

PERSPECTIVES

MARKOV CHAIN MONTE CARLO METHODS IN FINANCIAL ECONOMETRICS

Michael Verhofen (michael.verhofen@unisg.ch)

Swiss Institute of Banking and Finance, University of St. Gallen,

Rosenbergstrasse 52, 9000, St. Gallen, Switzerland

Tel.: +41 71 224 7028, Fax: +41 71 224 7088

Abstract. Markov Chain Monte Carlo (MCMC) methods have become very popular in financial econometrics during the last years. MCMC methods are applicable where classical methods fail. In this paper, we give an introduction to MCMC and present recent empirical evidence. Finally, we apply MCMC methods to portfolio choice to account for parameter uncertainty and to incorporate different degrees of belief in an asset pricing model.

1. Introduction

Asset pricing theory uses arbitrage and equilibrium arguments to derive the functional relationship between asset prices and the fundamentals of the economy: state variables, structural parameters, and market prices of risk. Asset pricing models enable researchers to derive solutions for objects of interest, such as prices or portfolio weights. The suggested models are also appealing from an empirical perspective: through a judicious choice of drift, diffusion, jump intensity and jump distribution, they accommodate a wide range of dynamics for state variables and prices.

The empirical analysis of asset pricing models deals with the inverse problem, i.e., to extract information about latent state variables, structural parameters, and market prices of risk from observed prices. The Bayesian solution to this inference problem is the distribution of the param-

eters and state variables, denoted by θ , conditional on observed prices, D . This posterior distribution, $P(\theta|D)$, combines the information in the model and the observed prices and is the key to inference on parameters and state variables.

Bayesian approaches have become popular in finance during the last years. The main building block is a fundamental theorem of probability, Bayes' theorem. Bayes' theorem shows how to incorporate prior information into decision making and statistical analysis. Formally,

$$P(\theta|D) = \frac{P(D|\theta)}{\int P(D|u)P(u)du},$$

where θ and D are two random variables, $P(\cdot)$ denotes a probability distribution, and $P(\cdot|\cdot)$ a conditional probability distribution. In Bayesian statistics, θ denotes the vector of parameters and D the observed data. $P(\theta)$ is commonly called the prior and represents the researcher's beliefs. These prior beliefs are combined with the likelihood, $P(D|\theta)$, to obtain the posterior distribution, $P(\theta|D)$. In this paper, we give an overview about Bayesian methods in empirical finance. We give an introduction to so-called Markov Chain Monte Carlo (MCMC) methods and present recent empirical evidence. Finally, we show the application of MCMC methods in portfolio choice. For an introduction to Bayesian statistics, we refer to ZELLNER (1971) and BERNAR and BERNARDO and SMITH (2000). MCMC methods are covered in

CARLIN and LOUIS (2000) and JOHANNES and POLSON (2003).

2. Markov Chain Monte Carlo

Characterizing $P(\theta|D)$ in asset pricing models is difficult for a variety of reasons. First, prices are observed discretely while most models are formulated in continuous time. Second, state variables are often latent from a researcher's perspective. Third, $P(\theta|D)$ is typically of very high dimension. Fourth, models generate transition probabilities for price and state variables that are non-normal, complicating standard estimation methods such as maximum likelihood or generalized methods of moments. Fifth, in term structure and option pricing models, parameters enter nonlinearly as the implicit solution to differential equations. Markov Chain Monte Carlo (MCMC) methods can tackle all of these issues.

MCMC methods are particularly well-suited for finance applications, in particular for continuous time models, for several reasons. First, continuous-time asset pricing models specify that prices and state variables solve stochastic differential equations, which are built from Brownian motions, Poisson processes, or other i.i.d. shocks. When discretized, these models take the form of time series models with normal, discrete mixture of normal, or scale mixture of normal error distributions. Therefore, standard tools of Bayesian inference are directly applicable. Second, MCMC is a unified estimation procedure which simultaneously estimates parameters and latent variables. MCMC directly computes the distributions of the latent variables and parameters given the observed data. This is a strong alternative to the usual approach of applying approximate filters or latent variable proxies. Third, MCMC allows researchers to quantify estimation and model risk. Finally, MCMC is based on conditional simulation without any optimization.

