Methods for Accurate and Efficient Bayesian Analysis of Time Series

Borowska, A.

2019

Chapter 1

Introduction

This thesis investigates Bayesian inference over time series models with the emphasis put on applications in economics and finance. We note, however, that the methods developed are general and can be employed in various fields. We adopt simulation-based techniques which are necessary in any nontrival problem in this setting. The main motivation behind the presented research is to increase the efficiency and accuracy of these computationally intensive methods in several different contexts. One of the main topics addressed is efficient and precise risk estimation, or rare event analysis. Another problem studied below is the efficiency of various sampling algorithms, in particular importance sampling (IS) and Markov chain Monte Carlo (MCMC) algorithms. Finally, we address the issue of forecasting, from a single model as well as from a combination of models.

A Bayesian approach provides a flexible, coherent and convenient framework for the analysis of time series for a number of reasons, see Robert (2007, Ch. 11) for "a defence of the Bayesian choice". Of these, one of the most relevant in practice is capturing of parameter uncertainty. Treating parameters as random variables and inference based on posterior distributions allows us to easily deal with the task of uncertainty quantification. This is of particular importance in the context of risk analysis, where the objective is precise estimation of rare events and where even a small degree of incorrectness might have tremendous and serious consequences (Chapters 2 and 3). Another advantage of a Bayesian approach is that it highly facilitates dealing with complex data and combining information stemming from different sources (Chapters 4 and 5). A related aspect is that standard Bayesian methods can be easily and naturally extended to develop hybrid approaches or schemes exceeding typical inference problems to allow e.g. for built-in optimisation elements (Chapters 3, 4 and 5).

1.1 Time series

The problem of identification of properties of processes evolving in time to characterise the observed patterns and to make predictions about their future realisations is ubiquitous in science. In economics and finance time series analysis has a well-established position and it comes naturally given the studied phenomena such as business cycle or stock market trends. There have been several approaches to investigate time-series, depending on the research area, e.g. signal processing approach (including spectral analysis) in physics and engineering; function mapping approach (including Gaussian processes and neural networks) in machine learning; statistical and econometric approach prevailing in economics and finance. We adopt on the latter methodology as it provides a useful explanation of the nature of the modelled processes and allows for a structural interpretation of the estimation results.

As far as the statistical analysis of time series is concerned, we focus on time-varying parameter models, which following Cox (1981) can be grouped into two classes of models with distinct advantages and disadvantages: observation-driven models and parameter-driven models. The former specify model parameters as deterministic functions of observations allowing for a perfect one-step-ahead predictability of the parameters given the current information set. The latter allow for idiosyncratic innovations governing the parameters dynamics and can be represented as state space models, see Durbin and Koopman (2012) for an extensive exposition of the state space methodology. In consequence, the likelihood of an observation-driven model is available in closed form while the likelihood of a parameter-driven model is typically analytically intractable. This makes observation-driven models easier to work with and faster to estimate but also gives extra flexibility and an intuitive structure to parameter-driven models. Well known examples of observation-driven models in econometrics include the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986) as well as the more recent generalized autoregressive score (GAS) models of Creal et al. (2013); a quintessential parameter-driven model in econometrics is the stochastic volatility (SV) model of Taylor (1994), with other important instances being the dynamic factor model of Geweke (1977), see also Stock and Watson (2002), and the factor-augmented vector autoregression (FAVAR) model of Bernanke et al. (2005) and Stock and Watson (2005). Given distinct merits of both classes we do not see it necessary to limit our attention to one certain class in advance. Depending on the main idea and the goal of each chapter we consider a specification which is more natural and convenient in the given context.

