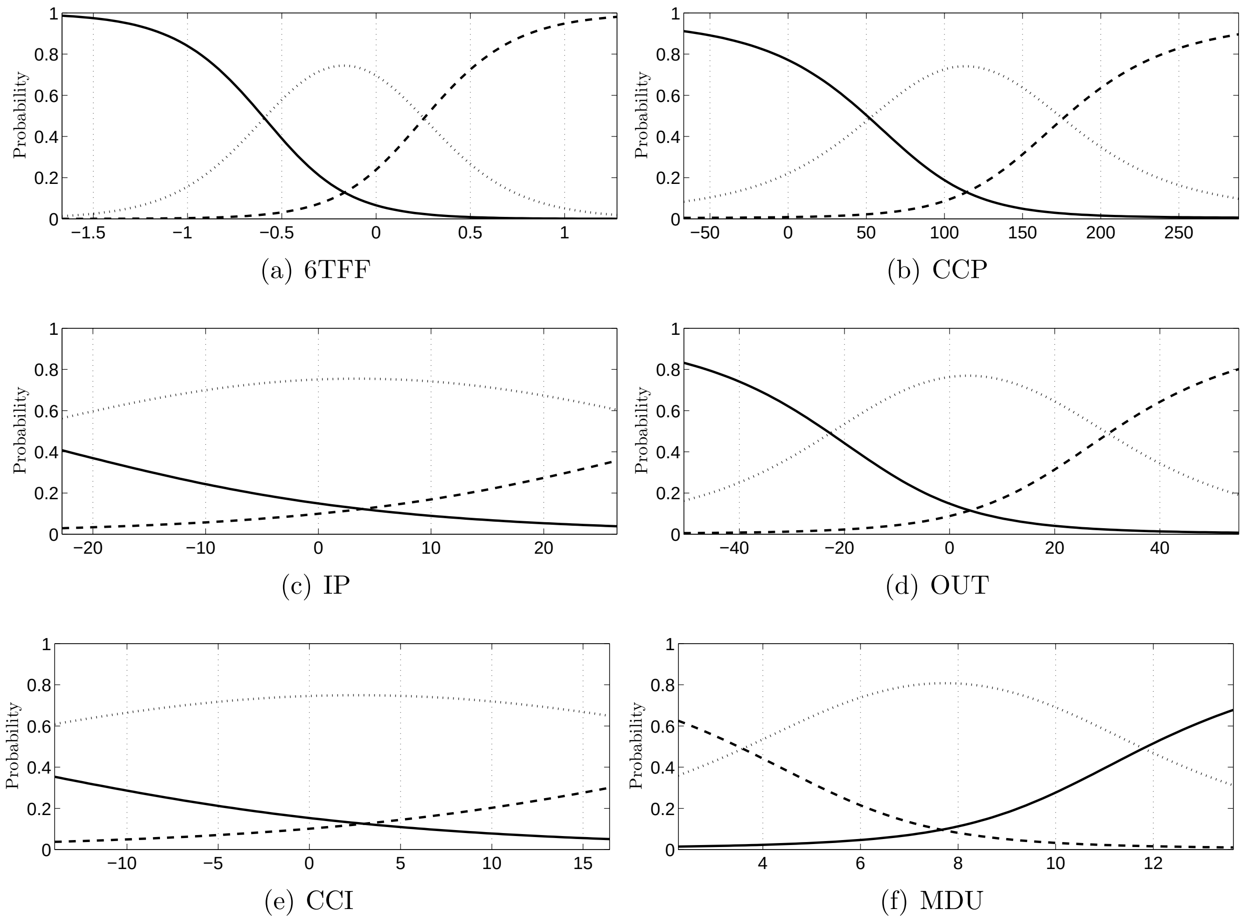
The posterior most likely combinations of covariates are obviously combinations of the posterior most likely included predictors discussed above. For the dynamic probit model, the specification that includes the interest rate spread (6TFF), the present situation consumer confidence (CCP) and TCB's coincident economic index (CEI) is the most likely with a posterior probability of 0.07. The second most likely model substitutes industrial production for CEI. Most combinations consist of a measure of current economic activity, such as industrial production, the coincident index or the output gap, complemented with survey-based information like the SPF three-month T-bill rate forecast, consumer confidence indexes or households' expectations about the duration of unemployment. The autoregressive component in the dynamic probit model accounts for a large part of the dynamics in the latent target rate r_t^*. The static ordered probit specification, though, needs more explanatory variables to describe these dynamics. We expect the main predictors which are important in the static model but less in the dynamic variant (10TFF, 3TE, CSC and IP) to pick up the strong dynamics in r_{t-1}^*. We find that these four covariates jointly explain about 50 percent of the total variation in the lagged latent target rate. For both the dynamic and the static specification we find small probabilities of individual models, indicating that model uncertainty is substantial. Consequently, averaging over models for descriptive or forecasting purposes may be preferable compared to relying upon a single specification with a particular choice of explanatory variables.

We measure the effects of the covariates on the target rate decisions by the marginal posteriors of the parameters β_k. We focus on the effect of an explanatory variable conditional on being included in the model. We therefore consider the properties of the marginal conditional densities $p(\beta_k \,|\, \mathbf{y}, \gamma_k = 1)$ as reported in Table 2.3. Obviously, we can only draw meaningful conclusions about these effects if the specific variable is incorporated frequently enough, that is, if $\mathsf{Pr}[\gamma_k = 1 \,|\, \mathbf{y}]$ is reasonably large. Table 2.3 shows the mean, standard deviation and the 5th and 95th percentiles of the posterior distributions conditional on predictor inclusion. We compute the unconditional posterior mean of β_k by using $\mathsf{E}[\beta_k \,|\, \mathbf{y}] = \mathsf{E}[\beta_k \,|\, \mathbf{y}, \gamma_k = 1] \cdot \mathsf{Pr}[\gamma_k = 1 \,|\, \mathbf{y}]$, because $\mathsf{E}[\beta_k \,|\, \mathbf{y}, \gamma_k = 0] = 0$.

Economic activity measures all have a positive effect, that is, larger values of these variables imply a higher likelihood of a target rate increase. This corresponds well with the idea that the FOMC tries to temper economic activity during expansionary periods by setting a higher target rate, in order to prevent inflation from becoming too high. If households expect that it will become more difficult to find a job in the coming months, this points at a slowdown in economic growth. The likelihood that the Fed will intervene and stimulate the economy by setting a lower target rate will increase, which explains why $\mathsf{Pr}[\beta_{\text{MDU}} < 0 \,|\, \mathbf{y}, \gamma_{\text{MDU}} = 1] \approx 1$. A similar explanation holds for the effect of households' debt outstanding (HCM).

The negative effect of the spread between the ten-year T-bond rate and the effective federal funds rate in the static probit model may seem strange at first. However, we can interpret this variable as a proxy for long-run inflation expectations, as pointed out in, *e.g.*, Estrella and Mishkin (1997). As such, it may not be a variable the FOMC reacts upon directly, but it rather measures market expectations of what the FOMC will decide at

Figure 2.2 *Effect of predictor variables, dynamic model*

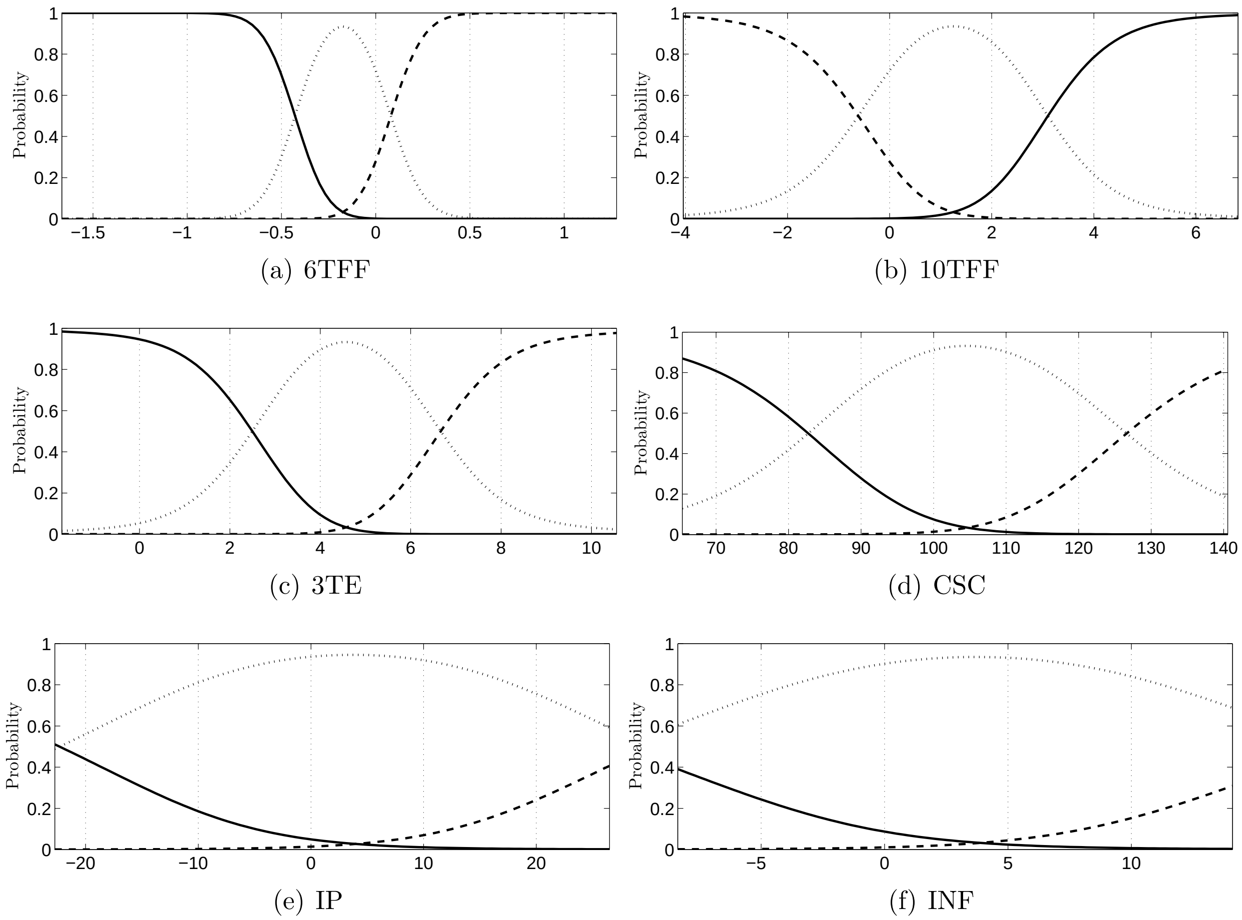


Notes: The graphs show the functional relation of the unconditional probabilities for the different types of FOMC decisions with the predictor variable. The solid, dotted and dashed lines represent the probabilities of a decision to decrease, leave unchanged, and increase the target rate, respectively. The probabilities are obtained while keeping constant all other predictor variables at their sample means. The probabilities can be interpreted as long term forecasts, since temporal dependence is integrated out such that $r_t^* \,|\, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\beta}'\mathbf{x}_t, (1-\varphi^2)^{-1})$. The displayed effects are conditional upon including the variable in the model ($\gamma_k = 1$); further parameter uncertainty is integrated out.

upcoming meetings and what the long-run impact of its decisions will be. If people expect inflation to be high for a relatively long period of time as the result of a relatively low target rate, long-term interest rates should increase to compensate. The ten-year T-bond rate thus serves more as a predictor of FOMC outcomes than an indicator to which the FOMC reacts. This reasoning is contrary to the explanation of the effect of a short-term spread like 6TFF, which represents short-term market expectations about inflation/economic activity to which the FOMC does react. Estrella and Mishkin (1997) also find different predictive natures of spreads with different maturities with regard to forecasting inflation and output growth. Mishkin (1990a, b) discusses similar issues as well.

To get a better understanding of the effects of the explanatory variables, we display the three category probabilities as functions of a predictor value in Figures 2.2 and 2.3 for the dynamic and static probit models, respectively. We show effects of the six posterior most relevant predictors. The steeper the graphs are, the stronger the effect of the explanatory

Figure 2.3 *Effect of predictor variables, static model*

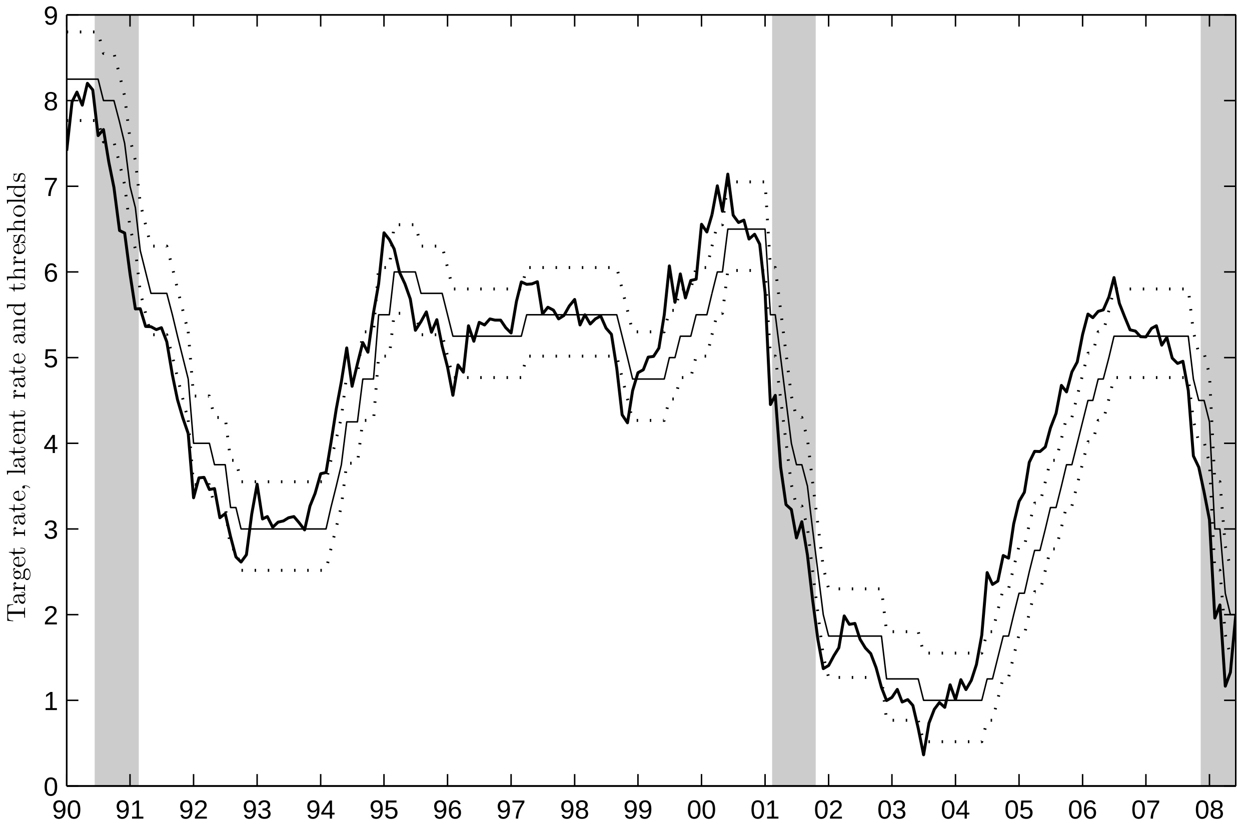


Notes: In the static model we use $r_t^* \,|\, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\beta}'\mathbf{x}_t, 1)$. For further notes, see Figure 2.2.

variable. We see that 6TFF, CCP and OUT have the largest impact on the type of FOMC decision. In Figure 2.3 we observe that 6TFF has an even stronger effect in the static probit model. Due to the negative effect of MDU, the graphs for a target rate increase (dashed line) and decrease (solid line) switch roles compared to variables with a positive influence. We further note that the effects in these figures are marginal effects, *i.e.*, holding constant the other variables at their sample means. However, we have previously seen that relevant predictors tend to come in complementary sets. Hence, the depicted marginal effects provide a lower bound for the joint impact.

A third finding from Table 2.3 is that the posterior results for the thresholds α_1 and α_2 also reveal a difference between the static and dynamic models. In the latter model, the posterior means of the thresholds are almost twice as large (in absolute value), substantially expanding the no-change region for the latent target rate. This is an implication of the difference in model specification. In the dynamic model, r_t^* will be more variable in an absolute sense because the unconditional variance of the error u_t equals $1/(1-\varphi^2) > 1$. To check whether the FOMC decides "symmetrically" with regard to upward or downward target rate adjustments, we compute $\mathsf{Pr}[\alpha_1 + \alpha_2 > 0 \,|\, \mathbf{y}]$. In case of symmetric behavior $\alpha_1 + \alpha_2$ would equally likely be positive or negative. In the dynamic probit model this probability is 0.60 and the static model has a value of even 0.85. This suggests that the

Figure 2.4 *Federal funds target rate, latent target rate and thresholds*



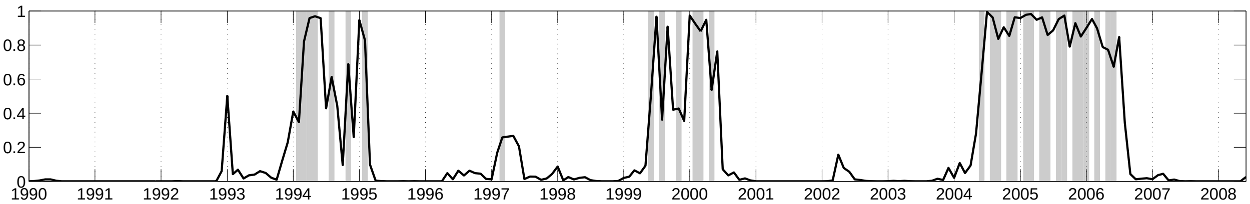
Notes: The graph shows monthly observations of the (one month lagged) federal funds target rate (black graph), the latent target rate (bold black graph) and the thresholds (dotted graphs). The shaded areas correspond with U.S. recessions, according to the NBER business cycle turning points.

FOMC is more reluctant to increase the target rate than to decrease it (as r_t^* must take more extreme values in order to trigger an increase of the target rate).[14]

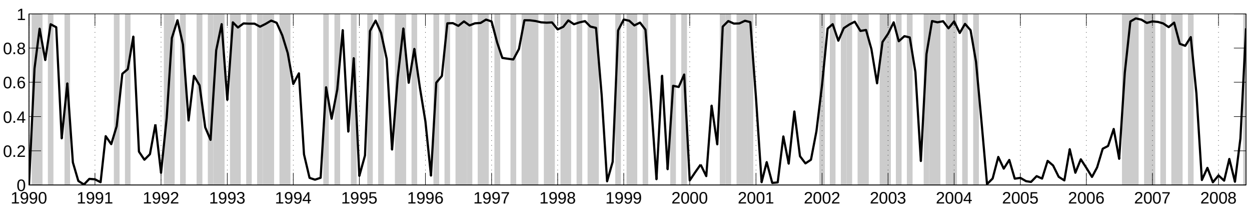
Finally, Figure 2.4 shows the posterior time-path of the latent target rate r_t^* and the thresholds that determine what decision will be taken. In order to facilitate the interpretation of this figure, we have pinned down the upper threshold at 0.50 and scaled down the other cut-off value and r_t^* correspondingly. The latent target leads the announced target, with a maximum sample correlation (0.70) for a twelve-month lead time. This point is well illustrated when we focus on the period following the millennium change. Before the target was lowered in 2001, we observe that the latent rate had already been decreasing for several months. A similar pattern appears for the upward target rate adjustments from mid-2004 onwards. This leading characteristic is in line with previous research, see Hu and Phillips (2004). The graph also shows that large changes in the target rate ($|\Delta r_t| > 0.25$) coincide with relatively large (absolute) values of r_t^*. For example, the rapid and pronounced decline in the latent rate during the second half of 2007 resulted in downward target changes of

[14]If we had a proper prior for the threshold parameters, we could compute the Savage–Dickey density ratio to formally test FOMC's symmetric decision-making restriction $\alpha_1 = -\alpha_2$.

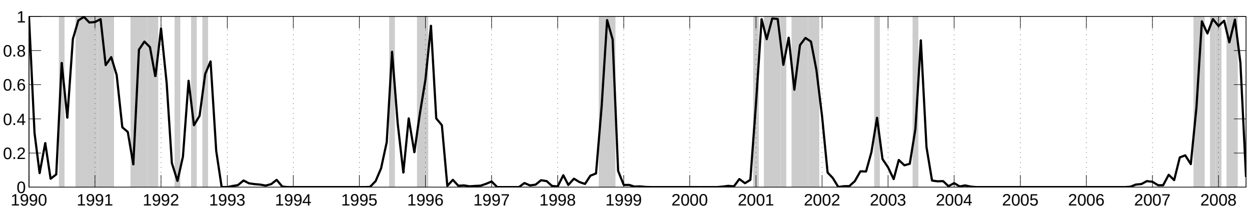
Figure 2.5 *Smoothed in-sample probabilities, dynamic model*



(a) Target rate increase



(b) No target rate change



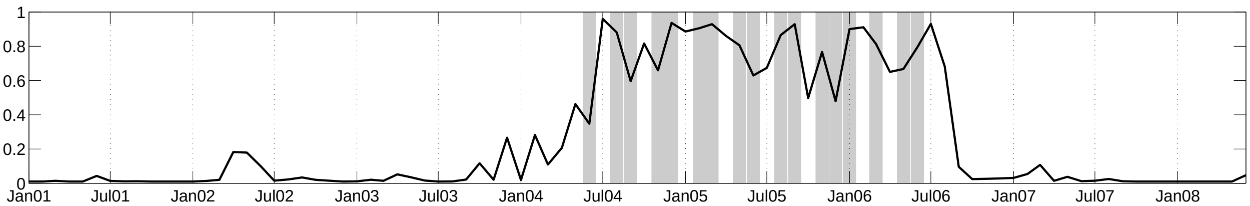
(c) Target rate decrease

0.50 and even 1.25 percentage points at single meetings. We note that this leading property of the latent target rate is quite promising for the ability of the dynamic probit model to actually predict future FOMC target rate decisions, which is the topic we turn to next.

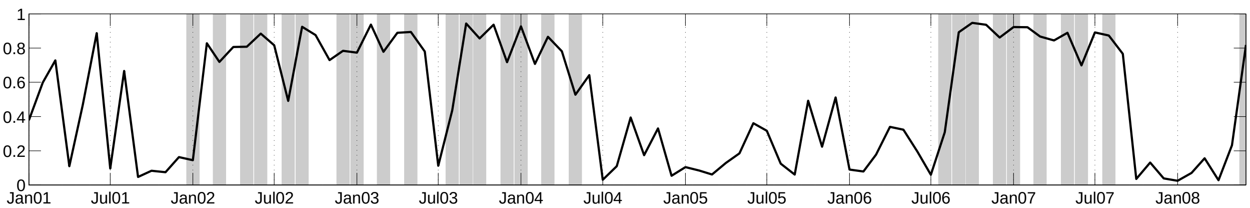
2.4.2 Forecasting

We examine the predictive ability of our dynamic ordered probit model in two ways. First, we compute the smoothed one-step-ahead probabilities of a decrease, no-change or increase of the target rate for each month in the full sample period January 1990–June 2008. Specifically, we compute the mean of $\mathsf{Pr}[y_t = j \,|\, \boldsymbol{\theta}, r^*_{t-1}] = \Phi(\alpha_j + r_{t-1} - \mu_t) - \Phi(\alpha_{j-1} + r_{t-1} - \mu_t)$, $(j = 1, 2, 3)$, over simulated values from the full-sample posterior $p(\boldsymbol{\theta}, \mathbf{r}^* \,|\, \mathbf{y})$. We use that $\mu_t = \boldsymbol{\beta}'\mathbf{x}_t + \varphi(r^*_{t-1} - \boldsymbol{\beta}'\mathbf{x}_{t-1})$ is the conditional mean of r^*_t. Figure 2.5 displays the results, with the shaded bars representing the actual FOMC decisions. The dynamic probit model fits well, achieving a hit rate of 90 percent. The hit rate is defined as the percentage of correctly predicted target rate decisions, in which the forecast is that decision that has the highest posterior probability. If we consider the static probit model, the in-sample hit rate deteriorates with 16 percentage points to 74 percent. The dynamic probit probabilities show pronounced behavior, assigning most of the weight to a single decision. The static probit model generally gives more moderate probability forecasts for target rate increases and decreases and assigns more weight to the no-change decision.

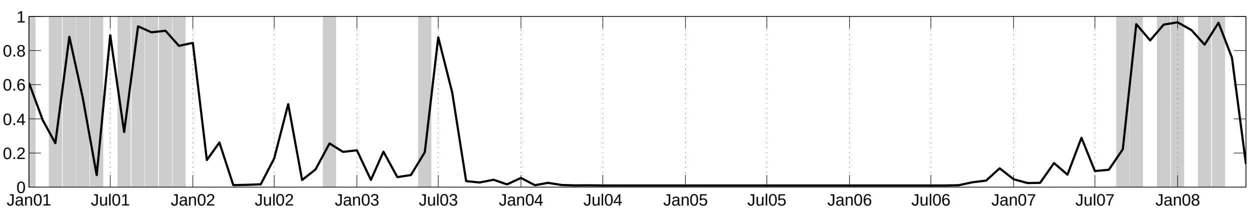
Figure 2.6 *Real-time one-month-ahead out-of-sample probabilities, dynamic model*



(a) Target rate increase



(b) No target rate change



(c) Target rate decrease

Second, we employ our real-time Bayesian forecasting procedure and obtain out-of-sample posterior predictive distributions for the period January 2001–June 2008. We depict the one-step-ahead probabilities in Figure 2.6. We first focus on the target increases in the upper panel of Figure 2.6. The dynamic probit model succeeds in predicting the cluster of target rate increases from mid-2004 till mid-2006 rather well. Already at the start of 2004 the probability of a target rate hike gradually increases. In June, when the FOMC decided to increase the target rate for the first time, it was up to 0.5. For all other target rate increases in this period the probability forecast is in the range $[0.6; 0.9]$. For the first meeting after this period of consecutive target rate increases, in August 2006, the model predicts that a rise and no change would realize with probabilities 0.7 and 0.3, respectively (see middle panel of Figure 2.6). Hence, the probability of a target rate increase was already declining considerably.

The probability forecasts for the beginning of the cluster of target rate decreases in 2001 are not very accurate, see the lower panel of Figure 2.6. This holds especially for the decrease of June 2001 with a probability forecast of only 10 percent. Possibly the posterior was not yet precise enough to generate accurate forecasts, with only 88 meetings available in our sample before 2001. In-sample these meetings are fitted quite well though, as seen in Figure 2.5. Closer inspection of the predictors reveals that some of the posterior most likely variables fail to indicate the occurrence of a recession and thus the need for lower interest rates. In

Table 2.5 *Hit rates and predictive likelihoods*

Model	Data type	In-sample H.R.	Out-of-sample H.R.	Pr.Li.
Dynamic	Real-time	89.8 (141/157)	82.3 (51/62)	−11.2
Static	Real-time	73.9 (116/157)	77.4 (48/62)	−17.4
Dynamic	Latest-available	88.5 (139/157)	80.6 (50/62)	−11.6
Static	Latest-available	77.7 (122/157)	72.6 (45/62)	−14.6
Benchmark I: Time-invariant		54.8 (86/157)	43.6 (27/62)	−30.4
Benchmark II: Pure AR(1)		81.5 (128/157)	85.5 (53/62)	−14.4

Notes: Hit rates ("H.R.") are equal to the percentage of correctly predicted target rate decisions, when the forecast of the decision is given by $\hat{y}_t = \arg\max\{\Pr[y_t = j] : j = 1, 2, 3\}$. The benchmarks are (I) an ordered probit model with time-invariant probabilities $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)'$ (effectively, $\boldsymbol{\beta} = \mathbf{0}_K$ in (2.3) and $\varphi = 0$ in (2.4); we use the uniform Dirichlet distribution as prior for $\boldsymbol{\pi}$) such that the dependent variables are temporally independent, and (II) a pure AR(1) model ($\boldsymbol{\beta} = \mathbf{0}_K$ in (2.3)) for the latent target rate r_t^*. "In-sample" refers to hit rates for probability forecasts obtained with the full-sample data covering January 1990–June 2008; "Out-of-sample" refers to hit rates for probability forecasts for January 2001–June 2008. The column with the heading "Pr.Li." contains the $\log_{10}$-valued predictive likelihoods for the out-of-sample period.

particular, throughout 2001 MDU and CCP even point towards an expansion. MDU only starts to increase in January 2002, while CCP increases until December 2001. During the following period, running from 2002 till the end of 2003, the target rate remained fairly stable with two isolated FOMC decisions to lower the target (November 2002 and June 2003). A prolonged period of consecutive decreases of the target rate started again at the end of 2007. During the six months before, the associated probability forecast already increases. Following the final decrease in May 2008 this probability quickly returns to zero. We further note that for the whole sample, in periods of an increasing target rate, the probability of a decrease is virtually zero, and vice versa.

For the static probit model we find less pronounced probability forecasts. Especially during the period of consecutive increases in 2004–06, the no-change decision is given considerable probability. For the period July 2002–December 2003 the static model considers target rate decreases more likely than the dynamic model. Also for the static probit model, we find a substantial and persistent rise in the probabilities of a target increase or decrease already a few months before a period of such target rate changes actually starts.

A researcher with a zero-one loss function is interested in hit rates (Geweke, 2005, Ch. 2). The corresponding point forecasts are the predictive modes of $y_s \,|\, \mathbf{y}^{1,s-1}$, $(s = \tau + 1, \ldots, T)$, with $\tau + 1$ as January 2001. We report out-of-sample hit rates in Table 2.5. Point forecasts of the dynamic probit model predict 82 percent of the meetings correctly. The static probit model achieves an out-of-sample hit rate of 77 percent. A formal Bayesian out-of-sample comparison of our models requires the computation of the predictive likelihood $p(\mathbf{y}^{\tau+1,T} \,|\, \mathbf{y}^{1,\tau})$ though. For the dynamic model we find a ($\log_{10}$) value of −11.2, compared to −17.4 for the static model, which shows that the dynamic variant finds decisively more support in the data with a predictive Bayes factor exceeding 2 (see the final column of Table 2.5).

The final two rows in Table 2.5 display the results for two benchmark models: a static model without any explanatory variables and a model with an AR(1) specification only. The first benchmark model is labelled "time-invariant" as its three category probabilities do not depend upon t. Clearly this model is outclassed by all other models, both in terms of in-sample fit as well as out-of-sample forecasting performance. In the model with an AR(1) specification but without explanatory variables, the persistence of the latent target rate is estimated to be even stronger than in the model with macro and financial predictors. As a result, the pure AR(1) model delivers the same *point* forecasts as the simple forecasting rule "next month's decision equals this month's." This results in an out-of-sample hit rate of 85.5 percent, outperforming both the static and dynamic probit models. In terms of predictive likelihood, however, the dynamic probit provides decisively better forecasts than this benchmark with a predicitive Bayes factor of 3.2 in its favor.

Finally, we examine the difference between using revised data and data available in real time. Visual inspection of the probability plots of the dynamic model reveals no significant differences. In Table 2.5 we see that the predictive likelihood using real-time data is slightly greater than in the case with latest-available data. The Bayes factor is around 0.4, indicating weak support for preferring real-time to revised data. We observe an improvement in the in-sample fit of the static model if we switch from real-time to revised data. We find such an improvement in terms of the out-of-sample performance too, with the predictive likelihood increasing from -17.4 to -14.6. However, at the same time the out-of-sample hit rate declines from 77.4 to 72.6 percent if we substitute real-time for latest-available data.

To give an insight into the advantages of model averaging, we also estimate the models with a fixed set of regressors. For both the dynamic and static ordered probits we choose the posterior most likely predictor combination. Unreported results show that the fixed-regressor models have better predictive likelihoods because they generally assign more probability to the most frequently sampled no-change category. The advantage of model averaging becomes most visible during the isolated target decreases in November 2002 and especially in June 2003 (Figure 2.6). These decisions are considered more likely compared to the fixed-predictor model, since model averaging incorporates an additional source of uncertainty which is expressed in a more dispersed predictive distribution. To forecast future FOMC decisions we could use this information and combine forecasts of the model with the given set of predictors with forecasts from a specification with endogenous variable selection.

2.4.3 Robustness checks

We assess the robustness of our empirical results in two respects, namely the choice of prior for the predictor inclusion probability π and the "averaging" period m used for constructing our predictor variables.

By replacing the uninformative Beta prior with a $\mathcal{B}e(3, 1)$ (prior mean for π of 0.75 but relatively large variance) or a $\mathcal{B}e(10, 10)$ (prior mass relatively tightly concentrated around $\pi = 0.5$), the posterior mean of π increases from 0.15 to 0.25 and 0.30, respectively. The type of posterior most likely variables does not change though. These larger models also do not result in better forecasting results. Our findings support the idea that we should be careful with interpreting the posterior inclusion probabilities. These can easily be inflated via the prior without enhancing the fit of the model, as the prior simply forces to include irrelevant predictors more frequently. On the other hand, preferring smaller models *a priori* by using $\mathcal{B}e(1, 10)$ and $\mathcal{B}e(1, 4)$ (with prior means for π of 0.09 and 0.20), we obtain exactly the same ten most likely included predictors, in both cases. Posterior means of π become 0.11 and 0.13, respectively. Also in terms of forecasting performance we do not find crucial differences with $\log_{10}$ predictive likelihoods of -11.1 and -11.3. Not surprisingly, since these prior beliefs are more in line with the information from the data and therefore close to the posterior results with an uninformative prior.

The choice of the averaging period m does have consequences. If we set it equal to three months, similar predictors as for $m = 1$ are selected. However, the forecasting performance deteriorates, mainly because the reversal decisions become harder to forecast. The smoothing of the predictors makes the probabilities react more sluggishly. An even more extreme choice of nine months has the result that hardly any explanatory variables provide useful information. A pure AR(1) process for r_t^* is then the posterior most likely model. As a result, the probabilities become very smooth with always considerable mass assigned to the no-change class. Obviously, since in this model no exogenous information is incorporated, changing patterns are difficult to predict and we simply extrapolate the past. However, the predictive likelihood of a pure AR(1) for the latent target rate takes a value of -14.4, which still results in an improvement over a benchmark model with time-invariant probabilities and a static ordered probit model, see Table 2.5. We attribute this to the strong dynamics in r_t^*, as outcomes of future meetings in the short run will likely be of the same type as y_t.

2.5 Conclusion

In this chapter we model the discrete changes in the federal funds target rate during the period January 1990–June 2008. We focus on the direction of change as decided by the FOMC during their meetings held approximately eight times per year, using ordered probit models. The key modeling issue that we address concerns the question which economic indicators help to predict the FOMC decisions. For this reason we use an endogenous variable selection procedure. We consider a set of 33 potential predictors, including macroeconomic and financial series and forward-looking variables obtained from the Survey of Professional Forecasters.

Our empirical results show strong evidence for persistence in the target rate decisions, beyond the persistence caused by the (strongly) autocorrelated covariates. Most predictive ability is found for, first, economic activity measures like industrial production, the output gap and the coincident index, and, second, term structure variables like interest rate spreads. In addition, SPF-based forecasts for the three-month T-bill rate and different survey-based variables that measure current and future consumer confidence have predictive content.

We also propose a Bayesian forecasting scheme on a real-time basis to construct out-of-sample probability forecasts, efficiently using all information available at the time of producing the forecasts. For this purpose, we construct a real-time data set and update the posterior parameter beliefs with importance sampling techniques each time a new observation comes in. The Bayesian approach makes sure that we can appropriately deal with parameter and model uncertainty to end up with a parameter- and model-free forecast.

Meetings in the period January 2001–June 2008 are predicted well. 82 percent of the outcomes during the out-of-sample period 2001–08 are correctly predicted. In-sample we even achieve a hit rate of 90 percent. Changes in the economy and subsequently in the outcome of FOMC meetings are quickly incorporated in the forecasts. For example, before the FOMC started increasing the target rate in mid-2004, its probability forecast was already rising for four months. The dynamic ordered probit model improves forecasting performance substantially compared to a static model. The latter achieves hit rates of 77 (in-sample) and 74 percent (out-of-sample).

2.A Technical details

In this appendix we provide details of the Bayesian analysis of our dynamic ordered probit model. We discuss the posterior simulation procedure in Sections 2.A.1–2, and the forecastig methods in Section 2.A.3.

2.A.1 Prior distribution and complete data likelihoood

With the specification of the individual prior distributions in Section 2.2.2 we obtain the joint prior density

$$p(\boldsymbol{\theta}) \propto \exp\Big\{-\frac{1}{2}(\boldsymbol{\psi}-\mathbf{a})'\mathbf{A}^{-1}(\boldsymbol{\psi}-\mathbf{a})\Big\}\exp\Big\{-\frac{1}{2B}(\varphi-b)^2\Big\}\times\mathbb{I}_{\{\varphi\in S\}} \\ \times\pi^{c_1+N(\boldsymbol{\gamma})-1}(1-\pi)^{c_2+K-N(\boldsymbol{\gamma})-1}\times\mathbb{I}_{\{\alpha_1<\alpha_2\}}, \tag{2.A.1}$$

in which K is the total number of potential predictors, and $N(\boldsymbol{\gamma})=\sum_{k=1}^{K}\gamma_k$ is the size of the model, conditional on the variable inclusion indicators $\boldsymbol{\gamma}$.

To derive the complete data likelihood of our model (2.1)–(2.4) we collect the latent target rates in $\mathbf{r}^*=(r_1^*,\ldots,r_T^*)'$ and the target rate decisions in the vector $\mathbf{y}=(y_1,\ldots,y_T)'$. The complete data likelihood is decomposed as $p(\mathbf{y},\mathbf{r}^*\,|\,\boldsymbol{\theta})=p(\mathbf{y}\,|\,\mathbf{r}^*,\boldsymbol{\theta})p(\mathbf{r}^*\,|\,\boldsymbol{\theta})$. The first pdf on the left follows from the mapping rule in (2.2) and noting that conditional on r_t^*, y_t is degenerate. To derive the second pdf, $p(\mathbf{r}^*\,|\,\boldsymbol{\theta})$, we use the first-order Markov property of $\{r_t^*\}$ by combining (2.3) and (2.4). The first latent variable in the sample, r_1^*, is modeled by its marginal distribution (conditional on its predictors) $r_1^*\,|\,\boldsymbol{\theta}\sim\mathcal{N}(\boldsymbol{\beta}'\mathbf{x}_1,(1-\varphi^2)^{-1})$. If we combine both densities we obtain the complete data likelihood

$$p(\mathbf{y},\mathbf{r}^*\,|\,\boldsymbol{\theta})\propto(1-\varphi^2)^{\frac{1}{2}}\exp\Big\{-\frac{1}{2}(1-\varphi^2)(r_1^*-\boldsymbol{\beta}'\mathbf{x}_1)^2\Big\} \tag{2.A.2}\\ \times\exp\Big\{-\frac{1}{2}\sum_{t=2}^{T}\big[r_t^*-\boldsymbol{\beta}'\mathbf{x}_t-\varphi(r_{t-1}^*-\boldsymbol{\beta}'\mathbf{x}_{t-1})\big]^2\Big\} \\ \times\prod_{t=1}^{T}\Big(\sum_{j=1}^{3}\mathbb{I}_{\{y_t=j\}}\mathbb{I}_{\{r_t^*-r_{t-1}\in(\alpha_{j-1},\alpha_j]\}}\Big).$$

2.A.2 Posterior simulation steps

We set up an MCMC sampler to simulate from the augmented posterior $p(\boldsymbol{\theta},\mathbf{r}^*\,|\,\mathbf{y})$ which is proportional to the product of (2.A.1) and (2.A.2). However, we first reparameterize the model to improve the mixing of the Markov chain (see, *e.g.*, Cowles, 1996, Liu and Sabatti, 2000, for difficulties with MCMC in ordered probit models). We define the transformations $\beta_0^*=-\alpha_1$, $\alpha_2^*=\alpha_2-\alpha_1$ and $w_t^*=r_t^*-\alpha_1$, $(t=1,\ldots,T)$, such that we eliminate the lower threshold parameter. We apply an MCMC sampler to generate a sample from $p(\boldsymbol{\theta}^*,\mathbf{w}^*\,|\,\mathbf{y})$ in which $\mathbf{w}^*=(w_1^*,\ldots,w_T^*)'$ and $\boldsymbol{\theta}^*=\{\beta_0^*,\alpha_2^*,\pi,\varphi,\boldsymbol{\psi},\boldsymbol{\gamma}\}$. To keep this reparameterization consistent with the original model, the implied prior on the transformed parameters consists of $p(\alpha_2^*)\propto\mathbb{I}_{\{\alpha_2^*>0\}}$ and $p(\beta_0^*)\propto 1$. We note that if we have a sample from $p(\boldsymbol{\theta}^*,\mathbf{w}^*\,|\,\mathbf{y})$, we also have a sample from the required $p(\boldsymbol{\theta},\mathbf{r}^*\,|\,\mathbf{y})$. The details of the six steps of the simulation scheme in Section 2.2.3 are as follows.

Step 1: Threshold parameter

In the reparameterized model we have to sample only one threshold parameter. We sample α_2^* from its full conditional posterior which is a uniform distribution on the interval $[\bar{a}_l, \bar{a}_u]$. That is, we simulate

$$\alpha_2^* \,|\, \{\mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\alpha_2^*}\} \sim \mathcal{U}([\bar{a}_l, \bar{a}_u]),$$
$$\bar{a}_l = \max\left\{\max_{t:y_t=2}\{w_t^* - r_{t-1}\}, 0\right\}, \quad \bar{a}_u = \min_{t:y_t=3}\{w_t^* - r_{t-1}\}.$$

Step 2: Autoregressive parameter

Due to the initialization of the latent target rate, we implement an independence Metropolis step to simulate φ. Conditional on $\mathbf{w}^*$ and $\boldsymbol{\theta}^*_{-\varphi}$ we know the original latent targets $\mathbf{r}^*$ and we compute $u_t = r_t^* - \boldsymbol{\beta}'\mathbf{x}_t$, $(t = 1, \ldots, T)$. Hence the full conditional posterior reads

$$p(\varphi \,|\, \mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\varphi}) \propto \exp\left\{-(\mathbf{u}^{2,T} - \varphi\mathbf{u}^{1,T-1})'(\mathbf{u}^{2,T} - \varphi\mathbf{u}^{1,T-1})/2\right\}$$
$$\times (1-\varphi^2)^{\frac{1}{2}} \exp\left\{-(1-\varphi^2)u_1^2/2\right\} \exp\left\{-(\varphi - b)^2/(2B)\right\} \mathbb{I}_{\{\varphi \in (-1,1)\}},$$

which is the exact likelihood specification of a Gaussian AR(1) model. Following Chib and Greenberg (1994), we simulate a truncated normal candidate value $\varphi^{(m)} \,|\, \{\mathbf{w}^*, \boldsymbol{\theta}^*_{-\varphi}\} \sim \mathcal{N}(\bar{b}, \bar{B}) \times \mathbb{I}_{\{\varphi^{(m)} \in (-1,1)\}}$, in which the two parameters are

$$\bar{b} = \bar{B}(\mathbf{u}^{2,T\prime}\mathbf{u}^{1,T-1} + b/B), \quad \bar{B} = 1/(\mathbf{u}^{1,T-1\prime}\mathbf{u}^{1,T-1} + B^{-1}).$$

We accept the candidate value $\varphi^{(m)}$ as the next draw with probability

$$\eta = \min\left\{1, \left(\frac{1-\varphi^{(m)2}}{1-\varphi^{(m-1)2}}\right)^{\frac{1}{2}} \exp\left\{-u_1^2\left[(1-\varphi^{(m)2}) + (1-\varphi^{(m-1)2})\right]/2\right\}\right\},$$

and with probability $1-\eta$ we keep the current draw and set $\varphi^{(m)} = \varphi^{(m-1)}$.

Step 3: Variable inclusion probability

We sample the "overall" probability of predictor inclusion from its full conditional posterior, which is the Beta distribution

$$\pi \,|\, \{\mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\pi}\} \sim \mathcal{B}e(c_1 + N(\boldsymbol{\gamma}), c_2 + K - N(\boldsymbol{\gamma})).$$

$N(\boldsymbol{\gamma})$ indicates the number of included predictors in the current iteration of the sampler.

Step 4: Latent data

With the intercept stacked with the regression parameters in $\boldsymbol{\beta}^* = (\beta_0^*, \boldsymbol{\beta}')'$, and augmenting the vector with predictors accordingly to $\mathbf{x}_t^* = (1, \mathbf{x}_t')'$, the transformed latent target behaves like $w_t^* - \boldsymbol{\beta}^{*\prime}\mathbf{x}_t^* = \varphi(w_{t-1}^* - \boldsymbol{\beta}^{*\prime}\mathbf{x}_{t-1}^*) + \varepsilon_t$ with $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. The temporal dependence in combination with the truncation (caused by conditioning on the observed decisions) makes the simulation of all latent variables in one block impossible. Instead, we sample them consecutively from their full conditional posteriors, see Albert and Chib (1993).

Since $\{w_t^*\}$ is first-order Markov, we use the two equations $w_1^* = \boldsymbol{\beta}^{*\prime}\mathbf{x}_1^* + (1-\varphi^2)^{-\frac{1}{2}}\varepsilon_1$ and $w_2^* = \boldsymbol{\beta}^{*\prime}\mathbf{x}_2^* + \varphi(w_1^* - \boldsymbol{\beta}^{*\prime}\mathbf{x}_1^*) + \varepsilon_2$ to derive the full conditional posterior of the initial w_1^*. We write them as a two-observations normal linear regression model with unknown parameter w_1^*. In this auxiliary regression we have dependent and independent variable vectors

$$\mathbf{y}_1^* = \begin{bmatrix} (1-\varphi^2)^{\frac{1}{2}}\boldsymbol{\beta}^{*\prime}\mathbf{x}_1^*, \\ w_2^* + \boldsymbol{\beta}^{*\prime}(\varphi\mathbf{x}_1^* - \mathbf{x}_2^*) \end{bmatrix}, \quad \mathbf{X}_1^* = \begin{bmatrix} (1-\varphi^2)^{\frac{1}{2}} \\ \varphi \end{bmatrix},$$

respectively. Therefore, we sample from the truncated normal distribution

$$w_1^* \,|\, \{\mathbf{y}, \mathbf{w}_{-1}^*, \boldsymbol{\theta}^*\} \sim \mathcal{N}(\mathbf{X}_1^{*\prime}\mathbf{y}_1^*, 1) \times \Big(\mathbb{I}_{\{y_1=1\}}\mathbb{I}_{\{w_1 - r_0 \le 0\}} + \mathbb{I}_{\{y_1=2\}}\mathbb{I}_{\{0 < w_1 - r_0 \le \alpha_2^*\}} + \mathbb{I}_{\{y_1=3\}}\mathbb{I}_{\{w_1 - r_0 > \alpha_2^*\}}\Big).$$

For w_t^*, $(t = 2, \ldots, T-1)$, we derive the full conditional posteriors in a similar fashion. We use the two equations $w_s^* = \boldsymbol{\beta}^{*\prime}\mathbf{x}_s^* + \varphi(w_{s-1}^* - \boldsymbol{\beta}^{*\prime}\mathbf{x}_{s-1}^*) + \varepsilon_s$, $(s = t, t+1)$, to obtain the auxiliary regression with unknown parameter w_t, and variables

$$\mathbf{y}_t^* = \begin{bmatrix} \varphi w_{t-1}^* + \boldsymbol{\beta}^{*\prime}(\mathbf{x}_t^* - \varphi\mathbf{x}_{t-1}^*) \\ w_{t+1}^* + \boldsymbol{\beta}^{*\prime}(\varphi\mathbf{x}_t^* - \mathbf{x}_{t+1}^*) \end{bmatrix}, \quad \mathbf{X}_t^* = \begin{bmatrix} 1 \\ \varphi \end{bmatrix}.$$

Then, we simulate from the truncated normal distribution

$$w_t^* \,|\, \{\mathbf{y}, \mathbf{w}_{-t}^*, \boldsymbol{\theta}^*\} \sim \mathcal{N}((1+\varphi^2)^{-1}\mathbf{X}_t^{*\prime}\mathbf{y}_t^*, (1+\varphi^2)^{-1}) \times \Big(\mathbb{I}_{\{y_t=1\}}\mathbb{I}_{\{w_t - r_{t-1} \le 0\}} + \mathbb{I}_{\{y_t=2\}}\mathbb{I}_{\{0 < w_t - r_{t-1} \le \alpha_2^*\}} + \mathbb{I}_{\{y_t=3\}}\mathbb{I}_{\{w_t - r_{t-1} > \alpha_2^*\}}\Big).$$

The final latent variable in the sample is simulated from the truncated normal distribution

$$w_T^* \,|\, \{\mathbf{y}, \mathbf{w}_{-T}^*, \boldsymbol{\theta}^*\} \sim \mathcal{N}(\boldsymbol{\beta}^{*\prime}\mathbf{x}_T^* + \varphi(w_{T-1}^* - \boldsymbol{\beta}^{*\prime}\mathbf{x}_{T-1}^*), 1) \times \Big(\mathbb{I}_{\{y_T=1\}}\mathbb{I}_{\{w_T - r_{T-1} \le 0\}} + \mathbb{I}_{\{y_T=2\}}\mathbb{I}_{\{0 < w_T - r_{T-1} \le \alpha_2^*\}} + \mathbb{I}_{\{y_T=3\}}\mathbb{I}_{\{w_T - r_{T-1} > \alpha_2^*\}}\Big).$$

In case of no FOMC decision in month t, that is, y_t is missing, there is no truncation and we sample from the derived normal distribution without truncation restrictions.

Steps 5: Variable inclusion indicators

Instead of deciding upon a variable's inclusion conditional on its regression effect, as Kuo and Mallick (1998) do, we integrate out the regression effects analytically in order to substantially decrease dependence in the Markov chain. We first rewrite the model to correct for the autocorrelation in w_t^*. We construct the auxiliary dependent z_t and independent $\mathbf{v}_t$ variables by collecting terms involving $\boldsymbol{\beta}^*$:

$$z_t \equiv w_t^* - \varphi w_{t-1}^* = \boldsymbol{\beta}^{*\prime}(\mathbf{x}_t^* - \varphi\mathbf{x}_{t-1}^*) + \varepsilon_t = \boldsymbol{\beta}^{*\prime}\mathbf{v}_t + \varepsilon_t, \qquad (t = 2, \ldots, T),$$
$$z_1 \equiv (1-\varphi^2)^{\frac{1}{2}} w_1^* = \boldsymbol{\beta}^{*\prime}\mathbf{x}_1^*(1-\varphi^2)^{\frac{1}{2}} + \varepsilon_1 = \boldsymbol{\beta}^{*\prime}\mathbf{v}_1 + \varepsilon_1.$$

Since $\boldsymbol{\beta}^{*\prime} = (\beta_0^*, \boldsymbol{\psi}') \odot (1, \boldsymbol{\gamma}') \equiv \boldsymbol{\psi}^{*\prime} \odot (1, \boldsymbol{\gamma}')$, we obtain the linear regression setting

$$\mathbf{z} = \mathbf{V}(\boldsymbol{\gamma})\boldsymbol{\psi}^* + \boldsymbol{\varepsilon}, \quad \mathbf{V}(\boldsymbol{\gamma}) = (\mathbf{v}_1, \ldots, \mathbf{v}_T)' \odot \Big(\boldsymbol{\iota}_T \otimes (1, \boldsymbol{\gamma}')\Big),$$

with error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_T)' \sim \mathcal{N}(\mathbf{0}_T, \mathbf{I}_T)$ and a $\boldsymbol{\gamma}$-dependent design matrix.

As we have a conditionally conjugate prior for $\boldsymbol{\psi}^*$ we can marginalize the conditional joint posterior of the inclusion indicators and regression effects. We define $g(\boldsymbol{\gamma}) = \pi^{N(\boldsymbol{\gamma})}(1-\pi)^{K-N(\boldsymbol{\gamma})}$ and obtain

$$\begin{aligned} p(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\{\boldsymbol{\psi}^*,\boldsymbol{\gamma}\}}) &\propto g(\boldsymbol{\gamma}) \int \exp\Big\{ - \Big[(\mathbf{z} - \mathbf{V}(\boldsymbol{\gamma})\boldsymbol{\psi}^*)'(\mathbf{z} - \mathbf{V}(\boldsymbol{\gamma})\boldsymbol{\psi}^*) \\ &\qquad + (\boldsymbol{\psi} - \mathbf{a})'\mathbf{A}^{-1}(\boldsymbol{\psi} - \mathbf{a})\Big]/2\Big\}\, \mathsf{d}\boldsymbol{\psi}^* \\ &\propto g(\boldsymbol{\gamma})|\overline{\mathbf{A}}|^{\frac{1}{2}} \exp\Big\{-[(\mathbf{z} - \mathbf{V}(\boldsymbol{\gamma})\overline{\mathbf{a}})'(\mathbf{z} - \mathbf{V}(\boldsymbol{\gamma})\overline{\mathbf{a}}) + (\overline{\mathbf{a}} - \mathbf{a}^*)'\mathbf{A}^{*-1}(\overline{\mathbf{a}} - \mathbf{a}^*)]/2\Big\}, \end{aligned} \tag{2.A.3}$$

in which the prior parameters are $\mathbf{a}^* = (0, \mathbf{a}')'$ and $\mathbf{A}^{*-1}$ is defined as $\mathbf{A}^{-1}$ expanded with a left column and a top row consisting of zeros. The two posterior parameters are

$$\overline{\mathbf{a}} = \overline{\mathbf{A}}(\mathbf{V}(\boldsymbol{\gamma})'\mathbf{z} + \mathbf{A}^{*-1}\mathbf{a}^*), \quad \overline{\mathbf{A}} = (\mathbf{V}(\boldsymbol{\gamma})'\mathbf{V}(\boldsymbol{\gamma}) + \mathbf{A}^{*-1})^{-1}. \tag{2.A.4}$$

We simulate each inclusion indicator γ_k, $(k = 1, \ldots, K)$, from the distribution with probability mass function $p(\gamma_k \,|\, \mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\{\boldsymbol{\psi}^*,\gamma_k\}})$, which is proportional to (2.A.3). That is, we compute

$$q_k \propto p(\gamma_k = 1, \boldsymbol{\gamma}_{-k} \,|\, \mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\{\boldsymbol{\psi}^*,\boldsymbol{\gamma}\}}), \quad 1 - q_k \propto p(\gamma_k = 0, \boldsymbol{\gamma}_{-k} \,|\, \mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\{\boldsymbol{\psi}^*,\boldsymbol{\gamma}\}}),$$

and sample from the Bernoulli distribution $\gamma_k \,|\, \{\mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\{\boldsymbol{\psi}^*,\gamma_k\}}\} \sim \mathcal{B}er(q_k)$ in which we obtain the probability of success q_k after normalization.

Step 6: Regression parameters

If we have an updated $\boldsymbol{\gamma}$, we simulate the regression parameters in $\boldsymbol{\psi}^*$ (which also contains the intercept/transformed lower threshold) in one block from their joint full conditional posterior $\boldsymbol{\psi}^* \,|\, \{\mathbf{y}, \mathbf{w}^*, \boldsymbol{\theta}^*_{-\boldsymbol{\psi}^*}\} \sim \mathcal{N}(\overline{\mathbf{a}}, \overline{\mathbf{A}})$, with the Gaussian mean and variance given in (2.A.4).

2.A.3 Bayesian forecasting

During the real-time Bayesian forecasting exercise we need to generate the series of one-step-ahead predictive densities $p(y_s \,|\, \mathbf{y}^{1,s-1})$, $(s = \tau+1, \ldots, T)$, which are given by equation (2.5). As discussed in Section 2.2.3, with a sample from $p(\mathbf{r}^{*;1,s-1}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1})$ we are able to evaluate the required integrals. In order to avoid calculating posterior results for every month s, we opt for sequential importance sampling techniques (see, for example, Robert and Casella, 2004, for a discussion).

We start with the following decomposition of the distribution we have to sample from,

$$p(\mathbf{r}^{*;1,s-1}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1}) = p(r^*_{s-1} \,|\, \mathbf{y}^{1,s-1}, \mathbf{r}^{*;1,s-2}, \boldsymbol{\theta}) p(\mathbf{r}^{*;1,s-2}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1}). \tag{2.A.5}$$

Given a sampled value $\{\mathbf{r}^{*,1,s-2}, \boldsymbol{\theta}\}$ from the distribution with density to the right of (2.A.5), we simulate r^*_{s-1} from $p(r^*_{s-1} \,|\, \mathbf{y}^{1,s-1}, \mathbf{r}^{*;1,s-2}, \boldsymbol{\theta})$. The latter is the pdf of a normal distribution with mean $\mu_{s-1} = \boldsymbol{\beta}'\mathbf{x}_{s-1} + \varphi(r^*_{s-2} - \boldsymbol{\beta}'\mathbf{x}_{s-2})$, unit variance and, due to the given FOMC decision y_{s-1}, truncated to the interval $(r_{s-2} + \alpha_{y_{s-1}-1},\, r_{s-2} + \alpha_{y_{s-1}}]$.

To sample from $p(\mathbf{r}^{*;1,s-2}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1})$ we use importance sampling. As importance function we apply the posterior density using data up to and including period $s-2$, that is, the pdf $p(\mathbf{r}^{*;1,s-2}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-2})$. We write the target density as

$$p(\mathbf{r}^{*;1,s-2}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-1}) = \frac{p(y_{s-1} \,|\, \mathbf{y}^{1,s-2}, \mathbf{r}^{*;1,s-2}, \boldsymbol{\theta})}{p(y_{s-1} \,|\, \mathbf{y}^{1,s-2})} p(\mathbf{r}^{*;1,s-2}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,s-2}), \tag{2.A.6}$$

which shows that the importance weights are given by

$$w_s(\mathbf{r}^{*;1,s-2}, \boldsymbol{\theta}) = p(y_{s-1} \,|\, \mathbf{y}^{1,s-2}, \mathbf{r}^{*;1,s-2}, \boldsymbol{\theta}) / p(y_{s-1} \,|\, \mathbf{y}^{1,s-2}).$$

Because of the appearance of the posterior with data $\mathbf{y}^{1,s-2}$ in (2.A.6), we apply the importance sampling decomposition recursively and start with a sample from $p(\mathbf{r}^{*;1,\tau}, \boldsymbol{\theta} \,|\, \mathbf{y}^{1,\tau})$. Hence, for $s = \tau+2, \ldots, T$, we simulate the latent variables r^*_{s-1}, and use Monte Carlo integration with importance-weight corrections to compute the predictive density of y_s as follows.

Step 1. Simulate the latent target rate from the truncated normal distribution

$$r^*_{s-1} \,|\, \{\mathbf{y}^{1,s-1}, \mathbf{r}^{*;1,s-2}, \boldsymbol{\theta}\} \sim \mathcal{N}(\mu_{s-1}, 1) \times \mathbb{I}_{\{(r^*_{s-1} - r_{s-2}) \in (\alpha_{y_{s-1}-1},\, \alpha_{y_{s-1}}]\}},$$

with $\mu_{s-1} = \boldsymbol{\beta}'\mathbf{x}_{s-1} + \varphi(r^*_{s-2} - \boldsymbol{\beta}'\mathbf{x}_{s-2})$;

Step 2. Update the importance weight $w_s(\mathbf{r}^{*;1,s-2}, \boldsymbol{\theta})$ using the initialization $w_{\tau+1} = 1$ and the recursive relation

$$w_s(\mathbf{r}^{*;1,s-2}, \boldsymbol{\theta}) \propto p(y_{s-1} \,|\, \mathbf{y}^{1,s-2}, \mathbf{r}^{*;1,s-2}, \boldsymbol{\theta}) w_{s-1}(\mathbf{r}^{*;1,s-3}, \boldsymbol{\theta});$$

Step 3. Evaluate the weighed conditional probabilities

$$\mathsf{Pr}[y_s = j \,|\, r^*_{s-1}, \boldsymbol{\theta}] \cdot w_s = [\Phi(\alpha_j + r_{s-1} - \mu_s) - \Phi(\alpha_{j-1} + r_{s-1} - \mu_s)] \cdot w_s,$$

using the simulated value $\{r^*_{s-1}, \boldsymbol{\theta}\}$, for all possible categories ($j = 1, 2, 3$), with $\Phi(\cdot)$ the standard normal cumulative distribution function;

Step 4. Repeat Steps 1–3 M times (= number of simulation runs), and compute $p(y_s \,|\, \mathbf{y}^{1,s-1})$ by using the simulation-sample average of the values generated in Step 3.

2.B Convergence checks

In this appendix we provide convergence diagnostics for our MCMC sampler. To obtain posterior results, we use a long single run of the chain. After a burn-in period, during which the effect of the initialization of the Markov chain dies out, continuation of the chain generates a dependent sample $\{\mathbf{r}^{*;(m)}, \boldsymbol{\theta}^{(m)}\}_{m=1}^{M}$ from $p(\mathbf{r}^*, \boldsymbol{\theta} \,|\, \mathbf{y})$. Hence, we need to determine this burn-in period and the number of runs M that serve as input to the computation of posterior characteristics of the model parameters (see Geyer, 1992, Brooks and Roberts, 1998).

Table 2.B *MCMC convergence diagnostics*

	β_{6TFF}	β_{CCP}	β_{IP}	β_{OUT}	β_{CEI}	β_{MDU}	β_{3TE}	β_{HCM}
Geweke test 1, p-value	0.14	0.12	0.89	0.51	0.69	0.63	0.85	0.09
Geweke test 2, p-value	0.49	0.18	0.78	0.05	0.68	0.16	0.59	0.58
Inefficiency factor	0.01	0.04	0.01	0.01	0.01	0.02	0.03	0.02
N.S.E.	0.03	0.02	0.01	0.03	0.01	0.02	0.04	0.02
	π	α_1	α_2	φ	ω_1	ω_2	ω_3	ω_4
Geweke test 1, p-value	0.44	0.27	0.20	0.65	0.40	0.71	0.10	0.36
Geweke test 2, p-value	0.26	0.79	0.77	0.95	0.27	0.24	0.94	0.67
Inefficiency factor	0.02	0.01	0.01	0.02	0.02	0.01	0.03	0.35
N.S.E.	0.00	0.03	0.05	0.00	0.00	0.09	0.07	0.00

Notes: Functions to examine convergence of the joint distribution: $\omega_1 = \pi\varphi$, $\omega_2 = \beta_{6TFF}\beta_{MDU}$, $\omega_3 = r^*_{Jan90} r^*_{Jan91}$ and $\omega_4 = \log|\alpha_2| - \log|\alpha_1|$. N.S.E. and inefficiency factors for the regression effects β_k are conditional on inclusion of the corresponding predictor, such that we can directly use them in association with Table 2.3.

We start with running the MCMC sampler for a period of 50,000 runs. The first 10,000 are considered as burn-in of the chain and the next 40,000 are used to test whether convergence of the sampler has been achieved. Geweke (1992, 2005) provides a method to check convergence of sample means, which is a necessary condition for convergence in distribution of the chain. We use the first and last 15,000 runs of the 40,000-period to construct two (approximately) mutually independent samples. For a given function h we compute the two sample means $\overline{h}_i = M_P^{-1} \sum h(\mathbf{r}_i^{*;(m)}, \boldsymbol{\theta}_i^{(m)})$ and the consistently estimated autocovariance times $\hat{\tau}_i^2$, $(i = 1, 2)$, which take into account the correlation within the subsamples. Then, under the null of equal means in the two subpopulations, $M_P(\overline{h}_1 - \overline{h}_2)^2/(\hat{\tau}_1^2 + \hat{\tau}_2^2) \xrightarrow{d} \chi^2(1)$, for M_P large. In the first row of the upper panel in Table 2.B we report p-values for this test with h the identity function, for the regression parameters of the eight most relevant predictors. The first four entries in the lower panel display the outcomes of the test for the thresholds, the autoregressive parameter and the "overall" inclusion probability. We find no indication of non-convergence based on these marginal properties. In addition we consider four functions that measure jointness between parameters and therefore could indicate non-convergence in the joint distribution. These functions are $\omega_1 = \pi\varphi$, $\omega_2 = \beta_{6TFF}\beta_{MDU}$, $\omega_3 = r^*_{Jan90} r^*_{Jan91}$ and $\omega_4 = \log|\alpha_2| - \log|\alpha_1|$. The final part of the first row in the lower panel shows that we cannot reject convergence in terms of these jointness measures either.

Next, we run the chain for another $M = 100,000$ periods on which we base our empirical results. As a double-check we implement the Geweke test again for this period. We use the first and last 25,000 draws. The second rows of both panels in Table 2.B depict the p-values. The parameter associated with the output gap attains the smallest value of 0.05. Based on these results we can safely use the draws for inference.

To examine the degree of Markov chain-induced dependence in the simulation sample we compute inefficiency factors (Robert and Casella, 2004, Ch. 12). They are the inverses of the autocorrelation times of $\{h(\mathbf{r}^{*;(m)}, \boldsymbol{\theta}^{(m)})\}$ and therefore enable comparison to simulation estimates based on an uncorrelated sample. The third rows of the two panels in Table 2.B display the inefficiency factors for the same functions as we used to test convergence. With respect to the marginal measures, all factors are of the order 10^{-2}, which is common to this type of models, see, *e.g.*, Cowles (1996). It means that, for example, the reported posterior mean of φ in Table 2.3 has the same precision as its simulation estimate based on an uncorrelated sample of $2,000$ draws. We report this precision of the estimated means, as measured by the numerical standard error (N.S.E.) and estimated by $\widehat{\tau}/\sqrt{M}$, in the fourth rows of Table 2.B. For the fourth jointness measure we observe that this transformation eliminates the correlation substantially with an efficiency factor close to one third, and hence small N.S.E. for the estimated $\mathsf{E}\left[\omega_4 \,|\, \mathbf{y}\right] \approx 0.14$. We note that this latter result is in agreement with the estimated relative reluctance to increase the target rate, as we find in Section 2.4.

CHAPTER 3

Leading Indicators and Their Unknown Lead Times

3.1 Introduction

Economic activity displays pronounced cyclical movements, alternating between peaks and troughs. For business and policy makers alike it is highly relevant to forecast these fluctuations, as this may enable them to take precautionary measures to ease their possible negative consequences. Having available a set of economic indicators leading the state of the economy for a particular horizon is therefore of much use.

In the business cycle forecasting literature the use of leading indicators takes an interesting intermediate position. While it does specify interrelatedness of economic variables, it generally does not attempt to do so in a theory-implied structural way though, but rather in terms of exploiting reduced-form correlations, see Auerbach (1982) and Diebold (1998). In that spirit the leading indicators have survived the various debates and orientations in macroeconomic forecasting. The (monthly) releases of leading indicators and composite leading indexes like The Conference Board's (TCB) are still anxiously awaited by investors and policy makers.

A substantial part of the literature on leading indicators is about the evaluation of a composite leading index' ability to forecast turning points, see Hymans *et al.* (1973), Neftçi (1982), Zarnowitz and Moore (1982), Diebold and Rudebusch (1989, 1991b) and Lahiri and Wang (1994), among others, and to forecast economic variables like output growth, see, for example, Stekler and Schepsman (1973), Diebold and Rudebusch (1991a). The relation of the business cycle with the cyclical movements in the composite leading index forms another important research topic. Filardo (1994), Hamilton and Perez-Quiros (1996), Camacho and Perez-Quiros (2002), Paap *et al.* (2009) and Çakmakli *et al.* (2011) assume a single cycle shared by coincident and leading variables, but it differs in terms of possibly regime-dependent phase shifts. Both strands of literature have in common that they perform *ex-post* analyses in the sense that they consider the composite leading economic index as given.

In this chapter we take a step back and propose a model-based approach to extract an alternative leading index which serves as a better instrument to signal business cycle regime changes, see Marcellino (2006) for a survey of alternative approaches. We consider two relaxations of the relatively simple way in which TCB constructs its composite index from the individual leading indicators, see The Conference Board (2001). First of all, TCB computes contemporaneous averages of the ten leading indicators, therefore assuming that the lead times are the same for all indicators. Clements and Galvão (2006) consider each of the leading variables in isolation and find that forecasting performance for a given horizon depends on the specific leading variable used. This suggests heterogeneity across lead times of individual indicators. We therefore relax TCB's assumption of homogenous lead times and allow the indicators to differ with respect to the number of months they lead the business cycle. Second, after standardizing the variances of the ten leading variables, TCB assigns equal weights to each of the leading indicators in the construction of the composite index. We, instead, estimate the relative importance of individual indicators and allow for potential unequal weighing of the indicators in our alternative leading index.

To determine the appropriate indicator-specific lead times and the weights to build our alternative composite index, we need a proxy of the business cycle itself. It is generally believed that coincident indicators share the same underlying cycle without phase shifts, which is assumed to correspond to the business cycle. Furthermore, the decisions of the National Bureau's of Economic Research (NBER) business cycle dating committee are widely accepted to indicate the turning points in the unobserved business cycle, see, for example, Issler and Vahid (2006). Based on these considerations we use a bivariate model in which we link the ten individual leading indicators to both TCB's composite *coincident* index and the zero-one recession indicator as released by the NBER. We set up the model in such a way that each leading indicator is allowed to have its own lead time and relative weight. By jointly considering the composite coincident index and the zero-one recession indicator, we extract a stronger leading signal compared to a univariate analysis, in which we solely consider either the coincident index or the NBER turning points.

Since we regard the lead times and the weights of the individual indicators as unknown parameters we have a discrete-continuous parameter space. By adopting a Bayesian analysis this issue is straightforwardly dealt with. Moreover, Bayes' approach provides the natural context for generating updated real-time forecasts of recession probabilities, which are useful in out-of-sample analyses.

The contributions to the literature are summarized as follows. First, we combine business-cycle information from two different sources, that is, TCB's composite coincident index and the NBER turning points. We relate this information to the ten leading macroeconomic variables and allow their lead times to be different across variables. By employing Bayesian methods we estimate the unknown lead times. Second, we examine whether some indicators are more important in leading the business cycle than others by estimating their respective weights in our new alternative leading index. We formally test both relaxations using Bayes factors. Third, we use both fully revised data and the leading variables as they were historically available (see, for example, Croushore, 2006, 2011, for forecasting implications). Although the most recent release of TCB's data is attentively watched, its value can be questioned (Diebold and Rudebusch, 1991a, b). Our real-time approach, however, allows us to recombine the predictive information from the individual indicators in such a way that we obtain a composite leading index that is actually available at the time of making a business-cyle forecast and has improved predictive performance. Fourth, our Bayesian methods enable to extract an "optimal" leading index for a given forecasting horizon, available in real time, if wanted. We average out parameter uncertainty about lead times and indicator weights according to the posterior distribution updated up to and including the month in which we want to forecast.

The main results of our approach are as follows. First, posterior results show decisive evidence in favor of heterogeneity in the lead times of the ten individual leading indicators. Specifically, two distinct groups of macroeconomic variables emerge. On the one hand, several leading indicators have a rather short lead time up to three months, including claims on unemployment benefits (one month), new orders for consumer goods (two months) and new orders for capital goods (three months). On the other hand, we also find a group of leading indicators which have much longer lead times of six months or more, including stock prices (six months), money supply (nine months) and the interest rate spread (twelve to thirteen months). Second, we find strong differences in weights of individual variables in our alternative leading index. For example, the interest rate spread is almost twice as important as most other indicators, and, on the contrary, weekly hours in manufacturing makes an almost negligible contribution. Third, Bayes factors are decisively in favor of our alternative leading index compared to a composition with equal lead times and weights (TCB's). Fourth, both in-sample and out-of-sample analyses show that our methods produce more accurate forecasts of recession probabilities. The evidence for indicator-specific weights is less convincing in the out-of-sample experiment compared to using our alternative leading index with equal weights (but with heterogenous lead times). This result is however largely due to the mediocre fit of the 2001 recession.