Consider an inference problem with parameter vector θ and data D . To make inference, we need

to know the posterior distribution $P(\theta|D)$. The key to Markov chain simulation is to create a Markov process whose stationary distribution is a specified $P(\theta|D)$, and to run the simulation sufficiently long so that the distribution of the current values of the process is close enough to the stationary transition distributions. It turns out that, for a given $P(\theta|D)$, many Markov Chains with the desired property can be constructed. Methods that use Markov chain simulation to obtain the distribution $P(\theta|D)$ are called Markov Chain Monte Carlo (MCMC) methods.

In the context of MCMC, the term "parameter" is used in a very general sense. A missing data point, a latent state, or an unobservable time-varying coefficient can be regarded as a parameter in the MCMC framework.

Gibbs sampling (GEMAN and GEMAN (1984)) is perhaps the most popular MCMC method. We introduce the idea of Gibbs sampling by using a simple problem with three parameters, denoted θ_1 , θ_2 , and θ_3 . The goal is to estimate the parameters of a certain model. For example, in a standard linear model, θ_1 might be the intercept and θ_2 and θ_3 the slope parameters. Suppose that the likelihood of the model is hard to obtain, but the three conditional distributions of the single parameters given the other parameters are available. That is, we assume that the conditional distributions $P(\theta_1|\theta_2, \theta_3, D)$, $P(\theta_2|\theta_1, \theta_3, D)$, and $P(\theta_3|\theta_1, \theta_2, D)$ are known. The Gibbs sampler proceeds as follows. Let $\theta_{2,0}$ and $\theta_{3,0}$ denote two arbitrary starting values of θ_2 and θ_3 . The Gibbs sampler draws random values from the conditional distributions. First, the sampler draws $\theta_{1,1} \sim P(\theta_1|\theta_{2,0}, \theta_{3,0}, D)$, then $\theta_{2,1} \sim P(\theta_2|\theta_{1,1}, \theta_{3,0}, D)$, and finally $\theta_{3,1} \sim P(\theta_3|\theta_{1,1}, \theta_{2,1}, D)$. This completes one Gibbs iteration.

Then, using the new parameters as starting values, another Gibbs iteration is completed. Repeating this procedure n times, we obtain a sequence of random draws for the parameters $\{\theta_{1,j}, \theta_{2,j}, \theta_{3,j}\}_{j=1}^n$. Under some regularity conditions, it can be shown that, for a sufficiently large sample, this sequence is approximately equivalent to a random

draw from the joint distribution of the parameters $P(\theta_1, \theta_2, \theta_3 | D)$. The regularity conditions require essentially that for an arbitrary starting value the Gibbs iterations have a chance to visit the full parameter space.

The sequence of random draws of the parameters can be used for inference. In practice, the first m random draws, usually called burn-in period, are discarded from the Gibbs sample to avoid biases from the choice of starting values. Estimates about the mean of a variable can be obtained by computing the sample average of a particular variable, $\hat{\theta}_i = \frac{1}{n-m} \sum_{j=m+1}^n \theta_{i,j}$. Estimates about other statistics of interest such as the standard error can be obtained analogously.

Gibbs sampling decomposes a high-dimensional estimation problem into several lower dimensional ones. At the extreme, a high-dimensional problem with N parameters can be solved iteratively by using N univariate conditional distributions.

In practice, convergence of a Gibbs sampler can be an important issue. Theory only suggest that the convergence occurs when the number of iterations is sufficiently high. CARLIN and LOUIS (2000) provide an overview about convergence diagnostics for Markov Chain Monte Carlo methods.

A number of other algorithms which are more general or more efficient have been proposed over time. Well-known samplers include the Metropolis–Hastings algorithm and the Griddy Gibbs sampler. For a discussion of different samplers, we refer to ROBERT and CASELLA (2005).

3. Empirical Applications of MCMC

MCMC methods are applicable in classical settings such as in regressions and stochastic volatility models. In such cases, the use of MCMC might be helpful for a number of reasons, for example, if estimation risk (parameter uncertainty) or prior information should be taken into account. The Markov Chain generated by MCMC algorithm can be used to analyze the sensitivity of financial models in the context of estimation risk. More-

over, Bayesian methods have exact finite sample properties in contrast to classical approaches.