1.2 Bayesian inference

As mentioned above, we adopt a Bayesian approach to statistical inference by which we understand estimation and prediction from a given model (as well as model comparison and selection). We refer to Robert (2007) for an in-depth exposition of the Bayesian principles including philosophical foundations of this inference paradigm and to Gelman et al. (2013) for a more practical treatment adhering to common-sense merits of Bayesian thinking. It is worth mentioning that Gelman et al. (2013) understand Bayesian methods more broadly, as consisting of three steps: (1) model building, (2) inference conditional on the model, (3) model checking. To our view their points (2) and (3) both belong to the inference problem, while step (1) does not necessarily need to be related to "Bayesian" statistics, in the sense that Bayesian reasoning can be applied to any model (even to a non-statistical one). The strength of the Bayesian paradigm is that its basic principles and rules are universal, regardless of the exact model specification, in particular the choice between observation-driven and parameter-driven models in our case. In this work we are less concerned with the modelling stage of data analysis but rather with developing accurate and efficient techniques to enhance Bayesian inference over existing models.

The key principle of Bayesian statistics is conditioning on the observed data $\mathbf{y}$, which is formalised by the *likelihood principle*[1]. Hence, the observed data are seen as fixed, which stays in contrast to classical statistics treating the data as random realisations of a sampling process. On the other hand, Bayesian statistics sees the model parameters (and all unobserved quantities) $\boldsymbol{\theta}$ as random variables and makes statements about them in terms of probability distributions – differently than frequentist reasoning which assumes that parameters are fixed. Further, Bayesian analysis allows for incorporating prior beliefs about these unknown quantities, e.g. expert knowledge or information stemming from complimentary sources – but also lack of any knowledge – into the inference process via the *prior distribution* $p(\boldsymbol{\theta})$. After recording the data, these initial beliefs are updated using the *likelihood* $p(\mathbf{y}|\boldsymbol{\theta})$, or the data distribution, to form the *posterior distribution* $p(\boldsymbol{\theta}|\mathbf{y})$. This step is formalised by applying the Bayes' theorem

[1]The likelihood principle states that for inference on an unknown parameter all of the evidence from any observation is entirely contained in the likelihood function of this observation, see Robert (2007, p. 15–16) and Gelman et al. (2013, p. 6).

to obtain the relationship between these three distributions as follows

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{1.2.1}$$

$$\propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{1.2.2}$$

The denominator of (1.2.1) is called the *marginal likelihood* and is a normalising constant (as it does not depend on $\boldsymbol{\theta}$). It is often analytically intractable and even hard to estimate, hence one typically needs to use the unnormalised posterior distribution (1.2.2), also known as the *kernel* of the posterior distribution. The posterior distribution is then used to obtain estimators of $\boldsymbol{\theta}$, which in the Bayesian setting are formally represented as integrals (see Robert and Casella, 2004, Ch. 1.2). This implies the associated computational problem of integration, often in high dimensions, which turns out not analytically solvable in many practical applications. For this reason most of Bayesian analysis is concerned with simulation-based computations, which frequently are computationally intensive.

1.3 Simulation methods and numerical efficiency

The Bayesian paradigm is conceptually clear and intuitive however it was not until the "*Markov chain Monte Carlo* revolution" in the 1990s that it has gained broader popularity, see Robert (2007) and Robert and Casella (2011). Prior to that Bayesian analysis was mostly limited to the use of conjugate[2] prior distributions (Green et al., 2015). It is crucial to point out that even though the popularisation of the MCMC methods in statistics and econometrics has triggered the growing acceptance and usage of Bayesian methods in these fields, current Bayesian computations are not necessarily limited to MCMC algorithms. Alternatives to MCMC include IS (Hammersley and Handscomb, 1964), sequential Monte Carlo (SMC, Doucet et al., 2001), approximate Bayesian computation (ABC, Marin et al., 2012) or variational Bayes (VB, Blei et al., 2017). All these methods but the last one are sampling-based, predominantly adopting Monte Carlo (MC) methods. A detailed treatment of MC techniques is beyond the scope of this chapter and we refer to Robert and Casella (2004) for a comprehensive examination of this broad area as well as to Green et al. (2015) for a review of the history and the current state of Bayesian computations.

[2]A family of prior distributions is said to be conjugate for the likelihood function if the resulting posterior distribution belongs to this family. Conjugacy allows for obtaining the posterior distribution simply by updating of the hyperparameters of the prior distribution.