The remainder of this chapter is organized as follows. We discuss the bivariate model and the methods to extract and test our alternative leading index in Section 3.2. In Section 3.3 we introduce the set of macroeconomic variables and the methods of The Conference Board to construct its composite leading and coincident economic indexes. We provide both full-sample posterior results and the outcomes of the real-time out-of-sample forecasting experiment in Section 3.4. Finally, in Section 3.5 we conclude. We relegate technical details of the computational procedures to Appendix 3.A.

3.2 Methodology

In this section we put forward our novel model-based approach to extract an alternative composite leading economic index. Following the model specification described in Section 3.2.1, we discuss the Bayesian procedures used for the statistical analysis of the model in Section 3.2.2. Finally, we describe the out-of-sample forecasting experiment to check the usefulness of the proposed leading index, in terms of predicting the state of the business cycle, in Section 3.2.4.

3.2.1 Model specification

The current state of the economy is usually interpreted as the state of the cyclical component which coincident economic variables have in common. The Conference Board aggregates four of these coincident variables to obtain an impression of that state, represented by the composite coincident economic index (CEI). In addition, the NBER turning-point dates provide qualitative information about the business cycle. We jointly model both business cycle variables, which allows us to extract a more powerful leading signal for our alternative index compared to single-variable models. If we would only take the recession indicator as coincident measure, we had very limited information as it is a mere zero-one variable.

Moreover, the NBER recession-dating procedure is not mechanical, which can raise questions about why a particular month is labeled as recessionary and not as part of an expansion. By considering the additional (continuous) coincident measure, we partly tackle this issue.

We use the monthly symmetric growth rate of TCB's composite coincident index and the recession indicator released by the NBER as dependent variables that describe the state of the economy in month t. We relate these two business-cycle measures to lagged values of the $J = 10$ individual leading variables that form the input to The Conference Board's composite leading indicator. The construction of our alternative leading index is motivated from a forecasting perspective. At the end of month t we have realizations of the ten leading indicators for that month and their past values at our disposal for predicting the state of the business cycle next month.[1] On the basis of these data, by computing a weighed average of lead-time-dependent values, we obtain a forecast for the state of the economy in month $t+1$. This weighed average forms our index for one-month-ahead prediction. In general, we obtain the h-months-ahead leading index by establishing a direct link between the leading indicators up to and including month t and the coincident indicators in month $t+h$, for any $h \geq 1$. The thus computed compositions of individual leading indicators lead the state of the business cycle optimally in terms of correlation by h months.

We have a two-dimensional dependent vector with both a continuous and a binary random variable, and denote it $\mathbf{y}_t = (y_{1t}, y_{2t})'$, $(t = 1, \ldots, T)$, with y_{1t} the symmetric growth rate of the CEI in month t and y_{2t} indicating whether month t corresponds to a recession ($y_{2t} = 1$) or an expansion ($y_{2t} = 0$). In addition, we define the unobserved continuous variable z_{2t} that determines the binary variable y_{2t} in the familiar probit setting, that is, $y_{2t} = \mathbb{I}_{\{z_{2t}<0\}}$ (see, for example, Albert and Chib, 1993). The functional form of our general bivariate model without any parameter restrictions is given by

$$y_{1t} = \beta_0 + \sum_{j=1}^{J} \beta_j x_{j,t-\kappa_j} + \varepsilon_{1t} = \mu_{1t} + \varepsilon_{1t}, \tag{3.1}$$

$$z_{2t} = \alpha_0 + \sum_{j=1}^{J} \alpha_j x_{j,t-\kappa_j} + \varepsilon_{2t} = \mu_{2t} + \varepsilon_{2t}, \tag{3.2}$$

with the (standardized) symmetric growth rates $x_{j,t}$, $(j = 1, \ldots, J)$, of the leading indicators as explanatory variables. Since these variables are leading, the lead times $\kappa_j \geq 1$. Conditional on the model parameters (and explanatory variables), we assume the innovations $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ to be independent over time and to follow a bivariate normal distribution with zero mean and a covariance matrix of which the variance term of ε_{2t} is fixed at 1 for identification purposes. We note that we pool information by assuming that the lead times $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_J)'$ in both equations are the same, motivated by the fact that both dependent variables share the same contemporaneous cyclical component. Our alternative leading indexes follow from further pooling by imposing various restrictions across indicator weights $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)'$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)'$, as we cover next.

TCB assumes equal lead times and equal weighing of all ten leading variables to construct its composite leading economic index (LEI), which we further discuss in Section 3.3. In our

[1] For the moment we ignore the publication delay of several leading economic variables. That is, month t's vintage in the real-time data matrix does not contain values for some of the variables in month t yet. In our real-time data analysis we actually build a forecast with the month-t vintage for the business cycle in month t. Which is useful nevertheless since the first release of the latter is only in month $t+1$. We refer to Bańbura *et al.* (2013) for such “nowcasting”-related issues.

model in (3.1)–(3.2) this corresponds with $\kappa_j = \ell \geq 1$ (but unknown what value ℓ takes), $\beta_j = \bar{\beta}$, and $\alpha_j = \bar{\alpha}$, $(j = 1, \ldots, J)$. We first investigate to what extent the individual variables differ in the number of months they lead the business cycle, while maintaining the restriction of equal weights. We examine this case by setting $\beta_j = \bar{\beta}$ and $\alpha_j = \gamma\bar{\beta}$, $(j = 1, \ldots, J)$. Under these parameter restrictions we define the growth rate of our alternative leading index as $\bar{\beta}\sum_{j=1}^{J} x_{j,t-\kappa_j}$.

We cast the specification with equal weights in a bivariate seemingly unrelated (SUR) model (see Zellner, 1962, Geweke, 2005) in which each equation has the same regressor $\sum_j x_{j,t-\kappa_j}$ associated with the common parameter $\bar{\beta}$. The imposed restrictions provide for the identification of the variance of the latent process, since we write the equivalent of (3.2) as $z_{2t}^* = \alpha_0/\gamma + \bar{\beta}\sum_{j=1}^{J} x_{j,t-\kappa_j} + \varepsilon_{2t}^* = \mu_{2t}^* + \varepsilon_{2t}^*$. In this format, $\boldsymbol{\varepsilon}_t^* = (\varepsilon_{1t}, \varepsilon_{2t}^*)' \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}_2, \boldsymbol{\Sigma})$, $(t = 1, \ldots, T)$, with the reparameterized covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1/\gamma \\ \rho\sigma_1/\gamma & \gamma^{-2} \end{bmatrix}, \tag{3.3}$$

and ρ the correlation between the two error terms.[2] The intercept of the latent variable is now reparameterized as $\alpha_0^* = \alpha_0/\gamma$. In our empirical application though, we report results about α_0 to keep the interpretation in terms of the original model.

In addition to indicator-specific lead times, we also investigate the relative importance of the ten individual indicators. The Conference Board adds up the ten standardized leading variables and therefore ignores that some leading variables may contain a stronger predictive business-cycle signal than others. If we identify the relative strength of the individual indicators, we can suppress noise and extract an improved leading index. We impose a relative-importance restriction across the two equations. For example, if a variable is twice as important as another in signaling a recession, this variable is twice as important as the other in leading the coincident index as well, again for reasons of sharing a cycle without phase shifts. Formally the restriction means that we set $\boldsymbol{\alpha} = \gamma\boldsymbol{\beta}$. With this specification we define the growth rate of the alternative leading index as $\sum_{j=1}^{J} \beta_j x_{j,t-\kappa_j}$. Analogous to the equal-weights case, this restriction also allows us to rewrite the bivariate model in the SUR framework, each equation having the same ten explanatory variables $x_{j,\kappa_{t-j}}$, $(j = 1, \ldots, J)$, and the covariance matrix is again unrestricted.[3]

3.2.2 Statistical analysis

We perform the statistical analysis of our models in the Bayesian framework. One of the advantages of taking the Bayesian approach is that it straightforwardly deals with the mixed continuous-discrete parameter space. In this section we derive the likelihood function of the model's parameters, and we discuss prior specifications and the posterior simulation procedure.

[2]The main advantage of this reparameterization is that the covariance matrix is now unrestricted, which facilitates its sampling in the posterior simulation procedure needed for the Bayesian analysis of the model.

[3]The indicator-weight parameters are independent of the corresponding lead times. If we relax this assumption and use lead-time-specific parameters instead, the number of parameters increases dramatically and most of them are unidentified by the data anyway, especially if a lead time is *a posteriori* tightly concentrated around one value. Visually inspecting the marginal posteriors of the weights for multimodality is a good alternative to assess any lead-time dependence. We further note the close relation between β_j and κ_j, *i.e.*, the lead time which generates the highest correlation with the business cycle gets assigned most posterior weight. Therefore, the other lead times are associated with smaller weights (lower correlations).

Likelihood function

We collect the unknown parameters of (the various restricted versions of) the model in (3.1)–(3.3) in $\boldsymbol{\theta} = \{\boldsymbol{\kappa}, \bar{\beta}(\boldsymbol{\beta}), \boldsymbol{\Sigma}, \beta_0, \alpha_0^*\}$, and we derive the likelihood contributions $p(\mathbf{y}_t \,|\, \boldsymbol{\theta})$, $(t = 1, \dots, T)$. To that end, we first define the partly unobserved auxiliary variable $\mathbf{y}_t^* = (y_{1t}, z_{2t}^*)'$. Then we obtain

$$\begin{aligned} p(\mathbf{y}_t \,|\, \boldsymbol{\theta}) &= \int p(\mathbf{y}_t \,|\, z_{2t}^*, \boldsymbol{\theta}) p(z_{2t}^* \,|\, \boldsymbol{\theta}) \, \mathsf{d}z_{2t}^* = \int p(y_{2t} \,|\, z_{2t}^*, \boldsymbol{\theta}) p(\mathbf{y}_t^* \,|\, \boldsymbol{\theta}) \, \mathsf{d}z_{2t}^* \\ &\propto \int \left(\mathbb{I}_{\{y_{2t}=1\}} \mathbb{I}_{\{z_{2t}^*<0\}} + \mathbb{I}_{\{y_{2t}=0\}} \mathbb{I}_{\{z_{2t}^*\geq 0\}} \right) \\ &\qquad \times |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{ -(\mathbf{y}_t^* - \boldsymbol{\mu}_t^*)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t^* - \boldsymbol{\mu}_t^*)/2 \right\} \mathsf{d}z_{2t}^*, \end{aligned} \tag{3.4}$$

with the two regression means stacked in $\boldsymbol{\mu}_t^* = (\mu_{1t}, \mu_{2t}^*)'$. With the properties of the multivariate normal distribution we evaluate the integral analytically. We have $z_{2t}^* \,|\, \{y_{1t}, \boldsymbol{\theta}\} \sim \mathcal{N}(\mu_{2|1,t}, \sigma_{2|1}^2)$ with the conditional mean equal to one of the two following expressions

$$\begin{aligned} \mu_{2|1,t} &= \alpha_0^* + \bar{\beta} \sum\nolimits_{j=1}^J x_{j,t-\kappa_j} + \frac{\rho}{\gamma\sigma_1} \Big(y_{1t} - \beta_0 - \bar{\beta} \sum\nolimits_{j=1}^J x_{j,t-\kappa_j} \Big), \\ \mu_{2|1,t} &= \alpha_0^* + \sum\nolimits_{j=1}^J \beta_j x_{j,t-\kappa_j} + \frac{\rho}{\gamma\sigma_1} \Big(y_{1t} - \beta_0 - \sum\nolimits_{j=1}^J \beta_j x_{j,t-\kappa_j} \Big), \end{aligned} \tag{3.5}$$

(equal indicator weights, or heterogenous indicator weights, respectively), and conditional variance

$$\sigma_{2|1}^2 = \gamma^{-2}(1 - \rho^2). \tag{3.6}$$

The conditional probability of a recession given the CEI's growth rate in month t (and the values of the leading indicators) is therefore analytically available and is given by $\mathsf{Pr}\,[y_{2t} = 1 \,|\, y_{1t}, \boldsymbol{\theta}] = \Phi(-\mu_{2|1,t}/\sigma_{2|1})$, in which $\Phi(\cdot)$ is the cumulative distribution function of the univariate standard normal distribution. With these results, the likelihood contribution of the t-th observation in (3.4) simplifies to

$$p(\mathbf{y}_t \,|\, \boldsymbol{\theta}) = \Big[\Phi(-\mu_{2|1,t}/\sigma_{2|1}) \mathbb{I}_{\{y_{2t}=1\}} + \Phi(\mu_{2|1,t}/\sigma_{2|1}) \mathbb{I}_{\{y_{2t}=0\}} \Big] \sigma_1^{-1} \phi([y_{1t} - \mu_{1t}]/\sigma_1), \tag{3.7}$$

and $\phi(\cdot)$ is the density function of the univariate standard normal distribution. Finally, the integrand in (3.4) provides the contributions to the complete data likelihood function which we use for posterior simulation.

Prior distribution

In the model with equal weights as well as in the specification with indicator-dependent weights we regard the lead times as *a priori* mutually independent and specify a uniform discrete distribution for each κ_j. Thus, we have the densities

$$p(\kappa_j) \propto 1, \quad \kappa_j \in \{1, 2, \dots, K_j^{\max}\}, \quad (j = 1, \dots, J), \tag{3.8}$$

and for nine of the ten leading indicators we set a maximum lead time of twelve months. However, because we know that interest rate spreads generally lead up to one year or more

(Estrella and Mishkin, 1998), we set the maximum lead time of the spread between the ten-year T-bond minus the federal funds rate to 24.[4]

In the model with equal weights ($\beta_j = \bar{\beta}$ and $\alpha_j = \gamma\bar{\beta}$, for all ten indicators) we use a conditionally conjugate prior for the single weight parameter,

$$\bar{\beta} \sim \mathcal{N}(b, B), \tag{3.9}$$

with mean $b = 0$ and relatively large variance $B = 10$. This prior specification allows us to analytically marginalize the full conditional posterior of $\{\boldsymbol{\kappa}, \bar{\beta}\}$ with respect to $\bar{\beta}$, because conditionally we have a Gaussian linear regression model. For the remaining parameters we take conditionally conjugate priors in the format with the reparameterized latent variable z^*_{2t}. This corresponds to an inverted Wishart distribution for variance matrix $\boldsymbol{\Sigma}$ and a multivariate normal distribution for intercepts $\boldsymbol{\Pi} = (\beta_0, \alpha^*_0)' \,|\, \boldsymbol{\Sigma}$. We set these as follows,[5]

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\nu, \mathbf{S}) \tag{3.10}$$

$$\boldsymbol{\Pi} \,|\, \boldsymbol{\Sigma} \sim \mathcal{N}(\mathbf{p}', q\boldsymbol{\Sigma}), \tag{3.11}$$

with $\mathbf{p}' = \mathbf{0}_2$, $q = 10$, $\mathbf{S} = \text{diag}(15, 2)$ and $\nu = 6$. The two diagonal elements of parameter $\mathbf{S}$ of the inverted Wishart prior differ as the second diagonal element implies the marginal prior for proportionality parameter γ. In fact, we impose a Gamma prior for its square, $\gamma^2 \sim \mathcal{G}a((6-1)/2, 2/2)$. Since our two dependent variables are proxies for the business cycle, we assume the sign of the effect of our leading index on both of them to be positive and we safely consider $\gamma > 0$.

For the model with indicator-specific weights ($\boldsymbol{\alpha} = \gamma\boldsymbol{\beta}$) we stick to (3.10)–(3.11) for the intercepts and the variance matrix. In this model specification we have J different weights though, and we take the multivariate analogue of (3.9),

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{b}, \mathbf{B}). \tag{3.12}$$

We choose its hyperparameters such that we have a relatively uninformative prior distribution and we assume the weights to be independent. Therefore we set the J means equal to zero and the variance matrix is $\mathbf{B} = 10\mathbf{I}_J$.

Posterior simulation

To derive posterior results we apply simulation techniques. We first stack the observations in $\mathbf{Y}_i = (y_{i1}, \ldots, y_{iT})'$, $(i = 1, 2)$, and define the $T \times 2$ matrix $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$. We combine the likelihood $p(\mathbf{Y} \,|\, \boldsymbol{\theta})$ and the specified prior density $p(\boldsymbol{\theta})$, to obtain the kernel of the posterior density $p(\boldsymbol{\theta} \,|\, \mathbf{Y}) \propto p(\mathbf{Y} \,|\, \boldsymbol{\theta})p(\boldsymbol{\theta})$. By taking the product of the terms in (3.7) we compute the likelihood function in its closed form. However, in order to facilitate sampling from the posterior distribution, we implement a Markov chain Monte Carlo (MCMC) scheme with data augmentation (Robert and Casella, 2004), for which the complete data likelihood

[4] Priors that take into account mutual dependence (for example of the "grouping" type in variable-selection models as in Chipman, 1996) or non-uniformity (for instance of the Litterman type with decreasing probabilities for longer lead times) are alternatives for our lead-times prior without any serious computational complications

[5] The intercepts are conditional on $\boldsymbol{\Sigma}$ such that we sample both the intercepts and the variance matrix in one block. We further note that this prior specification implies that the original intercept of the probit equation $\alpha_0 \,|\, \boldsymbol{\Sigma} \sim \mathcal{N}(0, q)$, which does not depend upon proportionality parameter γ.

function is required. If we stack the partly unobserved vectors in $\mathbf{Y}^* = (\mathbf{y}_1^*, \ldots, \mathbf{y}_T^*)'$ and additionally define $\mathbf{z}_2^* = (z_{21}^*, \ldots, z_{2T}^*)'$, then we simulate from the latent-variable augmented posterior

$$p(\mathbf{z}_2^*, \boldsymbol{\theta} \,|\, \mathbf{Y}) \propto p(\mathbf{Y}_2 \,|\, \mathbf{z}_2^*, \boldsymbol{\theta}) p(\mathbf{Y}^* \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta}),$$

in which the first two densities on the right form the complete data likelihood function. The MCMC procedure consists of the following four steps.

Step 1. Sample the latent variable z_{2t}^* from its full conditional posterior $p(z_{2t}^* \,|\, \mathbf{Y}, \boldsymbol{\theta}, \mathbf{z}_{2,-t}^*) = p(z_{2t}^* \,|\, \mathbf{Y}, \boldsymbol{\theta})$, for $t = 1, \ldots, T$ (truncated conditional normal distributions);

Step 2. Sample the lead time κ_j from $p(\kappa_j \,|\, \mathbf{Y}, \mathbf{z}_2^*, \boldsymbol{\Pi}, \boldsymbol{\Sigma}, \boldsymbol{\kappa}_{-j})$, for $j = 1, \ldots, J$ (discrete distributions);

Step 3. Sample $\bar{\beta}$ ($\boldsymbol{\beta}$) from its full conditional posterior (normal distribution);

Step 4. Sample $\{\boldsymbol{\Sigma}, \boldsymbol{\Pi}\}$ in one block from its full conditional posterior (inverted Wishart distribution for the covariance matrix and a normal distribution for $\boldsymbol{\Pi}$ given $\boldsymbol{\Sigma}$).

The first step is similar to a standard simulation step for latent variables in a probit model (Albert and Chib, 1993). The difference is that we need to account for the dependence between y_{1t} and z_{2t}^* and therefore simulate from truncated *conditional* normal distributions. In the second step we first integrate out the indicator weight(s) analytically before sampling the J lead times. Because we expect the model to generate strong dependence between the lead times and the indicator weight(s), implementing this step improves the mixing of the MCMC sampler substantially. The two final steps of our simulation procedure are employed by using the well-known properties of the Bayesian normal regression model. We refer to Appendix 3.A for derivations and details about this simulation scheme.

3.2.3 Testing alternative indexes

To statistically assess whether our alternative leading indexes better predict the state of the economy relative to the commonly used TCB composite leading index, we propose a number of hypotheses and a way to test these. The basic principle of the testing procedure is to examine whether our methods are capable of retrieving information from the ten individual leading indicators relevant for business cycle forecasting, *in addition* to LEI's predictive power.

As we operate in a Bayesian setting, we compare models based on their relative posterior model probabilities, see, for example, Kass and Raftery (1995) and Geweke (2005). Our testing problem is formulated in terms of comparing nested models. We augment the two regression equations (3.1) and (3.2) with $\text{LEI}_{t-\ell}$, TCB's composite leading index $\ell \geq 1$ months lagged. If in this augmented model the coefficient(s) of our index is(are) different from zero, our method extracts business-cycle leading information which is not incorporated in $\text{LEI}_{t-\ell}$.

We define the benchmark model M_1 as the bivariate model only including an intercept and TCB's LEI ℓ months lagged. Models M_2 and M_3 are the equal-weights model and the indicator-dependent-weights specification, respectively, both augmented with regressor $\text{LEI}_{t-\ell}$. To complete the prior specifications of all three models, we define $\widetilde{\boldsymbol{\Pi}} = (\boldsymbol{\Pi}, \boldsymbol{\Pi}_{\text{LEI}})'$ and extend the prior for the intercepts in (3.11) to also cover the two parameters $\boldsymbol{\Pi}_{\text{LEI}} =$

$(\beta_{\text{LEI}}, \alpha^*_{\text{LEI}})'$ associated with LEI. As a result, we obtain the matricvariate normal prior distribution[6]

$$\widetilde{\mathbf{\Pi}} \,|\, \mathbf{\Sigma} \sim \mathcal{MN}(\mathbf{0}_{2\times 2}, 10\mathbf{\Sigma} \otimes \mathbf{I}_2). \tag{3.13}$$

The priors for the other parameters $\{\boldsymbol{\kappa}, \mathbf{\Sigma}, \bar{\beta}(\boldsymbol{\beta})\}$ are the same as before. We compare models M_2 and M_3 to benchmark specification M_1. This corresponds to the hypotheses $\bar{\beta} = 0$ in M_2 and $\boldsymbol{\beta} = \mathbf{0}_J$ in M_3. Under these hypotheses, both γ and $\boldsymbol{\kappa}$ are nuisance parameters. However, since they have proper prior distributions in the nested M_1 this forms no problem for Bayesian model comparison (see Koop and Potter, 1999, for this Davies' problem when computing Bayes factors).

In our prior specifications it holds that $p(\boldsymbol{\theta}_{-\bar{\beta}} \,|\, \bar{\beta} = 0, M_2) = p(\boldsymbol{\theta}_{-\bar{\beta}} \,|\, M_1)$ and $p(\boldsymbol{\theta}_{-\boldsymbol{\beta}} \,|\, \boldsymbol{\beta} = \mathbf{0}_J, M_3) = p(\boldsymbol{\theta}_{-\boldsymbol{\beta}} \,|\, M_1)$, which ensues from our stronger assumption of prior independence between the leading-indicator weight(s) and all other parameters. With these criteria satisfied, we rely on the Savage–Dickey (SD) density ratio (Dickey, 1971, Verdinelli and Wasserman, 1995) to compute the two required Bayes factors

$$BF_{2,1} = \frac{p(\mathbf{Y} \,|\, M_2)}{p(\mathbf{Y} \,|\, M_1)} = \frac{p(\bar{\beta} = 0 \,|\, M_2)}{p(\bar{\beta} = 0 \,|\, \mathbf{Y}, M_2)}, \tag{3.14}$$

$$BF_{3,1} = \frac{p(\mathbf{Y} \,|\, M_3)}{p(\mathbf{Y} \,|\, M_1)} = \frac{p(\boldsymbol{\beta} = \mathbf{0}_J \,|\, M_3)}{p(\boldsymbol{\beta} = \mathbf{0}_J \,|\, \mathbf{Y}, M_3)}. \tag{3.15}$$

The advantage of the use of the SD density ratio is that a Bayesian analysis of model specifications M_2 and M_3 suffices and that we only have to evaluate the marginal posterior density of the parameters of interest in the hypothesized values. The height of the prior densities in the numerators in (3.14) and (3.15) are easily computed given our prior specification in (3.9) and (3.12), respectively. To obtain the denominators we adjust Step 4 of our posterior simulation scheme to sample $\widetilde{\mathbf{\Pi}}$ instead of $\mathbf{\Pi}$ (see Appendix 3.A), and use the MCMC output to Rao–Blackwellize (Gelfand and Smith, 1990) the conditional posteriors of $\bar{\beta}$ and $\boldsymbol{\beta}$ following from Step 3 of the simulation routine. In our application we compute the Bayes factors of M_2 and M_3 versus M_1 for different lead times of LEI, that is, $\ell = 1, 2, \ldots, 6$.

Finally, the model with merely $\text{LEI}_{t-\ell}$ as explanatory variable is nested in both M_2 and M_3. Moreover, the prior for the parameters in M_1 under the nesting in M_2 is the same as in the nested version of M_3, *i.e.*, $p(\boldsymbol{\theta}_{-\bar{\beta}} \,|\, M_1) = p(\boldsymbol{\theta}_{-\boldsymbol{\beta}} \,|\, M_1)$. Thus, the marginal likelihoods of M_1 in the two Bayes factors cancel out and we use (3.14) and (3.15) to compute $BF_{3,2}$ for testing the equal-weights restriction.

3.2.4 Forecasting recessions

We demonstrate the practical usefulness of our new indexes by implementing an out-of-sample forecasting procedure to predict recessionary periods. At the end of month $s-1$ we construct the one-month-ahead recession-probability forecast for next month s. This procedure requires two main inputs. The first consists of the set of leading-indicator data upon which the leading index is based. We use both the revised and real-time variants with

[6] As for the intercept of the probit equation, we also specify a prior for the parameter associated with LEI in terms of its reparameterized form. *I.e.*, $\alpha^*_{\text{LEI}} = \alpha_{\text{LEI}}/\gamma$, and for the original parameter it therefore holds that $\alpha_{\text{LEI}} \sim \mathcal{N}(0, 10)$, which no longer depends upon the proportionality parameter.

the notable difference in interpretation as we discuss in Section 3.3. The second input is the posterior distribution in month $s-1$, which has density $p(\boldsymbol{\theta}\,|\,\mathbf{Y}^{1,s-1})$.

In the Bayesian framework, the probability forecast we are interested in is the likelihood of a recession in month s weighed according to the posterior distribution updated up to and including month $s-1$. This way we integrate out parameter uncertainty both with respect to the unknown lead times and the leading-indicator weights. For $s=\tau+1,\ldots,T$ we compute the series of marginalized recession probabilities

$$\begin{aligned}\mathsf{Pr}\left[y_{2s}=1\,|\,\mathbf{Y}^{1,s-1}\right] &= \int\int \mathsf{Pr}\left[y_{2s}=1\,|\,z^*_{2s}\right] p(\mathbf{y}^*_s\,|\,\boldsymbol{\theta})p(\boldsymbol{\theta}\,|\,\mathbf{Y}^{1,s-1})\,\mathsf{d}\mathbf{y}^*_s\,\mathsf{d}\boldsymbol{\theta} \\ &= \int \Phi(-\gamma\mu^*_{2s})p(\boldsymbol{\theta}\,|\,\mathbf{Y}^{1,s-1})\,\mathsf{d}\boldsymbol{\theta}. \end{aligned} \tag{3.16}$$

For this expression we first analytically marginalize $p(\mathbf{y}^*_s\,|\,\boldsymbol{\theta})$ with respect to the growth rate of TCB's composite coincident index y_{1s} and obtain $z^*_{2s}\,|\,\boldsymbol{\theta}\sim\mathcal{N}(\mu^*_{2s},\gamma^{-2})$, resulting in the probit expression in (3.16).

In order to avoid rerunning the MCMC scheme each month to update posterior beliefs, we use importance sampling techniques to monthly obtain the posterior distribution. We use $p(\boldsymbol{\theta}\,|\,\mathbf{Y}^{1,\tau})$ as importance density and for each forecast we recursively compute the importance weights according to

$$w_s(\boldsymbol{\theta})\propto p(\mathbf{y}_{s-1}\,|\,\boldsymbol{\theta})w_{s-1}(\boldsymbol{\theta}),\quad (s=\tau+2,\ldots,T),$$

with initialization $w_{\tau+1}=1$. We note that in each iteration we evaluate the likelihood contributions of the *bivariate* model in (3.4) to obtain the importance weights, despite the fact that we are primarily interested in the recession-probability forecast.

For out-of-sample model comparison, we also make recession forecasts with a model only containing TCB's leading index, with unknown lead time ℓ though. To compare the forecasting performance of our leading indexes to the LEI's, we compute their respective marginalized predictive likelihoods. As we are particularly interested in just one scalar element (recession indicator) of the bivariate dependent variable, we marginalize with respect to the other variable (CEI's growth rate) and evaluate

$$\begin{aligned}p(\mathbf{Y}_2^{\tau+1,T}\,|\,\mathbf{Y}^{1,\tau}) &= \int\int p(\mathbf{Y}^{\tau+1,T}\,|\,\boldsymbol{\theta})p(\boldsymbol{\theta}\,|\,\mathbf{Y}^{1,\tau})\,\mathsf{d}\mathbf{Y}_1^{\tau+1,T}\,\mathsf{d}\boldsymbol{\theta} \\ &= \int \prod_{s=\tau+1}^{T}[\Phi(-\gamma\mu^*_{2s})\mathbb{I}_{\{y_{2s}=1\}}+\Phi(\gamma\mu^*_{2s})\mathbb{I}_{\{y_{2s}=0\}}]p(\boldsymbol{\theta}\,|\,\mathbf{Y}^{1,\tau})\,\mathsf{d}\boldsymbol{\theta},\end{aligned}$$

in the final $T-\tau$ recession-expansion observations of our sample.

In addition to the one-month-ahead models, we also make forecasts for three months and six months ahead. We obtain these forecasts in the "direct forecast" set-up as described in Marcellino *et al.* (2006). That is, we directly link the dependent variable h months ahead to information available at the end of month t. Equivalently, we say that we model month t's state of the business cycle using the real-time data vintage of month $t-h$, and we have restricted prior support $\kappa_j\geq h$ for the J lead times since more recent observations are not available yet. As a result we also obtain alternative leading economic indexes "optimized" for a given forecasting horizon h, instead of the single leading index as released by The Conference Board.

Table 3.1 *Leading and coincident economic indicators*

	Indicator's description	Ab.	TCB mn.	Tr.
	Panel A: Leading indicators			
1.	Number of hours average workweek, production workers	WHM	USHKIM..O	*sgr*
2.	Average weekly initial claims, state unemployment insurance	CUI	USUNINSCE	*–sgr*
3.	Manufacturers' new orders, consumer goods and materials	NOC	USCNORCGD	*sgr*
4.	Vendor performance, slower deliveries diffusion index	VPI	USVENDOR	*sgr*
5.	Manufacturers' new orders, nondefense capital goods	NOK	USNOIDN.D	*sgr*
6.	Number of newly issued building permits, private housing	NBP	USHOUSATE	*sgr*
7.	Stock prices, index 500 common stocks	SP5	US500STK	*sgr*
8.	Money supply, M2	M2	USM2....D	*sgr*
9.	Interest rate spread, 10-year Treasury bonds less federal funds	10TFF	USYSTNFF	*le*
10.	Index of consumer expectations	CCE	USUMCONEH	*sgr*
11.	**Composite Leading Index**	LEI	USCYLEAD	*sgr*
	Panel B: Coincident indicators			
12.	Employment	EMP	USEMPNAGE	*sgr*
13.	Industrial Production	IP	USINPRODG	*sgr*
14.	Personal income less transfer payments	PI	USPILESTD	*sgr*
15.	Manufacturing and trade sale	MTS	USBSSALED	*sgr*
16.	**Composite Coincident Index**	CEI	USCOININ	*sgr*

Notes: The table lists the ten individual leading economic variables in Panel A and the four variables that represent the current state of the economy in Panel B. "Ab." contains the indicator's abbreviation we use; "TCB mn." reports the variable's mnemonic applied by TCB; the final column ("Tr.") shows whether we transform the raw data series to symmetric growth rates (sgr) or use its level (le). The growth rate of initial claims of unemployment insurance is multiplied by -1 because of its inverse relation with the state of the economy compared to the other indicators.

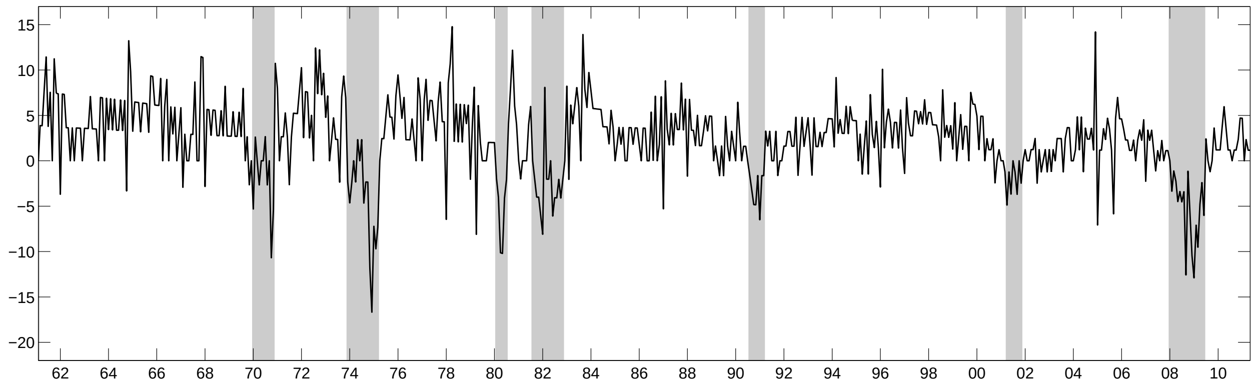
3.3 Data and TCB's methods

The empirical business cycle literature distinguishes three types of macroecnomic variables. A variable either leads the business cycle, coincides with the current state of the economy, or reacts sluggishly to economic conditions and therefore lags today's economic situation. In this chapter we use indicators from the first two classes. We apply coincident indicators and the NBER-announced recession dates as measures of the current economic conditions and we link them to predictive information available at earlier points in time, contained in the leading variables. Well-known leading variables are interest rate spreads, stock market returns and the number of new building permits for private housing.

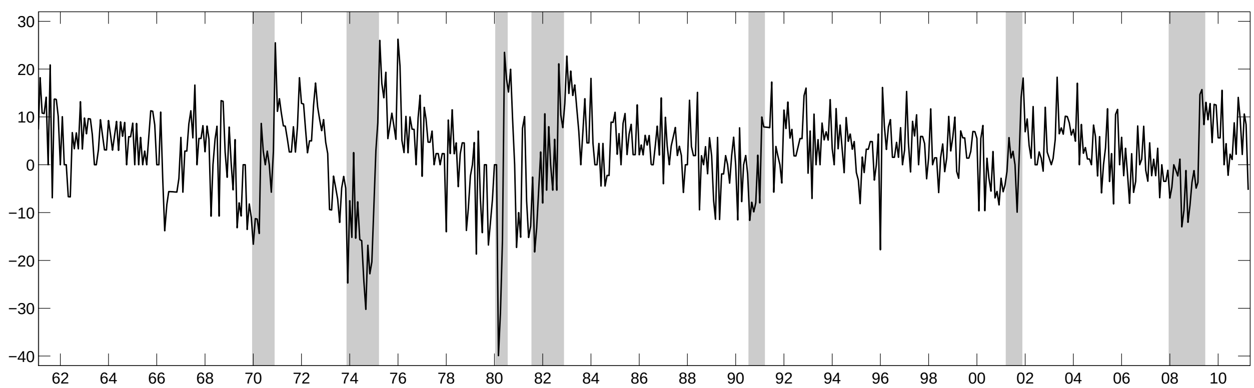
In Panel A of Table 3.1 we list the ten variables that build TCB's composite leading economic index (LEI).[7] We base our alternative leading index on this same set of variables. Panel B of the table shows the four macroeconomic series that jointly form TCB's composite coincident index (CEI), namely employment, industrial production, personal income and manufacturing sales.

[7] The Conference Board changed the definition of LEI in January 2012, replacing new orders for nondefense capital goods, vendor performance, M2 and consumer expectations with alternative variables. Primarily to prevent the use of *ex-post* knowledge and to establish a fair comparison between real-time and final data, we base our analysis on the set of variables that constituted LEI up to December 2011.

Figure 3.1 *TCB's composite economic indexes, February 1961–May 2011*



(a) One-month symmetric growth rates of the coincident index



(b) One-month symmetric growth rates of the leading index

In order to summarize (part of) the information contained in the individual variables, TCB applies a simple method to construct the single index (The Conference Board, 2001). First, the individual series are transformed by computing month-to-month symmetric growth rates, except for the long-minus-short interest rate spread which is used in levels. Second, the transformed series are standardized by division by their sample standard deviations. Because most U.S. macroeconomic time series show a large and sudden drop in volatility halfway the 1980s (see, for example, Kim *et al.*, 2008), these standardization factors are computed separately for the pre- and post-1983 subsamples. Third, the transformed and standardized series are summed to obtain the single-index growth rate. For the sample period February 1961–May 2011 we present the one-month symmetric growth rates of CEI and LEI in Figures 3.1(a) and (b), respectively. The shaded areas depict the recession periods as decided upon by the NBER business cycle dating committee.

For the construction of the composite *coincident* index, The Conference Board takes on a straightforward strategy with a clear interpretation. The composite coincident index at time t reflects the state of the economy for that month. The interpretation of the composite leading index is however less clear and its construction is restrictive. First, it is not immediately clear how many months LEI leads the business cycle. Second, it may well be that the lead times of the ten individual indicators differ and hence a simple contemporaneous average does not necessarily contain the strongest leading signal possible. Third and finally, all ten indicators are regarded as equally important, in the sense that they all get equal weight (of course apart from the standardization). Attributing relatively large weights to less informative leading indicators adds noise to the composite leading index and blurs the predictive signal.

We extract an alternative leading index both with fully revised and with real-time data on the leading indicators. For reasons of data availability as well as to compare our alternative indexes fairly to TCB's, we use the same set of individual leading variables. In addition to the practical issue of data availability, we also have two distinct interpretations of the results dependent upon the type of data used. With the real-time data we estimate the direct link between information on the leading variables as it was historically available at the actual time of the forecast construction and the true state of the economy. Since we are interested in forecasting "truth," we do use final data for dependent variable CEI (Diebold and Rudebusch, 1991a).

At the moment of forecasting we use the most recently available releases of the ten leading variables. Hence, using leading indicator j's κ_j-months-lagged values, means using partly revised data since we extract the κ_j-th off-diagonal of its real-time data matrix. Therefore, in this scenario we examine the predictive power of real-time and partly revised data releases, and we derive a leading index that provides practically useful forecasts in real time and that is optimal in terms of highest correlation with the business cycle measures. On the other hand, if we use the final data instead, we estimate the "true," reduced-form relation between leading indicators and the current state of the business cycle, of course under the assumption that data revisions lead to the true values.

3.4 Results

In this section we first present the results of the testing procedure to compare the various alternative leading indexes. Next, we discuss posterior properties of the model parameters, including lead times and relative weights. Third, we assess the outcomes of the out-of-sample forecasting experiment, and we conclude with a descriptive statistical analysis of the alternative indexes.

3.4.1 Testing

The top row of Panel A of Table 3.2 shows the Bayes factors of the equal-weights model with heterogenous lead times (M_2), versus the model with LEI ℓ months lagged as single regressor, when we use latest-available data. These factors are on a $\log_{10}$-scale, so we use Jeffreys' guidelines to assess whether our method improves on TCB's, using 2 as a "critical value." Considering the composition of LEI, if we use its ℓ-months lag, we test whether all ten leading indicators have equal lead time ℓ by using the $BF_{2,1}$'s. For all six lags of LEI we obtain Bayes factors in the range 52–63. Hence, for all values of ℓ, there is always at least one leading indicator which has a lead time different from ℓ. This result provides statistically convincing evidence for heterogenous lead times across the ten leading indicators.

Testing whether also relaxing the equal-weights restriction extracts a stronger leading signal for the business cycle results in the Bayes factors in the second row of Panel A. For this model with both heterogenous lead times and indicator weights (M_3) we obtain factors in a range 58–70. This again provides overwhelming evidence in favor of our new leading index relative to TCB's LEI. To compare M_3 to M_2 we subtract their respective Bayes factors versus the LEI model. The resulting factors are displayed in the third row of Panel A. For

Table 3.2 *Bayes factors nested model testing*

Model pair	LEI's lag ℓ					
	1	2	3	4	5	6
Panel A: Revised data						
M_2 vs. M_1	63.7	56.6	52.4	57.1	58.4	55.8
M_3 vs. M_1	70.1	67.6	61.3	58.7	64.0	66.5
M_3 vs. M_2; both contain $\text{LEI}_{t-\ell}$	6.4	11.0	8.9	1.6	5.6	10.7
Panel B: Real-time data						
M_2 vs. M_1	70.0	66.3	66.7	71.3	62.5	71.2
M_3 vs. M_1	73.4	79.3	72.6	78.1	79.9	82.8
M_3 vs. M_2; both contain $\text{LEI}_{t-\ell}$	3.4	13.0	5.9	6.7	17.4	11.7

Notes: These Bayes factors for nested model testing are the result of applying the Savage–Dickey density ratio. We compare the following models: M_1 is the bivariate model with as single predictor TCB's leading index ℓ months lagged; M_2 is the bivariate model with heterogenous lead times, but each leading variable is assumed to be equally important, and augmented with LEI ℓ months lagged; M_3 is the bivariate model with both heterogenous weights and lead times, again augmented with LEI ℓ months lagged. Bayes factors are on the $\log_{10}$-scale.

five out of the six lead times, the Bayes factors indicate that allowing for indicator-specific weights decisively improves over the case with only heterogenous lead times. Since the Bayes factor is partly a trade-off between the fit of the data (as measured by the likelihood) and the parsimony of the model specification (as expressed in the prior-posterior difference), we observe that introducing nine additional parameters in M_3 relative to M_2 results in a decisively better predictive description of the business cycle. The smallest Bayes factor we find is $BF_{3,2} = 1.6$, for the LEI four months lagged. Hence, following Jeffreys, this represents not decisive but still very strong support for heterogenous weights of the indicators relative to all leading variables equally important.

If we apply the same set of tests to our leading indexes constructed with real-time data, we obtain the results depicted in Panel B of Table 3.2. The first row shows that our model with indicator-specific lead times retrieves important additional business-cycle-leading information compared to the LEI available in real time, for all lags considered. It does so even more decisively than with fully revised data.

We find similar results for M_3 against the ℓ-months-lagged LEI. The second row in Panel B shows Bayes factors ranging between 72 and 82, assigning virtually zero weight to the model with only LEI. The final row of the table shows that, also for real-time data, some leading variables turn out to be decisively more important than others in predicting the business cycle. In summary, these testing results show overwhelming support for a leading index with both indicator-specific lead times and weights. It extracts additional predictive power from the set of ten leading indicators, relative to LEI. This improvement is largest is we apply data of the real-time type.

3.4.2 Posterior results

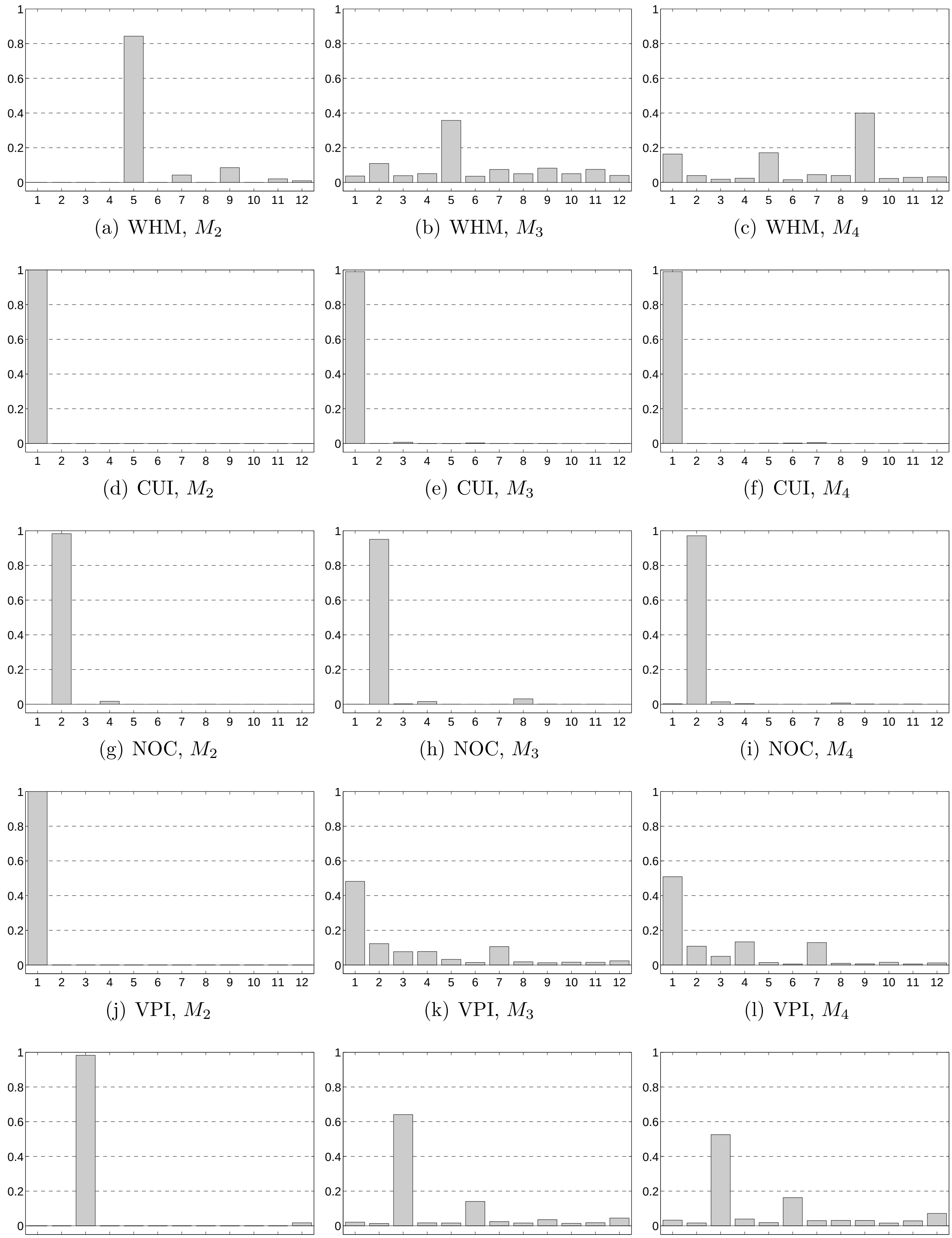
The left columns of Figures 3.2–3.3 show the marginal posterior distributions of the lead times of the ten leading indicators in the equal-weights model, obtained with latest-available data. All ten lead times have nearly all posterior weight concentrated at one particular value, with the notable exception of the interest rate spread (10TFF) in Figure 3.3(j). We distinguish two groups of indicators. One group consists of leading variables having relatively short lead times up to three months. Unemployment benefit claims (CUI), new orders for consumer goods (NOC), vendor performance index (VPI), new orders for capital goods (NOK) and new building permits (NBP) belong to this group. The other group of indicators have (relatively) long lead times. Weekly hours worked (WHM), S&P500 index (SP5), money supply growth (M2), consumer expectations (CCE) and 10TFF lead the business cycle at least five months. The latter has a more dispersed posterior lead-time distribution with probability mass distributed over the range of twelve to seventeen months, with the mode at 13.

If we look at the marginal posterior properties of the other parameters of model M_2 in Panel A of Table 3.3, we see that we estimate the indicator weight $\bar{\beta} \approx 0.60$. The fact that in M_2 all ten indicators are forced to have this same effect causes the very pronounced posteriors of the lead times. For example, if a few indicators are more important than others, they leverage the value of $\bar{\beta}$ and for the less important indicators the lead time corresponding closest to this weight is selected. We infer whether this issue is indeed affecting the results from the middle columns of Figures 3.2–3.3, which display the marginal posteriors of the ten lead times in M_3, the model with a potentially unequal weighing of the individual leading variables. We make the following four remarks with regard to the change in the posteriors of the lead times and the rank of the individual indicators in terms of relative importance, due to the relaxation of the uniform-importance restriction when going from M_2 to M_3.

First, we identify two lead times (WHM and VPI) that show spiked posteriors in M_2, but which have more dispersed distributions in M_3. Figures 3.4(a) and 3.4(j) depict the marginal posterior densities of the respective weights β_{WHM} and β_{VPI} and show that these leading variables are less important in constructing our new leading index, with considerable posterior mass around 0. We note that if the weight $\beta_j = 0$ for indicator j, then its lead-time parameter is not identified by the data and κ_j's posterior closely resembles its uniform prior. Equivalently, if the posterior of κ_j shows pronounced support for particular lead times, its corresponding weight is different from 0. For weekly hours worked we find an almost uniformly distributed posterior [Figure 3.2(b)] and vendor performance's posterior assigns slowly decreasing weights to lead times ranging from one month to seven months [Figure 3.2(k)].

Second, three lead times (NOK, NBP and CCE) have bi-modal marginal posteriors in M_3. The lead time of new orders for capital goods changes from a clear three-month lead to either three or six months. For new building permits posterior mass partly shifts from the relatively short lead time of three months in M_2 to longer periods of nine to eleven months in M_3. The lead time of consumer expectations goes the opposite direction: it is sharply estimated at ten months in M_2, but part of the posterior mass is generously redirected to $\kappa_{\text{CCE}} = 2$ in M_3. In Figures 3.4(m), 3.5(a) and 3.5(m) we display the weights of these three leading indicators in building our alternative leading index. Although all three have substantial positive weight with posterior modes around 0.30, there is minor support for 0 as well.

Figure 3.2 *Marginal posteriors of lead times, revised data*



(a) WHM, M_2 (b) WHM, M_3 (c) WHM, M_4

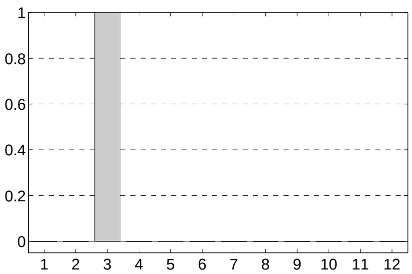
(d) CUI, M_2 (e) CUI, M_3 (f) CUI, M_4

(g) NOC, M_2 (h) NOC, M_3 (i) NOC, M_4

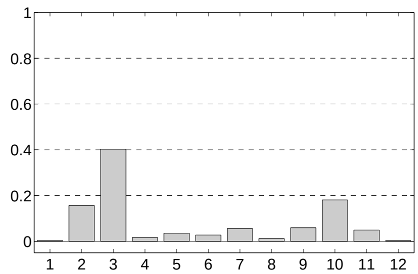
(j) VPI, M_2 (k) VPI, M_3 (l) VPI, M_4

(m) NOK, M_2 (n) NOK, M_3 (o) NOK, M_4

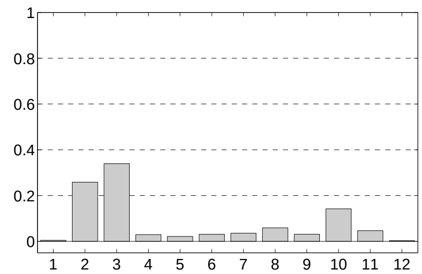
Figure 3.3 *Marginal posteriors of lead times, revised data,* Cont.



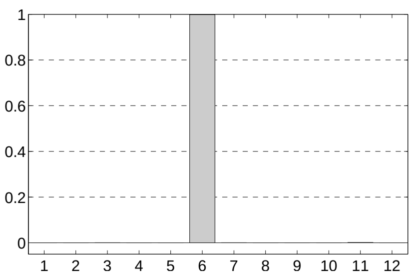
(a) NBP, M_2



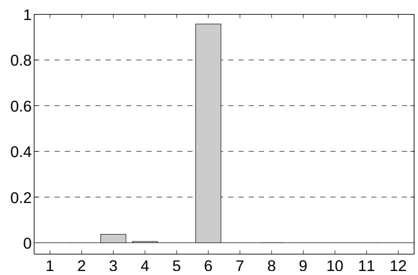
(b) NBP, M_3



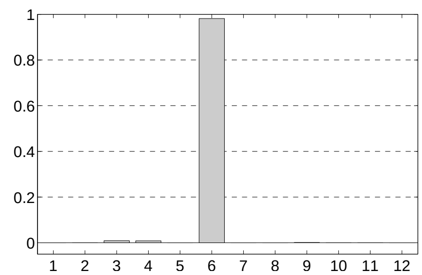
(c) NBP, M_4



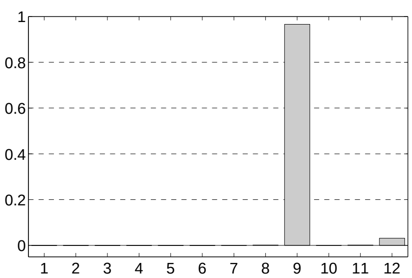
(d) SP5, M_2



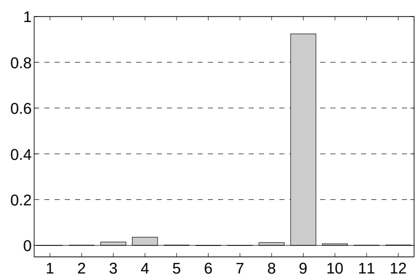
(e) SP5, M_3



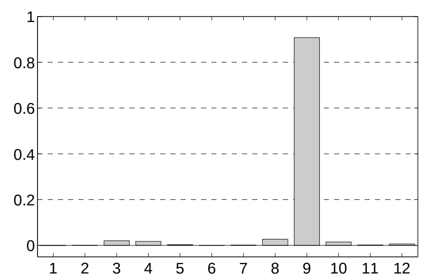
(f) SP5, M_4



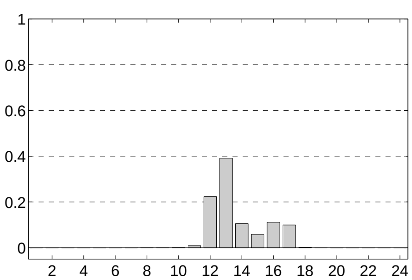
(g) M2, M_2



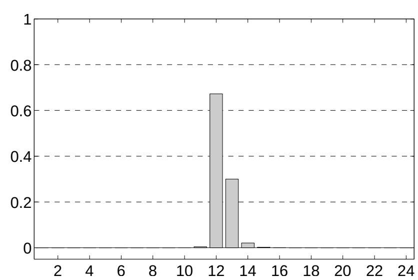
(h) M2, M_3



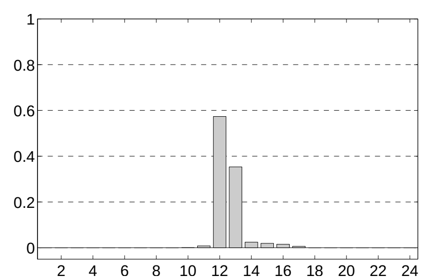
(i) M2, M_4



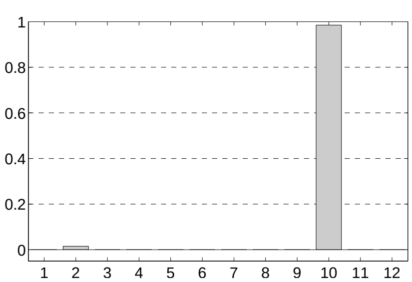
(j) 10TFF, M_2



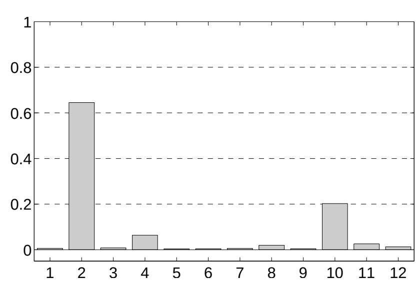
(k) 10TFF, M_3



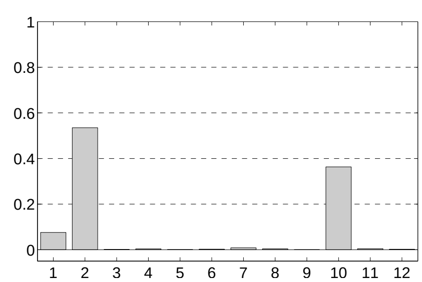
(l) 10TFF, M_4



(m) CCE, M_2

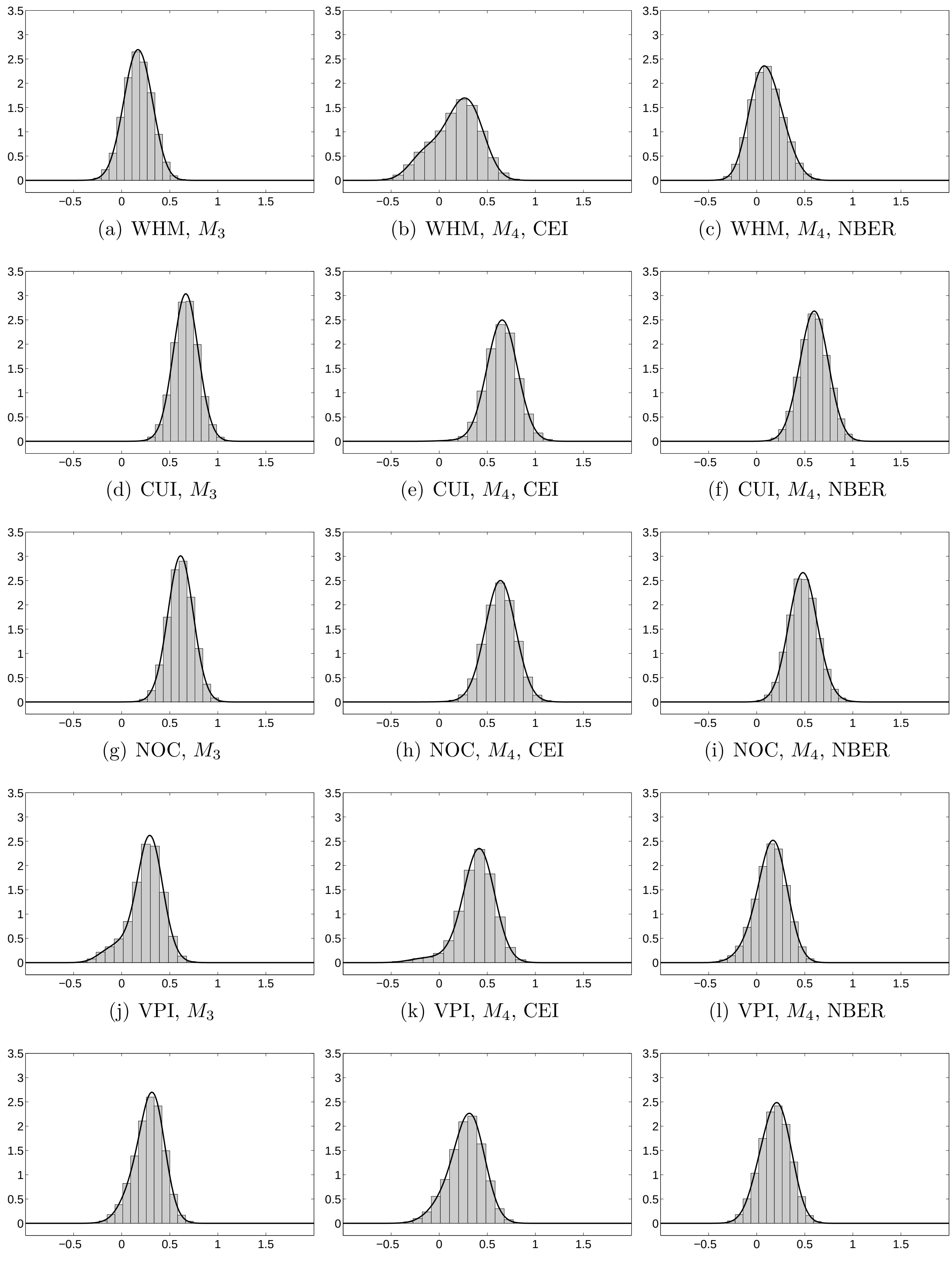


(n) CCE, M_3



(o) CCE, M_4

Figure 3.4 *Marginal posteriors of leading-indicator weights, revised data*



(a) WHM, M_3

(b) WHM, M_4, CEI

(c) WHM, M_4, NBER

(d) CUI, M_3

(e) CUI, M_4, CEI

(f) CUI, M_4, NBER

(g) NOC, M_3

(h) NOC, M_4, CEI

(i) NOC, M_4, NBER

(j) VPI, M_3

(k) VPI, M_4, CEI

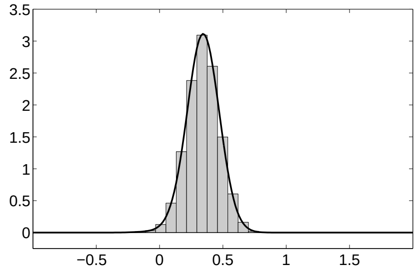
(l) VPI, M_4, NBER

(m) NOK, M_3

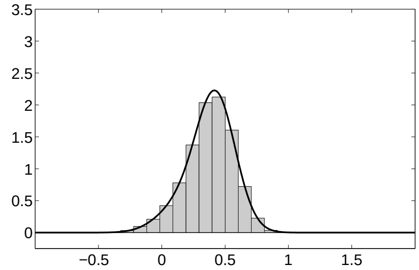
(n) NOK, M_4, CEI

(o) NOK, M_4, NBER

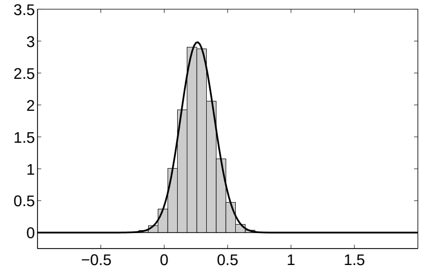
Figure 3.5 *Marginal posteriors of leading-indicator weights, revised data,* Cont.



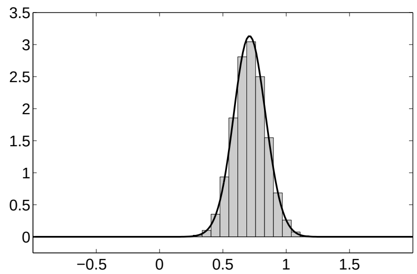
(a) NBP, M_3



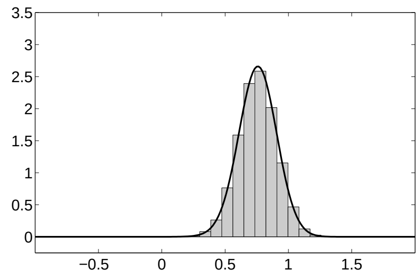
(b) NBP, M_4, CEI



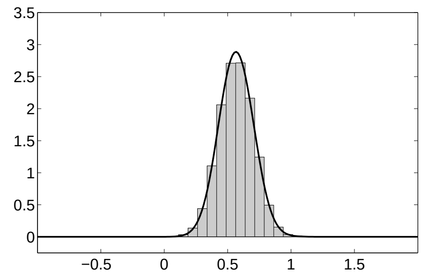
(c) NBP, M_4, NBER



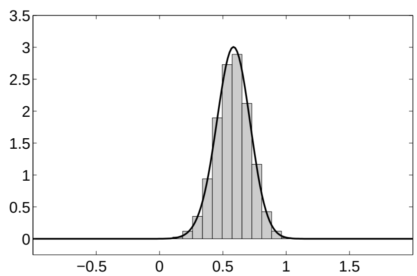
(d) SP5, M_3



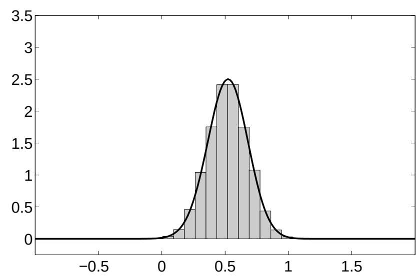
(e) SP5, M_4, CEI



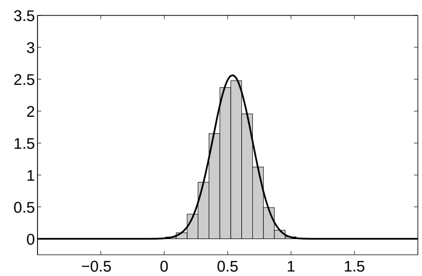
(f) SP5, M_4, NBER



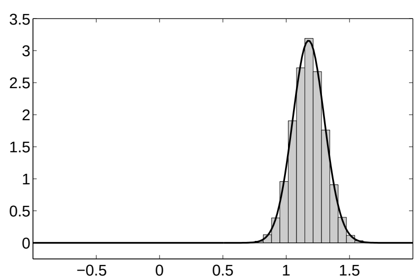
(g) M2, M_3



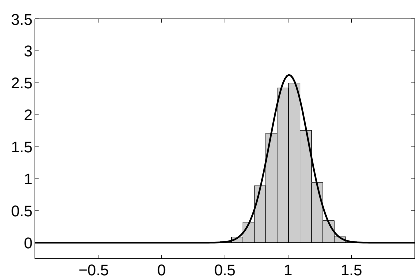
(h) M2, M_4, CEI



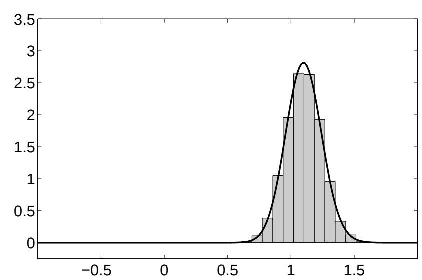
(i) M2, M_4, NBER



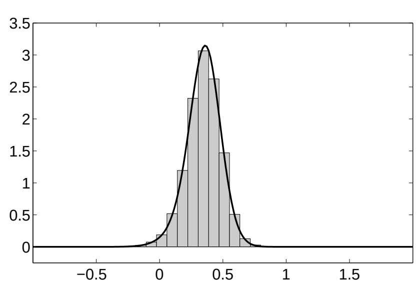
(j) 10TFF, M_3



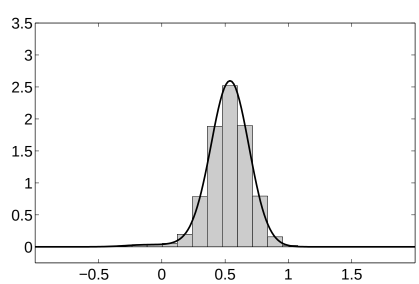
(k) 10TFF, M_4, CEI



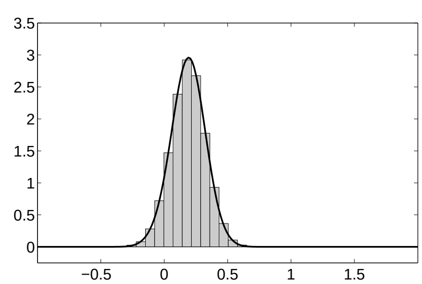
(l) 10TFF, M_4, NBER



(m) CCE, M_3



(n) CCE, M_4, CEI



(o) CCE, M_4, NBER

Table 3.3 *Properties of marginal posterior distributions*

Parameter	Mean	St.D.	Percentiles 5th	Percentiles 95th	Mean	St.D.	Percentiles 5th	Percentiles 95th
	Panel A: Revised data				*Panel B: Real-time data*			
Model M_2								
β_0	2.39	0.14	2.16	2.62	2.39	0.14	2.15	2.63
α_0	1.50	0.10	1.33	1.68	1.55	0.12	1.36	1.76
$\bar{\beta}$	0.58	0.04	0.52	0.65	0.59	0.04	0.52	0.66
σ_1^2	12.24	0.71	11.11	13.46	12.58	0.74	11.40	13.85
ρ	0.71	0.05	0.63	0.78	0.70	0.05	0.62	0.77
γ	0.50	0.05	0.43	0.59	0.54	0.06	0.44	0.64
Model M_3								
β_0	2.39	0.14	2.15	2.62	2.39	0.14	2.15	2.63
α_0	1.64	0.12	1.45	1.85	1.63	0.12	1.44	1.84
β_{WHM}	0.17	0.15	−0.08	0.41	0.18	0.15	−0.08	0.42
β_{CUI}	0.67	0.13	0.45	0.88	0.71	0.13	0.49	0.93
β_{NOC}	0.61	0.13	0.39	0.83	0.39	0.14	0.16	0.62
β_{VPI}	0.25	0.18	−0.10	0.51	0.43	0.14	0.20	0.66
β_{NOK}	0.28	0.16	0.01	0.52	0.35	0.16	0.07	0.58
β_{NBP}	0.34	0.13	0.13	0.55	0.35	0.13	0.14	0.57
β_{SP5}	0.71	0.13	0.50	0.92	0.69	0.13	0.48	0.90
β_{M2}	0.58	0.14	0.36	0.80	0.51	0.13	0.29	0.73
β_{10TFF}	1.18	0.13	0.97	1.39	1.22	0.13	1.01	1.43
β_{CCE}	0.35	0.14	0.11	0.56	0.34	0.14	0.11	0.56
σ_1^2	12.28	0.73	11.13	13.52	12.56	0.75	11.38	13.85
ρ	0.72	0.05	0.63	0.79	0.72	0.05	0.64	0.79
γ	0.57	0.05	0.48	0.66	0.57	0.05	0.48	0.66

Notes: Model M_2 is the bivariate model in which all leading indicators have equal weight $\bar{\beta}$; in Model M_3 each of the ten leading variables has its own weight β_j; both specifications allow for heterogenous lead times.

Third, a group of four leading variables (CUI, NOC, SP5 and M2) have the same spiked lead-time posteriors in both models. If we look at model M_3's parameters in Panel A of Table 3.3 and the plots of the marginal posterior densities in Figures 3.4 and 3.5, we also observe that these four variables are in-group equally important in our new leading index. With posteriors of their β_j's clearly shifted away from 0 and all four modes close to 0.60, they are more dominant than NOK, NBP and CCE, and certainly than WHM and VPI.

Fourth, the spread between the 10-year T-bond and the effective federal funds rate is evidently the most dominant constituent of our leading index. Figure 3.5(j) shows that β_{10TFF} has a posterior mode of 1.2, which makes this spread twice as important as CUI, NOC, SP5 and M2. The fact that 10TFF is allowed to have its own weight makes the posterior of its lead time also more pronounced. In Figure 3.3(k) we observe that the interest rate spread leads the business cycle exactly one year or thirteen months.

The left columns of Figures 3.6–3.7 show the posterior distributions of the lead times in the equal-weights model when we use real-time data instead. While we find evidence for a single lead-time value in M_2 with revised data for VPI, M2 and CCE, with real-time data the posterior mass is distributed among two lead-time values. The index of vendor performance leads one or two months, money supply four or eleven months, and consumer expectations four or ten months. If we apply the model with indicator-specific weights, these three leading variables keep this feature, see Figures 3.6(k), 3.7(h) and 3.7(n).

In Panel B of Table 3.3 we observe that WHM also has small relative importance in determining our leading index with real-time data. NOC loses importance compared to the leading index based on revised data, with its weight declining from 0.6 to 0.4. Both weekly hours worked and new orders consumer goods have a more diffuse lead-time posterior, as we see in Figures 3.6(b) and 3.6(h). On the other hand, VPI has a more peaked lead-time posterior compared to its revised-data counterpart, and it also gains additional weight. The posterior mean of β_{VPI} increases from 0.25 to 0.43.

