Time Varying Parameter Models for Inflation and Exchange Rates

ISBN 90 517 0829 7

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

This book is no. **256** of the Tinbergen Institute Research Series. This series is established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Time Varying Parameter Models for Inflation and Exchange Rates

Tijdsvariërende parameter modellen
voor inflatie en wisselkoersen

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof.dr.ir. J.H. van Bemmel
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
donderdag 13 september 2001 om 16.00 uur

door

Charles Steven Bos

geboren te Lima, Peru

Promotiecommissie

Promotoren:	Prof.dr. H.K. van Dijk
	Prof.dr. P.H.B.F. Franses
Overige leden:	Prof.dr. L. Bauwens
	Prof.dr. T. Kloek
	Prof.dr. M.J.C.M. Verbeek

To Laure

Acknowledgements

Questions and answers, those are what life at the university seems to be about. In this thesis I tried to answer questions on econometric issues, but a question left unanswered is why I started it all, or better: why it was fun.

It is not that I started the adventure of writing a thesis without much thought. It must have been Daniel Peña, with his stories about chicken and influence measures, and Ivo Steyn, who planted the seed of the idea to do research, while I was writing my master's thesis. But even their enthusiasm was not enough to make me fall for university life straight away: It was only while working for a year as an applied labour market researcher at the GAK that I chose to continue as a Ph.D. student at the Erasmus University. In those days, the thought of returning to university was running through my mind. When I later saw a vacancy for a Ph.D. project, in my own field of Econometrics, in a research group with a good reputation, I could not resist the temptation.

And this proved to be a good decision: Working on the articles, and later on the thesis itself, was fun throughout the years, thanks to the support of many people. First of all, I would like to thank my supervisors. Herman van Dijk was always there to listen to my rants and raves, to help find solutions for problems, to improve papers to satisfy the referees, and mainly to get the thesis finished. At the few moments I was looking for a new project to take up, Philip Hans Franses had some interesting ideas which fit nicely into my work. In my daily work, I was given lots of support on practical, scientific, computational or musical issues by Marius Ooms, who was a third member of the supervising committee: Thank you.

Back at the university, or more specifically at the Econometric Institute and both locations of the Tinbergen Institute, many people crossed my way. My roommate (whenever I was not in Amsterdam, and he was not abroad), Dick van Dijk, clearly deserves special mention, as he bore with me for such a long time, answering my queries, and also co-authoring several concepts of papers. One day we'll finish one of those.

At the Tinbergen Institute, Louise was always in for coffee, Silva for a swim, and Jeroen provided me with computing tricks and power, plus technical jokes. At the Econometric Institute, Richard was my main source of solutions for sampling problems, and Christiaan taught me to teach. Outside these Institutes, I benefitted from discussions with my co-authors Ronald Mahieu and Luc Bauwens, and from talks with Michel Lubrano, and many seminar participants. I enjoyed the company and support of many others whom I fail to mention here, and I'm thankful for your help and friendship.

Even though I had a good time at work, I could not have coped without some time off. Outside, Gosse and Annigje persuaded me often to turn off the computer and have a drink, a chat, or an ice-cream. In that respect I also want to thank Liesbeth, Eric,

Ingrid, Wendelien and Hilde, and the gang from the Lorca choir: It was great to have that fixed point in my week, singing every Monday night no matter what the deadline at work was, meeting with René, Margreet, Ruud, Noeke and all the others. Apart from this frequent outing, there were less frequent but not less important visits to the Madrilenian gang, who would have places to live, games of squash, and also very international parties arranged for me: Gracias por todo! Along with the international contacts I can also mention my family, as so many of us live abroad. Smoking a cigar in Angola, drinking a pint in London, and escaping to Eindhoven (which according to some lies in a foreign country as well) to visit Hanna, Michael and Thirza, Jan Willem, or my parents. Without your help, the moments of rest far away from Econometrics, finishing this thesis would have been a lot harder.

Amsterdam, June 2001

Contents

Chapter 1

Introduction

1.1 Motivation

In the economy, the relations between economic agents are subject to change. As the knowledge of production techniques improves, as the means of transportation allow for long-distance trade, as the society changes its preferences for certain goods or services, the structure of the economy varies accordingly. One of the goals of econometrics is to describe this structure of the economy, and to track possible changes in the relationships between the economic variables. If the basic structure of (a part of) the economy is understood, this structure can be used to construct a decision-theoretic framework for policy analysis.

Ever since econometrics started as a separate discipline in the 1930s, it has tried to describe the relations between economic agents. Originally, only small data sets with linear relations could be handled, but quickly knowledge and technology improved, allowing for the analysis of larger regression models, or of time series using e.g. stochastic difference equations and distributed lag models.

Gradually models became more involved, more flexible, and, as a result, they were better able to describe more intricate relations between economic variables. A simpler model might e.g. assume a constant exchange rate between the currencies of two countries, and break down if after a number of months one currency would depreciate strongly with respect to the other currency. A more elaborate model could be able to faithfully describe a larger part of the economic structure, and be robust against changes in the value of the currency of a country.

One can distinguish between two types of changes in a model of the economy: First, there can be a change within the *structure* of the model, necessitating the use of a more elaborate model, to describe correctly this more general structure. Second, there can be a change in the *data space*, meaning that a parameter in the model changes in value, due to a change in a variable. In this thesis, the focus is on changes in time series models, of the structural type in the first chapters, and in the data space in chapter 5.

The next sections provide a short introduction to the research questions of the thesis. They also outline why more standard models and methods of analysing may not be sufficient for the question at hand, and why the choices proposed in this thesis promise to provide better, more detailed, results.

1.2 Inflation: Long memory, changing means and changing variances

In recent decades a vast literature has been published on the subject of deciding whether a macroeconomic time series is stationary (also called 'integrated of order 0, I(0)') or non-stationary (often meaning 'integrated of the first order, I(1)'). The discussion started by the article by Nelson and Plosser (1982), and continues in e.g. Perron (1989) and Perron and Vogelsang (1992). This discussion is closely connected to the Box-Jenkins (1970, 1994) methodology of modelling time series. In this methodology, the time series y_t is modelled as a linear function of past observations $y_{t-1}, .., y_{t-p}$ and present and past disturbances $\epsilon_t, .., \epsilon_{t-q}$ $(p, q \geq 0)$. If the time series y_t is non-stationary, first differences are taken as often as necessary to get to a stationary series.

Part of the success of the ARIMA framework can be attributed to the fact that it provides a common framework with clear notation which can be used for many kinds of time series. In the continuation of this introduction, and in the first chapters of the thesis, the notation is used extensively. Therefore we introduce it here briefly. For a more detailed explanation, see e.g. Franses (1998).

The lag operator L is central in the notation. It serves to lag an observation y_t one time period, i.e. $Ly_t = y_{t-1}$. With the lag operator, taking first differences of the data is written as $y_t - y_{t-1} = y_t - Ly_t = (1-L)y_t$. The first difference operation can be applied d times to get dth order differences: $(1-L)^d y_t$. The autoregressive (AR) part of the model filters the observations in a (linear) manner such that a function of only disturbances results, e.g. $f(\epsilon_t, .., \epsilon_{t-q}) = y_t - \phi_1 y_{t-1} - .. - \phi_p y_{t-p} = (1 - \phi_1 L - .. - \phi_p L^p)y_t$. Past and present disturbances influence the observation through the so-called moving average (MA) part of the model, with $f(\epsilon_t, .., \epsilon_{t-q}) = \epsilon_t + \theta_1 \epsilon_{t-1} + .. + \theta_q \epsilon_{t-q} = (1 + \theta_1 L + .. + \theta_q L^q)\epsilon_t$. Lastly, the basic assumption is that the disturbances ϵ_t are independently and identically distributed with variance σ_ϵ^2. With all of the elements taken together, the notation of an ARIMA(p, d, q) model is

$$\begin{aligned} \Phi(L)(1-L)^d y_t &= \Theta(L)\epsilon_t, \\ \Phi(L) &= 1 - \phi_1 L - \cdots - \phi_p L^p, \\ \Theta(L) &= 1 + \theta_1 L + \cdots + \theta_q L^q, \\ \epsilon_t &\sim i.i.d.(0, \sigma_\epsilon^2). \end{aligned} \tag{1.1}$$

For ease of notation, the autoregressive and moving average polynomials are denoted by $\Phi(L)$ (of degree p) and $\Theta(L)$ (of degree q) respectively. These polynomials have their roots outside the unit circle, see e.g. Box, Jenkins and Reinsel (1994) for details.

The parameter d governs the stationarity of the series. In the original framework of Box and Jenkins (1970), d was restricted to be integer, with $d = 0$ for stationary series or $d = 1$ for non-stationary, I(1) series. In a simple ARIMA(0,d = 0,0) model, $y_t = \epsilon_t$, and clearly the model is stationary. When $d = 1$, the model becomes $(1-L)y_t = \epsilon_t$, or $y_t = y_{t-1} + \epsilon_t = \sum_{i=0}^{t} \epsilon_i$. Such a model is clearly nonstationary, as the effect of a shock never dies out. In Granger and Joyeux (1980) the generalization to non-integer values of d was introduced. It can be derived that, for both integer[1] and non-integer d, the dth

[1]For integer d, it is not trivial to see that equation (1.2) holds. Courant (1954, p. 335) elaborates

order difference operator can be written as

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} (-1)^k L^k. \tag{1.2}$$

In Beran (1994) the effect of fractional integration on the long term autocorrelation is given: If $0 < d < \frac{1}{2}$, the autocorrelation of an ARFIMA(0,d,0) model at long lags k can be approximated as

$$\rho(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} |k|^{2d-1}, \qquad (k \to \infty). \tag{1.3}$$

The autocorrelations at long term lags go down to zero at a (slow) hyperbolic rate. This contrasts with the original case of the stationary ARIMA$(p, 0, q)$ model, where the autocorrelations for large k decrease at a (quicker) exponential rate,

$$\rho(k) \sim c\gamma^k \qquad (k \to \infty), \tag{1.4}$$

see e.g. Greene (1990).

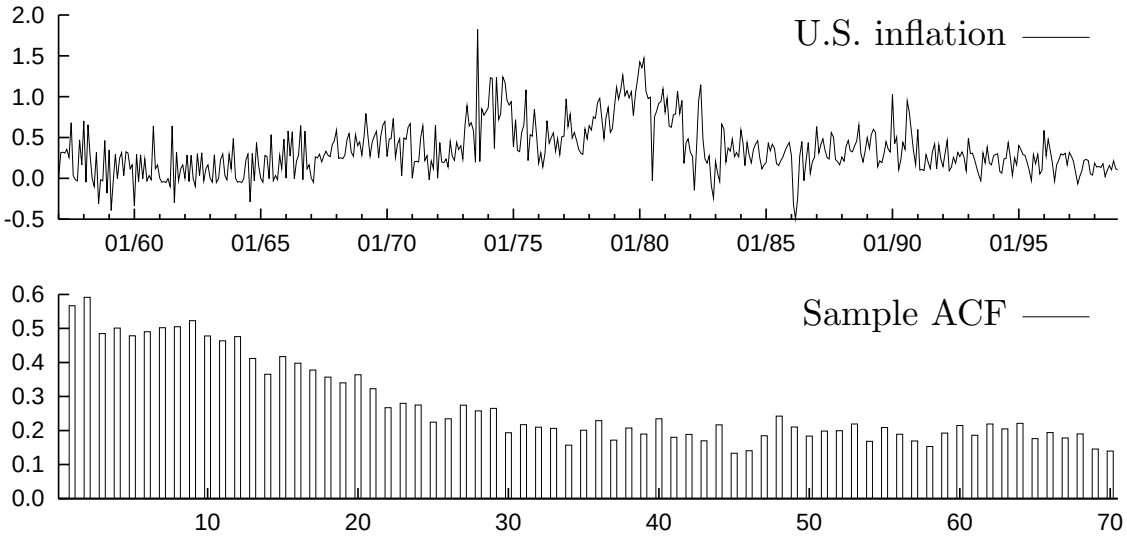

Figure 1.1: U.S. monthly inflation, all items, 1959–1999, with autocorrelation

For inflation series such as the U.S. inflation[2] in the top panel of figure 1.1, it is difficult to decide if the series is stationary (in the sense that $d = 0$ in (1.1)) or not. If it were stationary, then observing a prolonged period with higher inflation as in the seventies is not likely. On the other hand, if inflation were I(1), then the implication is that (the logarithm of) prices are integrated of the second order. Except for periods of

the necessary theorems to show that indeed the equation is correct for ordinary first or second order differences as well.

[2]Source: Bureau of Labor Statistics, series `CUUR0000SA0`, consumer price index U.S. city average. The series was transformed to monthly inflation figures adjusted for seasonality, see chapter 2.

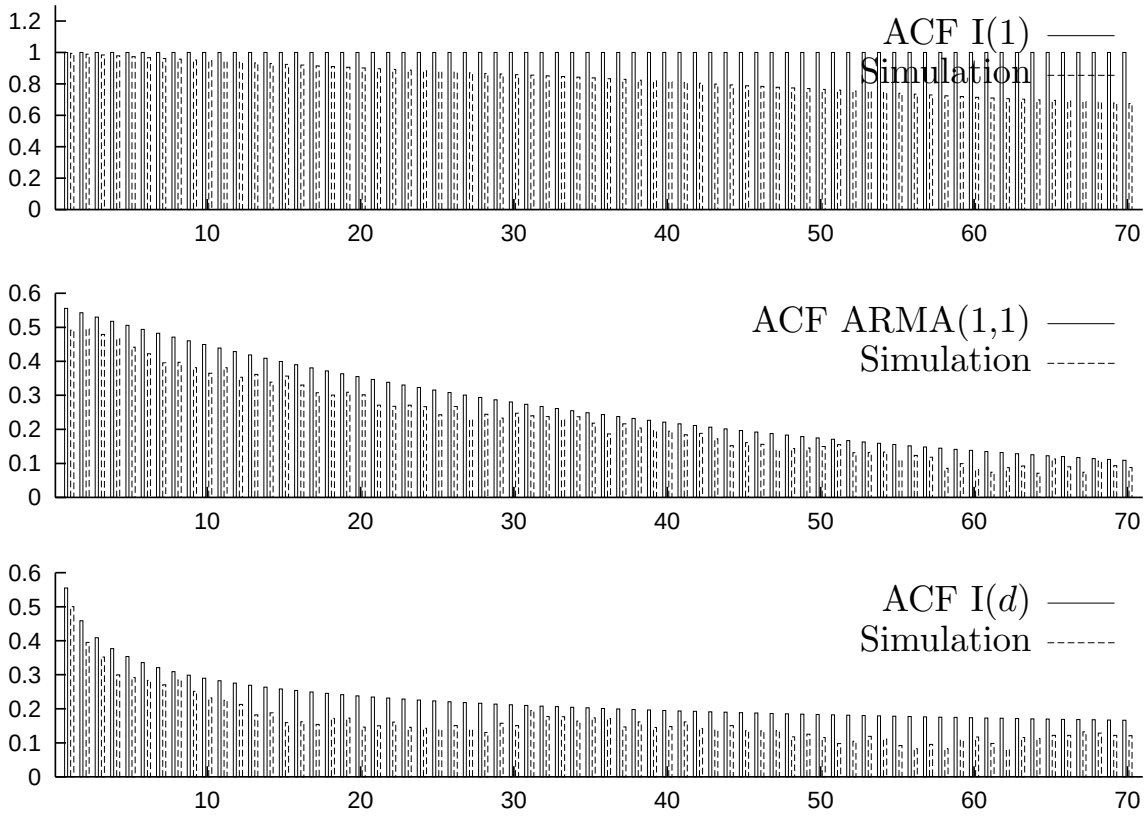

Figure 1.2: Theoretical and simulated autocorrelation function of I(1), ARMA(1,1) and I(d) series

hyperinflation, such behaviour is usually not observed for price indices (see Hassler and Wolters 1995).

The autocorrelation function (ACF) of the series (see the bottom panel of figure 1.1) can often help in getting an impression whether the series is stationary. Clearly the underlying inflation series is not integrated of the first order, as in that case the autocorrelation function stays at one perpetually, at least in theory. This theoretical ACF is drawn in the top panel of figure 1.2, together with the empirical sample autocorrelation function of a sample of length $T = 2000$ of an I(1) model.[3] Short memory ARMA models (that is, models which are integrated of order $d = 0$) display autocorrelation functions which go down to zero at an exponential rate, see equation (1.4). By picking special values for ϕ and θ the autocorrelation function may diminish rather slowly. The second panel of figure 1.2 displays both the theoretical and a simulated sample ACF for an ARMA model with parameters $\hat{\phi} = 0.98$ and $\hat{\theta} = -0.78$ as estimated from the inflation data. Comparing this second panel of figure 1.2 to the sample autocorrelation function in figure 1.1 we see that the shape of the autocorrelation function is not yet quite right: At short horizons, both figures seem to display reasonably similar correlation, but at larger horizons, the

[3]For highly correlated time series, the sample autocorrelation functions tends to underestimate the autocorrelation function of the data generating process, as is seen from the panels in figure 1.2.

correlation according to the ARMA model decreases too fast.

The long term correlation can be adjusted by allowing for fractional integration (see Granger and Joyeux 1980), with d taking non-integer values. For values of $d < \frac{1}{2}$ the resulting ARFIMA(p, d, q) model is still stationary (meaning that the series has a finite mean and variance), but with longer lasting correlations than in the short memory ARIMA$(p, 0, q)$ model. When $d \geq \frac{1}{2}$ the model is non-stationary. The third panel in figure 1.2 shows the autocorrelation a sample of an I(d) model, with d estimated at 0.36. Especially on longer horizons this model fits the sample ACF of inflation well; for the short-term correlation, the fit can be improved using AR and MA parameters.

In chapter 2 the ARFIMA framework is used for analysing inflation the G7 countries. Indeed the parameter d is found to be significantly different from zero. However, in these series changes in the mean seem to have occurred. The contribution of the chapter is the investigation into the effect of these possible changes in the mean, due to the oil crises, on the amount of integration present in the data.

The chapter starts with a simulation exercise, using a simple AR(1) and ARFIMA(1,d,0) model as data generating process, to investigate what could be the effect of a neglected break in the generated series on the estimate for the parameter d. It is found that neglecting a break leads to a positive bias in the estimate $\hat{d}$. As a consequence, when we take the changes in the mean of inflation during the 1970s into account for the G7 countries, we find less evidence for fractional integration.

The importance of estimating the degree of integration correctly lies in the consequence for predictions. Policy makers are in general interested in getting an impression of future inflation, both at short (e.g. 1 month) and longer (say two-year) horizons. With a stationary short memory series one can be quite sure that even at longer horizons the series will stay close to the overall mean: Deviations are temporary as the effect of shocks dies out quickly. For non-stationary series, the uncertainty grows with the horizon: The series may well wander off indefinitely, moving further away from the present value for longer horizons. Long memory time series, with $0 < d < 1$, take up intermediate positions between I(0) and I(1) models, with the variance increasing unboundedly with the horizon if $d > \frac{1}{2}$.

This effect of the degree of integration on predictions of inflation is the topic of chapter 3. U.S. core inflation is predicted up to 24 months ahead, using a range of ARFIMA models, with and without explanatory variables serving as leading indicators. The predictions are compared both on the basis of the precision of the point forecasts, as on the widths of the forecast intervals. We find that the width of forecast intervals derived from ARFIMA models appears to be correct, though we have to allow for temporary shifts in the mean inflation in the seventies and for lower variance of inflation in the Volcker-Greenspan period, after 1983. Non-stationary I(1) models result in forecast intervals which are too wide.

1.3 Exchange rate: Varying trend and variance

From inflation and the internal devaluation of money it is a small step towards exchange rates, governing the external value of money (Bos 1969). The exchange rates of the

German DMark and Japanese Yen vis-a-vis the U.S. Dollar (1982 ≡ 1) are displayed in figure 1.3, together with the cumulative interest rate differentials. Where there is a great controversy about the degree of integration of inflation rates, the opposite is true for exchange rates: Economists agree that (the logarithms of) exchange rates are integrated of the first order. Also, as with many financial time series, it is accepted that exchange rate returns are heteroskedastic, with periods of high volatility followed by more tranquil periods. There is less agreement on the correlation structure of the returns. Theoretically, the expected return should equal the interest rate differential between the home and foreign countries, with no further predictability. However, this uncovered interest rate parity is not found to hold convincingly on daily return data; also in figure 1.3 no clear link between the evolution of the exchange rates and the interest rate differential is found. Does this mean that there is no correlation in the exchange rate returns at all?

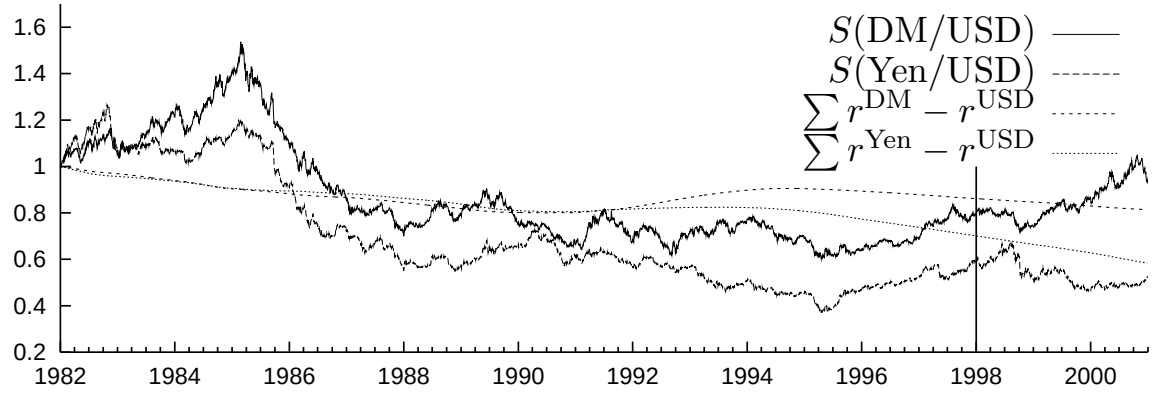

Figure 1.3: Scaled DM/USD and Yen/USD exchange rates, together with the cumulative return of the interest rate differential

Table 1.1: Estimation results for daily DM/USD exchange rate return data

Parameter	AR	FI	ARMA	ARFIMA
ϕ	0.016 (0.015)		0.857 (0.200)	0.993 (0.802)
d		0.021 (0.012)		0.016 (0.037)
θ			−0.844 (0.209)	−0.991 (0.832)
σ_ϵ^2	0.460	0.460	0.460	0.460
$\rho(\phi,\theta)$			−0.9992	−0.9840

Note: Maximum likelihood estimation results on demeaned daily DM/USD exchange rate returns 1/1/1982–31/12/1997, with standard errors, calculated using the ARFIMA package (Doornik and Ooms 1999). Last row reports the correlation between estimates of ϕ and θ.

Table 1.1 reports the maximum likelihood results of estimating basic ARFIMA models on the exchange rate returns. From the estimate of ϕ for the AR model it is seen that there is very little autocorrelation in the returns. Estimating the parameter d of an ARFIMA(0,d,0) model shows that indeed the order of integration of the returns is approximately 0, no clear sign of long memory is found in the data. The results of the ARMA

model deserve some extra attention. At first sight, results are significant with large values for the AR and MA parameters. At second sight, it may strike the researcher that this is a clear case of (near) root cancellation, with the AR polynomial $(1-\phi L)$ (nearly) cancelling against the MA polynomial $(1+\theta L)$.[4] Allowing for fractional integration as in chapters 2-3 does not alter this result.

In chapter 5 (chapter 4 is referred to below), exchange rates and the interest rates of the home and foreign countries are taken as input for deciding if a company is expected to be better off hedging its currency exposure or not. Hedging is advisable for companies with large exposures in a foreign currency when the expected return of the currency is low (taking the uncertainty of the return into account); when a positive return is expected, it is often better not to hedge. For building such a decision framework, the AR(F)IMA models do not serve well, as we saw from the imprecise parameter estimates in table 1.1. Therefore, two alterations in the modelling strategy are made:

i. We use a structural model to model the varying local mean of the exchange rate returns (or the local trend of the logarithm of the exchange rate);

ii. We use a Bayesian approach to incorporate the large parameter uncertainty concerning the parameters modelling the local mean/trend.

The first change is meant to provide a better description of the important characteristics of the data: For a hedging decision, we want to separate underlying trending[5] behaviour from the irregular component of the disturbances. Also, in the structural model it is possible to have only the variance of the observation equation varying over time, leaving the transition equation (1.6) untouched. The exchange rate returns are modelled as

$$s_t = \mu_t + \epsilon_t, \qquad \text{Return} = \text{Expected return} + \text{Disturbance}, \tag{1.5}$$

$$\mu_t = \rho\mu_{t-1} + \eta_t, \qquad \text{Expected return} = \rho \times \text{Previous expectation} + \text{Disturbance}. \tag{1.6}$$

With a so-called structural model (Harvey 1989), the time varying mean return contains information on the trending of the exchange rate levels. When the disturbances are Gaussian with constant variance, the structural model is an alternative representation for an ARIMA model.[6]

In chapter 5 we allow for time variation of the variance of the disturbance term ϵ_t; details are provided in the chapter. With such varying variance, or with a heavy-tailed disturbance distribution, the direct link between the state space model and the ARIMA model no longer holds.

[4] Not only the parameter ϕ is estimated close to $-\theta$, but also the correlation between the two parameters is extreme, see the bottom row of the table. In such a case, often the estimation procedure is not robust. Using a different optimization method may well lead to different outcomes, e.g. EViews 3.1 reports estimates of $\phi = 0.927$ (0.064) and $\theta = -0.917$ (0.068) for the same ARMA(1,1) model.

[5] Notice that the (local) trending of the exchange rates S_t in figure 1.3 corresponds to a (temporarily) non-zero mean of the exchange rate returns s_t.

[6] The structural model can be rewritten as $s_t = \frac{\eta_t}{1-\rho L} + \epsilon_t \Leftrightarrow (1-\rho L)s_t = \eta_t + (1-\rho L)\epsilon_t$. This model displays the same correlation structure as the ARMA(1,1) model $(1-\phi L)s_t = (1+\theta L)v_t$. The ARMA model in table 1.1 corresponds to a GLL model with parameters $\rho = \phi = 0.857, \sigma_\epsilon = 0.673, \sigma_\eta = 0.044$.

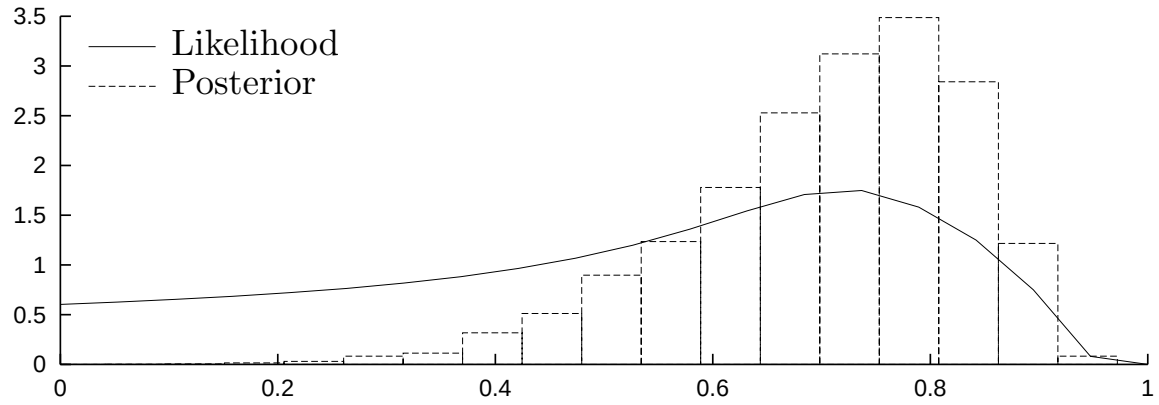

Figure 1.4: Uncertainty for the parameter ρ governing autocorrelation in equation (1.6). The figure displays the likelihood of the parameter ρ marginal of parameters σ_ϵ and σ_η, and the marginal posterior density of ρ.

The second alteration in comparison to chapters 2-3 is the type of statistical analysis that is used. Where in the earlier chapters the data contained enough information to find estimates for the parameters, in chapter 5 the exchange rate data is little informative about the parameters in equation (1.6) describing the local trend. In section 5.6.1 a signal-to-noise ratio is reported, which describes the strength of the signal μ_t in comparison to the noise ϵ_t. This ratio is computed as around 2%, indicating the scarcity of information on μ_t in the data.

In a classical statistical analysis, inference based on a vector of parameter estimates (possibly adjusted for results of a sensitivity analysis). The parameters are often estimated through the method of maximum likelihood. Figure 1.4 displays the likelihood[7] of the parameter ρ governing the autocorrelation in (1.6). The classical analysis would use the top of the likelihood function, disregarding the large spread of the likelihood over values of ρ.

With a Bayesian analysis, a posterior density of the parameters is constructed (see the histogram in figure 1.4). The uncertainty about the value of ρ is taken up in the procedure for deciding whether we want to hedge or not, as the decision is made over all possible values of the parameters, integrating them out with respect to their posterior density.[8]

In chapter 5 extensive use is made of a range of Bayesian simulation techniques. These are explained in detail in chapter 4. Apart from the simulation techniques, the chapter provides information on methods useful for comparing models in a Bayesian fashion, for assessing convergence, and especially on implementing the algorithms. The theory in this

[7]Actually, the likelihood displayed is the likelihood marginal of the parameters σ_ϵ and σ_η in the Generalized Local Level model (1.5)-(1.6), with normally distributed disturbances ϵ_t and η_t. The top of the likelihood function lies at $\hat{\rho} = 0.864, \hat{\sigma}_\epsilon = 0.673, \hat{\sigma}_\eta = 0.044$, even though the top of the likelihood function after marginalising over σ_ϵ and σ_η lies at $\rho = 0.72$.

[8]Using bootstrap methods and an elaborate sensitivity analysis, part of the uncertainty concerning the parameters can be taken into account in a classical analysis. However, this tends to be harder than using a Bayesian approach.

chapter is put to action in an elaborate example.

1.4 Overview

The previous sections of the introduction provided the gist of what is to come in following chapters. In short, this thesis deals with changing parameter models for economic time series of inflation and exchange rates. In chapter 2 (based on Bos, Franses and Ooms 1999, published in Empirical Economics) the relation between a changing mean and the degree of integration in inflation series is investigated. Building on these results, chapter 3 (Bos, Franses and Ooms 2001, published in the International Journal of Forecasting) introduces explanatory variables in the ARFIMA model, to estimate the uncertainty of predicting inflation along the lines prescribed by the Phillips curve. Chapter 4 prepares the Bayesian methodology to be used in chapter 5, with an example concerning a state space model with an unobserved changing mean component. Part of the chapter, concerning the Adaptive Polar Sampler (section 4.3.4) is based on Bauwens, Bos and Van Dijk (2000). In chapter 5 (part of the material in this chapter is published in the articles of Bos, Mahieu and Van Dijk (2000*a*, 2000*b*), in the Journal of Applied Econometrics and the proceedings of ISBA 2000), the example is elaborated into a full-fledged range of models fitting time varying trend and volatility behaviour of exchange rates. These models are used for deciding whether it is worthwhile to hedge daily currency risk or not, for DM/USD, Yen/USD, USD/DM and USD/Yen currency exposures. For this purpose, extensive use is made of the Bayesian toolkit of chapter 4.

In the thesis, choices have been made regarding the topics of research. Necessarily many other interesting topics have been left aside. Chapter 6 briefly summarizes what has been done, and provides a list of topics for possible future research.

Chapter 2

Long Memory and Level Shifts: Re-Analysing Inflation Rates

2.1 Introduction

A key application of long memory time series models concerns inflation. For example, Hassler and Wolters (1995) and Baillie, Chung and Tieslau (1996) find convincing evidence for the presence of long memory characteristics in, especially, inflation rates in the G7 countries. Long memory implies that shocks have a long-lasting effect. Similar to the arguments in the literature on unit roots versus mean shifts, see Perron (1989) and Perron and Vogelsang (1992), it may however be that empirical evidence for long memory is caused by neglecting one or more level shifts. Since such level shifts are not unlikely for inflation, where these may be caused by sudden oil price shocks, we examine whether evidence for long memory (indicated by the relevance of an ARFIMA model) in G7 inflation rates is spurious or exaggerated.

The outline of this chapter[1] is as follows. In section 2.2 we start with a brief motivation by having a closer look at monthly U.S. inflation, thereby extending some recent results summarized in Ooms (1996). In section 2.3 we put forward the relevant theory for testing for long memory and structural level shifts. Our results build on that of Hidalgo and Robinson (1996), who showed that the Wald test is applicable to testing for breaks in a long memory model and on Cheung (1993) since we put forward LM and Wald tests. Section 2.4 deals with a simulation study of the practical performance of the tests. In section 2.5 we apply our tests to monthly G7 inflation rates, where we assume that structural level shifts concur with substantial oil price changes. Our main findings are that apparent long memory is quite resistant to mean shifts, although for a few inflation rates we find that evidence that long memory disappears. In section 2.6, we summarize our findings and relate them to research that appeared after the publication of the article Bos et al. (1999) underlying this chapter.

[1] This chapter is a slightly adapted version of the article Bos, Franses and Ooms (1999), which appeared in Empirical Economics. It is reprinted with kind permission of the publisher.

2.2 A motivation

Consider the monthly U.S. inflation rate in figure 2.1. The data cover January 1957–December 1995 and concern all commodities (source: Bureau of Labor Statistics, series SA0). In the same graph, we draw straight lines that suggest that U.S. inflation has undergone four different regimes. First, until approximately 1967, inflation is stable at a low level. Then the Vietnam war exerts its effect on prices. Inflation is higher, but still quite stable (around this higher level). At the time of the first oil crisis, inflation almost doubles, while at the same time starting to display higher variability. This period of high inflation ends approximately halfway 1981. The final subsample shows a return to earlier inflation levels, although the variability of inflation stays high.

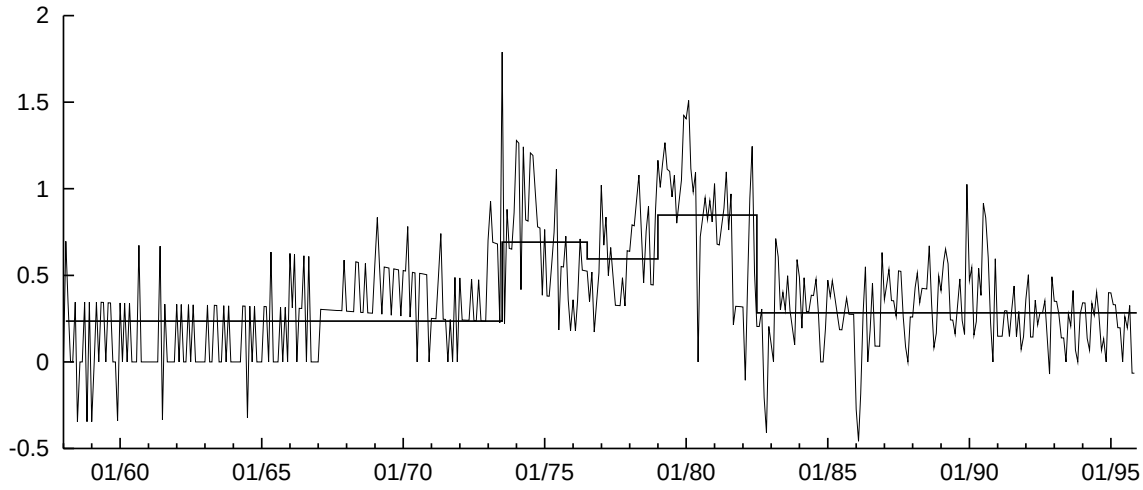

Figure 2.1: U.S. inflation levels

The apparent level shifts might reflect genuine long memory properties intrinsic to inflation. However, the level shifts may also be caused by exogenous events such as the oil crisis. It may also be that the data are better described by a long memory model with mean shifts.

Table 2.1: Estimated values of fractional parameter d and mean shifts β in an ARFIMA model for U.S. inflation

	Number of breaks					
Parameter	0		2		4	
d	0.50	(0.05)	0.40	(0.06)	0.38	(0.06)
$\beta_{1973:07}$			1.85	(0.35)	1.65	(0.40)
$\beta_{1976:07}$					−0.25	(0.44)
$\beta_{1979:01}$					0.97	(0.43)
$\beta_{1982:07}$			−1.71	(0.35)	−2.10	(0.39)

Note: Standard errors are given in parentheses.

To examine the impact of including mean shifts, we estimate an ARFIMA model (of the type discussed in section 2.3 below) for U.S. inflation (adjusted for seasonal means).

We allow for zero, two or four breaks in our ARFIMA model. The timing of the breaks is fixed exogenously and corresponds with shortly before and after the first oil crisis (1973:07 and 1976:07) and shortly before and after the second oil crisis (1979:01 and 1982:07). In table 2.1 we give some key results concerning the fractional differencing parameter d in an ARFIMA model and the parameters for the mean shifts.

The behaviour of the parameter d indicating the degree of fractional integration is interesting. Under the assumption of no mean shifts, we find clear indication of long memory, with an estimate for d even at the border of the non-stationary region. Allowing for two breaks we observe that $\hat{d}$ reduces considerably. Finally, allowing for four breaks does not seem to change much, since $\hat{d}$ obtains about the same value as in the case of two shifts.

2.3 Some theoretical results

The empirical results in table 2.1 evoke interest in the following issues. The first concerns how one would formally address modelling breaks and fractional integration jointly. The second concern is with test statistics that are useful to examine if structural shifts in an ARFIMA model are statistically plausible. The asymptotic distribution of these statistics is then relevant, but also their small sample performance. In this section we deal with these issues, except for the simulation evidence which we postpone to section 2.4.