Amongst the above mentioned methods for Bayesian computations IS deserves a special mention, as we argue below. IS is a relatively old technique dating back to Kahn and Marshal (1953) and Marshall (1956), with a long tradition in econometrics originating from Kloek and van Dijk (1978). Not only serves it as a building block for the SMC methods and provides intuition for MCMC algorithms, but also it can be used as a variance reduction method. Variance reduction is of key interest in risk analysis and Chapters 2 and 3 are concerned with this topic.

Even though performance capabilities of modern computers have been continuously increasing and the limitations in computations faced in the past are less of an issue nowadays, numerical accuracy and efficiency are still of major interest in statistics and econometrics. One of the reasons for this is that currently decisions need to be made quicker than ever, often in real time, whether it is for central bankers, stock market analysts or financial institution managers. Algorithms such as IS, MCMC or SMC are theoretically sound and deliver *exact* estimates when the number of simulations diverges to infinity. However, they all require a distribution to sample from, known as an importance, proposal or candidate distribution. The quality of this distribution is crucial for the ultimate performance of the algorithm. Hence different variants of the same algorithm, only differing with respect to the choice of this sampling distribution, are likely to result in dissimilar outcomes in terms of the uncertainty of the associated estimator. Therefore, methods which are reliable only "in the limit" (after excessively long simulation runs) are not suited for the purpose of real time decision making, with fast and precise methods being naturally preferred. We address the issue of constructing efficient sampling based algorithms – in different contexts – in all the chapters of this thesis, in particular in Chapter 2 and 4.

Furthermore, typical modelling and inference methods are designed to explain *average* scenarios. Since "all models are wrong", as famously stated by George Box, we cannot expect that our inference *conditional* on a model will be fully accurate. Nevertheless, because "some models are useful" though, we can aim at finding valuable aspects of a model or models at hand, while hedging against its or their potential shortcomings. For instance, suppose our ultimate goal is inference over a particular region of the posterior distribution, such as its tail. We then suggest to still use standard models due to their well-documented ability to capture stylised facts of the data, but simply in a problem-adjusted manner, with a tail-focused estimation. Chapter 3 is dedicated to this problem. Alternatively, suppose that our aim is to construct a profit-maximising portfolio while still caring about the associated investment risk. We cannot expect that there is a universally dominating (e.g. over time) single model and a single investment strategy but we can construct the portfolio based on an appropriately specified combination of models and strategies. We discuss and illustrate this idea in Chapter 5.

1.4 Thesis outline

This thesis consists of four self-contained chapters all related to Bayesian inference in time-series models. Below we present their overview.

Chapter 2 is titled "Bayesian Risk Evaluation for Long Horizons" and is based on joint work with Lennart Hoogerheide and Siem Jan Koopman. We present an accurate and efficient method for Bayesian estimation of two financial risk measures, Value-at-Risk and Expected Shortfall, for a given volatility model. We obtain precise forecasts of the tail of the distribution of returns not only for the 10-days-ahead horizon required by the Basel Committee but even for long horizons, like one-month or one-year ahead. The latter has recently attracted considerable attention due to the different properties of short term risk and long run risk. Precise forecasts of the tail of the distribution can also be useful for option pricing. The key insight behind our proposed IS based approach is the sequential construction of marginal and conditional importance densities for consecutive periods. For robustness, these importance densities are efficiently constructed as mixtures of Student's t densities. By oversampling the extremely negative scenarios and giving them lower importance weights, we achieve a much higher precision in characterising the properties of the left tail. We report substantial accuracy gains for all the considered horizons in empirical studies on two datasets of daily financial returns, including a highly volatile period of the recent financial crisis. We analyse two workhorse models used by financial practitioners, GARCH(1,1)-t and GAS(1,1)-t. To illustrate the flexibility of the proposed construction method, we present how it can be adjusted to the frequentist case, for which we provide counterparts of both Bayesian applications.