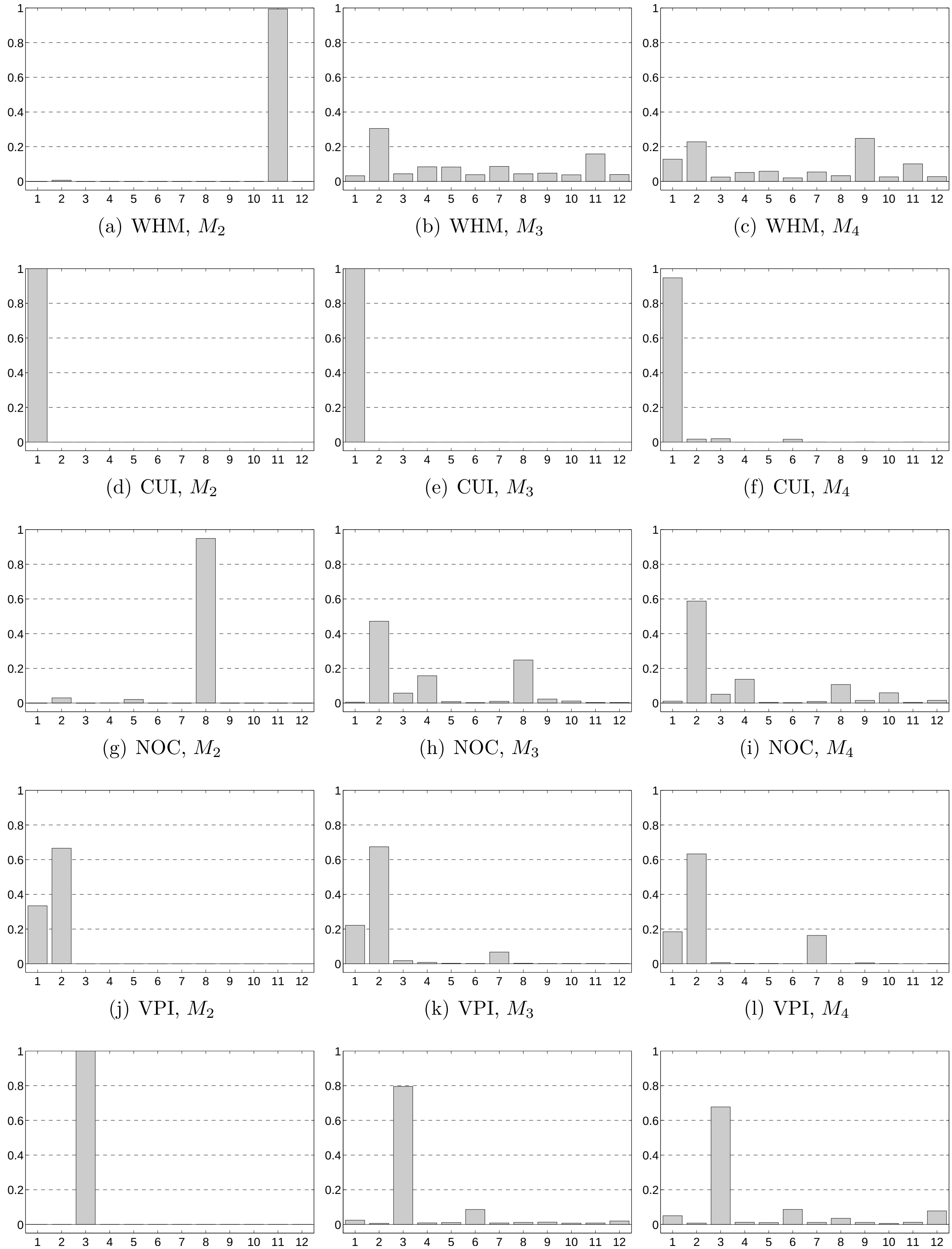
For the data measured in real time, we make the following classification in terms of relative importance by looking at the marginal posterior density plots of the indicator weights in the left columns of Figures 3.8–3.9. From less important to most important we start again with WHM. Next, a sizeable group consisting of NOC, VPI, NOK, NBP, M2 and CCE take substantial weights all around 0.4. The pair CUI and SP5 are more important, each having a weight of 0.7. Finally we see again that 10TFF, which is not liable to any revisions, dominates the other leading indicators with $\beta_{\text{10TFF}} \approx 1.2$.

We recall that we use the bivariate dependent variable with both the recession indicator and the growth rate of the coincident index to gather more information about the business cycle and hence extract a stronger leading signal. Since CEI is a continuous variable it contains more information due to its greater variability, and therefore implicitly has a larger weight in identifying the relative importance of the J leading variables. We examine as follows whether some leading indicators are more important in leading the coincident index than leading the recession indicator. First, we define model M_4 as the unrestricted bivariate model, *i.e.*, we do not impose any restrictions on the $2J$ indicator weights $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in (3.1) and (3.2). Second, we check the relative importance of the ten indicators across the two equations by comparing the posterior distributions of the two vectors with regression weights in M_4.[8]

First, the posteriors of the lead times are very close to the ones we obtain in M_3, as we see in the right columns of Figures 3.2–3.3 (revised data) and 3.6–3.7 (real-time data). Figures 3.4–3.5 (revised data) and 3.8–3.9 (real-time data) show in their middle and right columns the marginal posteriors of the elements of $\boldsymbol{\beta}$ and $\hat{\gamma}^{-1}\boldsymbol{\alpha}$, respectively (with $\hat{\gamma} = \mathsf{E}[\gamma \,|\, \mathbf{Y}, M_3] = 0.57$). If $\hat{\alpha}_j/\hat{\beta}_j \approx \hat{\gamma}$, the equal-relative-importance restriction across the two equations is likely to hold (with $\hat{\vartheta}_j = \mathsf{E}[\vartheta_j \,|\, \mathbf{Y}, M_4]$, $\vartheta \in \{\alpha, \beta\}$). With latest-available data the two smallest ratios we find belong to VPI and CCE and take values slightly above 0.2. The weights of these variables in the two equations show substantially different posteriors indeed, as we see in Figures 3.4(k) and 3.4(l) for VPI, and in Figures 3.5(n) and 3.5(o) for CCE. 10TFF has the largest ratio of 0.64. The other seven are all in the range 0.35–0.60. For real-time data we obtain very similar results. In summary, the effect of vendor performance and consumer expectations is larger in leading the coincident index than in leading the

[8]We parameterize this bivariate model such that the intercept of the latent regression equation is fixed at 1 and the variance matrix is unrestricted. We refer to Appendix 3.A.3 for further details about this model specification and issues related to posterior simulation.

Figure 3.6 *Marginal posteriors of lead times, real-time data*



(a) WHM, M_2 (b) WHM, M_3 (c) WHM, M_4

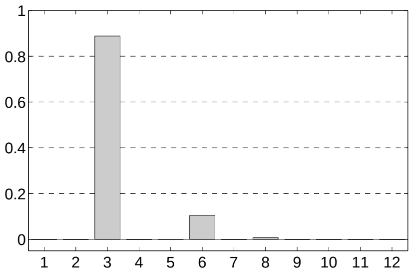
(d) CUI, M_2 (e) CUI, M_3 (f) CUI, M_4

(g) NOC, M_2 (h) NOC, M_3 (i) NOC, M_4

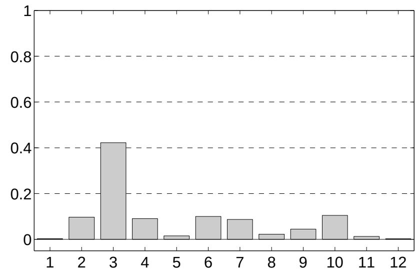
(j) VPI, M_2 (k) VPI, M_3 (l) VPI, M_4

(m) NOK, M_2 (n) NOK, M_3 (o) NOK, M_4

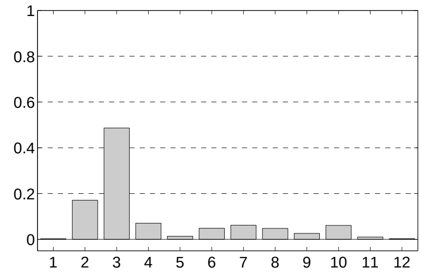
Figure 3.7 *Marginal posteriors of lead times, real-time data,* Cont.



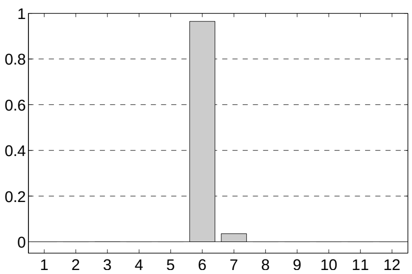
(a) NBP, M_2



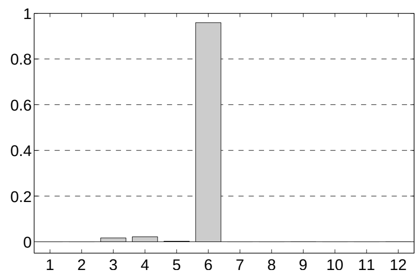
(b) NBP, M_3



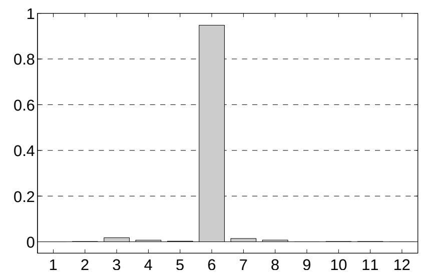
(c) NBP, M_4



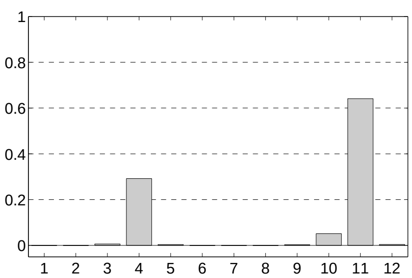
(d) SP5, M_2



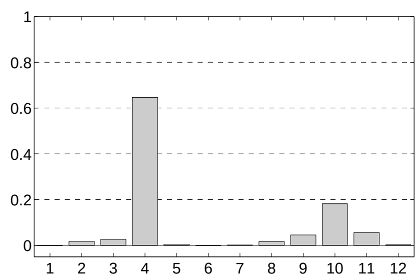
(e) SP5, M_3



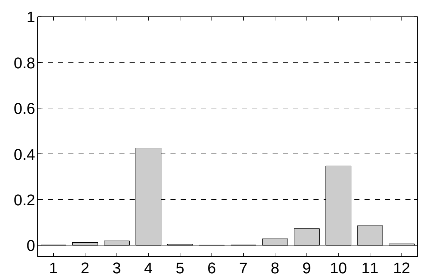
(f) SP5, M_4



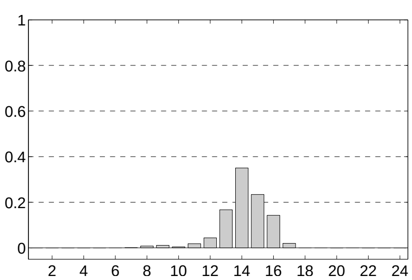
(g) M2, M_2



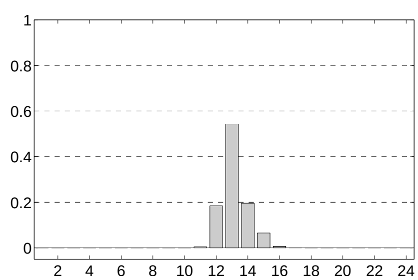
(h) M2, M_3



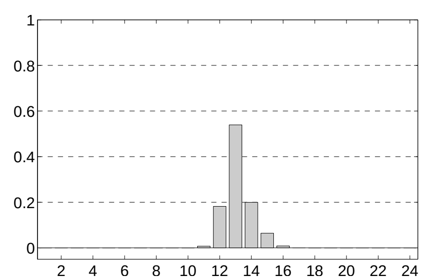
(i) M2, M_4



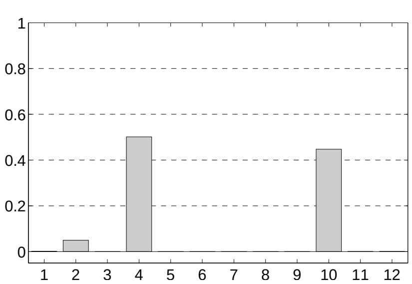
(j) 10TFF, M_2



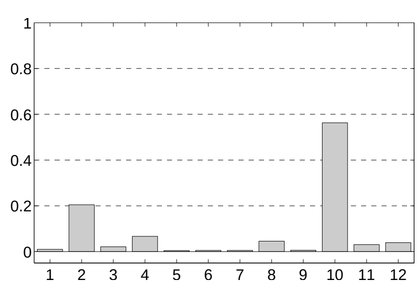
(k) 10TFF, M_3



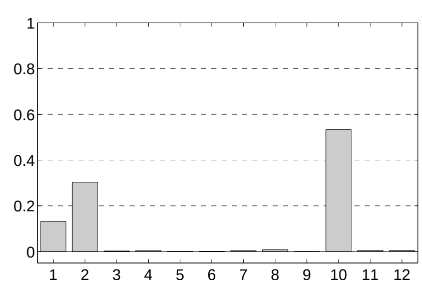
(l) 10TFF, M_4



(m) CCE, M_2

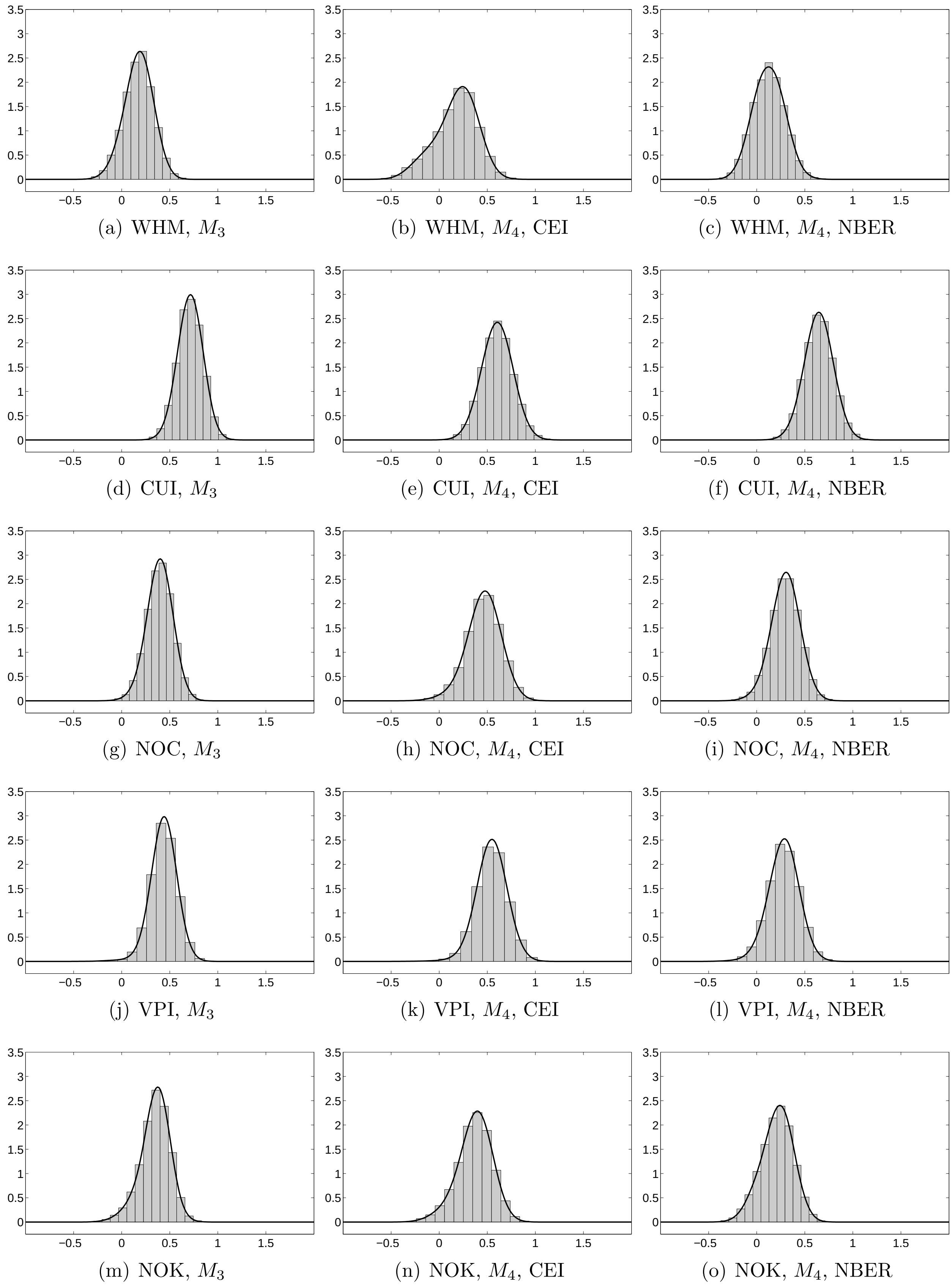


(n) CCE, M_3



(o) CCE, M_4

Figure 3.8 *Marginal posteriors of leading-indicator weights, real-time data*



(a) WHM, M_3 (b) WHM, M_4, CEI (c) WHM, M_4, NBER

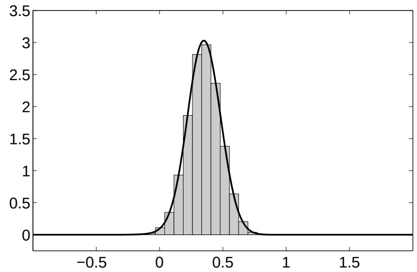
(d) CUI, M_3 (e) CUI, M_4, CEI (f) CUI, M_4, NBER

(g) NOC, M_3 (h) NOC, M_4, CEI (i) NOC, M_4, NBER

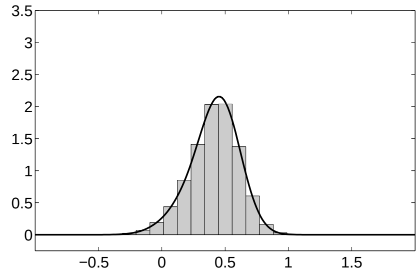
(j) VPI, M_3 (k) VPI, M_4, CEI (l) VPI, M_4, NBER

(m) NOK, M_3 (n) NOK, M_4, CEI (o) NOK, M_4, NBER

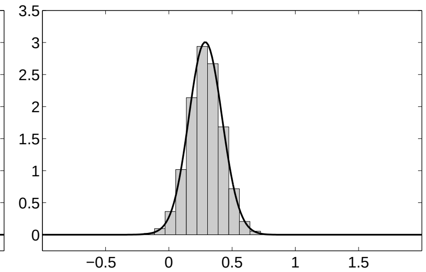
Figure 3.9 *Marginal posteriors of leading-indicator weights, real-time data,* Cont.



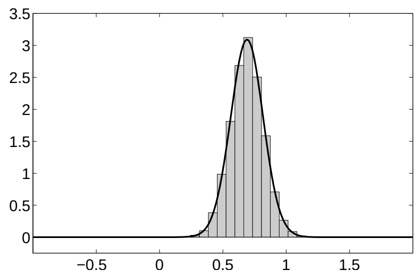
(a) NBP, M_3



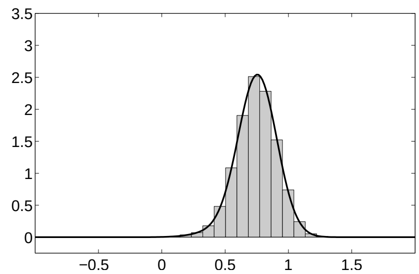
(b) NBP, M_4, CEI



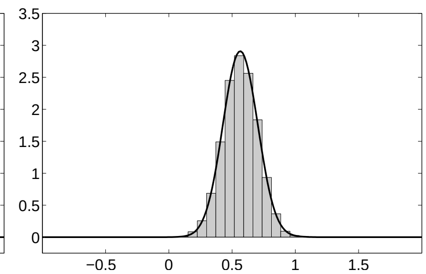
(c) NBP, M_4, NBER



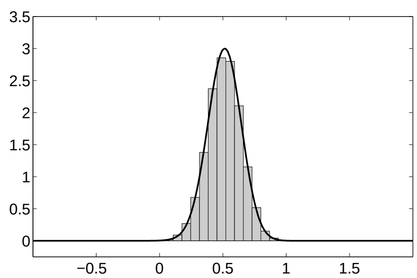
(d) SP5, M_3



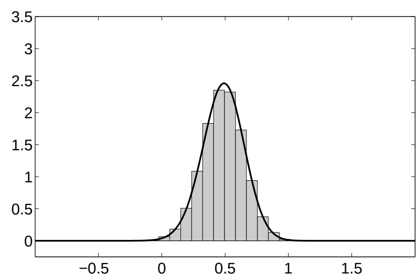
(e) SP5, M_4, CEI



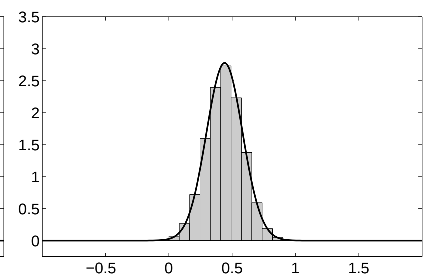
(f) SP5, M_4, NBER



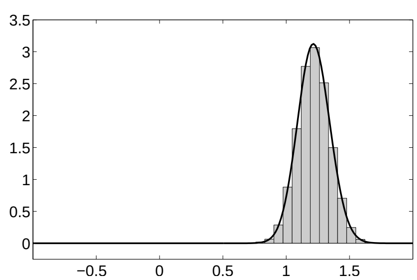
(g) M2, M_3



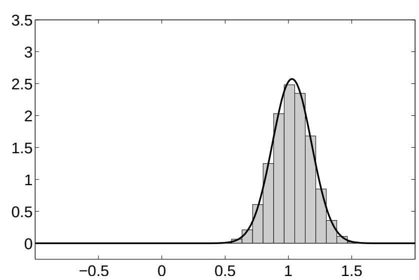
(h) M2, M_4, CEI



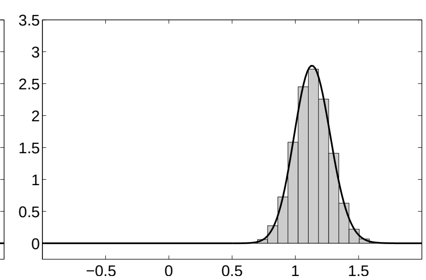
(i) M2, M_4, NBER



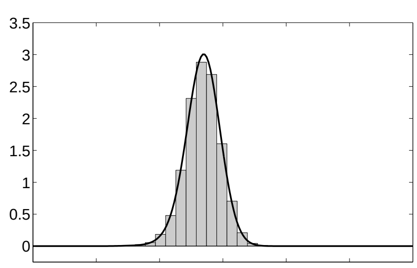
(j) 10TFF, M_3



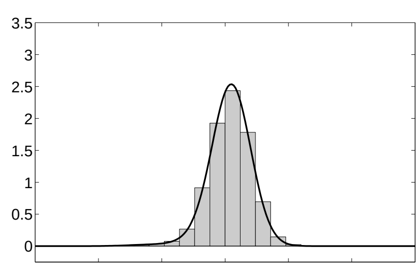
(k) 10TFF, M_4, CEI



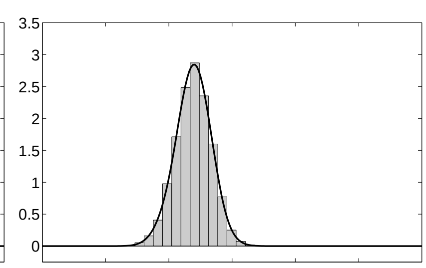
(l) 10TFF, M_4, NBER



(m) CCE, M_3

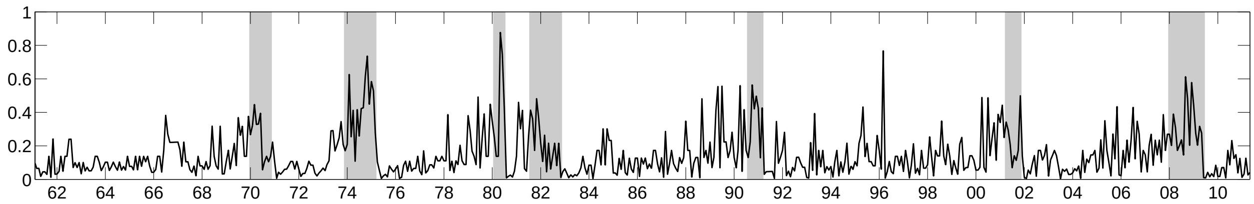


(n) CCE, M_4, CEI

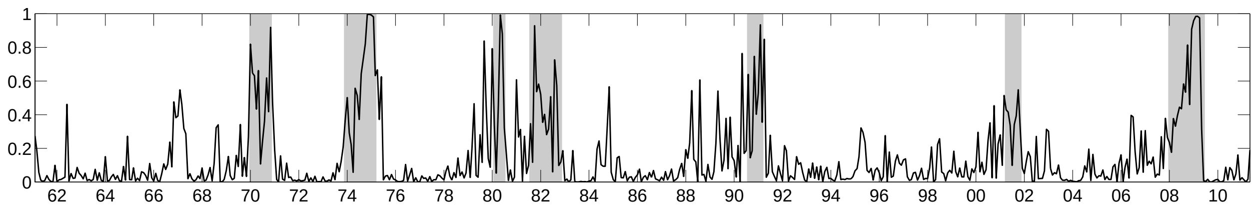


(o) CCE, M_4, NBER

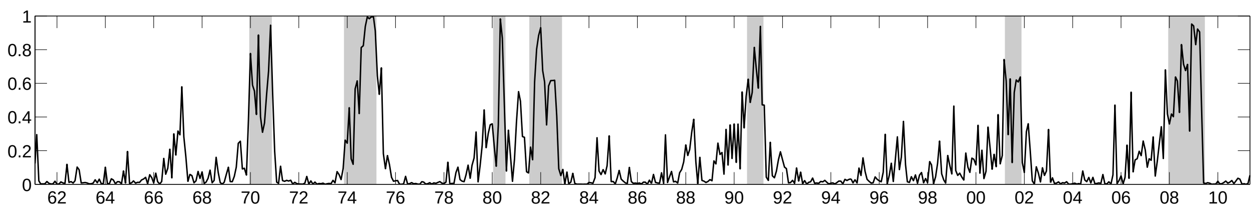
Figure 3.10 *In-sample one-month-ahead recession probabilities, revised data*



(a) TCB's LEI



(b) M_2



(c) M_3

NBER indicator. Second, the interest rate spread seems slightly more important in leading the zero-one state of the economy. Despite the small differences, these results sustain our assumption of equal relative importance of the indicators across the two equations in M_3. Moreover, in the next section we further see that relaxing this assumption, leading to model M_4, does not pay off in terms of improved forecasting of recessions.

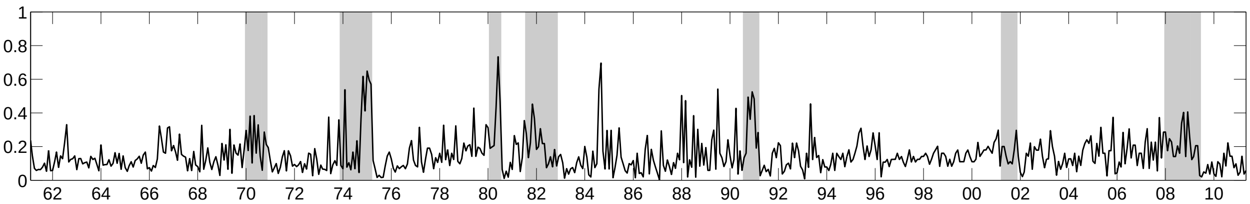
3.4.3 Forecasting

We start with an inspection of the in-sample forecasting performance of our leading indexes. We compute in-sample recession probabilities by mixing the marginalized (with respect to CEI) probability of a recession $\Phi(-\gamma\mu^*_{2t})$ over the full-sample posterior distribution $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$, such that we average out uncertainty about lead times and indicator weight(s).[9]

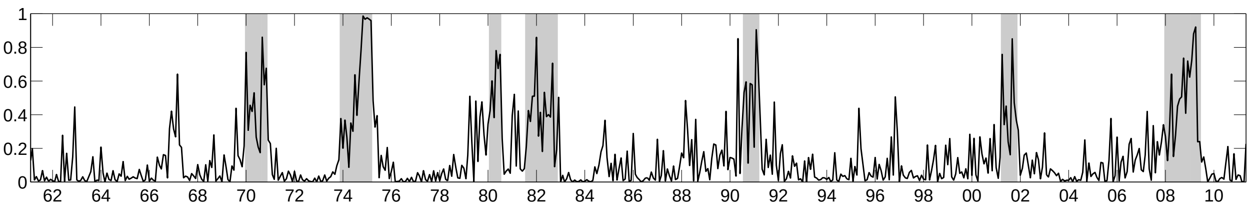
In Figure 3.10 we depict the in-sample recession probabilities for the model including solely TCB's LEI, and our alternative leading-index models, all obtained with revised data. If we compare Figure 3.10(a) to 3.10(b), we observe that M_2 provides a substantially better fit of the recessionary months which we indicate by the shaded areas. The in-sample forecasts of

[9] Additionally, we also have the *conditional* in-sample probabilities $\mathsf{Pr}\,[y_{2t} = 1 \,|\, y_{1t}]$, given CEI's growth rate. As the first release of month t's CEI is issued months before the NBER announces its recession decision, with our model we generate an updated predictive probability for month t. These updated probabilities are more pronounced since we find a correlation of $\rho \approx 0.7$ between the innovations in the two equations.

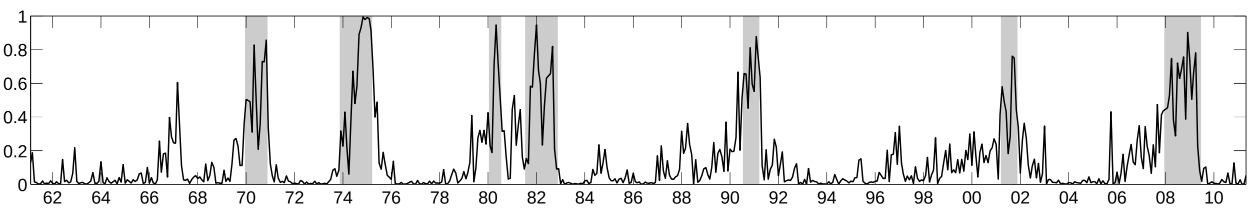
Figure 3.11 *In-sample one-month-ahead recession probabilities, real-time data*



(a) TCB's LEI



(b) M_2



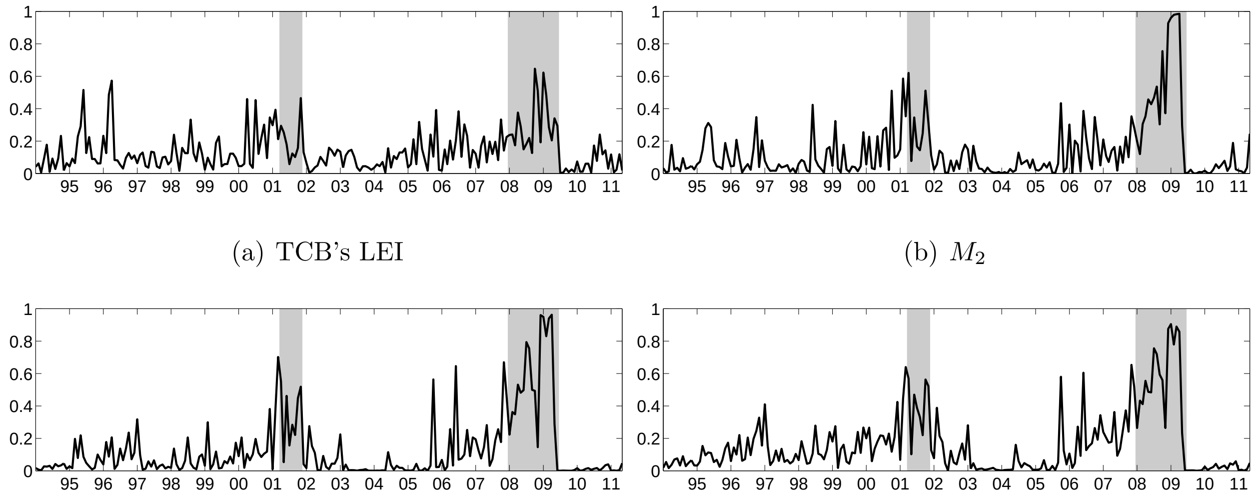
(c) M_3

M_3 in Figure 3.10(c) are even more pronounced compared to M_2's, especially during the first five recessionary periods. This finding visually confirms our testing results in Section 3.4.1. With all models the 2001 recession turns out to be the most difficult to accurately predict. Using the unrestricted-weights model M_4 does not change this either, since its recession probabilities are almost equal to M_3's (and therefore not reported). The probability forecasts for the 2001 slow-down period obtained with our indexes do behave relatively volatile and are steadily increasing about a year in advance of the recession.

We plot the in-sample probabilities with real-time data in Figure 3.11. Allowing each of the ten leading variables to have its own lead time results in an even stronger improvement over LEI's predictive power compared to the revised-data setting. The real-time variant of TCB's leading indicator does a reasonable job when predicting the 1974, 1980 and 1990 recessions, but otherwise its performance is modest. Both our new real-time indexes better anticipate recessionary periods. Comparing Figures 3.10 and 3.11 we find only small differences between applying latest-available data or leading indicators as they were historically available for one-month-ahead forecasting, at least visually.

Since we do not explicitly model persistence in the dependent variable in addition to the temporal dependence induced by the autodependent leading indicators, our *smoothed* recession probabilities are not as clear-cut as Markovian regime-switching models' (Hamilton, 1989, 1994, Ch. 22). However, the *forecast* probabilities of these regime-switching models

Figure 3.12 *Out-of-sample one-month-ahead recession probabilities, revised data*



(a) TCB's LEI

(b) M_2

(c) M_3

(d) M_4

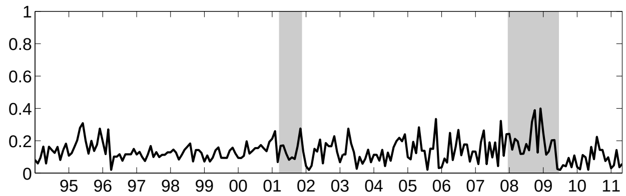
Table 3.4 *Marginalized predictive likelihoods and predictive Bayes factors*

	Forecasting horizon								
	$h = 1$			$h = 3$			$h = 6$		
Model	P.L.	$BF_{i,1}$	$BF_{3,2}$	P.L.	$BF_{i,1}$	$BF_{3,2}$	P.L.	$BF_{i,1}$	$BF_{3,2}$
Panel A: Revised									
TCB's LEI (M_1)	−27.8	–	–	−27.7	–	–	−26.5	–	–
M_2	−22.7	5.0	–	−22.1	5.6	–	−25.1	1.4	–
M_3	−21.1	6.7	1.6	−23.8	3.8	−1.7	−24.4	2.1	0.7
Panel B: Real-time									
TCB's LEI (M_1)	−32.2	–	–	−32.4	–	–	−31.6	–	–
M_2	−22.0	10.2	–	−22.9	9.5	–	−25.5	6.1	–
M_3	−24.3	7.9	−2.3	−26.5	6.0	−3.5	−24.3	7.3	1.2

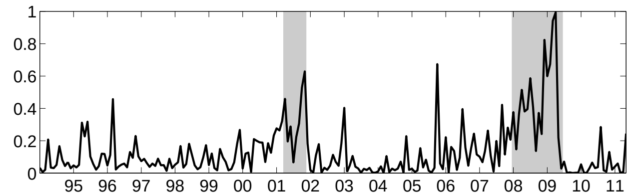
Notes: "TCB's LEI" is the bivariate model with LEI as single explanatory variable, but with unknown lead time; M_2 is the bivariate model with heterogenous lead times, and each leading variable has the same weight; M_3 is the bivariate model with both heterogenous weights and lead times. For all three forecasting horizons, the first forecast we always construct at the end of December 1993, *i.e.*, for $h = 1$ the first forecast concerns January 1994; $h = 3$, March 1994; $h = 6$, June 1994. Predictive likelihoods ("P.L.") are marginalized with respect to CEI's growth rate: $p(\mathbf{Y}_2^{\tau+1,T} \mid \mathbf{Y}^{1,\tau}) = \int p(\mathbf{Y}^{\tau+1,T} \mid \mathbf{Y}^{1,\tau})\, \mathsf{d}\mathbf{Y}_1^{\tau+1,T}$ and, as the Bayes factors, on the $\log_{10}$-scale. Panel A shows results for revised and Panel B for real-time leading-indicator data.

have difficulty in signalling a change in regime because the prevailing state is most likely to persist with estimated probability almost 1. On the contrary, we find out-of-sample probabilities close to their smoothed in-sample counterparts. If we implement the recession forecasting procedure with fully revised data, we obtain the one-month-ahead probability forecasts in Figure 3.12 for the period January 1994–May 2011.

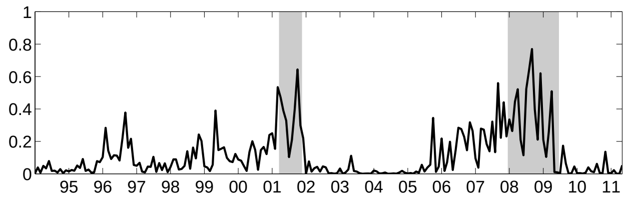
Figure 3.13 *Out-of-sample one-month-ahead recession probabilities, real-time data*



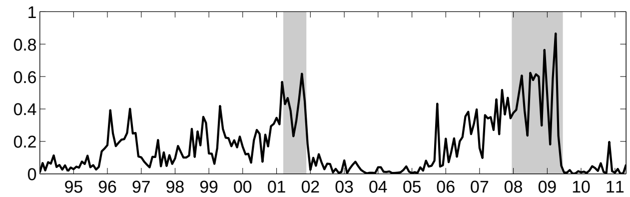
(a) TCB's LEI



(b) M_2



(c) M_3



(d) M_4

We report marginalized predictive likelihoods and Bayes factors in Panel A of Table 3.4 for three forecasting horizons, using latest-available data. For all three horizons both our alternative leading indexes outperform The Conference Board's, with Bayes factors decisively in favor of M_2 and M_3 for $h = 1, 3$, and very strong to decisive support for the six-month forecasting horizon. For one month and six months ahead the indicator-specific-weights index produces strongly to very strongly improved forecasts compared to our index with leading indicators equally weighed.

From a practical point of view the real-time probability forecasts are the most useful. We obtain these forecasts with our leading indexes as they were historically available and the posterior beliefs about the model parameters up to and including the month in which we produce the forecast. In Figure 3.13 we plot $\mathsf{Pr}\,[y_{2t} = 1 \,|\, \mathbf{Y}^{1,t-1}]$, $(t = \tau + 1, \ldots, T)$, as they were actually computable at the end of months December 1993–April 2011.

In Figure 3.13(a) we see very modest real-time forecasting performance of TCB's composite leading indicator. The probabilities just calmly fluctuate around 0.1, with a small increase for the 2008 recession. We remark that the last two recessions are generally harder to predict, as LEI does a better job in-sample for the other recessionary periods. Our leading index with equal weights for the ten indicators generates real-time forecasts depicted in Figure 3.13(b), which obviously bring forth a substantial improvement over LEI's. If we compare the respective marginal predictive likelihoods for $h = 1$ in Panel B of Table 3.4 we obtain a predictive Bayes factor of 10, decisively favoring M_2's leading index. Allowing for indicator-specific weights, we still convincingly outperform the LEI, but $BF_{3,2}$ favors the equal-weights case. This is largely due to the forecasts of M_3 during the most recent recession. Figure 3.13(c) shows that the recession probabilities start decreasing during 2008, whereas for M_2 the probability forecast suddenly drops at the end of the recession. The fact that we predict the 2008 recession more accurately in-sample [Figure 3.11(c)] suggests that this long-lasting recessionary period contains strong information for identifying the parameters of M_3.

Table 3.5 *Cross correlations leading indexes versus business cycle*

Leading index	Business cycle measure h months ahead: 1	2	3	4	5	6	7	8	9
Panel A: Revised data	CEI's growth rate								
LEI	0.27	0.38	0.34	0.35	0.29	0.31	0.26	0.27	0.30
M_3, $h = 1$	0.55	0.47	0.46	0.42	0.37	0.38	0.36	0.31	0.33
M_3, $h = 3$	0.44	0.47	0.51	0.45	0.42	0.44	0.38	0.37	0.38
M_3, $h = 6$	0.38	0.42	0.44	0.44	0.43	0.48	0.41	0.40	0.41
	NBER indicator								
LEI	−0.35	−0.41	−0.43	−0.42	−0.39	−0.40	−0.40	−0.39	−0.35
M_3, $h = 1$	−0.65	−0.64	−0.59	−0.54	−0.49	−0.46	−0.42	−0.37	−0.33
M_3, $h = 3$	−0.59	−0.63	−0.65	−0.62	−0.59	−0.57	−0.54	−0.49	−0.43
M_3, $h = 6$	−0.54	−0.58	−0.59	−0.59	−0.59	−0.61	−0.59	−0.55	−0.51
Panel B: Real-time data	CEI's growth rate								
LEI	0.16	0.29	0.24	0.21	0.15	0.15	0.13	0.12	0.12
M_3, $h = 1$	0.55	0.47	0.46	0.43	0.37	0.41	0.36	0.31	0.32
M_3, $h = 3$	0.42	0.46	0.50	0.44	0.41	0.43	0.38	0.37	0.38
M_3, $h = 6$	0.34	0.39	0.41	0.42	0.41	0.44	0.38	0.39	0.39
	NBER indicator								
LEI	−0.24	−0.29	−0.28	−0.24	−0.21	−0.20	−0.18	−0.16	−0.13
M_3, $h = 1$	−0.65	−0.63	−0.58	−0.53	−0.49	−0.46	−0.42	−0.37	−0.32
M_3, $h = 3$	−0.57	−0.61	−0.63	−0.60	−0.57	−0.55	−0.53	−0.49	−0.44
M_3, $h = 6$	−0.45	−0.51	−0.54	−0.56	−0.57	−0.59	−0.57	−0.53	−0.49

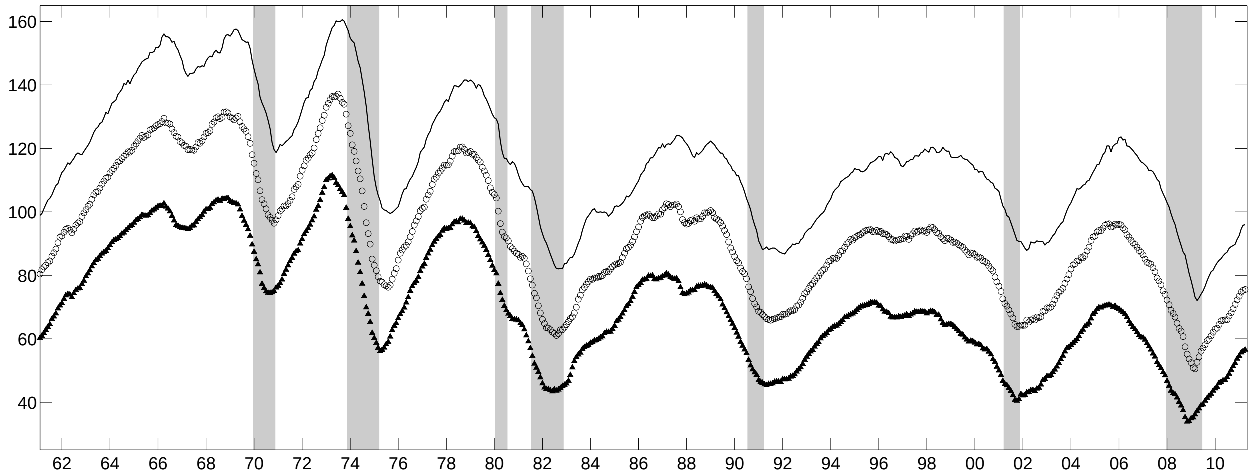
For three months ahead we obtain similar results. Both our alternative indexes produce better recession forecasts than TCB's, but M_3 does not improve over M_2 out-of-sample, with $BF_{3,2} = -3.5$. If we derive an index leading the business cycle half a year, the model with indicator-specific weights does provide a stronger signal as indicated by the predictive Bayes factor of 1.2 for M_3 relative to M_2.

We conclude the forecasting experiment with two further remarks. That our leading indexes more accurately anticipate changes in the state of the economy out-of-sample than LEI is evident from the predictive Bayes factors. Comparing M_2 to M_3 is however more delicate. We see that M_3 does better in-sample, but since it does slightly worse compared to M_2 in forecasting the 2008 recession, the predictive Bayes factors choose the equal-weights leading index for three out of six cases. Second, as seen in-sample, the 2001 recession is difficult to predict regardless of the model used, it is however one of just two recessions in our hold-out sample to assess forecasting performance.

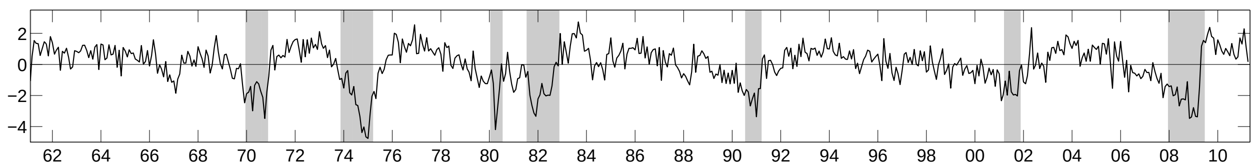
3.4.4 Properties alternative indexes

To analyze the time-series properties of our leading indexes extracted with model M_3, we plot both its growth rates and levels in Figures 3.14 and 3.15 for the final and real-time data, respectively, and for one, three and six months ahead. These time series clearly ex-

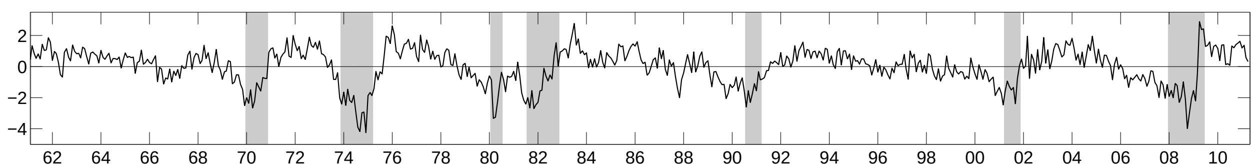
Figure 3.14 *Alternative model-based (M_3) leading economic indexes, revised data*



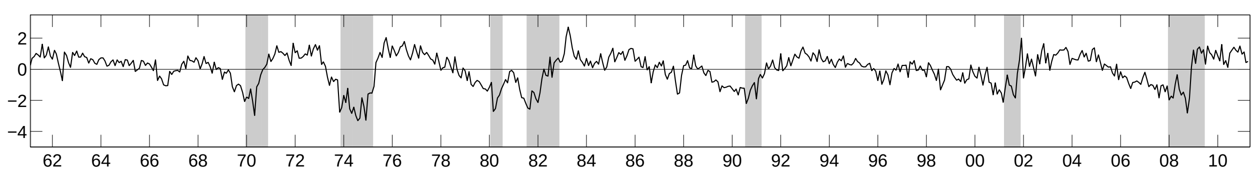
(a) Detrended indexes leading one month (solid line), three months ("○"), six months ("▲")



(b) Demeaned growth rates one-month-ahead index



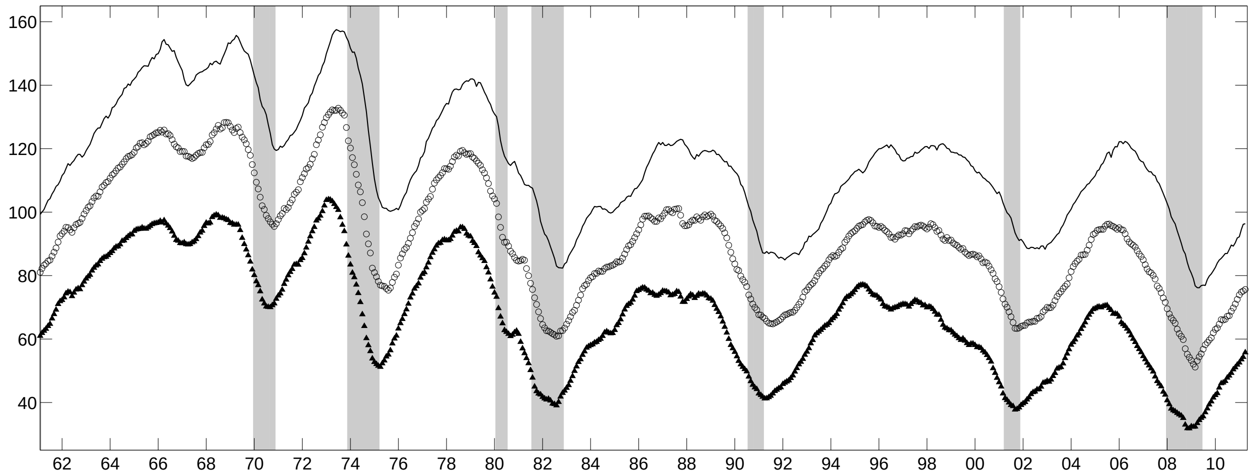
(c) Demeaned growth rates three-months-ahead index



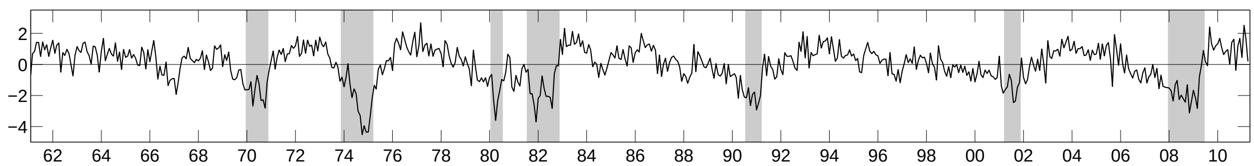
(d) Demeaned growth rates six-months-ahead index

hibit cyclical behavior, with recessions starting on average when the index changes from negatively to positively decreasing (second derivative changes sign). We observe three periods of pronounced decline of the index, but not persistent enough though to lead to the announcement of a recession (1967, tensions in the Middle East; 1987, stock market crash; 1997 Asian financial crisis). From these figures we also learn why the 2001 recession is difficult to forecast. Although the index is decreasing about six months in advance of the start of the recession, the decline is not as severe as for the other recessions.

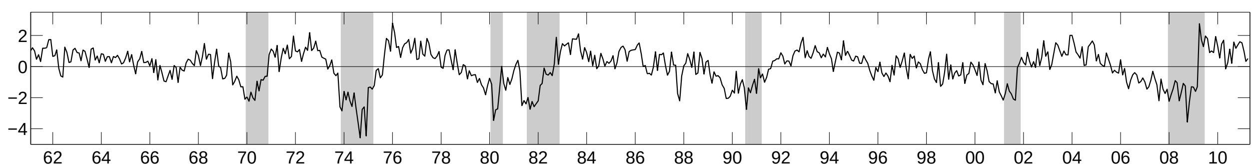
Figure 3.15 *Alternative model-based (M_3) leading economic indexes, real-time data*



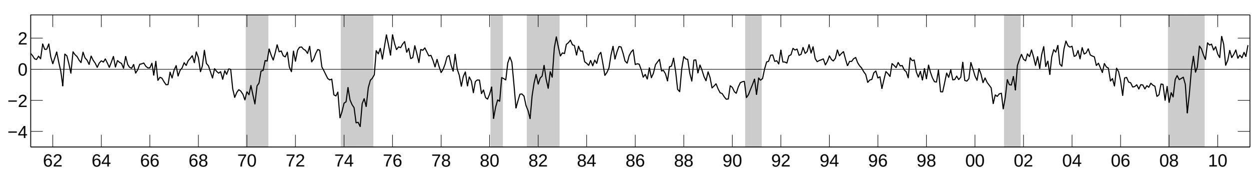
(a) Detrended indexes leading one month (solid line), three months ("○"), six months ("▲")



(b) Demeaned growth rates one-month-ahead index



(c) Demeaned growth rates three-months-ahead index



(d) Demeaned growth rates six-months-ahead index

Our methods allow us to compose indexes leading the business cycle for a given number of months. From the figures we derive that the six-months leading index signals changes in business conditions earlier than the one-month index. This feature is most notable when the economy comes out of a recession. During the 1974 recession for example, we see the six-months index reaching its through first. The sample correlations reported in Table 3.5 establish this numerically. They show that for a given h we automatically find the composition of leading variables with the highest correlation with the business cycle. Both in terms of leading CEI and the recession indicator, our h-months leading index realizes the highest correlation with these two business cycle measures h months ahead. TCB's LEI is most useful for forecasting two/three months ahead. Nonetheless, our methods offer a better alternative as they are especially designed for user-specified forecasting horizons, and they achieve substantially higher correlations with the business cycle. Naturally, the longer the horizon, the more difficult it becomes to foresee changes in the state of the economy as shown by the decreasing maximum correlations for h increasing. Final data provide stronger leading signals, as for all h their indexes attain higher correlations. Finally, the table corroborates our previous finding that especially LEI measured in real time can be improved upon substantially.

3.5 Conclusion

The construction of the commonly used and intently watched composite leading index as monthly released by The Conference Board involves two rather stringent assumptions. Its ten constituents are assumed to have identical lead times and are assigned equal weights. Obviously, individual leading indicators may differ in the number of months they lead the business cycle. Moreover, some leading indicators may maintain a stronger relation with the business cycle than others. We find empirical evidence supporting these two relaxations of TCB's methods and we propose an improved leading index that better signals changes in business conditions. Especially the enhancements of real-time forecasting are useful since first releases of TCB's leading index are known for their modest predictive power (Diebold and Rudebusch, 1991a).

We use a bivariate model to combine both quantitative (TCB's composite coincident index) and qualitative (NBER's turning points) measures of the state of the economy and relate these to the ten individual leading indicators. Each leading variable has its own lead time and relative weight. We deal with the resultant discrete-continuous parameter space in a Bayesian set-up. This approach also provides the natural context to construct real-time predictions for changes in the business cycle. That is, we employ real-time data (in addition to final data) and each month we recursively update the model parameters' posterior distribution.

We find decisive support for heterogeneity in the lead times and strong empirical evidence for differences in relative importance of the individual leading indicators. Our resulting alternative composite indexes outperform TCB's both in-sample and out-of-sample, providing more accurate recession-probability forecasts. Especially the improvement in real-time forecasting is important as it offers a framework that can actually be operationalized in practice.

Further research should reveal whether lead times and relative weights of the individual leading indicators are dependent on the state of the business cycle as well. Previous research finds that TCB's composite leading index has a longer lead time for the business cycle when the latter is in expansion than when it is in contraction (Paap *et al.*, 2009).

3.A Technical details

In this appendix we provide the technical details of the Bayesian analysis of our models. First, we introduce some additional notation to facilitate the exposition of the methods, next we discuss the derivation of the posterior distribution and the simulation routines to obtain a sample thereof.

We sample T times the bivariate observation $\mathbf{y}_t = (y_{1t}, y_{2t})'$, $(t = 1, \ldots, T)$, (continuous CEI and binary recession indicator). To model the contemporaneous dependence between these two variables we define the unobserved continuous random variables z_{2t}^*, $(t = 1, \ldots, T)$, that determine the value of the recession indicator. We use the reparameterized model specification in which the variance of these latent variables is identified due to the additional restriction imposed on the indicator weights (see Section 3.2.1). We define the vector $\mathbf{y}_t^* = (y_{1t}, z_{2t}^*)'$, which contains the two continuous variables of which the second component is not observed as opposed to the first. Realizations of the leading variables are arranged in the $T \times J$ lead-time-dependent matrix

$$\mathbf{X}(\boldsymbol{\kappa}) = \begin{bmatrix} \mathbf{x}_1^{1-\kappa_1, T-\kappa_1} & \mathbf{x}_2^{1-\kappa_2, T-\kappa_2} & \cdots & \mathbf{x}_J^{1-\kappa_J, T-\kappa_J} \end{bmatrix},$$

with vectors $\mathbf{x}_j^{1-\kappa_j, T-\kappa_j} = (x_{j,1-\kappa_j}, \ldots, x_{j,T-\kappa_j})'$, containing leading indicator j's lagged observations. Finally, we define the following three $T \times 2$ matrices

$$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)', \quad \mathbf{Y}^* = (\mathbf{y}_1^*, \ldots, \mathbf{y}_T^*)', \quad \boldsymbol{\varepsilon}^* = (\boldsymbol{\varepsilon}_1^*, \ldots, \boldsymbol{\varepsilon}_T^*)'. \tag{3.A.1}$$

We write the model for the continuous dependent variables in matrix notation as

$$\mathbf{Y}^* = \boldsymbol{\iota}_T \boldsymbol{\Pi}' + \mathbf{X}(\boldsymbol{\kappa})(\boldsymbol{\iota}_2' \otimes \boldsymbol{\beta}) + \boldsymbol{\varepsilon}^*, \quad \boldsymbol{\varepsilon}^* \sim \mathcal{MN}(\mathbf{0}_{T\times 2}, \boldsymbol{\Sigma} \otimes \mathbf{I}_T),$$

in which the two intercepts are stacked in $\boldsymbol{\Pi} = (\beta_0, \alpha_0^*)'$. Thus, we obtain a bivariate SUR model for $\mathbf{y}_t^*$, in which each equation has the same regressors and the regressors have identical partial effects across the two equations. The probit equations $y_{2t} = \mathbb{I}_{\{z_{2t}^* < 0\}}$, $(t = 1, \ldots, T)$, complete the model specification.

3.A.1 Prior distribution and complete data likelihood

We discuss the more general case (M_3) in which each leading variable has its own weight β_j. The model with equal weights $\beta_j = \bar{\beta}$ (M_2) follows as a special case. The prior specification from Section 3.2.2 leads to the prior density function

$$\begin{aligned} p(\boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2q}(\boldsymbol{\Pi} - \mathbf{p}')'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Pi} - \mathbf{p}')\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{b})'\mathbf{B}^{-1}(\boldsymbol{\beta} - \mathbf{b})\right\} \\ \times |\boldsymbol{\Sigma}|^{-(\nu+N+1)/2} \exp\left\{-\frac{1}{2}\mathsf{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{S}\right]\right\} \times \prod_{j=1}^{J} \sum_{k=1}^{K_j^{\max}} \mathbb{I}_{\{\kappa_j = k\}}, \end{aligned} \tag{3.A.2}$$

with $N = 2$ the dimension of $\mathbf{y}_t$.

To compose the complete data likelihood we first vectorize the SUR model in (3.A.1) such that we obtain the normal linear regression model

$$\mathbf{w} = \mathsf{vec}\left[\mathbf{Y}^* - \boldsymbol{\iota}_T\boldsymbol{\Pi}'\right] = \left[\boldsymbol{\iota}_2 \otimes \mathbf{X}(\boldsymbol{\kappa})\right]\boldsymbol{\beta} + \mathsf{vec}\left[\boldsymbol{\varepsilon}^*\right], \quad \mathsf{vec}\left[\boldsymbol{\varepsilon}^*\right] \sim \mathcal{N}(\mathbf{0}_{2T}, \boldsymbol{\Sigma} \otimes \mathbf{I}_T), \tag{3.A.3}$$

in which we denote the lead-time-dependent design matrix as $\mathbf{V}(\boldsymbol{\kappa}) = \boldsymbol{\iota}_2 \otimes \mathbf{X}(\boldsymbol{\kappa})$. With the latent data stacked in $\mathbf{z}_2^* = (z_{21}^*, \ldots, z_{2T}^*)'$, we obtain the complete data likelihood funtion

$$\begin{aligned} p(\mathbf{Y}, \mathbf{z}_2^* \,|\, \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}|^{-T/2} \exp\left\{-\frac{1}{2}[\mathbf{w} - \mathbf{V}(\boldsymbol{\kappa})\boldsymbol{\beta}]'(\boldsymbol{\Sigma} \otimes \mathbf{I}_T)^{-1}[\mathbf{w} - \mathbf{V}(\boldsymbol{\kappa})\boldsymbol{\beta}]\right\} \\ \times \prod_{t=1}^{T} \left(\mathbb{I}_{\{y_{2t}=1\}}\mathbb{I}_{\{z_{2t}^* < 0\}} + \mathbb{I}_{\{y_{2t}=0\}}\mathbb{I}_{\{z_{2t}^* \geq 0\}}\right). \end{aligned} \tag{3.A.4}$$

3.A.2 Posterior simulation steps

If we combine (3.A.2) with (3.A.4), we obtain the kernel of the augmented posterior distribution $p(\boldsymbol{\theta}, \mathbf{z}_2^* \,|\, \mathbf{Y})$, from which we sample by implementing an efficient MCMC scheme. We expect the dependence between the lead times and the regression weights of the individual leading indicators to be strong and therefore to slow down the mixing of a full Gibbs sampler considerably. To resolve this issue we marginalize with respect to the regression weights and sample the lead times independent from $\boldsymbol{\beta}$. The four steps in Section 3.2.2 are implemented as follows.

Step 1: Latent data

Sample each latent variable from $p(z_{2t}^* \,|\, \mathbf{Y}, \boldsymbol{\theta})$, $(t = 1, \ldots, T)$. Conditional on y_{1t} and $\boldsymbol{\theta}$, z_{2t}^* is independent from all other y_{1i} and z_{2i}^*. It depends on the contemporaneous CEI growth rate though. If we use the properties of the multivariate normal distribution, we sample each z_{2t}^* from a univariate truncated conditional normal distribution with moments as in (3.5)–(3.6). Thus, we update $\mathbf{z}_2^*$ by sampling each of its T components as

$$z_{2t}^* \,|\, \{\mathbf{Y}, \boldsymbol{\theta}\} \sim \mathcal{N}(\mu_{2|1,t}, \sigma_{2|1}^2) \times (\mathbb{I}_{\{y_{2t}=1\}}\mathbb{I}_{\{z_{2t}^*<0\}} + \mathbb{I}_{\{y_{2t}=0\}}\mathbb{I}_{\{z_{2t}^*\geq 0\}}).$$

Step 2: Lead times

Sample each lead time from $p(\kappa_j \,|\, \mathbf{Y}, \mathbf{z}_2^*, \boldsymbol{\Pi}, \boldsymbol{\Sigma}, \boldsymbol{\kappa}_{-j}) \propto p(\boldsymbol{\kappa} \,|\, \mathbf{Y}, \mathbf{z}_2^*, \boldsymbol{\Pi}, \boldsymbol{\Sigma})$, $(j = 1, \ldots, J)$. Due to the conditionally conjugate prior for the indicator weights, we analytically marginalize with respect to $\boldsymbol{\beta}$. First, we collect the terms involving $\boldsymbol{\beta}$ from (3.A.4) and (3.A.2), then we apply the decomposition rule (see Greenberg, 2008) and integrate the kernel of the normal distribution with mean and variance

$$\overline{\mathbf{b}}(\boldsymbol{\kappa}) = \overline{\mathbf{B}}(\boldsymbol{\kappa})[\mathbf{V}(\boldsymbol{\kappa})'(\boldsymbol{\Sigma} \otimes \mathbf{I}_T)^{-1}\mathbf{w} + \mathbf{B}^{-1}\mathbf{b}], \tag{3.A.5}$$

$$\overline{\mathbf{B}}(\boldsymbol{\kappa}) = [\mathbf{V}(\boldsymbol{\kappa})'(\boldsymbol{\Sigma} \otimes \mathbf{I}_T)^{-1}\mathbf{V}(\boldsymbol{\kappa}) + \mathbf{B}^{-1}]^{-1}. \tag{3.A.6}$$

Keeping all terms involving $\boldsymbol{\kappa}$, we obtain the marginalized density

$$\begin{aligned} p(\boldsymbol{\kappa} \,|\, \mathbf{Y}, \mathbf{z}_2^*, \boldsymbol{\Pi}, \boldsymbol{\Sigma}) &\propto \int p(\boldsymbol{\theta}, \mathbf{z}_2^* \,|\, \mathbf{Y}) \, \mathrm{d}\boldsymbol{\beta}, \\ &\propto \left|\overline{\mathbf{B}}(\boldsymbol{\kappa})\right|^{1/2} \exp\{-\mathbf{r}(\boldsymbol{\kappa})'\mathbf{r}(\boldsymbol{\kappa})/2\} \textstyle\prod_{j=1}^{J} \sum_{k=1}^{K_j^{\max}} \mathbb{I}_{\{\kappa_j=k\}}, \end{aligned} \tag{3.A.7}$$

in which

$$\begin{aligned} \mathbf{r}(\boldsymbol{\kappa})'\mathbf{r}(\boldsymbol{\kappa}) = {} & [\mathbf{w} - \mathbf{V}(\boldsymbol{\kappa})\overline{\mathbf{b}}(\boldsymbol{\kappa})]'(\boldsymbol{\Sigma} \otimes \mathbf{I}_T)^{-1}[\mathbf{w} - \mathbf{V}(\boldsymbol{\kappa})\overline{\mathbf{b}}(\boldsymbol{\kappa})] \\ & + [\overline{\mathbf{b}}(\boldsymbol{\kappa}) - \mathbf{b}]'\mathbf{B}^{-1}[\overline{\mathbf{b}}(\boldsymbol{\kappa}) - \mathbf{b}]. \end{aligned}$$

We evaluate the kernel in (3.A.7) for all supported values of κ_j to obtain

$$\bar{p}_{\kappa_j,k} = p(\kappa_j = k, \boldsymbol{\kappa}_{-j} \,|\, \mathbf{Y}, \mathbf{z}_2^*, \boldsymbol{\Pi}, \boldsymbol{\Sigma}), \quad (k = 1, \ldots, K_j^{\max}).$$

After normalization we draw from the discrete distribution with probabilities $\bar{p}_{\kappa_j,k}/\sum_i \bar{p}_{\kappa_j,i}$, $(k = 1, \ldots, K_j^{\max})$. For the equal-weights restricted model we substitute $\{\mathbf{V}(\boldsymbol{\kappa})\boldsymbol{\iota}_J, b, B\}$ for $\{\mathbf{V}(\boldsymbol{\kappa}), \mathbf{b}, \mathbf{B}\}$ in the above derivations.

Step 3: Indicator weights

Sample indicator weights in one block from $p(\boldsymbol{\beta}\,|\,\mathbf{Y},\mathbf{z}_2^*,\boldsymbol{\theta}_{-\boldsymbol{\beta}})$. If we use the auxiliary regression model with $2T$ observations in (3.A.3) and the Gaussian prior, we sample

$$\boldsymbol{\beta}\,|\,\{\mathbf{Y},\mathbf{z}_2^*,\boldsymbol{\theta}_{-\boldsymbol{\beta}}\} \sim \mathcal{N}\left(\overline{\mathbf{b}}(\boldsymbol{\kappa}),\overline{\mathbf{B}}(\boldsymbol{\kappa})\right),$$

with the mean and variance matrix as in (3.A.5)–(3.A.6). To sample $\bar{\beta}$ in M_2, we replace again the explanatory variables matrix and the prior mean and variance with their counterparts as in Step 2.

Step 4: Intercepts and covariance matrix

Sample intercepts and covariance matrix in one block from $p(\boldsymbol{\Sigma},\boldsymbol{\Pi}\,|\,\mathbf{Y},\mathbf{z}_2^*,\boldsymbol{\beta},\boldsymbol{\kappa})$. We obtain a conditional multivariate normal linear regression model and combined with the natural conjugate prior choice for $\{\boldsymbol{\Sigma},\boldsymbol{\Pi}\}$ we simulate $\boldsymbol{\Sigma}$ from the inverted Wishart distribution

$$\begin{aligned}
&\boldsymbol{\Sigma}\,|\,\{\mathbf{Y},\mathbf{z}_2^*,\boldsymbol{\beta},\boldsymbol{\kappa}\} \sim \mathcal{IW}(T+\nu,\overline{\mathbf{S}}),\\
&\overline{\mathbf{S}} = \mathbf{S} + (\mathbf{W}-\boldsymbol{\iota}_T\overline{\mathbf{p}})'(\mathbf{W}-\boldsymbol{\iota}_T\overline{\mathbf{p}}) + q^{-1}(\mathbf{p}-\overline{\mathbf{p}})'(\mathbf{p}-\overline{\mathbf{p}}),\\
&\mathbf{W} = \mathbf{Y}^* - \mathbf{X}(\boldsymbol{\kappa})(\boldsymbol{\iota}_2'\otimes\boldsymbol{\beta}), \quad \overline{\mathbf{p}} = (Tq+1)^{-1}(q\boldsymbol{\iota}_T'\mathbf{W}+\mathbf{p}),
\end{aligned}$$

and $\boldsymbol{\Pi}$ conditional on $\boldsymbol{\Sigma}$ from the bivariate normal distribution

$$\boldsymbol{\Pi}\,|\,\{\mathbf{Y},\mathbf{z}_2^*,\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\Sigma}\} \sim \mathcal{N}\left(\overline{\mathbf{p}}', q(Tq+1)^{-1}\boldsymbol{\Sigma}\right).$$

For the SD-testing procedure in Section 3.2.3 we replace this step with the following. We define the auxiliary matrices $\widetilde{\mathbf{U}} = (\boldsymbol{\iota}_T, \mathrm{LEI}^{1-\ell,T-\ell})$ and $\widetilde{\mathbf{W}} = \mathbf{Y}^* - \mathbf{X}(\boldsymbol{\kappa})(\boldsymbol{\iota}_2'\otimes\boldsymbol{\beta})$. With the resulting multivariate normal regression model and its conjugate prior in (3.13), we sample the regression parameters and the covariance matrix $\{\widetilde{\boldsymbol{\Pi}},\boldsymbol{\Sigma}\}$ in one block as

$$\begin{aligned}
&\boldsymbol{\Sigma}\,|\,\{\mathbf{Y},\mathbf{z}_2^*,\boldsymbol{\beta},\boldsymbol{\kappa}\} \sim \mathcal{IW}(\nu+T,\widetilde{\mathbf{S}}),\\
&\widetilde{\boldsymbol{\Pi}}\,|\,\{\mathbf{Y},\mathbf{z}_2^*,\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\Sigma}\} \sim \mathcal{MN}(\widetilde{\mathbf{P}},\boldsymbol{\Sigma}\otimes\widetilde{\mathbf{Q}}).
\end{aligned}$$

The posterior parameters of these distributions are

$$\begin{aligned}
&\widetilde{\mathbf{S}} = \mathbf{S} + (\widetilde{\mathbf{W}}-\widetilde{\mathbf{U}}\widetilde{\mathbf{P}})'(\widetilde{\mathbf{W}}-\widetilde{\mathbf{U}}\widetilde{\mathbf{P}}) + q^{-1}\widetilde{\mathbf{P}}'\widetilde{\mathbf{P}},\\
&\widetilde{\mathbf{Q}} = (\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}} + q^{-1}\mathbf{I}_2)^{-1}, \quad \widetilde{\mathbf{P}} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}},
\end{aligned}$$

using the prior zero mean and $q = 10$. To sample the lead times and the indicator weights we substitute $\mathsf{vec}[\widetilde{\mathbf{W}}]$ for $\mathbf{w}$, and to simulate the latent data we use $\mathbf{Y}^* = \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Pi}} + \mathbf{X}(\boldsymbol{\kappa})(\boldsymbol{\iota}_2'\otimes\boldsymbol{\beta}) + \boldsymbol{\varepsilon}^*$ instead.

3.A.3 Unrestricted bivariate model

In this section we discuss the technical details of the general unrestricted form (M_4) of our model (3.1)–(3.2). We fix the intercept of the latent variable at 1 such that we have the reparameterized latent variable $z_{2t}^{\dagger}$ and the unrestricted errors' covariance matrix $\boldsymbol{\Sigma}^{\dagger}$.[10]

[10]The lower-right element of $\boldsymbol{\Sigma}^{\dagger}$ is α_0^{-2}. Expansions occur more frequently than recessions, hence $\alpha_0 > 0$.