The ARFIMA model

A fractionally integrated model aims to capture the long memory that is apparent in a time series. Where the influence of a shock in a stationary (I(0)) model disappears after a limited number of periods (depending on the short memory parameters in the autoregressive and moving average parts), and where the effect of a shock lasts forever in a unit root (I(1)) process, the fractionally integrated model (FI(d) with $d \in (0,1)$) takes up an intermediate position, see Granger and Joyeux (1980), Hosking (1981), and more recently, Baillie (1996) and Beran (1994).

The ARFIMA(p,d,q) model is written as

$$\Phi(L)(1-L)^d(z_t - \mu_z) = \Theta(L)\epsilon_t \qquad t = 1,..,T,$$

where z_t is the time series at time t, μ_z its mean, and $\Phi(L) = 1 - \phi_1 L - .. - \phi_p L^p$ is the stable autoregressive polynomial in the lag operator L and $\Theta(L) = 1 + \theta_1 L + .. + \theta_q L^q$ the invertible moving average part ($p, q \in \{0,1,2,..\}$). $\Phi(L)$ and $\Theta(L)$ together define the short memory characteristics of the model. We assume the noise process ϵ_t to be Gaussian, with expectation zero and variance σ_ϵ^2. The long memory behaviour is governed by the part $(1-L)^d$. If d is an integer, the I(d) process is non-fractional, and taking d^{th} differences of z_t leads to a (stationary) I(0) series. If $d \in (0,1)$, one says that z_t exhibits long memory behaviour.

Including a level shift

To allow for a level shift, after a fraction τ $(0 < \tau < 1)$ of the data, we write the observations y_t as the sum of an unobserved ARFIMA process and the term for the level shift:

$$y_t = z_t + \mu I_{\{t>\tau T\}}. \tag{2.1}$$

The parameter μ indicates the size of the level shift in the series y_t at time τT. We define the relative level shift as

$$\beta = \frac{\mu}{\sigma_z}, \tag{2.2}$$

with σ_z being the standard deviation of the ARFIMA process. If the level shift μ and the timing of the break τ are known, this standard deviation can be estimated directly using the empirical standard deviation of the underlying process z_t.

The extension of (2.1) to k breaks is straightforward. We define μ_r as the r-th shift in level, compared to the previous level, and we define the relative breaksize β_r similar to (2.2), where $r = 1, .., k$. When we allow for k level changes at prespecified fractions $0 < \tau_1 < .. < \tau_k < 1$, we can extend (2.1) to

$$z_t = y_t - \sum_{r=1}^{k} \mu_r I_{\{t>\tau_r T\}}.$$

The sample mean of the underlying process z_t is defined as

$$\overline{z} = \frac{1}{T}\sum_t y_t - \frac{1}{T}\sum_t \sum_{r=1}^{k} \mu_r I_{\{t>\tau_r T\}} = \overline{y} - \frac{1}{T}\sum_{r=1}^{k} \lfloor T(1-\tau_r)\rfloor \mu_r,$$

where $\lfloor\cdot\rfloor$ denotes the operator to take the integer part of the argument (the entier function).

In this paper we assume we know the values of $\tau_1, .., \tau_k$. It is possible to endogenize the timing of the breaks, as in Andrews (1993), Bai (1997) or Bai and Perron (1998), but this introduces all kinds of problems concerning the estimation method and the appropriate critical values of the test statistics. Hsu (2000) investigates the same data set and model, with endogenous breakpoints. The results are similar to results presented here.

The spectrum of an ARFIMA model

In order to derive the following results, we assume from now on that z_t is a zero mean, stationary and invertible ARFIMA process, which is obtained from the original data by filtering out the known level shifts, and, if needed, by appropriate differencing, and

subtracting the sample mean $\overline{z}$, i.e.,

$$z_t^a(\mu, \lfloor d + \frac{1}{2} \rfloor) = y_t - \sum_{r=1}^{k} \mu_r I_{\{t > \tau_r T\}}, \tag{2.3}$$

$$z_t^b(\mu, \lfloor d + \frac{1}{2} \rfloor) = (1-L)^{\lfloor d + \frac{1}{2} \rfloor} z_t^a, \tag{2.4}$$

$$z_t(\mu, \lfloor d + \frac{1}{2} \rfloor) = z_t^b - \overline{z^b}. \tag{2.5}$$

We assume that z_t in (2.5) can be described by an ARFIMA(p, d, q) model with d in $[-0.5, 0.5)$.

The autocovariance generating function (ACGF) of an ARFIMA(p, d, q) process is written as

$$g(z; \Psi) = \sigma_\epsilon^2 \left((1-z)(1-z^{-1})\right)^{-d} \frac{\Theta(z)\Theta(z^{-1})}{\Phi(z)\Phi(z^{-1})},$$

see Harvey (1989). To save notation, we use Φ as the set of parameters $\{\phi_1, .., \phi_p\}$ in the polynomial $\Phi(L)$, Θ likewise for the MA parameters and we write Ψ as shorthand for the ARFIMA parameters $\{\Phi, d, \Theta, \sigma_\epsilon^2\}$. We use μ to denote the level shifts μ_1 to μ_r. The spectral generating function (SGF) of the ARFIMA model is given by

$$g(\lambda; \Psi) = \sigma_\epsilon^2 \left|1 - e^{i\lambda}\right|^{-2d} \frac{\left|\Theta(e^{i\lambda})\right|^2}{\left|\Phi(e^{i\lambda})\right|^2}, \tag{2.6}$$

leading to the power spectrum, which is used extensively in the likelihood function as

$$f(\lambda; \Psi) = \frac{1}{2\pi} g(\lambda; \Psi). \tag{2.7}$$

The loglikelihood

With z_t defined as in (2.5), the loglikelihood of the ARFIMA model is

$$\ln \mathcal{L}(y|\Phi, d, \Theta, \mu, \sigma_\epsilon^2) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_T(\Psi)| - \frac{1}{2} z' \Sigma_T^{-1}(\Psi) z.$$

The covariance matrix of z is $\Sigma_T(\Psi) = [\gamma(j-l)]_{j,l=1}^T$, with $\gamma(j)$ the j-th autocovariance of the process z. The loglikelihood depends only on the level shifts through the change from observations y_t to the underlying process z_t. This dependence is not stressed in the notation.

Although it is possible to construct the exact likelihood function in the time domain (see Sowell 1992), we use an approximation in the frequency domain following Harvey (1989). The latter procedure is computationally simpler. In practice, the problem with the calculation of the loglikelihood function is found in the covariance matrix $\Sigma_T(\Psi)$, which is a $T \times T$ matrix. Calculation of its determinant and inverse is time-consuming. Harvey (1989, section 4.3) proposes to use the following approximations:

$$\log |\Sigma_T(\Psi)| = T \log 2\pi + \sum_{j=0}^{T-1} \log f(\lambda_j; \Psi), \tag{2.8}$$

with $f(\lambda;\Psi)$ the power spectrum of the process z_t at frequency λ, $\lambda_j = 2\pi j/T$, and

$$z'\Sigma_T(\Psi)z = \sum_{j=0}^{T-1} \frac{I_z(\lambda_j;\mu)}{f(\lambda_j;\Psi)}, \tag{2.9}$$

where $I_z(\lambda;\mu)$ denotes the break-adjusted periodogram of z_t at frequency λ. In this notation, the dependence of the periodogram $I_z(\lambda;\mu)$ on the level shifts μ is made explicit again, see appendix 2.A.

When calculating the elements of equations (2.8)-(2.9), elements at frequency zero are disregarded, as advocated in Beran (1994), hence the summations start at $j = 1$. Taking all above results together leads to Whittle's approximative loglikelihood function, denoted by

$$\begin{aligned}
&\ln \mathcal{L}(y|\Phi, d, \Theta, \mu, \sigma_\epsilon^2) \\
&\quad = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|\Sigma_T(\Psi)| - \frac{1}{2}z'\Sigma_T^{-1}(\Psi)z \\
&\quad \approx -\frac{T}{2}\log 2\pi - \frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{j=1}^{T-1}\log f(\lambda_j;\Psi) - \frac{1}{2}\sum_{j=1}^{T-1}\frac{I_z(\lambda_j;\mu)}{f(\lambda_j;\Psi)} \\
&\quad = -T\log 2\pi - \sum_{j=1}^{T^*}\delta_j \log f(\lambda_j;\Psi) - \sum_{j=1}^{T^*}\delta_j \frac{I_z(\lambda_j;\mu)}{f(\lambda_j;\Psi)}.
\end{aligned} \tag{2.10}$$

Here we use the fact that for the power spectrum it holds that $f(\lambda;\Psi) = f(-\lambda;\Psi)$ and that $f(\lambda;\Psi) = f(\lambda + 2\pi;\Psi)$. Furthermore, $T^* = \lfloor T/2 \rfloor$ and weights δ_j are defined as

$$\delta_j = \begin{cases} \frac{1}{2} & \text{if } j = \frac{T}{2} = \left\lfloor \frac{T}{2} \right\rfloor \\ 1 & \text{else} \end{cases} \tag{2.11}$$

for $j = 1, .., T^*$. The purpose of the weighting is only to make sure that the midpoint of the range of frequencies is not used twice if T is even.

Testing for breaks

Testing whether level shifts occur corresponds with testing a linear restriction on the parameters $\xi = \{\Psi, \mu\} = \{\Phi, d, \Theta, \sigma_\epsilon^2, \mu\}$. The relevant null hypothesis and alternative hypothesis are

$$H_0 : R\xi = \mu = 0, \qquad H_1 : R\xi = \mu \neq 0,$$

where the alternative implies that we are testing against k breaks at prespecified moments. R is the matrix to select the parameters to be restricted from the vector ξ. In this setting, the Wald, Lagrange Multiplier (LM) and Likelihood Ratio (LR) tests can be used. In our case, the parameters under the null hypothesis of no breaks are more easily calculated than under the alternative, as the inclusion of breaks in the likelihood function would imply that for every evaluation of the likelihood, the periodogram $I_z(\lambda;\mu)$ would have to be recalculated. Thus, the LM principle is our first choice for the next section, where the

simulations are presented. In section 2.5, where the analysis is done for the countries of the G7, both the LM and Wald test statistics are used.

Hidalgo and Robinson (1996) have proved that the Wald test statistic on a single structural change in a long memory environment follows a χ^2_1 distribution, when using a non-parametric estimator for d and an OLS estimate (disregarding the value of $\hat{d}$ or any short memory parameters) for the mean of the series before and after the break. Their proof cannot easily be translated to a setting where an iterative generalized least squares or (approximate) maximum likelihood procedure is used. However, as parametric estimators tend to converge at least as fast as non-parametric estimators (assuming a correct model specification under the null hypothesis), this asymptotic χ^2 distribution can be expected to hold in our case too. As the Wald and LM tests are asymptotically equivalent, the results of Hidalgo and Robinson can be translated to the LM test. Finally, the extension to multiple breaks is straightforward, as long as we condition on the number and the timing of the breakpoints. Asymptotically, we expect a χ^2_k distribution for the test on k breaks under the null hypothesis.

The calculation of the test statistics follows the familiar lines:

$$W_\mu = \hat{\mu}' \left(R \, J^{-1}(\hat{\Psi}, \hat{\mu}) R' \right)^{-1} \hat{\mu},$$

$$LM_\mu = \left. \frac{\partial \ln \mathcal{L}}{\partial \xi} \right|'_{\hat{\xi}_0} J^{-1}(\hat{\Psi}_0, 0) \left. \frac{\partial \ln \mathcal{L}}{\partial \xi} \right|_{\hat{\xi}_0},$$

with ξ as defined before. The $\hat{\Psi}$ and $\hat{\mu}$ are the unrestricted estimates of the parameters in the model, whereas $\hat{\xi}_0 = \{\hat{\Psi}_0, 0\}$ is the estimate of the parameters under the null hypothesis of no breaks. The $J(\Psi, \mu)$ denotes the information matrix

$$J(\Psi, \mu) = -\mathsf{E} \begin{bmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial \Psi \partial \Psi'} & \frac{\partial^2 \ln \mathcal{L}}{\partial \Psi \partial \mu'} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \mu \partial \Psi'} & \frac{\partial^2 \ln \mathcal{L}}{\partial \mu \partial \mu'} \end{bmatrix}.$$

In appendix 2.A we provide more details on the calculation of these test statistics.

2.4 Simulation evidence

In this section we report on some simulation evidence concerning the small sample properties of the estimators of the parameters and of the LM test for level shifts.

Two data generating processes

We generate 512 (=T) observations from an AR(1) process with $\phi = 0.8$ and from an ARFIMA$(1, d, 0)$ process with $\phi = 0.4$ and $d = 0.3$ (see appendix 2.B for information on the generator used for the ARFIMA model). The variance σ^2_ϵ of the disturbances is taken to be 1. Halfway the sample, we add a level shift of size $\beta = 1$, i.e., a shock of one time the standard deviation of the underlying process is added to the mean of the series after the observation at $\frac{1}{2}T$. Given the variance of the ARFIMA processes, which is $\sigma^2_z = 2.778$ for the AR(1) model and $\sigma^2_z = 2.357$ in the ARFIMA$(1, d, 0)$ case,

the parameter μ in (2.1) equals 1.667 and 1.535, respectively. For both time series, i.e., AR(1) and ARFIMA$(1, d, 0)$, with and without a level shift, we estimate the parameters of an ARFIMA$(1, d, 1)$ model. Our simulations are based on 5000 replications.

The empirical distribution of the estimators

In figure 2.2 the empirical distribution functions of $\hat{\phi}, \hat{d}$ and $\hat{\theta}$ are shown for data generated according to the AR(1) process without a break. Figure 2.3 shows the estimation results for the same parameters, but estimated for data which include a level shift. The boxes indicate the 5%, 50% and 95% quantiles, whereas the dashed lines cross at the original parameter value.

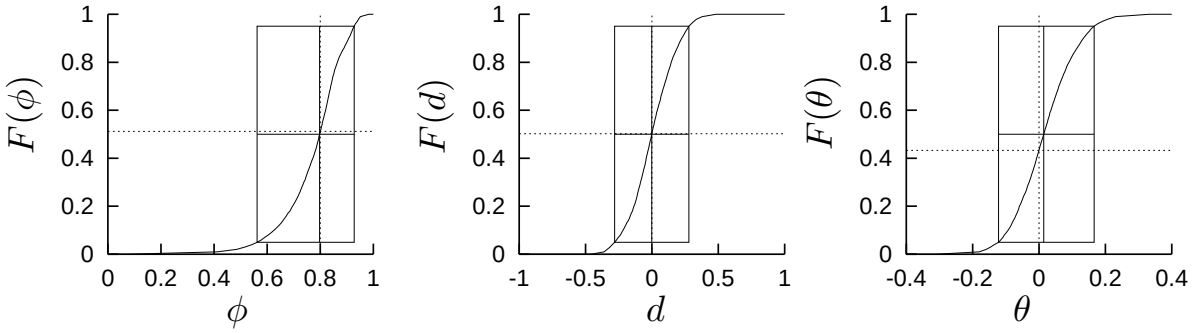

Figure 2.2: Estimating parameters in an ARFIMA$(1, d, 1)$ model. The DGP is an AR(1) model with $\phi = 0.8, \sigma_\epsilon^2 = 1$, without a level shift.

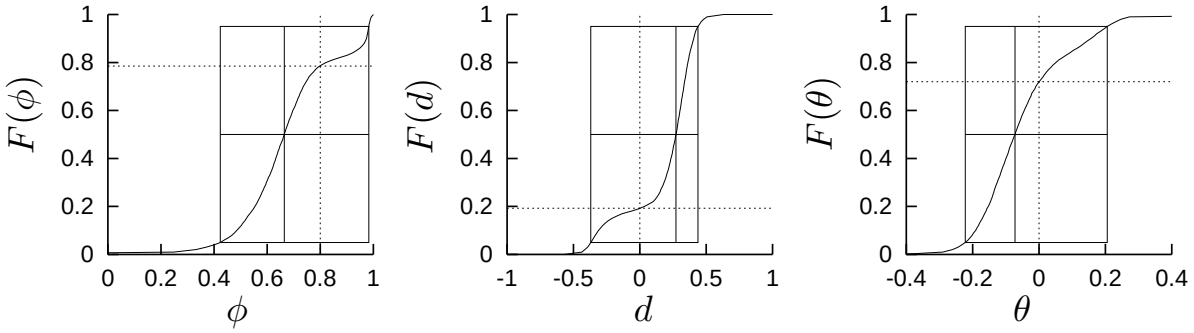

Figure 2.3: Estimating parameters in an ARFIMA$(1, d, 1)$ model. The DGP is an AR(1) model with $\phi = 0.8, \sigma_\epsilon^2 = 1$, with a level shift of size $\beta = 1$.

From the first graph we can conclude that the estimation procedure leads to consistent results. The estimated medians (with the corresponding values in the DGP between brackets) are $\hat{\phi}_{50\%} = 0.80$ [0.8], $\hat{d}_{50\%} = 0.00$ [0.0] and $\hat{\theta}_{50\%} = 0.01$ [0.0]. Including a break in the data generating process (DGP) leads to figure 2.3, which depicts the same estimators. In the left-hand graph we see that in a certain fraction of the simulations, inclusion of a level shift in the DGP leads to data with unit root properties ($\hat{\phi}$ close to 1). A closer examination of the estimation results shows that in these cases d is estimated

around or even below zero. In the other cases, ϕ was estimated below the 'true' value of 0.8, i.e. the median decreases to 0.66. Except for cases where a unit root was found, d is estimated at a value higher than in the previous simulations. Indeed, a correlation of -0.90 between $\hat{\phi}$ and $\hat{d}$ is found. The median $\hat{d}_{50\%}$ shifts upwards to a value of 0.27. Such a value for the degree of fractional integration is in general taken as a strong indication of long memory behaviour. The estimates of the parameter θ decrease somewhat, relative to the original value of zero, while their spread is higher than before. Clearly, neglecting a level shift in otherwise short memory data may lead one to believe that long memory resides in the data.

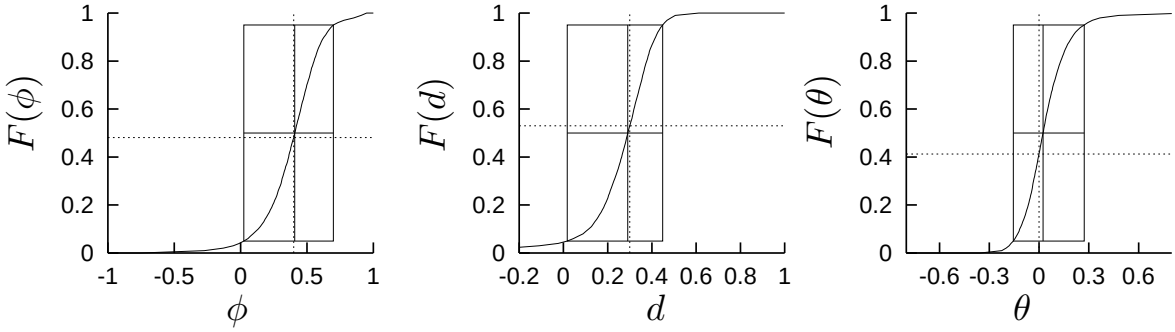

Figure 2.4: Estimating parameters in an ARFIMA$(1, d, 1)$ model. The DGP is an ARFIMA$(1, d, 0)$ model with $\phi = 0.4, d = 0.3, \sigma_\epsilon^2 = 1$, without a level shift.

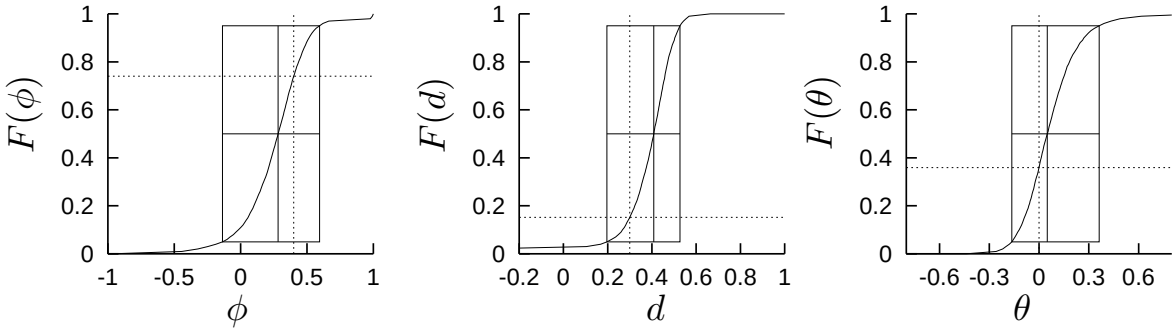

Figure 2.5: Estimating parameters in an ARFIMA$(1, d, 1)$ model. The DGP is an ARFIMA$(1, d, 0)$ model with $\phi = 0.4, d = 0.3, \sigma_\epsilon^2 = 1$, with a level shift of size $\beta = 1$.

When the ARFIMA$(1, d, 0)$ model is taken as the DGP, we obtain the results as given in figures 2.4 and 2.5. Figure 2.4 shows the consistency of the approximative Whittle estimator in the presence of long memory in the data. This result agrees with those reported in Hauser (1999), where encouraging results for this estimator are obtained from an extensive simulation study. The medians (with values of the DGP between brackets) found are $\hat{\phi}_{50\%} = 0.41$ [0.4], $\hat{d}_{50\%} = 0.29$ [0.3] and $\hat{\theta}_{50\%} = 0.02$ [0.0]. For the DGP where a break is included, we observe shifts in the empirical distribution of the estimators. Now, no indication of unit roots is found. The estimates of ϕ shift down to a distribution with 0.28 as a median. Estimated $\hat{d}$ values indicate even stronger long memory characteristics

than found before. The parameter $\hat{\theta}$ does not change much. Again, a positive bias in $\hat{d}$ is found when a level shift is neglected.

The empirical distribution of the LM test statistic for a level shift

Figure 2.6 depicts the empirical density and distribution function of the Lagrange Multiplier test statistic, when the presence of a break is tested in the series simulated according to the AR(1) DGP. The dashed curve in the same graph indicates the χ^2_1 distribution, which is supposed to be the asymptotic distribution of the statistic.

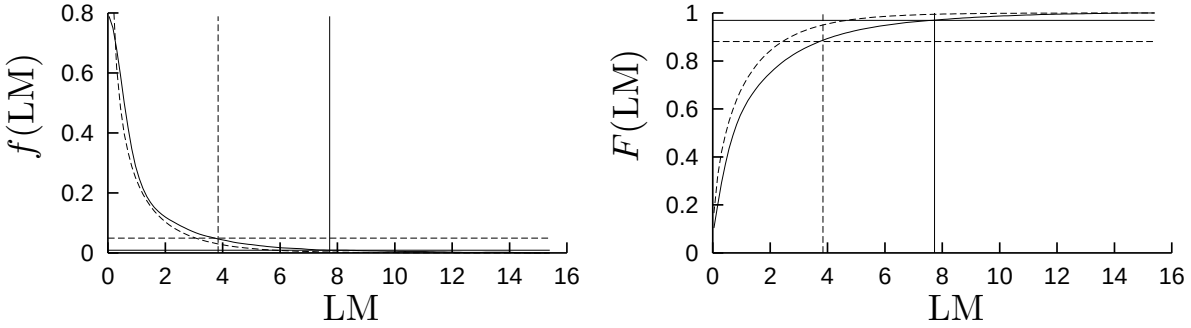

Figure 2.6: Empirical distribution of the LM test statistic. The DGP is an AR(1) model with $\phi = 0.8, \sigma^2_\epsilon = 1$, without a level shift.

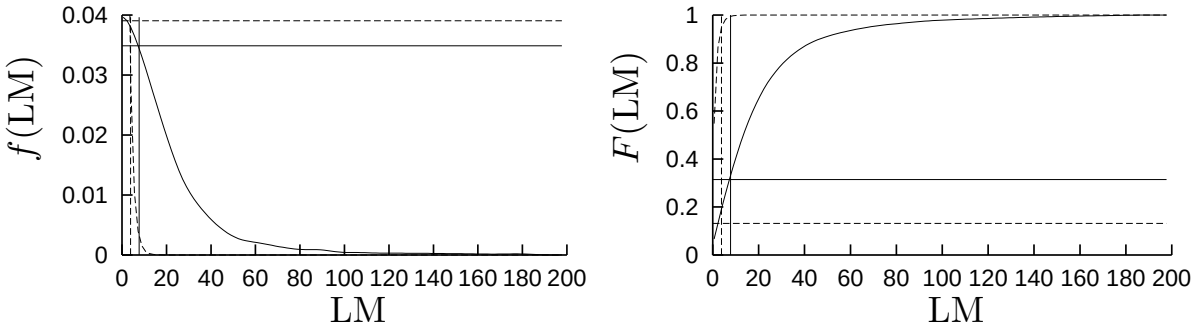

Figure 2.7: Empirical distribution of the LM test statistic. The DGP is an AR(1) model with $\phi = 0.8, \sigma^2_\epsilon = 1$, with a level shift of size $\beta = 1$.

The empirical density function exhibits the same shape as the χ^2 density. However, large values of the test statistic occur too often, as indicated by the heavier tail of the empirical distribution when compared to the χ^2 distribution. The standard 5% critical value for the χ^2 distribution is 3.84. Using this critical value leads to an empirical size of the test of 13%. The empirical 5% critical value is found at a value of the test statistic of 7.73.

The effects of the inclusion of a break in the DGP on the LM test statistic are summarized in figure 2.7. Although the shape of the density function still resembles a χ^2, much larger values of the test statistic are found. The standard χ^2-based critical value

would lead to a (correct) rejection of the null hypothesis of no break in 83% of the cases. Using the empirical critical value of 7.73 reduces the power to 69%, which is still quite reasonable.

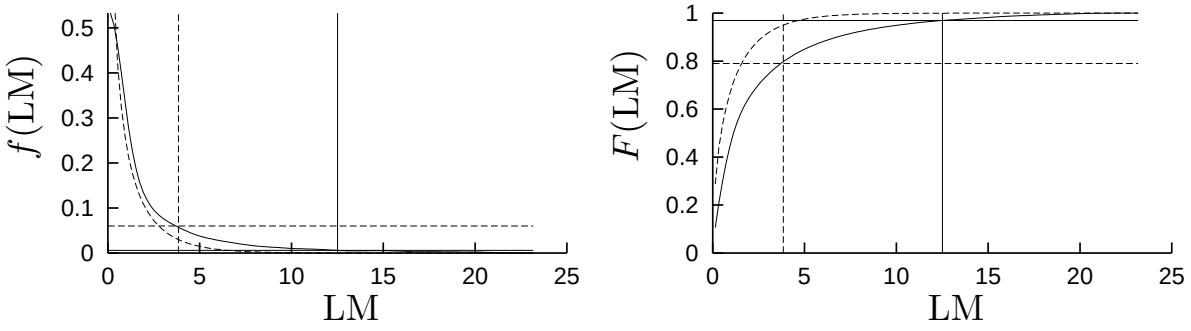

Figure 2.8: Empirical distribution of the LM test statistic. The DGP is an ARFIMA$(1, d, 0)$ model with $\phi = 0.4, d = 0.3, \sigma_\epsilon^2 = 1$, without a level shift.

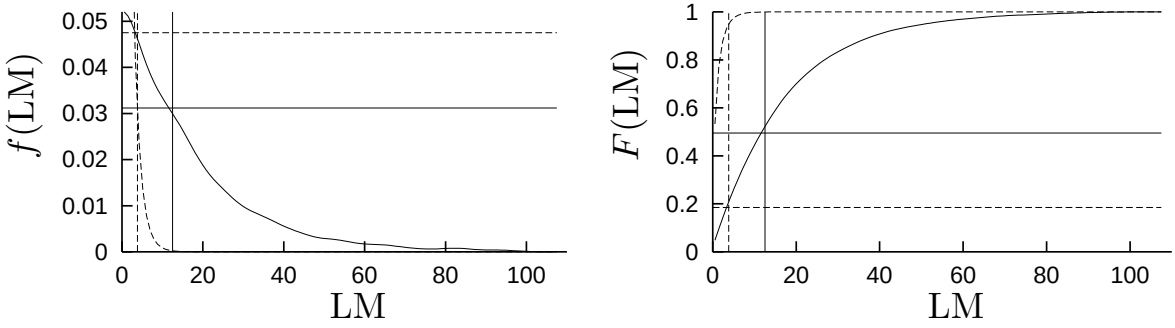

Figure 2.9: Empirical distribution of the LM test statistic. The DGP is an ARFIMA$(1, d, 0)$ model with $\phi = 0.4, d = 0.3, \sigma_\epsilon^2 = 1$, with a level shift of size $\beta = 1$.

When the data are generated according to the ARFIMA(1, d, 0) model, the empirical size of the test (at a nominal level of 5%) is 22% (see figure 2.8). Even when generating under the null, more large values of the test statistic are found. The empirical critical value at the 95% confidence level is 12.51, which is considerably higher than the original 3.84. Finally, when the DGP is the ARFIMA(1,d,0) model with a break, we obtain the results as in figure 2.9. If the critical value of 3.84 is used, the empirical power is still around 80%, as indicated by the horizontal dashed line in the graph on the right hand side of figure 2.9. Using the empirical 5% critical value in this case however lowers the power to 50%. The findings on size and power are summarized in table 2.2. The third column in this table reports the rejection rates for the Beran test for goodness-of-fit, advocated in Beran (1994), at a nominal level of 5%. We interpret the numbers in this column as that this test statistic does not signal important residual correlation in the models fitted to the data.

Our simulations lead to the conclusion that a neglected level shift has a substantial effect on the parameter estimates. The LM test seems to be able to detect a level shift, although the power can be low and some size distortions do occur.

Table 2.2: Empirical rejection frequencies of the LM test statistic for a level shift and the Beran test for residual white noise

DGP: AR(1), $\phi = 0.8$

β	LM > 3.84	LM > 7.73	$p_{\text{Beran}} < 0.05$
0	13.18	5.00	4.34
1	83.36	69.28	4.22

DGP: ARFIMA$(1, d, 0)$, $\phi = 0.4, d = 0.3$

β	LM > 3.84	LM > 12.51	$p_{\text{Beran}} < 0.05$
0	21.84	5.00	4.18
1	80.20	49.50	3.86

Note: Each series consists of 512 observations. The number of replications is 5000.

2.5 Inflation: Long memory and level shifts

In this section we re-analyse part of the series previously used by Baillie et al. (1996). The dataset consists of the Consumer Price Indices (CPI) for the countries of the G7: Canada, France, Germany, Italy, Japan, the United Kingdom and the United States. The data for the U.S. originates from the Bureau of Labor Statistics concerning the overall price index SA0, and it ranges from January 1957 until December 1995. Indices for the other countries are extracted from Citibase. Observations on the months January 1948 until March 1990 are available. Inflation rates are constructed from the price indices by taking $y_t = 100\Delta \log \text{CPI}_t$. As the inflation rates exhibit rather erratic behaviour in the first years of the sample, we only use the data starting in 1958. For the U.S., we have a sample of 456 observations, while for the other countries 387 observations are available. To account for part of the seasonality, the data y_t are first adjusted for seasonal means. The parameters in the ARFIMA models for the resulting series are estimated by optimizing the likelihood as described in section 2.3. As the estimation is done in the frequency domain, this adjustment for seasonality corresponds to putting the periodogram to zero at the seasonal frequencies (see Ooms and Hassler (1997)). The timing of the breaks corresponds with the oil price shocks (see table 2.1), and is taken to be equal for all countries.

We aim to consider an ARFIMA model. Several settings with different degrees in the AR and MA polynomials are tried. Our specification search results in a model with AR parameters ϕ_1, ϕ_{12} and ϕ_{13} together with the degree of integration d and residual variance σ_ϵ^2. This model appears to capture the short and long run correlations quite well, as is indicated by the Beran (1994) test for white noise. Specifically, adding moving average parameters often leads to root cancellation, and hardly improves the residual variance.

In table 2.3 the results of the estimations are presented. For each country we consider three models. First, the pure ARFIMA model is considered. The parameter $\hat{\phi}_{13}$ is significant only in the case of German inflation rates. For the other countries, the parameter can be omitted. The fractional integration parameter d is estimated around the commonly found value of 0.4. For the U.S., we find $\hat{d} = 0.501$. Theoretically, the approximative

Table 2.3: ARFIMA models (with and without level shifts) for monthly inflation rates in the G7 countries

Parameter	Canada		France		Germany		Italy		Japan		UK		USA	
$\hat{\phi}_1$	−0.140	(0.067)	−0.055	(0.073)	0.033	(0.080)	−0.231	(0.060)	−0.224	(0.067)	−0.157	(0.069)	−0.268	(0.060)
$\hat{\phi}_{12}$	0.228	(0.052)	0.139	(0.053)	0.096	(0.052)	−0.093	(0.050)	0.160	(0.049)	0.145	(0.050)	0.078	(0.046)
$\hat{\phi}_{13}$					0.144	(0.051)								
$\hat{d}$	0.324	(0.049)	0.357	(0.056)	0.207	(0.064)	0.445	(0.044)	0.267	(0.056)	0.385	(0.054)	0.501	(0.050)
$\hat{\sigma}_\epsilon$	0.304	(0.011)	0.320	(0.012)	0.257	(0.009)	0.514	(0.018)	0.694	(0.025)	0.530	(0.019)	0.239	(0.008)
$\hat{\sigma}_z$	0.382		0.407		0.286		0.672		0.737		0.646		0.328	
LL	−87.971		−107.836		−22.977		−290.133		−407.146		−302.569		7.598	
Beran/p_{Beran}	0.341	0.157	0.313	0.598	0.337	0.206	0.290	0.888	0.314	0.571	0.349	0.087	0.341	0.137
Two shifts	**Canada**		**France**		**Germany**		**Italy**		**Japan**		**UK**		**USA**	
$\hat{\phi}_1$	−0.021	(0.077)	0.056	(0.089)	0.065	(0.086)	−0.124	(0.074)	−0.200	(0.071)	−0.095	(0.077)	−0.205	(0.068)
$\hat{\phi}_{12}$	0.224	(0.052)	0.137	(0.052)	0.087	(0.052)	−0.118	(0.051)	0.158	(0.050)	0.127	(0.051)	0.066	(0.046)
$\hat{\phi}_{13}$					0.133	(0.051)								
$\hat{d}$	0.114	(0.062)	0.186	(0.076)	0.163	(0.072)	0.274	(0.063)	0.225	(0.062)	0.276	(0.066)	0.400	(0.059)
$\hat{\sigma}_\epsilon$	0.292	(0.011)	0.313	(0.011)	0.256	(0.009)	0.502	(0.018)	0.691	(0.025)	0.521	(0.019)	0.231	(0.008)
$\hat{\beta}_{1973:07}$	1.639	(0.214)	1.244	(0.292)	0.326	(0.261)	1.554	(0.276)	0.368	(0.239)	1.289	(0.294)	1.851	(0.350)
$\hat{\beta}_{1982:07}$	−1.258	(0.235)	−1.446	(0.287)	−0.468	(0.315)	−0.956	(0.302)	−0.521	(0.268)	−0.723	(0.344)	−1.708	(0.354)
$\hat{\sigma}_z$	0.304		0.334		0.275		0.532		0.716		0.557		0.259	
LM/p_{LM}	290.144	0.000	45.950	0.000	2.650	0.266	40.727	0.000	4.936	0.085	29.972	0.000	49.863	0.000
Wald/p_{Wald}	64.957	0.000	32.207	0.000	2.957	0.228	39.890	0.000	4.357	0.113	19.557	0.000	33.786	0.000
LL	−74.028		−99.602		−21.911		−282.305		−405.319		−295.851		21.117	
Beran/p_{Beran}	0.312	0.608	0.309	0.660	0.331	0.291	0.266	0.988	0.315	0.561	0.354	0.060	0.354	0.044
Four shifts	**Canada**		**France**		**Germany**		**Italy**		**Japan**		**UK**		**USA**	
$\hat{\phi}_1$	−0.023	(0.077)	0.075	(0.092)	0.051	(0.087)	−0.127	(0.075)	−0.126	(0.082)	−0.111	(0.079)	−0.202	(0.068)
$\hat{\phi}_{12}$	0.221	(0.052)	0.130	(0.052)	0.093	(0.053)	−0.117	(0.051)	0.140	(0.051)	0.135	(0.052)	0.059	(0.047)
$\hat{\phi}_{13}$					0.138	(0.052)								
$\hat{d}$	0.106	(0.062)	0.161	(0.080)	0.175	(0.076)	0.274	(0.065)	0.117	(0.075)	0.297	(0.075)	0.381	(0.058)
$\hat{\sigma}_\epsilon$	0.291	(0.010)	0.312	(0.011)	0.255	(0.009)	0.502	(0.018)	0.682	(0.025)	0.520	(0.019)	0.230	(0.008)
$\hat{\beta}_{1973:07}$	1.452	(0.279)	1.277	(0.347)	0.028	(0.352)	1.403	(0.358)	0.922	(0.247)	1.086	(0.410)	1.653	(0.402)
$\hat{\beta}_{1976:07}$	0.074	(0.331)	−0.227	(0.385)	0.114	(0.401)	0.107	(0.387)	−0.923	(0.308)	0.185	(0.493)	−0.248	(0.440)
$\hat{\beta}_{1979:01}$	0.432	(0.319)	0.443	(0.386)	0.597	(0.378)	0.247	(0.368)	−0.033	(0.296)	0.173	(0.408)	0.974	(0.435)
$\hat{\beta}_{1982:07}$	−1.529	(0.275)	−1.638	(0.335)	−0.763	(0.351)	−1.117	(0.345)	−0.335	(0.252)	−0.797	(0.379)	−2.097	(0.394)
$\hat{\sigma}_z$	0.302		0.329		0.277		0.531		0.689		0.570		0.255	
LM/p_{LM}	852.230	0.000	49.741	0.000	7.099	0.131	49.586	0.000	29.441	0.000	31.151	0.000	64.879	0.000
Wald/p_{Wald}	73.176	0.000	39.512	0.000	6.154	0.188	41.029	0.000	22.897	0.000	17.873	0.001	42.317	0.000
LL	−72.667		−98.978		−20.130		−281.860		−400.906		−295.601		23.560	
Beran/p_{Beran}	0.312	0.607	0.306	0.705	0.332	0.278	0.268	0.986	0.321	0.458	0.354	0.059	0.362	0.020

Note: Standard errors between parentheses.