Chapter 3 is titled "Partially Censored Posterior for Robust and Efficient Risk Evaluation" and is based on joint work with Lennart Hoogerheide, Siem Jan Koopman and Herman K. van Dijk. We introduce a novel approach to inference for a specific region of the predictive distribution. An important domain of application is accurate prediction of financial risk measures, where the area of interest is the left tail of the predictive density of logreturns. Our proposed approach originates from the Bayesian approach to parameter estimation and time series forecasting, however it is robust in the sense that it provides a more accurate estimation of the predictive density in the region of interest in case of misspecification. The first main contribution of this chapter is the novel concept of the Partially Censored Posterior (PCP), where the set of model parameters is partitioned into two subsets: for the first subset of parameters we consider the standard marginal posterior, for the second subset of parameters (that are particularly related to the region of interest) we consider the conditional censored posterior. The censoring means that observations outside the region of interest are censored: for those observations only the probability of being outside the region of interest matters. This approach yields more precise parameter estimation than a fully censored posterior for all parameters, and has more focus on the region of interest than a standard Bayesian approach. The second main contribution is that we introduce two novel methods for computationally efficient simulation: Conditional MitISEM, an MCMC method to simulate model parameters from the Partially Censored Posterior, and PCP-QERMit, an IS method that is introduced to further decrease the numerical standard errors of the Value-at-Risk and Expected Shortfall estimators. The third main contribution is that we consider the effect of using a time-varying boundary of the region of interest, which may provide more information about the left tail of the distribution of the standardized innovations. Extensive simulation and empirical studies show the ability of the introduced method to outperform standard approaches.

Chapter 4 is titled “Semi-Complete Data Augmentation for Efficient State Space Model Fitting” and is based on joint work with Ruth King. We propose a novel efficient model-fitting algorithm for state space models. State space models are an intuitive and flexible class of models, frequently used in practice. This flexibility, however, often comes at the price of substantially more complicated fi

tting of such models to data due to the associated likelihood being analytically intractable. For the general case a Bayesian data augmentation approach is often employed, where the true unknown states are treated as auxiliary variables and imputed within the MCMC algorithm. However, standard “vanilla” MCMC algorithms may perform very poorly due to high correlation between the imputed states and/or parameters, leading to the need for specialist algorithms. The proposed method circumvents the inefficiencies of traditional approaches by combining data augmentation with numerical integration in a Bayesian hybrid approach. This approach permits the use of standard “vanilla” updating algorithms that perform considerably better than the traditional approach in terms of considerably improved mixing and hence lower autocorrelation. We use the proposed Semi-Complete Data Augmentation algorithm in different application areas and associated types of models, leading to distinct implementation schemes and demonstrating efficiency gains in empirical studies.

Chapter 5 is titled "Forecast Density Combinations of Dynamic Models and Data Driven Portfolio Strategies" and is based on Baştürk, Borowska, Grassi, Hoogerheide, and van Dijk (2018). We propose a novel dynamic asset allocation approach in which model-based forecasts are directly combined with a set of data driven portfolio strategies, without the necessity to define a utility or other scoring function. The specification of the underlying models is motivated by findings of a scrupulous analysis of typical stylized facts of the time series of monthly returns of ten US industries over the period 1926M7–2015M6. The portfolio strategies are based on the practice in financial investing to take advantage of a positive or negative momentum in industry returns. In probabilistic terms, the resulting dynamic asset-allocation model is specified as a combination of return distributions stemming from multiple pairs of models and strategies. The combination weights are defined through feedback mechanisms that enable learning, to allow for cross-correlation and correlation over time. We base our Bayesian inference over the proposed model on its representation as a nonlinear non-Gaussian state space model. To increase the efficiency and robustness of the simulations we introduce a new nonlinear filter based on mixtures of Student's t distributions. Diagnostic analysis of posterior residuals gives insight into the model and strategy incompleteness or misspecification. An extensive empirical application reveals that a combination of a smaller set of flexible models outperforms a larger combination of basic model structures in terms of expected return and risk. We believe that dynamic patterns in combination weights and diagnostic learning provide useful signals from a risk-management perspective and can help enhancing modelling and policy.

Chapter 6 summarises the main findings and concludes the thesis.