We arrange the $2J$ weights in $\boldsymbol{\Lambda} = (\boldsymbol{\beta}, \boldsymbol{\alpha}^\dagger)$, with $\boldsymbol{\alpha}^\dagger = \alpha_0^{-1}\boldsymbol{\alpha}$. The model parameters are $\boldsymbol{\theta} = \{\boldsymbol{\kappa}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}^\dagger, \beta_0\}$. Defining the auxiliary matrix $\mathbf{W}(\beta_0) = \mathbf{Y}^\dagger - \boldsymbol{\iota}_T(\beta_0, 1)$, with $\mathbf{Y}^\dagger = (\mathbf{Y}_1, \mathbf{z}_2^\dagger)$, we have the complete data likelihood function

$$p(\mathbf{Y}, \mathbf{z}_2^\dagger \,|\, \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}^\dagger|^{-T/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{\dagger-1}\left(\mathbf{W}(\beta_0) - \mathbf{X}(\boldsymbol{\kappa})\boldsymbol{\Lambda}\right)'(\mathbf{W}(\beta_0) - \mathbf{X}(\boldsymbol{\kappa})\boldsymbol{\Lambda})\right]\right\} \times \prod_{t=1}^{T}\left(\mathbb{I}_{\{y_{2t}=1\}}\mathbb{I}_{\{z_{2t}^\dagger<0\}} + \mathbb{I}_{\{y_{2t}=0\}}\mathbb{I}_{\{z_{2t}^\dagger\geq 0\}}\right). \tag{3.A.8}$$

We choose the relatively uninformative conditionally conjugate prior

$$\boldsymbol{\Sigma}^\dagger \sim \mathcal{IW}(\nu, \mathbf{S}), \quad \boldsymbol{\Lambda} \,|\, \boldsymbol{\Sigma}^\dagger \sim \mathcal{MN}(\mathbf{P}, \boldsymbol{\Sigma}^\dagger \otimes \mathbf{Q}), \tag{3.A.9}$$

with parameters $\nu = 6$ and $\mathbf{S} = \mathrm{diag}(15, 1)$, and $\mathbf{P} = \mathbf{0}_{2\times 2}$ and $\mathbf{Q} = 10\mathbf{I}_2$. For the single intercept we opt for $\beta_0 \sim \mathcal{N}(b, B)$, with mean 0 and variance 10. The prior for the lead times is as in (3.8). All together, the density of the prior distribution in this unrestricted model is

$$p(\boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}^\dagger|^{-J/2}\,|\mathbf{Q}|^{-1} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{\dagger-1}(\boldsymbol{\Lambda} - \mathbf{P})'\mathbf{Q}^{-1}(\boldsymbol{\Lambda} - \mathbf{P})\right]\right\} \exp\left\{-\frac{1}{2B}(\beta_0 - b)^2\right\} \times |\boldsymbol{\Sigma}^\dagger|^{-(\nu+N+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{\dagger-1}\mathbf{S}\right]\right\} \prod_{j=1}^{J}\sum_{k=1}^{K_j^{\max}} \mathbb{I}_{\{\kappa_j=k\}}, \tag{3.A.10}$$

with $N = 2$ the dimension of $\mathbf{y}_t$.

The product of (3.A.10) and (3.A.8) results in the kernel of the augmented posterior distribution of which we obtain a dependent sample by implementing the following MCMC scheme. In this model we use an even more efficient MCMC procedure since we analytically integrate out both the indicator weights $\boldsymbol{\Lambda}$ and the variance matrix $\boldsymbol{\Sigma}^\dagger$.

Step 1: Latent data

Sample each latent variable from $p(z_{2t}^\dagger \,|\, \mathbf{Y}, \boldsymbol{\theta})$, $(t = 1, \ldots, T)$. This step is the same as for the restricted model, of course with the modified regression means.

Step 2: Lead times

Sample each lead time from $p(\kappa_j \,|\, \mathbf{Y}, \mathbf{z}_2^\dagger, \beta_0, \boldsymbol{\kappa}_{-j}) \propto p(\boldsymbol{\kappa} \,|\, \mathbf{Y}, \mathbf{z}_2, \beta_0)$, $(j = 1, \ldots, J)$. The conditionally conjugate prior in (3.A.9) allows for the analytical integration of $\{\boldsymbol{\Lambda}, \boldsymbol{\Sigma}^\dagger\}$ in this step. That is, we first collect terms involving $\boldsymbol{\Lambda}$ from the augmented posterior, apply the decomposition rule (see Greenberg, 2008) and integrate the kernel of the matricvariate normal $\mathcal{MN}(\overline{\mathbf{P}}(\boldsymbol{\kappa}), \boldsymbol{\Sigma}^\dagger \otimes \overline{\mathbf{Q}}(\boldsymbol{\kappa}))$ with parameters

$$\overline{\mathbf{Q}}(\boldsymbol{\kappa}) = (\mathbf{X}(\boldsymbol{\kappa})'\mathbf{X}(\boldsymbol{\kappa}) + \mathbf{Q}^{-1})^{-1}, \quad \overline{\mathbf{P}}(\boldsymbol{\kappa}) = \overline{\mathbf{Q}}(\boldsymbol{\kappa})(\mathbf{X}(\boldsymbol{\kappa})\mathbf{W}(\beta_0) + \mathbf{Q}^{-1}\mathbf{P}).$$

We save terms from this integral involving $\boldsymbol{\kappa}$ and $\boldsymbol{\Sigma}^\dagger$, which means that we multiply the remaining terms with $|\boldsymbol{\Sigma}^\dagger|^{J/2}|\overline{\mathbf{Q}}(\boldsymbol{\kappa})|^{N/2}\exp\left\{-\mathrm{tr}\left[\boldsymbol{\Sigma}^{\dagger-1}\mathbf{R}(\boldsymbol{\kappa})'\mathbf{R}(\boldsymbol{\kappa})/2\right]\right\}$, in which

$$\mathbf{R}(\boldsymbol{\kappa})'\mathbf{R}(\boldsymbol{\kappa}) = [\mathbf{W}(\beta_0) - \mathbf{X}(\boldsymbol{\kappa})\overline{\mathbf{P}}(\boldsymbol{\kappa})]'[\mathbf{W}(\beta_0) - \mathbf{X}(\boldsymbol{\kappa})\overline{\mathbf{P}}(\boldsymbol{\kappa})] + [\overline{\mathbf{P}}(\boldsymbol{\kappa}) - \mathbf{P}]'\mathbf{Q}^{-1}[\overline{\mathbf{P}}(\boldsymbol{\kappa}) - \mathbf{P}].$$

The deteminant terms with the $J/2$ powers cancel out, leaving us with the kernel of the inverted Wishart $\mathcal{IW}(T+\nu, \mathbf{R}(\boldsymbol{\kappa})'\mathbf{R}(\boldsymbol{\kappa})+\mathbf{S})$. This kernel integrates to the determinant of the inverted Wishart's matrix to the power $-(T+\nu)/2$. Thus, we end up with

$$\begin{aligned} p(\boldsymbol{\kappa}\,|\,\mathbf{Y},\mathbf{z}_2^\dagger,\beta_0) &\propto \int p(\boldsymbol{\theta},\mathbf{z}_2^\dagger\,|\,\mathbf{Y})\,\mathsf{d}\{\boldsymbol{\Lambda},\boldsymbol{\Sigma}^\dagger\} \\ &\propto \left|\overline{\mathbf{Q}}(\boldsymbol{\kappa})\right|^{N/2} \times \left|\mathbf{R}(\boldsymbol{\kappa})'\mathbf{R}(\boldsymbol{\kappa})+\mathbf{S}\right|^{-(T+\nu)/2} \times p(\boldsymbol{\kappa}). \qquad (3.A.11) \end{aligned}$$

We evaluate the kernel in (3.A.11) for all supported values for κ_j, normalize and sample from this discrete distribution.

Step 3: Indicator weights and covariance matrix

Sample the weights and variance matrix in one block from $p(\boldsymbol{\Lambda},\boldsymbol{\Sigma}^\dagger\,|\,\mathbf{Y},\mathbf{z}_2^\dagger,\beta_0,\boldsymbol{\kappa})$. From the derivations for Step 2 it follows that we simulate

$$\begin{aligned} &\boldsymbol{\Sigma}^\dagger\,|\,\{\mathbf{Y},\mathbf{z}_2^\dagger,\beta_0,\boldsymbol{\kappa}\} \sim \mathcal{IW}(T+\nu,\mathbf{R}(\boldsymbol{\kappa})'\mathbf{R}(\boldsymbol{\kappa})+\mathbf{S}), \\ &\boldsymbol{\Lambda}\,|\,\{\mathbf{Y},\mathbf{z}_2^\dagger,\beta_0,\boldsymbol{\kappa},\boldsymbol{\Sigma}^\dagger\} \sim \mathcal{MN}(\overline{\mathbf{P}}(\boldsymbol{\kappa}),\boldsymbol{\Sigma}^\dagger\otimes\overline{\mathbf{Q}}(\boldsymbol{\kappa})). \end{aligned}$$

Step 4: Intercept

Sample the single intercept from $p(\beta_0\,|\,\mathbf{Y},\mathbf{z}_2^\dagger,\boldsymbol{\theta}_{-\beta_0})$. We define the auxiliary regression model for $\mathbf{y}_t^\dagger=(y_{1t},z_{2t}^\dagger)'$, with $\mathbf{x}_t'(\boldsymbol{\kappa})$ the t-th row of $\mathbf{X}(\boldsymbol{\kappa})$:

$$\mathbf{y}_t^\dagger-\boldsymbol{\Lambda}'\mathbf{x}_t(\boldsymbol{\kappa})-(0,1)'=(\beta_0,0)'+\boldsymbol{\varepsilon}_t^\dagger,\quad \boldsymbol{\varepsilon}_t^\dagger\overset{i.i.d.}{\sim}\mathcal{N}(\mathbf{0}_2,\boldsymbol{\Sigma}^\dagger),\quad (t=1,\ldots,T), \qquad (3.A.12)$$

in which the errors are contemporaneously correlated. The latent $z_{2t}^\dagger$'s are informative about β_0, if and only if correlation $\rho\neq 0$. We manipulate this model such that we obtain an equivalent regression in which the errors are uncorrelated. If we define the Cholesky matrix $\mathbf{L}$ such that $\mathbf{L}\mathbf{L}'=\boldsymbol{\Sigma}^\dagger$ and multiply both sides of (3.A.12) by the inverse of the Cholesky, we have

$$\begin{aligned} &\mathbf{L}^{-1}[\mathbf{y}_t^*-\boldsymbol{\Lambda}'\mathbf{x}_t(\boldsymbol{\kappa})-(0,1)']=\mathbf{L}^{-1}(\beta_0,0)'+\mathbf{L}^{-1}\boldsymbol{\varepsilon}_t^\dagger \\ \iff\quad &\mathbf{w}_t=\mathbf{l}_1\beta_0+\boldsymbol{\eta}_t,\quad \boldsymbol{\eta}_t\overset{i.i.d.}{\sim}\mathcal{N}(\mathbf{0}_2,\mathbf{I}_2),\quad (t=1,\ldots,T), \end{aligned}$$

in which $\mathbf{l}_1$ is the first column of $\mathbf{L}^{-1}=(\mathbf{l}_1,\mathbf{l}_2)$. Therefore we get the univariate regression model with the $2T\times 1$ vector $\mathbf{w}=(\mathbf{w}_1',\ldots,\mathbf{w}_T')'$ with dependent variables, one column with regressor observations $\mathbf{V}=\boldsymbol{\iota}_T\otimes\mathbf{l}_1$, and all errors mutually uncorrelated with unit variance. We simulate the intercept as

$$\beta_0\,|\,\{\mathbf{Y},\mathbf{z}_2^\dagger,\boldsymbol{\theta}_{-\beta_0}\}\sim\mathcal{N}\Big((\mathbf{V}'\mathbf{V}+B^{-1})^{-1}(\mathbf{V}'\mathbf{w}+b/B),(\mathbf{V}'\mathbf{V}+B^{-1})^{-1}\Big).$$

CHAPTER 4

Parameter Changes in Empirical Macroeconomic Models

4.1 Introduction

Over the last two decades, empirical evidence showing that macroeconomic and financial time series are subject to changes in their statistical properties has mounted, see Stock and Watson (1996) and Andreou and Ghysels (2009), among many others. A prominent example in macroeconomics is the Great Moderation, referring to the large drop in volatility experienced by many macroeconomic time series in the first half of the 1980s, see, for example, McConnell and Perez-Quiros (2000), Stock and Watson (2002a), Sensier and van Dijk (2004) and Kim *et al.* (2008). Also in finance, the presence of structural breaks in predictive regression models for asset returns is by now well documented, see Pesaran and Timmermann (2002), Paye and Timmermann (2006), Rapach and Wohar (2006), Lettau and Van Nieuwerburgh (2008), Ravazzolo *et al.* (2008), Pettenuzzo and Timmermann (2011) and Dangl and Halling (2012).

Many empirical studies reporting evidence for structural changes in macroeconomic and financial time series make use of frequentist methods for detecting and dating such changes, as developed by Andrews (1993), Andrews and Ploberger (1994), Bai and Perron (1998), Bai *et al.* (1998) and Qu and Perron (2007), and we refer to Perron (2006) for a recent survey. These methods are classified as "historical" testing procedures (Andreou and Ghysels, 2009), in the sense that they are designed for testing for structural change and the identification of potential break dates *ex post* for time series observations spanning a given historical, in-sample period.[1] Out-of-sample forecasting in the presence of parameter changes has presented a much bigger challenge when relying upon frequentist methods, see the survey of Clements and Hendry (2006). A Bayesian approach better suits this problem, in the sense that parameter change is made an inherent part of the statistical time series model, in particular including the possibility that parameter changes occur in the out-of-sample period. Surprisingly then, accounting for possible future parameter changes when constructing out-of-sample forecasts has not received much attention yet in the Bayesian literature, with the

[1] A different strand of literature concerns testing for structural change "in real time," *i.e.*, monitoring whether new, incoming observations are consistent with a previously specified model, see Chu *et al.* (1996) and Zeileis *et al.* (2005), among others.

notable exceptions of Pesaran *et al.* (2006), Koop and Potter (2007), Maheu and Gordon (2008), Geweke and Jiang (2011), Maheu and Song (2014) and Song (2014).

In this chapter we propose a Bayesian mixture approach based on Dirichlet process mixture (DPM) methods to deal with parameter changes in empirical macroeconomic (time-series) models, with an explicit focus on the implications for out-of-sample forecasting. The stochastic DPM specification to describe the dynamic behavior of the parameter has a simple and intuitively appealing interpretation. In each period, with a particular probability a parameter change occurs and in that case the new parameter value is generated by a so-called base prior distribution. Otherwise the parameter value is equal to one of the already existing values. Thus, the (conditional) distribution of the model parameter is a two-component mixture of which one component is the base prior distribution and the other component consists of parameter values from the elapsed period thus far. The mixing probability for the first component is the probability of a new-born parameter value.

The key advantage of the specification lies in the flexibility of the Bayesian procedures for estimation and forecasting to deal with a wide range of different models. For estimation purposes, we use the Markov chain Monte Carlo (MCMC) based algorithms by Neal (2000) to simulate from the posterior distribution of the model parameters. These posterior simulators boil down to sampling from continuous-discrete mixture distributions. For forecasting purposes, the predictive distributions of future observations are of a mixture type as well, with one component being the model under the no-change scenario and the other being the model integrated over the base prior in case of a change. If the forecast horizon grows, the probability of a break in the out-of-sample period increases and the latter mixture component gets more weight.

Our approach to parameter changes offers two key advantages compared to other existing methods. Both are closely related to the desirable properties of parameter-change models as formulated by Koop and Potter (2007). The first advantage is that our specification allows for an *a priori* unknown number and timing of changes. In particular, the specification naturally allows for the possibility that changes occur beyond the in-sample period. It is commonly recognized that allowing for future changes is important for realistic out-of-sample forecasting. Previous attempts to do so have certain limitations. Pesaran *et al.* (2006) propose an out-of-sample extension of the Markovian model of Chib (1998). In that approach (structural) changes are modeled by means of a non-recurring Markov process which requires the specification of the number of breaks that occur, both in- and out-of-sample, see also Koop and Potter (2007). Pesaran *et al.* (2006) circumvent this issue by applying Bayesian model averaging over distinct scenarios, each with a specific number of breaks in the out-of-sample period. However, this procedure is computationally cumbersome, and still requires a specific plausible choice of the maximum number of breaks to happen during the forecasting horizon. The DPM approach solves these problems by specifying the number of parameter changes to be stochastic, both in- and out-of-sample. Moreover, parameter values from the past can recur, allowing for temporary, reversible parameter shifts.

A second advantage of our approach lies its ability to deal with parameter changes in various types of models. Existing structural break approaches are confined to linear regression models (*e.g.* Maheu and Gordon, 2008, Geweke and Jiang, 2011) or models that can, at least conditionally, be written in Gaussian state-space form, as in the dynamic mixture innovation models advocated by Gerlach *et al.* (2000), Giordani *et al.* (2007) and Giordani and Kohn (2008). By contrast, our set-up includes different types of models (or sampling distributions) as well, including models for limited dependent variables, and copula models for describing non-standard dependence structures, within or between time series. This flexibility is due to the computational advantages offered by the posterior simulators of Neal (2000). Each run of his simulators only requires evaluations of one-observation likelihoods and sampling from simple mixture distributions. Other simulators require analytical integration with respect to the regime-specific parameters to establish convergence, see Geweke and Jiang (2011), or Gerlach *et al.* (2000) for a similar solution in a (conditional) Gaussian state-space specification. Their application is, therefore, limited to conditionally conjugate models, and their feasibility relies on the computational ease of the integration step.

The dynamic mixture innovation framework of Giordani *et al.* (2007) and Giordani and Kohn (2008) crucially depends upon the assumption that the model can be written in Gaussian state-space form (at least conditionally) with the parameters treated as the states. The state equations are specified such that the parameter values are sampled from a mixture of a degenerate and a Gaussian component. For computational reasons, this Gaussian state-space approach requires the change in the parameter to come from a (mixed) normal distribution, otherwise the sampling methods developed by Giordani and Kohn (2008) cannot be applied. Theoretically this would not be too restrictive as long as the support for the parameter is unrestricted. If, however, the support of the parameter is restricted or prior beliefs confine the supported region (*e.g.* by truncation), this approach breaks down. The DPM framework is flexible with respect to distributional assumptions of the parameters, as we just opt for a base prior that has the appropriate features. Furthermore, the dynamic mixture innovations approach not only requires the (mixed) Gaussian assumption of the state innovations, but also of the measurement equation. The DPM methods are applicable to any kind of parametric sampling distribution, including models for continuous and discrete time-series variables, or a combination of the two.

The base prior distribution forms an important part of the modeling process. In the case of a parameter change the new parameter value is drawn from this distribution, independently from previous values. Its choice is guided by the model opted for, in terms of support, and, second, by prior beliefs. By including a third layer in the model, the independence assumption is relaxed and we update the prior beliefs (formulated by the base prior's hyperparameters). Such an additional hierarchical step makes regimes from the past informative for future parameter values and this information is properly absorbed into the predictive distribution. We find a similar step for structural break models in Carlin *et al.* (1992), Pesaran *et al.* (2006), Geweke and Jiang (2011) and Song (2014).

The outline of the remainder of this chapter is as follows. In Section 4.2 we introduce the DPM specification to account for parameter changes, analyze its implications for out-of-sample forecasting and describe issued related to the choice of an appropriate base prior distribution and the probability of a parameter change. We discuss the methods to simulate from the posterior distribution in Section 4.3, and as a demonstration we analyze an example with simulated data from an autoregression augmented with explanatory variables. In addition to the example model, in Section 4.4 we further demonstrate the use and wide applicability of our approach, both for descriptive in-sample analysis and for constructing out-of-sample forecasts. We examine three macroeconomic applications involving different types of models, including a copula model, a Poisson predictive regression model, and a Markov regime-switching model with variance changes. We provide a conclusion in Section 4.5. Issues related to simulation from the posterior distributions and to out-of-sample forecasting are exposed in Appendix 4.A.

4.2 Model specification

In this section we introduce the modeling framework to account for changes in an empirical model's parameters. In Section 4.2.1 we discuss the hierarchical modeling procedure. We focus on the implications of the model specification for out-of-sample forecasting in Section 4.2.2. We conclude this section with a discussion on the two inputs of the procedure, *i.e.*, the base prior distribution (Section 4.2.3) and the smoothness parameter (Section 4.2.4).

4.2.1 Hierarchical model

We let y_t be the target variable, which is observed for $t = 1, \dots, T$, and we define $\mathbf{y}^{k,l} = (y_k, y_{k+1}, \dots, y_l)'$, $(1 \leq k < l \leq T)$. Thus, $\mathbf{y}^{1,T} = \mathbf{y}$ denotes the complete set of time-series observations in the in-sample period. The time series in period t is characterized by a distribution with probability density function (pdf) $p(y_t \,|\, \mathbf{y}^{1,t-1}, \theta_t)$, which is parameterized by the time-varying parameter θ_t.[2] This is the first layer of the model.

The second model layer specifies the stochastic process that assigns each θ_t its value. We account for heterogenous θ_t's by employing a Dirichlet process, see Ferguson (1973) and Antoniak (1974) for a (technical) introduction, and Escobar (1994) for its applications. In particular, θ_t, $(t = 1, \dots, T)$, is the outcome of a Dirichlet process with base distribution with pdf $f_0(\cdot; \lambda)$, and smoothness parameter α. This specification implies that with positive probability some θ_t's equal the same value θ_i^*, $(i = 1, 2, \dots, I_T^* \leq T)$, such that θ_t's cluster into a "regime" with parameter value θ_i^*.

The stochastic process which describes the evolution of the time-varying parameters behaves according to the following set of conditional densities

$$p(\theta_t \,|\, \boldsymbol{\theta}^{1,t-1}) = \alpha a_{t-1} f_0(\theta_t; \lambda) + a_{t-1} \sum_{s=1}^{t-1} \delta(\theta_t - \theta_s), \quad (t = 2, \dots, T), \tag{4.1}$$

with initialization by $p(\theta_1) = f_0(\theta_1; \lambda)$. δ is the Dirac delta function which concentrates unit point mass at zero,[3] and the integrating constants are $a_s = (\alpha + s)^{-1}$. At time t the process takes on either a newly f_0-generated value, with probability proportional to the smoothness parameter, or equals one of the $t - 1$ existing values.

The discrete part of the mixture specification in (4.1) implies that particular parameter values recur with positive probablity. Consequently, we partition a sample $\boldsymbol{\theta} = (\theta_1, \dots, \theta_T)'$ into I_T^* subsets indexed by i. Each subset contains N_i^* members and we denote the corresponding parameter value as θ_i^*. Therefore, after having "observed" $\boldsymbol{\theta}$, the first out-of-sample parameter θ_{T+1} is generated from the conditional distribution with density

$$p(\theta_{T+1} \,|\, \boldsymbol{\theta}) \propto \alpha f_0(\theta_{T+1}; \lambda) + \sum_{i=1}^{I_T^*} N_i^* \delta(\theta_{T+1} - \theta_i^*). \tag{4.2}$$

The number of unique parameter values is a random variable of which the distribution depends upon α and sample size T. We discuss further issues related to the implied prior for I_T^* in Section 4.2.4.

[2] The distribution's parameters may depend on explanatory variables $\mathbf{x}_t$ or time-constant ω as well, which we leave out from the conditioning set here for reasons of legibility.

[3] Furthermore, we use the point-mass measure and define $\int g(\theta)\delta(\theta)\,\mathsf{d}\theta \equiv g(0)$.

The model specification is applicable to a multivariate parameter setting as well. In that case we either impose simultaneous changes in all parameters or we allow individual parameters to change independently by adopting for each a different process as in (4.1). Similarly, although we restrict the notation to univariate time series here, the modeling framework can straightforwardly be extended to multivariate dependent variables.

4.2.2 Forecasting implications

One of the main reasons why times series models may perform poorly in terms of (out-of-sample) forecasting is the often incorrect assumption that model parameters are constant over time. As shown by Clements and Hendry (2001, 2006), among others, neglecting parameter changes that occur during the in-sample period yields biased forecasts. Various (frequentist and Bayesian) methods are available for detecting and modeling parameter instability in-sample. However, if parameter changes have occurred in the past, it is likely that further changes occur during the out-of-sample period as well. Not accounting for this possibility results in more concentrated predictive densities, though an essential source of uncertainty is ignored and the location of the predictive density is possibly misplaced. In this section we focus on the implications of our methods in terms of out-of-sample forecasting by examining how this uncertainty with regard to any future parameter changes affects the predictive densities.

A Bayesian forecast is a specific property of the target variable's posterior predictive distribution which depends on the loss function of the forecaster (or decision maker). The predictive distribution formalizes the combination of the model structure, prior beliefs about parameter values and information revealed by the data. At time τ, $y_{\tau+1}$'s predictive density $p(y_{\tau+1} \,|\, \mathbf{y}^{1,\tau})$ is demarginalized, by recalling the hierarchical steps of Section 4.2.1, as

$$p(y_{\tau+1} \,|\, \mathbf{y}^{1,\tau}) = \int\int p(y_{\tau+1} \,|\, \mathbf{y}^{1,\tau}, \theta_{\tau+1}) p(\theta_{\tau+1} \,|\, \boldsymbol{\theta}^{1,\tau}) \, \mathsf{d}\theta_{\tau+1} \; p(\boldsymbol{\theta}^{1,\tau} \,|\, \mathbf{y}^{1,\tau}) \, \mathsf{d}\boldsymbol{\theta}^{1,\tau}.$$

The second density under the integral is supplied by (4.2). Thus, given the parameters from the past, $\theta_{\tau+1}$ equals one of the I_τ^* parameter values with probability $N_i^*/(\alpha + \tau)$, and with probability $\alpha/(\alpha + \tau)$ a new regime occurs with a parameter value generated by $f_0(\theta_{\tau+1}; \lambda)$. We mix the conditional distribution of $y_{\tau+1}$ over this discrete-continuous mixture distribution, and finally, we integrate out parameter uncertainty with respect to the parameter values from the past.

A number of remarks are in order. First, all regimes that have taken place in the past are allowed to recur in the future. Second, the total duration (N_i^*) of a regime, possibly divided over separated time periods, determines the importance of that regime in the predictive distribution. For example, if a single change has occured, the weight of the pre-change regime dies out as time proceeds, and the new regime starts dominating the predictive distribution. The downside is also evident. In case of a parameter change near the end of the sample, the old regime is assigned most weight. However, each structural break model contains issues of this kind. If a structural break model is able even to detect the break near the end of the sample, it has only very few observations to reliably identify the new regime's parameter value, leading to $p(y_{\tau+1} \,|\, \mathbf{y}^{1,\tau}, \theta_{\tau+1})$ being mixed over a distribution close to the prior of $\theta_{\tau+1}$. In such a situation we at least use parameter values obtained from the data. We note that Pesaran and Timmermann (2002), Pesaran and Pick (2011) and Pesaran *et al.* (2013) propose frequentist methods to address this kind of location-scale trade-off issues.

Third, the method accounts for future parameter changes. At time τ, the one-period-ahead predictive distribution accounts for a parameter change with probability $\alpha/(\alpha+\tau)$ in period $\tau+1$. Regardless of the size of τ, with positive probability there will be parameter shifts in the future. However, the greater τ, the less likely a new regime occurs in the next period. There has been ample opportunity to "generate" new parameter values such that the need for additional regimes gets smaller and in case of a parameter change, we rather expect one of the existing regimes to recur.

Fourth, if we construct an h-period-ahead forecast, we expect $\sum_{j=0}^{h-1} \alpha/(\alpha+\tau+j)$ new parameter values to be generated in the next h periods from τ on.[4] Obviously, for fixed τ, the larger the forecasting horizon, the more new parameter values we expect and as a result, existing values get a smaller weight in the predictive density.

Finally, we compute the predictive distributions by Monte Carlo integration. With $\boldsymbol{\theta}^{1,\tau;(m)}$ as a draw from the posterior $p(\boldsymbol{\theta}^{1,\tau} \,|\, \mathbf{y}^{1,\tau})$, $(m=1,\ldots,M)$, we simulate M paths of the intermediate parameters $\boldsymbol{\theta}^{\tau+1,\tau+h}$, and data $\mathbf{y}^{\tau+1,\tau+h-1}$. We compute the Monte Carlo average

$$M^{-1}\sum_{m=1}^{M} p\Big(y_{\tau+h} \,|\, \mathbf{y}^{1,\tau+h-1;(m)}, \theta_{\tau+h}^{(m)}\Big), \quad (h=1,2,\ldots),$$

which is a simulation-consistent estimator of $p(y_{\tau+h} \,|\, \mathbf{y}^{1,\tau})$.

4.2.3 Base distribution

From (4.1), if a new parameter value is born, it is generated from the base distribution with pdf f_0. To operationalize the methods we need to choose (i) the type of distribution and (ii) its parameter λ.

When choosing the type of base prior distribution, two considerations are of importance. First, the distribution y_t is generated from indicates the parameter space of θ_t. The support of f_0 must coincide with this space. Second, for computational reasons we favor a base prior that forms a conjugate pair with the sampling distribution, though non-conjugate pairs can be dealt with equally well.

The setting of the hyperparameter λ partly depends on the ultimate goal of the research. If it is mostly exploratory, that is, if we merely want to check for the possibility of parameter changes in the past, it suffices to choose a relatively uninformative prior that covers regions with plausible values sufficiently. But if we want to optimally learn from the past in order to more accurately forecast future y_t's, we put a prior on λ.

Suppose there turn out to be I_T^* regimes during the in-sample period, each with parameter value θ_i^*. Each element of $\boldsymbol{\theta}^* = (\theta_1^*,\ldots,\theta_{I_T^*}^*)'$ is generated from the base distribution. Moreover, conditional on λ these I_T^* unique parameters are statistically independent. The first advantage of the additional model layer is that after marginalizing with respect to λ the regimes show dependence. Second, and more important, it allows for a data-updating step to learn about λ. Both advantages combined have the desirable effect that parameter values from the past provide information relevant for future regimes.

[4] The increase in the number of unique parameter values from time $\tau+1$ up to and including $\tau+h$ equals $\sum_{j=0}^{h-1} \mathbb{I}_{\{U_j<\alpha/(\alpha+\tau+j)\}}$, with $U_j \overset{i.i.d.}{\sim} \mathcal{U}([0,1])$. We note that the number of new parameter values is independent of the number generated during period 1–τ. Its variance equals $\sum_{j=0}^{h-1} \alpha^2/(\alpha+\tau+j)^2$. Hence, for $\tau\to\infty$ no new regimes are to be expected.

If we integrate out λ, we obtain the marginal base prior $\int f_0(\theta_t;\lambda)p(\lambda)d\lambda$. This distribution provides insights in what values for θ_t are *a priori* covered. We note that in our parameter-change setting there will be a limited number of regimes and, hence, a limited number of unique θ_i^* values. Since these contain all the information in the data relevant for learning about λ, $p(\lambda \,|\, \mathbf{y})$ is generally close to $p(\lambda)$.

4.2.4 Smoothness parameter

The smoothness parameter of the Dirichlet process governs the number of regimes the time series is subjected to. That is, the choice of α implies a prior distribution for the number of unique parameter values I_T^*. Beside α, the distribution of I_T^* automatically depends (positively) on the sample size T as well. Antoniak (1974) and Escobar and West (1995) provide the implied distribution of the number of regimes given α and T, which is

$$\mathsf{Pr}\left[I_T^* = i \,|\, \alpha\right] = c_T(i)\alpha^i \frac{\Gamma(\alpha)}{\Gamma(\alpha+T)}, \quad (i = 1, 2, \ldots, T), \tag{4.3}$$

with $c_T(i)$ the unsigned Stirling numbers of the first kind. I_T^* covers the entire imaginable support, ranging from no parameter changes at all to the most extreme case with a different parameter value in each period. The smoothness parameter determines the distribution of the probability mass over this range. A prior belief about the number of regimes helps to set this parameter. We propose to set α in terms of the prior mode of I_T^*. That is, if we *a priori* believe $\mathsf{mode}\left[I_T^*\right] = i^*$, $(1 \leq i^* \leq T)$, we choose the smoothness parameter as

$$\alpha_{i^*} = \frac{1}{2}\Big(\exp\left\{-\Delta c(i^*+1)\right\} + \exp\left\{-\Delta c(i^*)\right\}\Big), \tag{4.4}$$

with $\Delta c(k) = \log c_T(k) - \log c_T(k-1)$, and for the end points of the spectrum we define $\Delta c(1) = \log c_T(1)$ and $\Delta c(T+1) = \Delta c(T) - 1$. Here we will not apply the latter, which is only required for the case of a new-born parameter in each of the T periods.

As Conley *et al.* (2008) remark, fixing the value of the smoothness parameter can result in a too informative prior on the number of regimes, especially when α is small, which holds in our case. By putting a prior on α and using the data to learn, this issue can be dealt with. Escobar (1994), Escobar and West (1995) and Conley *et al.* (2008) each suggest a type of prior distribution. Both the first and the latter propose to discretize the support of α, whereas Escobar and West (1995) suggest the most elegant solution by using a Gamma distribution such that we sample α with an additional data augmentation step in a Gibbs sampling routine.

We proceed as follows. First, we consider what number of unique parameter values would be most likely, and next check whether the implied prior is not too informative. In case it is, we opt for a Gamma prior on α with a mode equal to the value in (4.4), implying the required mode of I_T^*, however with sufficient dispersion such that the prior probability mass of I_T^* is more evenly distributed. To *a priori* examine the effects of the prior for α on the number of mixture components, we compute the marginal prior probability mass function

$$p(I_T^*) = \int p(I_T^* \,|\, \alpha)p(\alpha)\, \mathsf{d}\alpha, \quad (I_T^* = 1, 2, \ldots, T),$$

with Monte Carlo integration. The first density under the integral is evaluated by applying the mass function (4.3) and $p(\alpha)$ is the pdf of a $\mathcal{G}a(a_\alpha, b_\alpha)$ distribution such that $a_\alpha/b_\alpha = \alpha_{i^*}$.

4.3 Posterior simulation

In this section we discuss the procedure to simulate from the posterior distribution with pdf $p(\boldsymbol{\theta}, \alpha, \lambda \,|\, \mathbf{y}) \propto p(\mathbf{y} \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \alpha, \lambda) p(\alpha, \lambda)$. We use simulation techniques from the class of Markov chain Monte Carlo (MCMC) methods, see, for example, Robert and Casella (2004). Section 4.3.1 deals with the core of the sampling method, *i.e.*, with sampling the time-varying parameters. In Sections 4.3.2 and 4.3.3 we provide the simulation methods for the smoothness parameter and the parameters of the base distribution, respectively.

4.3.1 Time-varying parameters

We decompose the likelihood function as $p(\mathbf{y} \,|\, \boldsymbol{\theta}) = \prod_{t=1}^{T} p_t(y_t; \theta_t)$, with individual contributions $p_t(y_t; \theta_t) = p(y_t \,|\, \mathbf{y}^{1,t-1}, \theta_t)$. $p(\boldsymbol{\theta} \,|\, \alpha, \lambda)$ is given in (4.1). The simulation scheme for the time-varying parameters consists of two steps (we leave out the explicit conditioning on $\{\alpha, \lambda\}$ for notational convenience).

Step 1. Simulate each parameter θ_t, $(t = 1, \ldots, T)$, from its full conditional posterior distribution with density $p(\theta_t \,|\, \mathbf{y}, \boldsymbol{\theta}_{-t}) \propto p_t(y_t; \theta_t) p(\theta_t \,|\, \boldsymbol{\theta}_{-t})$;

Step 2. In order to enhance convergence of the Markov chain, implement a so-called remix step as described by Escobar and West (1995) and Neal (1998) (published as Neal, 2000).

In the following we further discuss issues related to the implementation of both these steps.

Step 1: Single-move step

Ferguson (1973) shows that the T parameters in (4.1) have full conditional prior densities $p(\theta_t \,|\, \boldsymbol{\theta}_{-t})$ which are also of the discrete-continuous mixture type. Conditional independence of (the innovations of) observations y_t given the time-varying parameters, and applying Bayes' rule results in the full conditional posteriors

$$\begin{aligned} p(\theta_t \,|\, \mathbf{y}, \boldsymbol{\theta}_{-t}) &\propto p_t(y_t; \theta_t) p(\theta_t \,|\, \boldsymbol{\theta}_{-t}) \\ &\propto \alpha p_0(y_t; \lambda) p(\theta_t \,|\, y_t) + \sum_{s \neq t} p_t(y_t; \theta_s) \delta(\theta_t - \theta_s), \quad (t = 1, \ldots, T). \end{aligned} \tag{4.5}$$

Thus, θ_t equals one of the other $T-1$ values with probability proportional to the likelihood contribution of y_t under regime parameter θ_s, and with probability proportional to $\alpha p_0(y_t; \lambda) = \alpha \int p_t(y_t; x) f_0(x; \lambda) \, \mathrm{d}x$ a new regime is born and its value is a realization of the distribution with pdf $p(\theta_t \,|\, y_t) \propto p_t(y_t; \theta_t) f_0(\theta_t; \lambda)$. In Section 4.2.3 we favored the use of conjugate pairs likelihood-f_0, the computational reason is obvious from the preceding: we simulate θ_t from a distribution of a known type in case of a new-born regime, and the "one-observation" marginal likelihoods are available analytically. We refer to Example 1 and the illustrations in Section 4.4 for various specific cases.

In empirical applications in which the pair $p_t(y_t; \theta_t)$-$f_0(\theta_t; \lambda)$ is non-conjugate, we follow Neal (1998) and employ a Metropolis–Hastings (MH) sampler for this step of the simulation scheme. This sampler is particularly useful in, for instance, a copula approach to model the joint distribution of a number of variables, or, a Poisson model with heterogeneity in an explanatory variable's effect on y_t.

Neal's (1998) Algorithm 5 prescribes to sample a proposal value $\theta^{\#}$ for θ_t from the dominating part of the target distribution in (4.5). That is, we sample from the full conditional prior distribution with pdf

$$p(\theta^{\#} \,|\, \boldsymbol{\theta}_{-t}) \propto \alpha f_0(\theta^{\#}; \lambda) + \sum_{s \neq t} \delta(\theta^{\#} - \theta_s).$$

With probability $\alpha/(\alpha + T - 1)$ we sample a new regime from the base prior f_0 and with probability $N_i^*/(\alpha + T - 1)$ we select an existing regime's parameter value. N_i^* is the number of the other $T - 1$ observations assigned to regime i.

The missing term of the target density in the candidate-generating pdf is the likelihood contribution of y_t. Hence, in iteration $m + 1$ of the posterior simulation procedure, the acceptance probability is

$$\alpha_t(\theta^{\#}, \theta_t^{(m)}) = \min \left\{ p_t(y_t; \theta^{\#}) / p_t(y_t; \theta_t^{(m)}), 1 \right\}.$$

We set $\theta_t^{(m+1)} = \theta^{\#}$ with probability α_t and with probability $1 - \alpha_t$ we keep the current value and set $\theta_t^{(m+1)} = \theta_t^{(m)}$. When generating proposals the algorithm does not discount θ_t-information revealed by y_t. The dominance of a regime in terms of its duration is important, and, if a new regime is chosen, its parameter value just comes from the base prior. Neal (1998) therefore proposes to repeat, for fixed t, the previous MH step R times in order to give the algorithm a reasonable chance to arrive at fruitful proposal values.

Step 2: Remix step

As the first step of the sampler is of the single-move type, the Markov chain generally shows slow mixing. After running one iteration of Step 1, we obtain a value for $\boldsymbol{\theta}$. Conditional on these T parameter values we have I_T^* unique regimes, according to which we partition the sample of $\mathbf{y}$ into subsamples $\mathbf{y}_{(i)} = \{y_t : t \in T_{(i)}\}$, with $T_{(i)} = \{t : \theta_t = \theta_i^*\}$, $(i = 1, \ldots, I_T^*)$. Every unique parameter value is an independent realization from f_0 (conditional on $\{\alpha, \lambda\}$). In this second step we update each parameter value θ_i^* by simulating from

$$p(\theta_i^* \,|\, \mathbf{y}_{(i)}) \propto k_i(\theta_i^*) = f_0(\theta_i^*; \lambda) \prod_{t \in T_{(i)}} p_t(y_t; \theta_i^*), \quad (i = 1, \ldots, I_T^*). \tag{4.6}$$

In the conjugate setting, this density is the "multi-observation" version of the continuous mixture component $p(\theta_t \,|\, y_t)$ in Step 1, and hence it has a known form and we simulate from it directly.

For non-conjugate settings we implement a 2D slice sampler to simulate from (4.6), see Neal (1997) and Robert and Casella (2004, Ch. 8). Given a value $\theta^{\#}$ in the support of (4.6), we evaluate the kernel of this density and draw a value u_i uniformly distributed on $[0, k_i(\theta^{\#})]$. Then we find the interval $A(u_i) = \{x : k_i(x) \geq u_i\}$ and sample θ_i^* from the uniform distribution on $A(u_i)$. As long as the value $\theta^{\#}$ is in the support of the target distribution, the next draw of θ_i^* is exactly distributed as in (4.6). Since we do not know the posterior support yet, we use the current value of regime i's parameter as input to the slice sampler. In iteration $m + 1$ of the MCMC procedure we successively sample

$$u_i \sim \mathcal{U}\Big([0, k_i(\theta_i^{*(m)})]\Big), \qquad \theta_i^{*(m+1)} \sim \mathcal{U}(A(u_i)).$$

For multidimensional time-varying parameters we use the slice sampler to successively simulate from each univariate full conditional posterior distribution.

When the parameter space of θ is restricted, we implement the above slice sampler to an unrestricted transformation $g(\theta) \in (-\infty, \infty)$. Because the form of the target density is generally unknown, probability mass can be concentrated at boundary values such that it becomes difficult to determine the set $A(u_i)$. In these cases we sample the transformed parameters $\vartheta_i = g(\theta_i^*)$ by slice sampling with the kernels

$$k_i^g(\vartheta_i) = k_i(g^{-1}(\vartheta_i)) \times |\mathcal{J}_{\theta \mapsto g}(\vartheta_i)|, \quad (i = 1, \ldots, I_T^*). \tag{4.7}$$

The I_T^* updated regime parameters obtained in this second step serve as input to the next iteration of Step 1. In the following example we illustrate the approach and demonstrate the previous two simulation steps.

Example 1 (*A dynamic regression model with changing mean and variance parameter*)**:** We consider a first-order autoregression augmented with an exogenous explanatory variable[5]

$$y_t = \beta_0 + \varphi y_{t-1} + \beta_1 x_t + \sigma \varepsilon_t, \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad (t = 1, \ldots, T).$$

We accommodate this basic model to account for changes in the regression parameter associated with x_t and the conditional variance σ^2. We show that the approach easily allows for changes in both parameters to occur simultaneously. With existing structural break models this is not straightforward, especially not when parameters of different "types" are under consideration, such as a mean and a variance parameter. We choose a conjugate base prior distribution for $\boldsymbol{\theta} = (\beta_1, \sigma^2)'$ consisting of the normal-inverted Gamma–2 pair, that is,

$$f_0(\boldsymbol{\theta}): \quad \beta_1 \,|\, \sigma^2 \sim \mathcal{N}(b_\beta, \sigma^2 B_\beta), \quad \sigma^2 \sim \mathcal{IG}2(\nu_{\sigma^2}, S_{\sigma^2}).$$

For the other two parameters, $\boldsymbol{\omega} = (\beta_0, \varphi)'$, assumed to be constant over time, we take independent priors, consisting of an unrestricted and a truncated univariate normal distribution, $\beta_0 \sim \mathcal{N}(b, B)$, and $\varphi \sim \mathcal{N}(a, A) \times \mathbb{I}_{\{0<\varphi<1\}}$.

To simulate from the posterior $p(\boldsymbol{\theta}^{1,T}, \boldsymbol{\omega} \,|\, \mathbf{y})$, we first condition on $\boldsymbol{\omega}$ and compute $w_t = y_t - \beta_0 - \varphi y_{t-1}$, $(t = 1, \ldots, T)$. Both the T (conditional) one-observation marginal likelihoods and posterior distributions have known form. The former are given by

$$p_0(y_t \,|\, \boldsymbol{\omega}) = \frac{1}{\sqrt{\pi}} \frac{\Gamma((\nu_{\sigma^2}+1)/2)}{\Gamma(\nu_{\sigma^2}/2)} (S_{\sigma^2}(x_t^2 B_\beta + 1))^{-\frac{1}{2}} \left(1 + \frac{(w_t - b_\beta x_t)^2}{S_{\sigma^2}(x_t^2 B_\beta + 1)}\right)^{-\frac{1}{2}(\nu_{\sigma^2}+1)}, \tag{4.8}$$

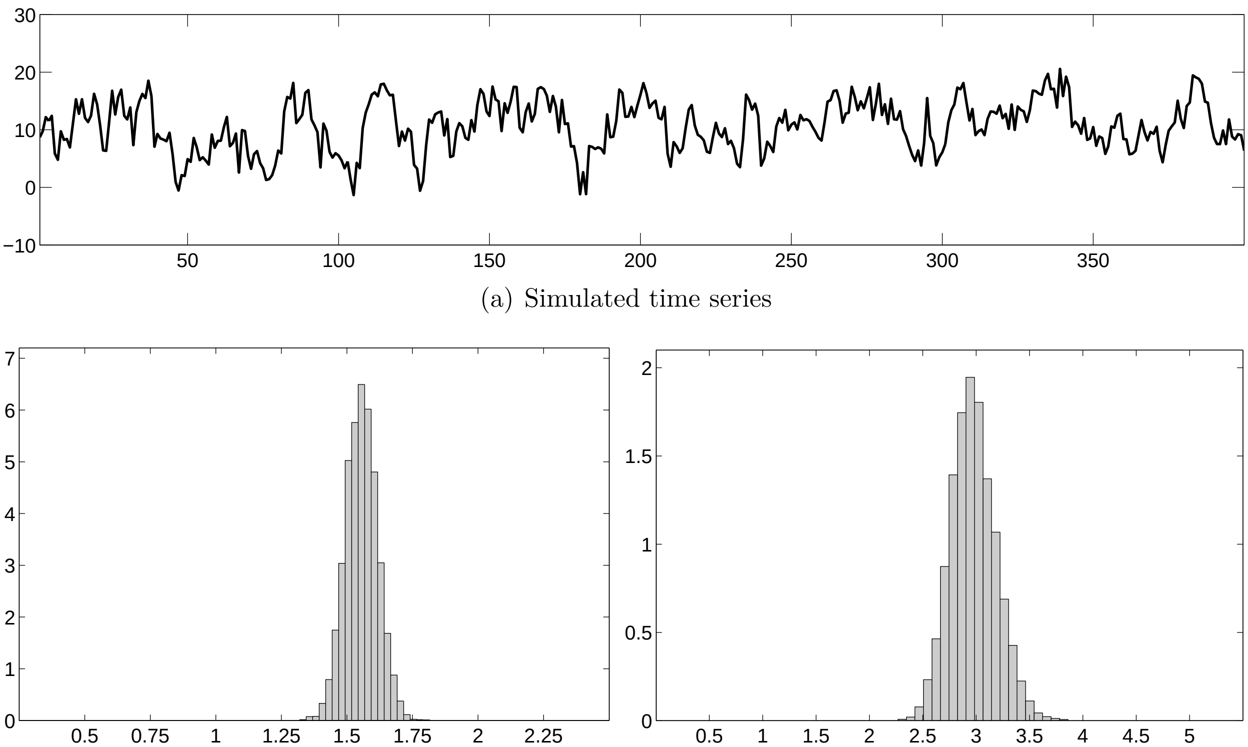
which are the densities of the Student's t distributions $\mathcal{T}(b_\beta x_t, (x_t^2 B_\beta + 1) S_{\sigma^2}/\nu_{\sigma^2}, \nu_{\sigma^2})$ evaluated in w_t. The latter, the one-observation posterior distributions, consist of the components

$$\beta_{1t} \,|\, \{y_t, \boldsymbol{\omega}, \sigma_t^2\} \sim \mathcal{N}(\bar{b}, \sigma_t^2 B_\beta / \bar{B}), \quad \sigma_t^2 \,|\, \{y_t, \boldsymbol{\omega}\} \sim \mathcal{IG}2(\nu_{\sigma^2} + 1, \bar{S}), \tag{4.9}$$

with parameters $\bar{b} = (w_t x_t B_\beta + b_\beta)/\bar{B}$, $\bar{B} = x_t^2 B_\beta + 1$, and $\bar{S} = S_{\sigma^2} + (w_t - b_\beta x_t)^2 / \bar{B}$. These two results form the inputs to the simulation steps of Section 4.3.1, and we notice that the distributions to sample from during the remix step are just the multi-observation equivalents of (4.9).

[5]We condition on the very first observation y_0.

Figure 4.1 *Posterior results dynamic regression example, no-change model*



(a) Simulated time series

(b) Posterior density β_1 (c) Posterior density σ^2

Given $\boldsymbol{\theta}^{1,T}$ we complete the posterior simulation scheme by simulating $\boldsymbol{\omega}$ in one block from its full conditional posterior. That is, we first simulate φ from its marginal (truncated Gaussian) and then β_0 conditional on φ (Gaussian). We refer to Appendix 4.A.1 for details.

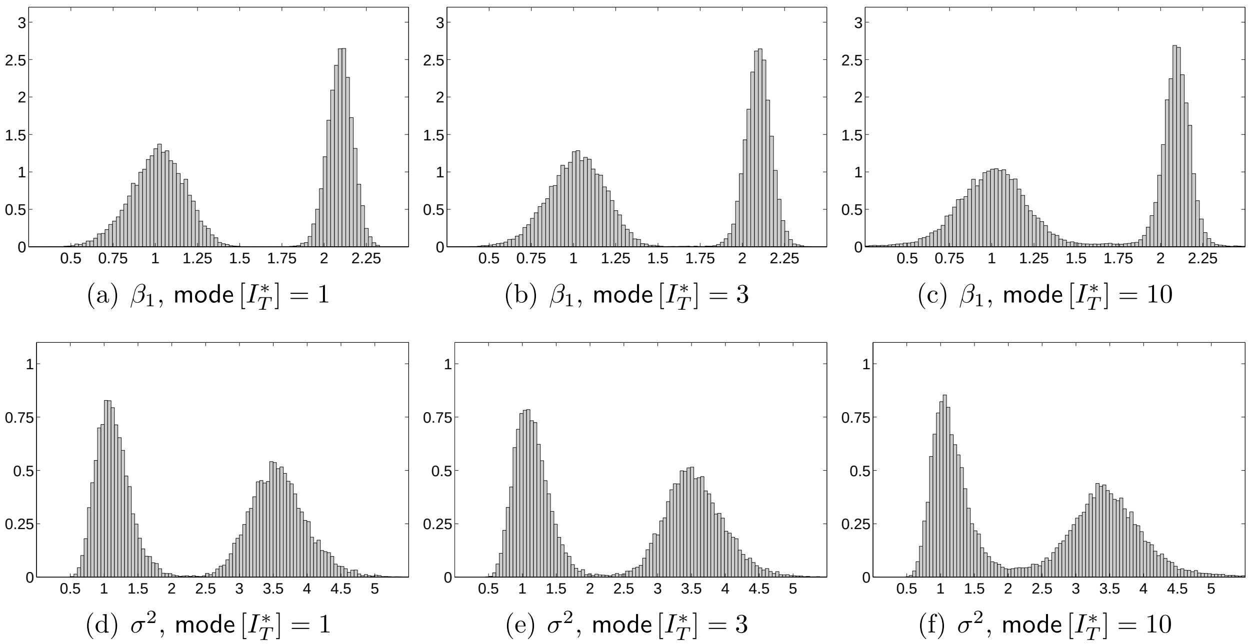
We compute a realization of the data-generating process using $\beta_0 = 2$, $\varphi = 0.8$, $x_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1.5^2)$, and we impose a change in the regression parameter and the conditional variance halfway the sample period: $\boldsymbol{\theta}_t = (2,1)'$ for $1 \leq t \leq \lfloor T/2 \rfloor$, and $\boldsymbol{\theta}_t = (1,4)'$ for $\lceil T/2 \rceil < t \leq T = 400$. The simulated time series is depicted in Figure 4.1(a), and ostensibly no parameter changes have taken place.

With the following values for the hyperparameters, $b_\beta = 0$, $B_\beta = 10$, $\nu_{\sigma^2} = 5$, $S_{\sigma^2} = 15$, and $a = b = 0$, $A = B = 10$ we compute posterior results.[6] If we totally ignore parameter changes (the special case with $\alpha = 0$), the posteriors of β_1 and σ^2 are obtained with f_0 as the regular prior. We depict the posterior densities in Figures 4.1(b)–(c), respectively. Obviously, both densities are concentrated around a value in between the respective two regime values. The ignored changes hardly have any impact on the posteriors of the time-constant parameters, since the regressors are orthogonal. We obtain $\mathsf{E}[\beta_0 \,|\, \mathbf{y}] = 2.10$, $(1.73; 2.47)$ and $\mathsf{E}[\varphi \,|\, \mathbf{y}] = 0.81$, $(0.77; 0.84)$, with the numbers in parentheses the 5th and 95th percentiles of the respective marginal posterior distributions.

We accomodate the model to incorporate parameter changes by examining the three scenarios $\alpha \in \{0.08; 0.40; 1.83\}$, each corresponding to a prior mode (1, 3 and 10) for the

[6]The implied base-prior properties are $\mathsf{E}\left[\sigma^2\right] = 15/(5-2) = 5$, $\mathsf{Var}\left[\sigma^2\right] = 5^2 \cdot 2/(5-4) = 50$, and $\mathsf{E}[\beta_1] = 0$, $\mathsf{Var}[\beta_1] = 15 \cdot 10/(5-2) = 50$, since marginally $\beta_1 \sim \mathcal{T}(b_\beta, (S_{\sigma^2}/\nu_{\sigma^2})B_\beta, \nu_{\sigma^2})$.

Figure 4.2 *Posterior densities dynamic regression example, parameter-change models*



(a) β_1, mode $[I_T^*] = 1$ (b) β_1, mode $[I_T^*] = 3$ (c) β_1, mode $[I_T^*] = 10$

(d) σ^2, mode $[I_T^*] = 1$ (e) σ^2, mode $[I_T^*] = 3$ (f) σ^2, mode $[I_T^*] = 10$

number of distinct regimes. Figures 4.2(a)–(c) show the posterior densities of β_1 and Figures 4.2(d)–(f) the posteriors of σ^2 for the three settings. For both parameters, in all three settings we obtain bi-modal distributions with posterior modes at the two regime values. When $\alpha = 1.83$ we see that the very generous prior heterogeneity creates a superfluous "connection" between the two modes. The posterior of β_0 is slightly more concentrated around the true value with $\mathsf{E}[\beta_0 \,|\, \mathbf{y}] = 2.02$, $(1.69; 2.34)$. The marginal posterior of φ is the same as in the $\alpha = 0$ case.

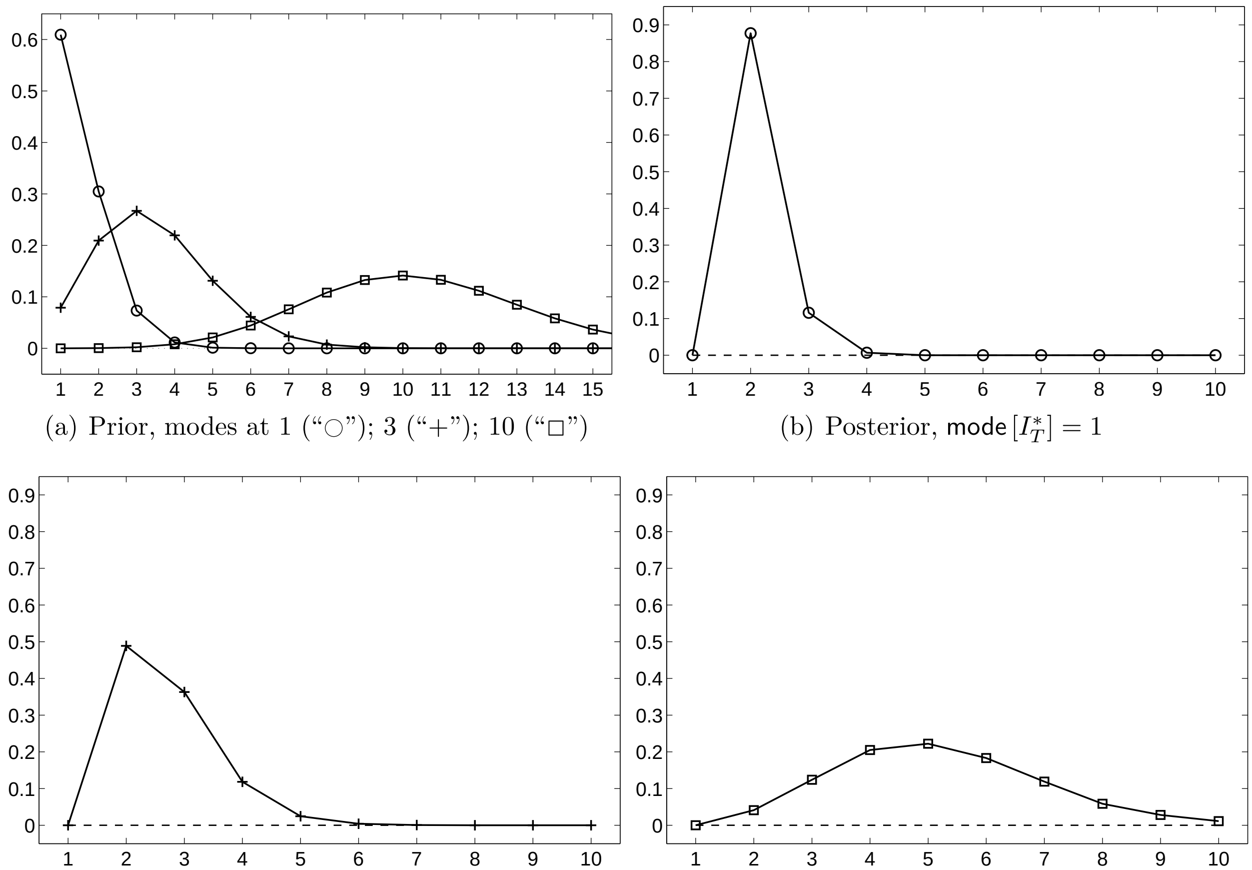
In Figure 4.3(a) we plot the prior probabilities of the number of distinct parameter values as described in Section 4.2.4. The three values of α are chosen such that the prior modes correspond to 1 ("○"), 3 ("+") and 10 ("□"). After updating with the data we have the posterior distributions of I_T^* which we show in Figures 4.3(b)–(d). In the first two cases the posterior mode is shifted 1 upwards and downwards, respectively, to 2, the true number of regimes. When we set α too extreme, we see that the data shift probability mass substantially downwards to more moderate values of I_T^*, and its mode is halved to 5. ∎

4.3.2 Smoothness parameter

If we use the prior of Escobar and West (1995) for the smoohtness parameter, we augment the MCMC procedure with a third step to simulate α. Its full conditional posterior only depends upon the categorization of the data according to their regime membership and the (implied) number of unique regimes I_T^*. Therefore, combining the "likelihood" in (4.3) with the Gamma prior of Section 4.2.4, we obtain

$$p(\alpha \,|\, \mathbf{y}, \boldsymbol{\theta}) = p(\alpha \,|\, I_T^*) \propto \alpha^{I_T^*} \frac{\Gamma(\alpha)}{\Gamma(\alpha + T)} \alpha^{a_\alpha - 1} \exp(-\alpha b_\alpha).$$

Figure 4.3 *Probabilities number of distinct parameter values, dynamic regression example*



(a) Prior, modes at 1 (“○”); 3 (“+”); 10 (“□”)

(b) Posterior, $\mathsf{mode}\,[I_T^*] = 1$

(c) Posterior, $\mathsf{mode}\,[I_T^*] = 3$

(d) Posterior, $\mathsf{mode}\,[I_T^*] = 10$

Escobar and West (1995) rewrite the ratio of gamma functions in terms of a beta function such that the target distribution's density becomes

$$p(\alpha \,|\, I_T^*) \propto \alpha^{a_\alpha + I_T^* - 2} \exp(-\alpha b_\alpha)(\alpha + T) \int_0^1 \eta^\alpha (1-\eta)^{T-1}\, \mathsf{d}\eta.$$

Thus, the target distribution is the marginal of the joint distribution of $\{\alpha, \eta\}$. With this demarginalization Escobar and West (1995) suggest a Gibbs step to simulate from this joint distribution by successively sampling from the two full conditionals $p(\eta \,|\, I_T^*, \alpha)$ and $p(\alpha \,|\, I_T^*, \eta)$. The former gives $\eta \,|\, \{I_T^*, \alpha\} \sim \mathcal{B}e(\alpha + 1, T)$, and the latter is a mixture of two Gamma distributions, *i.e.*, a $\mathcal{G}a(a_\alpha + I_T^*, b_\alpha - \log \eta)$ and a $\mathcal{G}a(a_\alpha + I_T^* - 1, b_\alpha - \log \eta)$, with the weight of the first being proportional to $a_\alpha + I_T^* - 1$ and that of the second to $T(b_\alpha - \log \eta)$.

4.3.3 Base prior parameters

In case of a prior on the parameter of the base distribution, we sample λ by extending the simulation scheme with a fourth step. Conditional on the sampled values for the parameters as obtained from Steps 1 and 2, we form the vector $\boldsymbol{\theta}^*$ containing the I_T^* unique values θ_i^*. According to the model specification in (4.1), these values are independent realizations from the base distribution. To update λ we therefore sample from the distribution with density

$$p(\lambda \,|\, \mathbf{y}, \boldsymbol{\theta}^*) \propto p(\lambda) \prod_{i=1}^{I_T^*} f_0(\theta_i^*; \lambda).$$

A prior $p(\lambda)$ that forms a non-conjugate pair with f_0 requires some tailoring to implement this simulation step, *e.g.* by employing a slice or a Metropolis–Hastings sampler. We refer to Section 4.4 for applications.

4.4 Macroeconomic applications

In this section we demonstrate the practical use of the approach by presenting three empirical macroeconomic applications. As we want to highlight the general applicability of the approach to different types of models, the illustrations involve a copula model, a predictive Poisson regression, and a Markovian regime-switching model with changes in the variance parameter. These three examples will touch on issues relevant with respect to the modeling process and posterior simulation, including prior specification and the computational steps of the approach in nonlinear models.

4.4.1 Clayton copula model for employment growth

To demonstrate the methods of Section 4.3.1 for non-conjugate settings we examine a copula application. The copula approach is becoming increasingly more popular, especially in empirical finance, to capture non-standard cross-sectional dependence (see, for example, McNeil *et al.*, 2005, Jondeau and Rockinger, 2006), or time-series dependence as in Chen and Fan (2006b). We adopt the copula-based approach of the latter and apply it to the monthly growth rate of the number of employees on nonagricultural payrolls (EMP), one of The Conference Board's coincident indicators.

Based on previous research (for example, Hamilton, 1989) we expect to find asymmetry in the statistical behavior of EMP. More in particular, we suspect a higher degree of persistence in the time series when in the lower quantiles of its marginal distribution. When economic conditions are adverse, they are so for a number of consecutive months, whereas when they are favorable, the series tends to fluctuate relatively more heavily in the upper quantiles. A Clayton copula function covers this type of dependence and we apply it to model the annualized monthly net log-growth rates $y_t = 1200(\log Y_t - \log Y_{t-1})$, for the period January 1960–December 2010, with Y_t the number of employees.

Methodology

We assume the EMP series to be first-order Markovian and its temporal dependence to be described by a bivariate Clayton copula function given by $C^{\mathrm{Cl}}(\mathbf{u}_t; \theta) = (u_t^{-\theta} + u_{t-1}^{-\theta} - 1)^{-1/\theta}$, with $\theta > 0$ and $\mathbf{u}_t = (u_t, u_{t-1})'$, (we condition on y_0, the growth rate in December 1959). The cumulative-distribution-function (CDF) transforms are marginally $u_t \sim \mathcal{U}([0,1])$, ($t = 0, 1, \ldots, T$), and we use the rescaled empirical CDF to compute the realizations.[7]

[7] That is, we do not make any assumptions about the form of the marginal distribution of y_t and simply compute the set of $T+1$ empirical-CDF transforms $u_t = \frac{1}{T+2} \sum_{s=0}^{T} \mathbb{I}_{\{y_s \leq y_t\}}$, such that $0 < u_t < 1$.

This copula model has the following characteristics. The parameter θ determines the strength of the temporal dependence between u_t and u_{t-1}, with higher values indicating stronger dependence. For example, Kendall's rank correlation, a measure of dependence that only depends on the copula function, is equal to $\theta/(\theta+2)$. Furthermore, the Clayton copula allows for consecutive "extreme" negative events, but independent positive events, which is formalized by lower tail dependence and upper tail independence, in the sense that

$$\lim_{q\downarrow 0} \mathsf{Pr}\left[u_t \le q \,|\, u_{t-1} \le q;\, \theta\right] = \lim_{q\downarrow 0} C^{\mathrm{Cl}}(q\boldsymbol{\iota}_2;\theta)/q = 2^{-1/\theta},$$
$$\lim_{q\uparrow 1} \mathsf{Pr}\left[u_t \ge q \,|\, u_{t-1} \ge q;\, \theta\right] = \lim_{q\uparrow 1} \Big[1-2q+C^{\mathrm{Cl}}(q\boldsymbol{\iota}_2;\theta)\Big]/(1-q) = 0.$$

To account for parameter changes in the Clayton copula's parameter we apply the model specification of Section 4.2 in which we use the following lognormal distribution as base prior,

$$f_0(\theta; a_\theta, A_\theta): \quad \theta \sim \mathcal{LN}(a_\theta, A_\theta).$$

The lognormal distribution (its generalized form) has the desirable property that it can be used as a prior for any parameter which is bounded from below/above. Evidently, this base prior and the likelihood contribution $p_t(u_t;\theta) = \partial^2 C^{\mathrm{Cl}}(\mathbf{u}_t;\theta)/(\partial u_t \partial u_{t-1})$ form a non-conjugate pair, so in Step 1 of the posterior simulation we use the Metropolis–Hastings sampler to simulate from the full conditional posteriors $p(\theta_t \,|\, \mathbf{u}, \boldsymbol{\theta}_{-t})$, with $\mathbf{u} = (u_0, u_1, \ldots, u_T)'$.

For Step 2 we implement a slice sampler to update each of the I_T^* distinct parameter values obtained in Step 1. For this slice-sampling step we take into account that $\theta > 0$, apply the transformation $\vartheta = \log(\theta)$ and, according to (4.6)–(4.7), we simulate each unrestricted ϑ_i^* from the distribution with pdf

$$p(\vartheta_i^* \,|\, \mathbf{u}_{(i)}) \propto \exp\left\{-(\vartheta_i^* - a_\theta)^2/(2A_\theta)\right\} \Big(\exp(\vartheta_i^*)+1\Big)^{|T_{(i)}|}$$
$$\times \prod_{t\in T_{(i)}} (u_t u_{t-1})^{-(\exp(\vartheta_i^*)+1)} \Big(u_t^{-\exp(\vartheta_i^*)} + u_{t-1}^{-\exp(\vartheta_i^*)} - 1\Big)^{-(\exp(-\vartheta_i^*)+2)}.$$

We note that the Jacobian term $\exp(\vartheta_i^*)$ cancels out against the first term of the base-prior lognormal density.

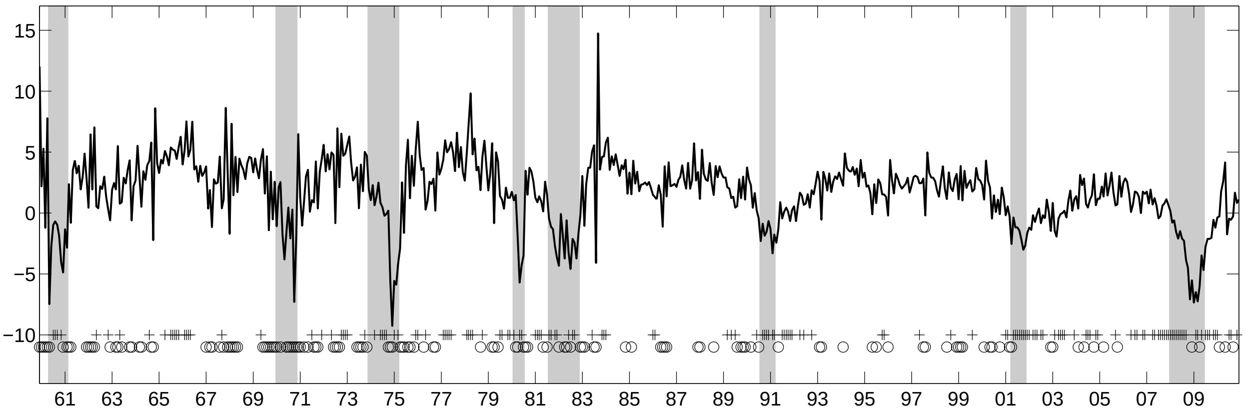
We follow Section 4.2.4 to specify the prior for the smoothness parameter α. We take the parameters of its Gamma distribution such that the prior mode of the number of distinct parameter values is 1, but we inflate the scale parameter to make the prior less informative, *i.e.*, we set $a_\alpha = 0.07 \times 25$ and $b_\alpha = 25$. Taking a small prior mode makes posterior determination of the number of regimes easier, because in Example 1 in Section 4.3.1 we saw that although the posterior number of regimes decreased considerably if we used a large value for α, a few prior-induced "irrelevant" ones remained.