Whittle estimator is only consistent in the range of $d \in [-0.5, 0.5)$, although it is known to be little biased if the degree of integration lies just outside the stationary region. The Beran test for the absence of residual correlation, which is reported along with its corresponding p-value, does not indicate strong correlation in the residuals. The $\hat{\sigma}_\epsilon$ reported is the estimated standard deviation of the disturbances. The $\hat{\sigma}_z$ denotes the standard deviation of the ARFIMA process, which in this no break case equals the standard deviation of the process z_t itself.

Allowing for a level shift in July 1973 and July 1982 leads to parameter estimates as reported in the second panel of table 2.3. Apart from the ARFIMA parameters, the sizes of the level shifts μ_r are estimated as well. Subtracting the estimated level shifts from the data leads to the underlying process z_t as in section 2.3. The empirical standard deviation $\hat{\sigma}_z$ of this underlying process is used to calculate the relative break sizes $\hat{\beta}_r$ as defined in equation (2.2). Reported standard deviations of the estimates of β_r are calculated from the original standard deviations of the level shifts $\hat{\mu}_r$, taking $\hat{\sigma}_z$ is given. As $\hat{\sigma}_z$ is not a given, fixed parameter, the true uncertainty about the β's is likely to be somewhat larger. Significant values of $\hat{\beta}_r$ are found in all countries except for Germany and Japan. For most countries, a considerably lower degree of fractional integration is found compared with the no break case. Also, the standard deviation of the residuals $\hat{\sigma}_\epsilon$ and of the underlying process $\hat{\sigma}_z$ is smaller, as expected after inclusion of extra parameters.

We also calculate the LM and Wald test statistics for the absence of level shifts. For each country, the value of the statistic and the corresponding p-value are reported. For the calculation of the p-value, it is assumed that the statistic follows a χ^2_k distribution, with k the hypothesized number of structural mean shifts. The LM and Wald test both point in the same direction. The hypothesis of no breaks seems to be rejected convincingly for five out of seven countries. Finally, notice that for the U.S. and the U.K., the Beran test statistic is getting worse. However, adding AR or MA components to our maintained model does not yield improvement.

The final panel of table 2.3 concerns two more breaks, in July 1976 and January 1979. For Canada or France, no dramatic changes occur (as compared to the two break case). For Germany it is interesting to see that a temporally higher inflation seems to be found between 1979 and 1982. The Wald and LM tests however do not reject the null of no breaks against the alternative of four breaks. For Italy and the U.K. these extra breakpoints do not lead to a strong change. Japan, however, seems to have undergone higher inflation in the period from 1973 until 1976, which is the first oil crisis period. The breaksizes at those moments are about equal in absolute size and opposite in sign. Both tests point out that allowing for four breaks should be preferred to the assumption of no level shifts. For U.S. inflation the four breaks do seem to matter. The degree of fractional integration decreases a little further, and both LM and Wald test statistics obtain larger values, indicating strong evidence for level shifts. On the other hand, the problems indicated by the Beran test statistic for residual serial correlation increase.

Allowing for level shifts is seen to have a huge effect on the degree of fractional integration. In Canada, two breaks suffice to have the degree of integration diminish to a level that is no longer significant. In Japan, the first two breakpoints chosen do not seem to fit well the moments at which the mean level of inflation underwent a change. However, in the setting of four breakpoints, a high inflation period is neutralized, and the

resulting z_t series displays no significant fractional integration. In France and Germany, $\hat{d}$ decreases to a level that usually is considered as not a strong indication of the presence of long memory, although the parameter itself is still significantly different from zero. In Italy, the estimate of d in a model with two level shifts is notably lower than in the pure ARFIMA model. This also holds true for the U.K. and U.S.

2.6 Conclusions

In this paper we investigated the effect a level shift can have on the apparent long memory characteristics. Especially for inflation rates, where long memory seems to exist, level shifts because of exogenous shocks may also have occurred. A framework for combining level shifts and long memory was put forward. In section 2.4, a simulation study was performed to investigate the possible effects of a level shift on the estimate of the long memory parameter d in the ARFIMA model. A shock of only one standard deviation of the underlying series already could lead to the erroneous impression that long memory was present in the data. Encouraged by the results of the simulations, an investigation of the inflation rates in the countries of the G7 was performed. Where a pure ARFIMA model replicates previous results, that is, a significant value of $\hat{d}$ is found in several countries, addition of a set of level shifts did decrease the degree of fractional integration in various countries.

The results of the simulations indicated a size distortion for the LM test. In the controlled environment of a simulation, adjusting for this distortion is possible. Even if the size is controlled, the power of the test is not impressive. This is a problem that is hard to escape. Fractional integration and the occurrence of level shifts can be quite hard to distinguish in samples of medium size. Indeed, an unreported investigation using a larger sample did lead to better results for the size and power of the test, though improvement was slow. A possible way of improving the small sample results would be to work with the exact Maximum Likelihood method, as propagated by Sowell (1992).

In our empirical work, timing of the breakpoints was taken to be fixed exogenously. With the parametric framework used here, the search for optimal breakpoints is computationally demanding. In a semiparametric setting, Hsu (2000) derives results for the G7 countries estimating the degree of integration jointly with the timing of (at most) two breakdates. These results confirm the findings reported here. Furthermore, we assumed the shifts took place suddenly, in a specific month. Allowing for a smooth transition between regimes, as in Van Dijk (1999, chapter 2) might provide a better fit, though it is questionable if the data is informative enough to estimate the extra parameters involved.

2.A Calculating the likelihood and the test statistics

In the calculation of the likelihood and the test statistics, several elements are combined. In this appendix the way the elements of these functions were calculated, are given.

In equation (2.7), the power spectrum was given as

$$f(\lambda;\Psi) = \frac{\sigma_\epsilon^2}{2\pi}\left|1-e^{i\lambda}\right|^{-2d}\frac{\left|\Theta(e^{i\lambda})\right|^2}{\left|\Phi(e^{i\lambda})\right|^2}. \tag{2.6)+(2.7}$$

The factors in this formula are calculated using

$$e^{-i\lambda} = \cos\lambda - i\sin\lambda \quad \Leftrightarrow \quad e^{i\lambda} = \cos\lambda + i\sin\lambda, \tag{2A.1}$$

$$\begin{aligned}\left|1-e^{i\lambda}\right|^{-2d} &= \left|1-\cos\lambda - i\sin\lambda\right|^{-2d} \\ &= \left(1-2\cos\lambda+\cos^2\lambda+\sin^2\lambda\right)^{-d} = (2-2\cos\lambda)^{-d},\end{aligned} \tag{2A.2}$$

$$\left|\Theta(e^{i\lambda})\right|^2 = \left|1+\sum_{j=1}^{q}\theta_j e^{ij\lambda}\right|^2 = \left(1+\sum_{j=1}^{q}\theta_j\cos j\lambda\right)^2 + \left(\sum_{j=1}^{q}\theta_j\sin j\lambda\right)^2, \tag{2A.3}$$

$$\left|\Phi(e^{i\lambda})\right|^2 = \left|1-\sum_{j=1}^{p}\phi_j e^{ij\lambda}\right|^2 = \left(1-\sum_{j=1}^{p}\phi_j\cos j\lambda\right)^2 + \left(-\sum_{j=1}^{p}\phi_j\sin j\lambda\right)^2. \tag{2A.4}$$

Before calculating the periodogram, the data following the underlying ARFIMA process are calculated from $y_t, t = 1,..,T$, the moments of the k breaks $\tau_r, r = 1,..,k$, and the sizes of the corresponding level shifts $\mu_r, r = 1,..,k$. Assuming d lies in the stationary realm (see remarks in section 2.3), and reiterating a slightly adjusted version of equations (2.3)-(2.5), we have

$$z_t^a(\mu,d) = z_t^b(\mu,d) = y_t - \sum_{r=1}^{k}\mu_r I_{\{t>\tau_r T\}}, \tag{2.3′}$$

$$z_t(\mu,d) = z_t^b - \overline{z^b}. \tag{2.5′}$$

Of this transformed data set, the periodogram is calculated as

$$\begin{aligned}I_z(\lambda;\mu) &= \frac{1}{2\pi T}\left|\sum_{t=1}^{T} z_t e^{it\lambda}\right|^2 \\ &= \frac{1}{2\pi T}\left((\sum_{t=1}^{T} z_t\cos t\lambda)^2 + (\sum_{t=1}^{T} z_t\sin t\lambda)^2\right).\end{aligned} \tag{2A.5}$$

Defining $T^* = \lfloor T/2\rfloor, \lambda_j = \frac{2\pi j}{T}$ and weights δ_j as in section 2.3,

$$\delta_j = \begin{cases}\frac{1}{2} & \text{if } j = \frac{T}{2} = \left\lfloor\frac{T}{2}\right\rfloor \\ 1 & \text{else}\end{cases} \qquad j = 1,..,T^* \tag{2.11′}$$

all elements in the likelihood function are known:

$$\begin{aligned}\ln \mathcal{L}(y|\Phi, d, \Theta, \mu, \sigma_\epsilon^2) &= \ln \mathcal{L}(y|\Psi, \mu) \\ &= -T \log 2\pi - \sum_{j=1}^{T^*} \delta_j \log f(\lambda_j; \Psi) - \sum_{j=1}^{T^*} \delta_j \frac{I_z(\lambda_j; \mu)}{f(\lambda_j; \Psi)}. \end{aligned} \tag{2.10′}$$

The analytical gradient of the loglikelihood with respect to the parameters of the model was used, during the optimization routine for the model and in the calculation of the test statistics. Separating the vector of ARFIMA parameters Ψ and the level shifts μ, we find

$$\frac{\partial \ln \mathcal{L}(y|\Psi, \mu)}{\partial \psi_r} = -\sum_{j=1}^{T^*} \delta_j \frac{1}{f(\lambda_j; \Psi)} \frac{\partial f(\lambda_j; \Psi)}{\partial \psi_r} + \sum_{j=1}^{T^*} \delta_j \frac{I_z(\lambda_j; \mu)}{f(\lambda_j; \Psi)^2} \frac{\partial f(\lambda_j; \Psi)}{\partial \psi_r}, \tag{2A.6}$$

$$\frac{\partial \ln \mathcal{L}(y|\Psi, \mu)}{\partial \mu_r} = -\sum_{j=1}^{T^*} \delta_j \frac{1}{f(\lambda_j; \Psi)} \frac{\partial I_z(\lambda_j; \mu)}{\partial \mu_r}. \tag{2A.7}$$

The derivatives of the power spectrum with respect to the ARFIMA parameters in Ψ are

$$\frac{\partial f(\lambda; \Psi)}{\partial \phi_r} = f(\lambda; \Psi) \times \frac{-1}{\left|\Phi(e^{i\lambda})\right|^2} \frac{\partial \left|\Phi(e^{i\lambda})\right|^2}{\partial \phi_r}, \tag{2A.8}$$

$$\frac{\partial f(\lambda; \Psi)}{\partial \theta_r} = f(\lambda; \Psi) \times \frac{1}{\left|\Theta(e^{i\lambda})\right|^2} \frac{\partial \left|\Theta(e^{i\lambda})\right|^2}{\partial \theta_r}, \tag{2A.9}$$

$$\frac{\partial f(\lambda; \Psi)}{\partial d} = f(\lambda; \Psi) \times (-\log(2 - 2\cos\lambda)), \tag{2A.10}$$

$$\frac{\partial f(\lambda; \Psi)}{\partial \sigma_\epsilon} = f(\lambda; \Psi) \times \frac{2}{\sigma_\epsilon}. \tag{2A.11}$$

Two derivatives are left unspecified in the equations (2A.8) and (2A.9). These follow, after some tedious but simple algebra, as

$$\frac{\partial \left|\Phi(e^{i\lambda})\right|^2}{\partial \phi_r} = -2 \left((1 - \sum_{k=1}^{p} \phi_k \cos k\lambda) \cos r\lambda - \sum_{k=1}^{p} \phi_k \sin k\lambda \sin r\lambda \right), \tag{2A.12}$$

$$\frac{\partial \left|\Theta(e^{i\lambda})\right|^2}{\partial \theta_r} = 2 \left((1 + \sum_{k=1}^{q} \theta_k \cos k\lambda) \cos r\lambda + \sum_{k=1}^{q} \theta_k \sin k\lambda \sin r\lambda \right). \tag{2A.13}$$

The last element that was left unknown in these equations was the derivative of the periodogram of z w.r.t. the value of the breaks. This last derivative is found to be

$$\begin{aligned}\frac{\partial I_z(\lambda, \mu)}{\partial \mu_r} = \frac{1}{\pi T} \Bigg(&\left(\sum_{t=1}^{T} z_t \cos t\lambda \right) \left(\sum_{t=1}^{T} \frac{\partial z_t}{\partial \mu_r} \cos t\lambda \right) \\ &+ \left(\sum_{t=1}^{T} z_t \sin t\lambda \right) \left(\sum_{t=1}^{T} \frac{\partial z_t}{\partial \mu_r} \sin t\lambda \right) \Bigg) \end{aligned} \tag{2A.14}$$

with

$$\frac{\partial z_t}{\partial \mu_r} = I_{\{\tau_r T > t\}} + \frac{T - \tau_r T}{T} \approx I_{\{\tau_r T > t\}} + 1 - \tau_r.$$

Constructing the gradient is done by combining the equations. Although it is possible to derive the analytical second derivative as well, this would become even harder. For calculation of the Hessian, the numerical first derivative of the (analytical) gradient is used.

2.B Generating an ARFIMA process

Beran (1994) describes a method, originating from Davies and Harte (1987), to generate T observations from a stationary Gaussian model, given the autocovariances $\gamma(0), .., \gamma(T-1)$. The following steps are taken:

i. Calculate the autocovariances of the model, $\gamma(0), \gamma(1), .., \gamma(T-2), \gamma(T-1)$, and define

$$\gamma^*(j) = \begin{cases} \gamma(j) & \text{if } j \leq T-1 \\ \gamma(T-j) & \text{if } T-1 < j \leq 2T-3 \end{cases} \qquad j = 0, .., 2T-3. \tag{2B.15}$$

ii. Calculate the Fourier transform of the $\gamma^*(\cdot)$,

$$g_k = \sum_{j=0}^{2T-3} \gamma^*(j) e^{ij\lambda_k},$$

for $k = 0, .., 2T-3$ and $\lambda_k = \frac{2\pi k}{2T-2}$. These g_k should all result to be positive.

iii. Generate random normals $U_1, .., U_{T-2}$ and $V_1, .., V_{T-2}$ all independent and with variance 1; simulate U_0 and U_{T-1} as independent of all other values and eachother, with variance 2. Define $V_0 = V_{T-1} = 0$. Construct random variables Z_k in the complex plane as

$$Z_k = \begin{cases} U_k + iV_k & k = 0, .., T-1 \\ U_{2T-k-2} - iV_{2T-k-2} & k = T, .., 2T-3 \end{cases}$$

iv. The observations y_t are now calculated as

$$y_t = \frac{1}{2\sqrt{T-1}} \sum_{k=0}^{2T-3} \sqrt{g_k} e^{i(t-1)\lambda_k} Z_k.$$

Our procedure for generating an ARFIMA(p, d, q) starts with generating an ARFIMA(0,d,0) as described above, using the autocovariance function of the process as follows (Gradshteyn and Ryzhik 1965, p. 372):

$$\gamma(k) = \sigma_\epsilon^2 \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}$$

for the covariances needed in (2B.15). Then, the y_t that are generated should be used instead of the usual disturbances in a routine generating an ARMA(p, q). Presample disturbances are taken to be zero. The result will be a series y'_t which is approximately distributed as a ARFIMA(p, d, q).

2.C Data sources, programs, justifications

All estimations and simulations reported in this article have been calculated using programs written by the authors, using the Gauss programming language. Except from the Gauss base program, version 3.2.15, the CML library, version 1.0.18, was used. As indicated in appendix 2.B, routines for generating the fractionally integrated observations are programmed using the procedure as described in Beran (1994). The likelihood of the ARFIMA model, incorporating breaks in the mean, was programmed using routines by Rolf Tschernig and Marius Ooms as guideline. For programming the analytical derivatives of the model, elements from the programs by Breidt, Crato and de Lima (1998) were used. Graphics are made using the GnuDraw package in Ox.

The dataset on the United States was obtained from the Bureau of Labor Statistics. The series concerning the 'Consumer Price Index-All Urban Consumers' with ID `CUUR0000SA0` was retrieved in January 1997, for the period of January 1957 until December 1995. The base period was 1982-1984. The series was not seasonally adjusted.

For Canada, France, Germany, Italy, Japan and the United Kingdom the dataset on Consumer Price Indices previously investigated by Baillie et al. (1996) was used. This dataset can be downloaded from the data-archive of the Journal of Applied Econometrics. It was originally constructed from data from the Citibase archive in January 1991. For the six countries the 'All Consumer Price Indices' were used, ranging from January 1948 until March 1990. Also these CPI's were not adjusted for seasonality.

Both programs and datasets are available on request from the author of this thesis.

Chapter 3

Inflation, Forecast Intervals and Long Memory Regression Models

3.1 Introduction

This chapter[1] concerns the usefulness of the ARFIMA model for U.S. inflation for out-of-sample forecasting. We consider both point and interval forecasts and we also examine the usefulness of explanatory variables for different forecast horizons. Therefore we do not only consider time variation in the coefficients for the mean of inflation, but also time variation in the forecast error variance.

The most useful explanatory variables for U.S. consumer price inflation are connected with the Phillips curve, with oil price shocks, and with changes in monetary policy. Gali and Gertler (2000) give references for relevant recent explanations. Unemployment, output gap variables and real unit labour costs correspond with the Phillips curve. Hooker (1999) summarises evidence on the effect of oil price shocks on postwar U.S. inflation. Ball and Mankiw (1995) stress the effect of the sectoral distribution of price changes.

The relevant literature reveals that some variables are important for the explanation of short term inflation dynamics, whereas others may help to explain the longer run dynamics. Moreover, the effects of some variables, like the effect of the oil price on overall inflation, seems to have changed significantly over time. After 1980, oil price shocks did not have the same impact as in the 1970s, and monetary policy seems to have decreased both the mean and the variance of inflation, ceteris paribus. The empirical part of the economic literature also shows the sluggishness of inflation adjustment in adjusting to fundamentals. Indeed, many lags of inflation are statistically significant in reduced form equations of economic models, which is consistent with long memory behaviour found in time series analysis of inflation series, as was shown in the previous chapter.

We examine the predictive ability of the dynamic regression models for several horizons, extending results of Stock and Watson (1999), who analysed only 12 month ahead forecasting. We confine our analysis to only a few relevant explanatory variables. Therefore we rule out large-scale leading indicator variables and sectoral asymmetry variables.

[1] The chapter is based on the article Bos, Franses and Ooms (2001), forthcoming in the International Journal of Forecasting.

Moreover, we take a statistical time series approach, where we derive multi-step forecasts from the likelihood for the model for one-step ahead predictions. In agreement with Stock and Watson (1999) we use a simulated out-of-sample forecasting framework, but we use fixed specifications for the recursive forecast evaluation period.

The remaining part of our paper is organised as follows. Section 3.2 starts with a recursive ARFIMAX analysis of monthly U.S. inflation and three leading indicators. We compare specifications for the error process up to an ARFIMAX(1,d,1) model and allow for deterministic regime changes. We use likelihood based time domain estimators based on the algorithm of Sowell (1992), see Doornik and Ooms (1999). This allows us to extend the model with macroeconomic leading indicators from the database developed by Stock and Watson (1999). Our starting point is a model with two level shifts in the period 1960-1999. We investigate the stability of the explanatory effects and we examine how they complement the simple level shift specification. We compare the forecasting ability of the models. We find that forecast intervals are too wide. Section 3.3 therefore analyses the results of weighted forecasting based on structural shifts in the variance. Section 3.4 employs statistical tests on the forecasts. Section 3.5 concludes.

3.2 Recursive ARFIMAX forecasting

We consider a monthly U.S. consumer price index, as provided by the Bureau of Labor Statistics' (BLS) website, July 2000. It concerns the influential core consumer price index, that is the U.S. city average items less food and energy, 1982-84=100, BLS code `CUUR0000SA0L1E`. We use data from 1960:04 to 1999:12 in our statistical analysis. Core inflation has not been affected by many outliers and it is therefore easier to interpret and analyse than other CPI indices.

3.2.1 Basic features of U.S. core inflation

Figure 3.1 shows a time series plot of monthly core inflation, measured as 100 times the first differences of the logarithms of the index. We seasonally adjust the series with two sets of centred seasonal dummies, allowing for a break in the seasonal pattern in 1984. This break roughly corresponds with a change in the seasonal pattern detected by the official seasonal adjustment procedure used by the BLS. The autocorrelation functions of inflation and changes in inflation establish the long memory property of inflation, that is the inflation series appears non-stationary, while the differenced inflation series appears to follow an MA(1) process. The combined information of figure 3.1 suggests an order of integration larger than zero, but probably smaller than one. The autocorrelations also show that seasonal variation has been removed. In order to concentrate on nonseasonal variation only, we condition our analysis on these seasonally adjusted data.

It is clear from figure 3.1 that the current mean of inflation is significantly lower than in the 1970s. Different explanations coexist, but the combination of oil price increases and an accommodating interest rate policy by the U.S. authorities is viewed as the most important cause of the exceptional inflation levels in the 1970s. The Volcker-Greenspan regime kept high inflation level at bay after 1983. Since we want to compare realistic

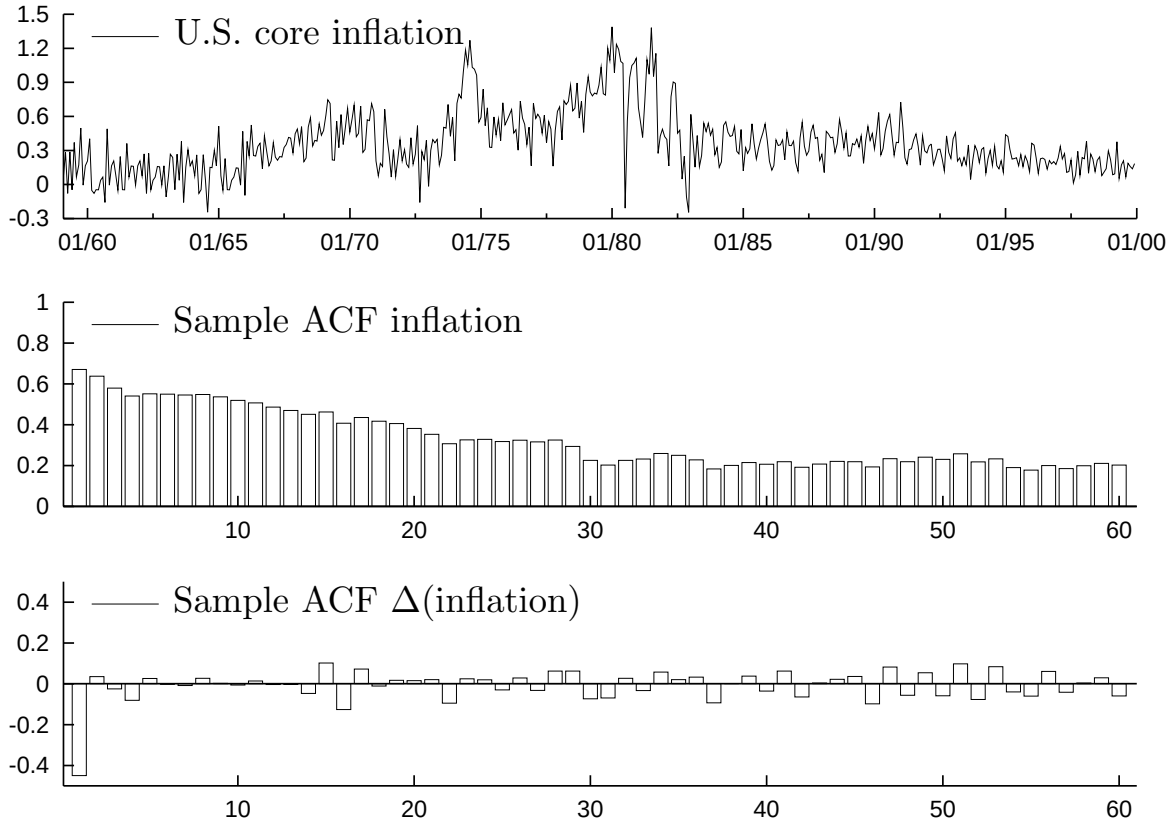

Figure 3.1: U.S. monthly core inflation, sample ACF of levels and differences

forecasting models, we have to allow for at least two structural breaks in the mean for all models, one upward break in 1973:07 and one downward break in 1982:07. Furthermore we allow for a separate mean for inflation in 1980:07, when U.S. prices were fixed in a high inflation period.

3.2.2 ARFIMAX modelling

The ARFIMA model has been introduced by Granger and Joyeux (1980) and Hosking (1981). Beran (1994) discusses the main asymptotic results for regression models with long memory errors. Hassler and Wolters (1995) and Ooms and Hassler (1997) showed that the ARFIMA(0,d,0) model with deterministic seasonality fits consumer price inflation series of many OECD countries rather well.

It is well known that ARMA models with an AR root close to the unit circle and an approximately cancelling MA root are also able to capture the low frequency characteristics typifying long memory. ARMA models can also be used to forecast long memory processes, see Brodsky and Hurvich (1999). ARIMA models are therefore interesting competitors in a forecasting study. The ARFIMA(1,d,1) model allows us to test ARIMA specifications by imposing restrictions. The last ACF of figure 3.1 may also indicate an ARIMA(0,1,1) specification.

The empirical macroeconomic literature suggests that macroeconomic leading indica-

tors can help to forecast yearly inflation rates out of sample, see Stock and Watson (1999). Hence, we extend the ARFIMA model with explanatory variables. As explanatory variables we use the U.S. unemployment rate u_t (seasonally adjusted, civilian labor force 16 years and older, BLS code `LFS21000000`), the short term interest rate r_t (we choose the rate on federal three month treasury bills) and the spread $s_t = R_t - r_t$ between long term and short term interest rates (for the long term interest rate R_t we use the return on a 10 year U.S. treasury bill with constant maturity).

The first explanatory variable, u_t, is used as a benchmark leading indicator for one-year-ahead forecasting of yearly inflation in Stock and Watson (1999). Its negative correlation with future inflation rates is associated in empirical macroeconometrics with the (old) Phillips curve. The second variable is the short term interest rate, r_t. In the short run, one observes a positive relationship between inflation and r_t. In the longer run, high interest rates are supposed to lead to lower inflation rates. The interrelationship is a complicated one, see Gali and Gertler (2000) for a recent analysis. The third leading indicator is the interest rate spread, s_t. A large spread is associated with higher inflationary expectations.

We specify the ARFIMAX model as

$$(1 - \phi_1 L)(1 - L)^d (y_t - x_t'\beta - v_{t-H}'\gamma_H) = (1 + \theta_1 L)\epsilon_t, \tag{3.1}$$

with L the lag operator ($Ly_t = y_{t-1}$), y_t the monthly inflation rate, x_t a vector of deterministic terms, v_t a k-vector of leading indicators, and H the forecast horizon. The zero mean disturbances ϵ_t are Gaussian white noise with standard deviation σ_ϵ. The parameter d specifies the order of integration. We require $0 \leq d \leq 1$, $|\phi_1| < 1$, $|\theta_1| < 1$ and $\phi_1 \neq -\theta_1$. Note that we need to lag the explanatory variables, v_t, at least H times in order to use them in an H-step-ahead forecasting model for inflation.

The column vector x_t consists of a constant, $x_{1,t} = 1$, measuring the autonomous mean of inflation, a level-shift for the high inflation period, $x_{2,t}$, and a single dummy for 1980:07, $x_{3,t}$: Observe the huge outlier in figure 3.1. The variable $x_{2,t}$ equals one in the period 1973:07-1982:06 and is zero otherwise. A separate level-shift for the Volcker-Greenspan period equalling one in the period after 1982:06 turned out to have approximately zero effect, even in a recursive analysis starting in 1984. Therefore our model incorporates the same inflation regime before 1973 and after 1982.

When $d = 1$, the original constant term, β_1, drops out of the model and we use the $\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$, $i = 2, 3$, $\Delta v_{j,t-H} = v_{j,t-H} - v_{j,t-H-1}$, $j = 1, \ldots, k$, as regressors for Δy_t. Note however, that all forecast results reported below refer to the same observations on y_t and partial sums of y_t, for all values of d.

The deterministic component of y_t is $x_t'\beta$. We conduct the time series analysis on y_t, corrected for the interventions $x_{2,t}\beta_2$ and $x_{3,t}\beta_3$, conditional on the information in the leading indicators.

Our three main specifications and their respective ranges for the order of integration for the stochastic component of inflation and log prices are summarised in table 3.1.

Tables 3.2a-3.2b present estimation results for four models for two samples, 1960:04-1984:01 and 1960:04-1999:11, for the forecast horizon $H = 1$. These were obtained by maximum likelihood as implemented by Doornik and Ooms (1999) in Ox (Doornik 1999), with graphical output in Gnuplot using the GnuDraw package of one of the authors.

Table 3.1: ARFIMA components and their interpretation

d	ϕ	θ	Inflation	Log prices
0	$<0,1>$	$<-1,0>$	Short memory	$I(1)$
1	$<0,1>$	$<-1,0>$	Non-stationary	$I(2)$
$<0,0.5>$	0	0	Long memory	$I(d+1)$

The maximum likelihood procedure is also available in the econometric software PcGive, see Doornik and Hendry (2001). Key statistics for the specifications in tables 3.2a and 3.2b are compared with statistics for other models in table 3.3. The results for the ARFIMAX(0,d,0) model in the first two columns of table 3.2a provide evidence of fractional integration in the stochastic component of inflation, even if we condition on u_{t-1} and r_{t-1}. That is, d is about 0.25, and it differs significantly from 0 and 0.5. This coefficient is markedly lower than the corresponding value in the plain unconditional ARFIMA(0,d,0) model, see table 3.3. Autonomous baseline inflation, β_1, in the first and the last part of the sample is estimated between 0.25 and 0.28 percent per month in this specification. The estimate of β_2 indicates that the average autonomous inflation level was more than twice as high in the 1970s as in the rest of the sample. The estimate for β_3 shows that there was an exceptional one-month drop in inflation in July 1980 of about 1 percent.

The third and fourth column of table 3.2a show that the ARMAX(1,1) model fits the inflation data about as well as the ARFIMAX(0,d,0) specification. The estimates for β_1 and β_2 correspond quite closely to the estimates for the ARFIMAX(0,d,0) model. The only marked difference for the initial sample is seen for the standard error of the constant term, β_1, which is smaller for the ARMAX model. Apparently, the uncertainty about the mean of inflation is lower if one assumes an ARMA model. This partly reflects the underlying assumption of short memory ARMA models on the speed of convergence of the ML estimate of the mean, $\widehat{\mu}_T$, to the population mean μ as the sample size, T increases. In ARMA models one implicitly assumes $\text{var}(\widehat{\mu}) = cT^{-1}, T \to \infty$. For ARFIMA models with $-0.5 < d < 0.5$ one allows for $\text{var}(\widehat{\mu}) = cT^{2d-1}, T \to \infty$, see Adenstedt (1974) and Beran (1994, Chapter 9). Note that this difference for the standard error of β_1 is smaller for the longer sample, as the AR parameter of the ARMAX model is closer to unity than in the initial sample. The estimate of the d parameter of the ARFIMAX model does not change by increasing the sample.

Table 3.2b shows results for the ARIMAX(1,1,1) model and the ARIMAX(0,1,1) specification. We include the results for the latter specification because it is a straightforward extension of the popular IMA(1,1) model for inflation rates, which leads to the well known exponentially weighted moving average forecasts of inflation. The models of table 3.2b do not incorporate a fixed mean for inflation, that is, β_1 is not identified. An additional constant measuring an autonomous trend of inflation turned out to be insignificant and is therefore omitted. The structural breaks in inflation are not as significant as for the ARFIMAX(0,d,0) and ARMAX(0,0) model. These specifications are able to pick up changes in the conditional mean without the introduction of shift-dummies in the model. The ARIMAX(1,1,1) and ARIMAX(0,1,1) provide similar fits in the initial sample, but in the longer sample the ARIMAX(1,1,1) model is clearly preferable.

The coefficients $\gamma_{1,u}$ and $\gamma_{1,r}$ in tables 3.2a-3.2b concern the effect of the leading indi-

cators. The coefficient $\gamma_{1,u}$ of lagged unemployment is negative for this forecast horizon, as expected, but this effect seems to have declined in the 1980s and 1990s. Lagged interest rates have a significant positive and stable effect as reflected in $\gamma_{1,r}$. These conclusions apply to all four models for the stochastic component of inflation.

Finally, diagnostic tests of tables 3.2a and 3.2b yield satisfactory outcomes for the period 1960-1983. For the period 1960-1999, there is evidence of heteroskedasticity, which is also reflected in the different estimates for σ_ϵ for the two samples. This may signal a permanent downward shift in the innovation variance in the 1980s and 1990s. This can also be the message of the ARCH test, see Lamoureux and Lastrapes (1990).

The relationships between y_t and the elements of v_{t-H} need to be relatively stable in order to be useful for forecasting. In theory one may even estimate a complete lag structure on all three variables simultaneously, but here we confine ourselves to one lag per variable at a time. We experimented with adding up to 3 extra lags and differences of the leading indicators, but this did not lead to significant improvements in either the in-sample fit or the out-of-sample forecasts.

Table 3.3 compares the models of tables 3.2a-3.2b with other AR(FI)MA(X) specifications. We present a summary of the main results for the stationary ARMA(1,1), ARFIMA(0,d,0), ARFIMA(1,d,0), ARFIMA(0,d,1), ARFIMA(1,d,1) models and the non-stationary ARIMA(1,1,1) and ARIMA(0,1,1) specifications, both with and without the explanatory variables. An ARFIMAX(1,d,1) specification does not provide a significantly better fit than either the ARFIMAX(0,d,0) model or the ARMAX(1,1) model. More importantly, the parameters of the ARFIMAX(1,d,1) are not well identified, since both the FI part and the ARMA part can capture the low frequency characteristics of the process. In table 3.3 we observe $\widehat{\phi}_1 = 0.98$ combined with $\widehat{d} = -0.63$, so that d cannot really be interpreted as the order of integration of the inflation process in this case.

The effect of the explanatory variables is similar across the different ARFIMAX models: unemployment has a negative effect, interest rates have a positive effect, the spread does not have an additional effect, and the introduction of the leading indicators lowers the estimate of d in the ARFIMA models and it decreases the estimate of ϕ in the ARMA models. The explanatory variables account partly for the persistence of inflation.

3.2.3 Recursive estimation and forecasting

Next we examine the point forecasts of the different models in a simulated out-of-sample experiment. We estimate parameters of a range of models recursively. Again, we compare univariate models with specifications with both single regressors and multiple regressors, that is, we vary the orders p, q and k in (3.1) and we either estimate d or put it equal to zero or unity.

We start with sample 1960:04-1984:01 and end with sample 1960:04-1999:11. We make point predictions of monthly inflation, $\widehat{y}_{t+H}$, for multiple horizons, $H = 1, 3, 6, 12, 24$. Moreover, we compute cumulative ("July 1998-June 1999") $\hat{z}_{t+H} = \sum_{i=1}^{H} \widehat{y}_{t+i}$ predictions for quarterly, half-yearly, yearly and two-yearly inflation. Cumulative inflation forecasts can be interpreted in other ways. They correspond with forecasts of the log price level H periods ahead, $\log(P_{t+H})$, minus todays log price level, $\log(P_t)$. For the univariate models, they also equal the forecasts of H times the future mean of inflation over the

Table 3.2a: ARFIMAX model estimates for U.S. core inflation

	ARFIMAX(0,d,0)				ARMAX(1,1)			
Parameter	60:04-84:01		60:04-99:11		60:04-84:01		60:04-99:11	
ϕ_1					0.757	(.10)	0.839	(.07)
d	0.245	(.05)	0.260	(.04)				
θ_1					−0.519	(.13)	−0.615	(.11)
β_1	0.256	(.11)	0.188	(.09)	0.251	(.08)	0.174	(.08)
β_2	0.296	(.06)	0.298	(.05)	0.322	(.05)	0.303	(.05)
β_3	−0.919	(.19)	−0.941	(.16)	−0.934	(.19)	−0.961	(.16)
$\gamma_{1,u}$	−0.038	(.01)	−0.023	(.01)	−0.040	(.01)	−0.022	(.01)
$\gamma_{1,r}$	0.045	(.01)	0.041	(.01)	0.047	(.01)	0.043	(.01)
σ_ϵ	0.191		0.166		0.191		0.166	
LL	66.680		180.136		68.085		180.362	
AIC	−119.360		−346.273		−120.170		−344.725	
Normality	6.616	0.04	18.165	0.00	6.752	0.03	17.998	0.00
ARCH	2.436	0.12	9.358	0.00	1.958	0.16	11.147	0.00
Box-Pierce	20.574	0.90	27.922	0.57	20.010	0.89	31.972	0.32

Note: Estimates of (3.1) for 4 specifications and two sample sizes, standard errors in parentheses. Diagnostic tests on residuals: Normality test jointly on third and fourth moment, ARCH test using 1 lag, Box-Pierce test using 36 lags and p-values of the asymptotic distributions under the assumption of correct specification, see Doornik and Ooms (1999).