More advanced applications of MCMC methods are in area of models with latent variables, such as time-varying coefficient models, regime switching models, or jump diffusion models. Overall, many current applications are not comparable to classical statistical approaches because a counterpart does not exist.

JOHANNES and POLSON (2003) show a number of applications of MCMC in financial econometrics. For example, they use MCMC methods to estimate volatility parameters jointly from the prices of options and the underlying, for a multivariate version of jump diffusion models, for time-varying equity premium models, and stochastic volatility models. For fixed income models, they extend the Vasicek model to a jump diffusion setting.

Many researchers have focussed on stochastic volatility models. JACQUIER et al. (1994) apply MCMC to stochastic volatility models. They estimate a model where volatility follows an autoregressive process and compare MCMC to other estimators. They find the Bayesian estimator to be superior. Similarly, JACQUIER et al. (2004) estimate a stochastic volatility model with level effects, fat-tails and correlated errors with a MCMC approach for equity index returns and exchange rates.

ERAKER et al. (2003) examine continuous-time stochastic volatility models for index returns with jumps in return and volatility. They analyze daily S&P 500 index returns from 1980 to 1999 and NASDAQ 100 index returns from 1985 to 1999. Their results suggest that the stochastic volatility and the stochastic volatility model with jumps in volatility are misspecified. The analysis based on Bayes factors suggests that the stochastic volatility model with independent jumps in returns and volatility performs relative well in contrast to the other models. The stochastic volatility model with correlated jumps in returns and volatility was identified as the second best model. Since multivariate

ARCH and GARCH models are usually hard to estimate with classical techniques, VRONTOS et al. (2003) focus on this issue. They find for the Athens exchange market that results from analytical solutions and numerical Bayesian methods are, in general, close to each other. They obtain different parameter estimates if a parameter is close to non-stationarity. Since volatility often behaves close to an unit root process, this is an important finding.

ERAKER (2001) illustrates how to use MCMC methods for the estimation of the constant elasticity of variance model and a stochastic volatility interest rate model. He finds that the stochastic volatility interest rate model provides a superior fit. Similar, KALIMIPALLI and SUSMEL (2004) estimate a regime switching stochastic volatility model to explain the behavior of short-term interest rates. In sample, they find support for this model compared to GARCH models, but out-of-sample results are mixed.

SFIRIDIS and GELFAND (2002) propose a Bayesian financial event study. The advantage of this sampling based approach is that assumptions of classical approaches such as the normality of returns can be easily dropped in a MCMC setting. Similarly, BRAV (2000) proposes a Bayesian test for long-term event studies because standard tests deliver biased signals when applied to long-term event studies. These biases result from cross-correlation and non-normality of returns. He illustrates his method by examining the long-horizon returns of initial public offerings and finds that the FAMA and FRENCH (1993) three-factor model is inconsistent with the observed performance, whereas a characteristic-based model cannot be rejected.

4. An Application to Portfolio Theory

In this section, we apply MCMC to portfolio theory. First, we analyze the effect of parameter uncertainty on portfolio choice. Then, we impose informative priors and show how the portfolio

selection problem changes. The example presented here is similar to the approach proposed by SCHERER (2004).

In a portfolio selection context, the performance of three different models for estimating expected stock returns, the historical mean, a shrinkage or Bayesian estimator and a CAPM-based estimator, is analyzed by JORIN (1991). He finds that the CAPM shows the best out-of-sample performance, the historical sample mean the worst performance, and the Bayesian estimator is in between. Similarly, DUMAS and JAQUILLAT (1990) analyze the performance of currency portfolios chosen by a Bayesian technique which takes estimation and portfolio risk into account. Their results are somehow ambiguous. For some settings they show an insignificant overperformance, while for other settings the overperformance disappears. POLSON and TEW (2000) analyze the portfolio selection problem under informative priors for expected returns and covariance matrices and dynamic rebalancing for the S&P 500 in the time period from 1970 to 1996. They find that the optimal portfolio outperforms the benchmark. Similar, LEDOIT and WOLF (2003) focus on the estimation of a Bayesian covariance matrix. They find in the time period 1972 to 1995 a significant lower out-of-sample variance than for a set of existing estimators.