Our base prior distribution requires the specification of the two parameters $\boldsymbol{\lambda} = \{a_\theta, A_\theta\}$. To learn about their values we apply the two independent priors

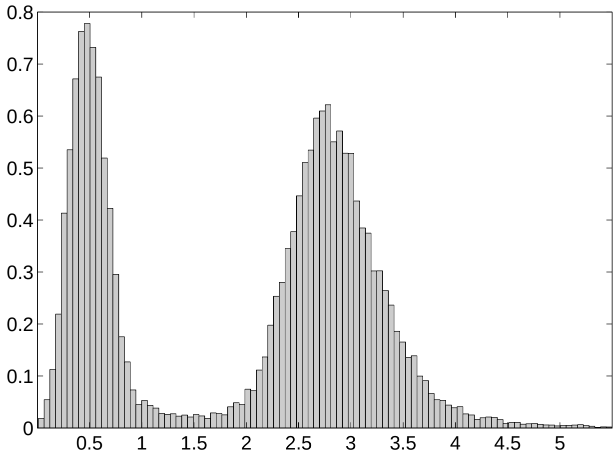
$$a_\theta \sim \mathcal{N}(b_0, B_0), \quad A_\theta \sim \mathcal{IG}2(\nu_0, S_0),$$

which are both conditionally conjugate, since $\log(\theta) \,|\, \boldsymbol{\lambda} \sim \mathcal{N}(a_\theta, A_\theta)$. We choose to set the hyperparameters relatively uninformative and we use prior simulation to set $b_0 = 1$, $B_0 = 1$,

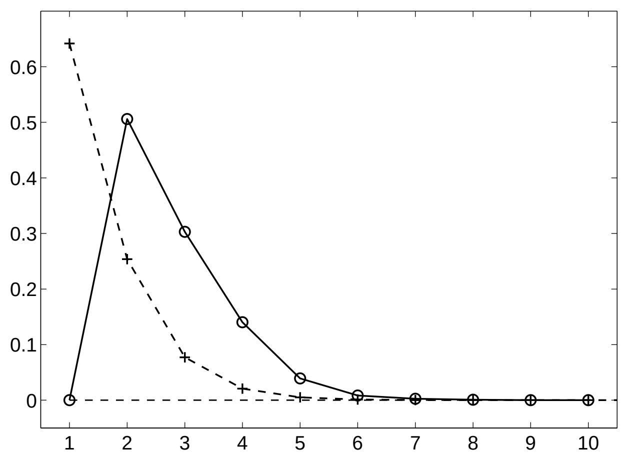
Figure 4.4 *Posterior results Clayton copula model for employment growth*



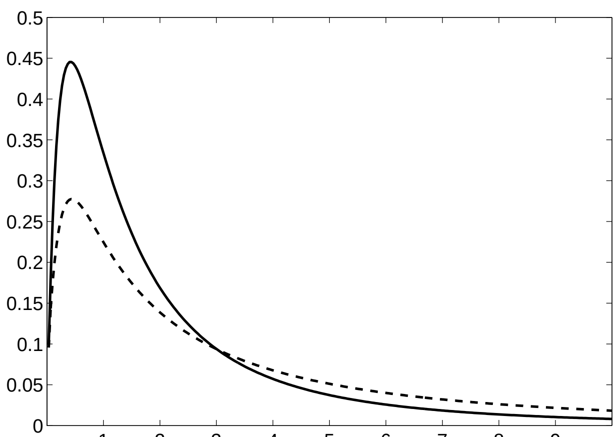
(a) Monthly employment growth rate, December 1959–December 2010



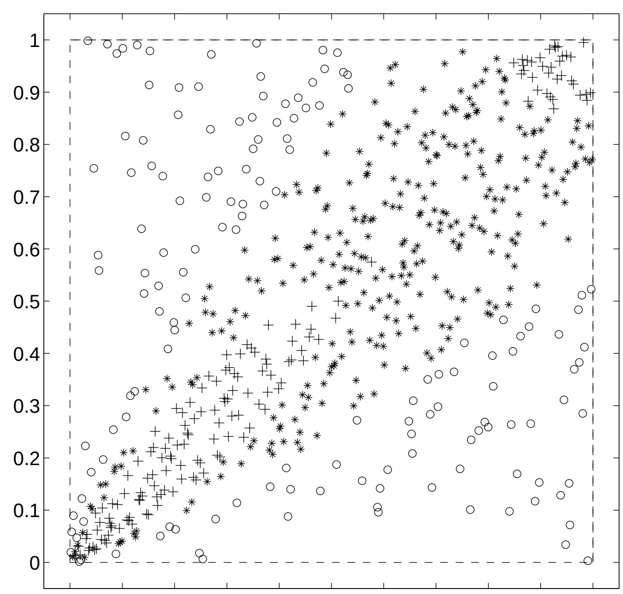
(b) Posterior density θ



(c) Probabilities I_T^*, prior (“+”); posterior (“○”)



(d) Base density θ, prior (dashed); posterior (solid)



(e) Scatter empirical-CDF transforms (u_{t-1}, u_t)

$\nu_0 = 12$, and $S_0 = 10$. In each iteration of the MCMC procedure we simulate

$$a_\theta \,|\, \{\mathbf{u}, \boldsymbol{\theta}^*, A_\theta\} \sim \mathcal{N}\Big(\Big[B_0 \textstyle\sum_i \log(\theta_i^*) + A_\theta b_0\Big]/(B_0 I_T^* + A_\theta),\ B_0 A_\theta/(B_0 I_T^* + A_\theta)\Big),$$
$$A_\theta \,|\, \{\mathbf{u}, \boldsymbol{\theta}^*, a_\theta\} \sim \mathcal{IG}2\Big(\nu_0 + I_T^*,\ S_0 + \textstyle\sum_i (\log(\theta_i^*) - a_\theta)^2\Big),$$

that is, we successively sample both parameters from their respective full conditional posterior distributions as in Section 4.3.3.

Results

We run the MCMC simulation scheme for 50,000 iterations and establish convergence of the chain after a burn-in period of 5,000 runs. With the draws after the burn-in period we obtain the following posterior results. In Figure 4.4(b) we show the marginal posterior density of the copula parameter $p(\theta_{T+1} \,|\, \mathbf{u})$, which shows clear bimodality. One regime imposing strong dependence between consecutive observations with a parameter value close to 3, and hence a Kendall's rank correlation of about 0.6. The other regime is, with a parameter value of 0.5, much closer to the independence copula. These two dominant regimes also show up when we look at the posterior distribution of the number of unique components, depicted in Figure 4.4(c). Although the prior assigns most mass to just one component, the posterior places the mode at 2 with probability 0.5. The presence of the strong-dependence regime is also reflected in the updated base distribution that generates new parameter values. In Figure 4.4(d) we show the prior (obtained by Monte Carlo) and the posterior base density. The prior's right tail is considerably deflated in favor of posterior values in the range $[0, 3]$. Thus, future new parameter values are less likely to be greater than 3.

In Figure 4.4(a) we plot the EMP time series and we depict whether an observation is assigned to the strong-dependence ("+") or the weak-dependence ("○") regime according to the posterior mean of θ_t. Periods containing recessionary months (shaded areas) with decreasing employment growth rates are associated with increased temporal dependence in the series. If we further look at the scatter of the empirical-CDF transforms in Figure 4.4(e) we evidently observe asymmetric temporal dependence because the transforms are more concentrated in the lower quantiles than in the upper. A finding that we also visually extract from the plot of the time series. When y_t drops negative, it remains so for some time, while being positive, it traverses the positive range more randomly. In addition to the two obvious regimes, we also observe the effects of the "nonparametric" mechanism of the model specification. The pairs $\mathbf{u}_t$ that are about equally likely under both regimes ("✱") are alternately assigned to either one, which is a well-known property of a mixture model.

4.4.2 Predictive Poisson regression for weeks of unemployment

In this application we investigate possible parameter changes in a count data model, that is, a model for a time series that is positive integer-valued. We demonstrate the approach with these limited dependent variables and show how to model a joint change in multiple parameters. We use a Poisson regression model to describe the average number of weeks of unemployment (MDU) for the period January 1949–December 2013. Valletta (1998) discusses the importance of MDU for, for instance, labor market interventions and policy.

We formulate a predictive model for MDU by relating it to an indicator reflecting business cycle fluctuations. Based on previous research by, *e.g.*, Baker (1992) and Verho (2008) we know that the unemployment rate leads MDU and therefore we use it as predictor.[8] We use its twelve-month moving average, first, to obtain a smoother estimate of business cycle fluctuations, and second, to partly account for its unknown lead time. Finally, the thus imposed persistence in the explanatory variable is helpful to describe the dynamics in the number of weeks unemployed.

Unlike Mukoyama and Şahin (2009), who look into economic-theoretic explanations for changed charachteristics of MDU, we examine the case from a forecasting perpective and accommodate an empirical model to parameter changes to generate more accurate forecasts. The MDU time series is published to one decimal, but it only takes relatively small values, and, hence, a Poisson regression for the series rounded to the nearest integer suits better than a standard linear regression model.

Methodology

After defining y_t as the number of weeks of unemployment, and u_t as the unemployment rate in month t, we formulate the following predictive Poisson regression[9] such that we can produce one-step-ahead forecasts,

$$y_t \,|\, \psi_t \overset{ind.}{\sim} \mathcal{P}oi(\psi_t), \quad \log \psi_t = \beta_0 + \beta_1 \sum\nolimits_{j=1}^{12} u_{t-j}/12, \quad (t = 1, \ldots, T).$$

As an increase in β_1 changes the unconditional mean of MDU, we also allow the intercept to change to absorb the implied level shift. We therefore implement the parameter-change approach with potential joint changes in both regression parameters. The two model parameters are unrestricted and the natural choice for the base prior is the bivariate normal distribution

$$f_0(\boldsymbol{\beta}; \mathbf{b}_\beta, \mathbf{B}_\beta): \quad \boldsymbol{\beta} = (\beta_0, \beta_1)' \sim \mathcal{N}(\mathbf{b}_\beta, \mathbf{B}_\beta). \tag{4.10}$$

With a conditionally conjugate layer for its parameters $\boldsymbol{\lambda} = \{\mathbf{b}_\beta, \mathbf{B}_\beta\}$, consisting of a conditional bivariate normal distribution for the base-prior mean, and an inverted Wishart distribution for the base-prior variance matrix,

$$\mathbf{b}_\beta \,|\, \mathbf{B}_\beta \sim \mathcal{N}(\mathbf{b}_0, B_0 \mathbf{B}_\beta), \quad \mathbf{B}_\beta \sim \mathcal{IW}(\nu_0, \mathbf{S}_0), \tag{4.11}$$

we straightforwardly simulate $\boldsymbol{\lambda}$ during the MCMC routine. Considering the log-linear relation between y_t and its regressors, we use prior Monte Carlo simulation to set $\mathbf{b}_0 = (1, 1)'$, $B_0 = 1$, $\nu_0 = 12$, and $\mathbf{S}_0 = 5\mathbf{I}_2$.

The data cover a large time span (780 observations) which makes it more plausible that parameter changes have taken place. We take a $\mathcal{G}a(0.36 \times 15, 15)$ prior for the smoothness parameter such that the implied marginal prior distribution for the number of regimes is substantially dispersed about 3, with $\mathsf{E}\,[I_T^*] = 3.45$ and $\mathsf{Var}\,[I_T^*] = 3.30$.

[8] We use the following data series. MDU: Average (mean) duration of unemployment: Seasonally adjusted; UR: Civilian unemployment rate: Seasonally adjusted: Percent: The unemployment rate represents the number of unemployed as a percentage of the labor force. Data are taken from the Federal Reserve Bank of St. Louis' FRED2 database: `http://research.stlouisfed.org/fred2/categories/12`, and were originally collected by the Bureau of Labor Statistics of the U.S. Department of Labor by means of monthly household surveys.

[9] Overdispersion of the data relative to the model-implied variance, a feature common to many empirical Poisson regression models, is no issue here.

Since observation t's likelihood contribution $p_t(y_t; \boldsymbol{\beta}) = \exp\{y_t \mathbf{x}_t{}'\boldsymbol{\beta} - \exp(\mathbf{x}_t{}'\boldsymbol{\beta})\}/y_t!$ forms a non-conjugate pair with the normal base prior in (4.10), we rely on the MH method described in Section 4.3.1 to sample the time-dependent parameters $\boldsymbol{\beta}_t$. For the remix step we use slice sampling and successively sample $\beta^*_{0,i}$ and $\beta^*_{1,i}$, $(i = 1, \ldots, I^*_T)$, from their full conditional posteriors $p(\beta^*_{0,i} \,|\, \mathbf{y}_{(i)}, \boldsymbol{\lambda}, \beta^*_{1,i})$ and $p(\beta^*_{1,i} \,|\, \mathbf{y}_{(i)}, \boldsymbol{\lambda}, \beta^*_{0,i})$, respectively. Both these densities are proportional to

$$p(\boldsymbol{\beta}^*_i \,|\, \mathbf{y}_{(i)}, \boldsymbol{\lambda}) \propto \exp\left\{-(\boldsymbol{\beta}^*_i - \mathbf{b}_\beta)'\mathbf{B}_\beta^{-1}(\boldsymbol{\beta}^*_i - \mathbf{b}_\beta)/2 + \sum\nolimits_{t\in T_{(i)}} \left[y_t \mathbf{x}_t{}'\boldsymbol{\beta}^*_i - \exp(\mathbf{x}_t{}'\boldsymbol{\beta}^*_i)\right]\right\},$$

with the other parameter held fixed at its most recent update.

We sample the smoothness parameter as in Escobar and West (1995). Considering $\boldsymbol{\beta}^*_i$, $(i = 1, \ldots, I^*_T)$, as the "observations," we simulate $\boldsymbol{\lambda}$ in one block as in a conjugate multivariate normal regression model (see Appendix 4.A.2 for details of this step).

Results

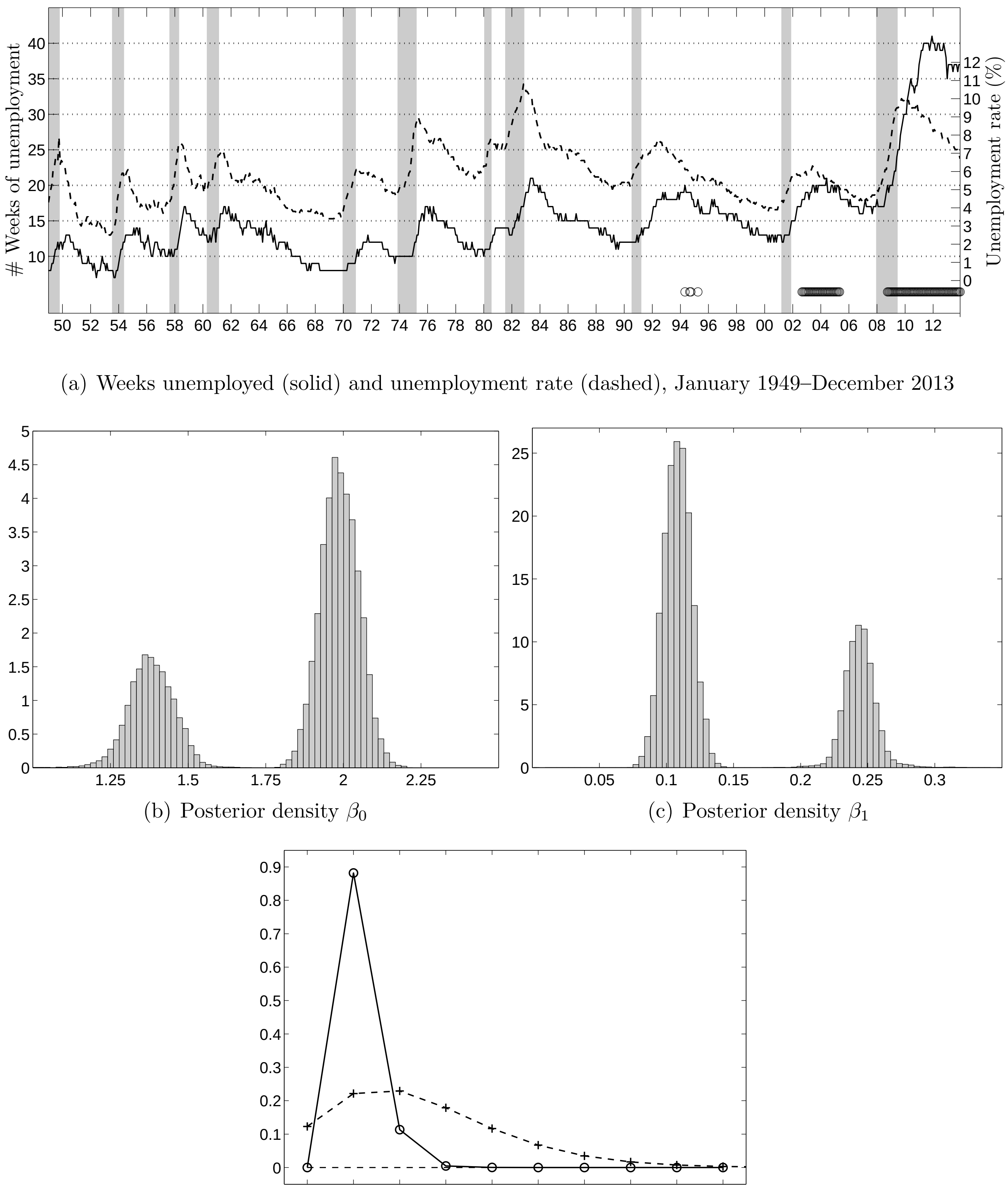
We warm up the Markov chain with 5,000 iterations to attain convergence, and use the next 50,000 draws as a dependent sample from the posterior $p(\boldsymbol{\beta}^{1,T}, \boldsymbol{\lambda}, \alpha \,|\, \mathbf{y})$. In Figure 4.5(b) we display the full-sample posterior density of the intercept β_0, and in Figure 4.5(c) we show the posterior of the effect of the unemployment rate on MDU, β_1. For both parameters two obvious regimes appear. This outcome is supported by the posterior probabilities for the number of distinct parameter values, which we plot in Figure 4.5(d) ("$\bigcirc$": posterior; "$+$": prior) and in which we observe an almost degenerate distribution for I^*_T at 2.

Plotting the time series gives us some more insights. Figure 4.5(a) shows the monthly series of the discrete dependent variable y_t (solid line) and the unemployment rate (dashed line). They clearly co-move with the unemployment rate leading MDU, and both are sluggishly lagging the business cycle (shaded areas reflect NBER-announced recessionary periods). However, since the two latest recessions the distance between the two series has changed while it has been relatively constant since the start of the sample in 1949 (see also Mukoyama and Şahin, 2009, who use Andrew's testing method to identify structural change). After the severe 2008 recession, the number of weeks of unemployment rose to an all-time high, and stayed there during 2011–13. Whereas, although being high, the unemployment rate did not display similar behavior, and its post-recession level is comparable to that of the early-1980s economic crisis.

The visual change in the relation between y_t and u_t is reflected in the increased responsiveness of MDU to the lagged unemployment rate. We find a β_1 which doubles from 0.11 to 0.24, and we label periods associated with this higher responsiveness by "$\bigcirc$" in Figure 4.5(a). In turn, the larger β_1 is accompanied with a smaller value for the intercept. We find two modes for β_0, one at 2 and the other at 1.35.

In order to check the effect of allowing for changes in the model parameters, we employ an out-of-sample forecasting experiment and compute the predictive likelihood $p(\mathbf{y}^{\tau+1,T} \,|\, \mathbf{y}^{1,\tau})$. We compare it to the forecasting results of a model with time-constant parameters, which is the nested variant obtained with a smoothness parameter $\alpha = 0$. Considering the interpretation of the predictive likelihood as the marginal likelihood of $\mathbf{y}^{\tau+1,T}$ under "prior" $p(\boldsymbol{\beta}^{1,\tau}, \boldsymbol{\lambda}, \alpha \,|\, \mathbf{y}^{1,\tau})$, we use the resulting predictive Bayes factors for model comparison (see Greenberg, 2008, Table 3.1).

Figure 4.5 *Posterior results Poisson model for weeks of unemployment*



(a) Weeks unemployed (solid) and unemployment rate (dashed), January 1949–December 2013

(b) Posterior density β_0

(c) Posterior density β_1

(d) Probabilities I_T^*, prior ("+"); posterior ("○")

In Table 4.1 we report the full-sample posterior properties of the time-constant parameters in Panel B. Both Poisson regression parameters are in between the two regime values which we found in the parameter-change specification. The predictive Bayes factor based

Table 4.1 *Posterior and forecasting results models for weeks of unemployment*

	Posterior				Forecasting	
			Percentiles		Predictive likelihood	
Parameter	Mean	St.D.	5th	95th	Jan 87–Dec 13	Jan 00–Dec 13
Panel A: Parameter-change model						
α	0.30	0.12	0.14	0.52	−500.6	−325.0
Panel B: No-change model						
β_0	1.68	0.03	1.63	1.74	−504.8	−329.3
β_1	0.17	0.01	0.16	0.18		

on the hold-out sample January 1987–December 2013 has a $\log_{10}$-value of 4.8, decisively favoring the model with parameter changes. To get a better understanding of the changed behavior of MDU, we also compute the predictive likelihoods for the final fourteen years which we show in the last column of Table 4.1. Then we find a Bayes factor of 4.3 which indicates that for the first thirteen years of the hold-out sample the parameter-change model is only weakly to substantially favored. Concluding, the decisive support is gathered in the latter part of the sample during which MDU reveals the atypical high-level behavior [Figure 4.5(a)], which demonstrates the efficacy of the approach to harnass an empirical model to parameter changes.

4.4.3 Forecasting U.S. quarterly GDP growth

With this final application we illustrate the parameter-change approach when forecasting the growth rate of U.S. gross domestic product (GDP). This economic series is extensively examined in the empirical literature and one of the main findings is the drop in its volatility during the 1980s, often referred to as the Great Moderation. See, amongst others, the studies of McConnell and Perez-Quiros (2000), Kim *et al.* (2008) and Geweke and Jiang (2011).

This macroeconomic application shows how our approach holds when we suspect that a single sudden structural change has taken place. Moreover, we demonstrate how the approach is easily applied to models proven useful for time series analysis. We model the quarterly growth rates with a Markovian switching mean and we implement the parameter-change approach to account for changes in GDP's conditional variance. We particularly focus on the consequences for out-of-sample forecasting.

Methodology

If Y_t represents real GDP in quarter t, then we define $y_t = 400(\log Y_t - \log Y_{t-1})$ as the annualized quarterly log-growth rate for the sample period 1948 Q1–2013 Q4. We follow Hamilton's (1989) seminal hidden-Markov approach with different mean-growth-rate regimes and the switching between them is directed by a latent homogenous first-order Markov chain $\{s_t\}$. In addition to Hamilton's (1989) two regimes, Boldin (1996) and Clements and Krolzig (2003) suggest a third, bounce-back regime to account for the peaks in growth rates immediately after a recession to relatively rapidly arrive back at the pre-recession level.[10] The first step of the modeling process results in

$$y_t = \beta_1 \mathbb{I}_{\{s_t=1\}} + \beta_2 \mathbb{I}_{\{s_t=2\}} + \beta_3 \mathbb{I}_{\{s_t=3\}} + \sigma\varepsilon_t, \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad (t = 1,\ldots,T), \qquad (4.12)$$

in which β_j is the mean growth rate of y_t in regime j. The unobserved process $\{s_t\}$ is first-order Markovian with transition-probabilities matrix $\mathbf{P} = (\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3)'$, $\mathbf{P}_i = (p_{i1}, p_{i2}, p_{i3})'$, and invariant distribution $\boldsymbol{\pi}(\mathbf{P})$, such that

$$\mathsf{Pr}\,[s_t = j \,|\, s_{t-1} = i, \mathbf{P}] = p_{ij}, \quad (i,j = 1,2,3), \quad (t = 2,\ldots,T), \qquad (4.13)$$

and the initialization of the process is given by $\mathsf{Pr}\,[s_1 = j \,|\, \mathbf{P}] = \pi_j(\mathbf{P})$, $(j = 1,2,3)$.

To complete the conditional-mean part of the model we specify the following priors. We choose the truncated trivariate normal distibution

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)' \sim \mathcal{N}(\mathbf{b}, \mathbf{B}) \times \mathbb{I}_{\{\beta_1 < \beta_2 < \beta_3\}},$$

for the three mean growth rates. The truncation serves to circumvent the label-switching problem. For interpretation, this prior makes that $s_t = 1$ is associated with a recessionary month, $s_t = 2$ with an expansion, and $s_t = 3$ with the highest mean growth rate and is therefore interpreted as the bounce-back state. We use the three independent Dirichlet distributions

$$\mathbf{P}_i \overset{ind.}{\sim} \mathcal{D}ir(a_{i1}, a_{i2}, a_{i3}), \quad (i = 1,2,3),$$

as prior for the elements of the transition matrix. We establish a prior close to uninformativeness with $\mathbf{b} = \mathbf{0}_3$, $\mathbf{B} = 15\mathbf{I}_3$, and $a_{ij} = 1$ for all i,j.

With the second part of the model specification we check for support of the hypothesis that the economy has structurally become less volatile. We apply the parameter-change method to the conditional variance of the GDP growth rates. Since $\sigma^2 > 0$ and the inverted Gamma–2 distribution forms a conditionally conjugate pair with the normal distribution of the innovations, we take as base prior

$$f_0(\sigma^2; \nu_{\sigma^2}, S_{\sigma^2}): \quad \sigma^2 \sim \mathcal{IG}2(\nu_{\sigma^2}, S_{\sigma^2}). \qquad (4.14)$$

As for the copula model in Section 4.4.1, we take a prior mode for the number of distinct variances equal to one and inflate the tails of I_T^* by using the prior $\mathcal{G}a(0.08 \times 25, 25)$ distribution for the smoothness parameter. If the posterior support of I_T^* is shifted away from 1, we have evidence for changes in GDP's variance parameter.

We further introduce a new prior for the parameters of the base prior, $\boldsymbol{\lambda} = \{\nu_{\sigma^2}, S_{\sigma^2}\}$. We suggest to use a Gamma distribution for the location parameter and a discrete distribution for the number of degrees of freedom, *i.e.*,

$$S_{\sigma^2} \sim \mathcal{G}a(a_0/2, 1/(2b_0)), \quad \mathsf{Pr}\left[\nu_{\sigma^2} = \underline{\nu}_j\right] = p_{\nu,j}, \quad (j = 1,2,\ldots,J). \qquad (4.15)$$

Aided by prior simulation, we set $a_0 = 4$, $b_0 = 8$, $\underline{\nu}_j = 2j + 4$, $p_{\nu,j} \propto 1$, and $J = 8$.

[10] During the posterior simulation procedure of a two-regime model with this sample, we find a recessionary regime that occasionally jumps to the expansion mean and the expansionary regime, in turn, jumps to a high-growth level. These computational findings sustain the motivation for the choice to use a third regime.

This prior specification results in an efficient and easy-to-implement posterior simulation step for $\boldsymbol{\lambda}$. In each iteration of the MCMC simulation we have I_T^* unique conditional variance parameters $\sigma_i^{2;*}$, and we update $\boldsymbol{\lambda}$ as in Section 4.3.3. That is, with the prior in (4.15) we sample both parameters in one block, for which we use[11]

$$p(\nu_{\sigma^2} \,|\, \mathbf{y}, \boldsymbol{\sigma}^{2;*}) \propto \frac{\Gamma((a_0 + \nu_{\sigma^2} I_T^*)/2)}{\Gamma(\nu_{\sigma^2}/2)^{I_T^*}} \prod_i \left(\sigma_i^{2;*}\right)^{-(\nu_{\sigma^2}+2)/2} \left(b_0^{-1} + \sum_i \sigma_i^{-2;*}\right)^{-(a_0+\nu_{\sigma^2} I_T^*)/2} \times \sum_{j=1}^{J} \mathbb{I}_{\{\nu_{\sigma^2}=\underline{\nu}_j\}} p_{\nu,j}, \tag{4.16}$$

$$S_{\sigma^2} \,|\, \{\mathbf{y}, \boldsymbol{\sigma}^{2;*}, \nu_{\sigma^2}\} \sim \mathcal{G}a\Big((a_0 + \nu_{\sigma^2} I_T^*)/2,\ \Big(b_0^{-1} + \sum_i \sigma_i^{-2;*}\Big)/2\Big). \tag{4.17}$$

One iteration of the complete posterior simulation scheme consists of (i) conditional on the mean parameters $\boldsymbol{\beta}$ and regime indicators $\mathbf{s} = (s_1, \ldots, s_T)'$, we execute the four steps of Section 4.3, and (ii) conditional on the T conditional variances $\boldsymbol{\sigma}^2$ we sample the elements of $\boldsymbol{\beta}$ (from their full conditional posteriors); $\mathbf{s}$ (in one block with the simulation smoother of Chib, 1996); and, successively, the three rows of $\mathbf{P}$ (with MH steps as in Geweke, 2005). In Appendix 4.A.3 we provide all steps in detail.

To compare our model with variance changes, we consider two additional model specifications. The first is the mean-switching model in (4.12)–(4.13), but with time-constant variance σ^2. To make a fair comparison, we use the same prior as in the variance-change setting and also allow for updating of the prior's parameters. That is, we have $\sigma^2 \,|\, \{\nu_{\sigma^2}, S_{\sigma^2}\} \sim \mathcal{IG}2(\nu_{\sigma^2}, S_{\sigma^2})$, and $\{\nu_{\sigma^2}, S_{\sigma^2}\}$ has the prior distribution in (4.15). We note that to update the latter pair in this model, we only have a single "observation" in each run of the simulation procedure.

The second benchmark model consists of (4.12)–(4.13) plus a single structural break in the variance during the sample period, located at unknown quarter ζ. We assume the two resulting variance parameters, σ_1^2 and σ_2^2, to be *a priori* independent and each has the base prior as prior distribution: $\sigma_j^2 \,|\, \{\nu_{\sigma^2}, S_{\sigma^2}\} \overset{i.i.d.}{\sim} \mathcal{IG}2(\nu_{\sigma^2}, S_{\sigma^2})$, $(j = 1, 2)$. Again for a fair comparison, we learn about the two hyperparameters as well, and now we have two "observations" in each MCMC iteration. For the unknown location of the single variance break we take the prior distribution

$$\mathsf{Pr}\,[\zeta = t] = p_{\zeta,t}, \quad (1 < t_1 \le t \le t_2 < T),$$

in which we exclude the break to occur during the first and last two observations in the sample. We consider the remaining quarters all equally likely to contain the break by setting $p_{\zeta,t} \propto 1$. We note that we efficiently sample $\{\sigma_1^2, \sigma_2^2, \zeta\}$ in one block, since we integrate out the variance parameters analytically (see Appendix 4.A.3).

For all three models we employ a forecasting exercise to compute the one-step-ahead predictive densities $p(y_t \,|\, \mathbf{y}^{1,t-1})$, for the hold-out period $t = \tau+1, \ldots, T$. In order to decrease the computational burden we do not re-run the entire posterior simulation routine for each one-observation-expanded period, but use importance sampling techniques instead. We only re-run the posterior sampler after each additional three years (= twelve observations) and correct for not using the exact posterior distribution by weighing the draws. Appendix 4.A.3 contains a detailed description of our forecasting methods.

[11] Analytically we integrate out S_{σ^2} and compute the marginal support for all (discrete) values $\underline{\nu}_j$. For derivations to arrive at this step we refer to Appendix 4.A.3.

Table 4.2 *Posterior results transition matrix GDP growth rate model*

State s_{t-1}	State s_t 1		2		3	
1	0.56	(0.32; 0.78)	0.27	(0.04; 0.57)	0.16	(0.02; 0.33)
2	0.07	(0.03; 0.11)	0.87	(0.78; 0.94)	0.07	(0.01; 0.13)
3	0.06	(0.00; 0.17)	0.29	(0.13; 0.45)	0.65	(0.50; 0.79)

Notes: The table reports the posterior means of the transition probabilities, and in parentheses the 5th and 95th percentiles of the respective marginal distributions. Due to rounding, the posterior means in the first two rows do not sum columnwise to exactly 1.

Table 4.3 *Posterior and forecasting results models for GDP growth rate*

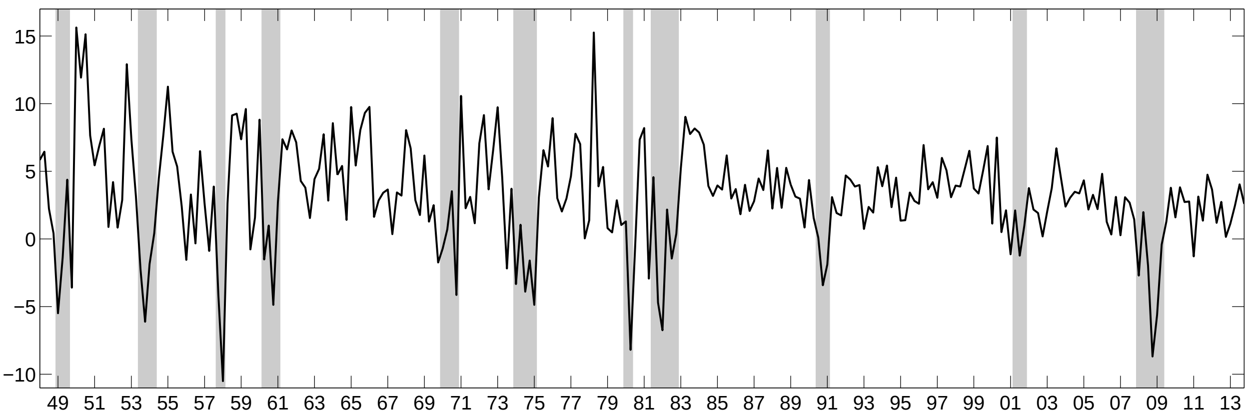
	Posterior				Forecasting	
			Percentiles		Predictive likelihood	
Parameter	Mean	St.D.	5th	95th	95 Q1–13 Q4	00 Q1–13 Q4
Panel A: Variance-change model						
β_1	−2.42	1.20	−4.29	−0.35	−78.8	−59.0
β_2	2.93	0.26	2.53	3.37		
β_3	7.56	0.54	6.71	8.39		
Panel B: No-change model						
σ^2	6.50	1.04	4.94	8.33	−79.7	−59.7
Panel C: One-break model						
σ_1^2	18.06	2.73	13.87	22.76	−76.3	−58.9
σ_2^2	3.25	0.57	2.42	4.27		

Results

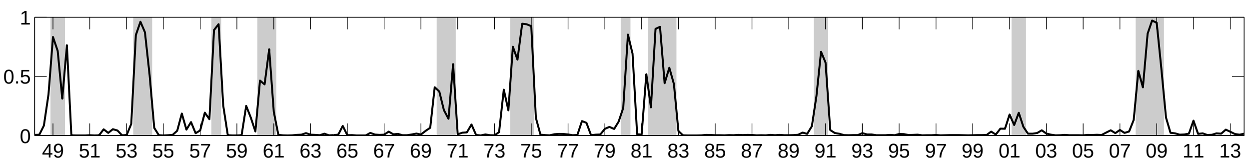
First we look at the full-sample posterior results which we compute with 50,000 runs of the MCMC routine after a warm-up batch consisting of the first 5,000 runs. In Figure 4.6(a) we plot the GDP time series and in Figures 4.6(b)–(d) we show the smoothed mean-growth-rate regime probabilites $\mathsf{Pr}\,[s_t = j \,|\, \mathbf{y}]$, $(t = 1, \ldots, T)$. We find the absence of a bounce-back effect for the three latest recessions, whereas the effect is prominently present directly after the pre-1990 recessions. We report posterior results for the transition matrix in Table 4.2, from which we derive that if GDP is in its bounce-back state, it either was in that state too during the previous quarter, or it comes out of the recessionary regime. In Panel A of Table 4.3 we see the more than doubled mean growth rate in the bounce-back regime compared to the expansion's.

In Figure 4.7(a) we depict the full-sample posterior of the variance parameter in the parameter-change model. We obviously find support for different variances, with support concentrated around the two modes of 3 and 10. We, therefore, find quarters with a standard deviation of GDP's growth rate that is more than twice as large as in the low-volatility

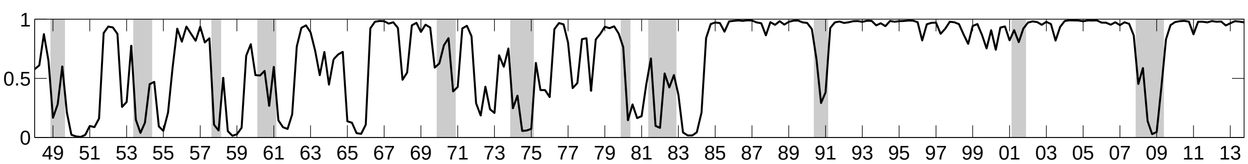
Figure 4.6 *Posterior results regime-switching mean GDP growth rate model*



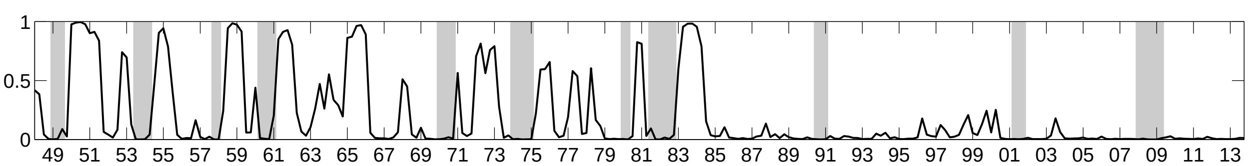
(a) Quarterly GDP growth rate, 1948 Q1–2013 Q4



(b) Smoothed probability recession regime



(c) Smoothed probability expansion regime

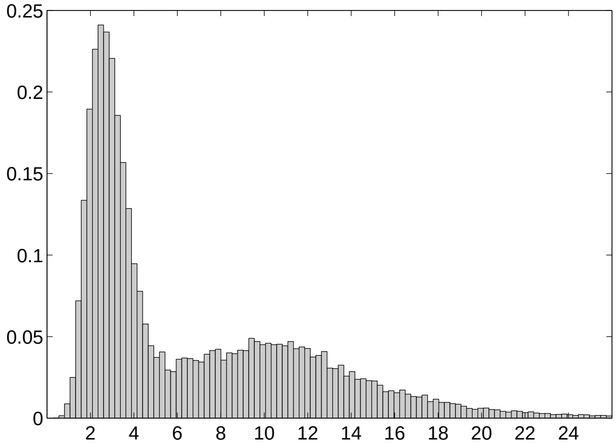


(d) Smoothed probability bounce-back regime

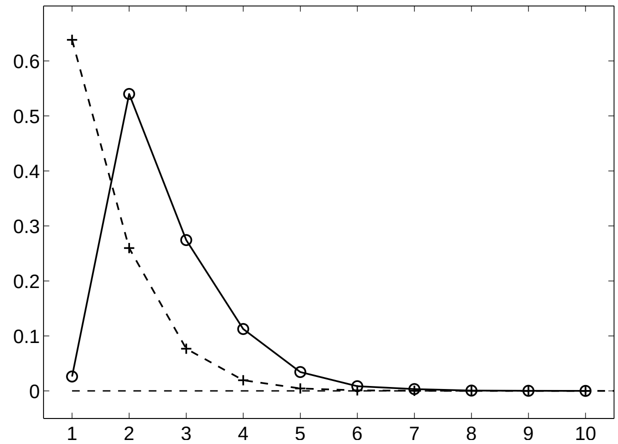
periods. Figure 4.7(b) shows the posterior of the number of unique variance parameter values, which is in line with this outcome. After updating with the data, the mode has changed from 1 ("+": prior) to 2 ("○": posterior), with considerable posterior mass at 3. The found segregation of the data according to the changed volatility sustains the Great Moderation hypothesis.

The posterior base distribution with density $\int f_0(\sigma^2; \boldsymbol{\lambda}) p(\boldsymbol{\lambda} \,|\, \mathbf{y})\, \mathsf{d}\boldsymbol{\lambda}$ is depicted (solid line) in Figure 4.7(c). It shows that the prior (dashed line) supports low values too much, since the posterior shifts to the right substantially, which is caused by the higher volatility during the first part of the sample.

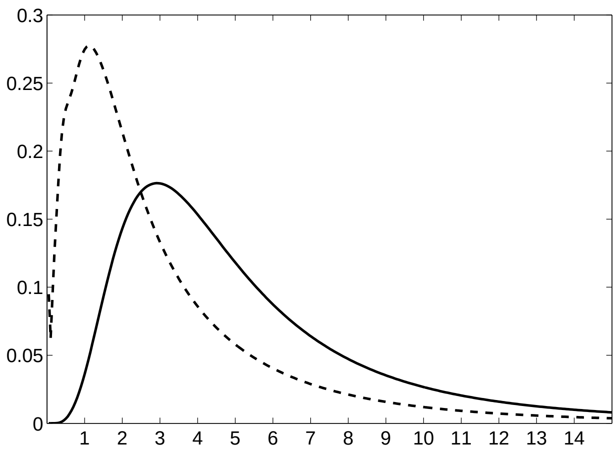
Figure 4.7 *Posterior results models for GDP growth rate with variance changes*



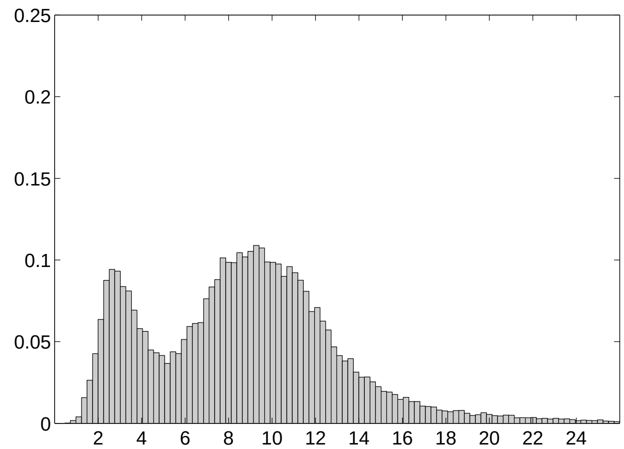
(a) Posterior density σ^2, 1948 Q1–2013 Q4



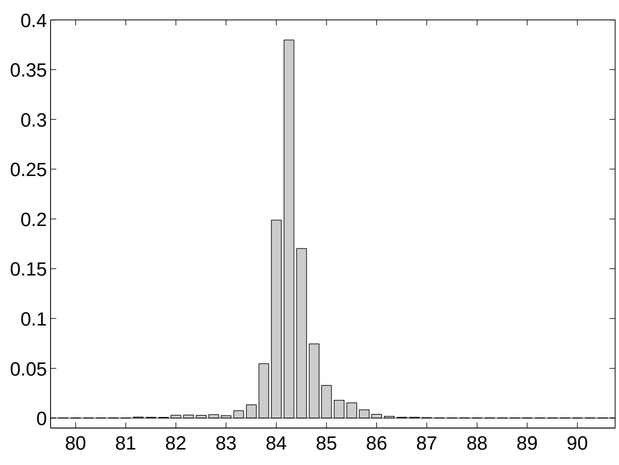
(b) Probabilities I_T^*, prior ("+"); posterior ("○")



(c) Base density σ^2, prior (dashed); posterior (solid)



(d) Posterior density σ^2, 1948 Q1–1994 Q4



(e) Posterior break date ζ, one-break model

To grasp the dynamic evolution of the variance parameter, we plot the posterior density of σ^2 based on data $\mathbf{y}^{1,\tau}$, $\tau = 1994$ Q4, in Figure 4.7(d). The high-volatility values are still dominant in this posterior, but a low-volatile regime has already been born and assigned a considerable number of observations. Next we examine what these results imply in terms of out-of-sample forecasting.

We compute the $\log_{10}$-valued predictive likelihoods $\sum_{t=\tau+1}^{T} \log_{10} p(y_t \mid \mathbf{y}^{1,t-1})$ to compare the forecasting performance of the three models, by using Jeffreys' model-comparison guidelines for the resulting predictive Bayes factors. If we compare our variance-change model to the model specification with time-constant variance, for the out-of-sample period 1995 Q1–2013 Q4 (penultimate column of Panels A and B of Table 4.3), we find a predictive Bayes factor of 0.9. Thus, we have substantial to strong evidence in favor of the model with variance changes. Comparing our model to the one-break specification, we obtain a Bayes factor decisively favoring the latter.

We attribute these findings to the following. First, as we see in Figure 4.7(e), in the one-break model the decline in variance is most likely located during 1984 Q1–3. Hence, before we start the forecasting experiment, we have about 40 observations to identify the post-break variance parameter which takes a value of about 3.25 (Panel C of Table 4.3). The first five years of the hold-out sample contain "stable" expansionary growth rates and therefore ideally fit into this low-volatility state.

Second, though we saw the already born smaller variance in the new regime in our variance-change model in Figure 4.7(d), the high-volatility values still have dominant weight in the posterior. With this posterior the first five calm years of the forecasting period are considerably less supported. On the contrary, the period thereafter contains two recessions and here the differences in forecasting performance are less prominent. To further illustrate this finding, we report predictive likelihoods for the last fourteen years in the final column of Table 4.3. Now, Bayes factors show hardly any difference between the one-break and the parameter-change model, and both still strongly dominate the no-change specification.

The analysis of the out-of-sample results may provide support for the hypothesis that the Great Moderation is over (Clark, 2009), or at least overestimated. On the other hand, simple visual inspection of Figure 4.6(a) points otherwise, that is, it suggests that growth rates have returned to the calm and steady pre-2008 state. We further note that with the one-break variance model we implicitly use data information, since we already know about the Great Moderation and that model is ideally tuned to it. It therefore better serves as best benchmark model and we rather check how close our parameter-change model comes.

4.5 Conclusion

In this chapter we demonstrate the use of Dirichlet process mixture models to accommodate an empirical model to changing parameters. This parameter-change approach is characterized by a number of desirable properties to account for issues of this kind.

The number of in-sample and out-of-sample changes are *a priori* unknown. The dynamic specification for parameter changes contains positive implications in terms of out-of-sample forecasting. In existing models, out-of-sample parameter changes are either neglected or require computationally demanding extensions. The approach we propose implies a random number of out-of-sample breaks, the distribution of which depends on the forecasting horizon and the smoothness parameter. The uncertainty regarding possible future changes is properly assimilated in the posterior predictive distribution.

The parameter-change models can handle both small and gradual, and sudden and large parameter changes. Moreover, it provides a model-implied location-dispersion trade-off for situations with a parameter change but few observations to identify the new regime's parameter. In such cases, the model parameter's posterior is smoothly and dynamically adapted.

The approach is flexible in the sense that the required posterior simulation methods do not impose any restrictions on the model under consideration. That is, it is not confined to linear regression models or models which are fit to a (mixed) Gaussian state-space representation. Nor are non-conjugate pairs of sampling distribution and prior excluded. The modeling stage involves the choice of a (time series) model, and a base prior distribution that generates the new parameter value in case of a change. The posterior simulation techniques are computationally less complex and intensive than existing methods for structural break models, which usually need Kalman recursions or filtering techniques.

We illustrate the approach both with macroeconomic applications and with artificially generated data. It turns out that the parameter-change approach provides robust results with respect to changes in the model parameters. It generates more accurate out-of-sample forecasts than models with time-constant parameters. As we effectively demonstrate with a model to forecast the duration of unemployment, if parameter change is slow and gradual the approach is especially useful. On the other hand, if we are *a priori* pretty sure a small and known number of sudden breaks have taken place, and we have available a fair amount of observations to identify the regimes' parameters, it proves more advantageous to opt for a structural break model. We show this with a time series model for GDP growth rates, a series known for its large and sudden drop in volatility. Though, for most empirical applications we do not have this kind of specific prior knowledge and it pays off to choose for the more robust parameter-change approach.

4.A Technical details

In this appendix we provide the details of the computational procedures for so far as they are not yet discussed in the text. In Sections 4.A.1 and 4.A.2 we discuss posterior simulation steps for the model in Example 1 (Section 4.3.1) and for the Poisson model (Section 4.4.2), respectively. In Section 4.A.3 we focus on both posterior simulation and the applied forecasting methods for the models for GDP (Section 4.4.3).

4.A.1 Dynamic regression example

The dynamic regression model in Example 1 contains two groups of parameters. The time-invariant model parameters $\boldsymbol{\omega} = (\beta_0, \varphi)'$ and the time-varying parameters $\boldsymbol{\theta}_t = (\beta_{1,t}, \sigma_t^2)'$, $(t = 1, \ldots, T)$. First, we implement the MCMC steps of Section 4.3.1 to simulate $\boldsymbol{\theta} = \{\boldsymbol{\theta}_t\}_{t=1}^T$ conditional on $\boldsymbol{\omega}$ and, second, we simulate the time-constant parameters conditional on $\boldsymbol{\theta}$.

Given a sampled value for $\boldsymbol{\omega}$, we compute the auxiliary dependent variables $v_t \equiv y_t - \beta_0 - \varphi y_{t-1} = \beta_1 x_t + \sigma\varepsilon_t$, $(t = 1, \ldots, T)$. To simulate the parameters liable to changes, we implement Step 1 of Section 4.3.1 with the marginal likelihoods as in (4.8) and the posterior distributions as in (4.9). For Step 2 we first construct the I_T^* subsamples $\{\mathbf{v}_{(i)}, \mathbf{x}_{(i)}\} = \{(v_t, x_t) : \theta_t = \theta_i^*, t = 1, \ldots, T\}$ indexed by $i = 1, \ldots, I_T^*$. We then update the unique parameter values in one block by simulating

$$\sigma_i^{2;*} \,|\, \{\mathbf{v}_{(i)}, \boldsymbol{\omega}\} \sim \mathcal{IG}2\Big(\nu_{\sigma^2} + |T_{(i)}|,\, S_{\sigma^2} + \mathbf{e}_{(i)}{}'(B_\beta \mathbf{x}_{(i)}\mathbf{x}_{(i)}{}' + \mathbf{I}_{|T_{(i)}|})\mathbf{e}_{(i)}\Big),$$

with $\mathbf{e}_{(i)} = \mathbf{v}_{(i)} - \mathbf{x}_{(i)} b_\beta$, and conditional on this sampled variance we simulate

$$\beta_1^* \,|\, \{\mathbf{v}_{(i)}, \boldsymbol{\omega}, \sigma_i^{2;*}\} \sim \mathcal{N}\Big((B_\beta \mathbf{x}_{(i)}{}' \mathbf{v}_{(i)} + b_\beta)/(B_\beta \mathbf{x}_{(i)}{}' \mathbf{x}_{(i)} + 1),\, \sigma_i^{2;*} B_\beta / (B_\beta \mathbf{x}_{(i)}{}' \mathbf{x}_{(i)} + 1)\Big).$$

To simulate the time-constant parameters given the sampled value for $\boldsymbol{\theta}$, we use the regression equations $v_t \equiv (y_t - \beta_{1t} x_t)/\sigma_t = \beta_0 \sigma_t^{-1} + \varphi y_{t-1} \sigma_t^{-1} + \varepsilon_t$. If we further define $\boldsymbol{\sigma}^{-1} = (\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_T^{-1})'$, we formulate the auxiliary regression model $\mathbf{v} = \mathbf{W}\boldsymbol{\omega} + \boldsymbol{\varepsilon}$, with design matrix $\mathbf{W} = (\boldsymbol{\iota}_2' \otimes \boldsymbol{\sigma}^{-1}) \odot (\boldsymbol{\iota}_T,\, \mathbf{y}^{0,T-1})$. With the prior parameters $\mathbf{b} = (b, a)'$ and $\mathbf{B} = \mathrm{diag}(B, A)$, the full conditional posterior density of $\boldsymbol{\omega}$ is

$$p(\boldsymbol{\omega} \,|\, \mathbf{y}, \boldsymbol{\theta}) \propto \exp\Big\{-(\boldsymbol{\omega} - \bar{\mathbf{b}})' \bar{\mathbf{B}}^{-1} (\boldsymbol{\omega} - \bar{\mathbf{b}})/2\Big\} \times \mathbb{I}_{\{0<\varphi<1\}}. \tag{4.A.1}$$

The expression on the right is the kernel of a one-sided truncated bivariate normal distribution. Its parameters are the matrix $\bar{\mathbf{B}} = (\mathbf{W}'\mathbf{W} + \mathbf{B}^{-1})^{-1}$ and the vector $\bar{\mathbf{b}} = \bar{\mathbf{B}}(\mathbf{W}'\mathbf{v} + \mathbf{B}^{-1}\mathbf{b})$.

Using the properties of the multivariate normal distribution, we decompose the quadratic term in (4.A.1) as the product of a conditional Gaussian kernel for $\beta_0 \,|\, \varphi$ and a marginal Gaussian kernel for φ. We integrate the former (which has unrestricted support) to obtain the marginal truncated normal distribution of φ.[12] We sample in one block from $p(\boldsymbol{\omega} \,|\, \mathbf{y}) = p(\beta_0 \,|\, \mathbf{y}, \varphi) p(\varphi \,|\, \mathbf{y})$. That is, with $\bar{b}_{(j)}$ and $\bar{B}_{(ij)}$ the j-th and (i,j)-th elements of $\bar{\mathbf{b}}$ and $\bar{\mathbf{B}}$, respectively, we simulate

$$\begin{aligned}
&\varphi \,|\, \{\mathbf{y}, \boldsymbol{\theta}\} \sim \mathcal{N}(\bar{b}_{(2)}, \bar{B}_{(22)}) \times \mathbb{I}_{\{0<\varphi<1\}},\\
&\beta_0 \,|\, \{\mathbf{y}, \boldsymbol{\theta}, \varphi\} \sim \mathcal{N}(\bar{b}_{(1)} + (\varphi - \bar{b}_{(2)}) \bar{B}_{(12)}/\bar{B}_{(11)},\, \bar{B}_{(11)} - \bar{B}_{(12)}^2/\bar{B}_{(11)}).
\end{aligned}$$

4.A.2 Predictive Poisson regression

For the parameter-change model for the number of weeks of unemployment in Section 4.4.2 we provide the details to implement the sampling of the base prior's parameters.

When we have a configuration obtained from Step 1 in Section 4.3.1, we have I_T^* distinct values $\boldsymbol{\beta}_i^*$ for the regression parameters. Each is generated from the bivariate normal base prior distribution in (4.10). First, we form the $(I_T^* \times 2)$-matrix $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*, \ldots, \boldsymbol{\beta}_{I_T^*}^*)'$. Next, with the prior for $\boldsymbol{\lambda} = \{\mathbf{b}_\beta, \mathbf{B}_\beta\}$ as in (4.11), we obtain the full conditional posterior density

$$\begin{aligned}
p(\boldsymbol{\lambda} \,|\, \mathbf{y}, \boldsymbol{\beta}^*) \propto{}& |\mathbf{B}_\beta|^{-I_T^*/2} \exp\Big\{-\mathrm{tr}\Big[\mathbf{B}_\beta^{-1} (\boldsymbol{\beta}^* - \boldsymbol{\iota}_{I_T^*} \mathbf{b}_\beta')' (\boldsymbol{\beta}^* - \boldsymbol{\iota}_{I_T^*} \mathbf{b}_\beta')\Big]/2\Big\}\\
&\times |\mathbf{B}_\beta|^{-(\nu_0+4)/2} \exp\Big\{-\mathrm{tr}\Big[\mathbf{B}_\beta^{-1}\Big(B_0^{-1}(\mathbf{b}_\beta - \mathbf{b}_0)(\mathbf{b}_\beta - \mathbf{b}_0)' + \mathbf{S}_0\Big)\Big]/2\Big\}.
\end{aligned}$$

Using computational results for a multivariate normal regression model with a conjugate prior, we obtain a decomposition of the full conditional posterior density of $\boldsymbol{\lambda}$ which allows for the one-block simulation

$$\begin{aligned}
&\mathbf{B}_\beta \,|\, \{\mathbf{y}, \boldsymbol{\beta}^*\} \sim \mathcal{IW}\Big(\nu_0 + I_T^*,\, \mathbf{S}_0 + (\boldsymbol{\beta}^* - \boldsymbol{\iota}_{I_T^*} \mathbf{b}_0')' (\mathbf{I}_{I_T^*} + B_0 \boldsymbol{\iota}_{I_T^*} \boldsymbol{\iota}_{I_T^*}')^{-1} (\boldsymbol{\beta}^* - \boldsymbol{\iota}_{I_T^*} \mathbf{b}_0')\Big),\\
&\mathbf{b}_\beta \,|\, \{\mathbf{y}, \boldsymbol{\beta}^*, \mathbf{B}_\beta\} \sim \mathcal{N}\Big((B_0 I_T^* + 1)^{-1}(B_0 \boldsymbol{\beta}^{*\prime} \boldsymbol{\iota}_{I_T^*} + \mathbf{b}_0),\, B_0 (B_0 I_T^* + 1)^{-1} \mathbf{B}_\beta\Big).
\end{aligned}$$

[12] The reverse is not possible since β_0 is marginally not Gaussian due to the truncation of φ, unless the two parameters are (*a posteriori*) independent, which in general they are not.

4.A.3 Models GDP growth rate

In this section we provide the steps to implement the MCMC scheme in the mean-switching variance-change model for the GDP growth rate as presented in Section 4.4.3. We also discuss posterior simulation for the two benchmark models. We end the section by describing the forecasting methods to compute the posterior predictive distributions.

Posterior simulation

The simulation algorithm is divided into two main parts. In the first part we sample the time-dependent variance parameters given the conditional-mean components $\boldsymbol{\omega} = \{\mathbf{s}, \boldsymbol{\beta}, \mathbf{P}\}$. In the second part we simulate $\boldsymbol{\omega}$ conditional on the variance parameters.

Part 1: Conditional variance

We start with defining the demeaned dependent variables

$$v_t \equiv y_t - \sum\nolimits_{j=1}^{3} \beta_j \mathbb{I}_{\{s_t=j\}} = \sigma_t \varepsilon_t, \quad (t = 1, \ldots, T),$$

which we arrange in the vector $\mathbf{v} = (v_1, \ldots, v_T)'$. These are the "data" with which we execute the two simulation steps of Section 4.3.1. The Gaussian sampling distribution and the inverted Gamma–2 base prior form a conjugate pair, so we compute the (conditional) marginal likelihood analytically as

$$\begin{aligned} p_0(\mathbf{y} \,|\, \boldsymbol{\omega}, \boldsymbol{\lambda}) &= (2\pi)^{-T/2} \frac{(S_{\sigma^2}/2)^{\nu_{\sigma^2}}}{\Gamma(\nu_{\sigma^2}/2)} \int (\sigma^2)^{-(\nu_{\sigma^2}+T+2)/2} \exp\left\{-(\mathbf{v}'\mathbf{v} + S_{\sigma^2})/(2\sigma^2)\right\} \mathsf{d}\sigma^2 \\ &= \pi^{-T/2} \Gamma((\nu_{\sigma^2}+T)/2)/\Gamma(\nu_{\sigma^2}/2) S_{\sigma^2}^{\nu_{\sigma^2}/2} (S_{\sigma^2} + \mathbf{v}'\mathbf{v})^{-(\nu_{\sigma^2}+T)/2}. \end{aligned} \tag{4.A.2}$$

For the remix step (Step 2) we simulate $\sigma_i^{2;*} \,|\, \{\mathbf{y}_{(i)}, \boldsymbol{\omega}, \boldsymbol{\lambda}\} \sim \mathcal{IG}2(\nu_{\sigma^2} + |T_{(i)}|,\, S_{\sigma^2} + \mathbf{v}_{(i)}{}'\mathbf{v}_{(i)})$, $(i = 1, \ldots, I_T^*)$. The one-observation marginal likelihoods and one-observation posteriors required to implement Step 1 are just special cases of the previous two expressions. With the T sampled values for the variances we define the vector $\boldsymbol{\sigma}^{-1} = (\sigma_1^{-1}, \ldots, \sigma_T^{-1})'$.

We simulate the base prior parameters $\boldsymbol{\lambda} = \{\nu_{\sigma^2}, S_{\sigma^2}\}$ in one block as in (4.16)–(4.17). To derive the two distributions we notice that the I_T^* unique variances $\boldsymbol{\sigma}^{2;*}$ are independent realizations from the inverted Gamma–2 base prior in (4.14). We combine their "likelihood" with the prior in (4.15) to obtain $\boldsymbol{\lambda}$'s full conditional posterior pdf

$$\begin{aligned} p(\boldsymbol{\lambda} \,|\, \mathbf{y}, \boldsymbol{\sigma}^{2;*}) &\propto \Gamma(a_0/2)^{-1}(2b_0)^{-a_0/2} S_{\sigma^2}^{a_0/2-1} \exp\{-S_{\sigma^2}/(2b_0)\} \cdot p(\nu_{\sigma^2}) \\ &\quad \times \prod\nolimits_{i=1}^{I_T^*} \Gamma(\nu_{\sigma^2}/2)^{-1} (S_{\sigma^2}/2)^{\nu_{\sigma^2}/2} (\sigma_i^{2;*})^{-(\nu_{\sigma^2}+2)/2} \exp\left\{-S_{\sigma^2}/(2\sigma_i^{2;*})\right\} \\ &\propto S_{\sigma^2}^{(a_0+\nu_{\sigma^2} I_T^*)/2-1} \exp\left\{-S_{\sigma^2}\left[\left(b_0^{-1} + \sum\nolimits_i \sigma_i^{-2;*}\right)/2\right]\right\} \qquad (4.A.3) \\ &\quad \times 2^{-\nu_{\sigma^2} I_T^*/2} \Gamma(\nu_{\sigma^2}/2)^{-I_T^*} \prod\nolimits_{i=1}^{I_T^*} (\sigma_i^{2;*})^{-(\nu_{\sigma^2}+2)/2} \cdot p(\nu_{\sigma^2}). \qquad (4.A.4) \end{aligned}$$

The expression in (4.A.3) is the kernel of the Gamma distribution $\mathcal{G}a((a_0 + \nu_{\sigma^2} I_T^*)/2,\, (b_0^{-1} + \sum_i \sigma_i^{-2;*})/2)$, which leads to $p(S_{\sigma^2} \,|\, \mathbf{y}, \boldsymbol{\sigma}^{2;*}, \nu_{\sigma^2})$ from wich we sample as in (4.17), after we have simulated ν_{σ^2} from the marginal $p(\nu_{\sigma^2} \,|\, \mathbf{y}, \boldsymbol{\sigma}^{2,*})$ as follows.

The kernel in (4.A.3) integrates to $[(b_0^{-1} + \sum_i \sigma_i^{-2;*})/2]^{-(a_0+\nu_{\sigma^2} I_T^*)/2} \Gamma((a_0 + \nu_{\sigma^2} I_T^*)/2)$, which we multiply with the expression in (4.A.4) to obtain the probability mass function in (4.16).

We evaluate this function in all supported values $\underline{\nu}_j$, normalize and simulate ν_{σ^2} from the resulting discrete marginal distribution.

To complete the simulation procedure for the variance-change part of the model, we simulate the smoothness parameter α as in Section 4.3.2.

Next, we discuss sampling of the variance parameters in the two benchmark models. For the no-change model we have $\sigma_t^2 = \sigma^2$ for all t. With the prior in Section 4.4.3, we simulate

$$\sigma^2 \,|\, \{\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}\} \sim \mathcal{IG}2(\nu_{\sigma^2} + T,\, S_{\sigma^2} + \mathbf{v}'\mathbf{v}).$$

Conditional on a draw of the single variance parameter we define $\boldsymbol{\sigma}^{-1} = \sigma^{-1}\boldsymbol{\iota}_T$, and we sample $\boldsymbol{\lambda}$ as in (4.16)–(4.17)

In the one-break model with unknown break location ζ, the conditional variance equals σ_1^2 when $t \leq \zeta$, and σ_2^2 for $t > \zeta$. Since the two variance parameters are *a priori* (conditionally) independent and their priors form a (conditionally) conjugate pair with the sampling distribution, we sample all three parameters in one block using the decomposition $p(\sigma_1^2, \sigma_2^2, \zeta \,|\, \mathbf{y}) = p(\sigma_1^2 \,|\, \mathbf{y}, \zeta)p(\sigma_2^2 \,|\, \mathbf{y}, \zeta)p(\zeta \,|\, \mathbf{y})$, (here, we temporarily suppress the conditioning on the other parameters).

Applying the result in (4.A.2), we obtain the probability mass function

$$\begin{aligned} p(\zeta \,|\, \mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}) &\propto p(\zeta) \int p(\mathbf{y} \,|\, \boldsymbol{\omega}, \sigma_1^2, \sigma_2^2, \zeta) p(\sigma_1^2 \,|\, \boldsymbol{\lambda}) p(\sigma_2^2 \,|\, \boldsymbol{\lambda}) \, \mathsf{d}\{\sigma_1^2, \sigma_2^2\} \\ &\propto \sum\nolimits_{t=t_1}^{t_2} \mathbb{I}_{\{\zeta = t\}} p_{\zeta,t} \, \Gamma((\zeta + \nu_{\sigma^2})/2)(S_{\sigma^2} + \mathbf{v}^{1,\zeta\prime}\mathbf{v}^{1,\zeta})^{-(\zeta+\nu_{\sigma^2})/2} \\ &\quad \times \Gamma((T - \zeta + \nu_{\sigma^2})/2)(S_{\sigma^2} + \mathbf{v}^{\zeta+1,T\prime}\mathbf{v}^{\zeta+1,T})^{-(T-\zeta+\nu_{\sigma^2})/2}. \end{aligned}$$

We evaluate this function for all supported break dates, normalize and sample ζ from the discrete marginal distribution. Conditional on the break date, the two variance parameters are independent and we simulate

$$\begin{aligned} \sigma_1^2 \,|\, \{\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \zeta\} &\sim \mathcal{IG}2(\nu_{\sigma^2} + \zeta,\, S_{\sigma^2} + \mathbf{v}^{1,\zeta\prime}\mathbf{v}^{1,\zeta}), \\ \sigma_2^2 \,|\, \{\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \zeta\} &\sim \mathcal{IG}2(\nu_{\sigma^2} + T - \zeta,\, S_{\sigma^2} + \mathbf{v}^{\zeta+1,T\prime}\mathbf{v}^{\zeta+1,T}). \end{aligned}$$

Finally, we define $\boldsymbol{\sigma}^{-1} = (\sigma_1^{-1}\boldsymbol{\iota}_\zeta',\, \sigma_2^{-1}\boldsymbol{\iota}_{T-\zeta}')'$, and with the two independent "realizations" from the $\mathcal{IG}2(\nu_{\sigma^2}, S_{\sigma^2})$ distribution, we sample $\boldsymbol{\lambda}$ according to (4.16)–(4.17).

Part 2: Conditional mean

In the second part of the MCMC routine we condition on the variance parameter(s) sampled in Part 1, and simulate the components of the conditional mean. This part is the same for all three models under consideration. We first rewrite the model in the regression form

$$v_t \equiv y_t/\sigma_t = \sigma_t^{-1} \sum\nolimits_{j=1}^{3} \mathbb{I}_{\{s_t = j\}} \beta_j + \varepsilon_t, \quad (t = 1, \ldots, T),$$

such that we obtain a switching regression with the vector of dependent variables $\mathbf{v}$, single-regressor vector $\mathbf{w} = \boldsymbol{\sigma}^{-1}$, and independently $\mathcal{N}(0, 1)$-distributed errors. Sampling $\boldsymbol{\omega}$ consists of three steps.

First, we simulate all T regime indicators $\mathbf{s}$ in one block from $p(\mathbf{s} \,|\, \mathbf{y}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \mathbf{P})$ with Chib's (1996) simulation smoother.

Second, due to the identifying restriction on the regression parameters, we sample them successively from their respective full conditional posteriors distributions, which are univariate truncated normal. If we define the subsample $\{\mathbf{v}_{(j)}, \mathbf{w}_{(j)}\} = \{(v_t, w_t) : s_t = j,\, t = 1, \ldots, T\}$, we successively $(j = 1, 2, 3)$ simulate

$$\beta_j \,|\, \{\mathbf{y}, \boldsymbol{\sigma}^2, \mathbf{s}, \boldsymbol{\beta}_{-j}\} \sim \mathcal{N}((B_{j|-j}\mathbf{v}_{(j)}{}'\mathbf{w}_{(j)} + b_{j|-j})/\widetilde{B}_j,\; B_{j|-j}/\widetilde{B}_j) \times \mathbb{I}_{\{\beta_{j-1}<\beta_j<\beta_{j+1}\}},$$

in which $\widetilde{B}_j = B_{j|-j}\mathbf{w}_{(j)}{}'\mathbf{w}_{(j)} + 1$. We additionally set $\beta_0 = -\infty$ and $\beta_4 = \infty$, and use the conditional prior parameters $b_{j|-j} = b_j + \mathbf{B}_j{}'(\boldsymbol{\beta}_{-j} - \mathbf{b}_{-j})/B_{jj}$, and $B_{j|-j} = B_{jj} - \mathbf{B}_j{}'\mathbf{B}_j/B_{jj}$, in which $\mathbf{B}_j$ is the j-th column of prior variance matrix $\mathbf{B}$, with the element B_{jj} deleted.

Third and finally, we sample the transition probabilities. The Markov property of $\{s_t\}$ provides the probability of a realization of the process, which is expressed as

$$p(\mathbf{s}\,|\,\mathbf{P}) = \Big(\sum_{j=1}^{3} \pi_j(\mathbf{P})\mathbb{I}_{\{s_1=j\}}\Big)\Big(\prod_{t=2}^{T} \sum_{i=1}^{3} \sum_{j=1}^{3} p_{ij}\mathbb{I}_{\{s_{t-1}=i \,\wedge\, s_t=j\}}\Big).$$

In combination with the independent Dirichlet priors for the three rows of $\mathbf{P}$, we obtain the full conditional posterior densities

$$p(\mathbf{P}_i\,|\,\mathbf{y}, \mathbf{s}, \mathbf{P}_{-i}) \propto \Big(\sum_{j=1}^{3} \pi_j(\mathbf{P})\mathbb{I}_{\{s_1=j\}}\Big)\Big(\prod_{j=1}^{3} p_{ij}^{a_{ij}+T_{ij}-1}\Big)\mathbb{I}_{\{\boldsymbol{\iota}_3'\mathbf{P}_i=1\}}, \quad (i = 1, 2, 3),$$

with $T_{ij} = \sum_{t=2}^{T} \mathbb{I}_{\{s_{t-1}=i \,\wedge\, s_t=j\}}$ as the number of transitions from state i to state j in the sampled value of $\mathbf{s}$, such that $\sum_{i,j} T_{ij} = T-1$. We note that $\mathbf{P}$'s rows are mutually dependent due to the initialization of the latent process. Nevertheless, the second term is the dominating part and equals the kernel of a Dirichlet distribution which does not depend upon the other rows. We follow Geweke (2005) and implement a Metropolis–Hastings sampler. In simulation run $m + 1$, we do the following successively for $i = 1, 2, 3$.

Step 1. Sample a proposal vector $\mathbf{P}_i^* \,|\, \mathbf{s} \sim \mathcal{D}ir(a_{i1} + T_{i1}, a_{i2} + T_{i2}, a_{i3} + T_{i3})$;

Step 2. Compute the invariant distribution $\pi(\mathbf{P}^*)$ under this proposal;

Step 3. Compute $\alpha(\mathbf{P}_i^*, \mathbf{P}_i^{(m)}) = \min\Big\{\sum_{j=1}^{3}[\pi_j(\mathbf{P}^*)/\pi_j(\mathbf{P}^{(m)})]\mathbb{I}_{\{s_1=j\}}, 1\Big\}$, and accept the proposal with probability α and update the transition matrix by replacing its i-th row by $\mathbf{P}_i^{*\prime}$; in case of rejection, maintain the current matrix and set $\mathbf{P}_i^{(m+1)} = \mathbf{P}_i^{(m)}$.

Forecasting

In this final section we discuss the computation of the predictive distributions, for which we use sequential importance sampling techniques (see, for example Robert and Casella, 2004).

We first demarginalize the one-quarter-ahead predictive density of y_t as

$$\begin{aligned} p(y_t\,|\,\mathbf{y}^{1,t-1}) = \int\!\!\int p(y_t\,|\,\mathbf{y}^{1,t-1}, \theta_t, s_t)p(\theta_t, s_t\,|\,\mathbf{y}^{1,t-1}, \boldsymbol{\theta}^{1,s-1}, \mathbf{s}^{1,t-1})\,\mathsf{d}\{s_t, \theta_t\} \\ \times\, p(\boldsymbol{\theta}^{1,t-1}, \mathbf{s}^{1,t-1}\,|\,\mathbf{y}^{1,t-1})\,\mathsf{d}\{\mathbf{s}^{1,t-1}, \boldsymbol{\theta}^{1,t-1}\}. \end{aligned} \tag{4.A.5}$$

The first term of the integrand equals $p_t(y_t; s_t, \theta_t)$. Since θ_t and s_t are *a priori* independent, the second term in (4.A.5) simplifies to $p(\theta_t\,|\,\boldsymbol{\theta}^{t-1})p(s_t\,|\,s_{t-1})$, in which the first pdf is given in (4.1) and the second is the transition density of the Markov process in (4.12). The final term is the posterior density in quarter $t-1$. Given a draw from the latter we proceed as follows to solve the first integral.