Table 3.2b: ARFIMAX model estimates for U.S. core inflation (continued)

	ARIMAX(1,1,1)				ARIMAX(0,1,1)			
Parameter	60:04-84:01		60:04-99:11		60:04-84:01		60:04-99:11	
ϕ_1	0.153	(.09)	0.167	(.06)				
θ_1	−0.879	(.07)	−0.897	(.04)	−0.793	(.05)	−0.821	(.04)
β_2	0.223	(.08)	0.255	(.07)	.191	(.09)	0.213	(.07)
β_3	−0.930	(.19)	−0.938	(.16)	−.961	(.19)	−0.964	(.17)
$\gamma_{1,u}$	−0.044	(.02)	−0.029	(.02)	−.046	(.03)	−0.029	(.02)
$\gamma_{1,r}$	0.039	(.01)	0.036	(.01)	.039	(.01)	0.038	(.01)
σ_ϵ	0.195		0.167		0.196		0.168	
LL	61.777		175.549		60.301		171.863	
AIC	−109.554		−337.098		−108.602		−331.73	
Normality	7.472	0.02	25.264	0.00	7.338	0.03	24.896	0.00
ARCH	3.574	0.06	12.518	0.00	7.719	0.01	25.095	0.00
Box-Pierce	25.339	0.71	33.327	0.31	29.18	0.56	43.694	0.07

Note: See table 3.2a for an explanation of the entries in the table.

Table 3.3: ARFIMAX model estimates

ϕ	d	θ	$\gamma_{1,u}$	$\gamma_{1,r}$	$\gamma_{1,s}$	AIC
.	.26	.	−.023	.041	.	−346.273
.06	.30	.	−.024	.041	.	−344.920
.	.28	.30	−.024	.041	.	−344.826
.84	.	−.62	−.022	.043	.	−344.725
.98	−.64	−.11	−.020	.043	.	−346.367
.	.26	.	−.027	.043	.007	−344.442
−.06	.29	.	−.026	.042	.005	−343.014
.	.29	−.05	−.026	.042	.005	−342.925
.83	.	−.60	−.031	.047	.014	−343.404
.97	−.63	−.11	−.028	.047	.013	−344.939
.	.35	.	.	.	.	−320.429
−.11	.42	.	.	.	.	−321.180
.	.41	−.11	.	.	.	−320.796
.92	.	−.65	.	.	.	−319.111
−.29	.43	.18	.	.	.	−319.459
.17	1	−.90	−.029	.036	.	−337.098
.	1	−.82	−.029	.038	.	−331.727
.17	1	−.90	−.026	.030	−.012	−335.599
.	1	−.82	−.028	.033	−.009	−330.000
.17	1	−.85	.	.	.	−317.565
.	1	−.76	.	.	.	−312.148

Note: Sample period: 1960:04-1999:11, a . indicates that the parameter is restricted to zero.

forecast horizon.

For the univariate models ($k = 0$ in (3.1)) we expect the longer run forecasts of the ARMA(1,1) specification to change the least as new inflation shocks enter the information set. The ARFIMA(0,d,0) forecasts allow for longer lasting deviations from the long run mean and are therefore expected to be more variable. The ARIMA(1,1,1) forecasts should show the largest variation. The differences between the three models in the variability of the long run forecasts should be reflected in their estimated forecast intervals. We summarise the asymptotic characteristics of the forecast intervals of the univariate models in table 3.4, where it is assumed that $H \to \infty, H/T \to 0$, where T is the sample size, so that the mean can be treated as known, see Beran (1994, §8.6). Table 3.4 makes clear that d is an influential parameter for long run interval forecasting of the log price level, at least theoretically. As the explanatory power of the leading indicators decreases for longer forecast horizons we can expect a similar behaviour for the long run forecast intervals of the models with regressors. It remains an empirical question whether the asymptotic results of table 3.4 provide a good indication for the sample sizes and forecast horizons in our study.

Table 3.4: Asymptotic rates of growth of variances

$d = 0$	$\mathsf{var}(\log(P_{T+H}/P_T)\|T)$	is $c \cdot H$,
$d = 1$	$\mathsf{var}(\log(P_{T+H}/P_T)\|T)$	is $c \cdot H^3$
$0 < d < 1$	$\mathsf{var}(\log(P_{T+H}/P_T)\|T)$	is $c \cdot H^{2d+1}$

Note: Indicated are the rates of growth of forecast intervals for the log prices given an order of integration d for inflation, for horizon $H \to \infty, H/T \to 0$, see Beran (1994, §8.6).

3.2.4 Recursive estimates

Before analysing the recursive out-of-sample forecasting performance of the models, we examine their recursive parameter estimates. We first address the stability of the dynamic parameters d, ϕ, θ and the mean parameter β_1 over our recursive estimation period. Figure 3.2 shows the recursive estimates of the ARFIMAX(0,d,0) model with u_{t-1} and r_{t-1} as leading indicators. Figure 3.3 shows the recursive estimates of the corresponding ARMA(1,1) model. The starting points and endpoints of these figures correspond to the results in table 3.2a. Both figures indicate that the effect of interest rates and unemployment has become less significant over the last decades, although the effect of interest rates has stabilised after 1993. The estimated d for the ARFIMAX(0,d,0) model is more stable than the corresponding ϕ and θ of the ARFIMAX(1,0,1) model. For comparison we present the recursive parameter estimate of the univariate ARFIMA(0,d,0) model in figure 3.4. The parameters of this simple model hardly change over time. This may be an advantage in forecasting. Note again that d is much higher in this univariate model.

The difference between the univariate specification and the model with leading indicators is smaller for multi-step forecasting at longer horizons. This is illustrated in figure 3.5 which shows the recursive parameter estimates of an ARFIMAX(0,d,0) model with

r_{t-24} and u_{t-24} as the leading indicators. These estimates for d are closer to the values obtained for the univariate model, as the explanatory power of r_{t-24} and u_{t-24} is considerably smaller than in the one-step-ahead forecasting model. Note also the negative sign of $\gamma_{24,r}$ as high short term interest rates are related with lower monthly inflation figures after two years. This also leads to a higher estimate for β_1 compared to results in figure 3.5 as $\gamma_{24,r} r_{t-24}$ has a negative mean over the sample whereas $\gamma_{1,r} r_{t-1}$ was clearly positive.

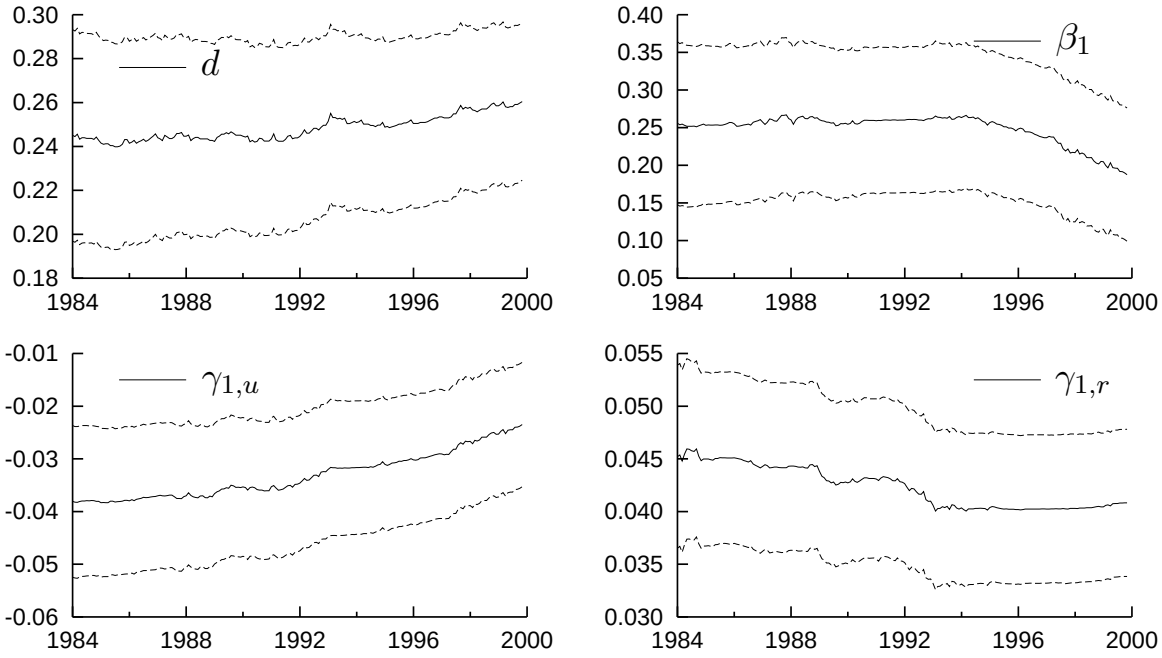

Figure 3.2: Recursive ARFIMAX(0,d,0) estimates ± one s.e., d, β_1, $\gamma_{1,u}$, $\gamma_{1,r}$

3.2.5 Recursive forecasting

Computing forecasts for an ARFIMA(p,d,q) model is not trivial. An ARFIMA(p,d,q) model cannot be written as a finite order ARMA model or as a finite dimensional state space model, so that standard methods do not apply. Here we use the optimal linear forecast of $y_{t+H} - x'_{t+H}\beta - v'_t\gamma_H$ given $y_1 - x'_1\beta - v'_{1-H}\gamma_H, \ldots, y_t - x'_t - v'_{t-H}\gamma_H$, see Beran (1994, §8.7) and Doornik and Ooms (1999) for the exact implementation. This forecast function explicitly assumes that forecasts are generated out of a finite sample. It takes the estimated covariance function of the stochastic part of the model as an input and results in a time dependent forecast function depending on all available observations. Forecast standard error estimates are computed accordingly. We use RMSEmod, that is the RMSE as derived from the model, to denote this forecast standard error estimate below. For the ARIMA model, $d = 1$, we cumulate the ARMA forecasts for the changes in inflation back to inflation forecasts. Forecast standard errors are adjusted for this cumulation as well.

We recursively compute the forecasts for the different horizons and compare them with actual values. Figure 3.6 illustrates two-year-ahead forecasting. It displays time series plots of the forecasts $\widehat{y}_{t+24|t}$, and the predetermined part of the forecast, $x'_{t+H}\widehat{\beta}_t$,

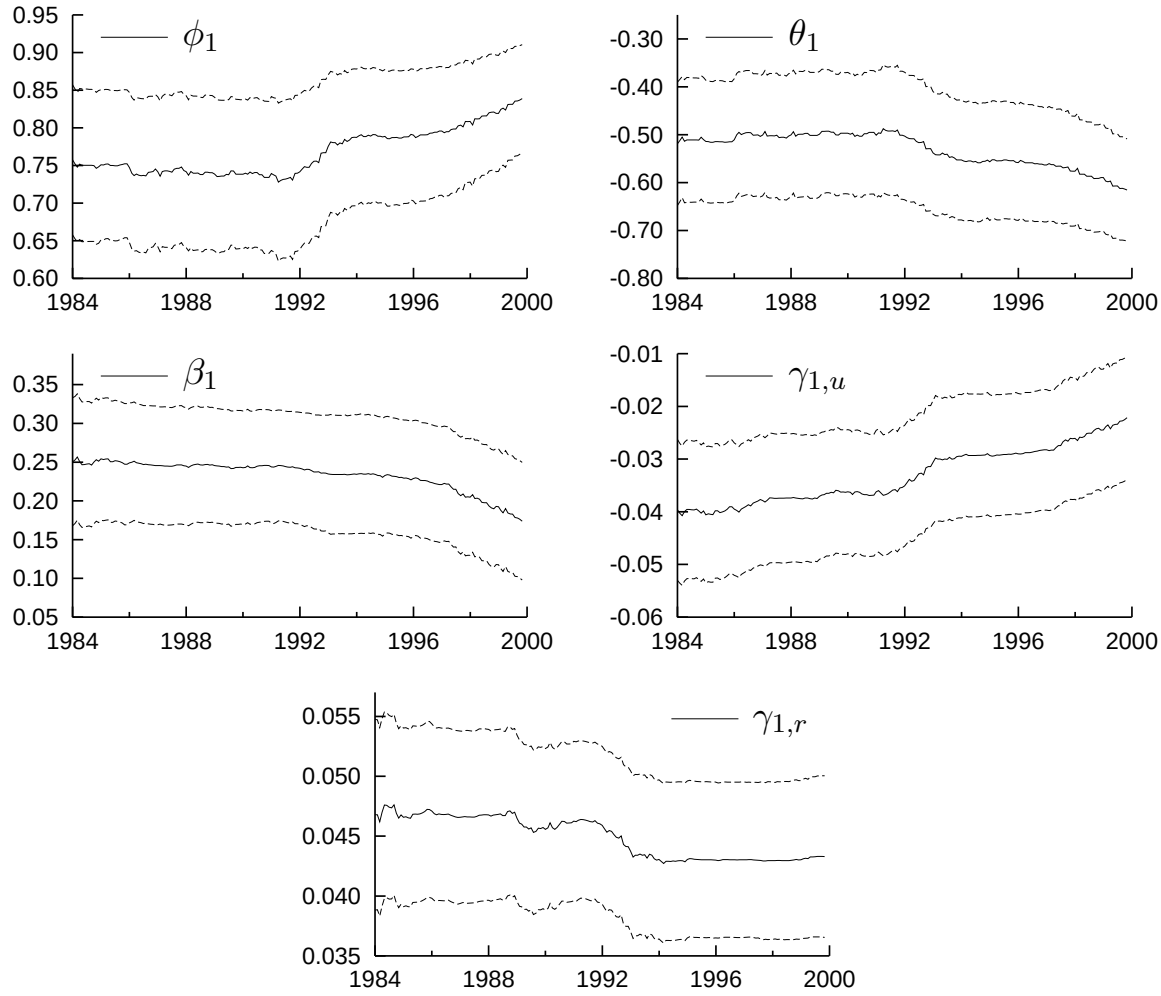

Figure 3.3: Recursive ARMAX(1,1) estimates ± one s.e., ϕ_1, θ_1, β_1, $\gamma_{1,u}$, $\gamma_{1,r}$

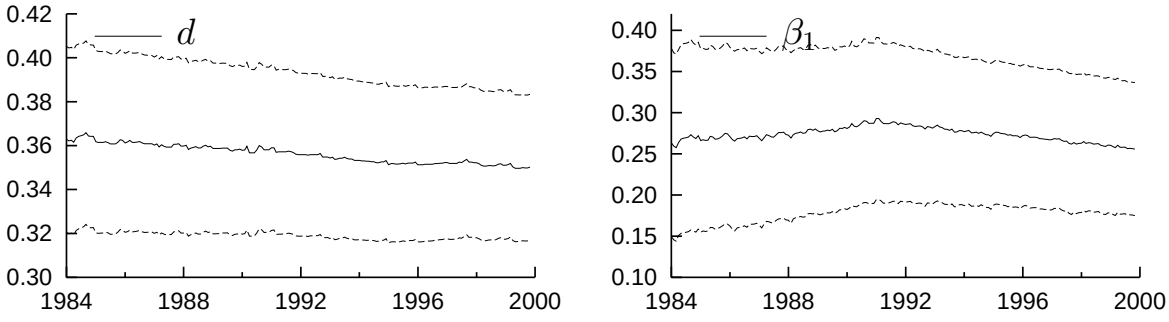

Figure 3.4: Recursive ARFIMA(0,d,0) estimates ± one s.e., d, β_1

together with the actual values y_{t+24}. We present only the forecasts of the univariate ARFIMA(0,d,0) model in graphical form. We give numerical evidence on all relevant models and all forecast horizons below. For the ARFIMA(0,d,0) model of figure 3.6 we observe a persistent deviation between $\widehat{y}_{t+24|t}$ and $x'_{t+24}\widehat{\beta}_t$, for t between 1984 and

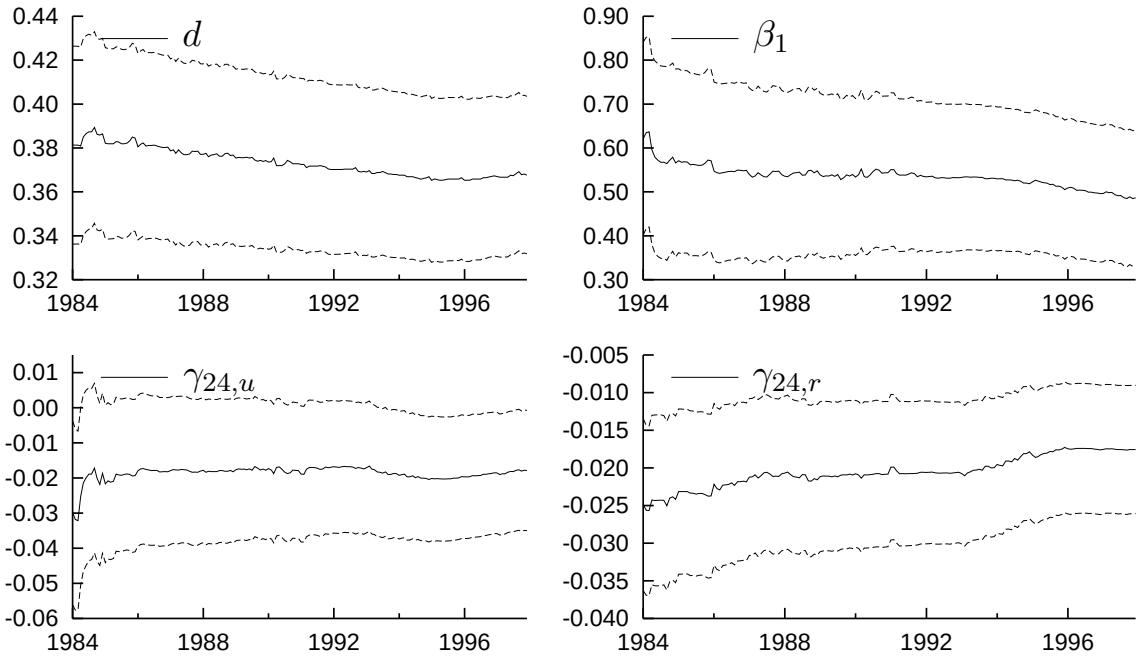

Figure 3.5: Recursive ARFIMAX(0,d,0) estimates ± one s.e., d, β_1, $\gamma_{24,u}$, $\gamma_{24,r}$

1994. This period corresponds to a period of persistently higher inflation, captured by the stochastic part of the ARFIMAX(0,d,0) model. The forecasts of the ARIMAX(1,1,1) model, not shown here, are not as smooth, but they follow the trend in inflation more closely. The bottom panel of figure 3.6 shows the cumulative forecast intervals for two-year inflation.

The ARFIMA(0,d,0) model tracks the downward swing in average inflation in the beginning of the 1990s more closely than the ARMA(1,1) model. The forecast intervals of the two models for this horizon are equally wide. This suggests that the differences between these two models are not as large in practice as asymptotic theory suggests. The cumulative forecasts of the ARIMA(1,1,1) model are again more volatile, but they naturally follow the persistent decline in inflation in the 1990s quite well. The forecast intervals are clearly much wider than for the ARFIMA(0,d,0) model. The leading indicator models show more volatility than the univariate models in their two-year-ahead forecasts, especially in the beginning of the evaluation period in 1984, when unemployment and interest rates were volatile and values for 1985 and 1986 were forecast.

Table 3.5 summarises the cumulative forecasting results for four univariate models, the ARMA(1,1), the ARFIMA(0,d,0), the ARFIMA(1,d,0) and the ARIMA(1,1,1) model. Results for the leading indicator models are presented below.

We present three measures of forecasting performance. First, we report the mean forecast error, MFE= $(t_2 - t_1 + 1)^{-1} \sum_{t=t_1}^{t_2} e_{t+H}$, with $e_{t+H} = \widehat{z}_{t+H} - z_{t+H}$, where $\widehat{z}_{t+H}$ is the cumulative inflation estimate for the period t to $t+H$, using the data and parameter estimates up to time t, $t+H$=(1984:01)+ $H \ldots$ 1999:12. Second, table 3.5 reports the root mean squared error, RMSE= $\left((t_2 - t_1 + 1)^{-1} \sum e_{t+H}^2\right)^{\frac{1}{2}}$, Third, we compute the mean absolute prediction error, MAPE= $(t_2 - t_1 + 1)^{-1} \sum |e_{t+H}|$. We also report the root of

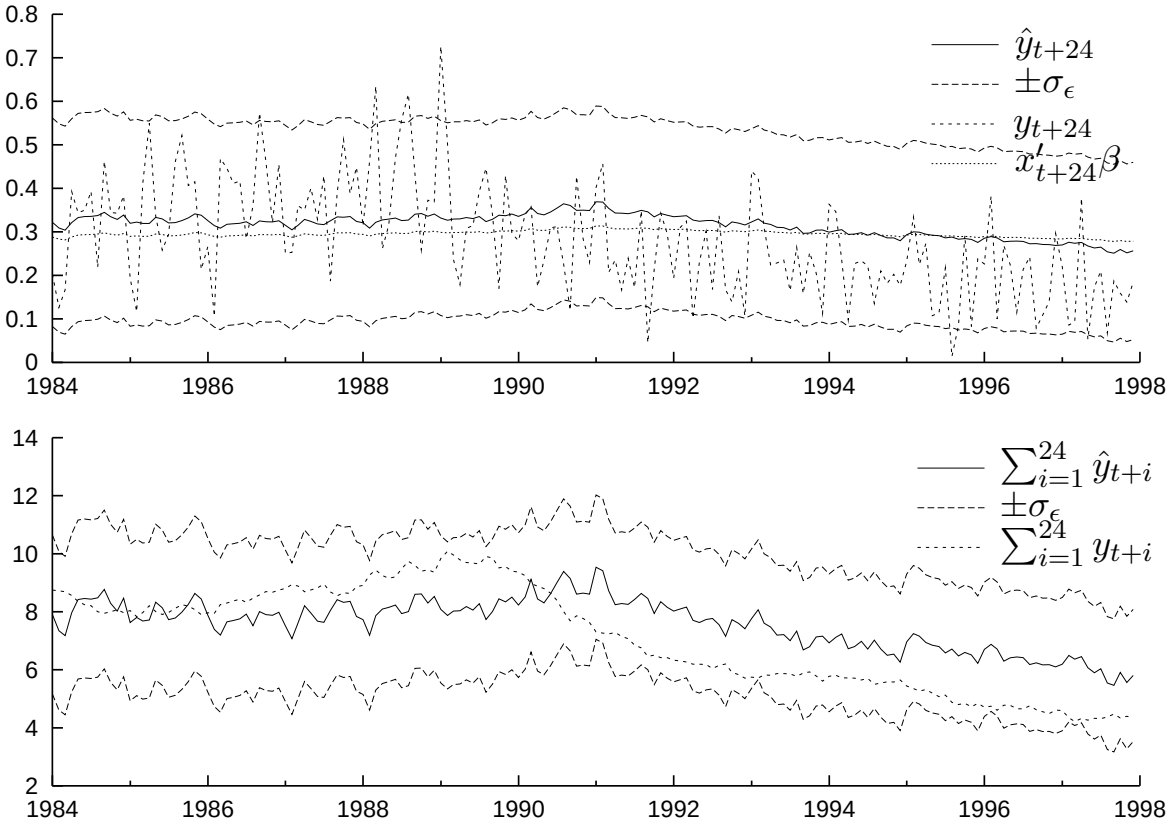

Figure 3.6: Recursive two-year-ahead forecasts ARFIMA(0,d,0) model ± one s.e., exogenous part of the forecast and actual inflation corresponding period. Bottom panel: Cumulative forecasts over a two-year period.

the H-step-ahead forecast error variance as predicted by the model, RMSEmod. The RMSEmod-values in the table are the mean values of the recursively computed model estimates over the evaluation sample. Finally we present the ratio of RMSEmod and RMSE to give a first indication of the coverage probability of the forecast intervals. We note that forecast intervals for the ARFIMAX(0,d,0) model and the ARMAX(1,1) specification are approximately equally wide, whereas the actual forecast RMSE for the ARFIMAX(0,d,0) model is considerably smaller for larger horizons.

The results for the best model for each forecast criterion are printed in boldface. It appears that the ARIMA(1,1,1) model provides the best forecasts overall at all horizons. The ARMA(1,1) model performs the worst. The relative differences in forecasting performance between the models increase with the forecast horizon. The one-step-ahead forecasts of the models are very similar. The estimated model forecast error variances (RMSEmods) agree very closely for horizons 1 to 6. Under correct specification these RMSEmods are expected to slightly underestimate the true forecast error variance in finite samples, because they neglect the effect of parameter uncertainty, but here we observe for all models that the RMSEmods are significantly higher than the actual RMSEs. In particular for the ARIMA(1,1,1) model, the estimated forecast error variances for the 12 and 24 period ahead cumulative forecasts are much too large. Refer also back to figure

Table 3.5: Simulated cumulative out-of-sample forecasts 1984-1999 for univariate models for U.S. core inflation

	Model			H				
	p	d	q	1	3	6	12	24
MFE	1	0	1	**0.001**	**0.001**	**0.015**	**0.059**	**0.280**
MFE	1	1	1	*0.006*	0.019	0.051	0.153	0.453
MFE	0	1	1	0.004	0.014	0.039	0.125	0.386
MFE	0	d	0	0.005	0.020	0.058	0.164	0.471
MFE	1	d	0	*0.006*	0.023	*0.064*	*0.179*	*0.501*
MFE	0	d	1	*0.006*	0.022	0.063	0.175	0.493
RMSE	1	0	1	*0.118*	*0.251*	*0.386*	*0.697*	*1.550*
RMSE	1	1	1	**0.115**	**0.233**	**0.339**	**0.552**	**1.228**
RMSE	0	1	1	0.117	0.244	0.367	0.615	1.355
RMSE	0	d	0	0.116	0.240	0.358	0.616	1.344
RMSE	1	d	0	0.117	0.241	0.355	0.600	1.317
RMSE	0	d	1	0.116	0.240	0.354	0.601	1.320
MAPE	1	0	1	*0.094*	*0.202*	*0.315*	*0.589*	*1.398*
MAPE	1	1	1	**0.091**	**0.182**	**0.266**	**0.439**	**0.969**
MAPE	0	1	1	0.093	0.190	0.284	0.479	1.050
MAPE	0	d	0	0.092	0.194	0.296	0.528	1.211
MAPE	1	d	0	0.092	0.194	0.289	0.510	1.182
MAPE	0	d	1	0.092	0.193	0.289	0.511	1.185
RMSEmod	1	0	1	0.183	0.415	0.761	1.455	2.530
RMSEmod	1	1	1	0.185	0.424	0.761	1.519	3.439
RMSEmod	0	1	1	0.186	0.413	0.782	1.719	4.296
RMSEmod	0	d	0	0.184	0.427	0.751	1.354	2.484
RMSEmod	1	d	0	0.183	0.417	0.748	1.389	2.611
RMSEmod	0	d	1	0.183	0.418	0.747	1.381	2.582
RMSEmod/RMSE	1	0	1	1.554	1.657	1.972	2.088	1.633
RMSEmod/RMSE	1	1	1	1.612	1.820	2.247	2.750	2.800
RMSEmod/RMSE	0	1	1	1.587	1.692	2.130	2.795	3.171
RMSEmod/RMSE	0	d	0	1.587	1.779	2.096	2.197	1.848
RMSEmod/RMSE	1	d	0	1.569	1.729	2.108	2.315	1.983
RMSEmod/RMSE	0	d	1	1.577	1.743	2.111	2.299	1.956

Note: Best ranking results in **boldface**, worst ranking results in *italics*.

3.6, which shows that all 2-year inflation outcomes lie in a one-σ forecast interval.

Table 3.6 contains a selection of the corresponding simulated out-of-sample forecasting results for a range of leading indicator models. We examined the forecasting performance of 6 specifications for the stochastic part and 5 specifications for the leading indicator part. We allowed for ARMA(1,1), ARFIMA(0,d,0), ARFIMA(1,d,0), ARFIMA(0,d,1), ARIMA(1,1,1), ARIMA(0,1,1) errors. We use three single leading indicator models with

u, r, s, one double leading indicator model with u and r (in tables 3.2a-3.2b) and finally a model with both u, r and s. We present the results for the best and for the worst of these 30 models for three forecasting criteria, MFE, RMSE and MAPE, and for 5 horizons: 1, 3, 6, 12, and 24 months. However, for the single leading indicator r_{t-H} we show the outcomes for all stochastic specifications, as the interest rates proves to be the most interesting explanatory variable.

Table 3.6 shows that r_{t-1} is the best leading indicator for one-step-ahead forecast-

Table 3.6: Simulated cumulative out-of-sample forecast performance 1984-1999 for leading indicator models for U.S. core inflation

	Model				H				
	X	p	d	q	1	3	6	12	24
MFE	r	1	0	1	0.003	0.004	**0.009**	0.048	0.486
MFE	r	1	1	1	**0.001**	0.010	0.048	0.157	1.042
MFE	r	0	1	1	**0.001**	0.009	0.042	0.140	0.938
MFE	r	1	d	0	**0.001**	0.007	0.044	0.143	0.955
MFE	r	0	d	0	**0.001**	0.006	0.035	0.120	0.815
MFE	r	0	d	1	**0.001**	0.007	0.041	0.137	0.938
MFE	u	1	0	1	0.002	**0.004**	0.013	**0.044**	**0.203**
MFE	u	1	1	1	*0.015*	0.066	0.175	0.368	0.567
MFE	urs	1	1	1	0.010	*0.070*	*0.215*	*0.512*	*1.636*
MFE	urs	1	d	0	0.006	0.025	0.088	0.246	1.187
RMSE	r	1	0	1	0.115	0.245	0.378	0.658	1.732
RMSE	r	1	1	1	**0.113**	**0.229**	**0.334**	0.559	2.066
RMSE	r	1	d	0	0.114	0.235	0.345	0.571	1.679
RMSE	r	0	d	0	0.114	0.234	0.344	0.575	1.632
RMSE	r	0	d	1	0.114	0.235	0.344	0.571	1.669
RMSE	s	1	1	1	0.116	0.247	0.359	**0.514**	1.701
RMSE	u	1	1	1	0.116	0.233	0.360	0.616	**1.245**
RMSE	ur	1	0	1	*0.119*	*0.268*	*0.432*	0.748	1.724
RMSE	urs	1	0	1	*0.119*	*0.268*	0.430	*0.753*	1.896
RMSE	urs	1	1	1	0.114	0.247	0.445	0.724	*2.407*
RMSE	urs	0	1	1	0.116	0.248	*0.451*	0.687	2.363
MAPE	r	1	0	1	0.092	0.199	0.307	0.546	1.538
MAPE	r	1	1	1	**0.090**	**0.183**	**0.263**	0.444	1.583
MAPE	r	1	d	0	0.091	0.191	0.277	0.483	1.404
MAPE	r	0	d	0	0.091	0.191	0.278	0.488	1.383
MAPE	r	0	d	1	0.091	0.191	0.277	0.483	1.396
MAPE	s	1	1	1	0.092	0.195	0.294	**0.415**	1.356
MAPE	u	1	1	1	0.092	0.187	0.281	0.499	**0.997**
MAPE	ur	1	0	1	0.096	*0.225*	*0.362*	0.616	1.466
MAPE	ur	1	1	1	0.092	0.195	0.284	0.536	*1.982*
MAPE	urs	1	0	1	*0.097*	*0.225*	*0.362*	*0.617*	1.623

Note: Best ranking results in **boldface**, worst ranking results in *italics*.

ing. Using only r_{t-1} brings the mean forecast error in inflation down to .001 percent for the ARFIMAX(0,d,0), ARFIMAX(1,d,0), ARFIMAX(0,d,1), ARIMAX(0,1,1) and ARIMAX(1,1,1) models. The RMSE is brought down to 0.113 percent per month. Using u_{t-1} or s_{t-1} does not lead to predictions that outperform the univariate models in short run forecasting. Using a combination of u_{t-1} and r_{t-1} does not improve upon the univariate models either. We observe similar results for forecast horizons of three and six months. Using only r_{t-3} and r_{t-6} respectively, leads to the most accurate forecasts. Forecast efficiency as measured by the RMSE is improved by two percent compared with univariate specifications.

For the one-year and two-year horizons we find that r_t is no longer the dominating leading indicator. At the one-year horizon, the ARIMAX(1,1,1) model with s_{t-12} as a single leading indicator provides the most accurate forecasts, although the other models with s_{t-12}, not shown in the table, are still dominated by their counterparts with r_{t-12}. At the two year horizon we see that u_{t-24} is the best leading indicator, despite the slowly declining effect we observed in the recursive parameter estimates in figure 3.5. However, the univariate models provide forecasts with better RMSEs.

Apparently, a substantial part of the long swings in inflation the 1980s and 1990s is successfully accounted for by the positive effect of lagged short term interest rates. On the other hand, an increase in short term interest rates can also be associated with a lower inflation in the longer run. This was illustrated by the recursive estimates of the negative coefficient γ_{24}, of r_{t-24}. This coefficient is less pronounced than the coefficient of r_{t-1}.

Overall, the ARIMA(1,1,1) model dominates the forecast performance of the other specifications for the error term. Although the persistent shifts in inflation can be modelled by explanatory variables, this does not entail a better forecasting performance for the ARFIMAX(0,d,0) model, compared with ARIMAX(1,1,1). Comparing the results of tables 3.5 and 3.6 overall, we do not observe large increases in forecasting precision by the addition of explanatory variables, but the regression variables do help for short run forecasting. For $H = 24$ the univariate models outperform the regression models on all criteria. Adding more regressors worsens the forecasting performance at all horizons. For short horizons, the precision loss is a few percent, but for two-year-ahead forecasting the differences are dramatic, resulting in RMSEs which are 20 percent larger than for univariate models.

Table 3.7 presents the predicted forecast root mean squared error, RMSEmod, of the leading indicator models and compares them with their actual forecast RMSEs. Comparing the RMSEmods for the ARFIMAX(0,d,0) model using r_{t-H} in table 3.7 with the RMSEmods for the univariate ARFIMA(0,d,0) model in table 3.5 we see lower values for shorter horizons and higher values for $H = 24$. These higher RMSEmods are connected with two effects, first the lower explanatory power of r_{t-24} compared with r_{t-1} and second the higher estimated d in the model with $H = 24$. Overall, the predicted forecast error standard deviations as measured by RMSEmod are much larger than actual RMSEs, especially for the ARIMAX(1,1,1) models. For horizon 1, RMSEmods are a factor 1.6 too large on average. For $H = 24$ this factor is even 2.7 for the ARIMAX(1,1,1) model with leading indicator u_{t-24}. The most likely reason for this overestimation of the scale of forecast intervals is the persistently low innovation variance over the period of the forecasting exercise, compared to the variance in the earlier estimation period.

Table 3.7: Predicted and actual root mean squared error of leading indicator models for U.S. core inflation

	Model				H				
	X	p	d	q	1	3	6	12	24
RMSEmod	r	1	0	1	0.179	0.406	0.741	1.381	3.117
RMSEmod	r	1	1	1	**0.181**	**0.414**	**0.754**	1.522	3.342
RMSEmod	r	0	1	1	0.186	0.413	0.782	1.719	4.296
RMSEmod	r	1	d	0	0.179	0.402	0.728	1.367	2.737
RMSEmod	r	0	d	0	0.179	0.405	0.722	1.328	2.561
RMSEmod	r	0	d	1	0.179	0.402	0.725	1.357	2.713
RMSEmod	s	1	1	1	0.182	0.404	0.698	**1.464**	3.382
RMSEmod	u	1	1	1	0.183	0.408	0.697	1.316	**3.367**
RMSEmod	ur	1	0	1	0.177	0.397	0.700	1.282	3.233
RMSEmod	ur	1	1	1	0.180	0.401	0.692	1.327	3.026
RMSEmod	urs	1	0	1	0.177	0.396	0.694	1.222	2.251
RMSEmod	urs	1	1	1	0.180	0.398	0.686	1.314	2.885
RMSEmod/RMSE	r	1	0	1	1.553	1.661	1.962	2.098	1.800
RMSEmod/RMSE	r	1	1	1	**1.600**	**1.808**	**2.257**	2.720	1.618
RMSEmod/RMSE	r	0	1	1	1.587	1.692	2.130	2.795	3.171
RMSEmod/RMSE	r	1	d	0	1.563	1.706	2.111	2.392	1.630
RMSEmod/RMSE	r	0	d	0	1.575	1.732	2.098	2.308	1.570
RMSEmod/RMSE	r	0	d	1	1.567	1.716	2.110	2.376	1.625
RMSEmod/RMSE	s	1	1	1	1.569	1.636	1.946	**2.851**	1.988
RMSEmod/RMSE	u	1	1	1	1.577	1.755	1.938	2.137	**2.705**
RMSEmod/RMSE	ur	1	0	1	1.493	1.481	1.622	1.713	1.875
RMSEmod/RMSE	ur	1	1	1	1.570	1.694	1.907	1.991	1.260
RMSEmod/RMSE	urs	1	0	1	1.482	1.478	1.612	1.622	1.187
RMSEmod/RMSE	urs	1	1	1	1.570	1.608	1.542	1.814	1.198

Note: Results for best ranking models in RMSE terms, see table 3.6, in **boldface**.

In the next section we apply weighted estimation to account for the level shifts in the innovation variance, which seem to have accompanied the level shifts in the mean of inflation. We shall see that this weighted estimation provides an adequate remedy for the overestimation of inflation forecast uncertainty in the 1980s and 1990s.

3.3 Recursive weighted ARFIMAX forecasting

So far we have not considered changes in the innovation variance. It is not unreasonable to assume that the Volcker-Greenspan regime in the 1980s and 1990s not only reduced the mean, but also the (innovation) variance of inflation. This was already indicated by the ARCH tests for the full-sample models in tables 3.2a-3.2b. As we are considering a level shift in variance, it is more natural to apply a heteroskedasticity test where the null of homoskedasticity is tested against an innovation variance depending on the regime. We apply a Breusch-Pagan test, which has a χ^2 distribution with 2 degrees of freedom

under the null. The resulting test statistic of around 24 strongly rejects the null in all the models examined so far. Figure 3.7 illustrates the weights for the 3 regimes that we used to model the level shifts in variance. These weights are based on subsample estimates of the innovation variance. The observations in the first regime receive a weight 1/2, the second regime a weight 1/3 and the last regime gets weight 1. Our key assumption is that the inflation rates after 1984 are more important than previous observations to estimate the current innovation variance. In the weighted estimation we use information from the full sample and the results can therefore no longer be interpreted as the analysis of "true" out-of-sample forecasts for the whole period 1984-1999. However, the analysis indicates that for the latter part of the forecast sample, weighted estimation could have significantly improved the interval forecasts.