BLACK and LITTERMAN (1992) suggested an approach to integrate explicitly subjective views about the future performance of certain stocks and equilibrium assumptions into portfolio selection. Starting with expected returns based on the CAPM and so-called views on the relative performance of different shares, they show that the posterior mean of expected returns is a weighted average of views and equilibrium returns. The weights are determined by the perceived degree of dispersion of expected returns from equilibrium and the confidence in views. If the investor has no views on expected returns, the equilibrium distribution, holding the market portfolio, should hold. Recently, SCHERER and MARTIN (2005) extended the

approach by BLACK and LITTERMAN (1992) to a non-normal setting using MCMC methods.

The following example bases on monthly data from January 1990 to October 2005 for 12 stock included in the Swiss Market Index (SMI). Results are reported for the first 8 stocks in the sample. The dataset is from Datastream. Implementation has been carried out in WinBUGS. For an applied introduction to Bayesian statistics based on WinBUGS, we refer to CONGDON (2001, 2003), and LANCASTER (2004). CARLIN and LOUIS (2000) review various Bayesian software packages.

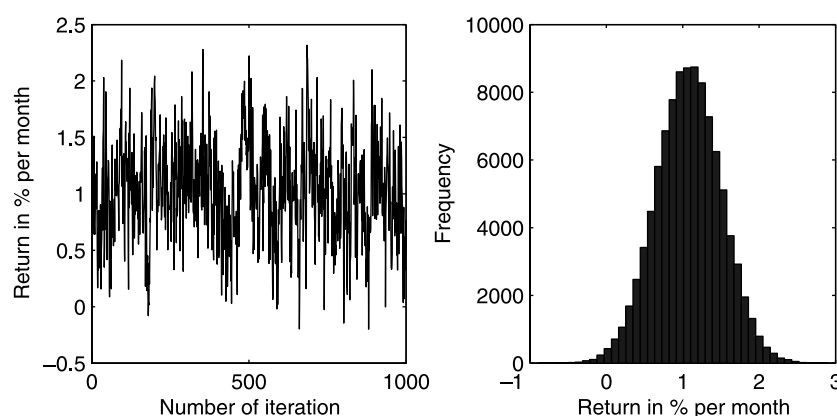
When applying mean-variance optimization in practice, it is necessary to form expected returns, expected volatilities, and expected correlations. An analysis of historical data is usually the starting point to form expectations.

We assume that continuously compounded returns r are multivariate-normally distributed, i.e., $r \sim MVN(\mu, \Sigma)$. To apply MCMC, priors on all parameter have to be imposed, i.e., the vector of mean returns μ and the variance-covariance-matrix Σ . As prior for the precision matrix, i.e., the inverse of the variance-covariance-matrix, a Wishart distribution is used, $\Sigma^{-1} \sim Wishart(R, \nu)$, where R is the prior precision matrix and ν the

degrees of freedom of the distribution. Since we follow for the variance-covariance an uninformative approach, we set R to the identity matrix multiplied by 0.001 and ν to 12, i.e., to the number of assets. This setting assures that no material prior information is incorporated. The usage of the Wishart distribution as prior for the variance-covariance is a common approach in Bayesian statistics. It ensures positive definiteness of the variance-covariance matrix and fast sampling.

The vector of mean returns is updated univariately. Let μ_i denote the mean return for asset i . We assume a normal distribution, $\mu_i \sim N(\mu_{i,prior}, \sigma_{i,prior})$ where $\mu_{i,prior}$ is the prior mean of μ_i , and $\sigma_{i,prior}$ the confidence in this prior estimate. We impose different values on $\mu_{i,prior}$ and $\sigma_{i,prior}^2$. First, we follow an uninformative approach. We set $\mu_{i,prior}$ to 0 and $\sigma_{i,prior}$ to 100. Then, we assume that a portfolio selector has some beliefs about future expected returns, but he does not want to trust fully in historical returns or in a asset pricing model. We assume that the investors centers his beliefs around the CAPM estimate, $\mu_{i,prior} = r_f + \beta_{i,prior} MRP$. For computations, the risk free rate r_f was set to 2% p.a., β_i is the OLS estimate, and expected market premium MRP was set to 5%.

Figure 1: Markov Chains and Histogram Generated by the Gibbs Sampler



The figure shows on the left the Markov chain generated by the Gibbs sampler for the expected return (in % per month) for Novartis. On the right, its histogram is displayed.