First, we simulate θ_t according to (4.12). Given that draw, we analytically integrate with respect to s_t and obtain the “predictive” likelihood contribution

$$p_{t\,|\,s_{t-1}=i}(y_t;\theta_t)=\sum\nolimits_{j=1}^{3}p_{ij}p_t(y_t;s_t=j,\theta_t).$$

To compute the second integral we apply sequential importance sampling. Instead of using a sample from $p(\boldsymbol{\theta}^{1,t-1},\mathbf{s}^{1,t-1}\,|\,\mathbf{y}^{1,t-1})$, we use one from the posterior one quarter back and compensate by importance sampling. We decompose the target posterior as

$$p(\boldsymbol{\theta}^{1,t-1},\mathbf{s}^{1,t-1}\,|\,\mathbf{y}^{1,t-1})=p(s_{t-1}\,|\,\mathbf{y}^{1,t-1},\boldsymbol{\theta}^{1,t-1},\mathbf{s}^{1,t-2})p(\boldsymbol{\theta}^{1,t-1},\mathbf{s}^{1,t-2}\,|\,\mathbf{y}^{1,t-1}). \tag{4.A.6}$$

We simulate from the first distribution on the right, which is the “updated” distribution of s_{t-1}. That is, we sample from the discrete distribution with probabilities obtained with

$$p(s_{t-1}\,|\,\mathbf{y}^{1,t-1},\boldsymbol{\theta}^{1,t-1},\mathbf{s}^{1,t-2})\propto p_{t-1}(y_{t-1};s_{t-1},\theta_{t-1})p(s_{t-1}\,|\,s_{t-2}). \tag{4.A.7}$$

We write the second term on the right in (4.A.6) as the product of three terms

$$\frac{p(y_{t-1}\,|\,\mathbf{y}^{1,t-1},\mathbf{s}^{1,t-2},\boldsymbol{\theta}^{1,t-1})}{p(y_{t-1}\,|\,\mathbf{y}^{1,t-2})}p(\theta_{t-1}\,|\,\boldsymbol{\theta}^{1,t-2},\mathbf{s}^{1,t-2},\mathbf{y}^{1,t-2})p(\boldsymbol{\theta}^{1,t-2},\mathbf{s}^{1,t-2}\,|\,\mathbf{y}^{1,t-2}).$$

The ratio equals the importance weight and the second term reduces again to the transition density $p(\theta_{t-1}\,|\,\boldsymbol{\theta}^{1,t-2})$, which we use to simulate θ_{t-1}. Since the third term is the posterior in quarter $t-2$, we repeat the previous for another quarter back, and we obtain a recursion for the importance weights.

To summarize the above in algorithmic form, we start with a sample from the posterior with pdf $p(\boldsymbol{\theta}^{1,\tau},\mathbf{s}^{1,\tau}\,|\,\mathbf{y}^{1,\tau})$, and for $t=\tau+1,\ldots,T$ we implement the following.

Step 1. Simulate θ_t according to (4.1);

Step 2. Simulate s_{t-1} according to (4.A.7);

Step 3. Update the importance weights by

$$w_t(\theta_{t-1},s_{t-2})=\frac{p_{t-1\,|\,s_{t-2}}(y_{t-1};\theta_{t-1})}{p(y_{t-1}\,|\,\mathbf{y}^{1,t-2})}w_{t-1}(\theta_{t-2},s_{t-3}),$$

with the predictive likelihood contribution already computed in Step 2; we start with $w_{\tau+1}=1$;

Step 4. Compute $p_{t\,|\,s_{t-1}}(y_t;\theta_t)\cdot w_t$, of which the simulation-sample average gives the Monte Carlo estimate of the predictive density $p(y_t\,|\,\mathbf{y}^{1,t-1})$.

CHAPTER 5

Economic Activity Nonparametrically Related to a Leading Indicator

5.1 Introduction

Cyclical changes in economic activity form an extensively studied topic in the empirical macroeconomics literature since the 1920s. Its interest originates both from a theoretical point of view to trace any cycle-driving structural causes, as well as from a forecasting perspective to anticipate future activity changes, as demonstrated in the seminal works of Persons (1919) and Burns and Mitchell (1946). In this field of research, (macro)economic variables leading the state of economic activity hold a central position. In spite of all economic-theoretical controversy and the invariably changing macroeconomic debate (Diebold, 1998, Qin, 2010), new releases of leading indicators have proven to be informative about the future state of the economy throughout the last century. The leading indcators are therefore intently watched and analyzed, mainly for forecasting purposes.

Inter alia, Diebold and Rudebusch (1991a), Lahiri and Wang (1994), Hamilton and Perez-Quiros (1996), Marcellino (2006) and the various chapters in Lahiri and Moore (1991) effectively demonstrate the usefulness of leading indicators for macroeconomic forecasting. The applied leading-indicator forecasting methods mainly address two issues. First, determining which leading variables are most useful for forecasting economic activity for a given number of months ahead (Clements and Galvão, 2006). This is also related to the identification of the different lead times of the indicators (see Paap *et al.*, 2009, and Chapter 3). The second issue concerns the functional form of the forecasting relation. Neftçi (1984), for example, established the stylized fact that many individual macroeconomic variables show statistical behavior dependent on the state of the business cycle. In this chapter we analyze issues of the second kind and investigate nonstandard relations between leading indicators and coincident variables. We apply a methodology that not only incorporates nonlinear relations, but which also covers any further nonstandard properties such as heteroskedasticity and rare large economic shocks which are insufficiently supported by a normal distribution.

If the distribution of the economic activity measure is of a nonstandard type (that is, non-Gaussian), ignoring this has serious consequences in terms of the reliability and usefulness of a business-cycle forecasting method. When the decisionmaker has a loss function different from standard squared loss, an adequate description of economic activity's whole conditional distribution is important, in addition to only modeling its first two moments. In such cases, special attention is often directed to less likely, but more influential events like an extremely slackening economy. In this chapter we apply The Conference Board's (TCB) composite coincident index to indicate the current state of economic activity and TCB's composite leading index as compound measure that leads the business cycle (The Conference Board, 2001). We nonparametrically estimate the joint distribution of these two economic variables for the period January 1965–December 2013. As such, we estimate both the unknown form of the predictive relation as well as the marginal distributions. For neither we impose any particular prior assumptions.

We examine nonlinearities to check whether the strength of the predictive relation changes with the value of the leading indicator. Moreover, the uncertainty surrounding a conditional-mean point forecast, as measured by the predictive variance of the future coincident-index growth rate, can vary with the value of the leading indicator as well. Finally, since the nonparametric method uses a mixture of normal distributions, the shocks to economic activity are non-Gaussian in general, thus it is able to more accurately estimate, and avoid underestimation of, the probability of more extreme changes in business conditions.

A Bayesian nonparametric method based on Dirichlet process mixture (DPM) models, as applied to univariate density-smoothing problems by Escobar (1994) and Escobar and West (1995), and generalized for the multivariate setting as in Müller *et al.* (1996), serves as our econometric instrument. DPM models have wide applicability and have recently gained increased interest due to advances in computational possibilities. Besides density estimation its applications also include survival analysis, hierarchical models and heterogeneity modeling. We refer to, amongst others, Bush and MacEachern (1996), Campolieti (2001), Gelfand and Kottas (2002), Müller and Quintana (2004) and Burda *et al.* (2009) for examples. In our application the approach amounts to estimating the unknown joint distribution with a mixture of multivariate normal distributions with an *a priori* unknown number of mixture components. From a forecasting perspective and for macroeconomic analysis our key interest lies in the conditional distribution of the coincident-index growth rate given the value of the leading index. This distribution naturally follows from the estimated joint and is of the mixture type too.

To examine what typical statistical features characterize the relation between the leading indicator and the coincident index, we make a comparison with threshold-regression models by computing Bayes factors. In threshold regressions, the various forms of the predictive relation are parametrically modeled such that they can explicitly vary with the value of the predictor. As such we use these models to help identifying which statistical properties of economic activity depend on the leading indicator, which is not immediately clear from the nonparametric approach because it estimates the whole distribution at once. We use threshold regressions for model comparison and not, for example, Markov-switching models, since in the latter the (strength of the) predictive relation is driven by an unobserved variable instead of the predictor variable as in our estimated conditional distribution of economic activity.

The set-up of our procedure is as follows. First, we check for any deviations from Gaussianity of the joint distribution of the growth rates of the coincident and leading economic indexes. We obtain the Bayes factors for this model comparison by applying the marginal likelihood methods of Basu and Chib (2003). Second, we compute the conditional distribution of economic activity given the growth rate of the leading indicator. This conditional distribution forms the basis for forecasts of economic activity. By building on Basu and

Chib's (2003) methods we derive and compute the marginal likelihood of this nonparametric model. Third, in the analyses we use both revised data of the leading index and data as they were historically available. Fourth, we estimate threshold-regression models with independent switching between regimes for the conditional mean and the conditional variance of economic activity. We compute their marginal likelihoods and Bayes factors and compare them to the nonparametric model's.

Various research has established significant nonlinear relations between macroeconomic variables. We mention Diebold and Rudebusch (1996), Fornari and Mele (1997), Morley and Piger (2012), and recently, examining Okun's law, Jardin and Stephan (2011) and Chinn *et al.* (2013). Instead of imposing a fixed functional form between the leading and coincident index, we let the data be leading in what form fits best. This contrasts with parametric alternatives to model a nonstandard joint distribution. One of these alternatives is copula modeling. Although we find most of its applications in finance, as for instance in Jondeau and Rockinger (2006), Chen and Fan (2006a) and Ausin and Lopes (2010), recently they also appear in marketing research (Smith *et al.*, 2010) and empirical macroeconomics (Dowd, 2008, Chollete and Ning, 2009). The chosen copula function generally imposes a stringent specific type of dependence. Moreover, the copula approach entails explicit modeling of the respective marginal distributions, whereas in the nonparametric approach these are simultaneously estimated.

Another alternative is to apply nonparametric regression techniques to flexibly model the conditional mean of the economic activity measure. From a frequentist perspective we mention Racine and Li (2004), and Bayesian alternatives are discussed by Smith and Kohn (1996, 1997), whereas Hamilton (2001) provides a flexible nonlinear regression method suited for both kinds of statistical inference. The Bayesian nonparametric method we apply is however not confined to the conditional mean, but incorporates possibly predictor-dependent higher moments as well.

One of the stylized facts of macroeconomic data is the increased uncertainty during downturns of the economy, which, though univariately, has been noted by Hamilton (1989). In a recent paper by Foerster (forthcoming), this type of time-varying volatility is further analyzed and empirically shown to be (asymmetrically) correlated to economic activity. To identify nonlinearities or any changing volatility in our macroeconomic application, both induced by the value of the leading indicator, we compare the nonparametric model to threshold regressions.

Finally, we note that the estimated conditional distribution of the coincident-index growth rate is described by a discrete mixture of normal distributions with an *a priori* unknown number of mixture components. Hence, the economic shocks, which are defined to be the deviations of the coincident-index growth rate from its conditional mean, are non-Gaussian in general. With one of the essential properties of normal mixtures (see, *e.g.*, Richardson and Green, 1997, Dellaportas and Papageorgiou, 2006, Frühwirth-Schnatter, 2006), the nonparametric model can capture any skewness or peakedness in economic activity's conditional distribution. This property is especially important for an adequate quantification of the likelihood of rare, but more extreme and influential shocks to economic activity. For example, recent research by Cúrdia *et al.* (2014) shows support for heavier-than-Gaussian tails of economic shocks.

We briefly summarize our findings as follows. We first find overwhelming empirical evidence in favor of the nonparametric model for economic activity compared to the linear Gaussian regression model. This preference is partly attributable to a nonlinear relation with the leading indicator. In particular, the latter's leading signal turns out to be stronger when it attains negative values, indicating a future downturn. With real-time data the forecasting rule almost breaks down completely during periods of normal economic activity. Second, the predictive variance of the coincident-index growth rate strongly varies with the growth rate of the leading index, both for real-time and fully revised data. A decline in the leading index, which anticipates a decline in economic activity, is associated with increased uncertainty about the amount of change in economic activity. Third, in addition to the leading-indicator-varying location and scale, we find empirical evidence for a non-Gaussian distribution of economic shocks. With revised data we estimate a relatively long and heavy left tail, which would be seriously underestimated if normality were imposed. The real-time leading index tends to overestimate future economic activity, since it has considerably fewer large negative observations, and as a result does a worse forecasting job.

The remainder of this chapter consists of the following. We start with a discussion of the data in Section 5.2. In Section 5.3 we introduce the Bayesian nonparametric method for density estimation and the testing procedure to compare the nonparametric model to threshold-regression parametric alternatives. We describe the empirical results in Section 5.4 and conclude with a discussion in Section 5.5. The two appendices present the technical details of the simulation methods applied for posterior analysis and of the marginal likelihood computations required for model comparison.

5.2 Data and real-time issues

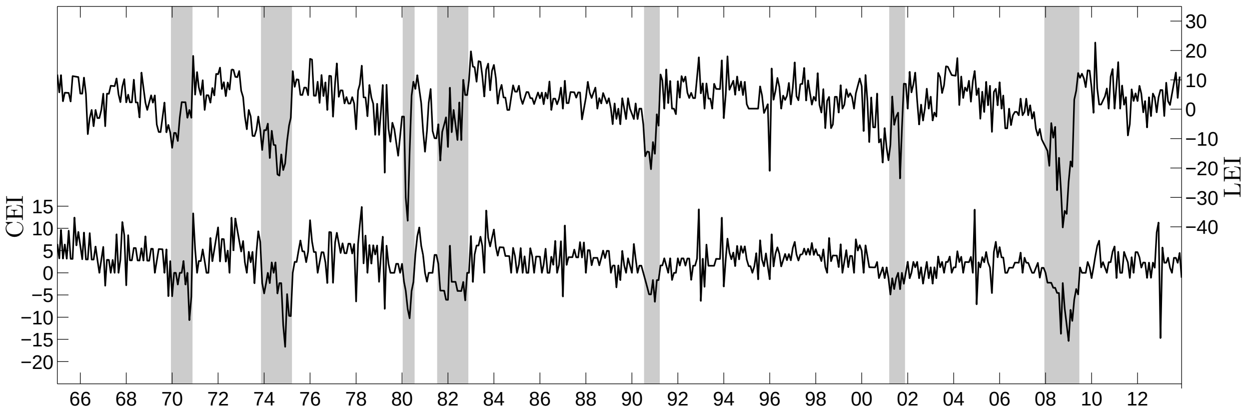
We use coincident indicators to measure economic activity in each month. Examples are sales, (industrial) production and employment. These economic variables are assumed to coincide with the business cycle as opposed to leading indicators which tend to anticipate changes in business conditions. We apply variables of the leading type, which include, for instance, financial variables and survey measures, to establish a predictive link with economic activity.

To obtain a less noisy signal of economic conditions it is good practice to report compound single economic indexes (leading as well as coincident). For example, TCB produces these indexes by summarizing information from fixed sets of individual indicators as follows (The Conference Board, 2001). First, the raw, trending series are transformed to (symmetric) growth rates (except interest rates, for which spreads are used), and these transformed series are standardized and summed to obtain the single-index growth rate.[1] For the sample period January 1965–Deceember 2013 we display the one-month symmetric growth rates of the thus constructed composite coincident (CEI) and leading economic (LEI) indexes in Figure 5.1(a). With the shaded areas we indicate the recessionary periods as decided upon by the National Bureau's of Economic Research (NBER) business cycle dating committee.

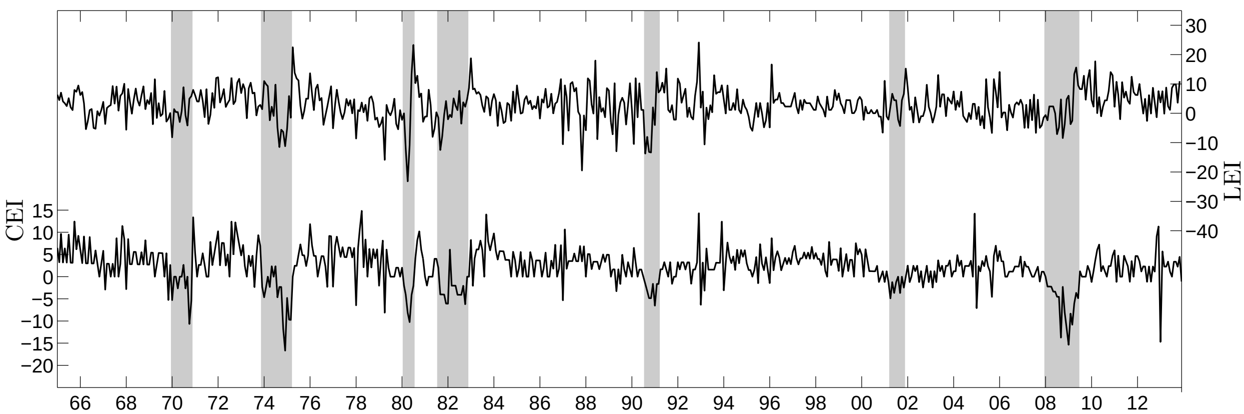
A number of leading indicators are liable to revisions after their initially released estimates, for example new orders and claims for unemployment insurance. In addition to the final data on the composite leading index [Figure 5.1(a)], we also apply our methods to the leading-indicator data as they were available in real time. We are interested in forecasting the "true" state of economic activity, and therefore we use final data for the dependent variable CEI in both settings (Diebold and Rudebusch, 1991a).

[1] To (partially) account for the large decrease in volatility of many macroeconomic series during the 1980s (Kim *et al.*, 2008), the standardization is computed separately for the pre-1983 subsample and for the period January 1984–December 2013.

Figure 5.1 *Composite coincident and leading indexes, January 1965–December 2013*



(a) Fully revised



(b) Real-time

In the real-time setting, at the end of month t we have data available on the LEI pertaining up to and including month $t-1$. If we define $\text{LEI}_s^{(t)}$ as the value of LEI in month s as released in month t, then at the end of month t, when we construct the real-time forecast of CEI_t, we can use $\text{LEI}_s^{(t)}$, $(s \leq t-1)$. The initial release for CEI_t is only available in next month $t+1$ and therefore we "predict" the present, which is termed nowcasting. We refer to Bańbura *et al.* (2013) for a survey of nowcasting and its methods. To check the predictive power of partly revised releases of LEI we use the ℓ-months lagged values $(\ell > 0)$ in month t's vintage (ℓ-th off-diagonal of the real-time data matrix). In this respect, if we use the fully revised LEI data, we obtain the "final" forecast of CEI_t by exploiting the true (reduced-form) relation between the leading index and economic activity.

Since the real-time LEI performed poorly in anticipating the two most recent recessionary periods of 2001 and 2008, as we can already derive from the modest leading-index decline in advance of these recessions in Figure 5.1(b) (see also Chapter 3), TCB has recently revised the construction of its leading index. This redefinition mainly consists of incorporating additional individual leading variables. However, in order to avoid the use of *ex-post* knowledge, for the pre-redefinition sample period we use the then prevailing construction of LEI.

5.3 Methodology

In Sections 5.3.1–5.3.2 we start with a description of the Bayesian nonparametric model, and we briefly discuss the simulation procedure to obtain a sample from the posterior distribution in Section 5.3.3. We describe the method to compare the nonparametric model to parametric alternatives in Section 5.3.4. We introduce these parametric alternatives, the threshold regressions, in Section 5.3.5.

5.3.1 Model specification

We define y_t as the symmetric monthly growth rate of CEI in month t. With $x_t = \mathrm{LEI}_{t-\ell}^{(t)}$ we denote the symmetric monthly growth of LEI in month $t-\ell$, ($\ell \geq 1$), as available in t, which serves as predictor of y_t. The two variables are collected in the vector $\mathbf{y}_t = (y_t, x_t)'$, ($t = 1, \ldots, T$). The matrix $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)' = (\mathbf{y}, \mathbf{x})$ contains all T bivariate observations.

We apply the density-smoothing methods as described in Müller *et al.* (1996) to estimate the joint distribution of CEI and LEI. This Bayesian nonparametric approach states that each of the bivariate observations is sampled from a multivariate normal distribution,

$$\mathbf{y}_t \,|\, \boldsymbol{\theta}_t \overset{ind.}{\sim} \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \quad (t = 1, \ldots, T), \tag{5.1}$$

with its own mean and covariance matrix arranged in $\boldsymbol{\theta}_t' = (\boldsymbol{\mu}_t', \mathsf{vec}\,[\boldsymbol{\Sigma}_t]')$. With a second layer of the model a structure is imposed onto the observation-specific parameters. Among others, Escobar (1994) and Escobar and West (1995) show that the DPM approach applies the set of conditional distributions with densities

$$p(\boldsymbol{\theta}_t \,|\, \boldsymbol{\theta}^{1,t-1}) = \frac{\alpha}{\alpha + t - 1} f_0(\boldsymbol{\theta}_t; \boldsymbol{\lambda}) + \frac{1}{\alpha + t - 1} \sum_{s=1}^{t-1} \delta(\boldsymbol{\theta}_t - \boldsymbol{\theta}_s), \quad (t = 2, \ldots, T), \tag{5.2}$$

in which $\boldsymbol{\theta}^{1,s} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_s)'$. With $\boldsymbol{\theta}_1 \sim F_0(\boldsymbol{\lambda})$, the joint distribution of all T parameters is completed. f_0 is the pdf of the so-called base distribution F_0, parameterized by hyperparameters $\boldsymbol{\lambda}$, and δ is the Dirac delta function indicating unit point mass at the origin.[2]

First, due to the point mass at existing values this particular structure implies that some of the $\boldsymbol{\theta}_t$'s equal each other with positive probability. Therefore, on average, some observations are generated from the same normal distribution. The number of unique normal distributions depends on both the smoothness parameter α and sample size T, both in a positive sense. If we take the sample size fixed, the smoothness of the estimated distribution is manipulated by choosing different values for α. The larger it is, the more unique mixture components are selected to more accurately fit the data. On the other hand, the more mixture components, the more bumpy the estimated distribution becomes.

Second, since the approach is nonparametric in a Bayesian way, the "parameters" do not have a clear-cut (economic) interpretation and simply serve as instruments to mould the model such that it describes the features of the data best. To choose the base distribution we therefore rely on the more practical considerations such as computational ease. We follow Escobar and West (1995) and take their multivariate equivalent which decomposes into an inverted Wishart distribution for the marginal of the covariance matrix and a conditional normal distribution for the vector of means. The main computational argument for the

[2] "Integration" of the discrete part of the parameters' pdf is with respect to the point-mass measure. We therefore define $\int g(\boldsymbol{\theta})\delta(\boldsymbol{\theta})\,\mathsf{d}\boldsymbol{\theta} \equiv g(\mathbf{0})$.

choice of this distribution is that it forms a conjugate pair with the sampling distribution. We specify the base distribution as

$$F_0(\boldsymbol{\lambda}): \quad \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma} \sim \mathcal{N}(\mathbf{b}_\mu, B_\mu \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} \sim \mathcal{IW}(\nu_\Sigma, \mathbf{S}_\Sigma). \tag{5.3}$$

Its key role in the model specification is to generate a new mean and covariance matrix when a new normal mixture component is activated.

Next, we show the power of this approach to flexibly model nonstandard statistical properties of the joint distribution of CEI and LEI. Given a value of $\boldsymbol{\theta} \equiv \boldsymbol{\theta}^{1,T}$, we have $I_T^* \leq T$ unique parameter values, a number in general much smaller than the sample size. Each unique parameter value $\boldsymbol{\theta}_j^*$, $(j = 1, \ldots, I_T^*)$, occurs $n_{j,T}$ times in $\boldsymbol{\theta}$, such that $\sum_j n_{j,T} = T$. After having observed $\mathbf{Y}$, we derive the posterior predictive distribution which is of the discrete-mixture type. We obtain its pdf

$$\begin{aligned} p(\mathbf{y}_{T+1} \,|\, \mathbf{Y}) &= \int \int p(\mathbf{y}_{T+1} \,|\, \boldsymbol{\theta}_{T+1}) p(\boldsymbol{\theta}_{T+1} \,|\, \boldsymbol{\theta}) \, \mathsf{d}\boldsymbol{\theta}_{T+1} \, p(\boldsymbol{\theta} \,|\, \mathbf{Y}) \, \mathsf{d}\boldsymbol{\theta} \\ &= \int \left[\frac{\alpha}{\alpha + T} p_0(\mathbf{y}_{T+1}; \boldsymbol{\lambda}) + \sum_{j=1}^{I_T^*} \frac{n_{j,T}}{\alpha + T} p(\mathbf{y}_{T+1} \,|\, \boldsymbol{\theta}_j^*) \right] p(\boldsymbol{\theta} \,|\, \mathbf{Y}) \, \mathsf{d}\boldsymbol{\theta}, \end{aligned} \tag{5.4}$$

after integrating out the future $\boldsymbol{\theta}_{T+1}$ analytically using (5.1)–(5.2). The one-observation marginal likelihood under prior F_0, $p_0(\mathbf{y}_{T+1}; \boldsymbol{\lambda}) = \int p(\mathbf{y}_{T+1} \,|\, \boldsymbol{\theta}) f_0(\boldsymbol{\theta}; \boldsymbol{\lambda}) \, \mathsf{d}\boldsymbol{\theta}$, is available in closed form due to the conjugate choice of base distribution. However, since its weight is relatively small in the mixture in (5.4), for convenience we consider a future observation to be generated from a mixture of multivariate normal distributions. The number of mixture components and corresponding mixture weights and components' parameters are weighed according to the posterior distribution. This amounts to a very flexible way to estimate an unknown distribution, due to the well-known richness of Gaussian mixtures, while at the same time the method circumvents the often cumbersome procedure to reliably determine the number of mixture components (see, among others, Richardson and Green, 1997, Frühwirth-Schnatter, 2006).

Since we model the joint distribution of (CEI,LEI), both the unknown dependence structure between, and the marginal distributions of the various elements of $\mathbf{y}_{T+1}$ (and combinations thereof) are estimated at the same time. In our economic application we have specific interest for one element (CEI) conditional on a predictive variable (LEI) and hence require a conditional distribution. To derive the conditional distribution of CEI given LEI, we use their joint distribution in (5.4). The mixture form is maintained and we obtain

$$\begin{aligned} p(y_{T+1} \,|\, \mathbf{Y}, x_{T+1}) &= p(\mathbf{y}_{T+1} \,|\, \mathbf{Y}) / p(x_{T+1} \,|\, \mathbf{Y}) \\ &= \int \left[w_0 p_0(y_{T+1} \,|\, x_{T+1}; \boldsymbol{\lambda}) + \sum_{j=1}^{I_T^*} w_j p(y_{T+1} \,|\, x_{T+1}, \boldsymbol{\theta}_j^*) \right] p(\boldsymbol{\theta} \,|\, \mathbf{Y}) \, \mathsf{d}\boldsymbol{\theta}, \end{aligned} \tag{5.5}$$

for its pdf, in which the mixture weights are given by

$$w_0 = \frac{\alpha p_0(x_{T+1}; \boldsymbol{\lambda})}{(\alpha + T) p(x_{T+1} \,|\, \mathbf{Y})}, \quad w_j = \frac{n_{j,T} p(x_{T+1} \,|\, \boldsymbol{\theta}_j^*)}{(\alpha + T) p(x_{T+1} \,|\, \mathbf{Y})}, \quad (j = 1, \ldots, I_T^*).$$

For the denominators of these weights we derive $p(x_{T+1} \,|\, \mathbf{Y}) = \int p(\mathbf{y}_{T+1} \,|\, \mathbf{Y}) \, \mathsf{d}y_{T+1}$, which is the marginal posterior predictive pdf of LEI evaluated in x_{T+1}. This density is given by

$p(x_{T+1} \,|\, \mathbf{Y}) = \int [\frac{\alpha}{\alpha+T} p_0(x_{T+1}; \boldsymbol{\lambda}) + \sum_{j=1}^{I_T^*} \frac{n_{j,T}}{\alpha+T} p(x_{T+1} \,|\, \boldsymbol{\theta}_j^*)] p(\boldsymbol{\theta} \,|\, \mathbf{Y}) \, \mathsf{d}\boldsymbol{\theta}$. All individual densities required to compute the conditional pdf in (5.5) are just conditionals and marginals of the bivariate densities in (5.4). Hence, all but two are Gaussian with the familiar conditional and marginal parameters. The exceptions are $p_0(y_{T+1} \,|\, x_{T+1}; \boldsymbol{\lambda})$ and $p_0(x_{T+1}; \boldsymbol{\lambda})$, which are univariate Student's t densities (see Appendix 5.A.1). We define the model of CEI given LEI which coincides with the derived conditional distribution as the nonparametric model of economic activity.

To examine nonlinear effects of the leading indicator on economic activity we need the regression function, which describes the conditional mean of CEI given LEI. Moreover, since we have a specification for the whole conditional distribution, (5.5) enables us to compute any expectation of a function of CEI given LEI. For example, we also use it to examine the degree of heteroskedasticity. The expression of the posterior predictive expectation of $f(y_{T+1})$ (conditional on $x_{T+1} = x$) becomes

$$\begin{aligned} \bar{f}(x) &\equiv \mathsf{E}\,[f(y_{T+1}) \,|\, \mathbf{Y}, x] \\ &= \int \Big[w_0 \mathsf{E}\,[f(y_{T+1}) \,|\, x; \boldsymbol{\lambda}] + \sum_{j=1}^{I_T^*} w_j \mathsf{E}[f(y_{T+1}) \,|\, x, \boldsymbol{\theta}_j^*] \Big] p(\boldsymbol{\theta} \,|\, \mathbf{Y}) \, \mathsf{d}\boldsymbol{\theta}. \end{aligned} \tag{5.6}$$

Thus, we evaluate and weigh the expectations of f under each of the I_T^*+1 distinct "regimes." The more observations assigned to a regime, the larger the weight of its expectation. Also, the likelier the predictor value under regime j (as measured by $p(x \,|\, \boldsymbol{\theta}_j^*)$), again the larger the weight of that regime's expectation. We further note that the (for the dominant part, Gaussian) mixture form of the conditional predictive distribution can straightforwardly incorporate nonnormality due to, for example, infrequent large economic shocks.

5.3.2 Prior distribution

The parameter α drives the number of mixture components to estimate the joint distribution of CEI and LEI, given fixed sample size T. *In extremis*, $\alpha = 0$ and we only have a single component, reducing the model to a standard bivariate Gaussian distribution with prior F_0. Equation (5.2) shows that with increasing α it becomes more likely that a new mixture component is generated. Antoniak (1974) derives the prior distribution of the total number of unique components I_T^* as a function of α. This result is important for dealing with the smoothness parameter in our application. Instead of using a fixed value, Escobar and West (1995) propose an extension consisting of a prior on α, to avoid a too informative prior for I_T^*. They opt for the Gamma prior

$$\alpha \sim \mathcal{G}a(a_\alpha, b_\alpha). \tag{5.7}$$

We consider two of such priors and check the dependence of the results on the implied priors on I_T^*. First, we locate prior modes of I_T^* at 5 and 10, and next, we fatten the tails of I_T^* with the two prior distibutions $\mathcal{G}a(0.72 \times 15, 15)$ and $\mathcal{G}a(1.68 \times 15, 15)$ for α. These steps result in $\mathsf{Var}\,[I_T^*] = 5.5; 10.7$, respectively, which we obtain by prior Monte Carlo simulation.

Escobar (1994) proposes to also put a prior on the base distribution's hyperparameters, for two main reasons. First, it results in a smoother density estimate as it directs the hyperparameters towards data-supported values. Second, a computational argument, it improves the mixing of the Markov chain used for posterior simulation as it induces additional

Table 5.1 *Effects* ceteris paribus *of prior parameters' settings*

	$\mathbf{b}_0$	$\mathbf{B}_0$	S_0	ν_0	$S_{w,ii}$	$S_{w,ij}$	ν_w	ν_Σ
Effect on								
Location	+	0	0	0	0	0	0	0
Scale	0	+	+	−	+	0	+	−
Dependence	0	0	0	0	−	+	0	0
Value	$\mathbf{0}_2$	$5\mathbf{I}_2$	10	5	10	0.5	8	6

Notes: The table shows the effects on the three reported properties of the (joint) distribution of CEI and LEI, when the prior parameter concerned is increased. To determine the effects on the dependence we use the standard correlation coefficient.

"shocks" to the chain. Moreover, it may be difficult to choose sensible values for $\boldsymbol{\lambda}$ *a priori*. We follow Müller *et al.* (1996) and choose for a prior on the two parameters of the Gaussian part and the matrix of the inverted Wishart component in (5.3). The first consists of a normal-inverted Gamma–2 distribution, and for the second part we choose a Wishart distribution. Thus, we specify this prior as

$$B_\mu \sim \mathcal{IG}2(\nu_0, S_0), \quad \mathbf{b}_\mu \,|\, B_\mu \sim \mathcal{N}(\mathbf{b}_0, B_\mu \mathbf{B}_0), \tag{5.8}$$

$$\mathbf{S}_\Sigma \sim \mathcal{W}(\nu_w, \mathbf{S}_w). \tag{5.9}$$

The components of this prior are conditionally conjugate with the base distribution. With adding levels to the model it becomes increasingly more difficult to set the parameters of the prior distributions. Therefore we adopt the approach by Geweke (2010) and use a type of prior predictive analysis to set the prior's parameters in (5.8)–(5.9). This results in a number of guidelines which we show in Table 5.1. Following these rules of thumb, we choose the parameter values for our empirical analysis as reported in the last line of this table.

5.3.3 Posterior simulation

When we have observed the data $\mathbf{Y}$, we use Markov chain Monte Carlo (MCMC) methods to obtain a sample from the posterior distribution of $\{\boldsymbol{\theta}, \boldsymbol{\lambda}, \alpha\}$. Neal (2000) lists a number of MCMC algorithms for DPM models to simulate the elements of $\boldsymbol{\theta}$. We implement his routine for conjugate models as, for instance, Escobar and West (1995), Bush and MacEachern (1996) and Campolieti (2001) do in their applications.

The MCMC simulation procedure consists of the four following steps.

Step 1. Sample each $\boldsymbol{\theta}_t$, $(t = 1, \ldots, T)$ from its full conditional posterior

$$\begin{aligned} p(\boldsymbol{\theta}_t \,|\, \mathbf{Y}, \boldsymbol{\theta}_{-t}, \boldsymbol{\lambda}, \alpha) &\propto p(\mathbf{y}_t \,|\, \boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t \,|\, \boldsymbol{\theta}_{-t}, \boldsymbol{\lambda}, \alpha) \\ &\propto \alpha p_0(\mathbf{y}_t; \boldsymbol{\lambda}) p(\boldsymbol{\theta}_t \,|\, \mathbf{y}_t, \boldsymbol{\lambda}) + \sum_{s \neq t} p(\mathbf{y}_t \,|\, \boldsymbol{\theta}_s) \delta(\boldsymbol{\theta}_t - \boldsymbol{\theta}_s), \end{aligned}$$

for which we use (5.2), and the prior exchangeability of $\boldsymbol{\theta}_t$ to obtain its full conditional prior. $p(\boldsymbol{\theta}_t \,|\, \mathbf{y}_t, \boldsymbol{\lambda}) \propto p(\mathbf{y}_t \,|\, \boldsymbol{\theta}_t) f_0(\boldsymbol{\theta}_t; \boldsymbol{\lambda})$ is the one-observation posterior under the base prior;

Step 2. After one run of Step 1, Neal (2000) shows that the single-move sampler is substantially improved in terms of mixing speed if we add a remix step. We have I_T^* unique values in $\boldsymbol{\theta}$ and the corresponding configuration of data. If $\mathbf{Y}_{(j)}$ contains all observations for which it holds that $\boldsymbol{\theta}_t = \boldsymbol{\theta}_j^*$, we simulate each unique parameter value ($j = 1, \ldots, I_T^*$) from

$$p(\boldsymbol{\theta}_j^* \,|\, \mathbf{Y}_{(j)}, \boldsymbol{\lambda}) \propto f_0(\boldsymbol{\theta}_j^*; \boldsymbol{\lambda}) \prod_{t:\boldsymbol{\theta}_t = \boldsymbol{\theta}_j^*} p(\mathbf{y}_t \,|\, \boldsymbol{\theta}_j^*),$$

which is the multi-observation posterior with data $\mathbf{Y}_{(j)}$ under the base prior;

Step 3. We sample $\boldsymbol{\lambda}$ as in any hierarchical model (Rossi *et al.*, 2005), noticing that each of the unique values $\boldsymbol{\theta}_j^*$ is generated from the base prior. Hence, we sample from

$$p(\boldsymbol{\lambda} \,|\, \mathbf{Y}, \boldsymbol{\theta}) \propto p(\boldsymbol{\lambda}) \prod_{j=1}^{I_T^*} f_0(\boldsymbol{\theta}_j^*; \boldsymbol{\lambda}),$$

with the prior as in (5.8)–(5.9). We sample all three hyperparameters in one block;

Step 4. With the prior in (5.7), Escobar and West (1995) show that the smoothness parameter α can be sampled from a two-component mixture of Gamma distributions once an auxiliary variable has been sampled from a Beta distribution. We use their approach.

In Appendix 5.A.1 we show that all necessary components for Steps 1–2 are available analytically and we provide expressions. Moreover, we derive the distributions to sample from in Step 3, and give details for the implementation of Step 4.

5.3.4 Model comparison

To compare the nonparametric model to parametric specifications, we follow the Bayesian literature (Kass and Raftery, 1995) and compute the respective marginal likelihoods and the implied Bayes factors. The implemention of the marginal likelihood methods of Chib (1995) or Chib and Jeliazkov (2001) in the nonparametric model involves an analytically intractable integral to evaluate a likelihood ordinate. Basu and Chib (2003) provide the tools, based on collapsed sequential importance sampling, to estimate the likelihood ordinate and we implement their routine.

First we compute the marginal likelihood $p(\mathbf{Y}) = \int p(\mathbf{Y} \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \boldsymbol{\lambda}, \alpha) p(\boldsymbol{\lambda}, \alpha)\, \mathsf{d}\{\boldsymbol{\theta}, \boldsymbol{\lambda}, \alpha\}$ and compare it to the Gaussian model's which is the special case under the restriction $\alpha = 0$. We note that with this comparison we consider all statistical properties at once (since we use the whole joint distribution), and test for any deviation from (joint) normality of the (CEI,LEI) data.

Second, because central economic interest lies in the future state of economic activity given the predictive leading indicator, we also need $p(\mathbf{y} \,|\, \mathbf{x})$, the marginal likelihood of the nonparametric model for CEI. To compute it we use the relation

$$p(\mathbf{y} \,|\, \mathbf{x}, A) = p(\mathbf{Y} \,|\, A) / p(\mathbf{x} \,|\, A), \tag{5.10}$$

in which we condition on the model assumptions A explicitly to denote that both parts of this ratio are to be computed under the same set of conditions. That is, the *multivariate* model specification in (5.1)–(5.3) and the priors (5.7)–(5.9). The numerator in (5.10) is the output of Basu and Chib's (2003) method as before. The denominator is the "marginalized" marginal likelihood of the multivariate $\mathbf{Y}$ with respect to $\mathbf{y}$, *i.e.*, $p(\mathbf{x} \mid A) = \int p(\mathbf{Y} \mid A)\,\mathsf{d}\mathbf{y}$.

Using the sampling distribution in (5.1), we integrate each y_t in the first model level and obtain the marginalized likelihood contributions of the predictor

$$p(x_t \mid \boldsymbol{\theta}_t) = \int p(\mathbf{y}_t \mid \boldsymbol{\theta}_t)\,\mathsf{d}y_t = \phi(x_t; \mu_{2,t}, \Sigma_{22,t}) = p(x_t \mid \boldsymbol{\theta}_{2,t}), \quad \boldsymbol{\theta}_{2,t} = (\mu_{2,t}, \Sigma_{22,t})'.$$

This shows that there is no *direct* sample information for parameters $\boldsymbol{\theta}_{1,t} = (\mu_{1,t}, \Sigma_{11,t}, \Sigma_{12,t})'$ if we only observe $\mathbf{x}$. These parameters drop from the model's first level with the result that we can integrate them analytically in the second level. We integrate out all $\boldsymbol{\theta}_{1,t}$ from (5.2), starting with the final $\boldsymbol{\theta}_{1,T}$, and obtain the conditional densities $p(\boldsymbol{\theta}_{2,t} \mid \boldsymbol{\theta}^{1,t-1}) = \int p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}^{1,t-1})\,\mathsf{d}\boldsymbol{\theta}_{1,t} = p(\boldsymbol{\theta}_{2,t} \mid \boldsymbol{\theta}_2^{1,t-1})$. Thus we arrive at

$$p(\boldsymbol{\theta}_{2,t} \mid \boldsymbol{\theta}_2^{1,t-1}) \propto \alpha f_{0,2}(\boldsymbol{\theta}_{2,t}; \boldsymbol{\lambda}) + \sum_{s=1}^{t-1} \delta(\boldsymbol{\theta}_{2,t} - \boldsymbol{\theta}_{2,s}), \quad (t = 2, \ldots, T),$$

and $\boldsymbol{\theta}_{2,1} \sim F_{0,2}(\boldsymbol{\lambda})$. The second marginal of our base prior distribution ($F_{0,2}$) has known form and only depends on hyperparameters $\boldsymbol{\lambda}_2 = \{b_{\mu,2}, B_\mu, S_{\Sigma,22}\}$. In the third level of the model (5.8)–(5.9), we integrate the hyperparameters $\boldsymbol{\lambda}_1$ to arrive at $p(\boldsymbol{\lambda}_2) = \int p(\boldsymbol{\lambda})\,\mathsf{d}\boldsymbol{\lambda}_1$. To summarize, "everything" involving the marginal parameters of y_t and the parameters governing the jointness of y_t and x_t is integrated out analytically and we end up with a marginal counterpart of the multivariate model.

This marginal model's sampling distribution is given by

$$x_t \mid \boldsymbol{\theta}_{2,t} \overset{ind.}{\sim} \mathcal{N}(\mu_{2,t}, \Sigma_{22,t}), \quad (t = 1, \ldots, T),$$

in which the parameters $\boldsymbol{\theta}_{2,t}$ are the result from a Dirichlet process with smoothness parameter α and base prior with density $f_{0,2}$. Using properties of the normal and the inverted Wishart distribution, this base distribution is specified as

$$F_{0,2}(\boldsymbol{\lambda}_2): \quad \mu_2 \mid \Sigma_{22} \sim \mathcal{N}(b_{\mu,2}, B_\mu \Sigma_{22}), \quad \Sigma_{22} \sim \mathcal{IW}(\nu_\Sigma - 1, S_{\Sigma,22}) = \mathcal{IG}2(\nu_\Sigma - 1, S_{\Sigma,22}).$$

The third level completes the marginal model and provides the prior $p(\boldsymbol{\lambda}_2)$ for the hyperparameters of $F_{0,2}$. This prior consists of the components

$$\begin{aligned} &B_\mu \sim \mathcal{IG}2(\nu_0, S_0), \quad b_{\mu,2} \mid B_\mu \sim \mathcal{N}(b_{0,2}, B_\mu B_{0,22}), \\ &S_{\Sigma,22} \sim \mathcal{W}(\nu_w, S_{w,22}) = \mathcal{G}a(\nu_w/2, S_{w,22}^{-1}/2). \end{aligned}$$

By applying the MCMC routines and the DPM marginal likelihood methods to this (univariate) model specification we obtain $p(\mathbf{x} \mid A)$, the denominator in (5.10). In Appendix 5.B.1 we provide the details of our marginal likelihood computations.

5.3.5 Threshold regressions

The nonparametric method has proven to be a powerful tool to estimate an unknown (nonstandard) distribution in various empirical applications. On the other hand, typical features such as nonlinearity are not explicitly modeled, but they are implicitly accounted for. From an economic point of view it is important to identify which of these features are present in the nonparametrically estimated distribution. We compare the nonparametric model to parametric model specifications which take the isolated features of interest explicitly into account. Hence, these alternative models mainly serve as instruments for inference, more than as serious competitors to the nonparametric specification.

In Section 5.3.1 we derived the nonparametrically estimated distribution of economic activity conditional on the leading indicator. In general, the statistical properties of CEI therefore vary with the value of LEI. For example, in case of a nonlinear predictive relation, a decline in the leading index can contain a stronger predictive signal about economic activity than when LEI is positive. Threshold-regression models (Tong and Lim, 1980) account for nonstandard relations between predictor and target variable in a similar fashion, though in a parametric manner. In particular, in a threshold regression the value of the leading index determines the parameter values that link LEI to CEI. This contrasts with Markov-switching models in which an unobserved process governs the relation between predictor and target variable, and which are therefore less suited for an interpretable comparison to the nonparametric model.

We specify the relation between economic activity and the leading indicator as

$$y_t = \boldsymbol{\beta}_t{}'\mathbf{x}_t + \sigma_t\varepsilon_t, \quad \mathbf{x}_t = (1, x_t)', \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad (t = 1, \ldots, T). \tag{5.11}$$

The most general form of the threshold regression we use is the above specification with two conditional-mean and two conditional-variance regimes,

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_1\mathbb{I}_{\{x_t\leq\gamma_1\}} + \boldsymbol{\beta}_2\mathbb{I}_{\{x_t>\gamma_1\}}, \qquad \sigma_t^2 = \sigma_1^2\mathbb{I}_{\{x_t\leq\gamma_2\}} + \sigma_2^2\mathbb{I}_{\{x_t>\gamma_2\}}. \tag{5.12}$$

This model is substantially restrictive if the predictive relation is highly nonlinear, which is however unlikely to be the case with macroeconomic data. If nonlinear or heteroskedastic features, of whatever form, are in the data, we expect the Bayes factor to favor the threshold regression compared to the linear, homoskedastic regression. Applying the threshold regressions in (5.11)–(5.12), we check for typical econometric features as nonlinearity, heteroskedasticity, and nonnormality of economic shocks as follows.

With the homoskedasticity restriction $\sigma_1 = \sigma_2 \equiv \sigma$, we first examine any nonlinearities in the relation between the coincident and leading index. Second, the linearity restriction $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 \equiv \boldsymbol{\beta}$ serves to check whether the conditional variance of economic activity varies with the value of LEI. Finally, if the nonparametric model is favored over the threshold model without any of the two restrictions, we can attribute this to either more complex nonlinear/heteroskedastic relations, which the threshold regression can only partly describe, or, the presence of infrequent but large economic shocks which are insufficiently supported by the Gaussian distribution.

We complete the threshold-regression model specification by formulating our prior distribution for its parameters. We consider the four regime parameters *a priori* statistically independent with the conditionally conjugate distributions

$$\boldsymbol{\beta}_j \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{b}_\beta, \mathbf{B}_\beta), \quad \sigma_j^2 \overset{i.i.d.}{\sim} \mathcal{IG}2(\nu_\sigma, S_\sigma), \quad (j = 1, 2). \tag{5.13}$$

For the two threshold parameters we specifiy the two independent discrete prior distributions

$$\Pr\Big[\gamma_j = \underline{\gamma}_{j,s}\Big] = p_{\gamma_j,s}, \quad (j = 1, 2;\ s = 1, \ldots, S), \tag{5.14}$$

with uniform support for the hundred equally-spaced points in the sample range of LEI, disregarding the five smallest and the five largest LEI values.[3]

To ensure a fair comparison of the parametric models to the nonparametric one, such that *a priori* none is given preference to, we follow Florens *et al.* (1996). Their principle prescribes that with just one observation it should be impossible to distinguish between the two models. From Section 5.3.1 we know the prior belief of an observation (given the leading index), $p(y_1 \,|\, x_1) = \int p_0(y_1 \,|\, x_1; \boldsymbol{\lambda}) p(\boldsymbol{\lambda})\, \mathsf{d}\boldsymbol{\lambda}$, in which the first density is the conditional Student's t given by $y_t \,|\, \{x_t, \boldsymbol{\lambda}\} \sim \mathcal{T}(\mathbf{b}'_{y|x}\mathbf{x}_t, S_{y|x_t}, \nu_\Sigma)$.[4] In the threshold-regression setting we have a conditonal for y_1 that looks like, but is not Student's t. That is, integrating the (first-layer) model parameters we obtain $p(y_t \,|\, x_t, \boldsymbol{\psi}) = \int \phi(y_t; \mathbf{b}'_\beta \mathbf{x}_t,\, \sigma^2[\mathbf{x}_t{}'\mathbf{B}_\beta \mathbf{x}_t/\sigma^2 + 1]) p(\sigma^2)\, \mathsf{d}\sigma^2$.

In order to establish a close match between the two prior beliefs, we first also put a prior on the threshold-regression model's hyperparameters $\boldsymbol{\psi} = \{\mathbf{b}_\beta, \mathbf{B}_\beta, S_\sigma\}$. We take conditionally conjugate distributions consisting of a normal-inverted Wishart for the parameters of the normal distribution, and a Gamma distribution for the inverted Gamma–2's parameter in (5.13), *i.e.*,

$$\mathbf{b}_\beta \,|\, \mathbf{B}_\beta \sim \mathcal{N}(\underline{\mathbf{b}}, \underline{B}\mathbf{B}_\beta), \quad \mathbf{B}_\beta \sim \mathcal{IW}(\underline{\nu}, \underline{\mathbf{S}}), \quad S_\sigma \sim \mathcal{G}a(\underline{a}_S/2,\, \underline{b}_S^{-1}/2). \tag{5.15}$$

Second, we use Monte Carlo simulation to set the parameters in (5.15) such that the first two prior moments of y_t and its dependence with x_t match the nonparametric model's. To (approximately) match higher moments, we also calibrate such that the prior tail probability $\mathsf{Pr}\,[y_t < -10]$ is the same in both models.[5]

We implement a Gibbs sampler with only two blocks to sample the threshold-regression model's parameters from their posterior distribution. For both the regression part $\boldsymbol{\omega}_1 = \{\gamma_1, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$ and the conditional-variance part $\boldsymbol{\omega}_2 = \{\gamma_2, \sigma_1^2, \sigma_2^2\}$ we sample the threshold and the two regime parameters in one block. We alternately sample from $p(\boldsymbol{\omega}_1 \,|\, \mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\omega}_2)$ and $p(\boldsymbol{\omega}_2 \,|\, \mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\omega}_1)$. Given the values for these regime parameters, we sample hyperparameters $\boldsymbol{\psi}$ in one block from $p(\boldsymbol{\psi} \,|\, \mathbf{y}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$. We refer to Appendix 5.A.2 for derivations and distributions. We obtain marginal likelihoods with the method by Chib (1995), of which we give details in Appendix 5.B.3.

5.4 Results

This section is divided into two parts. In Section 5.4.1 we describe the outcomes of the model comparison to reveal any nonstandard features of economic activity. We discuss the posterior results obtained with the nonparametric model in Section 5.4.2.

5.4.1 Testing nonstandard features

First we compare the estimated nonparametric distribution for the bivariate vector $\mathbf{y}_t$ to the multivariate Gaussian which we attain by taking $\alpha = 0$ in the DPM setting. In Panel A of

---

[3]In practice only the sampled x_t, $(t = 1, \ldots, T)$ influence the likelihood, and point mass at $x_{(t)}$ is uniformly distributed over $(x_{(t)}, x_{(t+1)})$; any two points strictly in between two consecutive order statistics of the explanatory variable are equally likely.

[4]We use the result that the conditionals of a multivariate Student's t distribution are also Student's t. In this case we have $\mathbf{b}_{y|x} = (b_{\mu,1} - b_{\mu,2}S_{\Sigma,12}/S_{\Sigma,22},\ S_{\Sigma,12}/S_{\Sigma,22})'$, $S_{y|x_t} = \{1 + [(x_t - b_{\mu,2})^2/S_{\Sigma,22} - 1]/\nu_\Sigma\}(S_{\Sigma,11} - S_{\Sigma,12}^2/S_{\Sigma,22})$.

[5]This procedure leads to $\underline{\mathbf{b}} = (0, 1/10)'$, $\underline{B} = 1.5$, $\underline{\mathbf{S}} = 4\mathbf{I}_2$, $\underline{\nu} = 16$, $\nu_\sigma = 12$, $\underline{a}_S = 200$, and $\underline{b}_S = 2$.

Table 5.2 *Marginal likelihoods and Bayes factors nonparametric joint models*

	Lag leading indicator							
	$\ell = 1$		$\ell = 3$		$\ell = 6$		$\ell = 12$	
Model	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$
Panel A: Revised data								
M_1: Prior 1	−1 595.8	–	−1 590.4	–	−1 597.0	–	−1 614.4	–
M_2: $\alpha = 0$	−1 627.5	31.7	−1 621.8	31.4	−1 632.2	35.2	−1 650.1	35.7
M_3: Prior 2	−1 595.1	−0.7	−1 589.6	−0.7	−1 596.4	−0.6	−1 613.7	−0.7
Panel B: Real-time data								
M_1: Prior 1	−1 527.5	–	−1 532.9	–	−1 534.9	–	−1 533.7	–
M_2: $\alpha = 0$	−1 544.7	17.3	−1 547.2	14.3	−1 550.5	15.6	−1 547.6	13.9
M_3: Prior 2	−1 526.9	−0.5	−1 532.2	−0.7	−1 534.3	−0.7	−1 533.0	−0.7

Notes: The table contains the marginal likelihoods ("M.L.") and Bayes factors on a $\log_{10}$-scale. Three models for the multivariate observations $\mathbf{y}_t = (y_t, x_t)'$ are considered, and the reported marginal likelihoods are $\log_{10} p(\mathbf{Y} \mid M_i)$. Bayes factors are computed with M_1 as benchmark, the mulivariate nonparametric model with the prior on α such that $\mathsf{mode}\,[I_T^* \mid M_1] = 5$. M_2 is the nested case with $\alpha = 0$, which reduces to a multivariate Gaussian model. M_3 is the nonparametric model with prior $\mathsf{mode}\,[I_T^* \mid M_3] = 10$ and incorporates more mixture components *a priori*.

Table 5.2 we report marginal likelihoods and Bayes factors with respect to the model with the "moderate" prior on α (M_1), for varying lags of the fully revised predictor variable LEI. For all four lag values we find Bayes factors greater than 30, which forms overwhelming evidence in favor of joint nonnormality of the bivariate data. Using additional support for more mixture components with the second prior on the smoothness parameter, we find substantial evidence in favor of that model (M_3) compared to M_1, with Bayes factors about 0.7 as reported in the third line of Panel A.

We display results for real-time data in Panel B. Again, the nonparametric model is decisively preferred to the multivariate Gaussian model, with Bayes factors in the range 14–17. Here as well, *a priori* allowing for the selection of more mixture components (M_3) gives even a substantially better fit of the data. We note that we cannot compare statistics between the two panels. Though CEI is the same in both, LEI is generally not. These first results clearly show that the two variables are certainly not jointly Gaussian.

For the output in Table 5.2, both the type of cross-variable dependence and the form of the two marginal distributions are taken into account at once. To exploit the predictive relation between CEI and LEI for forecasting purposes, we shift attention to the conditional distribution of CEI given LEI. In Table 5.3 we report the marginal likelihoods of the nonparametric conditional model for economic activity to examine the predictive qualities of the leading indicator. The conditional variant of model M_2 is the classical linear Gaussian regression model of CEI given LEI. That is, the conditional mean of the coincident index always varies linearly with the value of the leading index, and the innovations have constant variance and are generated from a Gaussian distribution.

The first two lines of Panel A in Table 5.3 give the marginal likelihoods and Bayes factors of M_1 with respect to the linear regression model. The Bayes factors are greater than 8 for LEI lagging one or three months, which provides decisive evidence in favor of the

Table 5.3 *Marginal likelihoods and Bayes factors nonparametric conditional models*

	Lag leading indicator							
	$\ell = 1$		$\ell = 3$		$\ell = 6$		$\ell = 12$	
Model	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$
Panel A: Revised data								
M_1: Prior 1	−698.7	–	−692.8	–	−699.2	–	−714.9	–
M_2: $\alpha = 0$	−707.3	8.7	−701.3	8.5	−711.4	12.2	−727.4	12.5
M_3: Prior 2	−698.4	−0.3	−692.5	−0.3	−699.1	−0.1	−714.7	−0.3
Panel B: Real-time data								
M_1: Prior 1	−719.8	–	−717.0	–	−720.7	–	−723.9	–
M_2: $\alpha = 0$	−732.6	12.8	−727.9	10.9	−731.4	10.7	−732.6	8.7
M_3: Prior 2	−719.7	−0.2	−716.6	−0.4	−720.4	−0.3	−723.6	−0.2

Notes: The table contains the marginal likelihoods ("M.L.") and Bayes factors on a $\log_{10}$-scale. Three models for CEI (y_t) given LEI (x_t) are considered, and the reported marginal likelihoods are $\log_{10} p(\mathbf{y} \mid \mathbf{x}, M_i) = \log_{10} p(\mathbf{Y} \mid M_i) - \log_{10} p(\mathbf{x} \mid M_i)$. Bayes factors are computed with M_1 as benchmark. See the notes accompanying Table 5.2 for definitions of the various models.

nonparametric model. For lags six and twelve months, the Bayes factors are even greater than 12, virtually nullifying the weight of the normal linear regression model in a model-averaging prediction. The third entry of Panel A shows that we find weak preference for the more flexible nonparametric model M_3, though for $\ell = 6$ almost none. This outcome suggests that the necessary additional mixture components we found in Table 5.2 are mainly required to better fit the marginal distribution of LEI. Panel B of Table 5.3 depicts the output when we use the real-time variant of the leading index instead. Bayes factors evidently opt again for the nonparametric model, though with increasing lags of LEI, the factors are now decreasing. The comparison of M_1 to M_3 gives similar results as when we use fully revised data.

Since the dependent variable is the same for all settings in Table 5.3, we can infer with which predictor we describe CEI best. The leading index is most strongly correlated with economic activity when it is lagged three months (see also Chapter 3). This holds both for the completely revised data in Panel A and the partly revised LEI in Panel B. Keeping the lag of LEI constant, we see that the fully revised explanatory variable always results in a greater marginal likelihood. Per lag the Bayes factors range from 9 to 20 (not shown in the table, but simply the difference between the marginal likelihoods of the models with fully revised and real-time LEI). Furthermore, the worst-performing lag of twelve months with revised data still outclasses the best-predicting real-time LEI ($\ell = 3$) with a Bayes factor of 2.1.

From the results sofar we can safely reject the linear Gaussian regression model for CEI in favor of the nonparametric model. Since the latter automatically, but implicitly, incorporates three potentially essential characteristics, *i.e.*, nonlinearity, heteroskedasticity and nonnormality, we apply the threshold-regression models from Section 5.3.5 to find out which features determine this manifest rejection. In Table 5.4 we show the marginal likelihoods and Bayes factors of M_1, the nonparametric conditional model, with respect to the threshold regressions. We first check for a nonlinear relation between LEI and CEI with $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, and

Table 5.4 *Marginal likelihoods and Bayes factors threshold-regression models*

	Lag leading indicator							
	$\ell = 1$		$\ell = 3$		$\ell = 6$		$\ell = 12$	
Model	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$	M.L.	$BF_{1,i}$
Panel A: Revised data								
M_4: Non-lin.; Hom.	−706.8	8.1	−701.8	9.0	−712.5	13.3	−728.2	13.2
M_5: Lin.; Het.	−706.2	7.6	−701.5	8.7	−707.5	8.3	−722.2	7.2
M_6: Non-lin.; Het.	−705.3	6.6	−701.2	8.4	−707.8	8.6	−722.8	7.8
Panel B: Real-time data								
M_4: Non-lin.; Hom.	−732.3	12.4	−726.5	9.5	−731.9	11.2	−733.2	9.3
M_5: Lin.; Het.	−729.5	9.6	−726.5	9.5	−731.6	10.9	−733.3	9.4
M_6: Non-lin.; Het.	−728.7	8.8	−725.4	8.5	−731.6	10.9	−733.5	9.6

Notes: The table contains the marginal likelihoods ("M.L.") and Bayes factors on a $\log_{10}$-scale. The three models are variants of the general threshold-regression model: M_4 is the nonlinear homoskedastic threshold regression ($\sigma_1 = \sigma_2$); M_5 is the linear heteroskedastic threshold regression ($\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$); M_6 is the general threshold regression allowing for both nonlinearity and heteroskedasticity. Bayes factors are computed with M_1 as benchmark, the nonparametric model with marginal likelihoods as in Table 5.3.

we consider the errors to be homoskedastic. The first entries of Panels A and B give the results for the revised and real-time leading index, respectively. We compare these Bayes factors to the ones of M_2 (the linear, homoskedastic regression model) in Table 5.3, since both are with respect to M_1.

With the latest-available one-month-lagged leading index we obtain a Bayes factor of 0.6 in favor of the nonlinear threshold-regression model. For the other three lags we weakly ($\ell = 3$) to substantially ($\ell = 6, 12$) prefer the linear model, with factors for M_4 relative to M_2 equal to −0.5, −1.1 and −0.7 respectively. In Panel B, for the partly revised LEI, Bayes factors indicate a substantial nonlinear relation for short-term lags of LEI of one and three months. A leading index lagged six and twelve months does not exhibit a nonlinear relation to CEI, with Bayes factors of 0.5 and 0.6 in favor of M_2.

Next we check whether a LEI-varying conditional variance of CEI is a reason to reject the normal regression model. Model M_5 in Table 5.4 denotes the threshold-regression model with linear conditional mean, but $\sigma_1 \neq \sigma_2$. In Panel A we observe strong evidence ($BF_{5,2} = 1.1$) for heteroskedasticity of this type when we take the leading index from the previous month as explanatory variable. Lags of six and twelve months result in $BF_{5,2} = 3.9; 5.3$ respectively, providing decisive support for two distinct variance regimes. For $\ell = 3$ however, we only find weak evidence in favor of homoskedastic errors. With real-time predictor values we obtain Bayes factors of 3.2 and 1.4 for one- and three-month lags (Panel B). With LEI lagging half a year or a complete year though, allowing for different variance regimes does not improve the statistical description of the conincident index' growth rate.

Finally, we apply the threshold-regression model without any kind of restriction (M_6). Incorporating both nonlinearity and heteroskedasticity we find the following pattern. For the two most moderate lag lengths, Bayes factors increase (substantially to strongly) compared two M_5. However, for the two longer horizons, this additional flexibility is mildly penalized in terms of marginal likelihoods, because its improvement in describing CEI is not substantial

Table 5.5 *Posterior properties parameters threshold-regression models*

Mod.	Regression part			Conditional-variance part		
	β_1	β_2	γ_1	σ_1^2	σ_2^2	γ_2
Panel A: Revised data						
M_4	0.28	0.13	−6.9	14.4	–	–
	(0.22; 0.36)	(0.07; 0.19)	(−13.9; 7.4)	(13.1; 15.9)		
M_5	0.19	–	–	37.1	13.8	−15.0
	(0.16; 0.23)			(24.0; 54.6)	(12.5; 15.3)	(−19.0; −12.1)
M_6	0.28	0.13	−6.6	33.9	13.8	−15.8
	(0.19; 0.37)	(0.06; 0.19)	(−14.4; 8.3)	(21.4; 51.2)	(12.4; 15.1)	(−20.3; −11.7)
Panel B: Real-time data						
M_4	0.22	0.03	6.3	17.7	–	–
	(0.11; 0.31)	(−0.09; 0.14)	(−1.3; 13.6)	(16.0; 19.4)		
M_5	0.08	–	–	36.8	15.7	−3.4
	(0.03; 0.14)			(27.8; 48.1)	(14.1; 17.4)	(−4.1; − 2.6)
M_6	0.21	0.02	6.3	35.4	15.4	−3.4
	(0.09; 0.33)	(−0.09; 0.13)	(−1.3; 14.0)	(26.7; 46.2)	(13.9; 17.1)	(−4.1; − 2.6)

Notes: Per model ("Mod.") the table reports the posterior means, and (between parentheses) the 5th and 95th percentiles of the marginal posterior distributions. See the notes of Table 5.4 for the definitions of the three applied threshold regressions. The leading indicator is lagged one month ($\ell = 1$). We only report results for the regression parameter associated with the leading indicator.

enough. Since LEI is generally applied as a short-term predictor of business-cycle variables, this result is not surprising. The composite leading index tends to have only weak predictive power for horizons greater than a half year and considering more rich functional relations between CEI and LEI cannot change this.

In Table 5.5 we report posterior properties of the parameters in the three threshold-regression models. In Panels A and B we show posterior means and percentiles of the marginal posterior distributions for latest-available and real-time LEI, both lagged one month. The nonlinear homoskedastic model reveals two regression regimes with mutually exclusive posterior density regions for the two parameters.[6] The effect of LEI on CEI is twice as large in the first regime relative to the second. The marginal posterior of the threshold value indicates substantial uncertainty about its value, though most posterior mass is located at negative growth of the leading index.[7] Therefore, the leading index contains a stronger signal for anticipating downturns than for an expanding economy. Applying the real-time predictor shows a similar asymmetric effect on CEI, and in this case LEI hardly has any predictive power for CEI once it has reached values above γ_1. Moreover, the threshold value

[6] We note that we have not imposed any inequality restrictions on the regression (nor on the variance) parameters, which are often applied in mixture models with an unobserved switching variable to counter their label-switching problem.

[7] The relatively widespread posterior support for γ_1 can also point at a more smooth transition between the two regression regimes.

is considerably greater compared to its value in Panel A. We attribute this to the fact that the real-time LEI attains substantially greater values than the revised, especially in advance of a recession, as we also see in Figure 5.1.

If we apply the linear but heteroskedastic model, we find two very distinct variance regimes. The threshold value is in the range $[-19; -12]$, and if LEI drops below γ_2, the conditional variance of CEI more than doubles. In Panel B we also find such results for the leading index measured in real time, in the entry of M_5. The uncertainty about the location of the regime change is strongly diminished though, compared to the application of the revised LEI, with γ_2 about -3.5. Here as well, for the same reason, the estimated threshold value is substantially greater in the real-time case. These outcomes are in line with the stylized fact that a business-cycle variable tends to exhibit increased volatility during recessionary periods.

With model M_6 we incorporate both nonlinearity and heteroskedasticity, and we display its parameters' posterior properties in the final entries of the two panels in Table 5.5. Posterior supported values of all six parameters are very similar for the two types of predictor variables compared to the previously discussed nested cases. Especially with respect to the conditional variance of CEI it is important to allow for different regimes. Keeping σ^2 constant over the sample space of LEI causes a serious underestimation of volatility of the economic activity variable, principally in times when we are particularly interested in what the future state of the economy will be, that is, when LEI signifies a downturn.

Finally, we compare the nonparametric model to the threshold model without restrictions. The Bayes factors $BF_{1,6}$ in the third lines of the two panels in Table 5.4 indicate that despite incorporating nonlinearity and heteroskedasticity with model M_6, the nonparametric model still decisively outclasses the threshold-regression model, regardless of the number of months LEI is lagged. Therefore, in addition to the nonstandard relation between the coincident and the leading index as displayed in the first two moments of the conditional distribution, we have firm evidence for assuming further nonstandard distributional properties, with non-Gaussian economic shocks in particular.

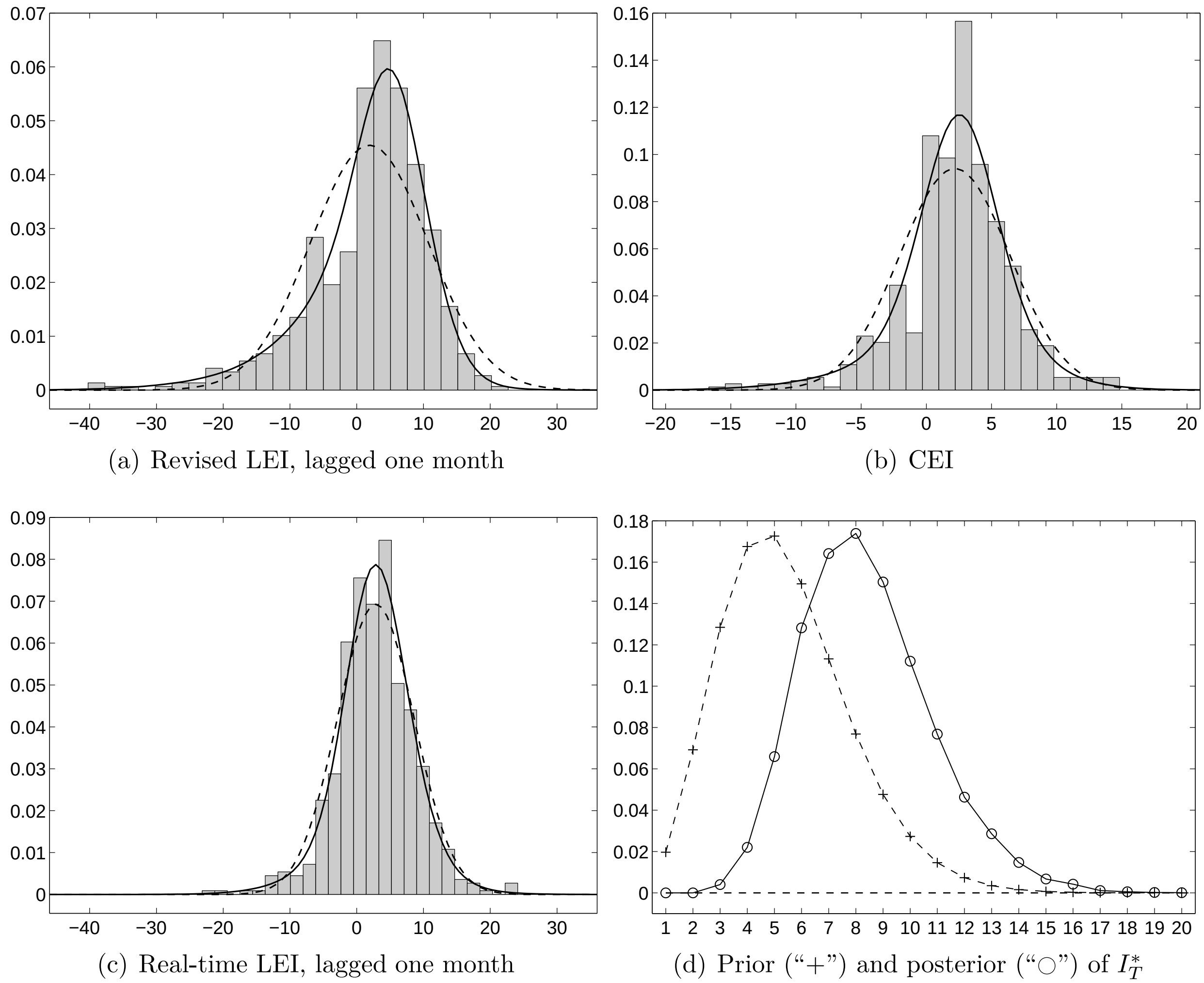
To summarize the results of our model comparison, we find evidence for nonlinearity, the most emphatic when we apply short-term lags of LEI of one to three months. Second, heteroskedasticity is undoubtedly present in completely revised data and when we apply the shorter lag lengths of the real-time LEI. Regarding further distributional assumptions, both predictor types and all values of ℓ strongly direct towards nonnormal shocks.

5.4.2 Posterior results

In this section we discuss the posterior results of the nonparametric model for economic activity conditional on the leading indicator. First we look at the marginal posterior predictive distributions of the two macroeconomic variables which result from the Bayesian joint density smoothing.