A more extreme solution would have been to use only the observations of the last regime, but this also would have made the recursive out-of-sample forecasting analysis of the long memory model practically impossible. In that case, recursive estimates of the RMSEmods would only be relevant for the last few years of the forecast sample. However, some eight years of observations from the beginning of the estimation sample can be deleted. We experimented with recursively deleting the observations from the first regime. The key parameters remained relatively stable, even when we used only data from 1970 onward.

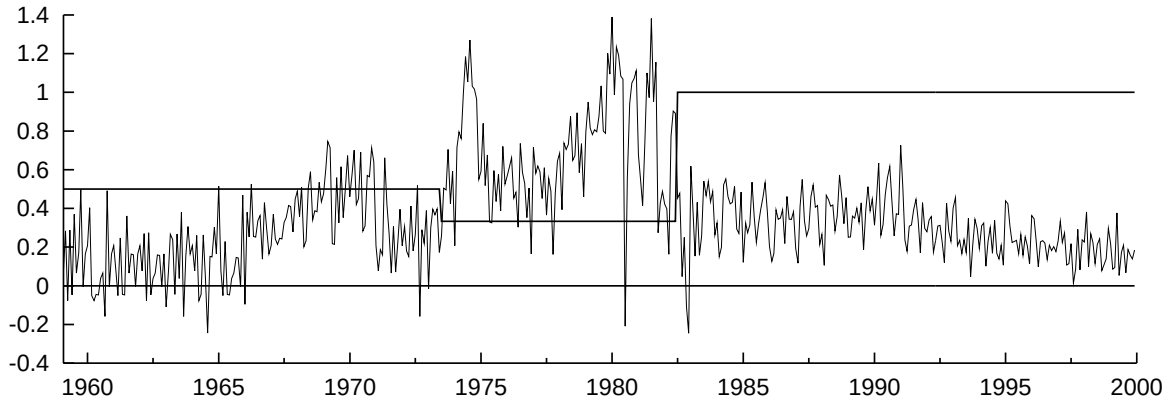

Figure 3.7: Weights used in weighted estimation

The introduction of the different weights has a beneficial influence on the estimated variance of the parameter estimates and on the (scaled) estimate of the innovation variance. However, the recursive estimates of the dynamic parameters, not shown here, do not change significantly by the weighting, although they naturally become more variable as the most recent observations have the highest weights. The weighted estimation leads to similar or lower forecast errors. Estimated standard errors are significantly lower and forecast intervals prove to be more realistic over our forecasting period. Tables 3.8a-3.8b present the same statistics as provided in tables 3.5 and 3.6, but allowing for weighting. The main purpose of the weighting, reducing the overall difference between RMSEmods and RMSEs, is clearly achieved, in particular for the ARFIMAX(0,d,0) model and the ARMAX(1,1) model. The RMSEmods of the ARIMAX(1,1,1) model for $H = 24$ still seem too large.

Table 3.8a: Simulated cumulative out-of-sample forecasts 1984-1999 for univariate models and regression models for U.S. core inflation, using weighted estimation

	Model				H				
	X	p	d	q	1	3	6	12	24
MFE		1	0	1	0.008	0.025	0.068	0.196	0.582
MFE		1	1	1	0.007	0.023	0.058	0.167	0.482
MFE		0	1	1	0.006	0.018	0.048	0.142	0.426
MFE		0	d	0	0.008	0.030	0.078	0.196	0.478
MFE	r	1	0	1	0.002	**-0.001**	**0.010**	**0.029**	0.442
MFE	r	1	1	1	**0.000**	0.014	0.060	0.146	0.924
MFE	r	0	1	1	0.001	0.013	0.053	0.129	0.822
MFE	r	1	d	0	**-0.000**	0.005	0.041	0.086	0.775
MFE	r	0	d	0	**-0.000**	0.004	0.033	0.074	0.681
MFE	r	0	d	1	-0.000	0.005	0.040	0.084	0.770
MFE	s	1	0	1	0.002	0.004	**0.010**	**0.029**	0.705
MFE	u	1	0	1	0.007	0.019	0.032	0.065	**0.221**
MFE	u	1	1	1	*0.017*	0.068	0.162	0.274	0.563
MFE	ur	0	1	1	0.005	0.043	0.133	0.172	*1.301*
MFE	urs	1	0	1	**0.000**	-0.010	0.122	0.041	0.871
MFE	urs	1	1	1	0.008	*0.071*	*0.210*	*0.397*	1.178
RMSE		1	0	1	0.117	0.242	0.368	0.653	1.525
RMSE		1	1	1	0.114	0.228	0.330	0.541	**1.214**
RMSE		0	1	1	0.116	0.233	0.342	0.563	1.259
RMSE		0	d	0	0.116	0.241	0.368	0.648	1.417
RMSE	r	1	0	1	0.114	0.238	0.366	0.618	1.519
RMSE	r	1	1	1	**0.113**	**0.225**	**0.328**	0.519	1.716
RMSE	r	0	1	1	0.114	0.230	0.347	0.554	1.741
RMSE	r	1	d	0	0.115	0.234	0.349	0.577	1.525
RMSE	r	0	d	0	0.114	0.233	0.350	0.582	1.525
RMSE	r	0	d	1	0.114	0.234	0.348	0.577	1.523
RMSE	s	1	1	1	0.115	0.246	0.344	**0.500**	1.573
RMSE	s	1	d	0	*0.118*	0.256	0.377	0.623	1.616
RMSE	u	1	1	1	0.115	0.229	0.345	0.552	1.246
RMSE	urs	1	0	1	0.117	*0.260*	0.388	*0.681*	1.723
RMSE	urs	1	1	1	0.114	0.255	*0.418*	0.627	*2.049*

Note: Best ranking results in **boldface**, worst ranking results in *italics*.

Table 3.8b: Simulated cumulative out-of-sample forecasts 1984-1999 for univariate models and regression models for U.S. core inflation, using weighted estimation (continued)

	Model				H				
	X	p	d	q	1	3	6	12	24
MAPE		1	0	1	0.093	0.196	0.299	0.560	1.363
MAPE		1	1	1	0.091	0.180	0.263	0.432	**0.970**
MAPE		0	1	1	0.092	0.182	0.270	0.446	0.988
MAPE		0	d	0	0.092	0.196	0.305	0.564	1.285
MAPE	r	1	0	1	0.092	0.194	0.296	0.511	1.373
MAPE	r	1	1	1	**0.090**	**0.179**	**0.261**	0.418	1.286
MAPE	r	0	1	1	0.092	0.182	0.272	0.438	1.308
MAPE	r	1	d	0	0.092	0.192	0.283	0.487	1.320
MAPE	r	0	d	0	0.091	0.191	0.286	0.491	1.333
MAPE	r	0	d	1	0.091	0.191	0.283	0.487	1.317
MAPE	s	1	1	1	0.092	0.194	0.279	**0.405**	1.256
MAPE	u	1	1	1	0.091	0.184	0.272	0.454	1.012
MAPE	ur	1	0	1	*0.095*	0.213	0.320	0.551	1.402
MAPE	ur	0	1	1	0.092	0.184	0.277	0.459	*1.546*
MAPE	urs	1	0	1	*0.095*	*0.213*	0.310	*0.561*	1.485
MAPE	urs	1	1	1	0.091	0.203	*0.335*	0.526	1.504
MAPE	urs	1	d	0	*0.095*	0.208	0.306	0.504	1.504
MAPE	urs	0	d	0	*0.095*	0.207	0.310	0.512	1.524
MAPE	urs	0	d	1	*0.095*	0.208	0.306	0.505	1.506

Note: Best ranking results in **boldface**, worst ranking results in *italics*.

Table 3.9: Predicted and actual forecast root mean squared error, 1984-1999, of univariate models and regression models for U.S. core inflation, using weighted estimation

	Model				H				
	X	p	d	q	1	3	6	12	24
RMSEmod		1	0	1	0.119	0.252	0.445	0.839	1.482
RMSEmod		1	1	1	0.119	0.263	0.458	0.881	**1.931**
RMSEmod		0	1	1	0.120	0.251	0.450	0.972	2.199
RMSEmod		0	d	0	0.119	0.268	0.460	0.809	1.444
RMSEmod	r	1	0	1	0.115	0.246	0.443	0.823	1.757
RMSEmod	r	1	1	1	**0.116**	**0.258**	**0.456**	0.877	1.904
RMSEmod	r	0	1	1	0.116	0.245	0.453	0.928	2.172
RMSEmod	r	1	d	0	0.115	0.254	0.449	0.804	1.576
RMSEmod	r	0	d	0	0.115	0.255	0.448	0.793	1.501
RMSEmod	r	0	d	1	0.115	0.254	0.449	0.802	1.573
RMSEmod	s	1	1	1	0.117	0.252	0.435	**0.845**	1.910
RMSEmod	u	1	1	1	0.118	0.255	0.429	0.829	1.893
RMSEmod/RMSE		1	0	1	1.017	1.041	1.244	1.285	1.029
RMSEmod/RMSE		1	1	1	1.044	1.154	1.387	1.628	**1.591**
RMSEmod/RMSE		0	1	1	1.038	1.079	1.314	1.647	1.747
RMSEmod/RMSE		0	d	0	1.026	1.112	1.250	1.248	1.019
RMSEmod/RMSE	r	1	0	1	**1.006**	**1.036**	**1.212**	1.331	1.156
RMSEmod/RMSE	r	1	1	1	1.024	1.147	1.388	1.688	1.110
RMSEmod/RMSE	r	0	1	1	1.018	1.065	1.306	1.675	1.247
RMSEmod/RMSE	r	1	d	0	1.003	1.084	1.288	1.393	1.033
RMSEmod/RMSE	r	0	d	0	1.011	1.095	1.281	1.363	0.984
RMSEmod/RMSE	r	0	d	1	1.004	1.086	1.288	1.390	1.033
RMSEmod/RMSE	r	1	1	1	1.018	1.065	1.306	1.675	1.247
RMSEmod/RMSE	s	1	1	1	1.012	1.026	1.267	**1.691**	1.214
RMSEmod/RMSE	u	1	1	1	1.024	1.117	1.242	1.500	1.519

Note: Results for best ranking models in RMSE terms, see table 3.8a, in **boldface**.

3.4 Recursive ARFIMAX forecast tests

In this section we evaluate the statistical performance of the recursive forecast intervals. We test for the adequacy of unconditional coverage, following Christoffersen (1998). We expect the test to reject models where the predicted forecast error variance deviates too much from the actual forecast error variance.

Table 3.10 reports the empirical unconditional coverage probabilities for the best forecasting models. After including the weights in the estimation, the 60% unconditional coverage is approximately correct for the one-step forecasts of all reported models. However, the ARIMA(1,1,1) and ARIMAX(1,1,1) models lead to forecast intervals which are often too wide for multi-step forecasts. Note there are only 7 non-overlapping two-year inflation periods in the evaluation period, so the power of tests based on the coverage probabilities for $H = 24$ is not high.

Table 3.10: Unconditional coverage test of 60% forecast intervals, U.S. core inflation, using weighted estimation

Model				H				
X	p	d	q	1	3	6	12	24
	1	0	1	0.58	0.75*	0.71	0.60	0.29
	1	1	1	0.63	0.81*	0.90*	0.80	0.86
	0	1	1	0.61	0.78*	0.94*	0.87	1.00*
	1	d	0	0.58	0.76*	0.77*	0.60	0.57
	0	d	0	0.59	0.76*	0.74	0.60	0.43
r	1	0	1	0.56	0.71	0.71	0.60	0.43
r	1	1	1	0.57	0.78*	0.94*	0.87*	0.71
r	0	1	1	0.56	0.71	0.94*	0.87*	0.71
r	1	d	0	0.57	0.71	0.77*	0.60	0.43
r	0	d	0	0.57	0.71	0.71	0.60	0.43
s	1	1	1	0.57	0.73*	0.84*	0.87*	0.71
u	1	1	1	0.59	0.73*	0.84*	0.87*	0.86
L_c				0.530	0.478	0.425	0.349	0.236
U_c				0.669	0.718	0.766	0.834	0.925
N				191	63	31	15	7

Note: Empirical coverage probabilities which differ significantly (at 95% confidence level) from the 60% theoretical coverage are indicated by an asterisk. Bottom rows indicate the bounds of the acceptance region. N is the number of non-overlapping forecast intervals, 1984-1999.

Subsequently we investigate whether there is serial correlation in the scale of the actual forecast error distribution using the independence test also suggested by Christoffersen (1998). We expect power for this test against persistent changes in the forecast error variance over the evaluation period.

The results of the independence test are given in table 3.11. Reported are the likelihood ratio test statistics, which have a $\chi^2(1)$ limiting distribution. Since the forecasts are correlated for horizons $H > 1$ we use only one out of every H forecast errors in the independence test. None of the independence tests rejects.

Finally we examine the possibility of improving the forecasts by using combinations of models in a forecast encompassing framework, see e.g. Harvey, Leybourne and Newbold (1998) for a recent review and Harvey and Newbold (2000) for the extension to multivariate forecast encompassing. West (2001) discusses forecast encompassing in the context of forecasts from recursive regressions. The basic idea is to construct a combined forecast f_{ct} as a weighted average of the forecasts of a baseline model 1, f_{1t} and the forecasts of other non-nested models 2 and 3, say f_{2t} and f_{3t}, that is

$$f_{ct} = (1 - \lambda_2 - \lambda_3)f_{1t} + \lambda_2 f_{2t} + \lambda_3 f_{3t}, \qquad 0 \leq \lambda_1, \lambda_2, \lambda_3 \leq 1. \tag{3.2}$$

For testing purposes one rewrites (3.2) as

$$e_{1t} = c + \lambda_2(e_{1t} - e_{2t}) + \lambda_3(e_{1t} - e_{3t}) + u_t, \tag{3.3}$$

with $e_{it} = y_t - f_{it}$ and $u_t = y_t - f_{ct}$. The constant c equals zero if all forecasts are

Table 3.11: Independence test of coverage of 60% forecast intervals, U.S. core inflation, using weights

	Model			H				
X	p	d	q	1	3	6	12	24
	1	0	1	0.40	0.05	.	0.47	.
	1	1	1	0.01	0.35	.	0.50	.
	0	1	1	0.07	0.07	.	2.14	.
	1	d	0	0.40	0.23	.	0.47	.
	0	d	0	1.07	0.23	0.10	0.47	.
r	1	0	1	0.95	0.39	0.20	0.30	.
r	1	1	1	0.30	0.07	.	2.14	.
r	0	1	1	0.05	0.03	.	2.14	.
r	1	d	0	0.13	0.39	.	0.30	.
r	0	d	0	0.01	0.39	0.20	0.30	.
s	1	1	1	0.45	0.75	0.39	.	0.59
u	1	1	1	1.07	0.20	.	2.14	.
N				191	63	31	15	7

Note: Likelihood ratio test on independence. which asymptotically is $\chi^2(1)$ distributed, see Christoffersen (1998). Included in the test are only N independent forecasts at a distance of H months. A . indicates there are not enough relevant observations.

unbiased. This assumption does not always hold, so it is good practice to add a constant if one employs equation (3.3) as a test regression.

When λ_1 and λ_2 nearly add up to unity, this indicates that the alternative forecasts perform better. When all forecast errors are zero mean Gaussian without serial correlation, standard statistical regression theory applies, at least asymptotically. This analysis provides an easy-to-compute statistical measure of the relative forecasting performance of the models under scrutiny. The procedure provides an extra measure of the usefulness of the different explanatory variables, where not only the in-sample fit, but also the stability of the explanatory power over the period 1984-1999 plays a role.

The first rows of table 3.12 provide bilateral model comparisons, where models with the best leading indicator, r_{t-1}, are chosen as benchmark model 1. Each row in the table corresponds to one (multiple) forecast encompassing test. It appears from test nr. 6 that the ARIMAX(1,1,1) model with u_{t-1} can help to improve the forecasts of the ARFIMAX(0,d,0) with r_{t-1}, but the p-value of this test is not very low. The last rows of table 3.12 show that the forecasts of ARFIMAX models with r_{t-1} cannot be improved by combining them with corresponding ARFIMAX models with other explanatory variables: none of the estimated λ_i's differs significantly from zero.

3.5 Conclusion

Statistical analysis shows that U.S. postwar inflation is long memory, with an order of integration of around 0.3, even after allowing for a structural shift in the mean and variance to capture the high inflation period in the 1970s. Recursive estimation shows

Table 3.12: Multivariate one-step-ahead forecast encompassing tests

nr.	Model 1	Model 2	Model 3	F	p_F
1	$r(0,d,0)$	$r(1,0,1)$			
		0.17 [0.57]		0.32	0.57
2	$r(0,d,0)$	$s(1,0,1)$			
		0.16 [0.61]		0.37	0.54
3	$r(0,d,0)$	$u(1,0,1)$			
		0.17 [0.59]		0.35	0.55
4	$r(0,d,0)$	$r(1,1,1)$			
		0.76 [1.74]		3.01	0.08
5	$r(0,d,0)$	$s(1,1,1)$			
		0.40 [1.78]		3.16	0.08
6	$r(0,d,0)$	$u(1,1,1)$			
		0.52 [1.93]		3.74	0.05
7	$r(0,d,0)$	$s(0,d,0)$			
		0.08 [0.30]		0.09	0.76
8	$r(0,d,0)$	$u(0,d,0)$			
		0.15 [0.49]		0.24	0.62
9	$r(1,0,1)$	$u(1,0,1)$	$s(1,0,1)$		
		0.08 [0.20]	0.13 [0.38]	0.22	0.80
10	$r(1,1,1)$	$u(1,1,1)$	$s(1,1,1)$		
		0.33 [1.07]	0.18 [0.66]	1.31	0.27
11	$r(0,1,1)$	$u(0,1,1)$	$s(0,1,1)$		
		0.22 [0.69]	0.19 [0.70]	0.85	0.43
12	$r(1,d,1)$	$u(1,d,1)$	$s(1,d,1)$		
		0.12 [0.28]	0.01 [0.03]	0.07	0.93
13	$r(0,d,0)$	$u(0,d,0)$	$s(0,d,0)$		
		0.17 [0.39]	−0.02 [−0.06]	0.12	0.88

Note: The columns under model 2 and model 3 give the parameter estimates of λ_2 and λ_3 in (3.3), with corresponding t-values in brackets. Models are indicated by their single leading indicator and the orders p, d, q. Results compare forecasts of model 2 and model 3 with the baseline model 1. F-statistics for $\lambda_2 (= \lambda_3) = 0$ and corresponding p-values presented under F and p_F.

that the order of integration has remained quite stable. Statistical analysis of dynamic regression models for inflation conditioning on lags of unemployment and interest rates shows a stable effect of short lags of short term interest rates. The errors of the regression models are still long memory. An ARMAX(1,1) model and an ARIMAX(1,1,1) model provide a similar in-sample fit as the ARFIMAX(0,d,0) model.

We performed a recursive out-of-sample forecasting exercise for the period 1984-1999 for cumulative inflation forecasting up to a two-year horizon, using univariate models and long memory regression models. The ARIMA(1,1,1) model performs better than the ARFIMA(0,d,0) model regarding the precision of point forecasts. The introduction of conditioning variables improves forecasting precision at short horizons. With regard to forecast interval estimation, downweighting the observations in the 1970s turns out to

be essential in order to get realistic intervals for inflation in the 1990s. The empirical multi-step prediction intervals for the ARIMA(1,1,1) model with or without explanatory variables are too wide, also when weighting is applied. The multi-step forecast intervals of the ARFIMAX(0,d,0) model prove to be more realistic.

Chapter 4

Bayesian Sampling Methods

4.1 Introduction

After applying classical statistical methods in chapters 2 and 3, the subsequent chapters are dedicated to Bayesian approaches to estimating the posterior distributions of model parameters, and making decisions using the parameter distributions.

The difference between the classical and the Bayesian approach to statistics is a fundamental change of paradigm. In the classical analysis, it is assumed that we know the correct structure of the model, and want to estimate the true value of the model parameters. The important concept here is this 'true value of parameters'. In classical statistics, it is assumed that a specific true value of the parameters exists, but that we do not know it. There is no sense in claiming that a certain parameters lies with probability α between bounds c_1 and c_2; instead in classical statistics it should be stated that the region $[c_1, c_2]$ contains with probability α the true parameter value. The data and statistics based on the data like the bounds of a confidence interval are considered to be stochastic, whereas the parameters are fixed.

From a Bayesian viewpoint, there is no such thing as a true parameter value. All we have is a collection of data, which from every practical viewpoint is given, fixed, and therefore not random. Furthermore we may have some prior idea about what value of the parameters could be expected, summarized in a (possibly non- or little informative) prior density of the parameters. Combining the prior density with the likelihood of observing the data a posterior density of the parameters is constructed. This posterior density is a quantitative, probabilistic description of the knowledge about the parameters in the model.

The advantage of a Bayesian analysis is that the complete posterior distribution of the parameters can be used for further analysis: In chapter 5 we are interested in the optimal hedging decision under uncertainty. As we can construct the complete probability distribution of the parameters, we can compute the (probability) distribution of the utility that is derived for a specific hedging decision, and choose the hedging decision which maximizes expected utility. In contrast, a classical analysis would usually optimize utility conditional on a vector of parameters (derived e.g. through maximizing the likelihood of the model), disregarding the parameter uncertainty.

Even though the Bayesian approach has these clear advantages in terms of the manner

in which estimation and decision-making are natural counterparts, the approach only became mainstream in the last decade of the previous century. Several reasons can be recognized for the slow adaption of this alternative statistical framework.

First, and most importantly of all, the paradigm change which is needed when moving from classical to Bayesian statistics is not done light-heartedly by statisticians and econometricians. After regarding parameters as unknown, fixed, quantities for centuries, Bayes (1763) presented a parameter as a random variable, with an associated distribution. When observing data, we move from a state of relative ignorance[1] concerning the parameter to a state where we are informed, corresponding to a posterior distribution for the parameter which is concentrated on the region where the parameter is expected to be found.

This switch, towards viewing parameters as random instead of fixed was just one obstacle for the Bayesian approach to gain ground. A second obstacle was the computational effort connected to the derivation of posterior distributions. Only in special cases an analytical solution for the posterior density is available, in other cases higher order integrals have to be solved numerically.

The second problem was relieved with the advent of Monte Carlo sampling techniques for calculating integrals. In the econometric literature importance sampling (Kloek and Van Dijk 1978) was the first available method for using sampling techniques for approximating densities and integrals. The method itself dates back to Hammersley and Handscomb (1964). In the engineering literature, sampling methods were used decades earlier than in econometrics. E.g. the article by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) introduced a method which is now known under the name of Metropolis-Hastings (MH for short) sampling, partly named after Hastings (1970) who introduced a generalized version of the original method in the statistical literature. A special case of the MH algorithm, the Gibbs sampler, was introduced by Geman and Geman (1984) and applied to the reconstruction of images.[2] Gelfand and Smith (1990) explained the work of Geman and Geman (1984) and of Tanner and Wong (1987), starting off a further boom in research on sampling methods in econometrics.

With the advent of ever faster computers, the research into the field of modern sampling methods bloomed, and continues to do so. New sampling methods keep being introduced, often building forth on older methods (see e.g. Durbin and Koopman (2000), who apply the importance sampler to improve on a method called the simulation smoother (De Jong and Shephard 1995)).

In the present chapter, section 4.2 provides a short review of the aforementioned integration/sampling methods (see e.g Bauwens, Lubrano and Richard 1999, for a more elaborate description). The exposition of these basic methods is followed by two brief

[1] We may have some prior ideas on the parameter, therefore it could be said that we initially are in a state of *relative* ignorance.

[2] In the application of the reconstruction of images (or also of the reconstruction of sound) the target is to retrieve the original image, or the original sound of an old recording. All scratches, missing pieces, have to be filled in in a manner which seems to be most plausible given the data at hand. This 'data at hand' in this case is the damaged image/recording, the prior information of the researcher can consist of the manner in which one sound wave can follow from the previous one, or what a picture might roughly look like. Using the Gibbs sampler, there are techniques to sample the posterior 'density' of the picture, getting rid of implausible scratches in the picture or recording.

sections indicating lines of proof of convergence of the Monte Carlo methods, and with some points of attention in practical situations. A range of not-so-standard extensions of the algorithms is given in section 4.3, most importantly a concise description of the technique of Metropolis-within-Gibbs (section 4.3.1), the multistep Gibbs sampler (section 4.3.2) and the method of Adaptive Polar Sampling (section 4.3.4).

The chapter continues with a description of the concept of marginal likelihood and the closely connected posterior odds; this section 4.4 concentrates on aspects of the marginal likelihood which are specific to the use of large data sets (as in chapter 5). It is followed in section 4.5 by examples of some of the methods explained here, and concluding remarks in section 4.6.

4.2 Basic sampling methods

Most of the scientific or more practical research question posed can be be written in the format "What is the expected value of $g(\theta)$?", where $g(\theta)$ may be an inflation figure, precipitation, maximum loss for a large investor, or percentage of votes for a political party, possibly depending on a set of parameters θ. In mathematical terms, the object of interest is

$$\mathsf{E}(g(\theta)) = \int_\theta g(\theta) p_\theta(\theta) d\,\theta. \tag{4.1}$$

The basis for all Monte Carlo integration methods is the approximation of the integral[3] through a sample mean,

$$\mathsf{E}(g(\theta)) \approx \frac{1}{N} \sum_i g(\theta^{(i)}). \tag{4.3}$$

In this equation $\theta^{(i)}, i = 1, \ldots, N$ is a sample from the distribution p_θ. As it is often not possible to sample directly from p_θ (coined the target distribution in the following), the algorithms mentioned in section 4.1 come into play.

4.2.1 Importance sampling

In Kloek and Van Dijk (1978) the method of importance sampling is introduced. It is aimed at calculating integrals of the form (4.1) when a sample θ from the target density $p_\theta(\theta)$ is not available. An approximating candidate density $q_\theta(\theta)$ is used, which relates to

[3]When only the kernel of the density $p_\theta(\theta)$ is known, the integral in equation (4.1) is divided by the integral over the kernel. This would lead to a formula

$$\mathsf{E}(g(\theta)) = \frac{\int_\theta g(\theta)\kappa_\theta(\theta) d\,\theta}{\int_\theta \kappa_\theta(\theta) d\,\theta}, \tag{4.2}$$

and similar straightforward modifications to subsequent formulas in this chapter. For simplicity of notation, the integrating constant is assumed known.

the target according to the weight function $w(\theta) = p_\theta(\theta)/q_\theta(\theta)$, see figure 4.1. From the observation that

$$\mathsf{E}(g(\theta)) = \int_\theta g(\theta)\frac{p_\theta(\theta)}{q_\theta(\theta)}q_\theta(\theta)d\,\theta = \int_\theta g(\theta)w(\theta)q_\theta(\theta)d\,\theta, \tag{4.1′}$$

the approximation is calculated as

$$\mathsf{E}(g(\theta)) \approx \frac{1}{N}\sum_i g(\theta^{(i)})w(\theta^{(i)}). \tag{4.3′}$$

The sample $\theta^{(i)}, i = 1, \ldots, N$ is drawn not from the target density $p_\theta(\theta)$ but from the candidate density $q_\theta(\theta)$, adapting for the difference between the two through the use of weights in (4.3′). The method of integrating through importance sampling works well if the candidate density $q_\theta(\theta)$ approximates the target closely, i.e. when the weight function $w(\theta)$ is close to 1. In the tails of the density, this is often not attainable; care should be taken in that case to choose a candidate density with heavier tails than those of the target density, as the weights $w(\theta)$ could get very large in the opposite case. Also, the expectation of the weight function, $\mathsf{E}(w(\theta))$, must be finite.

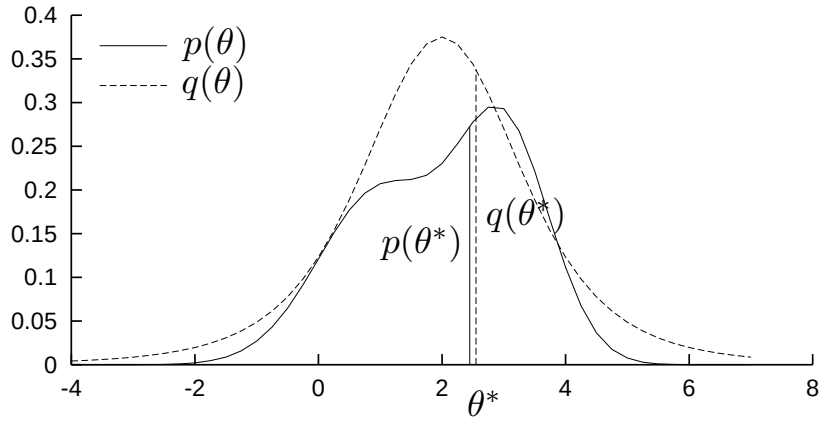

Figure 4.1: Sampling using the importance sampler, with target and candidate density

4.2.2 Sampling/Importance Resampling

The method of importance sampling from the previous section delivers not a sample from the target density, but only a weighted sample. Rubin (1987) introduces the technique of Sampling/Importance Resampling (SIR), which basically uses the weights of the importance sample as probabilities of retaining a specific sampled $\theta^{(i)}$, i.e.

i. Sample $\theta^{(1,*)}, \ldots, \theta^{(R,*)}$ from a candidate density $q(\theta)$, computing weights $w(\theta) = p(\theta)/q(\theta)$

ii. Resample a sample $\theta^{(1)}, \ldots, \theta^{(M)}$ of size $M \ll R$ from the parameter vectors sampled in step i, each with probability $p_{\mathrm{SIR}}(\theta) = w(\theta)/\sum w(\theta)$.

Only when the candidate density approximates the target well, resulting in probabilities $p_{\mathrm{SIR}}(\theta)$ with low variance, can the SIR sampling method be expected to be relatively efficient. If the probabilities are very different, a large sample R in the first step is needed, afterwards disregarding a large part of the sampled $\theta^{(i,*)}$ with low weights.

4.2.3 Acceptance-rejection sampling

The importance sampling algorithm results in a sample $\theta^{(i)} \sim q_\theta(\theta)$ from the candidate density with corresponding weights, which can be used to calculate the integral in (4.1). In many situations it is convenient to have a sample which comes from the correct target distribution, such that without applying any weight the approximation in (4.3) can be used.

For this sampling method, the target density has to be covered by a kernel of the candidate density, i.e. a constant c has to be found such that

$$p_\theta(\theta) \leq c\, q_\theta(\theta) \qquad \forall\theta.$$

Instead of drawing directly from the target density p_θ (as the algorithm is only used in cases where drawing from p_θ is impossible or impracticle) a value is drawn from the candidate, say θ^*. This drawing is accepted with probability

$$\alpha_{\mathrm{AR}} = \frac{p_\theta(\theta)}{c q_\theta(\theta)} \tag{4.4}$$

and rejected otherwise (see figure 4.2 for an example). If a sample of N elements is needed, sampling candidates and evaluating the acceptance-rejection probability continues until N elements are accepted. The sample of accepted parameter vectors is a sample from the correct target density, no further weighting is needed.

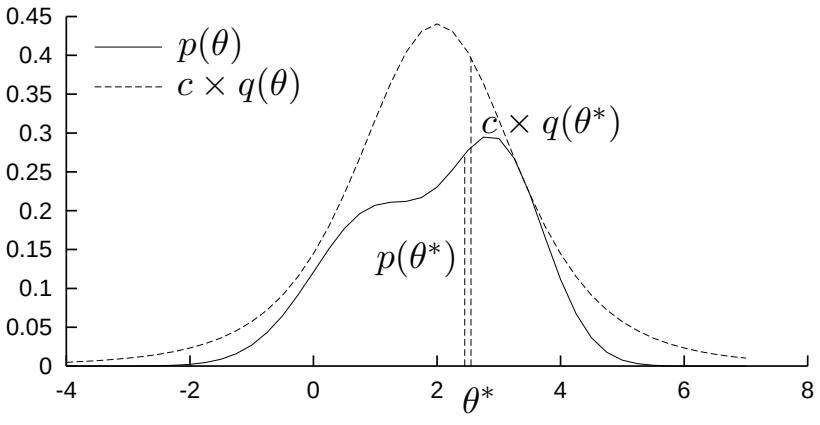

Figure 4.2: Sampling using the acceptance/rejection sampler, with target density and enveloping candidate kernel

Like the importance sampling algorithm, this method is good in its simplicity. Drawbacks of this algorithm are that a candidate density is needed together with a multiplication factor c such that the kernel $cq_\theta(\theta)$ covers the target density. If the factor c is much

larger than 1, the number of accepted drawings will be low and the algorithm therefore little efficient. In such a case the importance sampling algorithm would still be able to use those drawings, though with small weights having little influence on the final value of the integral.

4.2.4 The Metropolis-Hastings algorithm

The origins of the Metropolis-Hastings algorithm date back to the 1950s, when the algorithm was introduced by Metropolis et al. (1953). Hastings (1970) reintroduced it to the econometric community, though a clear exposition had to wait until Smith and Roberts (1993) and later Chib and Greenberg (1995).

The algorithm is a true sampling algorithm (like the acceptance-rejection algorithm of the previous section) in the sense that it results in a sample $\Theta = (\theta^{(1)}, \theta^{(2)}, \ldots)$ which may be used for evaluating all kinds of objective functions $g_1(\theta), g_2(\theta)$ etc. As in the case of both the acceptance-rejection and importance sampling, the target density $q(\theta|\theta^{(i)})$[4] is compared to a candidate density. The algorithm consists of the following steps:

i. Initialize, start with a drawing $\theta^{(0)}$, set $i = 0$.

ii. Generate a candidate draw $\theta^* \sim q(\theta|\theta^{(i)})$.

iii. Calculate the acceptance probability

$$\alpha_{\text{MH}}(\theta^{(i)}, \theta^*) = \min\left[\frac{p_\theta(\theta^*)q(\theta^{(i)}|\theta^*)}{p_\theta(\theta^{(i)})q(\theta^*|\theta^{(i)})}, 1\right]. \tag{4.5}$$

iv. With probability $\alpha_{\text{MH}}(\theta^{(i)}, \theta^*)$ set $\theta^{(i+1)} = \theta^*$, else retain $\theta^{(i+1)} = \theta^{(i)}$.

v. Increase i.

vi. Repeat steps ii-v until a sufficiently large sample is collected (see sections 4.2.6-4.2.7).

Note that this algorithm results in a chain of drawings, where there is correlation between drawing $\theta^{(i+1)}$ and $\theta^{(i)}$.[5]. This chain is a Markov chain, depending only on the state $\theta^{(i)}$ in the previous period. See also section 4.2.6.

Without the intention of being anywhere near exhaustive (see e.g. Geweke (1999) for a more elaborate exposition), some remarks can be made. When the candidate density does not depend on the state of the chain, an independence chain results, with transition probability

$$\alpha_{\text{MH}}(\theta^{(i)}, \theta^*) = \min\left[\frac{p_\theta(\theta^*)q(\theta^{(i)})}{p_\theta(\theta^{(i)})q(\theta^*)}, 1\right] = \min\left[\frac{w(\theta^*)}{w(\theta^{(i)})}, 1\right]. \tag{4.6}$$

[4]The candidate density may depend on the last vector of parameter $\theta^{(i)}$ drawn in the chain. This last vector of parameters is often called the state of the chain.

[5]There are two sources of correlation. Firstly, when a candidate draw is rejected, the old element $\theta^{(i)}$ is duplicated. Secondly, the candidate density may depend on the previous drawing $\theta^{(i)}$, resulting in correlated draws

Note how in this case the acceptance probability depends on the weights which were also used in the Importance Sampler, section 4.2.1. Often such an independence chain, with e.g. $q(\theta^*) = t(\hat{\theta}, \alpha, \nu)$ a Student-t density with as expectation $\hat{\theta}$ a preliminary estimate of the mode of the posterior density and variance $\frac{\nu}{\nu-2}\alpha^2$ works well on unimodal posterior densities.

When the candidate density is symmetric ($q(\theta^a|\theta^b) = q(\theta^b|\theta^a)$), it cancels from the acceptance probability equation, leaving

$$\alpha_{\text{MH}}(\theta^{(i)}, \theta^*) = \min\left[\frac{p_\theta(\theta^*)}{p_\theta(\theta^{(i)})}, 1\right]. \tag{4.7}$$

This is the original algorithm as proposed in Metropolis et al. (1953). A special case is the candidate density which only depends on the distance between θ^* and θ, i.e. $q(\theta^*|\theta) = q(\theta^* - \theta)$. The resulting chain is known under the name of Random Walk Metropolis chain. A simple choice is to use $\theta^* \sim \mathcal{N}(\theta, \sigma_q^2)$; the variance of the candidate density can be calibrated to take steps which are reasonably close to θ such that the probability of accepting the candidate is not too low, but with a stepsize large enough to ensure sufficient mixing of the chain.

4.2.5 Gibbs sampling

The Gibbs sampler is possibly the sampling technique which is used most frequently. In the statistical physics literature it was (and still is) known as the heat bath algorithm, but Geman and Geman (1984) christened it in the mainstream statistical literature as the Gibbs sampler. It was popularized by Casella and George (1992) among econometricians, and consists of splitting up the parameter space into blocks of parameters for which it is possible to specify the full conditionals: Let θ be the parameter vector, and let it be subdivided into $\theta = \{\theta_1, \ldots, \theta_k\}$. Then the algorithm proceeds as follows:

i. Initialize, start with a drawing $\theta^{(0)}$, set $i = 0$.

ii. Given the i-th drawing $\theta^{(i)}$, the next one is found by simulating

$$\begin{aligned}
\theta_1^{(i+1)} &\sim \pi(\theta_1|\theta_2^{(i)}, \ldots, \theta_k^{(i)}),\\
\theta_2^{(i+1)} &\sim \pi(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_k^{(i)}),\\
&\vdots\\
\theta_k^{(i+1)} &\sim \pi(\theta_k|\theta_1^{(i+1)}, \ldots, \theta_{k-1}^{(i+1)}).
\end{aligned}$$

iii. Increase i.

iv. Repeat steps ii-iii until convergence (see sections 4.2.6-4.2.7).