We impose two different values on $\sigma_{i,prior}^2$, 1 and 0.2, representing a weak and a strong belief in the CAPM. In a more analytical setting, this issue was recently addressed by AVRAMOV (2004).

Figure 1 shows for one asset, the Novartis share, a part of the Markov Chain generated by the Gibbs sampler (left) and its histogram (right). We used 110,000 iterations and discarded the first 10,000 iterations. We estimate a mean return of 1.08% per month with a standard error of 0.42% for the uninformative case; this equals the OLS estimate. However, this result is not a random finding, but generally holds if underlying assumptions for OLS and MCMC are identical. Similar Markov Chains are created for each parameter in the model, i.e., for the mean returns for all assets and for all elements of the variance-covariance matrix.

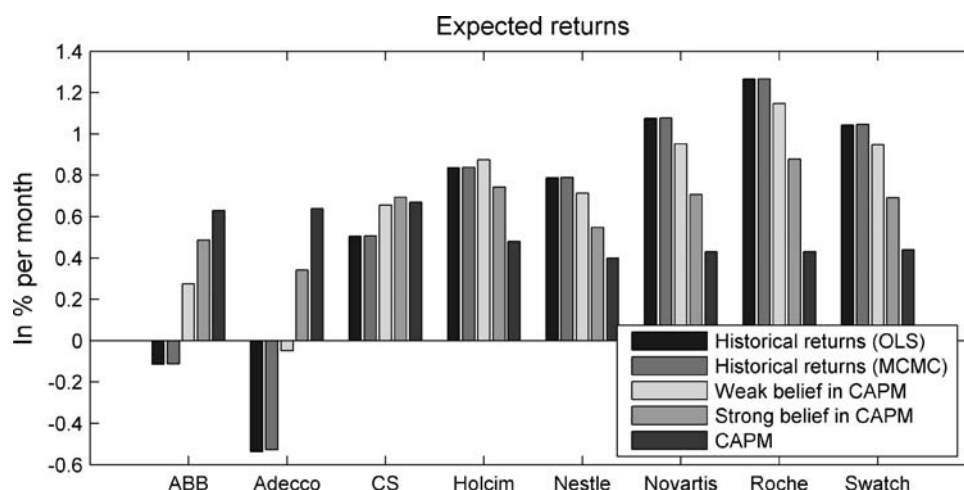
Figure 2 shows the expected returns obtained from an univariate OLS estimation, the MCMC sampler with (1) uninformative priors, (2) a small degree of confidence in the CAPM, and (3) a strong degree of confidence in the CAPM. Finally, we show the CAPM estimated return. Let us note two things. First, the OLS and MCMC results are for the

uninformative case quasi equal. This is expected and indicates that the implementation was carried out properly. Second, the posterior estimate is for the informative cases—as expected—the closer to the CAPM returns the higher the degree of confidence in the prior estimate.

Figure 3 shows the implications for portfolio choice. The benchmarking model is the standard mean-variance optimization approach ignoring parameter uncertainty, and without any prior beliefs. Figure 3 shows on the upper left the distribution of portfolio weights in the tangency portfolio, its mean and the standard error. For each of the 100,000 iterations of the Gibbs sampler, we take the parameter draw and pursue a portfolio optimization. The resulting distribution of portfolio weights therefore reflects the uncertainty in portfolio choice due to parameter uncertainty. The uncertainty due to parameter uncertainty is material.