In Figure 5.2 we show the histograms of the two series of data and the marginal predictive densities $p(z_{T+1} \mid \mathbf{Y})$, $(z = y, x)$ (solid graphs). Figure 5.2(a) provides visual evidence for clear rejection of normality of the fully revised LEI growth rates. For constructing the statistics in Table 5.3 we used the marginal likelihoods of the joint model and the marginal LEI model, which means that we have available the Bayes factor of the marginal nonparametric LEI model against the Gaussian's. With a value of $\log_{10} p(\mathbf{x} \mid M_1) - \log_{10} p(\mathbf{x} \mid M_2) = 23.0$ the Bayes factor clearly rejects the Gaussian distribution for LEI in favor of the nonparametric distribution. The long and heavy left tail is severely underestimated by fitting a Gaussian

Figure 5.2 *Properties marginal posterior predictive distributions*



(a) Revised LEI, lagged one month

(b) CEI

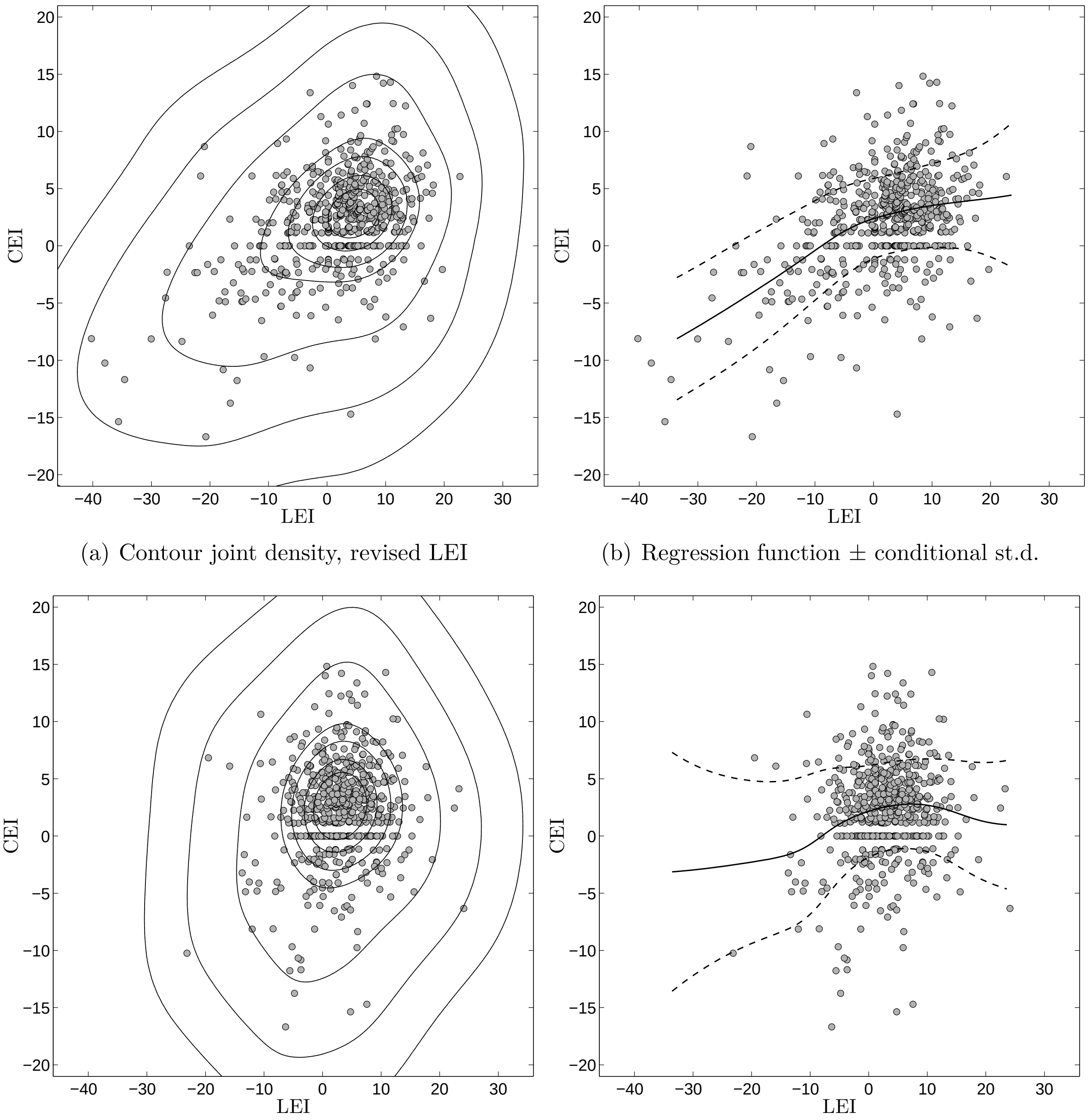
(c) Real-time LEI, lagged one month

(d) Prior ("+") and posterior ("○") of I_T^*

distribution, as we see in the figure (dashed graph). The marginal predictive density of the real-time leading index is reproduced in 5.2(c) and we immediately observe the thinner left tail. Although we have a number of large negative growth rates up to -20 for the real-time LEI, the Bayes factor is, with a value of 4.5, considerably smaller than for its latest-available series, but still undoubtedly in favor of a non-Gaussian distribution.

From Figures 5.2(a) and 5.2(c) we also remark the profound revision the leading index is liable to. Particularly, initial negative LEI growth, anticipating a recessionary period, is generally adjusted further downward during the revision process. We also remind TCB's recently updated methodology to partly explain the fatter left tail of revised LEI. The adjusted methods result in *ex-post* more extreme decreases in the leading index, which better anticipate the two latest recessions of 2001 and 2008. For example, in the six months leading up to the 2001 recession LEI before the adjustment showed a decrease of nearly 3%, whereas with the new methodology it was almost 7%. Likewise for 2008, according to the original composition the leading index experienced a modest 1% decline during July–December 2008, while in the same months the new LEI decreased 4%.

Figure 5.3 *Properties joint and conditional posterior predictive distributions*



(a) Contour joint density, revised LEI

(b) Regression function ± conditional st.d.

(c) Contour joint density, real-time LEI

(d) Regression function ± conditional st.d.

We plot the marginal posterior predictive density of CEI in Figure 5.2(b). Again we notice extra peakedness and the more support for negative values compared to the fitted normal distribution. This nonnormality of the marginal distribution is partially attributable to the location-scale mixture resulting from the nonlinear and heteroskedastic relation with its predictor LEI, and, as we saw in Section 5.4.1, to non-Gaussian economic shocks.

Prior and posterior probabilities of the number of normal mixture components used to estimate the unknown joint distribution of (CEI,LEI) are displayed in Figure 5.2(d). These probabilities are associated with the “modest” prior on α which results in a prior mode at

5 for I_T^*. After updating with the data a considerable number of additional components are required, shifting probability mass to the right with a posterior mode at 8. This figure visually sustains our findings from Table 5.2, in which we saw the substantially greater support in the data for the more flexible prior on the smoothness parameter. To describe the characteristics of the relation between CEI and LEI, the topic we turn to next, the two priors lead to very similar posterior results though.

In Figures 5.3(a) and (c) we show the contours of the joint posterior predictive density $p(\mathbf{y}_{T+1} \,|\, \mathbf{Y})$, which we derived in (5.4), superimposed upon the scatter plots of the latest-available and real-time data, respectively. If we look at the former we notice a long joint left tail in the negative quadrant. On the other hand, for positive values of both LEI and CEI data are more randomly scattered.

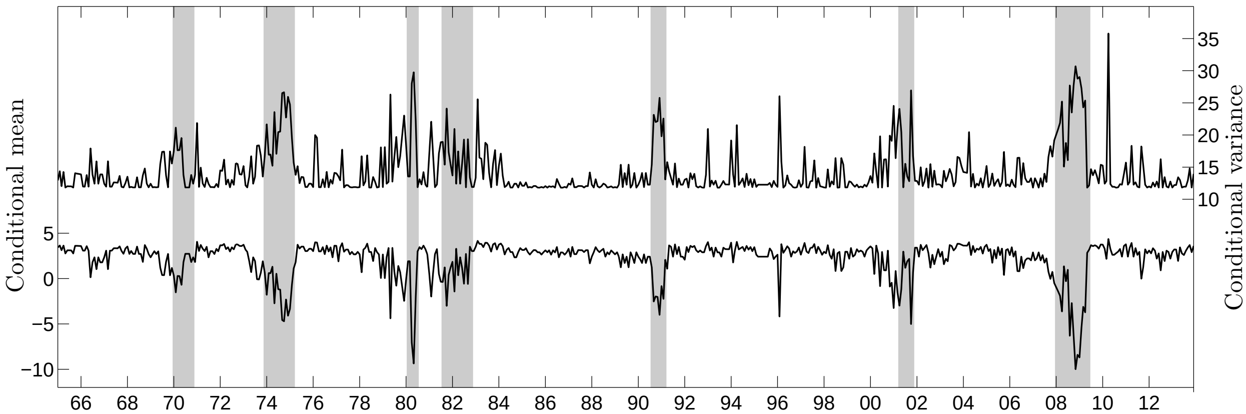
We further illustrate these two outcomes in Figure 5.3(b), which depicts the regression function (solid graph) of the coincident-index growth rate given last month's growth rate of the leading index. We obtain this nonparametric relation by applying (5.6) with the identity function $f_1(y_{T+1}) = y_{T+1}$. The dashed lines represent the one-standard-deviation deviations of the conditional mean, *i.e.*, $\bar{f}_1(x) \pm [\bar{f}_2(x) - \bar{f}_1^2(x)]^{1/2}$, with $f_2(y_{T+1}) = y_{T+1}^2$. The relation between the two variables obviously changes depending upon the value of the leading index. If LEI is negative (certainly when smaller than -5) its relation with CEI is stronger. That is, the conditional mean of CEI reacts more vehemently to changes in LEI compared to positive leading-index growth. The regession function becomes less steep once LEI is greater than zero, with its slope halved. Moreover, a negative LEI almost surely leads to below-average growth of CEI, whereas a positive LEI merely coincides with a random variation of CEI about its expansion average of 3. These results form a generalization (a smoother transition between regimes) of our findings with the nonlinear threshold regressions in Panel A of Table 5.5.

If we use initial releases of the leading index, we obtain the LEI-CEI relation as displayed in Figure 5.3(d). We find strong similarities with the results in Table 5.5 (Panel B). Although the relation is stronger too when the leading index decreases, if LEI turns positive, the relation effectively breaks down. We further remark the infrequent joint negative events compared to the figures with revised data. It is this outcome that makes the real-time leading index considerably inferior to the revised LEI in terms of forecasting economic activity.

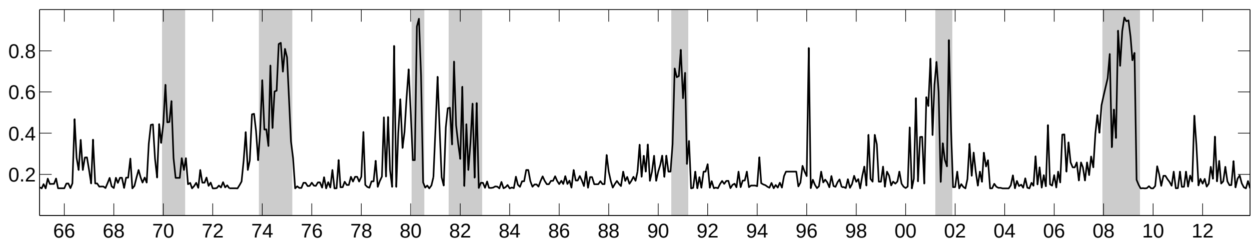
In Figure 5.4 we show the time-series graphs of properties of the predictive distribtution of CEI given LEI. Figure 5.4(a) depicts the evolution of its conditional mean and variance over time. The plot of the latter clearly exhibits the importance of incorporating heteroskedasticity. During expansionary periods, when the leading-index growth is positive and steady, the variance moves in the range $[12, 15]$, whereas during recessions it attains peaks of 25 and above. From the graph displaying the predictive first moment of CEI's growth rate we also observe different behavior over the business cycle. Obvious downturns in LEI prelude large negative growth of CEI. On the other hand, when the leading index grows, next month's CEI will be positive as well, with values between 2 and 3, but the magnitude of LEI is significantly less important. This result straightforwardly sprouts from the nonlinear regression we saw previously, with a less strong relation for positive LEI.

Figure 5.4(b) reproduces the probability of negative CEI given LEI. That is, we apply (5.6) with $f_3(y_t) = \mathbb{I}_{\{y_t \leq 0\}}$, which amounts to the graphed $\bar{f}_3(x_t)$, $(t = 1, \ldots, T)$. Recessions are accompanied with large predictive probabilities, and regardless the degree of positive LEI growth, there always is a substantial probability slightly below 0.2 of a decline in the coincident index. This outcome stems from the estimated heavier-than-Gaussian left tail

Figure 5.4 *Properties CEI given revised LEI, January 1965–December 2013*



(a) Conditional mean and variance



(b) Predictive probability negative CEI

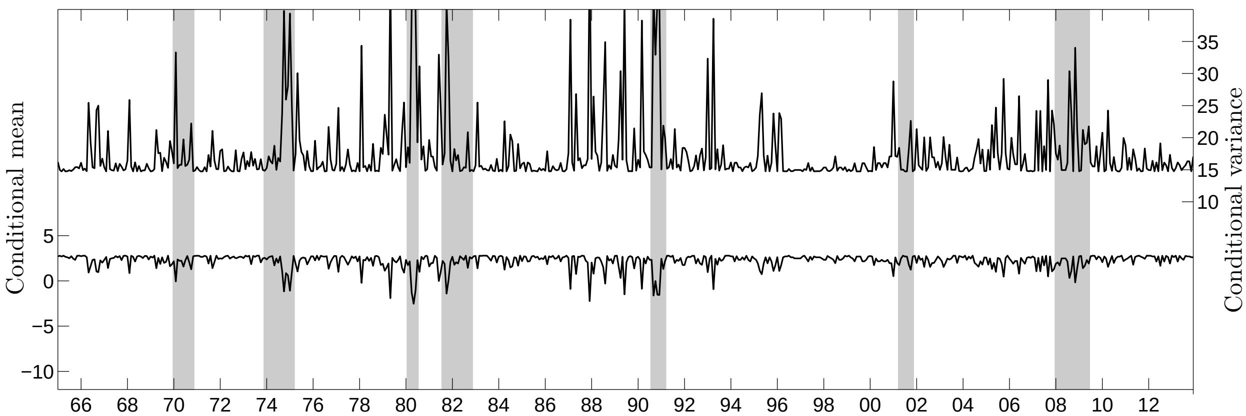
of CEI's distribution. The computed probabilities could serve as the input to recession-probability forecasting. For example, with similar models for CEI h months ahead, we could estimate probabilities of events like "consecutive negative growth of CEI in the next h months."

Figure 5.5 displays the real-time versions of the properties of CEI's predictive distribution. The differences are immediately clear if we compare with the revised LEI. We already saw the absent events of joint large negative growth rates. Although the predictive mean of CEI declines for the first five recessions in our sample, the two most recent contractions in 2001 and 2008 are not well anticipated by the real-time leading index. The probability of negative CEI growth is also less pronounced. From 2006 up to 2008 and during the recession, we do find increased volatility. Clear is the inferior forecasting performance of the real-time LEI, which is in line with the findings of Diebold and Rudebusch (1991a). To conclude, for both predictor types all three properties of future economic activity are strongly affected by the leading index, an essential result which is ignored in normal linear regression models.

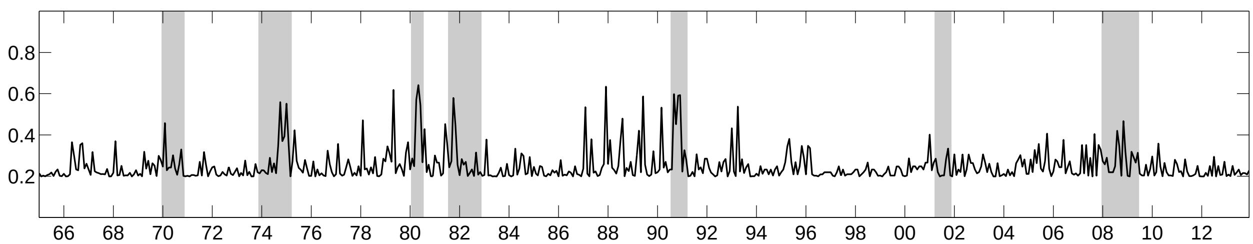
5.5 Conclusion

For decades, identifying economic variables that lead the state of economic activity has formed a substantial part of the empirical macroeconomics literature. Having available a set of such variables, it is still a challenge to formulate a statistical model that adequately relates them to the business cycle fluctuations and, subsequently, obtain a practically useful

Figure 5.5 *Properties CEI given real-time LEI, January 1965–December 2013*



(a) Conditional mean and variance



(b) Predictive probability negative CEI

specification for short-term forecasting. For example, the strength of the relation may be dependent upon the state of the business cycle. Moreover, not only the location of the forecast is important to accurately determine, but other statistical properties such as heteroskedasticity and peakedness of the variable to forecast are relevant as well, especially if the forecaster is particularly interested in uncertainty or more extreme changes in economic activity. We find empirical evidence sustaining the assumptions of nonstandard distributional properties of economic activity.

With a Bayesian nonparametric method based on multivariate Dirichlet process mixtures we model the joint distribution of economic activity and a leading indicator. We use The Conference Board's coincident-index and the lagged leading-index growth rates, respectively. The joint distribution is estimated as a mixture of multivariate normal distributions with an unknown number of mixture components. First we perform a model comparison using Bayes factors to examine potential non-Gaussianity of this joint distribution. Next we compute the estimated conditional distribution which serves as the basis for producing forecasts. We suggest a procedure to compute the marginal likelihoods of the nonparametric conditional models. In turn, these marginal likelihoods serve as inputs to a comparison with threshold-regression models in order to find out what particular statistical properties of economic activity are nonstandard.

We apply both latest-available and real-time data of TCB's leading index and for both we find decisive support for nonnormality of the joint distribution. If we "decompose" the nonparametric conditional distribution by comparing it to threshold-regression models with different regimes for the conditional mean and the conditional variance, we find that a substantial part of the nonnormality can be attributed to nonlinear dependence between economic activity and the leading index, and to a conditional variance that depends on the value of the predictive variable. More in particular, the predictive relation is stronger when the leading index signals a downturn, and uncertainty, as measured by the predictive variance, increases under the same circumstances. With real-time data we establish a similar pattern, with the notable difference that the leading index is only very modestly informative about the expected growth rate of economic activity when it indicates an expansion. Finally, empirical evidence decisively supports the assumption of nonnormal economic shocks, which is especially important to forecasters with a loss function different from the standard squared loss type.

Further research should reveal the potential gains from using the Bayesian nonparametric methodology for forecasting economic activity and recessionary periods. Especially in settings in which the forecaster is particularly concerned with rare events or uncertainty. Moreover, instead of using a single (composite) leading indicator we should examine the incorporation of more individual leading indicators, possibly with unknown lead times. Multivariate extensions of our bivariate setting soon start to suffer from the *curse of dimensionality* once the number of leading indicators increases. Bayesian nonparametric regression methods (for example as applied in Salimans, 2012) extended to cope with heteroskedasticity and non-Gaussian errors could provide a way out. Augmenting the model with explicit dynamics of the dependent variable could further improve forecasting results.

5.A Posterior simulation steps

In this appendix we provide a detailed overview of the computational steps to obtain a sample from the posterior distribution for the various models. In Section 5.3.1 we discuss the multivariate nonparametric model and in Section 5.3.5 the threshold-regression models.

5.A.1 Multivariate nonparametric model

We model a vector $\mathbf{y}_t$, $(t = 1, \dots, T)$, of K unrestricted random variables. The special univariate case evolves naturally by taking $K = 1$. Each vector is sampled from its own multivariate normal distribution parameterized by $\boldsymbol{\theta}_t$. $\boldsymbol{\theta}_t = (\boldsymbol{\mu}_t', \mathsf{vec}\,[\boldsymbol{\Sigma}_t]')'$ consist of a vector of means and a covariance matrix. The DPM approach implies that some of the T parameters equal the same value with positive probability. Each of these unique values is independently sampled from the base distribution with density f_0. The MCMC scheme in Section 5.3.3 consists of four steps.

Step 1: Individual parameters

We derive the one-observation marginal likelihood $p_0(\mathbf{y}_t; \boldsymbol{\lambda}) = \int p(\mathbf{y}_t \,|\, \boldsymbol{\theta}) f_0(\boldsymbol{\theta}; \boldsymbol{\lambda})\, \mathsf{d}\boldsymbol{\theta}$, and the one-observation posterior density $p(\boldsymbol{\theta}_t \,|\, \mathbf{y}_t, \boldsymbol{\lambda})$. We start with the latter.

For each t, with a probability proportional to $\alpha p_0(\mathbf{y}_t; \boldsymbol{\lambda})$ we sample $\boldsymbol{\theta}_t$ from $p(\boldsymbol{\theta}_t \,|\, \mathbf{y}_t, \boldsymbol{\lambda}) \propto p(\mathbf{y}_t \,|\, \boldsymbol{\theta}_t) f_0(\boldsymbol{\theta}_t; \boldsymbol{\lambda})$. This is the posterior counterpart of the prior in (5.3) (see, for example, Geweke, 2005, and Appendix B) and consists of

$$\boldsymbol{\Sigma}_t \,|\, \{\mathbf{y}_t, \boldsymbol{\lambda}\} \sim \mathcal{IW}(\nu_\Sigma + 1, \overline{\mathbf{S}}), \quad \boldsymbol{\mu}_t \,|\, \{\mathbf{y}_t, \boldsymbol{\lambda}, \boldsymbol{\Sigma}_t\} \sim \mathcal{N}(\overline{\mathbf{b}}, B_\mu (B_\mu + 1)^{-1} \boldsymbol{\Sigma}_t). \tag{5.A.1}$$

The mean of the Gaussian distribution equals $\overline{\mathbf{b}} = (B_\mu + 1)^{-1}(B_\mu \mathbf{y}_t + \mathbf{b}_\mu)$, and the matrix of the inverted Wishart distribution is $\overline{\mathbf{S}} = \mathbf{S}_\Sigma + (\mathbf{y}_t - \overline{\mathbf{b}})(\mathbf{y}_t - \overline{\mathbf{b}})' + B_\mu^{-1}(\overline{\mathbf{b}} - \mathbf{b}_\mu)(\overline{\mathbf{b}} - \mathbf{b}_\mu)' = \mathbf{S}_\Sigma + (1 + B_\mu)^{-1}(\mathbf{y}_t - \mathbf{b}_\mu)(\mathbf{y}_t - \mathbf{b}_\mu)'$.

To obtain the one-observation marginal likelihood we use this distribution, and first integrate the Gaussian kernel and then the inverted Wishart's. Accounting the integrating constants, we obtain the marginal likelihood

$$\begin{aligned} p_0(\mathbf{y}_t; \boldsymbol{\lambda}) = {} & \Gamma_K((\nu_\Sigma + 1)/2)/\Gamma_K(\nu_\Sigma/2)\pi^{-K/2}(1 + B_\mu)^{-K/2}|\mathbf{S}_\Sigma|^{\nu_\Sigma/2} \\ & \times |\mathbf{S}_\Sigma + (1 + B_\mu)^{-1}(\mathbf{y}_t - \mathbf{b}_\mu)(\mathbf{y}_t - \mathbf{b}_\mu)'|^{-(\nu_\Sigma+1)/2}. \end{aligned} \tag{5.A.2}$$

With the Cholesky decomposition $\mathbf{S}_\Sigma^{-1} = \mathbf{L}'\mathbf{L}$ we rewrite the latter determinant in (5.A.2) as $|\mathbf{S}_\Sigma||\mathbf{I}_K + (1 + B_\mu)^{-1}\mathbf{L}(\mathbf{y}_t - \mathbf{b}_\mu)(\mathbf{y}_t - \mathbf{b}_\mu)'\mathbf{L}'|$. We apply the computationally more efficient expression of the marginal likelihood which becomes

$$\begin{aligned} p_0(\mathbf{y}_t; \boldsymbol{\lambda}) = {} & \Gamma((\nu_\Sigma + 1)/2)/\Gamma((\nu_\Sigma - K + 1)/2)\pi^{-K/2}(1 + B_\mu)^{-K/2}|\mathbf{S}_\Sigma|^{-1/2} \\ & \times \left[1 + (1 + B_\mu)^{-1}(\mathbf{y}_t - \mathbf{b}_\mu)'\mathbf{S}_\Sigma^{-1}(\mathbf{y}_t - \mathbf{b}_\mu)\right]^{-(\nu_\Sigma+1)/2}, \end{aligned} \tag{5.A.3}$$

because $|\mathbf{I}_K + \mathbf{x}\mathbf{x}'| = 1 + \mathbf{x}'\mathbf{x}$. This is the density of the K-variate Student's t distribution $\mathcal{T}(\mathbf{b}_\mu, (1 + B_\mu)(\nu_\Sigma + 1 - K)^{-1}\mathbf{S}_\Sigma,\ \nu_\Sigma + 1 - K)$.

Step 2: Unique parameters

The distributions for this remix step are the multi-observation equivalents of (5.A.1). That is, if the configuration splits the data into subsamples $\mathbf{Y}_{(i)}$, $(i = 1, \ldots, I_T^*)$, each of length T_i, then each of the I_T^* mixture components' parameters has distribution

$$\begin{aligned} &\boldsymbol{\Sigma}_i^* \,|\, \{\mathbf{Y}_{(i)}, \boldsymbol{\lambda}\} \sim \mathcal{IW}(\nu_\Sigma + T_i, \overline{\mathbf{S}}), \\ &\boldsymbol{\mu}_i^* \,|\, \{\mathbf{Y}_{(i)}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}_i^*\} \sim \mathcal{N}(\overline{\mathbf{b}}, B_\mu(T_i B_\mu + 1)^{-1}\boldsymbol{\Sigma}_i^*), \end{aligned} \tag{5.A.4}$$

with parameters

$$\begin{aligned} \overline{\mathbf{b}} &= (T_i B_\mu + 1)^{-1}(B_\mu \mathbf{Y}_{(i)}' \boldsymbol{\iota}_{T_i} + \mathbf{b}_\mu), \\ \overline{\mathbf{S}} &= \mathbf{S}_\Sigma + (\mathbf{Y}_{(i)} - \boldsymbol{\iota}_{T_i}\overline{\mathbf{b}}')'(\mathbf{Y}_{(i)} - \boldsymbol{\iota}_{T_i}\overline{\mathbf{b}}') + B_\mu^{-1}(\overline{\mathbf{b}} - \mathbf{b}_\mu)(\overline{\mathbf{b}} - \mathbf{b}_\mu)' \\ &= \mathbf{S}_\Sigma + (\mathbf{Y}_{(i)} - \boldsymbol{\iota}_{T_i}\mathbf{b}_\mu')'(\mathbf{I}_{T_i} + B_\mu \boldsymbol{\iota}_{T_i \times T_i})^{-1}(\mathbf{Y}_{(i)} - \boldsymbol{\iota}_{T_i}\mathbf{b}_\mu'). \end{aligned}$$

The last equation provides more insight into the inverted Wishart's scale matrix, whereas the penultimate expression is computationally more attractive due to the lack of the large-matrix inversion.

Step 3: Base distribution's hyperparameters

Each of the I_T^* unique parameters $\boldsymbol{\theta}_i^*$ resulting from Steps 1–2 is an independent realization of the base distribution given the latter's hyperparameters $\boldsymbol{\lambda} = \{\mathbf{b}_\mu, B_\mu, \mathbf{S}_\Sigma, \nu_\Sigma\}$, *i.e.*,

$$\boldsymbol{\Sigma}_i^* \,|\, \boldsymbol{\lambda} \overset{i.i.d.}{\sim} \mathcal{IW}(\nu_\Sigma, \mathbf{S}_\Sigma), \quad \boldsymbol{\mu}_i^* \,|\, \{\boldsymbol{\lambda}, \boldsymbol{\Sigma}_i^*\} \overset{ind.}{\sim} \mathcal{N}(\mathbf{b}_\mu, B_\mu \boldsymbol{\Sigma}_i^*), \quad (i = 1, \ldots, I_T^*).$$

With the prior in (5.8) we obtain a (conditionally) conjugate "likelihood"-prior pair. Conditional on the configuration of the data we have "observations" $\boldsymbol{\mu}^* = \{\boldsymbol{\mu}_1^*, \dots, \boldsymbol{\mu}_{I_T^*}^*\}$ and $\boldsymbol{\Sigma}^* = \{\boldsymbol{\Sigma}_1^*, \dots, \boldsymbol{\Sigma}_{I_T^*}^*\}$. In order to obtain the distributions from which to simulate $\{B_\mu, \mathbf{b}_\mu\}$, we cast the setting in a normal regression framework. Interpreting the prior as an additional observation we have the $I_T^* + 1$ equations

$$\boldsymbol{\mu}_i^* = \mathbf{b}_\mu + \boldsymbol{\Sigma}_i^{*\frac{1}{2}} \boldsymbol{\eta}_i, \quad (i = 1, \dots, I_T^*), \quad \mathbf{b}_0 = \mathbf{b}_\mu + \mathbf{B}_0^{\frac{1}{2}} \boldsymbol{\eta}_0,$$

in which the errors are $\boldsymbol{\eta}_i \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}_K, B_\mu \mathbf{I}_K)$, $(i = 0, 1, \dots, I_T^*)$. We scale these equations to get $K(I_T^* + 1)$ observations with zero covariance. We define the auxiliary variables

$$\mathbf{v}_i = \boldsymbol{\Sigma}_i^{*-\frac{1}{2}} \boldsymbol{\mu}_i, \quad \mathbf{W}_i = \boldsymbol{\Sigma}_i^{*-\frac{1}{2}}, \quad (i = 1, \dots, I_T^*), \quad \mathbf{v}_0 = \mathbf{B}_0^{-\frac{1}{2}} \mathbf{b}_0, \quad \mathbf{W}_0 = \mathbf{B}_0^{-\frac{1}{2}},$$

and stack them as $\mathbf{v}' = (\mathbf{v}_0', \mathbf{v}_1', \dots, \mathbf{v}_{I_T^*}')$ and $\mathbf{W}' = (\mathbf{W}_0', \mathbf{W}_1', \dots, \mathbf{W}_{I_T^*}')$. With this set-up we sample the two parameters in one block from the joint full conditional posterior, which is decomposed as

$$\begin{aligned} B_\mu \,|\, \{\mathbf{Y}, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*\} &\sim \mathcal{IG}2(\nu_0 + K I_T^*,\, S_0 + (\mathbf{v} - \mathbf{W}\overline{\mathbf{b}})'(\mathbf{v} - \mathbf{W}\overline{\mathbf{b}})), \\ \mathbf{b}_\mu \,|\, \{\mathbf{Y}, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, B_\mu\} &\sim \mathcal{N}(\overline{\mathbf{b}}, B_\mu (\mathbf{W}'\mathbf{W})^{-1}), \end{aligned} \tag{5.A.5}$$

in which $\overline{\mathbf{b}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{v} = (\sum_i \boldsymbol{\Sigma}_i^{*-1} + \mathbf{B}_0^{-1})^{-1}(\sum_i \boldsymbol{\Sigma}_i^{*-1} \boldsymbol{\mu}_i^* + \mathbf{B}_0^{-1}\mathbf{b}_0)$.

To sample the hyperparameter of the inverted Wishart part of the base distribution, we use the prior on $\mathbf{S}_\Sigma$ in (5.9). Its full conditional posterior has known form and is independent of the other hyperparameters, hence all elements of $\boldsymbol{\lambda}$ susceptible to updating are sampled at once. The density of the full conditional posterior becomes

$$\begin{aligned} p(\mathbf{S}_\Sigma \,|\, \mathbf{Y}, \boldsymbol{\Sigma}^*) \propto p(\boldsymbol{\Sigma}^* \,|\, \mathbf{S}_\Sigma) p(\mathbf{S}_\Sigma) \propto |\mathbf{S}_\Sigma|^{\nu_\Sigma I_T^*/2} \exp\Big\{ - \mathsf{tr}\Big[\Big(\textstyle\sum_i \boldsymbol{\Sigma}_i^{*-1}\Big)\mathbf{S}_\Sigma\Big] /2\Big\} \\ \times |\mathbf{S}_\Sigma|^{(\nu_w - K - 1)/2} \exp\Big\{ - \mathsf{tr}\Big[\mathbf{S}_\Sigma \mathbf{S}_w^{-1}\Big] /2\Big\}. \end{aligned}$$

Thus, we sample from the Wishart distribution

$$\mathbf{S}_\Sigma \,|\, \{\mathbf{Y}, \boldsymbol{\Sigma}^*\} \sim \mathcal{W}\Big(I_T^* \nu_\Sigma + \nu_w,\, \Big(\textstyle\sum_i \boldsymbol{\Sigma}_i^{*-1} + \mathbf{S}_w^{-1}\Big)^{-1}\Big). \tag{5.A.6}$$

Step 4: Smoothness parameter

With the Gamma prior distribution for α in (5.7), Escobar and West (1995) show that its full conditional posterior only depends upon the number of unique mixture components. They sample α using the Gibbs principle by applying a data-augmentation step. We follow their procedure and first simulate the auxiliary variable $\eta \,|\, \{\mathbf{Y}, \boldsymbol{\theta}^*, \boldsymbol{\lambda}, \alpha\} = \eta \,|\, \{\mathbf{Y}, I_T^*, \alpha\} \sim \mathcal{B}e(\alpha + 1, T)$. Second, we simulate the smoothness parameter from the distribution with density $p(\alpha \,|\, \mathbf{Y}, \boldsymbol{\theta}^*, \boldsymbol{\lambda}, \eta) = p(\alpha \,|\, \mathbf{Y}, I_T^*, \eta)$, which is the two-component mixture of Gamma distributions

$$\begin{aligned} \alpha \,|\, \{\mathbf{Y}, I_T^*, \eta\} \sim\; & w \mathcal{G}a(a_\alpha + I_T^*,\, b_\alpha - \log \eta) \\ & + (1 - w)\mathcal{G}a(a_\alpha + I_T^* - 1,\, b_\alpha - \log \eta), \end{aligned} \tag{5.A.7}$$

with mixture weight $w = (a_\alpha + I_T^* - 1)/[a_\alpha + I_T^* - 1 + T(b_\alpha - \log \eta)]$.

5.A.2 Threshold-regression models

To simulate from the posterior distribution in the various threshold-regression models of Section 5.3.5 we implement a three-step Gibbs sampler by simulating $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$ and $\boldsymbol{\psi}$ from their full conditional posterior distributions. In this section we discuss technical details of the three steps.

Step 1: Regression part's parameters

We draw the three regression-part parameters in one block as follows. Conditional on $\boldsymbol{\omega}_2$ and $\boldsymbol{\psi}$ we obtain the unit-variance regression model $y_t/\sigma_t = \sigma_t^{-1}\mathbf{x}_t{}'\boldsymbol{\beta}_t + \varepsilon_t \Rightarrow v_t = \mathbf{w}_t{}'\boldsymbol{\beta}_t + \varepsilon_t$, with the respective priors on the regression coefficients and the threshold parameter as in (5.13) and (5.14). For each supported γ_1 we split up the sample according to whether $x_t > \gamma_1$ and create the two subsamples $\{\mathbf{v}_1(\gamma_1), \mathbf{W}_1(\gamma_1)\} = \{(v_t, \mathbf{w}_t), t : x_t \leq \gamma_1\}$ and $\{\mathbf{v}_2(\gamma_1), \mathbf{W}_2(\gamma_1)\} = \{(v_t, \mathbf{w}_t), t : x_t > \gamma_1\}$. With these definitions the full conditional posterior density of $\boldsymbol{\omega}_1 = \{\gamma_1, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$ is

$$\begin{aligned} p(\boldsymbol{\omega}_1 \,|\, \mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\omega}_2) \propto \exp\Big\{ -\textstyle\sum_{j=1}^{2} (\mathbf{v}_j(\gamma_1) - \mathbf{W}_j(\gamma_1)\boldsymbol{\beta}_j)'(\mathbf{v}_j(\gamma_1) - \mathbf{W}_j(\gamma_1)\boldsymbol{\beta}_j)/2 \Big\} \\ \times \exp\Big\{ -\textstyle\sum_{j=1}^{2} (\boldsymbol{\beta}_j - \mathbf{b}_\beta)'\mathbf{B}_\beta^{-1}(\boldsymbol{\beta}_j - \mathbf{b}_\beta)/2 \Big\} \cdot p(\gamma_1). \end{aligned}$$

We analytically integrate the regression parameters of the two regimes and obtain the marginal pdf of the threshold parameter

$$\begin{aligned} p(\gamma_1 \,|\, \mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\omega}_2) \propto \Big(\textstyle\sum_{s=1}^{S} \mathbb{I}_{\{\gamma_1 = \underline{\gamma}_{1,s}\}} p_{\gamma_{1,s}} \Big) \prod_{j=1}^{2} |\mathbf{W}_j(\gamma_1)'\mathbf{W}_j(\gamma_1) + \mathbf{B}_\beta^{-1}|^{-\frac{1}{2}} \\ \times \exp\Big\{ -\textstyle\sum_{j=1}^{2} \Big[(\mathbf{v}_j(\gamma_1) - \mathbf{W}_j(\gamma_1)\bar{\mathbf{b}}_j)'(\mathbf{v}_j(\gamma_1) - \mathbf{W}_j(\gamma_1)\bar{\mathbf{b}}_j) \\ + (\bar{\mathbf{b}}_j - \mathbf{b}_\beta)'\mathbf{B}_\beta^{-1}(\bar{\mathbf{b}}_j - \mathbf{b}_\beta) \Big]/2 \Big\}, \end{aligned} \tag{5.A.8}$$

with $\bar{\mathbf{b}}_j = (\mathbf{W}_j(\gamma_1)'\mathbf{W}_j(\gamma_1) + \mathbf{B}_\beta^{-1})^{-1}(\mathbf{W}_j(\gamma_1)'\mathbf{v}_j(\gamma_1) + \mathbf{B}_\beta^{-1}\mathbf{b}_\beta)$. Since the two regression parameters are conditionally independent we obtain (suppressing the conditioning on $\mathbf{y}$ and $\boldsymbol{\psi}$) $p(\boldsymbol{\omega}_1) = p(\gamma_1)p(\boldsymbol{\beta}_1 \,|\, \gamma_1)p(\boldsymbol{\beta}_2 \,|\, \gamma_1)$. We first sample γ_1 from the distribution given in (5.A.8), and next we simulate

$$\boldsymbol{\beta}_j \,|\, \{\mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\omega}_2, \gamma_1\} \overset{ind.}{\sim} \mathcal{N}(\bar{\mathbf{b}}_j, (\mathbf{W}_j(\gamma_1)'\mathbf{W}_j(\gamma_1) + \mathbf{B}_\beta^{-1})^{-1}), \quad (j = 1, 2). \tag{5.A.9}$$

Step 2: Conditional-variance part's parameters

In a similar fashion we derive the one-block-simulation step of the conditional-variance parameters. Conditional on $\boldsymbol{\omega}_1$ and $\boldsymbol{\psi}$ we write $v_t \equiv y_t - \mathbf{x}_t{}'\boldsymbol{\beta}_t = \sigma_t\varepsilon_t$. We split up the sample into the two subsamples $\mathbf{v}_1(\gamma_2)$ and $\mathbf{v}_2(\gamma_2)$; if $x_t \leq \gamma_2$ the corresponding v_t ends up in $\mathbf{v}_1$, otherwise it is assigned to the second subsample. Each subsample has length T_j, $T_1 + T_2 = T$. The conditional posterior density of $\boldsymbol{\omega}_2$ becomes

$$\begin{aligned} p(\boldsymbol{\omega}_2 \,|\, \mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\omega}_1) \propto (\sigma_1^2)^{-T_1/2}(\sigma_2^2)^{-T_2/2} \exp\Big\{ -[\mathbf{v}_1(\gamma_2)'\mathbf{v}_1(\gamma_2)/\sigma_1^2 + \mathbf{v}_2(\gamma_2)'\mathbf{v}_2(\gamma_2)/\sigma_2^2]/2 \Big\} \\ \times (\sigma_1^1\sigma_2^2)^{-(\nu_\sigma+2)/2} \exp\Big\{ -2^{-1}S_\sigma(\sigma_1^{-2} + \sigma_2^{-2}) \Big\} \cdot p(\gamma_2). \end{aligned}$$

We integrate out the variances after identifying the two kernels of the respective inverted Gamma–2 distributions and we obtain the threshold parameter's marginal density

$$p(\gamma_2 \,|\, \mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\omega}_1) \propto \Gamma((T_1 + \nu_\sigma)/2)[\mathbf{v}_1(\gamma_2)'\mathbf{v}_1(\gamma_2) + S_\sigma]^{-(T_1+\nu_\sigma)/2} \\ \times \Gamma((T_2 + \nu_\sigma)/2)[\mathbf{v}_2(\gamma_2)'\mathbf{v}_2(\gamma_2) + S_\sigma]^{-(T_2+\nu_\sigma)/2} \sum_{s=1}^{S} \mathbb{I}_{\{\gamma_2 = \underline{\gamma}_{2,s}\}} p_{\gamma_2,s}. \tag{5.A.10}$$

The two variances are conditionally independent and we simulate in one block using the decomposition $p(\boldsymbol{\omega}_2) = p(\gamma_2)p(\sigma_1^2 \,|\, \gamma_2)p(\sigma_2^2 \,|\, \gamma_2)$. First, we simulate the threshold parameter using (5.A.10) and, next, we simulate

$$\sigma_j^2 \,|\, \{\mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\omega}_1, \gamma_2\} \overset{ind.}{\sim} \mathcal{IG}2(T_j + \nu_\sigma, \mathbf{v}_j(\gamma_2)'\mathbf{v}_j(\gamma_2) + S_\sigma), \quad (j = 1, 2). \tag{5.A.11}$$

Posterior simulation for the threshold regressions with equality restrictions is a special case of the preceding. In such restricted cases the threshold concerned becomes a nuisance parameter and its posterior equals its prior.

Step 3: Second-level parameters

The final step of the MCMC routine involves the one-block sampling of the second-level parameters $\boldsymbol{\psi}$. $\{\mathbf{b}_\beta, \mathbf{B}_\beta\}$ and S_σ are conditionally independent and we use $p(\boldsymbol{\psi} \,|\, \mathbf{y}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = p(\mathbf{b}_\beta, \mathbf{B}_\beta \,|\, \mathbf{y}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)p(S_\sigma \,|\, \mathbf{y}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$. We start with the first pdf of this decomposition.

Depending on the imposed equality restrictions, we have $1 \leq T_\beta \leq 2$ independent draws for the regression parameters as in (5.13), which we collect in matrix $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{T_\beta})$. Combined with the prior in (5.15) we cast this submodel in the conjugate multivariate regression form. Thus, we sample in one block by simulating

$$\mathbf{B}_\beta \,|\, \{\mathbf{y}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2\} \sim \mathcal{IW}(\underline{\nu} + T_\beta, \underline{\mathbf{S}} + (\boldsymbol{\beta} - \underline{\mathbf{b}}\boldsymbol{\iota}_{T_\beta}')(\mathbf{I}_{T_\beta} + \underline{B}\boldsymbol{\iota}_{T_\beta}\boldsymbol{\iota}_{T_\beta}{}')^{-1}(\boldsymbol{\beta} - \underline{\mathbf{b}}\boldsymbol{\iota}_{T_\beta}')'), \tag{5.A.12}$$

$$\mathbf{b}_\beta \,|\, \{\mathbf{y}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \mathbf{B}_\beta\} \sim \mathcal{N}(\overline{\mathbf{b}}_\beta, (T_\beta + \underline{B}^{-1})^{-1}\mathbf{B}_\beta), \tag{5.A.13}$$

in which $\overline{\mathbf{b}}_\beta = (T_\beta + \underline{B}^{-1})^{-1}(\boldsymbol{\beta}\boldsymbol{\iota}_{T_\beta} + \underline{B}^{-1}\underline{\mathbf{b}})$.

To sample S_σ we combine the prior belief in (5.15) with the $1 \leq T_\sigma \leq 2$ "observations" σ_j^2 in (5.13). This results in the conditional posterior density

$$p(S_\sigma \,|\, \mathbf{y}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \propto (S_\sigma^{\nu_\sigma/2})^{T_\sigma} \exp\Big\{ - S_\sigma \sum_{j=1}^{T_\sigma} \sigma_j^{-2}/2 \Big\} S_\sigma^{\underline{a}_S/2-1} \exp\Big\{ -S_\sigma \underline{b}_S^{-1}/2 \Big\}.$$

We simulate from a Gamma distribution as follows,

$$S_\sigma \,|\, \boldsymbol{\omega}_2 \sim \mathcal{G}a\Big((T_\sigma\nu_\sigma + \underline{a}_S)/2, \Big(\sum_{j=1}^{T_\sigma} \sigma_j^{-2} + \underline{b}_S^{-1}\Big)/2\Big). \tag{5.A.14}$$

5.B Marginal likelihood computations

We discuss the derivations for and implementation of the collapsed sequential importance sampling procedure by Basu and Chib (2003) to compute the marginal likelihood for the nonparametric model in Section 5.B.1. In Sections 5.B.2 and 5.B.3 we consider the computations of the marginal likelihoods for the benchmark models, *i.e.*, the special case $\alpha = 0$ (linear Gaussian model) and the threshold-regression models, respectively.

5.B.1 Multivariate nonparametric model

Basu and Chib (2003) use the marginal density identity $\log_{10} p(\mathbf{Y}) = \log_{10} p(\mathbf{Y}\,|\,\boldsymbol{\omega}^\dagger) + \log_{10} p(\boldsymbol{\omega}^\dagger) - \log_{10} p(\boldsymbol{\omega}^\dagger\,|\,\mathbf{Y})$ (as originally proposed in Chib, 1995), with the ordinate $\boldsymbol{\omega}^\dagger = \{\alpha^\dagger, \boldsymbol{\lambda}^\dagger\}$. However, computing the likelihood ordinate $p(\mathbf{Y}\,|\,\boldsymbol{\omega}^\dagger) = \int p(\mathbf{Y}\,|\,\boldsymbol{\theta}, \boldsymbol{\omega}^\dagger) p(\boldsymbol{\theta}\,|\,\boldsymbol{\omega}^\dagger)\,\mathsf{d}\boldsymbol{\theta}$ in the DPM model involves an analytically intractable integral. Therefore, to solve the integral Basu and Chib (2003) use simulation techniques, in particular, a collapsed sequential importance sampling method.

The "collapsing" of the procedure exploits the equivalence of the DPM model in terms of the unique mixture-component parameters $\boldsymbol{\theta}^*_j$, $(j = 1, \ldots, I^*_T)$ and the DPM specification with mixture-component indicators $\mathbf{s}^{1,T} = (s_1, \ldots, s_T)'$, which is due to MacEachern (1994), to produce a more simulation-efficient likelihood-ordinate estimator. In the latter model variant the components' parameters are integrated analytically and the model is parameterized with the discrete-valued labels $s_t \in \{1, \ldots, I^*_T\}$, $(t = 1, \ldots, T)$, that indicate from which mixture component each observation is generated.

If we sequentially sample the mixture indicators from the distribution with pdf

$$p^*(\mathbf{s}^{1,T}\,|\,\mathbf{Y}, \boldsymbol{\omega}^\dagger) = p(s_1\,|\,\mathbf{y}_1, \boldsymbol{\omega}^\dagger) \prod_{t=2}^{T} p(s_t\,|\,\mathbf{Y}^{1,t}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger), \tag{5.B.15}$$

and use (5.B.15) as importance density for the posterior $p(\mathbf{s}^{1,T}\,|\,\mathbf{Y}, \boldsymbol{\omega}^\dagger)$, then the Monte Carlo average of the importance weights forms an estimate of the likelihood ordinate $p(\mathbf{Y}\,|\,\boldsymbol{\omega}^\dagger)$. In iteration m of the procedure we execute the following two steps for each t.

Step 1. Evaluate the t-th predictive likelihood contribution $p(\mathbf{y}_t\,|\,\mathbf{Y}^{1,t-1}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger)$;

Step 2. Simulate mixture indicator s_t from the t-th component in (5.B.15).

We start every iteration with the collapsing for $t = 1$ in Step 1, resulting in

$$u_1 = p(\mathbf{y}_1\,|\,\boldsymbol{\omega}^\dagger) = \int p(\mathbf{y}_1\,|\,\boldsymbol{\theta}) f_0(\boldsymbol{\theta}; \boldsymbol{\lambda}^\dagger)\,\mathsf{d}\boldsymbol{\theta} \equiv p_0(\mathbf{y}_1; \boldsymbol{\lambda}^\dagger).$$

This is the one-observation marginal likelihood of the first observation as provided in (5.A.3). Therefore, $\mathbf{y}_1\,|\,\boldsymbol{\omega}^\dagger \sim \mathcal{T}(\mathbf{b}^\dagger_\mu, (1 + B^\dagger_\mu)(\nu^\dagger_\Sigma + 1 - K)^{-1}\mathbf{S}^\dagger_\Sigma, \nu^\dagger_\Sigma + 1 - K)$, and we evaluate the density of a multivariate Student's t distribution. For Step 2 we start by setting $s_1 = 1$. In the remainder of this section we derive the implementation details for $t = 2, \ldots, T$.

Step 1: Predictive likelihood evaluation

The t-th predictive likelihood contribution involves the integral $p(\mathbf{y}_t\,|\,\mathbf{Y}^{1,t-1}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger) = \int p(\mathbf{y}_t\,|\,\boldsymbol{\theta}) p(\boldsymbol{\theta}\,|\,\mathbf{Y}^{1,t-1}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger)\,\mathsf{d}\boldsymbol{\theta}$. The first density under the integral is the t-th likelihood contribution from (5.1). The second density belongs to the distribution of a generic mixture-component parameter $\boldsymbol{\theta}$ given data and mixture indicators up to and including $t-1$. If $\mathbf{s}^{1,t-1}$ contains k_{t-1} distinct indicator values, the collapsing is with respect to the corresponding unique mixture-component parameters in $\boldsymbol{\theta}^*$. The second pdf is therefore expressed as

$$\begin{aligned} p(\boldsymbol{\theta}\,|\,\mathbf{Y}^{1,t-1}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger) &= \int p(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*, \boldsymbol{\omega}^\dagger) p(\boldsymbol{\theta}^*\,|\,\mathbf{Y}^{1,t-1}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger)\,\mathsf{d}\boldsymbol{\theta}^* \\ &\propto \alpha^\dagger f_0(\boldsymbol{\theta}; \boldsymbol{\lambda}^\dagger) + \sum_{j=1}^{k_{t-1}} n_{j,t-1} p(\boldsymbol{\theta}\,|\,\mathbf{Y}^*_{j,t-1}, \boldsymbol{\lambda}^\dagger). \end{aligned}$$

For this derivation we use the dynamics of (5.2) and the conditional independence properties of the distinct component parameters. Moreover, we define the matrix $\mathbf{Y}^*_{j,t-1}$ which contains all those observations $\mathbf{y}_i$, $(i \leq t-1)$, for which $s_i = j$, and $n_{j,t-1} = \sum_{i=1}^{t-1} \mathbb{I}_{\{s_i=j\}}$.

Thus, we marginalize with respect to $\boldsymbol{\theta}$ and compute the predictive likelihood

$$u_t = p(\mathbf{y}_t \,|\, \mathbf{Y}^{1,t-1}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger) = \frac{\alpha^\dagger}{\alpha^\dagger + t - 1} \int p(\mathbf{y}_t \,|\, \boldsymbol{\theta}) f_0(\boldsymbol{\theta}; \boldsymbol{\lambda}^\dagger)\, \mathsf{d}\boldsymbol{\theta} \tag{5.B.16}$$

$$+ \sum_{j=1}^{k_{t-1}} \frac{n_{j,t-1}}{\alpha^\dagger + t - 1} \int p(\mathbf{y}_t \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \mathbf{Y}^*_{j,t-1}, \boldsymbol{\lambda}^\dagger)\, \mathsf{d}\boldsymbol{\theta}. \tag{5.B.17}$$

The integral in (5.B.16) is the one-observation marginal likelihood $p_0(\mathbf{y}_t; \boldsymbol{\lambda}^\dagger)$. The k_{t-1} integrals in (5.B.17) specify the "marginal" likelihood of $\mathbf{y}_t$ under assigning it to component j and having observed data $\mathbf{Y}^*_{j,t-1}$ generated by that same mixture component. Under conjugacy also the latter integrals are available in closed form.

From the posterior simulation step in equation (5.A.4) we know that given data $\mathbf{Y}^*_{j,t-1}$ the parameters of the j-th mixture component have distribution

$$\boldsymbol{\Sigma}^*_j \,|\, \{\mathbf{Y}^*_{j,t-1}, \boldsymbol{\lambda}^\dagger\} \sim \mathcal{IW}(\overline{\nu}_j, \overline{\mathbf{S}}_j), \quad \boldsymbol{\mu}^*_j \,|\, \{\mathbf{Y}^*_{j,t-1}, \boldsymbol{\lambda}^\dagger, \boldsymbol{\Sigma}^*_j\} \sim \mathcal{N}(\overline{\mathbf{b}}_j, \overline{B}_j \boldsymbol{\Sigma}^*_j). \tag{5.B.18}$$

This posterior has exactly the same structure as the base prior in (5.3). Hence, for $j = 1, \ldots, k_{t-1}$, substituting the base-prior hyperparameters with the posterior parameters from (5.B.18) results in the multivariate Student's t distributions

$$\mathbf{y}_t \,|\, \{\mathbf{Y}^*_{j,t-1}, \boldsymbol{\lambda}^\dagger, s_t = j\} \sim \mathcal{T}(\overline{\mathbf{b}}_j, (1 + \overline{B}_j)/(\overline{\nu}_j + 1 - K)\overline{\mathbf{S}}_j, \overline{\nu}_j + 1 - K). \tag{5.B.19}$$

Step 2: Mixture-indicator sampling

To sample from the distribution with pdf the t-th component in (5.B.15), we derive the *updated* mixture indicator's density

$$p(s_t \,|\, \mathbf{Y}^{1,t}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger) \propto p(\mathbf{y}_t \,|\, \mathbf{Y}^{1,t-1}, \mathbf{s}^{1,t}, \boldsymbol{\lambda}^\dagger) p(s_t \,|\, \mathbf{s}^{1,t-1}, \alpha^\dagger). \tag{5.B.20}$$

The first density either equals one of the k_{t-1} Student's t densities in (5.B.19), or $p_0(\mathbf{y}_t; \boldsymbol{\lambda}^\dagger)$, dependent upon the value $s_t \in \{1, \ldots, k_{t-1}, k_{t-1}+1\}$. The second density is the result of the collapsed model representation of MacEachern (1994) and represented by the probabilities

$$\mathsf{Pr}[s_t = j \,|\, \mathbf{s}^{1,t-1}, \alpha^\dagger] = \begin{cases} \alpha^\dagger/(\alpha^\dagger + t - 1), & j = k_{t-1} + 1, \\ n_{j,t-1}/(\alpha^\dagger + t - 1), & j = 1, \ldots, k_{t-1}. \end{cases}$$

Thus, we sample from (5.B.20) by using the discrete distribution

$$\mathsf{Pr}[s_t = j \,|\, \mathbf{Y}^{1,t}, \mathbf{s}^{1,t-1}, \boldsymbol{\omega}^\dagger] \propto \begin{cases} \alpha^\dagger \int p(\mathbf{y}_t \,|\, \boldsymbol{\theta}) f_0(\boldsymbol{\theta}; \boldsymbol{\lambda}^\dagger)\, \mathsf{d}\boldsymbol{\theta}, & j = k_{t-1} + 1, \\ n_{j,t-1} \int p(\mathbf{y}_t \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \mathbf{Y}^*_{j,t-1}, \boldsymbol{\lambda}^\dagger)\, \mathsf{d}\boldsymbol{\theta}, & j = 1, \ldots, k_{t-1}. \end{cases}$$

We repeat this two-step recursion M times and the Monte Carlo average of the importance weights $M^{-1} \sum_{m=1}^{M} \prod_{t=1}^{T} u_t^{(m)}$ serves as simulation-consistent estimator of the likelihood ordinate. The two remaining parts to complete the computation of the marginal likelihood are the posterior and prior ordinates.

To evaluate the posterior ordinate $p(\boldsymbol{\omega}^\dagger \,|\, \mathbf{Y}) = \int p(\boldsymbol{\lambda}^\dagger, \alpha^\dagger \,|\, \mathbf{Y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \mathbf{Y})\, \mathsf{d}\boldsymbol{\theta}$ we follow Chib (1995). The prior independence of $\boldsymbol{\lambda}$ and α is maintained in the conditional posterior (given the parameters $\boldsymbol{\theta}$), see Steps 3–4 in Appendix 5.A.1. We therefore obtain

$$p(\boldsymbol{\omega}^\dagger \,|\, \mathbf{Y}) = \int p(\mathbf{b}_\mu^\dagger, B_\mu^\dagger \,|\, \mathbf{Y}, \boldsymbol{\theta}) p(\mathbf{S}_\Sigma^\dagger \,|\, \mathbf{Y}, \boldsymbol{\theta}) p(\alpha^\dagger \,|\, \mathbf{Y}, \boldsymbol{\theta}, \eta) p(\boldsymbol{\theta}, \eta \,|\, \mathbf{Y}) \, \mathsf{d}\{\boldsymbol{\theta}, \eta\}.$$

The first three densities have known form and are given in (5.A.5), (5.A.6) and (5.A.7), respectively. Since we have available a sample from the (augmented) posterior $p(\boldsymbol{\theta}, \eta \,|\, \mathbf{Y})$ as a result of the MCMC simulation, we do not have to run additional MCMC simulations as is generally the case (Basu and Chib, 2003). We use the Monte Carlo average of the product under the integral evaluated in the posterior draws as the required estimate.

Finally, since the prior ordinate consists of the components $p(\boldsymbol{\omega}^\dagger) = p(\alpha^\dagger) p(\mathbf{b}_\mu^\dagger, B_\mu^\dagger) p(\mathbf{S}_\Sigma^\dagger)$, we evaluate (5.7), (5.8) and (5.9) in $\boldsymbol{\omega}^\dagger$.

5.B.2 Linear Gaussian model

If $\alpha = 0$ in the nonparametric model, we have a single Gaussian mixture component from which all observations are sampled, and we obtain the multivariate model

$$\mathbf{y}_t \,|\, \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\} \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} \sim \mathcal{IW}(\nu_\Sigma, \mathbf{S}_\Sigma), \quad \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma} \sim \mathcal{N}(\mathbf{b}_\mu, B_\mu \boldsymbol{\Sigma}), \quad (t = 1, \ldots, T).$$

The second-layer prior on the hyperparameters is again as in (5.8) and (5.9). Due to conjugacy properties we analytically marginalize with respect to $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ to obtain the likelihood ordinate $p(\mathbf{Y} \,|\, \boldsymbol{\lambda}^\dagger)$, which is the density of a matricvariate Student's t distribution (see Appendix B). However, a computationally more attractive expression is provided by the decomposition $p(\mathbf{Y} \,|\, \boldsymbol{\lambda}^\dagger) = p(\mathbf{y}_1 \,|\, \boldsymbol{\lambda}^\dagger) \prod_{t=2}^{T} p(\mathbf{y}_t \,|\, \mathbf{Y}^{1,t-1}, \boldsymbol{\lambda}^\dagger)$. We compute the likelihood ordinate by iterative Bayesian updating.

The first density equals the one-observation marginal likelihood as expressed in (5.A.3), which results in $\mathbf{y}_1 \,|\, \boldsymbol{\lambda}^\dagger \sim \mathcal{T}(\mathbf{b}_\mu^\dagger, (1 + B_\mu^\dagger)(\nu_\Sigma + 1 - K)^{-1} \mathbf{S}_\Sigma^\dagger, \, \nu_\Sigma + 1 - K)$. For the remaining components we derive the updated posterior distribution

$$\boldsymbol{\Sigma} \,|\, \{\mathbf{Y}^{1,t-1}, \boldsymbol{\lambda}^\dagger\} \sim \mathcal{IW}(\overline{\nu}_{t-1}, \overline{\mathbf{S}}_{t-1}), \quad \boldsymbol{\mu} \,|\, \{\mathbf{Y}^{1,t-1}, \boldsymbol{\lambda}^\dagger, \boldsymbol{\Sigma}\} \sim \mathcal{N}(\overline{\mathbf{b}}_{t-1}, \overline{B}_{t-1} \boldsymbol{\Sigma}), \qquad (5.B.21)$$

with posterior parameters as in (5.B.18), but now with *all* data up to and including time $t - 1$. Thus, the t-th component of the likelihood ordinate is analytically obtained and we use $\mathbf{y}_t \,|\, \{\mathbf{Y}^{1,t-1}, \boldsymbol{\lambda}^\dagger\} \sim \mathcal{T}(\overline{\mathbf{b}}_{t-1}, (1 + \overline{B}_{t-1})(\overline{\nu}_{t-1} + 1 - K)^{-1} \overline{\mathbf{S}}_{t-1}, \, \overline{\nu}_{(t-1)} + 1 - K)$.

Concluding, to obtain the marginal likelihood we first iteratively update the four posterior parameters in (5.B.21) and evaluate the corresponding multivariate Student's t pdfs. Second, we compute the "penalizing" part $\log_{10} p(\boldsymbol{\lambda}^\dagger) - \log_{10} p(\boldsymbol{\lambda}^\dagger \,|\, \mathbf{Y})$ of the marginal likelihood in a similar way as described for the general DPM case.

5.B.3 Threshold-regression models

We follow Chib (1995) to calculate the marginal likelihood in the various threshold-regression models of Section 5.3.5. That is, we apply Bayes' marginal density identity (BMI) with the ordinate $\{\boldsymbol{\omega}_1^\dagger, \boldsymbol{\omega}_2^\dagger, \boldsymbol{\psi}^\dagger\}$ equal to the posterior mode of the parameters. The implementation of the method requires the evaluation of three densities. Additional simulation steps are needed to compute them. Next we discuss how to obtain these three inputs.

Likelihood ordinate

The first input of the BMI is the likelihood function evaluated in the ordinate. The sampling distribution specified in (5.11) results in

$$p(\mathbf{y}\,|\,\boldsymbol{\omega}_1^\dagger,\boldsymbol{\omega}_2^\dagger,\boldsymbol{\psi}^\dagger)=p(\mathbf{y}\,|\,\boldsymbol{\omega}_1^\dagger,\boldsymbol{\omega}_2^\dagger)=\prod_{t=1}^T\phi(y_t;\mathbf{x}_t{}'\boldsymbol{\beta}_t^\dagger,\sigma_t^{2\dagger}).$$

The parameter values are $\boldsymbol{\beta}_t^\dagger=\boldsymbol{\beta}_1^\dagger\mathbb{I}_{\{x_t\leq\gamma_1^\dagger\}}+\boldsymbol{\beta}_2^\dagger\mathbb{I}_{\{x_t>\gamma_1^\dagger\}}$ and $\sigma_t^{2\dagger}=\sigma_1^{2\dagger}\mathbb{I}_{\{x_t\leq\gamma_2^\dagger\}}+\sigma_2^{2\dagger}\mathbb{I}_{\{x_t>\gamma_2^\dagger\}}$, and $\phi(\cdot;a,A)$ denotes the density of the normal distribution with mean a and variance A.

Prior ordinate

BMI's second input is the prior density evaluated in the ordinate which, due to independence of the regression- and variance-part parameters, results in

$$p(\boldsymbol{\omega}_1^\dagger,\boldsymbol{\omega}_2^\dagger,\boldsymbol{\psi}^\dagger)=p(\boldsymbol{\omega}_1^\dagger\,|\,\boldsymbol{\psi}^\dagger)p(\boldsymbol{\omega}_2^\dagger\,|\,\boldsymbol{\psi}^\dagger)p(\boldsymbol{\psi}^\dagger).$$

Its first component further simplifies to $p(\boldsymbol{\omega}_1^\dagger\,|\,\boldsymbol{\psi}^\dagger)=p(\boldsymbol{\beta}_1^\dagger\,|\,\boldsymbol{\psi}^\dagger)p(\boldsymbol{\beta}_2^\dagger\,|\,\boldsymbol{\psi}^\dagger)p(\gamma_1^\dagger)$, and likewise for the second holds $p(\boldsymbol{\omega}_2^\dagger\,|\,\boldsymbol{\psi}^\dagger)=p(\sigma_1^{2\dagger}\,|\,\boldsymbol{\psi}^\dagger)p(\sigma_2^{2\dagger}\,|\,\boldsymbol{\psi}^\dagger)p(\gamma_2^\dagger)$. To compute these six individual priors we use (5.13)–(5.14). The final part concerns the second-layer prior $p(\boldsymbol{\psi}^\dagger)=p(\mathbf{b}_\beta^\dagger\,|\,\mathbf{B}_\beta^\dagger)p(\mathbf{B}_\beta^\dagger)p(S_\sigma^\dagger)$ for which we apply (5.15).

Posterior ordinate

Third and final input is the posterior density evaluated in the ordinate. Using the structure of the posterior simulation procedure we decompose it into the three blocks

$$p(\boldsymbol{\omega}_1^\dagger,\boldsymbol{\omega}_2^\dagger,\boldsymbol{\psi}^\dagger\,|\,\mathbf{y})=p(\boldsymbol{\omega}_1^\dagger\,|\,\mathbf{y})p(\boldsymbol{\omega}_2^\dagger\,|\,\mathbf{y},\boldsymbol{\omega}_1^\dagger)p(\boldsymbol{\psi}^\dagger\,|\,\mathbf{y},\boldsymbol{\omega}_1^\dagger,\boldsymbol{\omega}_2^\dagger).\tag{5.B.22}$$

To evaluate the first density to the right of (5.B.22) we compute the integral $p(\boldsymbol{\omega}_1^\dagger\,|\,\mathbf{y})=\int p(\boldsymbol{\omega}_1^\dagger\,|\,\mathbf{y},\boldsymbol{\omega}_2,\boldsymbol{\psi})p(\boldsymbol{\omega}_2,\boldsymbol{\psi}\,|\,\mathbf{y})\,\mathsf{d}\{\boldsymbol{\omega}_2,\boldsymbol{\psi}\}$ by Monte Carlo integration of the first-block density in (5.A.8)–(5.A.9) over the Gibbs draws.

For the second pdf in (5.B.22) we evaluate $p(\boldsymbol{\omega}_2^\dagger\,|\,\mathbf{y},\boldsymbol{\omega}_1^\dagger)=\int p(\boldsymbol{\omega}_2^\dagger\,|\,\mathbf{y},\boldsymbol{\omega}_1^\dagger,\boldsymbol{\psi})p(\boldsymbol{\psi}\,|\,\mathbf{y},\boldsymbol{\omega}_1^\dagger)\,\mathsf{d}\boldsymbol{\psi}$. In this case we need additional Gibbs runs for $\{\boldsymbol{\omega}_2,\boldsymbol{\psi}\}$ with $\boldsymbol{\omega}_1$ fixed at $\boldsymbol{\omega}_1^\dagger$ such that we obtain a sample from $p(\boldsymbol{\psi}\,|\,\mathbf{y},\boldsymbol{\omega}_1^\dagger)$. With these draws we compute the Monte Carlo average of the second-block density given in (5.A.10)–(5.A.11).

With the conditional independence properties as discussed in Section 5.A.2, the final density in (5.B.22) simplifies to $p(\mathbf{B}_\beta^\dagger\,|\,\mathbf{y},\boldsymbol{\omega}_1^\dagger)p(\mathbf{b}_\beta^\dagger\,|\,\mathbf{y},\boldsymbol{\omega}_1^\dagger,\mathbf{B}_\beta^\dagger)p(S_\sigma^\dagger\,|\,\mathbf{y},\boldsymbol{\omega}_2^\dagger)$. The three components of this decomposition are readily available from (5.A.12), (5.A.13) and (5.A.14), respectively.

APPENDIX A

Definitions Probability Distributions

Beta

A random variable Y follows a Beta distribution with parameters $\alpha > 0$ and $\beta > 0$, *i.e.*, $Y \sim \mathcal{B}e(\alpha, \beta)$, if its pdf is

$$p(y \,|\, \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha - 1}(1 - y)^{\beta - 1}\mathbb{I}_{\{y \in (0,1)\}}.$$

The mean and variance are

$$\mathsf{E}\,[Y] = \frac{\alpha}{\alpha + \beta}, \qquad \mathsf{Var}\,[Y] = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

Dirichlet

A random vector $\mathbf{Y} = (Y_1, \ldots, Y_K)'$ with elements $Y_k \in (0, 1)$ and $\sum_{k=1}^{K} Y_k = 1$, follows a Dirichlet distribution with parameters $\alpha_k > 0$, $(k = 1, \ldots, K)$, *i.e.*, $\mathbf{Y} \sim \mathcal{D}ir(\alpha_1, \ldots, \alpha_K)$, if its pdf is

$$p(\mathbf{y} \,|\, \alpha_1, \ldots, \alpha_K) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} \cdots y_K^{\alpha_K - 1} \mathbb{I}_{\{\mathbf{y} \in (0,1)^K \wedge \mathbf{y}'\boldsymbol{\iota}_K = 1\}}.$$

The Dirichlet is the multivariate generalization of the Beta distribution. Marginally, each element $Y_k \sim \mathcal{B}e(\alpha_k, \sum_{j \neq k} \alpha_j)$, and this result also provides the respective means and variances.

Gamma

A random variable Y follows a Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, *i.e.*, $Y \sim \mathcal{G}a(\alpha, \beta)$, if its pdf is

$$p(y \,|\, \alpha, \beta) = c \cdot y^{\alpha - 1} \exp\{-\beta y\} \mathbb{I}_{\{y > 0\}},$$

with the normalizing constant $c = \beta^\alpha/\Gamma(\alpha)$. The mean and variance are

$$\mathsf{E}[Y] = \alpha/\beta, \qquad \mathsf{Var}[Y] = \alpha/\beta^2.$$

Inverted Gamma–2

A random variable Y follows an inverted Gamma–2 distribution with $\nu > 0$ degrees of freedom and parameter $S > 0$, *i.e.*, $Y \sim \mathcal{IG}2(\nu, S)$, if its pdf is

$$p(y \mid \nu, S) = c^{-1} \cdot y^{-\frac{\nu+2}{2}} \exp\{-S/(2y)\} \, \mathbb{I}_{\{y>0\}},$$

in which the inverse of the integrating constant is $c = \Gamma(\nu/2)\left(\frac{2}{S}\right)^{\nu/2}$. The mean and variance are

$$\mathsf{E}[Y] = \frac{S}{\nu-2}, \quad \text{if} \quad \nu > 2, \qquad \mathsf{Var}[Y] = \frac{2}{\nu-4}\left(\mathsf{E}[Y]\right)^2, \quad \text{if} \quad \nu > 4.$$

Notes: (1) $Y \sim \mathcal{IG}2(\nu, S) \Leftrightarrow S/Y \sim \chi^2(\nu)$; (2) $Y \sim \mathcal{IG}2(\nu, S) \Leftrightarrow 1/Y \sim \mathcal{G}a(\nu/2, S/2)$.

Lognormal

A random variable Y follows a lognormal distribution with parameters μ and $\sigma^2 > 0$, *i.e.*, $Y \sim \mathcal{LN}(\mu, \sigma^2)$, if $X \equiv \log(Y) \sim \mathcal{N}(\mu, \sigma^2)$. Its pdf (using the transformation theorem) is

$$p(y \mid \mu, \sigma^2) = (2\pi\sigma^2 y^2)^{-\frac{1}{2}} \exp\left\{-\frac{(\log(y) - \mu)^2}{2\sigma^2}\right\} \mathbb{I}_{\{y>0\}}.$$

The mean and variance are (using the moment-generating function of a normal distribution)

$$\mathsf{E}[Y] = \exp\left\{\mu + \sigma^2/2\right\}, \qquad \mathsf{Var}[Y] = \exp\left\{2(\mu + \sigma^2)\right\} - \exp\left\{2\mu + \sigma^2\right\}.$$

Student's t

A random variable Y follows a Student's t distribution with location parameter μ, scale parameter $\sigma^2 > 0$ and $\nu > 0$ degrees of freedom, *i.e.*, $Y \sim \mathcal{T}(\mu, \sigma^2, \nu)$, if its pdf is

$$p(y \mid \mu, \sigma^2, \nu) = c \cdot \left(1 + \frac{(y-\mu)^2}{\sigma^2 \nu}\right)^{-\frac{\nu+1}{2}} \mathbb{I}_{\{-\infty<y<\infty\}},$$

with integrating constant $c = \Gamma\left(\frac{\nu+1}{2}\right)/\Gamma\left(\frac{\nu}{2}\right)(\sigma^2\nu\pi)^{-\frac{1}{2}}$. Its mean and variance are

$$\mathsf{E}[Y] = \mu, \quad \text{if} \quad \nu > 1, \qquad \mathsf{Var}[Y] = \nu\sigma^2/(\nu-2), \quad \text{if} \quad \nu > 2.$$

Truncated normal

A random variable Y follows a truncated normal distribution with parameters μ and $\sigma^2 > 0$ and support in the region $S \subseteq \mathbb{R}$, *i.e.*, $Y \sim \mathcal{N}(\mu, \sigma^2) \times \mathbb{I}_{\{Y \in S\}}$, if its pdf is

$$p(y \mid \mu, \sigma^2, S) = \frac{\phi(y; \mu, \sigma^2)}{\Lambda(S; \mu, \sigma^2)} \mathbb{I}_{\{y \in S\}},$$

with $\Lambda(S;\mu,\sigma^2) = \mathsf{Pr}[X \in S]$, if $X \sim \mathcal{N}(\mu,\sigma^2)$. With $S = [\gamma_1,\gamma_2]$, the mean is

$$\mathsf{E}[Y] = \mu + \Big[\phi(\gamma_1;\mu,\sigma^2) - \phi(\gamma_2;\mu,\sigma^2)\Big]/\Lambda(S;\mu,\sigma^2),$$

and the variance is always smaller than σ^2.