Given a starting vector of parameters $\theta^{(0)}$ in the support of the density, the algorithm proceeds by stepwise generating each element of the next $\theta^{(i+1)}$ from the full conditionals as above. The algorithm can be considered to 'pass through' intermediate points

$(\theta_1^{(i+1)}, \theta_2^{(i)}, \ldots, \theta_k^{(i)})$, $(\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_k^{(i)})$, $\ldots$, $(\theta_1^{(i+1)}, \ldots, \theta_{k-1}^{(i+1)}, \theta_k^{(i)})$ towards the new element $\theta^{(i+1)}$.

An important concept connected with the Gibbs sampler is the concept of data augmentation, already introduced by Tanner and Wong (1987). In many situations, the posterior density $p(\theta)$ is hard to sample from, but there is a conditional density $p(\theta|z)$ which is easily analysed. This often occurs in models with missing or unobserved data, e.g. Tobit or Probit models. If also the distribution of $z|\theta$ is of a known form, a Gibbs chain is easily built sampling from $z|\theta^{(i)}$ followed by a draw of $\theta^{(i+1)} \sim p(\theta|z)$. Disregarding the values of z sampled during the process, the $\theta^{(i)}$ can be shown to have the correct distribution $p(\theta)$ (see Casella and George (1992) for details).

4.2.6 On convergence: Theory

The Markov chain Monte Carlo methods all try to find a transition kernel to move from one iteration to the next, such that eventually a drawing from the target density results. This target density then is the invariant density $\pi(y)$, i.e.

$$\pi^*(dy) = \int_\Omega P(x, dy)\pi(x)\, dx.$$

In this formula, $P(x, A)$ is the transition kernel describing the probability of moving from a location $x \in \Omega$ to a location within the set $A \subset \Omega$. This equation states that starting from the invariant density $\pi(x)$, taking another step $P(x, dy)$ according to the transition kernel leads to the same invariant distribution π^* (note that $\pi(x)$ indicates the invariant density, $\pi^*(dx)$ the distribution with the corresponding Lebesgue measure).

For all MCMC methods the transition kernel is constructed as the sum of the probability of a move and of no move, thus

$$P(x, dy) = p(x, y)dy + r(x)I(x \in dy).$$

In this function, $r(x) = 1 - \int_\Omega p(x, y)dy$ is the probability of no move, triggered by the indicator function $I(x \in dy)$. With this notation introduced, Theorem 1 in Tierney (1994) states

Theorem 1 *Assume that $P(.,.)$ is π^*-irreducible and has π^* as its invariant distribution. Then P is positive recurrent and π^* is the unique invariant distribution of P. If P is also aperiodic, then for π^*-almost all $x \in \Omega$ and all sets $A \in \omega$*

$$||P^n(x, A) - \pi^*(A)|| \longrightarrow 0.$$

In this theorem $||\ ||$ denotes a total variance distance measure. The concept of irreducibility excludes cases where the chain can get caught in a subset $B \in \Omega$ with zero probability of reaching a location $x \notin B$ in a finite number of steps. Aperiodicity is needed to ensure that Ω cannot be split up into subsets B_1 and $B_2, B_1 \cup B_2 = B$ such that

$$P^{n-1}(x, B_2) = P^n(x, B_1) = \begin{cases} 1 & \text{if } n \text{ even} \\ 0 & \text{otherwise} \end{cases}$$

Theorem 3 in Tierney (1994), or the second part of Proposition 1 in Chib and Greenberg (1996), gives the result needed for convergence of quantities as in equation (4.3):

Theorem 2 *When the function g is real valued and π^*-integrable, the sample average*

$$\frac{1}{N}\sum g(\theta^{(i)}) \longrightarrow \int_\theta g(\theta)\pi_\theta(\theta)d\,\theta$$

as $N \to \infty$.

In general it is easily confirmed if the Metropolis chain fulfills the conditions of Theorem 1. When the support of the candidate density is at least as large as of the target density, and the chain is started at an interior point of the support, in most cases the convergence conditions are met. In practice one has to be careful that all parts of the sample space are visited regularly; if in some region the candidate probability is a lot smaller than the target probability, elements in this region have to be replicated often, leading to strong correlation in the sample. See Geweke (1999) for a more precise rendering of the conditions.

With the Gibbs sampler a practitioner should verify that indeed the target density is not degenerate. Proper full conditional densities do not guarantee a proper invariant distribution (see Hobert and Casella 1996). Secondly, strong correlation between elements of the target density can lead to even stronger correlation in successive elements of the chain, possibly resulting in a quasi- or total reducibility of the sample space.[6] The convergence of the Gibbs sampler is furthermore hampered by a large number of blocks. Each extra block introduces an extra full conditional distribution, which in its turn leads to an extra source of correlation.

4.2.7 On convergence: Practice

Theorem 1 of section 4.2.6 states that eventually the drawing from a Markov chain can be considered to be a drawing from the invariant distribution. Each subsequent drawing is also a drawing from the invariant distribution, but a dependent one: There can be a strong correlation between successive drawings. The following measures can be taken to counter this correlation:

i. **Single chain**: In some cases, correlation is not much of a practical problem. As long as the researcher realizes that the sample of size N is a dependent sample, and therefore only corresponds in information content to a sample of $N_1 < N$ drawings, results can be sufficiently good. With strong correlation, a larger sample is needed than in the case when the correlation is relatively weak.

ii. **Multiple chains**: A range of N different Markov chains of length N_2 can be run, saving only the last drawing of the chain. If the length N_2 is large enough for the chain such that the drawing has converged to the target distribution, the N parameter vectors at the end of each of the chains are independent.

[6] Or, more popularly stated: 'Correlation kills the Gibbs sampler'.

iii. **Interspersed chain**: An intermediate position is obtained when one long single chain of length $N \times N_3$ is run, saving only one drawing out of every N_3. The N resulting drawings have lower correlation than the sample from the single chain itself, and the computational overhead is less than with the multiple-chain method, as usually $N_3 \ll N_2$ is sufficient to get a considerable reduction in correlation.

For assessing convergence, several statistics have been devised (Geweke 1992, Yu and Mykland 1998). The basic idea behind most of them is to compare moments of the sampled parameters at different parts of the chain. The implementation of Yu and Mykland (1998) is most practical. They propose to consider a plot of the CUSUM path with

$$\text{CUSUM}_t = \frac{1}{t\sigma_\theta} \sum_{i=1}^{t} (\theta^{(i)} - \mu_\theta), \qquad t = 1, \ldots, N. \tag{4.8}$$

The cumulative sum of the drawings is adapted for the empirical mean and standard deviation of the complete sample. This is not necessary, but can be convenient to compare the CUSUM plots for different parameters. A plot of CUSUM_t against time which diverges from zero for a prolonged period of time is an indication of bad convergence.

Apart from the CUSUM plot it is often convenient to check convergence by visual means from a plot of the drawings themselves, of a moving average of e.g. the last 100 drawings, and of the cumulative or running mean of drawings $1, \ldots, t$.

Geweke (1992) promotes the use of the Relative Numerical Efficiency (RNE) as a measure for the quality of a correlated sample. It compares the empirical variance of the sample with a correlation-consistent variance estimator,

$$\text{RNE} = \frac{\sigma_\theta^2}{\sigma_{\text{NW},q}^2}, \tag{4.9}$$

where σ_θ^2 is a direct estimator of the variance, and $\sigma_{\text{NW},q}^2$ is the Newey-West (Newey and West 1987) variance estimator taking the correlation up to lags of $q\%$ of the size of the sample into account. Practical values for q can be 4, 8 or even 15%.

4.3 Extensions

In this section several extensions to the basic MCMC algorithms are presented. The aim of presenting these elaborations is to provide the researcher with ideas on how to implement an effective sampler for the case at hand. Each extension is presented in a concise manner, with enough detail for implementation, and references to relevant literature for further details.

The choice of extensions to the sampling methods made here is limited, to include the most common algorithms and especially the algorithms which are used in the example in section 4.5 and in chapter 5. Gilks, Richardson and Spiegelhalter (1996) describe a range of other sampling methods among which are the Hit-and-Run algorithm (Chen and Schmeiser 1998) and the class of Adaptive Direction Samplers (Gilks, Roberts and George 1994). Also not covered below are sampling methods which allow for a sampling

space which randomly changes in dimension, like the Simulated Tempering (Marinari and Parisi 1992, Liu and Sabatti 1998) and the Reversible Jump algorithms (Green 1995, Richardson and Green 1997).

4.3.1 Metropolis-within-Gibbs

The name of this section is, according to some, badly chosen, as is the name of the Gibbs sampler, in as far as the name seems to imply that the sampler applies a different algorithm than the Metropolis-Hastings sampler. Effectively, the Gibbs, and also the algorithm known as Metropolis-within-Gibbs, are special cases of the very flexible and powerful Hastings sampling methodology.

Chib and Greenberg (1995) present this point most clearly, in an elaboration of a 'block-at-a-time' algorithm. It is shown how a chain consisting of two or more full conditional subchains converges to a chain delivering drawings from the joint distribution, i.e. if

$$P(x_1, dy_1|x_2) \xrightarrow{\text{i.d.}} \pi^*(dy_1|x_2)$$

and

$$P(x_2, dy_2|x_1) \xrightarrow{\text{i.d.}} \pi^*(dy_2|x_1)$$

(with $\xrightarrow{\text{i.d.}}$ signifying 'has as invariant distribution'), then

$$P(x_1, dy_1|x_2) \times P(x_2, dy_2|x_1) \xrightarrow{\text{i.d.}} \pi^*(dy_1, dy_2). \tag{4.10}$$

When both the conditional transition kernels equal the respective conditional distributions (and therefore equal the invariant conditional distributions, without any need for convergence in the conditional steps), the pure Gibbs algorithm results. However, when one (or more) of the transition kernels is of the Metropolis-Hastings type, then an algorithm results which is sometimes called the Metropolis-within-Gibbs. This algorithm is especially useful when one of the full conditional densities cannot be sampled from directly, but it is possible to provide an approximating sampler. Note that the conditioning is always done on the latest available drawing of a block. When e.g. the MH step is used for sampling a new $x_2|x_1$, then the candidate density for the MH step can only use (condition on) the last $x_1^{(i)}$, and not a previous $x_1^{(i-1)}$, even though conditioning on a previous $x_1^{(i-1)}$ can be computationally more efficient when it comes to calculating the acceptance probability α_{MH}.

4.3.2 Multistep Gibbs samplers revisited

Under the heading of the Gibbs sampler the general case was already presented in section 4.2.5. Though conceptually the bivariate Gibbs sampler is most easily explained, in theory the number of full conditional densities in the chain is not limited, leading to (highly) multivariate Gibbs samplers. Similarly, the multistep Gibbs sampler could be introduced

along the lines of the Metropolis-within-Gibbs sampler, with the Metropolis part replaced by another Gibbs step: A Gibbs-within-Gibbs sampler ìs a multistep Gibbs sampler.

It seems there is nothing new under the sun on this matter. Even so, several practical issues connected to the implementation of a multivariate Gibbs sampler merit some discussion. First of all, there is the issue of the amount of preparation which is needed before the Gibbs sampler can be implemented (but see also section 4.3.3, on the Griddy Gibbs sampler). With the MH sampler, only the (analytic) posterior density function is needed, in a closed formula. For Gibbs, a separate full conditional density for each block has to be derived, including a method of sampling from the corresponding conditional distribution. In section 4.5 an example is given of a model of which parameters are sampled in a Gibbs chain. In this model, 5 different conditional densities have to be derived, including methods to sample from them. Each derivation is prone to errors, therefore using the Gibbs sampler on highly dimensional problems can become problematic.

A more serious problem is the correlation in the chain. With each conditional distribution of $\theta_i|\theta_{-i}$ (with θ_{-i} indicating the parameter vector θ excluding element i) extra dependence is introduced. In section 4.5.4 an example is given where indeed the correlation is very high, leading to a low quality of the sample drawn using the Gibbs sampler. Though the example is quite extreme, even in normal cases (see also chapter 5, where correlations starts to be problematic in a real-world situation) the correlation induced by a sequence of multiple Gibbs steps can occur.

4.3.3 Griddy Gibbs sampler

For some models the conditional density $p(\theta_i|\theta_{-i})$ is hard to derive or to sample from. In those cases the Griddy Gibbs sampler (Ritter and Tanner 1992, Bauwens, Lubrano and Richard 1999, section 3.4.3.2) can offer relief.

The Griddy Gibbs sampler constructs an approximation $\tilde{p}(\theta_i|\theta_{-i})$ to the conditional density numerically, by evaluating the (joint) posterior density on a grid over the support of values for $\theta_i|\theta_{-i}$ (that is, keeping the conditioning parameters θ_{-i} constant). As

$$p(\theta_i|\theta_{-i}) = \frac{p(\theta_i, \theta_{-i})}{\int_{\theta_i} p(\theta_i, \theta_{-i})}, \tag{4.11}$$

the conditional density of $\theta_i|\theta_{-i}$ is proportional to the joint density.

To sample from a general density function p with cumulative distribution function P, we can take a drawing u from a uniform distribution, and apply the inverse CDF P^{-1} to arrive at a drawing $\theta = P^{-1}(u)$ from the original distribution. This method can be used given a numerical approximation $\tilde{P}(\theta_i|\theta_{-i})$ to $P(\theta_i|\theta_{-i})$ to sample a new value of θ_i. Figure 4.3 displays a general density function $p(\theta)$ in the left panel, together with an approximation based on a linear spline connecting a number of support points. Using the linear approximation, the cumulative distribution $\tilde{P}(\theta_i|\theta_{-i})$ in the right panel was constructed. For a drawing $u = 0.62$ from the uniform density, the corresponding value of θ can be read from the x-axis.

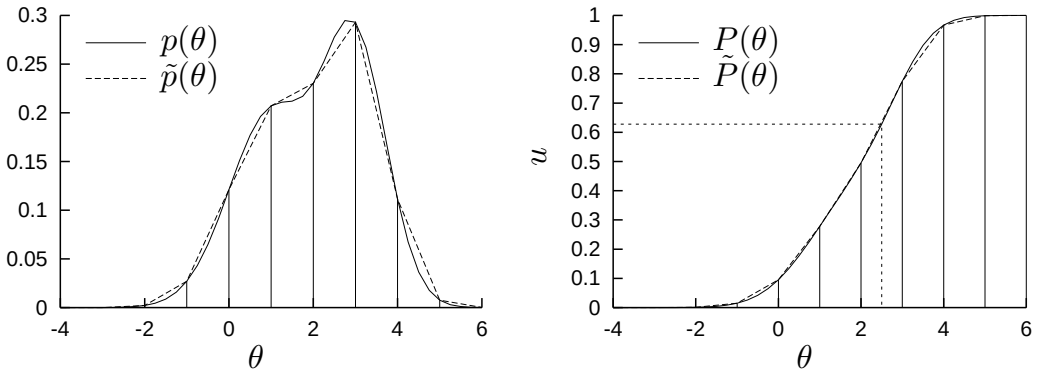

Figure 4.3: Sampling using the Griddy Gibbs sampler

4.3.4 Adaptive Polar Sampling

The Metropolis-Hastings and Gibbs samplers of sections 4.2.4 and 4.2.5 are flexible enough to sample from a broad range of posterior densities. In some situations, when e.g. no good approximating candidate density is available for the MH sampler, or when correlation in the Gibbs chain is very high, alternative sampling methods can help.

In this section the Adaptive Polar Sampling (APS) algorithm is explained. It was first presented in Bauwens, Bos and Van Dijk (1999) and builds upon the MIXIN algorithm (Van Dijk and Kloek 1980, Van Dijk, Kloek and Boender 1985). The idea behind the algorithm can be compared to the ideas behind the Adaptive Direction Samplers (Gilks et al. 1994). The algorithm is devised to be able to sample effectively also from multimodal posterior densities, even when the number and location of the modes is not known a priori; this is a case where the MH sampler often does not converge. Compared with the Gibbs sampler, APS is not hampered by strong correlation between the parameters of the model. Little information is needed except for the posterior density function itself.[7]

The main idea behind the APS algorithm is to split the problem in two: Sampling is done in a transformation of the original parameter space to polar coordinates. The algorithm applies a standard Metropolis algorithm on the directions η of the parameter vectors, and a univariate numerical procedure on the distances ρ. The idea is explained in figure 4.4. Panel **A** displays a bivariate mixture density with two separate modes. Standard sampling algorithms run into trouble sampling from this density. Conditional on a (possibly preliminary, very inaccurate) estimate of the location and the scale of the density the original parameter vector $\theta = (\theta_1, \theta_2)$ is transformed into the space of (y_1, y_2), having mean 0 and a unit covariance matrix (see panel **B**). Then, a second transformation to polar coordinates is applied, leading to (η, ρ). The complete transition is indicated as

$$(\eta, \rho) = T(\theta \,|\, \mu, \Sigma) = T_{y \to \eta,\rho}(y) \circ T_{\theta \to y}(\theta \,|\, \mu, \Sigma) \tag{4.12}$$

(the precise definition of the transformation is given in appendix 4.B). Panel **C** in figure 4.4 plots the marginal density $p(\eta)$ of the directions which results from the bimodal density,

[7] A drawback of the algorithm is that the analytic posterior density function is needed. When data augmentation is necessary to arrive at a tractable posterior density, APS can be used to sample from the augmented parameter space. However, we have at present little experience with such samplers.

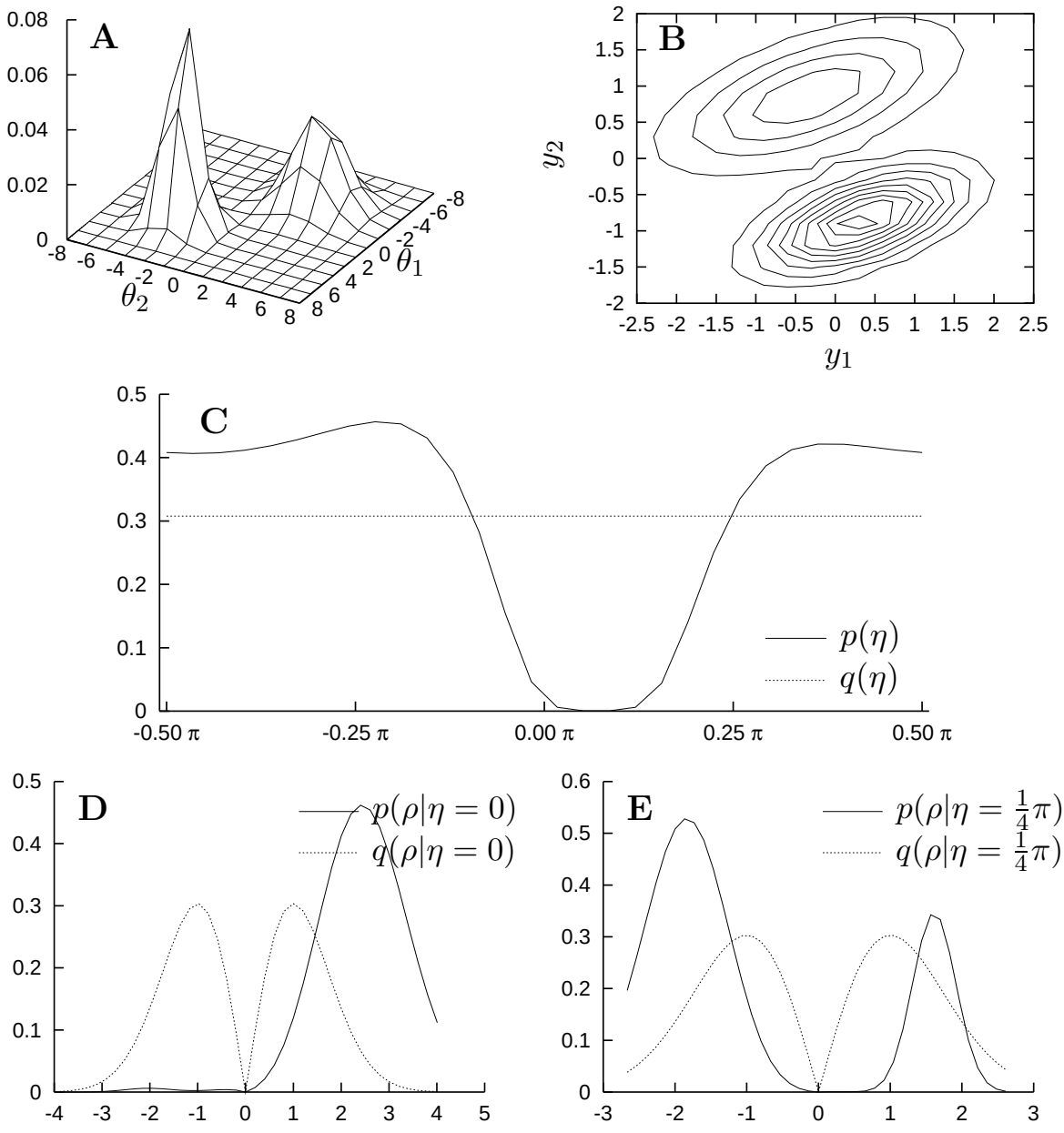

Figure 4.4: Transforming a bivariate mixture density to polar coordinates

and also the candidate density $q(\eta)$ which results from transforming the parameter space for a bivariate normal density. It is seen that for directions $\eta \in [-0.1\pi, 0.2\pi]$ little probability mass is found for the bimodal density, when compared to the (transformed normal) candidate density $q(\eta)$. For other directions, the difference between the density of η for the bimodal and the normal density is not large. On the other hand, the conditional density of $\rho|\eta$ depends strongly on the direction chosen. Panels **D** and **E** show two conditional densities of $\rho|\eta = 0$ and $\rho|\eta = \frac{1}{4}\pi$, together with the conditional densities resulting from the transformed normal density, indicated by $q(\rho|\eta)$. Here the differences in densities are larger, with the mass of the transformed bimodal density lying further away from the center of the distribution.

The sampling from directions and distances is based on the observation that the marginal density of directions η is well behaved compared to the density of the parameters in the original space. Problematic heavy tales, multimodality, or oddly shaped support for the parameters translates into an oddly shaped density for the conditional $\rho|\eta$, less so for the marginal density of η. Therefore, Bauwens, Bos and Van Dijk (1999) propose to sample a candidate direction η^* from the transformed normal density (corresponding to the uniform density for η in the bivariate case, as in panel **C** of figure 4.4). A Metropolis-Hastings step is applied on η, accepting the candidate draw η^* instead of the previous $\eta^{(i)}$ with probability

$$\alpha(\eta^{(i)}, \eta^*) = \min\left[\frac{p(\eta^*)q(\eta^{(i)})}{p(\eta^{(i)})q(\eta^*)}, 1\right] \tag{4.6'}$$

$$= \min\left[\frac{\int_\rho p_\theta\left(T^{-1}(\eta^*, \rho)\right)|J(\rho)|\,\partial\rho}{\int_\rho p_\theta\left(T^{-1}(\eta^{(i)}, \rho)\right)|J(\rho)|\,\partial\rho}, 1\right], \tag{4.13}$$

with $J(\rho)$ indicating the part of the Jacobian of the transformation depending on the distance parameter ρ.[8] Alternatively, the candidate direction η can be given a weight as in the importance sampler (section 4.2.1). The resulting Adaptive Polar Importance Sampling (APIS) algorithm does not reject any drawings, but may collect a large number of sampled parameter vectors with low weights if the candidate density is not very precise (i.e., especially when the estimates of location and scale are not very accurate).

During the numerical evaluation of the univariate integrals $\int_\rho p_\theta\left(T^{-1}(\eta^*, \rho)\right)|J(\rho)|\,\partial\rho$ and $\int_\rho p_\theta\left(T^{-1}(\eta^{(i)}, \rho)\right)|J(\rho)|\,\partial\rho$, information is collected to construct the conditional density $p(\rho|\eta^{(i+1)})$.[9] From the conditional distribution (see the method explained in section 4.3.3, on the Griddy Gibbs sampler) we draw a value of ρ, which together with $\eta^{(i+1)}$ defines a point $\theta = T^{-1}(\eta^{(i+1)}, \rho^{(i+1)})$ in the original parameter space.

Sampling new values of η and ρ (or, equivalently, of θ) continues until an improved estimate of the location and scale parameters μ and Σ can be derived from the drawings, or until the sample is large enough for other purposes. As the sampling of η is computationally intensive due to the numerical integral over ρ that has to be evaluated, it is advisable to sample multiple values of ρ for the same direction η. As long as a sufficient

[8] The transition from (4.6′) to (4.13) is explained in the appendix.

[9] Note that $\eta^{(i+1)}$ is either η^* or $\eta^{(i)}$, such that the conditional density of $\rho|\eta^{(i+1)}$ is proportional to $p(\eta^{(i+1)}, \rho)$ which is merely a transformation from $p_\theta\left(T^{-1}(\eta^*, \rho)\right)$.

number of different directions η have been sampled to cover the parameter space of the directions, this does not hamper convergence.

Practical experience tells us that good initial estimates of location and scale are not needed: Using one or more short initial sets of drawings, the estimates can easily be updated. During initial rounds, it is advisable to force acceptance in the Metropolis-Hastings step after a low (e.g. 5) number of rejections, to keep the chain moving. This leads to a sample which is not from the correct posterior distribution, but which serves well for improving the estimate of the location and scale parameters. In later rounds, forced acceptance should not (or only after a large number of rejections) occur. The structure of the APS algorithm ensures that the final sample converges to a sample from the posterior density function, irrespective of initial starting values for location and scale parameters. For an indication of the proof, see Bauwens, Bos and Van Dijk (1999).

4.4 Posterior odds and marginal likelihood

4.4.1 Posterior odds, the Bayes factor, and model choice

When models are estimated in a classical manner, they can be compared on the basis of the likelihood they attain. The likelihood function is evaluated in the point indicated by the parameter estimates, often at the location of maximum likelihood. In a Bayesian framework, there is not one parameter vector characterizing the fit of the model. Instead, based on the likelihood and the prior, the full posterior distribution of the parameters is derived (see section 4.2 on sampling the posterior distribution). Characteristic for the fit of a model M is in this case the expected or marginal likelihood $m(Y \mid M)$, where the expectation is taken over the likelihood $\mathcal{L}(Y; \theta \mid M)$ with respect to the prior distribution $\pi(\theta \mid M)$ of the parameters,

$$m(Y \mid M) = \int_{\theta} \mathcal{L}(Y; \theta \mid M) \pi(\theta \mid M) \, \partial\theta. \tag{4.14}$$

Based on this marginal likelihood the posterior odds can be calculated, summarizing the evidence in favour of or against a model, compared to a competing model. The posterior odds (PO) are

$$\text{PO} = \frac{m(Y \mid M_1)}{m(Y \mid M_2)} \frac{\pi(M_1)}{\pi(M_2)} = \text{Bayes factor} \times \text{prior odds} \tag{4.15}$$

with $m(Y \mid M_i)$ the marginal likelihood of data Y assuming model i, and $\pi(M_i)$ the prior probability attached to this model. Note that for the posterior odds and the Bayes factor to exist, the priors specified on the parameters have to be non-degenerate.

The Bayes factor (BF), which is the ratio of the marginal likelihoods of the competing models, reports the evidence in the data in favour of or against a model. It will be seen in later sections that the evidence as indicated by the Bayes factor can take on very large or very small values. Kass and Raftery (1995) propose to use a classification based on the logarithm of the Bayes factor, and consider the evidence in favour of a model M_1 as compared to model M_2 'not worth more than a bare mention' if $\log(\text{BF}) < 1$, to consider

it 'positive' if $1 \leq \log(\text{BF}) < 3$, 'strong' when $3 \leq \log(\text{BF}) < 5$ and 'very strong' for even larger values.

4.4.2 Calculating the marginal likelihood

Methods based on the posterior kernel

Only in very special cases, most notably for the exponential likelihood with conjugate priors, the marginal likelihood m can be analytically calculated as the integrating constant[10] of the posterior, as

$$p(\theta|Y) = \frac{1}{m}\mathcal{L}(Y;\theta)\pi(\theta). \tag{4.16}$$

For other models, numerical solutions have to be found. A good overview is given in Kass and Raftery (1995). The most important methods are described below.

The brute force method is to do numerical integration using Gaussian quadrature to integrate the product of likelihood and prior. This method may be labelled a 'brute force' method, as it can be very inefficient. Only in lower dimensional problems, Gaussian quadrature is sufficiently fast. Furthermore, in general the likelihood function is very peaked, leading to an integrand which only has a positive (not nearly zero) value in a small area. The strong contrast between the (large) region with virtually zero likelihood and the (small) region with high likelihood makes it hard to integrate in a numerically stable way.

A second method (McCulloch and Rossi 1992) calculates the marginal likelihood using simulation, as an approximation to the expectation of the likelihood with respect to the prior distribution of the parameters:

$$m = \mathsf{E}_{\pi(\theta)}\mathcal{L}(Y;\theta) \approx \frac{1}{N}\sum \mathcal{L}(Y;\theta^{(i)}) = m_{\text{Prior}} \tag{4.17}$$

with $\theta^{(1)}, .., \theta^{(N)}$ a sample of size N from the prior density. This method is not very efficient either, as the sampled values from the prior may well be in a region with relatively low likelihood, necessitating the use of a large sample size.

The efficiency of the simulation method can be improved using importance sampling (IS), i.e. by sampling from a density $\pi^*(\theta)$ with more mass in regions of high likelihood. Weights have to be used to adapt for the fact that we are not sampling from the correct distribution, leading to an estimator m_{IS} calculated as

$$m_{\text{IS}} = \frac{\sum w_i \mathcal{L}(Y;\theta^{(i)})}{\sum w_i}, \qquad \text{with weights } w_i = \frac{\pi(\theta^{(i)})}{\pi^*(\theta^{(i)})}. \tag{4.18}$$

It seems logical to sample from the posterior density $\pi^*(\theta) = p(\theta|Y) \propto \mathcal{L}(Y;\theta)\pi(\theta)$ if that is possible, as this favours drawings of the parameter vector in regions with high posterior mass. This leads to

$$m_{\text{HM}} = \frac{\sum m}{\sum \frac{m}{\mathcal{L}(Y;\theta^{(i)})}} = \frac{Nm}{m\sum \frac{1}{\mathcal{L}(Y;\theta^{(i)})}} = \left(\frac{1}{N}\sum \frac{1}{\mathcal{L}(Y;\theta^{(i)})}\right)^{-1} \tag{4.19}$$

[10] As we are calculating the marginal likelihood of a specific model M on a fixed data set Y in this section, dependence of m on these quantities is suppressed in the notation.

with HM indicating that this estimator uses the harmonic mean (Newton and Raftery 1994). Though this estimator is consistent, the variance $\mathsf{var}\big(1/\mathcal{L}(Y;\theta)\big)$ may be infinite, especially if $\mathcal{L}(Y;\theta)$ has thin tails. In such a case, some sampled parameter vectors $\theta^{(i)}$ can lead to huge contributions $1/\mathcal{L}(Y;\theta^{(i)})$. In several papers methods are sought to improve on the variance of this measure, leaving intact the consistency of the estimator. A possibility put forward in Newton and Raftery (1994) is to combine in the importance sampling density both the prior, with a certain (small) weight δ, and the posterior, with weight $1-\delta$. Then $\pi^*(\theta) = \delta\pi(\theta) + (1-\delta)p(\theta \mid Y)$ is used in calculating m_{IS} in (4.18). This estimator still is efficient and consistent, and can be shown to satisfy a Gaussian central limit theorem. A more recent solution to stabilize the harmonic mean estimator is presented in Satagopan and Newton (2000). They calculate the estimator m_{SHM} as

$$m_{\mathrm{SHM}} = \left(\frac{1}{N}\sum \frac{1}{p(Y \mid h(\theta^{(i)}))}\right)^{-1} \tag{4.20}$$

with $h(\theta)$ a function reducing the parameter space. The elements in the harmonic mean are marginal likelihoods in themselves, with part of the parameter space integrated out. The transformation $h(\theta)$ has to be chosen such that the marginalization to $h(\theta)$ is more easily calculated than the full marginalization (if the full marginalization were available, that would result in the marginal likelihood straight away). No clear rules for finding such a transformation are given for general situations.

As the marginal likelihood is the integrating constant of the posterior density, the idea arises to use an approximating density p_{App}. The relative difference in height between the approximating density and the kernel of the posterior is an estimate of the integrating constant. A range of methods can be derived from this observation, using

$$m_{\mathrm{App}} = \frac{\mathcal{L}(Y;\theta)\pi(\theta)}{p(\theta \mid Y)} \approx \frac{\mathcal{L}(Y;\theta)\pi(\theta)}{p_{\mathrm{App}}(\theta)}. \tag{4.21}$$

First, LaPlace's method approximates the logarithm of the posterior kernel at the mode $\tilde{\theta}$, $l(\tilde{\theta})$, using a quadratic expansion. The approximation is found to be a normal distribution with mean $\tilde{\theta}$ and covariance matrix $\tilde{\Sigma} = \left(\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}\right)\Big|_{\theta=\tilde{\theta}}$. Substituting the normal density in (4.21) gives

$$m_{\mathrm{LP}} = (2\pi)^{k/2}|\tilde{\Sigma}|^{1/2}\mathcal{L}(Y;\tilde{\theta})\pi(\tilde{\theta}). \tag{4.22}$$

Conditions for the accuracy of this method are found in Kass, Tierney and Raftery (1990).

Alternatively, when the prior is relatively uninformative, the kernel can be approximated at the location of the maximum likelihood estimator $\hat{\theta}$ and corresponding covariance matrix $\hat{\Sigma}$. The estimator $m_{\mathrm{LP(ML)}}$ is calculated according to (4.22), using the maximum likelihood estimate instead of the estimate at the posterior mode. Likewise, for symmetric posterior densities it is often simpler to find the location of the posterior mean or median from a sample, and calculate $m_{\mathrm{LP(Mean)}}$ or $m_{\mathrm{LP(Median)}}$ correspondingly.

When a sample from the posterior is available, the posterior density can be approximated using a kernel smoothing algorithm. When the resulting smoothed posterior density is made to integrate to one, the marginal likelihood m_{Kern} can be calculated comparing

the posterior kernel to the smoothed posterior density estimate in (4.21). For the location θ any point could be chosen, though best results are found at a location of high density, e.g. at the posterior mode $\tilde{\theta}$. Note that for higher dimensional models, getting a sufficiently accurate smoothed density estimate may require a very large sample. In our experience, results of this method are precise enough for comparisons between models when the dimension of the parameter vector is less than 10.

A method using the Gibbs sampler

The methods described above work well for small dimensional problems, in general for methods where the importance sampler or the Metropolis-Hastings algorithm could be used. In cases where data augmentation is needed to arrive at a tractable likelihood function, or when the full conditional densities are easier to sample from than using the IS or MH methods, the Gibbs sampler is used. Chib (1995) describes a method for consistently estimating the marginal likelihood based on the output of the Gibbs chain.

Again, the starting point for the calculations is the marginal likelihood identity, in a rewritten version of equation (4.16),

$$m(Y) = \frac{\mathcal{L}(Y;\theta)\pi(\theta)}{p(\theta|Y)}. \tag{4.23}$$

Equation (4.23) holds at every location θ in the parameter space.[11] The posterior kernel $\mathcal{L}(Y;\theta)\pi(\theta)$ is known in closed form, and through the Gibbs sampler we can generate drawings θ from the posterior $p(\theta|Y) \propto \mathcal{L}(Y;\theta)\pi(\theta)$. The posterior $p(\theta|Y)$ is the problematic element here, as we do not know the integrating constant (the integrating constant is $m(Y)$, exactly the element of interest of the whole exercise). Note how the posterior can be written as the product of conditional densities, for each of the building blocks θ_i of $\theta = (\theta_1, .., \theta_k)$:

$$p(\theta|Y) = p(\theta_1|Y) \prod_{i=2}^{k} p(\theta_i|\theta_1, .., \theta_{i-1}, Y). \tag{4.24}$$

When using the Gibbs sampler, not the above conditional densities but only the full conditional densities are known. Therefore we need the relation

$$\begin{aligned} p(\theta_i|\theta_1, .., \theta_{i-1}, Y) &= \int p(\theta_i|\theta_1, .., \theta_{i-1}, \theta_{i+1}, .., \theta_k, Y) \\ &\qquad \times p(\theta_{i+1}, .., \theta_k|\theta_1, .., \theta_{i-1}, Y)\, \partial\theta_{i+1}..\partial\theta_k \\ &\approx \frac{1}{N} \sum_j p(\theta_i|\theta_1, .., \theta_{i-1}, \theta_{i+1}^{(j)}, .., \theta_k^{(j)}, Y), \qquad i = 1, .., k-1. \end{aligned} \tag{4.25}$$

The conditional density can be approximated averaging out over the full conditional density, with $\theta_{i+1}^{(j)}, .., \theta_k^{(j)}$ one of N drawings from the density $p(\theta_{i+1}, .., \theta_k|\theta_1, .., \theta_{i-1}, Y)$. These

[11] The parameter vector θ can include the elements resulting from data augmentation. Only at the end of the present section we distinguish between the parameters of interest and the nuisance parameters.

drawings can be sampled using the original Gibbs sampler, with this difference that elements $\theta_1, .., \theta_{i-1}$ are kept fixed. Note how in this sample θ_i is sampled, but not used except for keeping the chain moving.

For each of the conditional densities $p(\theta_i|\theta_1, .., \theta_{i-1}, Y), i = 1, .., k-1$ in (4.25) a separate run of the Gibbs sampler is needed, each sampling from $k - i$ different full conditional densities. Only the conditional density with $i = k$ is readily available, as it itself is the full conditional density.