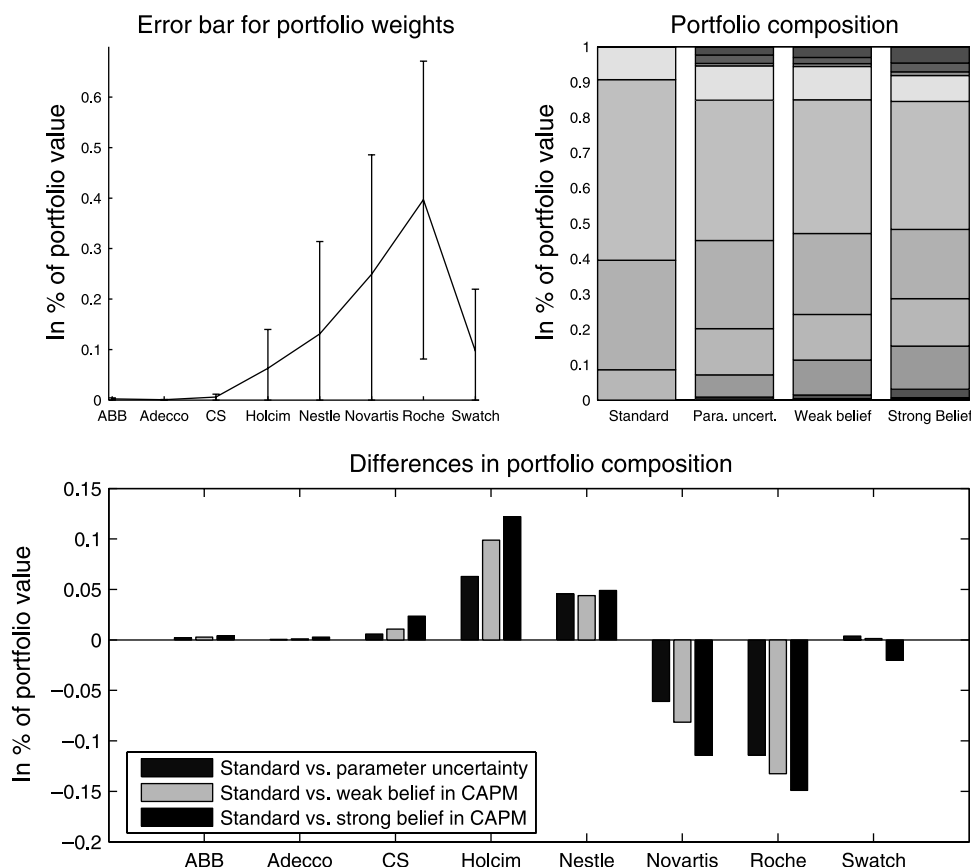
Figure 3 compares the four approaches in the upper right graph. It shows the distribution of assets among stocks for (1) a classical portfolio optimization ignoring estimation risk, (2) for a portfolio

Figure 2: Expected Returns with different Degrees of Belief



The figure shows the expected returns based on a pure historical perspective where for estimation purposes OLS and MCMC have been used, and with a weak, strong or dogmatic belief in the CAPM.

Figure 3: Effect of Parameter Uncertainty and Different Degrees of Confidence in the CAPM on Portfolio Choice



The upper left plot shows the mean of portfolio weights for the tangency portfolio (including the one sigma confidence interval) after accounting for parameter uncertainty. The right plot in the upper part shows different compositions of the tangency portfolio for a standard approach ignoring parameter uncertainty, after accounting for parameter uncertainty, and with a weak and strong belief in the CAPM. The lower plot shows the resulting changes in portfolio weights.

optimization taking estimation risk into account, and, additional to estimation risk, with a (3) weak belief in the CAPM, and (4) a strong belief in the CAPM. All non-classical approaches are more balanced than the classical approach ignoring estimation risk. The lower plot in Figure 3 shows the differences in portfolio weights. The differences are, in general, material and positively connected with the degree of confidence in the CAPM.

5. Conclusion

Bayesian econometric methods are a very powerful technique. They enable the estimation of parameters for almost any model, to account for estimation risk and to incorporate prior beliefs. Moreover, they have exact finite sample properties. Closed form solutions are available for rather simple models, whereas more advanced models require numerical schemes. Therefore, Markov Chain

Monte Carlo (MCMC) methods have drawn considerable attention.

The idea of Markov chain simulation is to simulate a Markov process of the parameter vector which converges to the posterior distribution. This Markov chain can be used to obtain estimates for the means and standard deviations of parameters.

In financial econometrics, a wide range of applications have emerged in the last years. For example, MCMC methods have been suggested for parameter estimation for stochastic volatility models, jump diffusion models and to account for model and parameter uncertainty.

Acknowledgements

I acknowledge helpful comments from an anonymous referee, Manuel Ammann, Bernd Brommundt, and Stephan Süß.