Multivariate Student's t

A $k \times 1$ random vector $\mathbf{Y}$ follows a multivariate Student's t distribution with $k \times 1$ location parameter $\boldsymbol{\mu}$, $k \times k$ positive definite symmetric scale matrix $\mathbf{S}$ and $\nu > 0$ degrees of freedom, *i.e.*, $\mathbf{Y} \sim \mathcal{T}(\boldsymbol{\mu},\mathbf{S},\nu)$, if its pdf is

$$p(\mathbf{y}\,|\,\boldsymbol{\mu},\mathbf{S},\nu) = c \cdot |\mathbf{S}|^{-1/2}\left\{\nu + (\mathbf{y}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}^{-\frac{\nu+k}{2}} \mathbb{I}_{\{\mathbf{y}\in\mathbb{R}^k\}},$$

in which the integrating constant equals $c = \dfrac{\nu^{\nu/2}\Gamma((\nu+k)/2)}{\pi^{k/2}\Gamma(\nu/2)}$. Its mean and variance are

$$\mathsf{E}[\mathbf{Y}] = \boldsymbol{\mu}, \quad \text{if} \quad \nu > 1, \qquad \mathsf{Var}[\mathbf{Y}] = \frac{\nu}{\nu-2}\mathbf{S}, \quad \text{if} \quad \nu > 2.$$

Notes: (1) It is the distribution of $\mathbf{Y}\,|\,\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu},\mathbf{Z})$ mixed over $\mathbf{Z} \sim \mathcal{IW}(\nu+k-1,\nu\mathbf{S})$. Or, $\mathbf{Y}\,|\,Z \sim \mathcal{N}(\boldsymbol{\mu},Z\mathbf{S})$ mixed over $Z \sim \mathcal{IG}2(\nu,\nu)$; (2) The conditional distribution of a subvector $\mathbf{Y}_1$ of length k_1 given the remaining elements $\mathbf{Y}_2$ is established with the regression-lemma decomposition

$$\begin{aligned}(\mathbf{y}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\mathbf{y}-\boldsymbol{\mu}) &= (\mathbf{y}_1-\boldsymbol{\mu}_{1|2})'\mathbf{S}_{1|2}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_{1|2}) + (\mathbf{y}_2-\boldsymbol{\mu}_2)'\mathbf{S}_{22}^{-1}(\mathbf{y}_2-\boldsymbol{\mu}_2)\\ &\equiv (\mathbf{y}_1-\boldsymbol{\mu}_{1|2})'\mathbf{S}_{1|2}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_{1|2}) + R(\mathbf{y}_2),\end{aligned}$$

resulting in $\mathbf{Y}_1\,|\,\{\mathbf{Y}_2 = \mathbf{y}_2\} \sim \mathcal{T}(\boldsymbol{\mu}_{1|2}, (\nu+k-k_1)^{-1}(\nu+R(\mathbf{y}_2))\mathbf{S}_{1|2}, \nu+k-k_1)$, with $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \mathbf{S}_{12}\mathbf{S}_{22}^{-1}(\mathbf{y}_2-\boldsymbol{\mu}_2)$ and $\mathbf{S}_{1|2} = \mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{12}'$. Marginally, $\mathbf{Y}_2 \sim \mathcal{T}(\boldsymbol{\mu}_2,\mathbf{S}_{22},\nu)$.

Inverted Wishart

A $k \times k$ random symmetric positive definite matrix $\mathbf{Y}$ is said to follow an inverted Wishart distribution with parameters a $k \times k$ positive definite symmetric matrix $\mathbf{S}$ and $\nu > k-1$ degrees of freedom, *i.e.*, $\mathbf{Y} \sim \mathcal{IW}(\nu,\mathbf{S})$, if its pdf is

$$p(\mathbf{Y}\,|\,\mathbf{S},\nu) = c \cdot |\mathbf{Y}|^{-(\nu+k+1)/2}\exp\left\{-\frac{1}{2}\mathsf{tr}\left[\mathbf{Y}^{-1}\mathbf{S}\right]\right\}\mathbb{I}_{\{\mathbf{Y}\in\boldsymbol{\Lambda}\}},$$

in which $\boldsymbol{\Lambda} \subseteq \mathbb{R}^{k\times k}$ is the space of symmetric positive definite matrices, and the integrating constant is $c = \dfrac{|\mathbf{S}|^{\nu/2}}{2^{\nu k/2}\Gamma_k(\nu/2)}$, with $\Gamma_p(a) = \pi^{p(p-1)/4}\prod_{i=1}^p \Gamma(a+(1-i)/2)$ as the p-variate gamma function. The mean and variances are

$$\begin{aligned}&\mathsf{E}[\mathbf{Y}] = (\nu-k-1)^{-1}\mathbf{S}, \quad \text{if} \quad \nu > k+1,\\ &\mathsf{Var}[Y_{ij}] = \frac{(\nu-k+1)S_{ij}^2 + (\nu-k-1)S_{ii}S_{jj}}{(\nu-k)(\nu-k-1)^2(\nu-k-3)}, \quad \text{if} \quad \nu > k+3.\end{aligned}$$

Notes: (1) For each diagonal block $\mathbf{Y}_{ii}$ of size k_1 it holds that marginally $\mathbf{Y}_{ii} \sim \mathcal{IW}(\nu-(k-k_1),\mathbf{S}_{ii})$, with $\mathbf{S}_{ii}$ the i-th diagonal block of $\mathbf{S}$; (2) If $k=1$, then $Y \sim \mathcal{IG}2(\nu, S)$.

Wishart

If $\mathbf{Y} \sim \mathcal{IW}(\nu,\mathbf{S})$, then $\mathbf{X}=\mathbf{Y}^{-1}$ follows a Wishart distribution with parameter $\mathbf{S}^*=\mathbf{S}^{-1}$ and degrees of freedom ν, *i.e.*, $\mathbf{X} \sim \mathcal{W}(\nu,\mathbf{S}^*)$. Its pdf results from the transformation:

$$p(\mathbf{X}\,|\,\mathbf{S}^*,\nu) = p(\mathbf{Y}\,|\,\mathbf{S},\nu)|_{\mathbf{Y}=\mathbf{X}^{-1}}\,|\mathcal{J}_{\mathbf{Y}\to\mathbf{X}}| \propto |\mathbf{X}|^{(\nu-k-1)/2}\exp\left\{-\frac{1}{2}\mathsf{tr}\left[\mathbf{X}\mathbf{S}^{*-1}\right]\right\}\mathbb{I}_{\{\mathbf{X}\in\mathbf{\Lambda}\}},$$

since $|\mathcal{J}_{\mathbf{Y}\to\mathbf{X}}| = |\mathbf{X}|^{-(k+1)}$. The integrating constant is the same as for the inverted Wishart, with $\mathbf{S}$ replaced by $\mathbf{S}^{*-1}$. The mean and variances are

$$\mathsf{E}\left[\mathbf{X}\right]=\nu\mathbf{S}^*, \qquad \mathsf{Var}\left[X_{ij}\right]=\nu(S_{ij}^{*2}+S_{ii}^*S_{jj}^*).$$

With the partitioning

$$\mathbf{X}=\begin{bmatrix}\mathbf{X}_{11} & \mathbf{X}_{12}\\ \mathbf{X}_{21} & \mathbf{X}_{22}\end{bmatrix}, \quad \mathbf{S}^*=\begin{bmatrix}\mathbf{S}_{11}^* & \mathbf{S}_{12}^*\\ \mathbf{S}_{21}^* & \mathbf{S}_{22}^*\end{bmatrix},$$

then marginally $\mathbf{X}_{ii} \sim \mathcal{W}(\nu,\mathbf{S}_{ii}^*)$, $(i=1,2)$. If $k=1$, $X \sim \mathcal{G}a(\nu/2, 1/(2S^*))$.

Matricvariate normal

A $k\times l$ random matrix $\mathbf{X}$ is said to follow a matricvariate normal distribution with $k\times l$ mean parameter $\mathbf{P}$ and covariance matrix $\mathbf{Q}\otimes\mathbf{R}$, with the first an $l\times l$ and the second a $k\times k$ matrix, *i.e.*, $\mathbf{X} \sim \mathcal{MN}\left(\mathbf{P},\mathbf{Q}\otimes\mathbf{R}\right)$, if its pdf is

$$p\left(\mathbf{X}\,|\,\mathbf{P},\mathbf{Q}\otimes\mathbf{R}\right) = c\cdot\exp\left\{-\frac{1}{2}\mathsf{tr}\left[\mathbf{Q}^{-1}\left(\mathbf{X}-\mathbf{P}\right)'\mathbf{R}^{-1}\left(\mathbf{X}-\mathbf{P}\right)\right]\right\}\mathbb{I}_{\{\mathbf{X}\in\mathbb{R}^{k\times l}\}}.$$

The constant is $c=(2\pi)^{-\frac{kl}{2}}\left|\mathbf{Q}\right|^{-\frac{k}{2}}\left|\mathbf{R}\right|^{-\frac{l}{2}}$. This distribution is a covariance-restricted special case of a normal distribution for the vectorized $\mathbf{X}$, that is, $\mathsf{vec}\left[\mathbf{X}\right] \sim \mathcal{N}\left(\mathsf{vec}\left[\mathbf{P}\right],\mathbf{Q}\otimes\mathbf{R}\right)$.

Matricvariate Student's t

A $k\times J$ random matrix $\mathbf{Z}$ is said to follow a matricvariate Student's t distribution with location parameter $\mathbf{M}$ $(k\times J)$, symmetric positive definite scale parameter matrices $\mathbf{S}$ $(J\times J)$ and $\mathbf{L}$ $(k\times k)$, and degrees of freedom ν, *i.e.*, $\mathbf{Z} \sim \mathcal{MT}(\mathbf{M},\mathbf{S},\mathbf{L},\nu)$, if its pdf is

$$p(\mathbf{Z}\,|\,\mathbf{M},\mathbf{S},\mathbf{L},\nu) = c\cdot|\mathbf{S}|^{\nu/2}|\mathbf{L}|^{J/2}\left|\mathbf{S}+(\mathbf{Z}-\mathbf{M})'\mathbf{L}(\mathbf{Z}-\mathbf{M})\right|^{-(\nu+k)/2}\mathbb{I}_{\{\mathbf{Z}\in\mathbb{R}^{k\times J}\}},$$

with $c=\pi^{-kJ/2}\dfrac{\Gamma_J((\nu+k)/2)}{\Gamma_J(\nu/2)}$ the integrating constant. The expectation and variance are

$$\mathsf{E}\left[\mathbf{Z}\right]=\mathbf{M}, \quad \text{if} \quad \nu>J, \qquad \mathsf{Var}\left[\mathsf{vec}\left[\mathbf{Z}\right]\right]=\frac{1}{\nu-J-1}(\mathbf{S}\otimes\mathbf{L}^{-1}), \quad \text{if} \quad \nu>J+1.$$

APPENDIX B

Useful Results for Regression Settings

B.1 Normal linear regression model

We have T observations y_t, $(t = 1, \dots, T)$ of the dependent variable. Their conditional means are linearly related to K explanatory variables $\mathbf{x}_t = (x_{1,t}, \dots, x_{K,t})'$. The errors ε_t are mutually independent as well as independent from the covariates, and each follows a normal (Gaussian) distribution. Then, the normal linear regression model in matrix form is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 \mathbf{I}_T),$$

with regression parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$, the observations stacked in $\mathbf{y} = (y_1, \dots, y_T)'$ and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_T)'$, and the errors in $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_T)'$.

B.1.1 Prior and posterior

If we use the prior $p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta} \,|\, \sigma^2) p(\sigma^2)$, of which the two composing distributions are

$$\sigma^2 \sim \mathcal{IG}2(\nu, S), \qquad \boldsymbol{\beta} \,|\, \sigma^2 \sim \mathcal{N}(\mathbf{b}, \sigma^2 \mathbf{B}), \tag{B.1}$$

then this particular form is maintained in the posterior distribution with density $p(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y})$. That is, the posterior distribution consists of

$$\sigma^2 \,|\, \mathbf{y} \sim \mathcal{IG}2(\bar{\nu}, \bar{S}), \qquad \boldsymbol{\beta} \,|\, \{\mathbf{y}, \sigma^2\} \sim \mathcal{N}(\bar{\mathbf{b}}, \sigma^2 \bar{\mathbf{B}}), \tag{B.2}$$

with the posterior parameters $\bar{\nu} = \nu + T$, $\bar{S} = S + (\mathbf{y} - \mathbf{X}\bar{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\bar{\mathbf{b}}) + (\mathbf{b} - \bar{\mathbf{b}})'\mathbf{B}^{-1}(\mathbf{b} - \bar{\mathbf{b}})$, $\bar{\mathbf{b}} = \bar{\mathbf{B}}(\mathbf{X}'\mathbf{y} + \mathbf{B}^{-1}\mathbf{b})$, and $\bar{\mathbf{B}} = (\mathbf{X}'\mathbf{X} + \mathbf{B}^{-1})^{-1}$. Because of the Gaussian-inverted Gamma–2 mixture, marginally the posterior of the regression parameters is the multivariate Student's t distribution

$$\boldsymbol{\beta} \,|\, \mathbf{y} \sim \mathcal{T}\left(\bar{\mathbf{b}}, (\bar{S}/\bar{\nu})\bar{\mathbf{B}}, \bar{\nu}\right).$$

This result for the joint posterior distribution means that the K marginal posteriors of the individual parameters in $\boldsymbol{\beta}$ are the univariate Student's t distributions

$$\beta_k \,|\, \mathbf{y} \sim \mathcal{T}(\bar{b}_k, (\bar{S}/\bar{\nu})\bar{B}_{kk}, \bar{\nu}), \quad (k = 1, \ldots, K),$$

with $\bar{b}_k$ and $\bar{B}_{kk}$ the k-th and (k,k)-element of $\bar{\mathbf{b}}$ and $\bar{\mathbf{B}}$, respectively.

B.1.2 Marginal and predictive likelihood

Since the prior in (B.1) forms a conjugate pair with the Gaussian likelihood, the marginal likelihood is available analytically, *i.e.*, we can evaluate the integral $p(\mathbf{y}) = \int p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2) \, \mathsf{d}\{\boldsymbol{\beta}, \sigma^2\}$ without using simulation techniques. First, we integate with respect to $\boldsymbol{\beta}$ by applying the decomposition rule (Greenberg, 2008) and then integrate the kernel of the Gaussian density in (B.2). Second, we integrate out σ^2 after recognizing the kernel of the inverted Gamma–2 density of (B.2). Keeping track of the integrating constants, these steps result in

$$p(\mathbf{y}) = \pi^{-\frac{T}{2}} \frac{\Gamma\left(\frac{T+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} S^{\frac{\nu}{2}} \left[Q(\mathbf{y}) + S\right]^{-\frac{T+\nu}{2}} |\mathbf{B}|^{-\frac{1}{2}} \left|\mathbf{X}'\mathbf{X} + \mathbf{B}^{-1}\right|^{-\frac{1}{2}}, \tag{B.3}$$

in which $Q(\mathbf{y}) = (\mathbf{y} - \mathbf{X}\bar{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\bar{\mathbf{b}}) + (\mathbf{b} - \bar{\mathbf{b}})'\mathbf{B}^{-1}(\mathbf{b} - \bar{\mathbf{b}})$. Since $\bar{\mathbf{b}}$ is linear in $\mathbf{y}$, $Q(\mathbf{y})$ is quadratic in $\mathbf{y}$, and the expression in (B.3) is the density of a multivariate Student's t. It is not obvious to derive its parameters from this expression though. To that end, we use the following alternative derivation.

If we condition on the variance σ^2, we only deal with Gaussian distributions in $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. That is, $\boldsymbol{\beta} \,|\, \sigma^2 \sim \mathcal{N}(\mathbf{b}, \sigma^2\mathbf{B})$ and $\boldsymbol{\varepsilon} \,|\, \sigma^2 \sim \mathcal{N}(\mathbf{0}_T, \sigma^2\mathbf{I}_T)$, which shows that $\mathbf{y} \,|\, \sigma^2$ is Gaussian as well. We obtain

$$\mathbf{y} \,|\, \sigma^2 \sim \mathcal{N}(\mathbf{X}\mathbf{b}, \sigma^2(\mathbf{X}\mathbf{B}\mathbf{X}' + \mathbf{I}_T)), \tag{B.4}$$

in which the variance matrix is the result of the regression parameters and the error terms being uncorrelated. Next, integrating out σ^2 means mixing the Gaussian distribution in (B.4) over the inverted Gamma–2 prior, such that marginally $\mathbf{y}$ has a Student's t distribution. We define $\boldsymbol{\Omega} = \mathbf{X}\mathbf{B}\mathbf{X}' + \mathbf{I}_T$, then we have

$$p(\mathbf{y}) \propto \int (\sigma^2)^{-\frac{T+\nu+2}{2}} \exp\left\{- \left[(\mathbf{y} - \mathbf{X}\mathbf{b})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}) + S\right] / (2\sigma^2)\right\} \, \mathsf{d}\sigma^2.$$

We integrate the kernel of the $\mathcal{IG}2(\nu + T, S + (\mathbf{y} - \mathbf{X}\mathbf{b})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}))$ distribution and keep the terms involving $\mathbf{y}$, such that,

$$p(\mathbf{y}) \propto \left([(\mathbf{y} - \mathbf{X}\mathbf{b})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}) + S]/2\right)^{-\frac{T+\nu}{2}}. \tag{B.5}$$

Since this is the kernel of a T-variate Student's t distribution, we obtain the marginal distribution $\mathbf{y} \sim \mathcal{T}(\mathbf{X}\mathbf{b}, (S/\nu)(\mathbf{X}\mathbf{B}\mathbf{X}' + \mathbf{I}_T), \nu)$. This implies that *a priori* y_t has expectation $\mathbf{b}'\mathbf{x}_t$ and variance $({\mathbf{x}_t}'\mathbf{B}\mathbf{x}_t + 1)S/(\nu - 2)$, provided that $\nu > 2$. Due to the large-matrix inversion in (B.5), to compute the marginal likelihood (B.3) is preferable.

Using the conjugacy property of the prior, we follow an analogous derivation to find the posterior predictive distribution of y_{T+1} given its covariates $\mathbf{x}_{T+1}$. We have $\boldsymbol{\beta} \,|\, \{\mathbf{y}, \sigma^2\} \sim \mathcal{N}(\bar{\mathbf{b}}, \sigma^2\bar{\mathbf{B}})$ such that $y_{T+1} \,|\, \{\mathbf{y}, \sigma^2\} \sim \mathcal{N}(\bar{\mathbf{b}}'\mathbf{x}_{T+1}, \sigma^2(\mathbf{x}_{T+1}{}'\bar{\mathbf{B}}\mathbf{x}_{T+1} + 1))$. We obtain the same result as for the marginal likelihood, though with the prior hyperparameters substituted with their posterior counterparts. We end up with

$$y_{T+1} \,|\, \mathbf{y} \sim \mathcal{T}(\bar{\mathbf{b}}'\mathbf{x}_{T+1},\, (\bar{S}/\bar{\nu})(\mathbf{x}_{T+1}{}'\bar{\mathbf{B}}\mathbf{x}_{T+1} + 1),\, \bar{\nu}).$$

For some applications (see Chapters 4 and 5) we need to compute the likelihood only marginalized with respect to either the regression parameters or the error's variance parameter. We start with the former.

Conditional on σ^2

We need $p(\mathbf{y} \,|\, \sigma^2) = \int p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta} \,|\, \sigma^2) \,\mathsf{d}\boldsymbol{\beta}$, in which $\boldsymbol{\beta} \,|\, \sigma^2 \sim \mathcal{N}(\mathbf{b}, \sigma^2\mathbf{B})$. If we define the auxiliary variables

$$\mathbf{v} = \begin{bmatrix} \mathbf{y} \\ \mathbf{B}^{-\frac{1}{2}}\mathbf{b} \end{bmatrix}, \qquad \mathbf{W} = \begin{bmatrix} \mathbf{X} \\ \mathbf{B}^{-\frac{1}{2}} \end{bmatrix},$$

the integrand becomes

$$p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta} \,|\, \sigma^2) = (2\pi\sigma^2)^{-\frac{T+K}{2}} \left|\mathbf{B}\right|^{-\frac{1}{2}} \exp\left\{-\left(\mathbf{v} - \mathbf{W}\boldsymbol{\beta}\right)' \left(\mathbf{v} - \mathbf{W}\boldsymbol{\beta}\right)/(2\sigma^2)\right\}.$$

Applying the decomposition rule

$$\left(\mathbf{v} - \mathbf{W}\boldsymbol{\beta}\right)' \left(\mathbf{v} - \mathbf{W}\boldsymbol{\beta}\right) = (\mathbf{v} - \mathbf{W}\bar{\mathbf{b}})'(\mathbf{v} - \mathbf{W}\bar{\mathbf{b}}) + (\boldsymbol{\beta} - \bar{\mathbf{b}})'\mathbf{W}'\mathbf{W}(\boldsymbol{\beta} - \bar{\mathbf{b}}),$$

and integrating the kernel of the $\mathcal{N}(\bar{\mathbf{b}}, \sigma^2 \left(\mathbf{W}'\mathbf{W}\right)^{-1})$ distribution results in

$$p(\mathbf{y} \,|\, \sigma^2) = (2\pi\sigma^2)^{-\frac{T}{2}} \left|\mathbf{B}\right|^{-\frac{1}{2}} \left|\mathbf{X}'\mathbf{X} + \mathbf{B}^{-1}\right|^{-\frac{1}{2}} \exp\left\{-Q(\mathbf{y})/(2\sigma^2)\right\}.$$

We note that this is the compuationally more attractive version of (B.4).

Conditional on $\boldsymbol{\beta}$

The required integral is $p(\mathbf{y} \,|\, \boldsymbol{\beta}) = \int p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) p(\sigma^2 \,|\, \boldsymbol{\beta}) \,\mathsf{d}\sigma^2$. The prior in (B.1) implies that $\sigma^2 \,|\, \boldsymbol{\beta} \sim \mathcal{IG}2(\nu + K,\, S + \left(\boldsymbol{\beta} - \mathbf{b}\right)' \mathbf{B}^{-1} \left(\boldsymbol{\beta} - \mathbf{b}\right))$. Combining this prior with the multivariate normal of $\mathbf{y} \,|\, \{\boldsymbol{\beta}, \sigma^2\}$ shows that the integrand is proportional to the kernel of the $\mathcal{IG}2\left(\nu + K + T,\, S + (\mathbf{v} - \mathbf{W}\boldsymbol{\beta})'(\mathbf{v} - \mathbf{W}\boldsymbol{\beta})\right)$ distribution. If we collect the integrating constants, we get

$$p(\mathbf{y} \,|\, \boldsymbol{\beta}) = \pi^{-\frac{T}{2}} \frac{\Gamma((\nu + K + T)/2)}{\Gamma((\nu + K)/2)} \left(S^*\right)^{-\frac{T}{2}} \left(\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)' \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)/S^* + 1\right)^{-\frac{\nu+K+T}{2}},$$

with $S^* = S + \left(\boldsymbol{\beta} - \mathbf{b}\right)' \mathbf{B}^{-1} \left(\boldsymbol{\beta} - \mathbf{b}\right)$. We notice that $\mathbf{y} \,|\, \boldsymbol{\beta} \sim \mathcal{T}\left(\mathbf{X}\boldsymbol{\beta},\, S^*/(\nu + K)\mathbf{I}_T,\, \nu + K\right)$. All T observations share the common unknown σ^2, and after marginalization the data are no longer independent; the diagonal scale matrix of the Student's t distribution is not sufficient for independence.

B.2 Multivariate normal linear regression model

We have T multivariate observations $\mathbf{y}_t = (y_{1t}, \dots, y_{Jt})'$, $(t = 1, \dots, T)$. The conditional mean of each observation consists of a linear combination of K explanatory variables $\mathbf{x}_t = (x_{1t}, \dots, x_{Kt})'$. The errors $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{Jt})'$ follow a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$, and are serially independent. Using matrix notation, the model is given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Pi} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{MN}(\mathbf{0}_{T\times J}, \boldsymbol{\Sigma} \otimes \mathbf{I}_T),$$

with the matrices $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_T)'$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T)'$. The matrix $\boldsymbol{\Pi} = (\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_J)$, with $\boldsymbol{\pi}_j = (\pi_{1j}, \dots, \pi_{Kj})'$, $(j = 1, \dots, J)$, contains the $K \times J$ regression parameters.

B.2.1 Prior and posterior

If we apply the natural conjugate prior distribution consisting of the components

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\nu, \mathbf{S}), \quad \boldsymbol{\Pi} \,|\, \boldsymbol{\Sigma} \sim \mathcal{MN}(\mathbf{P}, \boldsymbol{\Sigma} \otimes \mathbf{Q}),$$

we obtain the posterior distribution which is decomposed as

$$\boldsymbol{\Sigma} \,|\, \mathbf{Y} \sim \mathcal{IW}(\overline{\nu}, \overline{\mathbf{S}}), \quad \boldsymbol{\Pi} \,|\, \{\mathbf{Y}, \boldsymbol{\Sigma}\} \sim \mathcal{MN}(\overline{\mathbf{P}}, \boldsymbol{\Sigma} \otimes \overline{\mathbf{Q}}).$$

The posterior's parameters are $\overline{\mathbf{S}} = \mathbf{S} + (\mathbf{Y} - \mathbf{X}\overline{\mathbf{P}})'(\mathbf{Y} - \mathbf{X}\overline{\mathbf{P}}) + (\overline{\mathbf{P}} - \mathbf{P})'\mathbf{Q}^{-1}(\overline{\mathbf{P}} - \mathbf{P})$, $\overline{\nu} = \nu + T$, $\overline{\mathbf{P}} = \overline{\mathbf{Q}}\,(\mathbf{X}'\mathbf{Y} + \mathbf{Q}^{-1}\mathbf{P})$, and $\overline{\mathbf{Q}} = (\mathbf{X}'\mathbf{X} + \mathbf{Q}^{-1})^{-1}$.

B.2.2 Marginal likelihood

We integrate out $\boldsymbol{\Pi}$ by noting that $\mathbf{Y}$ given the variance matrix $\boldsymbol{\Sigma}$ is the sum of two mutually independent matricvariate normal variables and hence is matricvariate normal itself. Its conditional expectation is $\mathsf{E}\,[\mathbf{Y} \,|\, \boldsymbol{\Sigma}] = \mathbf{X}\mathsf{E}\,[\boldsymbol{\Pi} \,|\, \boldsymbol{\Sigma}] + \mathsf{E}\,[\boldsymbol{\varepsilon} \,|\, \boldsymbol{\Sigma}] = \mathbf{X}\mathbf{P}$. To derive its conditional variance we use the zero conditional covariance between $\boldsymbol{\Pi}$ and $\boldsymbol{\varepsilon}$, and the vectorization of $\mathbf{Y}$ as $\mathsf{vec}\,[\mathbf{Y}] = (\mathbf{I}_J \otimes \mathbf{X})\mathsf{vec}\,[\boldsymbol{\Pi}] + \mathsf{vec}\,[\boldsymbol{\varepsilon}]$. Then, the conditional variance is

$$\mathsf{Var}\,[\mathbf{Y} \,|\, \boldsymbol{\Sigma}] = (\mathbf{I}_J \otimes \mathbf{X})(\boldsymbol{\Sigma} \otimes \mathbf{Q})(\mathbf{I}_J \otimes \mathbf{X}') + \boldsymbol{\Sigma} \otimes \mathbf{I}_T = \boldsymbol{\Sigma} \otimes (\mathbf{I}_T + \mathbf{X}\mathbf{Q}\mathbf{X}').$$

Thus, we obtain the matricvariate normal $\mathbf{Y} \,|\, \boldsymbol{\Sigma} \sim \mathcal{MN}(\mathbf{X}\mathbf{P}, \boldsymbol{\Sigma} \otimes (\mathbf{I}_T + \mathbf{X}\mathbf{Q}\mathbf{X}'))$. If we combine this distribution's pdf with the inverted Wishart prior of $\boldsymbol{\Sigma}$, we obtain the kernel (in $\boldsymbol{\Sigma}$) of the inverted Wishart distribution $\mathcal{IW}(\nu + T, \mathbf{S} + (\mathbf{Y} - \mathbf{X}\mathbf{P})'\boldsymbol{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{P}))$, with $\boldsymbol{\Omega} = \mathbf{I}_T + \mathbf{X}\mathbf{Q}\mathbf{X}'$. Integrating the latter's kernel and accounting for all constants of integration, we end up with the marginal likelihood

$$\begin{aligned}\int p(\mathbf{Y} \,|\, \boldsymbol{\Pi}, \boldsymbol{\Sigma}) p(\boldsymbol{\Pi}, \boldsymbol{\Sigma})\,\mathsf{d}\{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\} = {} & \frac{\Gamma_J((\nu + T)/2)}{\Gamma_J(\nu/2)} \pi^{-TJ/2} |\boldsymbol{\Omega}|^{-J/2} |\mathbf{S}|^{\nu/2} \\ & \times |\mathbf{S} + (\mathbf{Y} - \mathbf{X}\mathbf{P})'\boldsymbol{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{P})|^{-(\nu+T)/2},\end{aligned}$$

which is the density of a matricvariate Student's t distribution, and shows that the prior belief about $\mathbf{Y}$ is formalized by $\mathbf{Y} \sim \mathcal{MT}(\mathbf{X}\mathbf{P}, \mathbf{S}, \boldsymbol{\Omega}^{-1}, \nu)$.

Samenvatting (*Summary in Dutch*)

Beslissingen op zowel het microniveau, genomen door consumenten en producenten, als op het institutionele niveau, bestaande uit centrale banken en andere beleidsinstellingen, hangen vaak sterk af van wat 'de economie' doet of gaat doen. In het eerste geval gaat het bijvoorbeeld om investeringsbeslissingen door producenten, consumenten die wel of niet een substantiële aankoop doen, of om werknemers die besluiten op zoek te gaan naar ander werk. In het tweede geval kan worden gedacht aan monetaire of fiscale beleidsmaatregelen. Ofschoon de een wat kwantitatiever te werk gaat dan de ander – de consument zal bijvoorbeeld wat meer op intuïtie handelen vergeleken met de centrale bank –, allen maken een inschatting van de toekomstige economische situatie omdat die een aanmerkelijke invloed zal hebben op de uitkomst hun respectievelijke beslissingen.

De onderwerpen in dit proefschrift beslaan verschillende aspecten van de econometrische benadering om tot voorspellingen van macro-economische variabelen te komen. Met dit doel voor ogen worden statistische modellen geformuleerd die de kenmerkende facetten van de geanalyseerde economische variabele zo realistisch mogelijk beschrijven. Met de tegenwoordig voorhanden zijnde, sterk toegenomen machinale rekenkracht en de daaruit voortvloeiende praktische implementeerbaarheid van intensieve rekenkundige methoden, worden macro-econometrische aspecten behandeld om tot modellen te komen die een beter beeld geven – en daarmee nauwkeurigere voorspellingen – van de (relaties tussen) economische variabelen dan tot dusver. De uitkomst van de hier toegepaste econometrische benadering behelst nauwkeurig gekwantificeerde uitspraken over bijvoorbeeld de toekomstige conjunctuurstand of aanstaande interventies door de centrale bank, welke de parameters vormen in het besluitvormingsproces van economische actoren.

In de besproken macro-econometrische toepassingen staat het bewerkstelligen van een bruikbare relatie tussen een te voorspellen variabele (doelvariabele) en een groep voorspellende variabelen (voorspellers) centraal. In dat opzicht vallen alle toepassingen onder de paraplu van de niet-structurele voorspelmethodologie. Dat houdt in dat in de gebruikte modellen geen aannames worden gedaan over het gedrag van economische actoren en alleen de in de economische data aanwezige correlatiestructuren worden gebruikt om voorspellingen te maken. Hoofdstukken 2 en 3 gaan grofweg over de identificatie van die economische variabelen die de grootste voorspellende kracht bevatten voor een bepaalde doelvariabele. De functionele vorm van de relatie tussen de twee typen variablen komt aan bod in de hoofdstukken 4 en 5. Beide kwesties zijn evident verbonden aan het doel om accuratere voorspellingen te produceren.

In hoofdstuk 2 is de doelvariabele de beslissing van het *Federal Open Market Committee* (FOMC) aangaande de *Federal funds target rate*, welke, in ieder geval gedurende de bestudeerde periode 1990–2008, een belangrijke rol speelt in het monetaire beleid in de VS. Deze beslissingen hebben daarmee ook een verregaande mondiale invloed. Er kan worden besloten de *target rate* te verlagen, te verhogen of onaangeroerd te laten. Met behulp van een probit-model waarin het ordinale karakter van de afhankelijke variabele alsmede de afhankelijkheid in tijd van opeenvolgende beslissingen worden opgenomen, wordt er gekeken welke variabelen uit een gegeven, uitgebreide verzameling macro-economische indicatoren het bruikbaarst zijn om de FOMC-beslissingen te voorspellen. De selectie van de best presterende voorspellende indicatoren is 'geëndogeniseerd', met als resultaat dat in plaats van het simpelweg wel of niet opnemen in het model, iedere voorspeller een kans krijgt toegedicht die de gradatie van relevantie aangeeft. Iedere combinatie van voorspellende variabelen heeft zodoende een kans toegewezen gekregen. Dergelijke combinaties worden geïnterpreteerd als verschillende modellen en gebruikmakend van hun modelkansen wordt de modelonzekerheid meegenomen in de voorspellingen.

Het blijkt dat indicatoren van economische activiteit en vooruitziende variabelen als rentespreidingsmaatstaven zeer relelvante voorspellers zijn. Ook is informatie verkregen met enquêtes, zoals de mate van consumentenvertrouwen, van substantieel belang. Het corrigeren voor de modelonzekerheid die het gevolg is van het niet zeker weten wat de beste voorspellers zijn, zorgt voor een aanzienlijke verbetering in het voorspellen van de FOMC-beslissingen. Voor de geanalyseerde periode wordt binnen de steekproef 90 procent en buiten de steekproef 82 procent van de beslissingen juist voorspeld.

Waar in de toepassing in hoofdstuk 2 helemaal geen aannames worden gedaan over welke macro-economische variabelen het best dienen als voorspellers, wordt er in hoofdstuk 3 uitgegaan van een vaste verzameling voorspellers die allen relevant worden geacht en dus allen worden opgenomen in de modelspecificatie. In dit hoofdstuk draait het om het voorspellen van conjunctuurfluctuaties. De doelvariabele in deze toepassing is bivariaat en bestaat, ten eerste, uit een continue variabele samengesteld uit indicatoren als industriële productie, werkgelegenheids-, inkomens- en verkoopcijfers die een afspiegeling vormt van de mate van economische activiteit en, ten tweede, uit een binaire recessie-indicator die door het *National Bureau of Economic Research* (NBER) wordt gepubliceerd. De bivariate conjunctuurmaatstaf wordt gerelateerd aan zogenaamde leidende macro-economische variabelen. Dergelijke variabelen staan erom bekend dat ze voorlopen op 'de economie' en daarom uitstekend geschikt zijn voor het construeren van voorspellingen. Wat echter niet bekend is, is de periode die zo'n voorspeller voorloopt. Voor een verzameling van tien leidende indicatoren worden deze periodes geschat aan de hand van gegevens die de jaren 1961–2011 bestrijken en wordt er gekeken in hoeverre die voorloopperiodes van elkaar verschillen. Bovendien wordt onderzocht of bepaalde leidende variabelen meer voorspellende kracht bezitten dan andere.

Empirische resultaten wijzen uit dat er grote verschillen zijn in zowel het aantal maanden dat een indicator voorloopt als in het gewicht dat aan een indicator wordt toegekend. Zo kan een tweedeling gemaakt worden bestaande uit een eerste groep met variabelen die één maand tot drie maanden voorlopen (bijv. werkloosheidsuitkeringsaanvragen en nieuwe orders voor zowel consumenten- als kapitaalgoederen) en een tweede groep indicatoren met voorloopperiodes tot één jaar (zoals consumentenvertrouwen en de interestspreidingsmaatstaf). Bovendien blijken financiële variabelen relatief gezien belangrijk. Gegeven deze gevonden heterogeniteit in eigenschappen van de toegepaste leidende variabelen, wordt een 'barometer' samengesteld die, gegeven een aantal maanden vooruit, aangeeft wat de verandering in de conjunctuurstand zal zijn. De aldus gevormde index is significant beter in staat (zowel binnen als buiten de steekproef) om de NBER-recessies te voorspellen dan een eenvoudig maandelijks gelijkgewogen gemiddelde van alle gebruikte leidende variabelen.

Om de onbekende parameters in de voorgestelde dynamische modellen te schatten worden in macro-econometrie vaak tijdreeksgegevens gebruikt. Aangezien deze verzameld worden voor opeenvolgende tijdsperioden, onstaat het risico dat de gespecificeerde relatie verandert naarmate de tijd voortschrijdt. In hoofdstuk 4 wordt een methode voorgesteld en toegepast om dit verschijnsel op te vangen. Ervan uitgaande dat de functionele vorm tussen doelvariabele en voorspeller gehandhaafd blijft en de veranderingen in de relatie te wijten zijn aan modelparameters die met de tijd wijzigen, wordt in dit hoofdstuk een model toegepast dat de dynamiek van de modelparameters beschrijft. In het bijzonder gaat het hier om een dynamische specificatie die aangeeft dat de waarde van de modelparameter in een bepaalde tijdsperiode of gelijk is aan een van de waarden uit het verleden, of een nieuwe waarde aanneemt. In dat laatste geval is de nieuwe waarde een realisatie uit een door de econometrist voorgestelde kansverdeling. De keuze van deze kansverdeling hangt af van het functionele verband tussen doelvariable en voorspeller, d.w.z., dit verband kan mogelijk restricties opwerpen voor de waarden die de modelparameter kan aannemen, en van de *a priori* waarschijnlijk geachte waarden, welke gebaseerd kunnen zijn op expertise of resultaten uit het verleden.

De aanpak in hoofdstuk 4 kent twee belangrijke krachten ten opzichte van bestaande methoden voor veranderingen in parameterwaarden. Ten eerste, de flexibiliteit garandeert dat de methode inzetbaar is voor in principe alle modelspecificaties (functionele verbanden) voor een te voorspellen variabele. Dat betekent dat ze niet beperkt is tot lineaire specificaties of continue doelvariabelen. Ten tweede, aangezien het hoogst onwaarschijnlijk is in de meeste toepassingen dat het aantal parameterveranderingen van te voren vaststaat, is het zeer wenselijk dat in de gebruikte dynamische specificatie voor de modelparameter dit aantal een kansvariabele is. Bovendien zijn parameterveranderingen buiten de steekproefperiode niet uitgesloten en het risico hierop wordt verwerkt in de voorspellingen. Beide punten worden nadrukkelijk naar voren gebracht in een viertal toepassingen waarvan een met kunstmatig gegenereerde data en drie macro-economische applicaties, waaronder het voorspellen van de groei in werkgelegenheid, bruto binnenlands product en de verwachte werkloosheidsduur. Deze toepassingen tonen aan dat de veronderstelling dat modelparameters constant over de tijd zijn, erg beperkend kan uitwerken en voorspellingen aanzienlijk verbeterd worden wanneer die aanname los wordt gelaten en de methode voor parameterveranderingen wordt toegepast.

In de hoofdstukken 2 t/m 4 wordt de vorm van de functionele relatie tussen de te voorspellen economische variabele en zijn voorspellers bekend verondersteld. In plaats van het opleggen van een bepaalde structuur, wordt in hoofdstuk 5 een methode toegepast die het mogelijk maakt om de verzamelde data te laten aangeven welk patroon er tussen de gebruikte variabelen bestaat. Op deze wijze wordt het risico van het verkeerd specificeren van het verband, dat wordt aangewend voor het maken van voorspellingen, aanzienlijk verkleind. De toepassing in dit hoofdstuk betreft het voorspellen van veranderingen in macro-economische activiteit voor de periode 1965–2013, waarbij, net als in hoofdstuk 3, een voorlopende macro-indicator dienst doet als voorspeller. De methode is dermate flexibel dat, gegeven een geobserveerde waarde van de leidende indicator, niet alleen een uitspraak gedaan kan worden over de verwachte verandering in economische activiteit, maar ook over zijn zogenaamde hogere momenten als de variabiliteit en de waarschijnlijkheid van gebeurtenissen zoals een buitengewoon snel krimpende economie. Uitspraken betreffende de laatste twee eigenschappen van toekomstige activiteit hebben in het bijzonder waarde wanneer de uitwerking van een beslissing uitermate gevoelig is voor relatief onwaarschijnlijke, maar wel zeer verregaande veranderingen in economische activiteit.

De in hoofdstuk 5 toegepaste methode is van de niet-parametrische soort, wat inhoudt dat er *a priori* (bijna) niets wordt aangenomen over de vorm van de relatie tussen (in dit geval) economische activiteit en een voorlopende macro-indicator, noch over het type kansverdeling van de doelvariabele. Door gebruik te maken van een mix van multivariate normale verdelingen, bestaande uit een bij voorbaat onbekend aantal componenten, wordt de onbekende gezamenlijke kansverdeling van doelvariabele en voorspeller gemodelleerd. Na invoer van de data wordt zodoende de gehele conditionele kansverdeling van de doelvariabele gegeven de voorspeller geschat en uitspraken over toekomstige economische activiteit zijn hierop gebaseerd. Er wordt een drietal bevindingen gedaan. Ten eerste zijn er sterke aanwijzingen voor een niet-lineair verband. In het bijzonder bevat de voorlopende macro-indicator een relatief sterker signaal voor neergaande veranderingen in macro-economische activiteit. Ten tweede, in de aanloop naar en tijdens economische neergang neemt de onzekerheid rond de verwachte verandering in activiteit toe. Ten laatste blijkt dat het risico op buitengewoon flinke krimp in activiteit serieus onderschat wordt als er simpleweg wordt gekozen voor een normale verdeling om de veranderingen in activiteit te modelleren.

De statistische analyses van alle toepassingen in dit proefschrift zijn Bayesiaans. Deze keuze wordt hoofdzakelijk gedreven door twee motivaties. De eerste is van conceptuele aard en de tweede richt zich op de rekenkundige voordelen. Bij het econometrisch modelleren van een doelvariabele onstaat noodzakelijkerwijs een aantal vormen van onzekerheid. In de respectievelijke hoofstukken spelen bijvoorbeeld de kwesties: welke voorspellers te gebruiken?; hoe lang is de periode dat een leidende indicator de conjunctuurstand leidt?; zijn modelparameters constant over de tijd of kennen ze waardeveranderingen?; en hoe ziet de functionele vorm van de relatie tussen doelvariable en voorspeller er nou precies uit? In het algemeen valt geen enkele van deze vragen met zekerheid te beantwoorden. Daarbij komt nog, gegeven een modelspecificatie, de onzekerheid überhaupt wat betreft de onbekende waarden van de modelparameters, een aspect dat zeker in macro-econometrie een niet onaanzienlijke rol speelt gezien de vaak bescheiden steekproeflengtes. Ofschoon er *ad hoc*-methoden bestaan om deze bronnen van onzekerheid mee te nemen in frequentistische voorspelanalyses, biedt de Bayesiaanse aanpak een elegante en eenvormige context waarin alle onzekerheden volgens de regels van de kansrekening in de voorspellingen verwerkt worden.

Ten tweede, dankzij de ontwikkeling van relatief snelle en krachtige machinale rekenmodules en nieuwe simulatiemethoden heeft de Bayesiaanse aanpak de laatste twee decennia ook vanuit praktisch oogpunt een enorme toevlucht genomen. De hiërarchische opbouw van moderne econometrische modellen maakt het mogelijk om een omvangrijk (en daarmee vaak realistischer) model te specificeren wat in blokken 'opgeknipt' kan worden om het vervolgens stapsgewijs te analyseren met behulp van geautomatiseerde simulatietechnieken. Zo ook in alle toepassingen in dit proefschrift. De verwachting is dat deze trend door zal zetten met als gevolg dat toekomstige (macro-)economische voorspellingen verder aan accuratesse en betrouwbaarheid zullen winnen.

Bibliography

Albert, J. H. and Chib, S. (1993). Bayesian analysis for binary and polychotomous response data, *Journal of the American Statistical Association* **88**: 901–921.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2007). Real-time price discovery in stock, bond and foreign exchange markets, *Journal of International Economics* **73**: 251–277.

Andreou, E. and Ghysels, E. (2009). Structural breaks in financial time series, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer-Verlag, Berlin, pp. 839–870.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**: 821–856.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* **62**: 1383–1414.

Ang, A., Bekaert, G. and Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better?, *Journal of Monetary Economics* **54**: 1163–1212.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian non-parametric problems, *Annals of Statistics* **2**: 1152–1174.

Auerbach, A. J. (1982). The index of leading indicators: "Measurement without theory" thirty-five years later, *Review of Economics and Statistics* **64**: 589–595.

Ausin, M. C. and Lopes, H. F. (2010). Time-varying joint distribution through copulas, *Computational Statistics & Data Analysis* **54**: 2383–2399.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes, *Econometrica* **66**: 47–78.

Bai, J. S., Lumsdaine, R. L. and Stock, J. H. (1998). Testing for and dating breaks in multivariate time series, *Review of Economic Studies* **65**: 395–432.

Baker, M. (1992). Unemployment duration: Compositional effects and cyclical variability, *American Economic Review* **82**: 313–321.

Bańbura, M., Giannone, D., Modugano, M. and Reichlin, L. (2013). Now-casting and the real-time data flow, *in* G. Elliott and A. Timmermann (eds), *Handbook of Economic Forecasting, Volume 2A*, North-Holland, Amsterdam, pp. 195–237.

Basu, S. and Chib, S. (2003). Marginal likelihood and Bayes factors for Dirichlet process mixture models, *Journal of the American Statistical Association* **98**: 224–235.

Bernanke, B. S. and Blinder, A. S. (1992). The federal funds rate and the channels of monetary transmission, *American Economic Review* **82**: 901–921.

Bernanke, B. S. and Kuttner, K. N. (2005). What explains the stock market's reaction to Federal Reserve policy?, *Journal of Finance* **60**: 1221–1257.

Boldin, M. D. (1996). A check on the robustness of Hamilton's Markov switching model approach to the economic analysis of the business cycle, *Studies in Nonlinear Dynamics and Econometrics* **1**: 35–46.

Brooks, S. P. and Roberts, G. O. (1998). Covergence assessment techniques for Markov chain Monte Carlo, *Statistics and Computing* **8**: 319–335.

Burda, M., Harding, M. and Hausman, J. (2009). A Bayesian mixed logit-probit model for multinomial choice, *Journal of Econometrics* **147**: 232–246.

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

Bush, C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs, *Biometrika* **83**: 275–285.

Çakmakli, C., Paap, R. and van Dijk, D. (2011). Modeling and estimation of synchronization in multistate Markov-switching models. Working paper no. 11-002/4, Tinbergen Institute Discussion Papers, Tinbergen Institute, Rotterdam.

Camacho, M. and Perez-Quiros, G. (2002). This is what the leading indicators lead, *Journal of Applied Econometrics* **17**: 61–80.

Campbell, S. D. and Diebold, F. X. (2009). Stock returns and expected business conditions: Half a century of direct evidence, *Journal of Business & Economic Statistics* **27**: 266–278.

Campolieti, M. (2001). Bayesian semiparametric estimation of discrete duration models: An application of the Dirichlet process prior, *Journal of Applied Econometrics* **16**: 1–22.

Carlin, B. P., Gelfand, A. E. and Smith, A. F. (1992). Hierarchical Bayesian analysis of changepoint problems, *Applied Statistics* **41**: 389–405.

Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification, *Journal of Econometrics* **135**: 125–154.

Chen, X. and Fan, Y. (2006b). Estimation of copula-based semiparametric time series models, *Journal of Econometrics* **130**: 307–335.

Chib, S. (1995). Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association* **90**: 1313–1321.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics* **75**: 79–97.

Chib, S. (1998). Estimation and comparison of multiple change-point models, *Journal of Econometrics* **86**: 221–241.

Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with ARMA(p,q) errors, *Journal of Econometrics* **64**: 183–206.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output, *Journal of the American Statistical Association* **96**: 270–281.

Chinn, M., Ferrara, L. and Mignon, V. (2013). Post-recession US employment through the lens of a non-linear Okun's law. Working paper no. 2013-12, EconomiX Working Papers, Université de Paris Ouest Nanterre La Défense.

Chipman, H. (1996). Bayesian variable selection with related predictors, *Canadian Journal of Statistics* **24**: 17–36.

Chollete, L. and Ning, C. (2009). The dependence structure of macroeconomic variables in the US. Working paper no. 2009/31, UiS Working Papers in Economics and Finance, University of Stavanger.

Chu, C. S. J., Stinchcombe, M. and White, H. (1996). Monitoring structural change, *Econometrica* **64**: 1045–1065.

Chulia-Soler, H., Martens, M. and van Dijk, D. (2010). Asymmetric effects of federal funds target rate changes on S&P100 stock returns, volatilities and correlations, *Journal of Banking and Finance* **34**: 834–839.

Clark, T. E. (2009). Is the Great Moderation over?, An empirical analysis, *Economic review*, Federal Reserve Bank of Kansas City.

Clements, M. P. and Galvão, A. B. (2006). Combining predictors and combining information in modelling: Forecasting US recession probabilities and output growth, *in* C. Milas, P. Rothman and D. van Dijk (eds), *Nonlinear Time Series Analysis of Business Cycles*, Elsevier, Amsterdam, pp. 55–73.

Clements, M. P. and Hendry, D. F. (2001). *Forecasting Non-Stationary Economic Time Series*, MIT Press, Cambridge, MA.

Clements, M. P. and Hendry, D. F. (2006). Forecasting with breaks, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting, Volume 1*, North-Holland, Amsterdam, pp. 605–657.

Clements, M. P. and Krolzig, H.-M. (2003). Business cycle asymmetries: Characterization and testing based on Markov-switching autoregressions, *Journal of Business & Economic Statistics* **21**: 196–211.

Conley, T. G., Hansen, C. B., McCulloch, R. E. and Rossi, P. E. (2008). A semi-parametric Bayesian approach to the instrumental variable problem, *Journal of Econometrics* **144**: 276–305.

Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models, *Statistics and Computing* **6**: 101–111.

Croushore, D. (1993). Introducing: the survey of professional forecasters, *Federal Reserve Bank of Philadelphia Business Review* **Nov/Dec**: 3–15.

Croushore, D. (2006). Forecasting with real-time macroeconomic data, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting, Volume 1*, North-Holland, Amsterdam, pp. 961–982.

Croushore, D. (2011). Frontiers of real-time data analysis, *Journal of Economic Literature* **49**: 72–100.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists, *Journal of Econometrics* **105**: 111–130.

Cúrdia, V., del Negro, M. and Greenwald, D. L. (2014). Rare shocks, great recessions, *Journal of Applied Econometrics* **29**: 1031–1052.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients, *Journal of Financial Economics* **106**: 157–181.

Dellaportas, P. and Papageorgiou, I. (2006). Multivariate mixtures of normals with unknown number of components, *Statistical Computing* **16**: 57–68.

Dickey, J. (1971). The weighted likelihood ratio, linear hypothesis on normal location parameters, *Annals of Statistics* **42**: 204–223.

Diebold, F. X. (1998). The past, present and future of macroeconomic forecasting, *Journal of Economic Perspectives* **12**: 175–192.

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

Diebold, F. X. and Rudebusch, G. D. (1991a). Forecasting output with the composite leading index: A real-time analysis, *Journal of the American Statistical Association* **86**: 603–610.

Diebold, F. X. and Rudebusch, G. D. (1991b). Turning point prediction with the composite leading index: A real-time analysis, *in* K. Lahiri and G. H. Moore (eds), *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge University Press, Cambridge, pp. 231–256.

Diebold, F. X. and Rudebusch, G. D. (1996). Measuring business cycles: A modern perspective, *Review of Economics and Statistics* **78**: 67–77.

Doppelhofer, G. and Weeks, M. (2009a). Jointness of growth determinants, *Journal of Applied Econometrics* **24**: 209–244.

Doppelhofer, G. and Weeks, M. (2009b). Jointness of growth determinants: Reply to comments by Rodney Strachan, Eduardo Ley and Mark F.J. Steel, *Journal of Applied Econometrics* **24**: 252–256.

Dowd, K. (2008). Copulas in macroeconomics, *Journal of International and Global Economic Studies* **1**: 1–26.

Dueker, M. (1999). Conditional heteroskedasticity in qualitative response models of time series: A Gibbs-sampling approach to the bank prime rate, *Journal of Business & Economic Statistics* **17**: 466–472.

Dueker, M. J. and Rasche, R. H. (2004). Discrete policy changes and empirical models of the federal funds rate, *Federal Reserve Bank of St. Louis Review* **Nov/Dec**: 61–72.

Ehrmann, M. and Fratzscher, M. (2007). Transparency, disclosure and the Federal Reserve, *International Journal of Central Banking* **3**: 179–225.

Eichengreen, B., Watson, M. W. and Grossman, R. S. (1985). Bank rate policy under the interwar Gold Standard: A dynamic probit model, *Economic Journal* **95**: 725–745.

Escobar, M. D. (1994). Estimating Normal means with a Dirichlet process prior, *Journal of the American Statistical Association* **89**: 268–277.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association* **90**: 577–588.

Estrella, A. and Mishkin, F. S. (1997). The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank, *European Economic Review* **41**: 1375–1401.

Estrella, A. and Mishkin, F. S. (1998). Predicting U.S. recessions: Financial variables as leading indicators, *Review of Economics and Statistics* **80**: 45–61.

Faust, J., Rogers, J. H., Wang, S.-Y. B. and Wright, J. H. (2007). The high-frequency response of exchange rates and interest rates to macroeconomic announcements, *Journal of Monetary Economics* **54**: 1051–1068.

Federal Reserve Board of Governors (2005). *The Federal Reserve System – Purposes and Functions*, Federal Reserve, Washington, DC.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *Annals of Statistics* **1**: 209–230.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics, *Journal of Business & Economic Statistics* **12**: 299–308.

Florens, J. P., Richard, J. F. and Rolin, J. M. (1996). Bayesian encompassing specification tests of a parametric model against a nonparametric alternative. Working paper, Department of Economics, University of Pittsburgh, Pittsburgh, PA.

Foerster, A. (forthcoming). The asymmetric effects of uncertainty, *Federal Reserve Bank of Kansas City, Economic Review* .

Fornari, F. and Mele, A. (1997). Asymmetries and non-linearities in economic activity, *Applied Financial Economics* **7**: 203–206.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*, Springer Science, New York, NY.

Garthwaite, P. (1994). An interpretation of partial least squares, *Journal of the American Statistical Association* **17**: 122–127.

Gelfand, A. E. and Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models, *Journal of Computational and Graphical Statistics* **11**: 289–305.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* **85**: 398–409.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling, *Journal of the American Statistical Association* **88**: 901–921.

Gerlach, R., Carter, C. and Kohn, R. (2000). Efficient Bayesian inference for dynamic mixture models, *Journal of the American Statistical Association* **95**: 819–828.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics, Volume 4*, Oxford University Press, Oxford, pp. 169–193.

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: Inference, development, and communication, *Econometric Reviews* **18**: 1–73.

Geweke, J. (2001). Bayesian econometrics and forecasting, *Journal of Econometrics* **100**: 11–15.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*, John Wiley & Sons, Hoboken, NJ.

Geweke, J. (2010). *Complete and Incomplete Econometric Models*, Princeton University Press, Princeton, NJ.

Geweke, J. and Jiang, Y. (2011). Inference and prediction in a multiple structural break model, *Journal of Econometrics* **163**: 172–185.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo, *Statistical Science* **7**: 473–511.

Giordani, P. and Kohn, R. (2008). Efficient Bayesian inference for multiple change-point and mixture innovation models, *Journal of Business & Economic Statistics* **26**: 66–77.

Giordani, P., Kohn, R. and van Dijk, D. (2007). A unified approach to nonlinearity, structural change and outliers, *Journal of Econometrics* **137**: 112–133.

Girard, P. and Parent, E. (2001). Bayesian analysis of autocorrelated ordered categorical data for industrial quality monitoring, *Technometrics* **42**: 180–191.

Grammig, J. and Kehrle, K. (2008). A new marked point process model for the federal funds rate target – Methodology and forecast evaluation, *Journal of Economic Dynamics & Control* **32**: 2370–2396.

Greenberg, E. (2008). *Introduction to Bayesian Econometrics*, Cambridge University Press, New York, NY.

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

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

Hamilton, J. D. (2001). A parametric approach to flexible nonlinear inference, *Econometrica* **69**: 537–573.

Hamilton, J. D. and Jordà, O. (2002). A model of the federal funds rate target, *Journal of Political Economy* **110**: 1135–1167.

Hamilton, J. D. and Perez-Quiros, G. (1996). What do the leading indicators lead?, *Journal of Business* **69**: 27–49.

Hayo, B. and Neuenkirch, M. (2010). Do Federal Reserve communications help predict federal funds target rate decisions?, *Journal of Macroeconomics* **32**: 1014–1024.

Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C. (1999). Bayesian model averaging: A tutorial, *Statistical Science* **14**: 382–417.

Hu, L. and Phillips, P. C. B. (2004). Dynamics of the federal funds target rate: A nonstationary discrete choice approach, *Journal of Applied Econometrics* **19**: 851–867.

Hymans, S. H., Greenspan, A., Shiskin, J. and Early, J. (1973). On the use of leading indicators to predict cyclical turning points, *Brookings Papers on Economic Activity* **2**: 339–384.

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

Jardin, M. and Stephan, G. (2011). How Okun's law is non-linear in Europe: A semiparametric approach. Working paper, CREM, Université de Rennes 1, Rennes, France.

Jondeau, E. and Rockinger, M. (2006). The copula-GARCH model of conditional dependencies: An international stock market application, *Journal of International Money and Finance* **25**: 827–853.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association* **90**: 773–795.

Kauppi, H. (2012). Predicting the direction of the Fed's target rate, *Journal of Forecasting* **31**: 47–67.

Kim, C.-J., Nelson, C. R. and Piger, J. (2008). The less-volatile U.S. economy: A Bayesian investigation of timing, breadth, and potential explanations, *Journal of Business & Economic Statistics* **22**: 80–93.

Kim, H., Jackson, J. and Saba, R. (2009). Forecasting the FOMC's interest rate setting behavior: A further analysis, *Journal of Forecasting* **28**: 145–165.

Koenig, E. F. (2002). Using the purchasing managers' index to assess the economy's strength and the likely direction of monetary policy, *Federal Reserve Bank of Dallas Economic & Financial Policy Review* **1**: 1–14.

Koenig, E. F., Dolmas, S. and Piger, J. (2003). The use and abuse of real-time data in economic forecasting, *Review of Economics and Statistics* **83**: 618–628.

Koop, G. and Potter, S. M. (1999). Bayes factors and nonlinearity: Evidence from economic time series, *Journal of Econometrics* **88**: 251–281.

Koop, G. and Potter, S. M. (2007). Estimation and forecasting in models with multiple breaks, *Review of Economic Studies* **74**: 763–789.

Kuo, L. and Mallick, B. (1998). Variable selection for regression models, *Indian Journal of Statistics* **60**: 901–921.

Lahiri, K. and Moore, G. (1991). *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge University Press, Cambridge.

Lahiri, K. and Wang, J. G. (1994). Predicting cyclical turning points with leading index in a Markov switching model, *Journal of Forecasting* **13**: 245–263.

Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence, *Review of Financial Studies* **21**: 1607–1652.

Liu, J. S. and Sabatti, C. (2000). Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation, *Biometrika* **87**: 901–921.

MacEachern, S. N. (1994). Estimating Normal means with a conjugate style Dirichlet process prior, *Communications in Statistics: Simulation and Computation* **23**: 727–741.

Maheu, J. M. and Gordon, S. (2008). Learning, forecasting and structural breaks, *Journal of Applied Econometrics* **23**: 553–583.

Maheu, J. M. and Song, Y. (2014). A new structural break model, with an application to Canadian inflation forecasting, *International Journal of Forecasting* **30**: 144–160.

Marcellino, M. (2006). Leading indicators: What have we learned?, *in* G. Elliot, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting, Volume 1*, North-Holland, Amsterdam, pp. 897–960.

Marcellino, M., Stock, J. H. and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, *Journal of Econometrics* **135**: 499–526.

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

McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton, NJ.

Mishkin, F. S. (1990a). The information in the longer-maturity term structure about future inflation, *Quarterly Journal of Economics* **55**: 815–828.

Mishkin, F. S. (1990b). What does the term structure tell us about future inflation?, *Journal of Monetary Economics* **25**: 77–95.

Monokroussos, G. (2011). Dynamic limited dependent variable modeling and U.S. monetary policy, *Journal of Money, Credit and Banking* **43**: 519–534.

Morgan, M. S. (1990). *The History of Econometric Ideas*, Cambridge University Press, Cambridge.

Morley, J. and Piger, J. (2012). The asymmetric business cycle, *Review of Economics and Statistics* **94**: 208–221.

Mukoyama, T. and Şahin, A. (2009). Why did average duration of unemployment become so much longer?, *Journal of Monetary Economics* **56**: 200–209.

Müller, P., Erkanli, A. and West, M. (1996). Bayesian curve fitting using multivariate normal mixtures, *Biometrika* **83**: 67–79.

Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis, *Statistical Science* **19**: 95–110.

Neal, R. M. (1997). Markov chain Monte Carlo methods based on "slicing" the density function. Technical Report No. 9722, Department of Statistics, University of Toronto.

Neal, R. M. (1998). Markov chain sampling methods for Dirichlet process mixture models. Technical Report No. 9815, Department of Statistics, University of Toronto.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics* **9**: 249–265.

Neftçi, S. N. (1982). Optimal prediction of cyclical downturns, *Journal of Economic Dynamics & Control* **4**: 225–241.

Neftçi, S. N. (1984). Are economic time series asymmetric over the business cycle?, *Journal of Political Economy* **92**: 307–328.

Paap, R., Segers, R. and van Dijk, D. (2009). Do leading indicators lead peaks more than troughs?, *Journal of Business & Economic Statistics* **27**: 528–543.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models, *Journal of Empirical Finance* **13**: 274–315.

Perron, P. (2006). Dealing with structural breaks, *in* K. Patterson and T. C. Mills (eds), *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, Palgrave Macmillan, Hampshire, pp. 278–352.

Persons, W. (1919). An index of general business conditions, *Review of Economic Statistics* **1**: 111–205.

Pesaran, M. H., Pettenuzzo, D. and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks, *Review of Economic Studies* **73**: 1057–1084.

Pesaran, M. H. and Pick, A. (2011). Forecast combination across estimation windows, *Journal of Business & Economic Statistics* **29**: 307–318.

Pesaran, M. H., Pick, A. and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks, *Journal of Econometrics* **177**: 134–152.

Pesaran, M. H. and Timmermann, A. (2002). Market timing and return prediction under model instability, *Journal of Empirical Finance* **9**: 495–510.

Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks, *Journal of Econometrics* **164**: 60–78.

Piazzesi, M. (2005). Bond yields and the Federal Reserve, *Journal of Political Economy* **113**: 311–344.

Poole, W. (2005). How predictable is Fed policy?, *Federal Reserve Bank of St. Louis Review* **Nov/Dec**: 659–668.

Qin, D. (2010). Econometric studies of business cycles in the history of econometrics. Working paper no. 669, School of Economics and Finance, Queen Mary, University of London.

Qu, Z. J. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions, *Econometrica* **75**: 459–502.

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data, *Journal of Econometrics* **119**: 99–130.

Rapach, D. E. and Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate U.S. stock returns, *Journal of Financial Econometrics* **4**: 238–274.

Ravazzolo, F., van Dijk, D., Paap, R. and Franses, P. H. (2008). Bayesian model averaging in the presence of structural breaks, *in* M. Wohar and D. E. Rapach (eds), *Frontiers of Economics and Globalization Series, Vol.3*, Emerald Group Publishing, Bingley, pp. 561–594.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society,* B **59**: 731–792.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd edn, Springer-Verlag, New York, NY.

Rossi, P. E., Allenby, G. M. and McCulloch, R. (2005). *Bayesian Statistics and Marketing*, John Wiley & Sons, Chichester.

Rudebusch, G. D. (2001). Is the Fed too timid? Monetary policy in an uncertain world, *Review of Economics and Statistics* **81**: 203–217.

Rudebusch, G. D. (2002). Term structure evidence on interest rate smoothing and monetary policy inertia, *Journal of Monetary Economics* **49**: 1161–1187.

Salimans, T. (2012). Variable selection and functional form uncertainty in cross-country growth regressions, *Journal of Econometrics* **171**: 267–280.

Scotti, C. (2011). A bivariate model of Federal Reserve and ECB main policy rates, *International Journal of Central Banking* **7**: 37–78.

Sensier, M. and van Dijk, D. (2004). Testing for volatility changes in U.S. macroeconomic time series, *Review of Economics and Statistics* **86**: 833–839.

Smith, A. F. M. (1973). A general Bayesian model, *Journal of the Royal Statistical Society,* B **35**: 67–75.

Smith, M. S., Gan, Q. and Kohn, R. (2010). Modeling dependence using skew t copulas: Bayesian inference and applications, *Journal of Applied Econometrics* **27**: 500–522.

Smith, M. S. and Kohn, R. (1996). Nonparametric regrssion using Bayesian variable selection, *Journal of Econometrics* **75**: 317–343.

Smith, M. S. and Kohn, R. (1997). A Bayesian approach to nonparametric bivariate regression, *Journal of the American Statistical Association* **92**: 1522–1535.

Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden Markov model, *Journal of Applied Econometrics* **29**: 825–842.

Stark, T. and Croushore, D. (2002). Forecasting with a real-time data set for macroeconomists, *Journal of Macroeconomics* **24**: 507Ű531.

Stekler, H. O. and Schepsman, M. (1973). Forecasting with an index of leading series, *Journal of the American Statistical Association* **342**: 291–296.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations, *Journal of Business & Economic Statistics* **14**: 11–30.

Stock, J. H. and Watson, M. W. (2002a). Has the business cycle changed and why?, *in* M. Gertler and K. Rogoff (eds), *NBER Macroeconomics Annual 2002*, MIT Press, Cambridge, MA, pp. 159–230.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices, *Journal of Economic Literature* **41**: 788–829.

Stock, J. and Watson, M. (2002b). Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* **20**: 147–162.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distibutions by data augmentation, *Journal of the American Statistical Association* **82**: 528–550.

Taylor, J. B. (1993). Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy* **39**: 195–214.

Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models, *in* G. Elliot, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting, Volume 1*, North-Holland, Amsterdam, pp. 413–458.

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

Tierney, L. (1994). Markov chains for exploring posterior distributions, *Annals of Statistics* **22**: 1701–1762.

Tong, H. and Lim, K. S. (1980). Threshold autoregressions, limit cycles, and data, *Journal of the Royal Statistical Society,* B **42**: 245–292.

Valletta, R. G. (1998). Changes in the structure and duration of U.S. unemployment, 1967–1998, *Federal Reserve Bank of San Francisco Economic Review* **3**: 29–40.

Van den Hauwe, S. (2015). A Bayesian nonparametric method for relating a leading indicator to economic activity. Working paper, Erasmus University, Rotterdam.

Van den Hauwe, S., Paap, R. and van Dijk, D. (2013). Bayesian forecasting of federal funds target rate decisions, *Journal of Macroeconomics* **37**: 19–40.

Van den Hauwe, S., Paap, R. and van Dijk, D. (2014). How much do the leading indicators lead? Working paper, Erasmus University, Rotterdam.

Van den Hauwe, S., Paap, R. and van Dijk, D. (2015). A Bayesian mixture approach to parameter changes in empirical macroeconomic models. Working paper, Erasmus University, Rotterdam.

Vanderhart, P. G. (2000). The Federal Reserve's reaction function under Greenspan: An ordinal probit analysis, *Journal of Macroeconomics* **22**: 631–644.

Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage–Dickey density ratio, *Journal of the American Statistical Association* **90**: 614–618.

Verho, J. (2008). Determinants of unemployment duration over the business cycle in Finland. Working paper no. 226, Helsinki Center for Economic Research, University of Helsinki.

Zarnowitz, V. and Moore, G. H. (1982). Sequential signals of recession and recovery, *Journal of Business* **55**: 57–85.

Zeileis, A., Leisch, F., Kleiber, C. and Hornik, K. (2005). Monitoring structural change in dynamic econometric models, *Journal of Applied Econometrics* **20**: 99–121.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* **57**: 348–368.

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

573. M.J. GERRITSE, *Policy Competition and the Spatial Economy*
574. A. OPSCHOOR, *Understanding Financial Market Volatility*
575. R.R. VAN LOON, *Tourism and the Economic Valuation of Cultural Heritage*
576. I.L. LYUBIMOV, *Essays on Political Economy and Economic Development*
577. A.A.F. GERRITSEN, *Essays in Optimal Government Policy*
578. M.L. SCHOLTUS, *The Impact of High-Frequency Trading on Financial Markets*
579. E. RAVIV, *Forecasting Financial and Macroeconomic Variables: Shrinkage, Dimension reduction, and Aggregation*
580. J. TICHEM, *Altruism, Conformism, and Incentives in the Workplace*
581. E.S. HENDRIKS, *Essays in Law and Economics*
582. X. SHEN, *Essays on Empirical Asset Pricing*
583. L.T. GATAREK, *Econometric Contributions to Financial Trading, Hedging and Risk Measurement*
584. X. LI, *Temporary Price Deviation, Limited Attention and Information Acquisition in the Stock Market*
585. Y. DAI, *Efficiency in Corporate Takeovers*
586. S.L. VAN DER STER, *Approximate feasibility in real-time scheduling: Speeding up in order to meet deadlines*
587. A. SELIM, *An Examination of Uncertainty from a Psychological and Economic Viewpoint*
588. B.Z. YUESHEN, *Frictions in Modern Financial Markets and the Implications for Market Quality*
589. D. VAN DOLDER, *Game Shows, Gambles, and Economic Behavior*
590. S.P. CEYHAN, *Essays on Bayesian Analysis of Time Varying Economic Patterns*
591. S. RENES, *Never the Single Measure*
592. D.L. IN 'T VELD, *Complex Systems in Financial Economics: Applications to Interbank and Stock Markets*

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

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

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

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

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

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

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

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

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

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

603. S. GUBINS, *Information Technologies and Travel*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

622. A. KAMM, *Political Actors Playing Games: Theory and Experiments*