The marginal likelihood identity (4.23) holds regardless of the location θ chosen. However, results can be expected to be more precise at locations with a higher density. When the problem at hand is of a high dimension (this often occurs when use was made of data augmentation, see also the example in section 4.5.5), the mass is dispersed over these many dimensions. In such a case it is advisable to treat the parameters, say the vector z, resulting from the data augmentation differently. As equation (4.23) is derived from

$$p(\theta|Y) = \frac{1}{m(Y)}\mathcal{L}(Y;\theta)\pi(\theta), \tag{4.16}$$

we can integrate out the elements z from the parameter vector $\theta = (\tilde{\theta}, z)$ on both sides, resulting in

$$m(Y) = \frac{\int \mathcal{L}(Y;\tilde{\theta}, z)\pi(\tilde{\theta}, z)\partial z}{\int p(\tilde{\theta}, z|Y)\partial z}. \tag{4.23$'$}$$

Note that both in the numerator and the denominator the integral over z is needed. In some cases (e.g. the state vector in a Gaussian state space model) this integral is known analytically. If it is not known, there is often no simple sampling approximation to it,[12] and the computation of the marginal likelihood should be repeated for a number of high-density vectors $z^{(i)}$, as

$$m(Y) = \frac{1}{q}\sum_{i=1}^{q} \frac{\mathcal{L}(Y;\tilde{\theta}, z^{(i)})\pi(\tilde{\theta}, z^{(i)})}{p(\tilde{\theta}, z^{(i)}|Y)}. \tag{4.23$''$}$$

This computation can become very time consuming, as for the computation of one marginal likelihood already $k - 1$ separate Gibbs chains have to be sampled, leading to a number of $q \times (k - 1)$ separate chains needed for the computation of (4.23$''$).

4.4.3 Calculating the Bayes factor

As indicated in section 4.4.1, the Bayes factor is a statistic indicating the evidence in favour of or against a model, as compared to a second model. The simplest way to calculate the BF is to compute the marginal likelihoods for both models M_1 and M_2 and write

$$\text{BF} = \frac{m(M_1)}{m(M_2)}. \tag{4.26}$$

[12] For the integral in the denominator, we can sample z along from the full conditional densities as in (4.25). For the numerator, $\int \mathcal{L}(Y;\tilde{\theta}, z)\pi(\tilde{\theta}, z)\partial z = \int \mathcal{L}(Y;\tilde{\theta}, z)\pi(\theta|z)\pi(z) \approx \frac{1}{N}\sum \mathcal{L}(Y;\tilde{\theta}, z)\pi(\theta|z^{(i)})$, with $z^{(i)} \sim \pi(z)$ a sample from the prior. See the discussion around the computation of m_{Prior} in (4.17), for a discussion of the problems concerning such a sampling approximation.

A second method can be used when we wish to test for a restriction of the parameter space, e.g. $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$, testing for M_1 with $\theta_2 = \tilde{\theta}_2$ against an unrestricted model M_2. In case the prior of the restricted model equals the restricted version of the prior of the general model,

$$\pi(\theta_1|M_1) = \pi(\theta_1, \theta_2|M_2, \theta_2 = \tilde{\theta}_2) = \pi(\theta_1|\theta_2 = \tilde{\theta}_2, M_2),$$

then it was already shown by Dickey (1971) that

$$\text{BF} = \frac{p(\tilde{\theta}_2|Y)}{\pi(\tilde{\theta}_2)} = \frac{\int p(\theta_1, \tilde{\theta}_2|Y)\partial\theta_1}{\pi(\tilde{\theta}_2)}. \tag{4.27}$$

Dickey attributed the ratio to Savage; nowadays it is known as the Savage-Dickey density ratio.

In cases when we test between two models which do not correspond in the parametrization of the priors, i.e. if

$$\pi(\theta_1|M_1) \neq \pi(\theta_1|\theta_2 = \tilde{\theta}_2, M_2),$$

then Verdinelli and Wasserman (1995) present the generalized Savage-Dickey density ratio. This generalization is computed as

$$\text{BF} = p(\tilde{\theta}_2|Y)\ \mathsf{E}_{\theta_1|\tilde{\theta}_2,M_2}\left[\frac{\pi(\theta_1|M_1)}{\pi(\theta_1, \tilde{\theta}_2|M_2)}\right] = \frac{p(\tilde{\theta}_2|Y)}{P}\pi(\tilde{\theta}_2)\ \mathsf{E}_{\theta_1|\tilde{\theta}_2,M_2}\left[\frac{\pi(\theta_1|M_1)}{\pi(\theta_1|\tilde{\theta}_2, M_2)}\right]. \tag{4.28}$$

This last equation holds if we assume that the prior $\pi(\theta_1, \tilde{\theta}_2|M_2)$ and the posterior $p(\tilde{\theta}_2|Y)$ are bounded for almost all θ_1 and that the expectation is finite.

4.5 Example: Sampling from a time series model

4.5.1 Introducing the model

In this last section of the chapter on sampling methods, an example is presented on which a range of the samplers can be applied. The data set used here is a simulated series similar to the exchange rate data used in chapter 5. The exchange rate data is modelled to exhibit a very erratic behaviour around a stochastic trend, with the trend component only weak compared to the strength of the disturbances. Such a series can be simulated from and modelled with a state space model (Harvey 1989), with exchange rate returns s_t distributed around a mean μ_t. The mean is assumed to follow a random walk (less strict assumptions will be made in chapter 5). The model is written down using a observation equation (4.29) and a transition equation (4.30),

$$s_t = \mu_t + \epsilon_t \tag{4.29}$$

$$\mu_t = \mu_{t-1} + \eta_t. \tag{4.30}$$

The disturbances η_t in the transition equation are assumed to be independent and normally distributed, $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right), t = 1, .., T$. For the observation disturbances, we either

assume $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$, or $\epsilon_t \sim t(0, \sqrt{(\nu-2)/\nu}\sigma_\epsilon, \nu)$.[13] Using the last option, we allow for a heavier tailed disturbance distribution, with $\nu > 2$ the number of degrees of freedom of the Student-t distribution with expectation 0 and variance σ_ϵ^2.

In the example we use values of $\sigma_\epsilon = 1, \sigma_\eta = 0.01, \nu = 4$ and two different sample lengths of $T = 100$ and $T = 1000$, see also table 4.1.

4.5.2 Likelihood and posterior

When the disturbances ϵ_t are normally distributed, the model is entirely linear and Gaussian. In that case, the Kalman filter equations (De Jong 1989, Harvey 1989) lead to a prediction-error decomposition. This decomposition filters out the prediction errors v_t which are normally distributed with variance F_t, conditional on all observations up to and including s_t and the parameters (see appendix 4.A for the Kalman equations). Indicating the past observations by $S_t = \{s_1, .., s_t\}$, the likelihood is

$$\mathcal{L}(S_T; \sigma_\epsilon, \sigma_\eta) = \prod_{t=1}^{T} (2\pi F_t)^{-1/2} \exp\left(-\frac{v_t^2}{2F_t}\right). \tag{4.31}$$

This leaves us with a likelihood which we can calculate analytically, though through a recursive formula.

When the disturbances in the observation equation (4.29) are modelled as Student-t distributed, the Kalman equations of the appendix do not hold, as the model is no longer purely Gaussian. This implied that closed-form (possibly recursive) formula for the likelihood can be found. It is however possible to introduce factors z_t such that the model is conditionally Gaussian:[14]

$$\begin{aligned} \epsilon_t \,|\, z_t &\sim \mathcal{N}\left(0, (\sigma_\epsilon z_t)^2\right), \\ z_t &\sim \text{IG-1}(\alpha_z = \frac{\nu}{2}, \beta_z = \frac{2}{\nu - 2}). \end{aligned}$$

[13] See appendix 4.C for an overview of the notation of the density functions applied here.

[14] In the present context it is more convenient to work with parameters $\sigma_\epsilon, \sigma_\eta$ and z_t instead of $\sigma_\epsilon^2, \sigma_\eta^2$ and z_t^2. This latter choice of parameters would lead to the use of the Inverted Gamma density instead of the Inverted Gamma-1 density. Otherwise, derivations and results are similar. Appendix 4.C states the formulation of this and other density functions.

Table 4.1: Parameters and priors

Parameter	Value	Prior density	Hyperparameters	
σ_ϵ	1	IG-1	$\alpha_\epsilon = 5$	$\beta_\epsilon = 0.2$
σ_η	0.01	IG-1	$\alpha_\eta = 5$	$\beta_\eta = 20$
ν	4	trunc. Cauchy	$\nu > 2$	
$z \mid \nu$	–	IG-1	$\alpha_z = \frac{\nu}{2}$	$\beta_z = \frac{2}{\nu-2}$

With this setup, the marginal density of ϵ_t is Student-t as intended, as we can derive that

$$\begin{aligned}
p(\epsilon_t) &= \int p(\epsilon_t \mid z) p(z) dz \propto \int z^{-1} \exp\left(-\frac{\epsilon_t^2}{2\sigma_\epsilon^2 z^2}\right) z^{-\left(2\frac{\nu}{2}+1\right)} \exp\left(-\frac{\nu-2}{2z^2}\right) dz \\
&= \int z^{-\left(2\frac{\nu+1}{2}+1\right)} \exp\left(-\frac{1}{z^2}\frac{\epsilon^2 + \sigma_\epsilon^2(\nu-2)}{2\sigma_\epsilon^2}\right) dz \\
&\propto \beta^\alpha \int f_{\text{IG-1}}\left(z; \alpha = \frac{\nu+1}{2}, \beta = \frac{2\sigma_\epsilon^2}{\epsilon^2 + \sigma_\epsilon^2(\nu-2)}\right) dz \\
&= \left(\frac{2\sigma_\epsilon^2}{\epsilon^2 + \sigma_\epsilon^2(\nu-2)}\right)^{\frac{\nu+1}{2}} \propto \left(1 + \frac{\epsilon^2}{\sigma_\epsilon^2(\nu-2)}\right)^{-\frac{\nu+1}{2}} \\
&\sim t\left(0, \sqrt{\frac{\nu-2}{\nu}}\sigma_\epsilon, \nu\right).
\end{aligned}$$

With the parameter vector augmented to include the elements $z_1, .., z_T$ the model is Gaussian again, and the likelihood can be written down for both the standard state space model as for the model with Student-t error terms. Combining the likelihood with a prior we get the kernel of a posterior density. We choose conjugate prior densities, to simplify the derivation of the full conditional Gibbs sampling densities at a later stage. For σ_ϵ and σ_η, the conjugate prior density is the Inverted Gamma-1 density. Hyperparameters α and β are chosen such that the prior expectation of the standard deviations σ_ϵ and σ_η is approximately correct. The standard deviation of the prior density is of similar size as the expectation and the true value (see table 4.1).

4.5.3 Sampling methods based on the likelihood function

With the likelihood function (and therefore also the posterior) known as a closed formula in the case of normal errors, we can apply methods like the importance sampler (section 4.2.1) and the Metropolis-Hastings sampler (section 4.2.4). These methods need an approximating density which is easy to sample from. A choice which often works reasonably well[15] is to use a Student-t density, calibrating the mean and covariance of the candidate to mimic those of the posterior density in the mode. Output of a standard maximization routine can be used to find the mode $\hat{\mu}$; the covariance matrix $\hat{\Sigma}$ can be approximated

[15] In cases when the posterior density is known to be unimodal with tails which are not in some directions much heavier than those of a Student-t density, this method can be expected to give a good approximation.

as the inverse of minus the matrix of second derivatives around the mode. The degrees-of-freedom parameter can be fixed at e.g. $\nu = 4$ if the posterior is expected to be clearly non-normal, or at higher values if the deviation from normality is not large. A further improvement is found when one takes the results from a preliminary round of the sampler based on the estimates $\hat{\mu}$ and $\hat{\Sigma}$, and adjusts these estimates to the sample mean and variance. In the following implementations of the IS and MH samplers, we initially use the estimated mode and covariance matrix from the optimization routine, and adapt these three times using increasing sample sizes, before collecting the final sample.

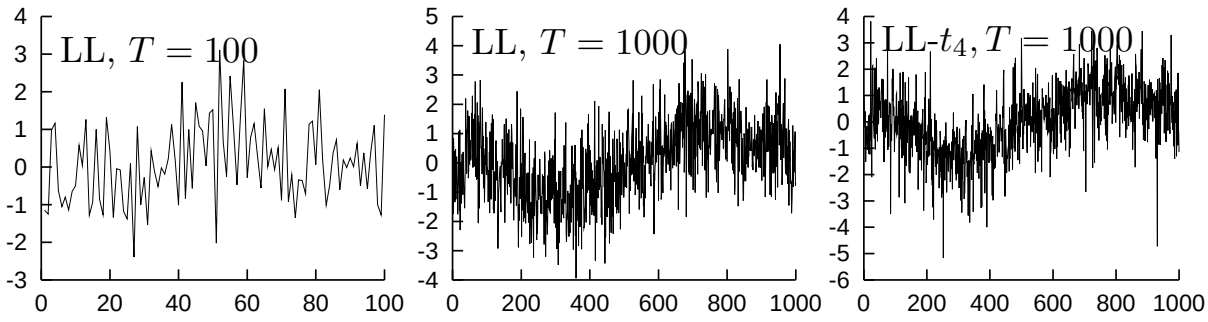

Figure 4.5: Simulated data from the Local Level model, with $T = 100$ (left panel) and $T = 1000$ (middle), and from the Local Level model with Student-t disturbances ϵ_t, $T = 1000$ (right)

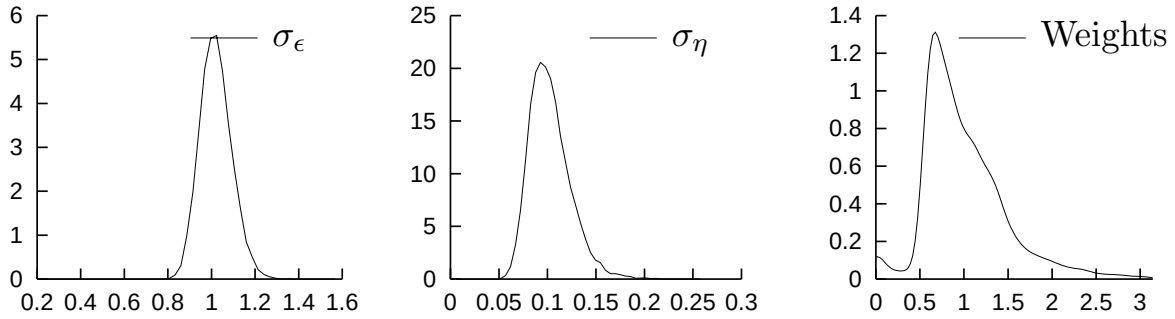

Figure 4.6: Posterior from importance sampler, with weights, $T = 100$

The output of the importance sampler is a series of drawings from the candidate density with corresponding weights. The closer these weights are to 1 (indicating a perfect candidate), the better are the results. Figure 4.5 plots, in the left panel, a sample of length $T = 100$ from the model described in section 4.5.1. This data set was used as input in sampling 10000 drawings using the importance sampler, with a Student-t candidate fitted to the mode of the posterior, with $\nu = 4$ degrees of freedom. The resulting posterior is plotted in figure 4.6, for the parameters σ_ϵ and σ_η. The third panel in the figure shows the distribution of the weights: The bulk of the weights is close to 1. Only a few observations are weighted heavily (at a weight of 4), and some other observations receive an almost

zero weight, leading to a small second mode at zero in the graph of the weights.[16]

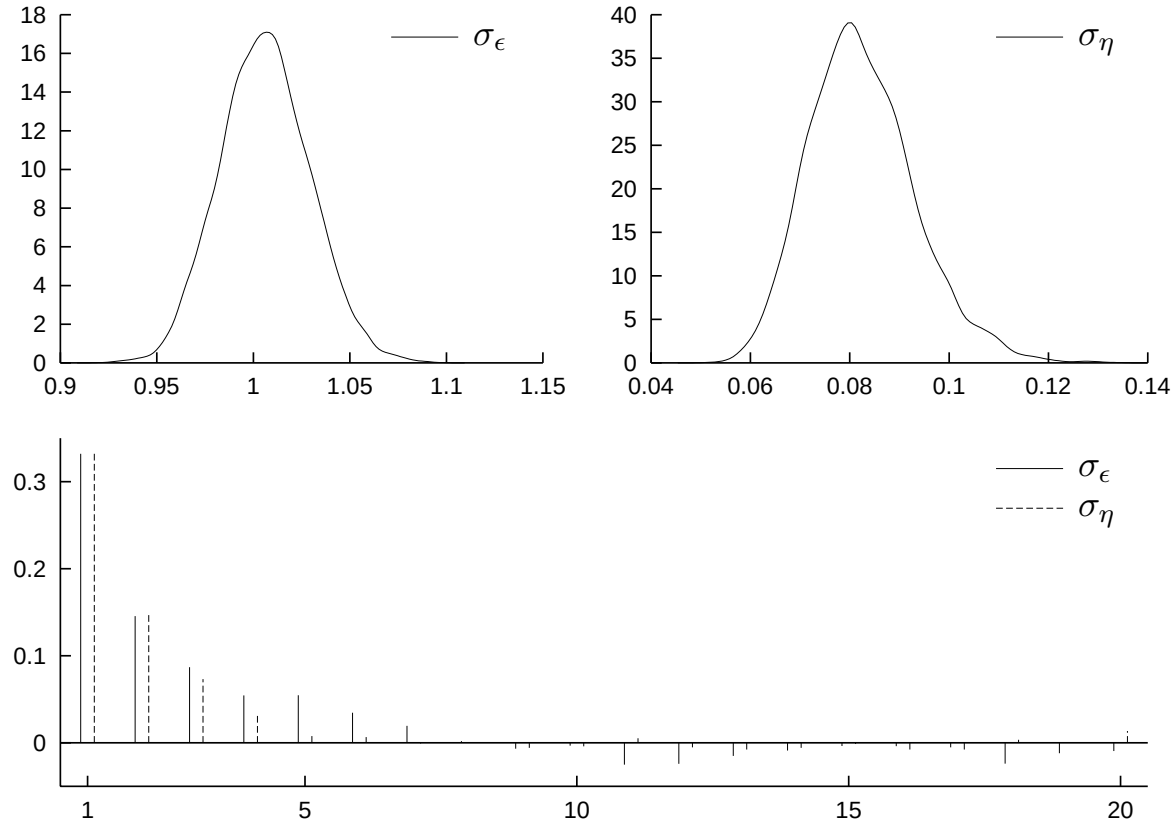

Figure 4.7: Posterior from Metropolis-Hastings sampler, with data from the Local Level model, $T = 1000$

When the sample size is enlarged to include $T = 1000$ observations (these data are plotted in the middle panel of figure 4.5), more information on the value of the parameters is available in the data. The posterior, this time applying a Metropolis-Hastings sampler, is given in figure 4.7. The posterior densities of both σ_ϵ and σ_η are indeed much more concentrated than in figure 4.6, where a data set of size $T = 100$ was used with the importance sampler. Note that the Metropolis-Hastings sampler leads to a correlated sample from the posterior density, in contrast the importance sampling algorithm. The bottom panel in the figure shows the correlation in the sample, which in this case is not large. The acceptance rate of the MH sampler was above 75%, implying that the Markov chain moved to a new location most of the time.

The adaptive polar sampling algorithm of section 4.3.4 can be applied instead of the Metropolis-Hastings sampler. For $T = 1000$ this algorithm was used in gathering 1000 accepted directions, with 10 drawings in each direction. This leads to a sample of a size comparable to the 10000 (accepted) drawings of the IS and MH samplers. The correlation of the sample from the APS algorithm is smaller than for the MH-based sample, as can be seen from figure 4.8 (drawn on the same scale as the correlation plot in figure 4.7). The

[16]The posteriors are constructed applying a kernel approximation to a histogram of the drawings.

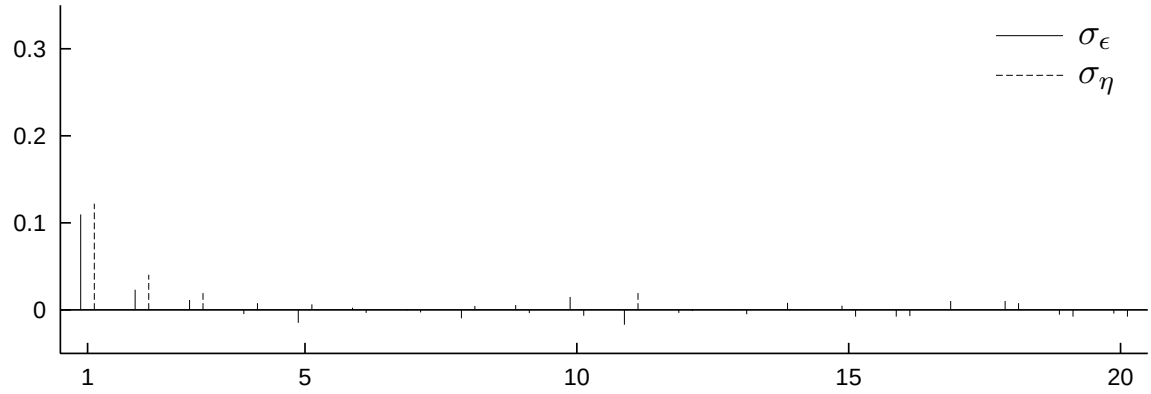

Figure 4.8: Correlation of drawings from the Adaptive Polar sampler, with data from the Local Level model, $T = 1000$

posterior density of the parameters is, of course, equal to the posterior of the Metropolis-Hastings sampler. Note however that the strongest correlation of the APS algorithm is not in the space of the parameters themselves, but in the space of the directions η (see section 4.3.4).

4.5.4 Sampling methods based on the conditional densities

For the Gibbs sampler (see section 4.2.5) the full conditional densities are needed. If we take the model with Student-t disturbances,[17] the augmented likelihood combined with the prior gives the following posterior kernel:

$$p(\sigma_\epsilon, \sigma_\eta, z, \nu, \mu \,|\, S) \propto \mathcal{L}(S; \sigma_\epsilon, \sigma_\eta, z, \nu, \mu)\pi(\sigma_\epsilon, \sigma_\eta, z, \nu) \propto$$
$$\left[\prod_{t=1}^{T}(\sigma_\epsilon z_t)^{-1} \exp\left(-\frac{1}{2}\frac{(s_t-\mu_t)^2}{\sigma_\epsilon^2 z_t^2}\right) z_t^{-\left(2\frac{\nu}{2}+1\right)} \exp\left(-\frac{\nu-2}{2z_t^2}\right)\right]$$
$$\times \sigma_\epsilon^{-(2\alpha_\epsilon+1)} \exp\left(-\frac{1}{\sigma_\epsilon^2\beta_\epsilon}\right) \times \sigma_\eta^{-(2\alpha_\eta+1)} \exp\left(-\frac{1}{\sigma_\eta^2\beta_\eta}\right) \times \frac{1}{1+\nu^2}. \quad (4.32)$$

The full conditional densities are proportional to the joint posterior, fixing the conditioning parameters at their respective values. Each of the full conditional densities has to be

[17] For this model, using the Metropolis-Hastings or the importance sampler is not feasible, as the posterior density is not analytically known without data augmentation. With data augmentation, the parameter vector of the model includes T elements μ_t and another T variance factors z_t. The resulting parameter vector is too large to be handled using MH or IS algorithms. The Gibbs sampler on the other hand can be used on the model with normally distributed disturbances. The sampler is constructed following the lines exposed here, with elements z_t fixed at 1.

derived before the Gibbs sampler can be used. For the parameter σ_ϵ this gives

$$p(\sigma_\epsilon \,|\, \sigma_\eta, z, \nu, \mu, S) \propto p(\sigma_\epsilon, \sigma_\eta, z, \nu, \mu \,|\, S) \propto$$
$$\sigma_\epsilon^{-(T+2\alpha_\epsilon+1)} \exp\left(-\frac{1}{\sigma_\epsilon^2}\left(\sum \frac{(s_t-\mu_t)^2}{2z_t^2} + \frac{1}{\beta_\epsilon}\right)\right)$$
$$\sim \text{IG-1}\left(\alpha = \frac{T}{2} + \alpha_\epsilon, \beta = \left(\sum \frac{(s_t-\mu_t)^2}{2z_t^2} + \frac{1}{\beta_\epsilon}\right)^{-1}\right). \quad (4.33)$$

Likewise, the conditional density of the standard deviation of the transition equation σ_η can be shown to be

$$p(\sigma_\eta \,|\, \sigma_\epsilon, z, \nu, \mu, S) \sim \text{IG-1}\left(\alpha = \frac{T}{2} + \alpha_\eta, \beta = \left(\sum \frac{(\mu_t-\mu_{t-1})^2}{2} + \frac{1}{\beta_\eta}\right)^{-1}\right). \quad (4.34)$$

The variance factors z_t can be sampled one-at-a-time, using

$$p(z_t \,|\, \sigma_\epsilon, \sigma_\eta, \nu, \mu, S) \propto z_t^{-(\nu+2)} \exp\left(-\frac{1}{z_t^2}\left(\frac{(s_t-\mu_t)^2}{2\sigma_\epsilon^2} + \frac{\nu-2}{2}\right)\right)$$
$$\sim \text{IG-1}\left(\alpha = \frac{\nu+1}{2}, \beta = \left(\frac{(s_t-\mu_t)^2}{2\sigma_\epsilon^2} + \frac{(\nu-2)}{2}\right)^{-1}\right). \quad (4.35)$$

The last parameter we need to sample is the degrees-of-freedom parameter ν. Collecting the factors in (4.32) containing ν we find

$$p(\nu \,|\, \sigma_\epsilon, \sigma_\eta, z, \mu, S) \propto \frac{1}{1+\nu^2} \times \prod_t \text{IG-1}\left(z_t; \alpha = \frac{\nu}{2}, \beta = \frac{2}{\nu-2}\right). \quad (4.36)$$

This conditional density of ν is not of a known functional form for which a direct sampling method would be possible. Instead, we can use a Metropolis-within-Gibbs step (section 4.3.1). Assume $\nu^{(i)}$ is the last degrees-of-freedom parameter sampled. Draw a candidate $\nu^* \sim \mathcal{N}(\nu^{(i)}, S_\nu)$, and calculate

$$\alpha_{\text{MH}} = \min\left[\frac{p(\nu^*|..)q(\nu^{(i)})}{p(\nu^{(i)}|..)q(\nu^*)}, 1\right]. \quad (4.37)$$

Note that the integrating constant in equation (4.36) is not needed, as it cancels from the Metropolis-Hastings acceptance probability in (4.37) anyhow. The parameter S_ν can be used to calibrate the Metropolis step: A small value leads to a high acceptance probability but slow mixing (or, put differently, strong correlation) between successive values of ν. A larger value for S_ν can lead to lower acceptance rates, but less correlation if a sufficient number of ν^* is accepted. A different solution, leading to less correlation but a higher computational burden, is to numerically construct the cumulative distribution function and use the Griddy Gibbs sampler for ν (see section 4.3.3).

Having found the conditional densities of the standard parameters $\sigma_\epsilon, \sigma_\eta$ and ν, and the variance factors z_t which were introduced to attain a conditionally Gaussian model, we are

left with the state parameter μ_t which also have to be sampled. One of the first available methods for sampling the states is presented in Carter and Kohn (1994). In this paper, a method is described to sample from the conditional density $p(\mu_t|y_t, \mu_{t-1}, \mu_{t+1}, \sigma_\epsilon, \sigma_\eta, z_t)$. Note that the past (future) state contains all possible information about the past (future) of the series; no extra information is contained in the observations y_{t-1} or y_{t+1}. With an uninformative prior density the posterior conditional density is normal,[18] such that is not difficult to sample successively $\sigma_\epsilon, \sigma_\eta, \mu_1, \mu_2, .., \mu_T$ from the full conditional densities.

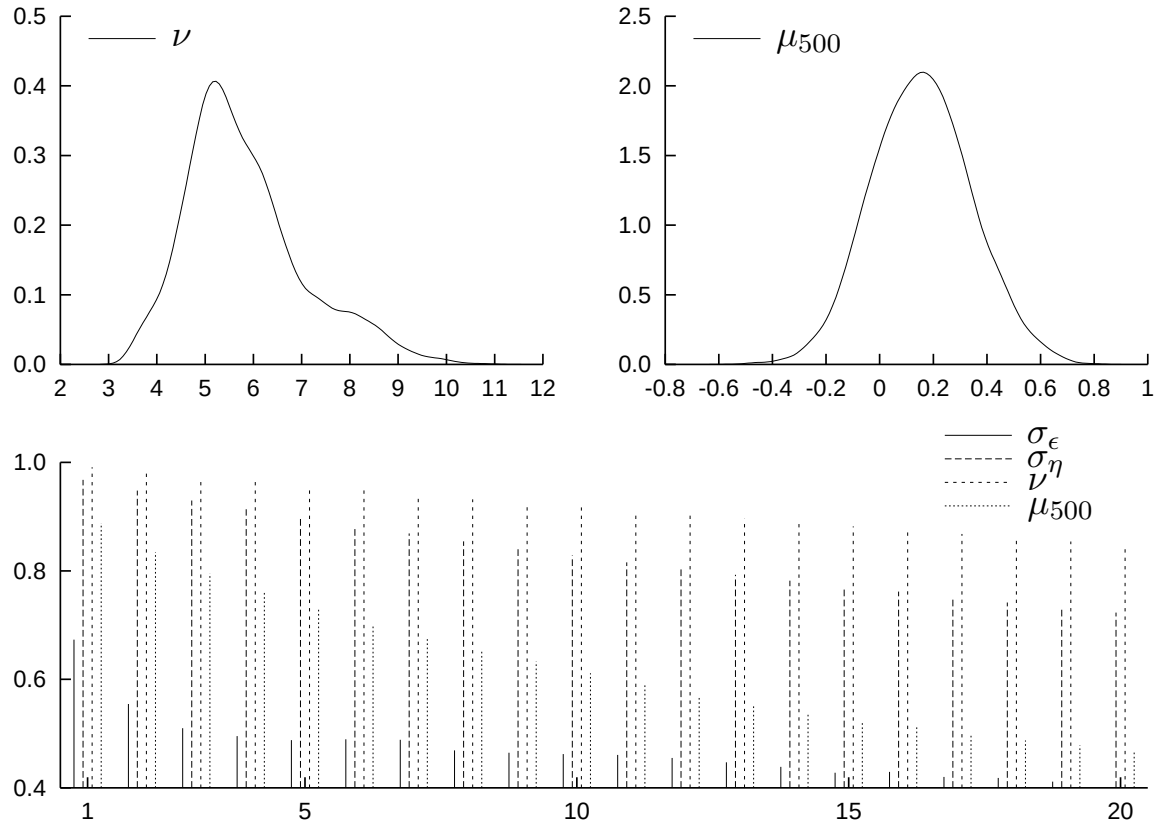

Figure 4.9: Simulating from a Local Level-Student-t model, with the posteriors of ν and μ_{500} and the correlations

The data set used in this example with the Gibbs sampler is displayed in the right panel of figure 4.5. The data was generated from the Local Level model, this time with Student-t disturbances ϵ_t, with $\nu = 4$ degrees of freedom, and is of length $T = 1000$. Using the Gibbs algorithm a sample of size 10000 was generated from the posterior. The posterior distributions of σ_ϵ and σ_η are similar to the ones found using the Metropolis-Hastings sampler, in figure 4.7. The first two panels of figure 4.9 show the posterior of the new parameter ν and of the values sampled for the state element μ_{500}, halfway the sampled data set. These two parameters (and effectively the whole state vector $\mu_1, .., \mu_T$) exhibit far stronger autocorrelation than found with the MH sampler, compare the autocorrelation in the top right panel of figure 4.9 with the correlations in figure 4.7.

[18] See appendix 4.A for the derivation.

For the parameter ν, the problem of strong autocorrelation is inherent to the setup of the sampling scheme, and it is connected to the size of the data set at hand. This can be explained as follows. We iteratively sample from the conditional distributions of $z_t|\nu$ $(t = 1, .., T)$ and $\nu|z_1, .., z_T$. When sampling a new set of z_t's there is enough freedom for each of the z's to move away from the previously sampled value, without much correlation. However, when the sample size T is large, this vector of z_t's contains rather precise information on the values of ν originally used in sampling them. Therefore, a very tight conditional density for $\nu|z_1, .., z_T$ is found for sampling a new value of the degrees-of-freedom parameter ν. Small steps and strong correlation in ν are the results. To lower the correlation, several steps can be taken. Three options come to mind:

i. A first option is to improve on the sampling of ν: As the normal candidate used in the MH step for sampling ν leads to an acceptance rate of around 50%, an improved sampling method with lower (preferably no) rejection can also reduce correlation in ν. If the Griddy Gibbs sampler is used, a direct drawing from the approximate conditional distribution can be made, thus lowering correlation due to rejected values for ν.

ii. A second, and probably more effective measure would be to find a sampling procedure for drawing a new value of ν and a new vector of z's jointly, conditional on the other parameters in the model. In general, it is always better if possible to draw parameters which are strongly correlated together from their joint distribution, otherwise the correlation between the parameters reflects itself in a strong correlation between successive drawings in the Gibbs chain.

iii. As it is in this case not simple to sample from the joint density as mentioned in ii, we can approximate the joint density using a (short) Gibbs chain. If we iterate a number G of times between drawing a value of $\nu|z^{(g-1)}, \sigma_\epsilon, \sigma_\eta$ and $z_1, .., z_T|\nu^{(g)}, \sigma_\epsilon, \sigma_\eta$ with $g = 1, .., G$, the final drawing $\nu^{(G)}, z^{(G)}$ is approximately drawn from the joint distribution of ν and z, conditional on the (fixed) other parameters. In general correlation in a Gibbs chain can be lowered by updating highly correlated parameters more often than others (Zeger and Karim 1991).

The strong correlation found between successive drawings of the state elements μ_t is not surprising. Conditional on the state-of-affairs of the previous period, μ_{t-1}, and the next period μ_{t+1}, the possible values for the present period state μ_t are narrowly defined. Therefore, also the μ_t's are strongly correlated over time, and as a consequence also between successive iterations in the Gibbs chain. As explained in ii, it is better to simulate elements which are strongly correlated together, from their joint distribution, conditional on the other parameters. This is exactly what the simulation smoother (De Jong and Shephard 1995, see appendix 4.A for the derivation) does. In this way, the number of conditional densities used in sampling is reduced greatly, and correlation between subsequent iterations of drawings of μ_t almost disappears.

Figure 4.10 shows the remaining correlation when the improvements mentioned above are applied to the sampling algorithm. More precisely, as improvements we sampled the values for $\nu|z, \sigma_\epsilon, \sigma_\eta$ using the Griddy Gibbs sampler; the sampling of $\nu|z, \sigma_\epsilon, \sigma_\eta$ and

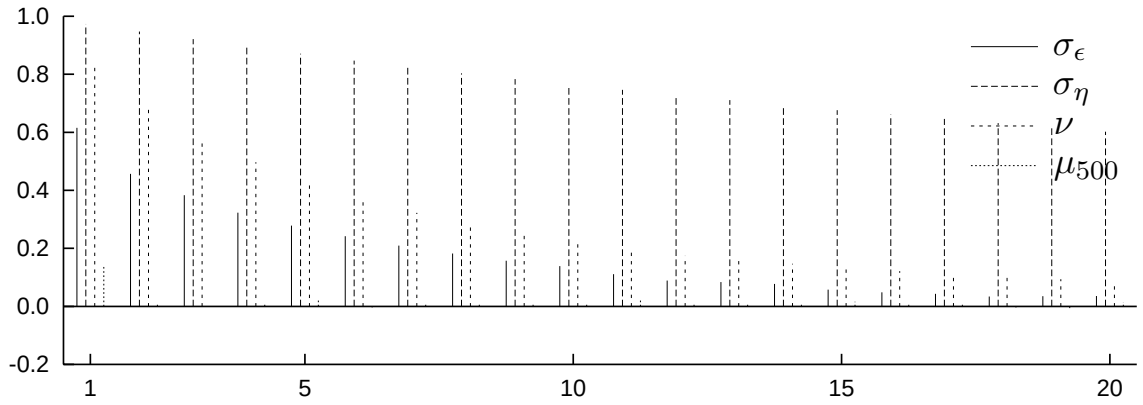

Figure 4.10: Correlation in the sample when using the improvements for the Gibbs sampler on the Local Level-Student-t model

$z|\nu, \sigma_\epsilon, \sigma_\eta$ is repeated $G = 10$ times within each iteration of the Gibbs sampler, continuing with the last value of ν and the vector z. For μ, the simulation smoother is used. Indeed, it is seen that there is virtually no correlation left between drawings of μ_{500}. The correlation in ν is reduced greatly, and also the correlation in σ_ϵ is not large. Only σ_η displays a correlation of 0.6 even at lag 20. This correlation is a result of the choice of parameters: As $\sigma_\eta \ll \sigma_\epsilon$, little information about the mean process and its variance is available in the data set. Again, the sampling of σ_η and of μ could be repeated to further reduce correlation in σ_η as well.

4.5.5 Calculating the marginal likelihood of the model

In section 4.4.2 a range of methods for calculating the marginal likelihood of a model was presented. In the present section the methods are applied, using the data set of length $T = 1000$ from the local level model. A first series of methods consisted in computing a weighted average of likelihood values. Section 4.4.2 started with a formula for m_{Prior}, where parameter vectors θ are sampled from the prior density, taking the average of the corresponding likelihood values. Table 4.2, in its upper left cell, reports a value of -1464.99 for the logarithm of the marginal likelihood (henceforth indicated by log-m). This value is calculated using a sample of 13112 drawings from the prior density.[19] To check the influence of the size of the sample, results were replicated 50 times for sizes of $n = 1000$. The second and third rows in table 4.2 (and in subsequent tables in this section) report the mean and standard deviation of the 50 log-m values calculated.