REFERENCES

- AVRAMOV, D. (2004): "Stock Return Predictability and Asset Pricing Models", *Review of Financial Studies* 17, pp. 699–738.
- BERNARDO, J. M. and A. F. SMITH (2000): *Bayesian Theory*: John Wiley & Sons, New York.
- BLACK, F. and R. LITTERMAN (1992): "Global Portfolio Optimization", *Financial Analysts Journal* 48, pp. 28–43.
- BRAV, A. (2000): "Inference in Long-Horizon Event Studies: A Bayesian Approach with Application to Initial Public Offerings", *Journal of Finance* 55, pp. 1979–2016.
- CARLIN, B. P. and T. A. LOUIS (2000): *Bayes and Empirical Bayes Methods for Data Analysis*: Chapman & Hall/CRC, New York.
- CONGDON, P. (2001): *Bayesian Statistical Modelling*: John Wiley & Sons, New York.
- CONGDON, P. (2003): *Applied Bayesian Modelling*: John Wiley & Sons, New York.
- DUMAS, B. and B. JACQUILLAT (1990): "Performance of Currency Portfolios Chosen by a Bayesian Technique", *Journal of Banking and Finance* 14, pp. 539–558.
- ERAKER, B. (2001): "MCMC Analysis of Diffusion Models with Applications to Finance", *Journal of Business & Economic Statistics* 19, pp. 177–191.
- ERAKER, B., M. JOHANNES and N. POLSON (2003): "The Impact of Jumps in Volatility and Returns", *Journal of Finance* 58, pp. 1269–1300.
- FAMA, E. and K. FRENCH (1993): "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics* 33, pp. 3–57.
- GEMAN, S. and D. GEMAN (1984): "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, pp. 721–741.
- JACQUIER, E., N. G. POLSON and P. E. ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business & Economic Statistics* 12, pp. 69–87.
- JACQUIER, E., N. G. POLSON and P. E. ROSSI (2004): "Bayesian Analysis of Stochastic Volatility

- Models with Fat-Tails and Correlated Errors", *Journal of Econometrics* 122, pp. 185–212.
- JOHANNES, M. and N. POLSON (2003): "MCMC Methods for Financial Econometrics", in: Y. Ait-Sahalia, (ed.), *Handbook of Financial Econometrics*.
- JORIN, P. (1991): "Bayesian and CAPM Estimators of the Mean: Implications for Portfolio Selection", *Journal of Banking and Finance* 15, pp. 717–727.
- KALIMIPALLI, M. and R. SUSMEL (2004): "Regime-Switching Stochastic Volatility and Short-Term Interest Rates", *Journal of Empirical Finance* 11, pp. 309–330.
- LANCASTER, T. (2004): *An Introduction to Modern Bayesian Econometrics*: Blackwell, Malden, MA.
- LEDOIT, O. and M. WOLF (2003): "Improved Estimation of the Covariance Matrix of Stock Returns with an Applications to Portfolio Selection", *Journal of Empirical Finance* 10, pp. 603–621.
- POLSON, N. and B. TEW (2000): "Bayesian Portfolios Selection: An Empirical Analysis of the S&P 500 Index", *Journal of Business & Economic Statistics* 18, pp. 164–173.
- ROBERT, C. R. and G. CASELLA (2005): *Monte Carlo Statistical Methods*, 2nd edn.: Springer, Berlin, Heidelberg, New York.
- SCHERER, B. (2004): "Resampled Efficiency and Portfolio Choice", *Financial Markets and Portfolio Management* 18, pp. 382–398.
- SCHERER, B. and D. MARTIN (2005): *An Introduction to Modern Portfolio Optimization*: Springer, Berlin, Heidelberg, New York.
- SFIRIDIS, J. M. and A. E. GELFAND (2002): "A Survey of Sampling-Based Bayesian Analysis of Financial Data", *Applied Mathematical Finance* 9, pp. 273–291.
- VRONTOS, D., P. DELLAPORTAS and D. PLOITIS (2003): "Inference for Some Multivariate ARCH And GARCH Models", *Journal of Forecasting* 22, pp. 427–446.
- ZELLNER, A. (1971): *An Introduction to Bayesian Inference in Econometrics*: John Wiley & Sons, New York.



Michael Verhofen studied Business Administration at the University of St. Gallen and received his masters degree in 2002. Since October 2002 he is a doctoral student at the University of St. Gallen and works as a research assistant at the Swiss Institute of Banking and Finance. His research interests are in the field of Financial Econometrics, Asset Pricing and Financial Derivatives.