From these first results we see that marginal likelihood values can easily become extreme: The marginal likelihood values themselves, without taking logarithms, are of size $\exp(-1465) \approx 0$ for this combination of model and data set. The log-m values depend linearly on the number of observations in the data set, exactly like the standard log-likelihood is of order T. As a consequence, also the logarithms of the posterior odds depend linearly on T: For large data sets, small per-observation evidence in favour of a model

[19]The odd size is chosen to equal the size of the sample from the Metropolis-Hastings algorithm, applied later in this section.

may well lead to very large or extremely small posterior odds. We will see an example of this later in this section.

After this digression, let us return to the calculation of the log-m values. Instead of sampling from the prior, the posterior sample can be applied, leading to the harmonic mean estimator log-m_{HM}, reported in the second column in table 4.2. Care has to be taken in the computation of this and other marginal likelihood measures, as one easily ends up with $\pm\infty$ or 0. The instability of the HM estimator is reflected in the larger standard deviation of the 50 repetitions of the small-sample computations. The difference in outcomes of log-m_{HM} and log-m_{Prior} is 1.3, implying that the marginal likelihood according to the HM estimator is $\exp(1.3) \approx 3.7$ times larger than the corresponding outcome using the prior estimator.

An intermediate position is taken by the measure m_{IS}, sampling parameter vectors from an importance function other than the prior or posterior density. The third column of table 4.2 reports the outcomes applying a Student-t importance density with $\nu = 4$ degrees-of-freedom, with mode and covariance fitted to the posterior mode. Even though this importance density has heavier tails than the posterior (needed for improving the convergence behaviour of m over the behaviour of m_{HM}), we do not seem to gain much stability; the (average) value of log-m falls in between log-m_{Prior} and log-m_{HM}, but the variability (in small samples) is 2-3 times larger. We will not pursue the search for a better importance sampling distribution at this time.

Table 4.2: Logarithms of marginal likelihoods, calculated using the prior, harmonic mean and importance sampling estimators, log-m_{Prior}, log-m_{HM} and log-m_{IS}

	log-m_{Prior}	log-m_{HM}	log-m_{IS}
log-m	−1464.99	−1463.71	−1464.43
μ(log-m)	−1465.04	−1463.68	−1464.25
σ(log-m)	0.202	0.295	0.760

Note: Reported are the log-m based on a sample of size $n = 13112$. μ(log-m) and σ(log-m) are the mean and standard deviation of 50 replications of the computation using sample size $n = 1000$.

With a LaPlace approximation to the posterior kernel, the location at which the log-m is computed matters. Table 4.3 reports in the columns the log-m_{LP} evaluated at the mode of the posterior and the location of maximum likelihood, both found by optimizing the corresponding function, and at the mean and median of the posterior sample of size $n = 13112$ from the MH sampling algorithm. The results correspond closely, as the posterior is symmetric, and the prior not very informative. As the locations of the mode and of the maximum likelihood parameter vector are found without using the sample, no variation over the smaller samples[20] occurs (see the third row in the table). The median is more robust as an estimator of the location, therefore it leads to lower variation in the log-m value than using the mean as a location estimator.

[20]The smaller samples of size $n = 1000$ were drawn randomly-with-replacement from the sample of size $n = 13112$ resulting from the MH sampling algorithm.

Instead of using a normal approximation to the posterior density, a kernel smoothing density can be fit to the posterior sample.[21] Again, at four locations the log-m_{Kern} was computed, with the results reported in table 4.4. Notice that the results for log-m_{Kern} correspond closely to the results found in log-m_{LP}. The variation in the results for the small samples is larger with log-m_{Kern}, as the kernel density smoother needs a rather large sample size to get a reasonable precision.

Table 4.3: Marginal likelihoods calculated using the LaPlace approximation

	Mode	ML	Mean	Median
log-m_{LP}	−1465.05	−1466.20	−1465.02	−1465.02
μ(log-m)	〃	〃	−1465.03	−1465.02
σ(log-m)	0	0	0.044	0.005

Note: See table 4.2 for an explanation of the entries in the table.

Table 4.4: Marginal likelihoods calculated using the kernel approximation

	Mode	ML	Mean	Median
log-m_{Kern}	−1465.01	−1465.08	−1465.03	−1465.03
μ(log-m)	−1464.89	−1465.04	−1464.92	−1464.92
σ(log-m)	0.087	0.140	0.080	0.083

Note: See table 4.2 for an explanation of the entries in the table.

The marginal likelihood under the hypothesis that the model is the Local Level model with Student-t observation disturbances is not so easily calculated, as the likelihood function and the posterior kernel cannot be evaluated without augmenting the parameter vector to a high dimension where the methods applied before loose all accuracy. Therefore, we switch to the method of Chib (1995), to compute the marginal likelihood from the output of the Gibbs sampler (see the second part of section 4.4.2).

Let $\tilde{\sigma}_\eta, \tilde{\sigma}_\epsilon$ and $\tilde{\nu}$ be the mean of the posterior density (obtained from the sample from the posterior), enlarged with a vector $\tilde{z}^{[j]}$. We calculate the average marginal likelihood over a range of q different vectors $\tilde{z}^{[j]}$, according to

$$m_{\text{Gibbs}} = \frac{1}{q}\sum_{j=1}^{q} \frac{\mathcal{L}(S;\tilde{\sigma}_\eta,\tilde{\sigma}_\epsilon,\tilde{\nu},\tilde{z}^{[j]})\pi(\tilde{\sigma}_\eta,\tilde{\sigma}_\epsilon,\tilde{\nu},\tilde{z}^{[j]})}{p(\tilde{\sigma}_\eta,\tilde{\sigma}_\epsilon,\tilde{\nu},\tilde{z}^{[j]}|S)}. \tag{4.38}$$

For the denominator, we decompose

$$p(\tilde{\sigma}_\eta,\tilde{\sigma}_\epsilon,\tilde{\nu},\tilde{z}^{[j]}|S) = p(\tilde{\sigma}_\eta|\tilde{\sigma}_\epsilon,\tilde{\nu},\tilde{z}^{[j]},S)p(\tilde{\sigma}_\epsilon|\tilde{\nu},\tilde{z}^{[j]},S)p(\tilde{\nu}|\tilde{z}^{[j]},S)p(\tilde{z}^{[j]}|S). \tag{4.39}$$

For each of the conditional densities in (4.39) a separate Gibbs chain is run, estimating

[21] We apply a Gaussian kernel with automatic bandwidth selection according to Silverman's (1986) rule.

i. $p(\tilde{\sigma}_\eta|\tilde{\sigma}_\epsilon, \tilde{\nu}, \tilde{z}^{[j]}, S) \approx \frac{1}{n}\sum p(\tilde{\sigma}_\eta|\mu^{(i)}, \tilde{\sigma}_\epsilon, \tilde{\nu}, \tilde{z}^{[j]}, S)$ sampling n elements $\mu^{(i)}$ and $\sigma_\eta^{(i)}$ from the chain

$$\begin{aligned}\mu^{(i)} &\sim \mu|\sigma_\eta^{(i-1)}, \tilde{\sigma}_\epsilon, \tilde{\nu}, \tilde{z}^{[j]}, S,\\ \sigma_\eta^{(i)} &\sim \sigma_\eta|\mu^{(i)}, \tilde{\sigma}_\epsilon, \tilde{\nu}, \tilde{z}^{[j]}, S;\end{aligned}$$

Note that we need to sample $\sigma_\eta^{(i)}$ here alongside with $\mu^{(i)}$. This way, the sampled $\mu^{(i)}$ come from the conditional density $\mu|\tilde{\sigma}_\epsilon, \tilde{\nu}, \tilde{z}^{[j]}, S$, marginal of the value of σ_η.

ii. $p(\tilde{\sigma}_\epsilon|\tilde{\nu}, S) \approx \frac{1}{n}\sum p(\tilde{\sigma}_\epsilon|\mu^{(i)}, \sigma_\eta^{(i)}, \tilde{\nu}, \tilde{z}^{[j]}, S)$ sampling n elements $\mu^{(i)}, \sigma_\eta^{(i)}$ and $\sigma_\epsilon^{(i)}$ from the chain

$$\begin{aligned}\mu^{(i)} &\sim \mu|\sigma_\eta^{(i-1)}, \sigma_\epsilon^{(i-1)}, \tilde{\nu}, \tilde{z}^{[j]}, S,\\ \sigma_\eta^{(i)} &\sim \sigma_\eta|\mu^{(i)}, \sigma_\epsilon^{(i-1)}, \tilde{\nu}, \tilde{z}^{[j]}, S,\\ \sigma_\epsilon^{(i)} &\sim \sigma_\epsilon|\mu^{(i)}, \sigma_\eta^{(i)}, \tilde{\nu}, \tilde{z}^{[j]}, S;\end{aligned}$$

iii. $p(\tilde{\nu}|\tilde{z}^{[j]}, S) \approx \frac{1}{n}\sum p(\tilde{\nu}|\mu^{(i)}, \sigma_\eta^{(i)}, \sigma_\epsilon^{(i)}, \tilde{z}^{[j]}, S)$ sampling n elements $\mu^{(i)}, \sigma_\eta^{(i)}, \sigma_\epsilon^{(i)}$ and $\nu^{(i)}$ from the chain

$$\begin{aligned}\mu^{(i)} &\sim \mu|\sigma_\epsilon^{(i-1)}, \sigma_\eta^{(i-1)}, \nu^{(i-1)}, \tilde{z}^{[j]}, S,\\ \sigma_\eta^{(i)} &\sim \sigma_\eta|\mu^{(i)}, \sigma_\epsilon^{(i-1)}, \nu^{(i-1)}, \tilde{z}^{[j]}, S,\\ \sigma_\epsilon^{(i)} &\sim \sigma_\epsilon|\mu^{(i)}, \sigma_\eta^{(i)}, \nu^{(i-1)}, \tilde{z}^{[j]}, S,\\ \nu^{(i)} &\sim \nu|\mu^{(i)}, \sigma_\eta^{(i)}, \sigma_\epsilon^{(i)}, \tilde{z}^{[j]}, S;\end{aligned}$$

iv. $p(\tilde{z}^{[j]}|S) \approx \frac{1}{n}\sum p(\tilde{z}^{[j]}|\mu^{(i)}, \sigma_\eta^{(i)}, \sigma_\epsilon^{(i)}, \nu^{(i)}, S)$ sampling n elements $\mu^{(i)}, \sigma_\eta^{(i)}, \sigma_\epsilon^{(i)}, \nu^{(i)}$ and $z^{(i)}$ from the chain

$$\begin{aligned}\mu^{(i)} &\sim \mu|\sigma_\epsilon^{(i-1)}, \sigma_\eta^{(i-1)}, \nu^{(i-1)}, \tilde{z}^{(i)}, S,\\ \sigma_\eta^{(i)} &\sim \sigma_\eta|\mu^{(i)}, \sigma_\epsilon^{(i-1)}, \nu^{(i-1)}, \tilde{z}^{(i)}, S,\\ \sigma_\epsilon^{(i)} &\sim \sigma_\epsilon|\mu^{(i)}, \sigma_\eta^{(i)}, \nu^{(i-1)}, \tilde{z}^{(i)}, S,\\ \nu^{(i)} &\sim \nu|\mu^{(i)}, \sigma_\eta^{(i)}, \sigma_\epsilon^{(i)}, \tilde{z}^{(i)}, S,\\ z^{(i)} &\sim z|\mu^{(i)}, \sigma_\eta^{(i)}, \sigma_\epsilon^{(i)}, \nu^{(i)}, S.\end{aligned}$$

The numerator of the marginal likelihood equation (4.38) can be calculated without further detours. With $\tilde{z}$ known the model is Gaussian and the Kalman equations can be used to compute the likelihood through the prediction-error decomposition (see also appendix 4.A). The functional form of the prior is given, and therefore its computation does not pose any problem.

In table 4.5 the results are given of applying the algorithm to both the data set with normal disturbances (in the first two columns) as to the data set with the heavy-tailed disturbance distribution (in columns three and four). For each data set results are calculated assuming that the Local Level model is correct (in columns 1 and 3, skipping

Table 4.5: Marginal likelihoods calculated using the Gibbs' conditional densities algorithm

Data	Local Level		Local Level-Student-t	
Model	LL	LLS	LL	LLS
log-m_{Gibbs}	−1465.00	−1468.83	−1417.51	−1391.49
μ(log-m)	−1465.01	−1469.79	−1417.53	−1393.66
σ(log-m)	0.186	0.959	0.227	2.687

Note: The value for log-m_{Gibbs} was calculated using (4.38) with $q = 10$ and chain length $n = 10000$. The bottom rows are calculated as the mean and standard deviation of 100 replications of (4.38) with $q = 1, n = 1000$.

the sampling of elements ν and z in the algorithm above) or making the assumption that Student-t disturbances are part of the model (indicated by the heading LLS). The first row contains results for the Gibbs output when each subsampler was run for $n = 10000$ iterations (see below equation (4.39)), evaluating the log-m at the mean of the posterior sample. When Student-t disturbances are introduced, a value of z for conditioning upon is sampled from the full conditional posterior density $\tilde{z} \sim p(z|\tilde{\theta}, Y)$, and the calculation is repeated $q = 10$ times in equation (4.23″). For the comparison of the stability of the marginal likelihood estimate, the computation was repeated $n = 100$ times for subsampler chain lengths of $n = 1000$, with reported μ(log-m) and σ(log-m) as the results.

For the LL data the log-m of the LL model is -1465.00, close to the previous results. When this result is compared to the outcome assuming the LLS model, of -1468.83, we see that the data favours the LL model. The difference is not very large, even though the LLS model has an extra parameter for fitting the data. For the second data set (columns three and four) we see that the data overwhelmingly favours the LLS model: The logarithm of the Bayes factor equals the difference in log-m's. This difference is approximately 26, indicating the the Local Level-Student-t model is deemed exp(26) times more plausible than the simpler Local Level model.

From the standard deviations σ(log-m) we see that the marginal likelihood calculation using the conditional densities method is not very stable. With larger sample sizes in the subsamplers stability is better.

4.5.6 Sampling setup and details

Throughout the example this section, all computations have been performed using programs written by the author, in Ox 2.20 (Doornik 1999). Extensive use was made of the SsfPack package by Koopman, Shephard and Doornik (1999). This section describes the initializations and other details of the setup that were used in these programs.

Importance sampling

As an importance density, a Student-t with $\nu = 4$ degrees-of-freedom was used. Four degrees of freedom ensure that the tails of the importance density are heavier than the tails of the target density at hand. As a first start, a maximization routine is used to find the mode of the posterior density. The Student-t density is transformed such that its mode

corresponds to the mode of the posterior; for the covariance, a numerical approximation to the covariance of the posterior around the mode is used.

Table 4.6: Sampling setup and results

						Acceptance			
Method	DGP	T	μ	ν	Burn-in	rate	Total size	max($\lvert\rho_{20}\rvert$)	Time
IS	$\mathcal{N}$	100			-	1	10000	0	4
IS	$\mathcal{N}$	1000			-	1	10000	0	18
MH	$\mathcal{N}$	1000			1000	0.76	13112	0.014	24
APS	$\mathcal{N}$	1000			100	0.96	10460	0.009	3:51
Gibbs	t	1000	CPS	$\mathcal{N}$	1000	0.55	10000	0.87	3:16
Gibbs	t	1000	SS	GG-10	1000	0.99	10000	0.50	59:25
log-m	$\mathcal{N}$	1000					13112		1:26
log-m_{Gibbs}	$\mathcal{N}$	1000	SS	-	0		10000		1:01
log-m_{Gibbs}	t	1000	SS	-	0		10000		1:01
log-m_{Gibbs}	$\mathcal{N}$	1000	SS	$\mathcal{N}$-10	0		10000		22:30
log-m_{Gibbs}	t	1000	SS	$\mathcal{N}$-10	0		10000		22:25

Note: For the sampling and log-m-computation methods this table reports the acceptance rate, total sample size, maximum correlation at lag 20 and the duration of the compuation, indicating in the first columns the DGP, length of the data set, the method for simulating μ (using the method of Carlin, Polson and Stoffer (1992) or the Simulation Smoother), the sampling candidate for ν (normal or with a Griddy Gibbs candidate), and the number of replications of sampling z and ν (either 1 or 10).

With the location and the scale parameters of the candidate set up, a first round of the importance sampler is run, collecting 1000 parameter vectors with corresponding weights. These are used to (re)estimate the mean and covariance of the posterior, and the importance density is adapted accordingly. The sampling and adapting of the location is done for two more rounds, of lengths of 2000 and 5000 drawings from the parameter vector, respectively. Then, in a final round, 10000 drawings from the importance density are gathered and used with their weights in the example.

Table 4.6 reports some key statistics and the computing time. The first line covers the importance sampling (column 1) on the model with normal disturbances (second column), of length $T = 100$. A burn-in period is not necessary, as the correlation in the sample (column 9) is zero by construction. All drawings are accepted (column 7). The time needed for this computation is a mere 4 seconds, using Ox 2.20 (see Doornik 1999) on a computer running a 900 MHz AMD processor using Linux. A similar calculation with a data set of length $T = 1000$ took 18 seconds, as reported in the second row of the table.

Metropolis-Hastings sampling

The setup for the candidate density was the same as described above for the importance density in the IS algorithm, with the difference that sampling was continued until a set of $n = 1000, 2000, 5000$ or 10000 accepted drawings were collected. As there is correlation in the chain, we allowed for 1000 (accepted) drawings as a burn-in period before collecting the final sample. The final total sample size was 13112 vectors large, corresponding to

an acceptance rate of 0.76. The maximum correlation after 20 lags was found for the σ_η parameter, and was of size $\rho_{20} = 0.014$. Essentially the MH sampling algorithm needs one posterior density evaluation for each drawing; as the final sample is 30% larger than for the IS algorithm, the computating time is also about 30% longer.

Adaptive polar sampling

The setting was mainly the same as for the MH sampling above. The only extra choice we have for the APS algorithm is the number of directions to sample. Results reported here use $n_\eta = 50, 100, 250$ and 1000 different accepted directions η, leading to a sample of $n/n_\eta = 20, 20, 20$ or 10 drawn parameter vectors in each direction throughout the 4 rounds. In the first round, the Metropolis-Hastings step for accepting or rejecting a new candidate was skipped; as a consequence, the first round does not give a sample from the correct posterior distribution, but does help to quickly improve on the location and scale parameters. Later rounds run the full-blown APS algorithm, resulting in a sample from the correct posterior density. As a burn-in period, 100 accepted directions were disregarded in the final sample before starting to collect the drawings.

Gibbs sampling

The initialization of the Gibbs sampler consisted in finding starting values for $\sigma_\epsilon, \sigma_\eta, \nu$ and z. The two standard deviations are initialized at the optimal values when searching for the mode of the posterior density of the Local Level model, neglecting the effect of the heavy tailed disturbance density. Likewise, the vector of z, multiplying the standard deviation σ_ϵ at each time period, is initially filled with ones, also corresponding to a normal disturbance density. The initial value of ν was chosen to be 4, the value used in generating the data with the Student-t disturbances.

In the initial Gibbs sampler, new values of ν were drawn from the full conditional density using a Metropolis-within-Gibbs step (see section 4.3.1). The candidate density for ν was normal with a mean equal to the previously drawn $\nu^{(i-1)}$ and a standard deviation of 0.3, resulting in a Random Walk Metropolis chain, see section 4.2.4. The standard deviation was fixed at this value after some calibration, to give a sufficient number of accepted drawings with a step size being not too small. This choice of candidate gave an acceptance rate for ν of 0.55 as reported table 4.6.

The sampler continued first for 1000 iterations to get rid of dependence on the starting conditions, and then for 10000 iterations more. When sampling from the Local Level-Student-t model using data generated from the model with normal disturbances, the parameter ν drifted away to larger and larger values. Theoretically, it should 'converge' to $\nu = \infty$, as the disturbances were normally distributed. Therefore, in this case the sampler did not (nor would it ever) converge, as there is no way that it could. However, within the 10000 sampled parameter vectors it became very clear that ν was drifting away indeed. The parameters σ_ϵ and σ_η were sampled from their correct distributions, as the Student-t density with any value $\nu > 20$ is hardly distinguishable from the normal density.

As the correlation in the sample was high, three measures were taken to improve the mixing of the sampler. In the table these are indicated by the letters SS, for Simulation Smoother, in the column indicating the sampling method of μ, and by the GG-10

in the next column, indicating that we use the Griddy Gibbs sampler on ν, repeating the sampling of z and ν 10 times within each iteration of the Gibbs sampler. An adaptive integration rule is used to find the integrating constant of the density of $\nu|z$. The points visited by the integration routine are used as support points for constructing the approximate density and cumulative distribution function. In order to adjust for the approximation error in the distribution function, the sampled value ν from the Griddy Gibbs algorithm was used as candidate drawing in a MH step. The acceptance rate of 0.99 indicates that the computed density function was very accurate.

The timings of the Gibbs sampler, taking about three minutes for the method using the Carlin et al. (1992) method for simulating μ and a normal candidate density for sampling ν and one hour when using the Simulation Smoother and Griddy Gibbs sampler with internal 10-fold replication, indicate clearly the increased computational load. Note that this increase in duration is entirely caused by the numerical integration inherent in the Griddy Gibbs algorithm, especially as it is repeated 10 times at each iteration.

Marginal likelihood computations

The calculation of the logarithms of marginal likelihoods using the likelihood function, log-m_{HM}, log-m_{Prior}, log-m_{IS}, log-m_{LP} and log-m_{Kern} can be done rather quickly, they could all be computed in little more than one minute. The marginal likelihood for the model with Student-t disturbances had to be calculated using the Chib-method, which is computationally more intensive. Instead of the Griddy Gibbs method we used the normal candidate density in a Metropolis step for sampling ν, to lower the time needed. Calculating the marginal likelihood of the model without Student-t disturbances is done without a problem, in one minute. Introducing the t_ν density for the disturbances led to a duration of more than two minutes for the DGP where the heavy tails were included for each repetition. As we repeated the calculation of $(4.23'')$ $q = 10$ times, the total duration was more than 22 minutes.

As a separate Gibbs chain has to be sampled for each element of the parameter vector, computations for the log-m_{Gibbs} can easily take a very long time (in section 5.6.2, computation of the marginal likelihood for some of the models could take 12 hours). As often the differences in values of log-m between models are large, one could choose to go for less precision and take smaller sample sizes. With a sample size of $n = 1000$, computation of one repetition of $(4.23'')$ took 14 seconds. This can easily be repeated very often, to get a more precise estimate of log-m_{Gibbs}.

4.6 Concluding remarks

This chapter introduced the basic algorithms for sampling from a Bayesian posterior density. The sampling algorithms themselves are presented easily enough, but often the step from theory to implementation is harder. Therefore, the algorithms have been put in action in an exemplifying application, showing the ins and outs of the different algorithms.

Algorithms as the importance sampler, sampling-importance resampling algorithm, the acceptance-rejection sampler, the Metropolis-Hastings sampler and the adaptive polar sampling can all be used when the posterior kernel function can be evaluated easily. For

the acceptance-rejection sampler, a envelopping density kernel is needed, which in practice is often hard to find. Other methods have their advantages and their disadvantages. When correlation in the posterior sample is a problem (e.g. when the continuation of the analysis, using the posterior sample, is costly), the importance sampler or the SIR algorithm can be used. Also the APS is a good option here, as it leads to a high quality sample from the posterior density. When a good candidate density is available, and the posterior density is well-behaved (not multimodal, relatively thin tails, etc.), the Metropolis-Hastings algorithm is easily implemented and can be expected to lead to good results.

For cases where data augmentation is needed to get tractable (conditional) densities, the Gibbs sampler is the solution, with possible MH steps delivering some of the draws from the conditional densities. The combination of the Gibbs sampler with the Griddy Gibbs sampler is quite simple to implement. However, for larger dimensions of the parameter vector the correlation in the chain can become very large, as was seen in the example when we used the original Carlin et al. (1992) approach to sampling from a state space model. A clever setup of the Gibbs sampler can help in overcoming the strong correlation.

A special field of research is the computation of Bayes factors and of the marginal likelihood of a model. Several articles on the topic have been summarized here. Though the theoretical advantages of the methods are known, we do not know of an empirical comparison of the whole range of methods for computing the marginal likelihoods as in section 4.5.5. It was found that on the model used here the simpler methods based on a LaPlace or kernel smoothing approximation to the sample of the posterior density work well. The conditional densities method has to be used whenever the posterior kernel cannot be easily evaluated. Results for this method displayed larger variability. Also the implementation is considerably more difficult than the implementation of the other methods.

4.A Sampling and the state space model

Section 4.5 presents a state space model, which is used as a workhorse for explaining all kinds of different sampling methods. This appendix provides extra background on the state space model and the algorithms connected to it. Following the notation in Koopman et al. (1999) closely the Kalman filter is introduced, which lies behind the calculations of the likelihood and posterior of the model in section 4.5.3. It is followed by methods by Carlin et al. (1992) and De Jong and Shephard (1995) for sampling from the states, as they are used in section 4.5.4.

A general linear state space model can be written as

$$\alpha_{t+1} = d_t + T_t\alpha_t + H_t\epsilon_t, \qquad \alpha_1 \sim \mathcal{N}(a, P), \tag{4A.1}$$

$$y_t = c_t + Z_t\alpha_t + G_t\epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, I_r). \tag{4A.2}$$

Equation (4A.1) is known as the transition equation, whereas (4A.2) is often referred to as the observation equation. Taking both equations together in larger matrices, the system can be written as

$$\begin{aligned}
&\begin{pmatrix}\alpha_{t+1}\\ y_t\end{pmatrix} = \delta_t + \Phi_t\alpha_t + u_t, \qquad (4A.3)\\
&u_t = \begin{pmatrix}H_t\\ G_t\end{pmatrix}\epsilon_t \sim \mathcal{N}(0, \Omega_t),\\
&\delta_t = \begin{pmatrix}d_t\\ c_t\end{pmatrix}, \qquad \Phi_t = \begin{pmatrix}T_t\\ Z_t\end{pmatrix}, \qquad \Omega_t = \begin{pmatrix}H_tH_t' & H_tG_t'\\ G_tH_t' & G_tG_t'\end{pmatrix}.
\end{aligned}$$

As long as the model is conditionally Gaussian, the first and second moments are sufficient statistics for the densities in the model. The Kalman filter consists of a set of filtering equations which, based on the observations and the parameter matrices, derive the expectations and variances of $\alpha_{t+1} \,|\, Y_t$, and also the expectation and variance of the predictive density of $y_{t+1} \,|\, Y_t$. The Kalman filter equations are

$$\begin{aligned}
v_t &= y_t - c_t - Z_ta_t,\\
F_t &= Z_tP_tZ_t' + G_tG_t',\\
K_t &= (T_tP_tZ_t' + H_tG_t')F_t^{-1}, \qquad (4A.4)\\
a_{t+1} &= d_t + T_ta_t + K_tv_t,\\
P_{t+1} &= T_tP_tT_t' + H_tH_t' - K_tF_tK_t', \qquad t = 1, .., T.
\end{aligned}$$

The recursion starts using $a_1 = a$ and $P_1 = P$, the starting values for the initial state $\alpha_1 \sim \mathcal{N}(a, P)$. Output from the Kalman filter includes v_t, the prediction error at time t given all previous observations and the corresponding variance F_t. These elements can be used in a prediction-error decomposition of the likelihood of the model, with $\mathcal{L}(Y;\theta) = \prod p(y_t \,|\, Y_{t-1}, \theta) = \prod p(v_t \,|\, Y_{t-1}, \theta) = \prod f_{\mathcal{N}(0,F_t)}(\nu_t)$. This is the likelihood used in section 4.5.3 to construct a closed-form formula for the posterior density of the parameters in a state space model.

When, in the Gibbs sampler, we need to sample a state α_t conditional on the data and the other states, using Bayes' theorem we find three relations inferring information

on α_t (Carlin et al. 1992): The relation between y_t and α_t, the one with α_{t-1} and the one with α_{t+1}. If we disregard possible correlation between the disturbances in the transition equation (4A.1) and those in the observation equation (4A.2)[22], we find a posterior $P(\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1}) \propto \mathcal{L}(y_t|\alpha_t)\pi(\alpha_t|\alpha_{t-1}, \alpha_{t+1})$ which can be written as

$$\begin{aligned}-2\log P(\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1}) &= c \\ &+ (\alpha_t - d_{t-1} - T_{t-1}\alpha_{t-1})' \left(H_{t-1}H_{t-1}'\right)^{-1} (\alpha_t - d_{t-1} - T_{t-1}\alpha_{t-1}) \\ &+ (y_t - c_t - Z_t\alpha_t)' \left(G_tG_t'\right)^{-1} (y_t - c_t - Z_t\alpha_t) \\ &+ (\alpha_{t+1} - d_t - T_t\alpha_t)' \left(H_tH_t'\right)^{-1} (\alpha_{t+1} - d_t - T_t\alpha_t).\end{aligned}$$

This formula can be manipulated such that it is written in the format $(\alpha_t - B_tb_t)'B_t^{-1}(\alpha_t - B_tb_t)$, indicating that $\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1} \sim \mathcal{N}(B_tb_t, B_t)$ with

$$B_t^{-1} = \left(H_{t-1}H_{t-1}'\right)^{-1} + Z_t' \left(G_tG_t'\right)^{-1} Z_t + T_t' \left(H_tH_t'\right)^{-1} T_t, \tag{4A.5}$$

$$\begin{aligned}b_t' = &(d_{t-1} + T_{t-1}\alpha_{t-1})' \left(H_{t-1}H_{t-1}'\right)^{-1} + (y_t - c_t)' \left(G_tG_t'\right)^{-1} Z_t \\ &+ (\alpha_{t+1} - d_t)' \left(H_tH_t'\right)^{-1} T_t.\end{aligned} \tag{4A.6}$$

The example in section 4.3.2 simplifies, as we have $Z_t = T_t = 1, H_tH_t' = \sigma_\eta^2, G_tG_t' = \sigma_\epsilon^2 z_t, H_tG_t' = d_t = c_t = 0$. Filling in the matrices leads to $\mu_t|y_t, \mu_{t-1}, \mu_{t+1} \sim \mathcal{N}(B_tb_t, B_t)$ with $b_t = (\mu_{t-1} + \mu_{t+1})/\sigma_\eta^2 + y_t/\sigma_\epsilon^2 z_t$ and $B_t^{-1} = 2/\sigma_\eta^2 + 1/\sigma_\epsilon^2 z_t$.

Instead of sampling one state element at a time, the simulation smoother (Carter and Kohn 1994, Frühwirth-Schnatter 1994, De Jong and Shephard 1995) manages to sample a complete set of vectors $\tilde{u}_1, .., \tilde{u}_T|Y_T, d, \Phi, \Omega$. Any particular vector $\tilde{u}_i = \Gamma u_i$ is a selection of the disturbances u_i in equation (4A.3) ensuring that no problems with nonsingularity occur (for conditions on Γ, see Koopman et al. 1999). In order to sample $\tilde{u}$, the simulation smoother first needs the output from the Kalman filter. In the notation of Koopman et al. (1999) again, write Γ^* for the matrix constructed from Γ, skipping rows with only zeroes. Then the equations of the simulation smoother are

$$\begin{aligned}C_t &= \Gamma^* \begin{pmatrix} H_t \\ G_t \end{pmatrix} (I - G_t'F_t^{-1}G_t - J_t'N_tJ_t) \begin{pmatrix} H_t \\ G_t \end{pmatrix}' \Gamma^{*\prime}, \\ W_t &= \Gamma^* \begin{pmatrix} H_t \\ G_t \end{pmatrix} (G_t'F_t^{-1}G_t - J_t'N_tJ_t), \\ r_{t-1} &= Z_t'F_t^{-1}v_t - W_t'C_t^{-1}\xi_t + L_tr_t \\ N_{t-1} &= Z_t'F_t^{-1}Z_t + W_t'C_t^{-1}W_t + L_t'N_tL_t, \\ \tilde{u}_t &= \Gamma^{*\prime} \left(\Gamma^* \begin{pmatrix} H_t \\ G_t \end{pmatrix} (G_t'F_t^{-1}v_t + J_t'r_t) + \xi_t \right), \\ L_t &= T_t - K_tZ_t, \qquad J_t = H_t - K_tG_t, \qquad \xi_t \sim \mathcal{N}(0, C_t).\end{aligned} \tag{4A.7}$$

These equations are successively calculated in a backward recursion, $t = T, .., 1$, starting using $r_n = 0$ and $N_n = 0$. In the example, the matrix $\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is chosen to gen-

[22] In most practical situations, correlation is fixed to be zero for identification of the model. Otherwise, the correlation is easily introduced in the equations in this appendix.

erate drawings from the disturbances of the transition equation. The states μ_t can be reconstructed using the recursion in the transition equation (4A.1).

4.B Derivations and distributions for APS

In section 4.3.4 the Adaptive Polar sampler was explained. This appendix reports the exact formula's for the transformations that are applied, and the resulting distributions.

The transformation to polar coordinates

$$(\eta, \rho) = T(\theta \,|\, \mu, \Sigma) = T_{y\to\eta,\rho}(y) \circ T_{\theta\to y}(\theta \,|\, \mu, \Sigma) \tag{4.12}$$

is comprised of two parts. The standardization uses the Choleski decomposition $\Sigma^{\frac{1}{2}}$ of the covariance matrix,

$$y = T_{\theta\to y}(\theta \,|\, \mu, \Sigma) = \Sigma^{-\frac{1}{2}}(\theta - \mu). \tag{4.12a}$$

The standardized parameters y are transformed to polar coordinates ρ and η. The signed distance measure ρ is based on the length $d = \sqrt{y'y}$ of y, but also indicates the sign of the first element of y:

$$\rho = \text{sgn}(y_1)\; d. \tag{4B.8}$$

The directions η_j are defined recursively by

$$\eta_j = \arcsin \frac{y_{n-j+1}}{\rho \prod_{i=1}^{j-1} \cos \eta_i} \in (-\frac{1}{2}\pi, \frac{1}{2}\pi], \qquad \text{for } j = 1, .., n-1, \tag{4B.9}$$

where by convention $\prod_{i=1}^{0} \cos \eta_i = 1$, such that $\eta_1 = \arcsin y_n/\rho$. The inverse transformation $T^{-1}_{y\to\eta,\rho}(\eta, \rho)$ from polar coordinates to y, is

$$y_j = T^{-1}_{y,j}(\eta, \rho) = \rho \sin \eta_{n-j+1} \prod_{i=1}^{n-j} \cos \eta_i, \qquad j = 1, .., n, \tag{4B.10}$$

when we define the sinus of the (otherwise undefined, and unused) η_n to be 1. The Jacobian of the transformation $T_{\eta,\rho}(\theta)$ from θ to (η, ρ) is

$$\begin{aligned} J(\eta, \rho) &= \det(\Sigma)^{\frac{1}{2}} \times |\rho|^{n-1} \times \left| \prod_{i=1}^{n-2} \cos^{n-i-1} \eta_i \right| \\ &= \det(\Sigma)^{\frac{1}{2}} \times |J(\rho)| \times |J(\eta)|. \end{aligned} \tag{4B.11}$$

In section 4.3.4 the marginal density of η resulting from a normal density in the original parameter space is used. Here, we present the derivation of the transformed candidate and target densities. Denote the normal density in the original space by

$$q(\theta) \propto \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta - \mu)'\Sigma^{-\frac{1}{2}}(\theta - \mu) \right),$$

and denote by $q(\eta,\rho), q(\eta)$ and $q(\rho)$ the joint and marginal densities of η and ρ after the polar transformation defined by (4B.8)-(4B.9). The following results hold:

$$\begin{aligned} q(\eta,\rho) &= |J(\eta,\rho,\Sigma)|\, q(T^{-1}(\eta,\rho)) \\ &\propto \det(\Sigma)^{\frac{1}{2}}\, |\rho|^{n-1}\, |\prod_{i=1}^{n-2} \cos^{n-i-1}\eta_i|\, \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta-\mu)'\Sigma^{-\frac{1}{2}}(\theta-\mu)\right) \\ &= |\prod_{i=1}^{n-2} \cos^{n-i-1}\eta_i| \times |\rho|^{n-1} \times \exp\left(-\frac{1}{2}\rho^2\right) \\ &= |J(\eta)| \times |J(\rho)| \times \exp\left(-\frac{1}{2}\rho^2\right), \\ q(\eta) &\propto |J(\eta)|, \qquad (4B.12) \\ q(\rho) &\propto |J(\rho)| \exp\left(-\frac{1}{2}\rho^2\right), \qquad (4B.13) \end{aligned}$$

see also equation (4B.11). For the target density, the independence between ρ and η does not hold in general. The marginal density of η for the target density $p_\theta(\theta)$ is

$$p(\eta) = \int p_\theta\left(T^{-1}(\eta,\rho)\right) |J(\eta,\rho,\Sigma)|\, d\rho = |J(\eta)|\, |J(\Sigma)| \int p_\theta\left(T^{-1}(\eta,\rho)\right) |J(\rho)|\, d\rho. \quad (4B.14)$$

In the acceptance probability (4.13), the Jacobians of η and Σ cancel, leaving only the parts under the integral, evaluated over both the candidate direction η^* and the previous direction $\eta^{(i)}$.

4.C Selected density functions

In the course of this chapter, several density functions are applied. Though all density functions are well known to most researchers, several notations are in use. In order to create clarity as to the notation applied here, a list of densities is provided.

i. The Normal density:

$$\begin{aligned} f_{\mathcal{N}(\mu,\sigma^2)}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \\ \mathsf{E}(x) &= \mu, \qquad \mathsf{var}(x) = \sigma^2, \qquad \mathsf{mode}(x) = \mu. \end{aligned}$$

ii. The Student-t density:

$$\begin{aligned} f_{t(\mu,\alpha,\nu)}(x) &= B\left(\frac{\nu}{2},\frac{\nu}{2}\right)(\nu\alpha^2)^{-1/2}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\alpha}\right)^2\right)^{-\frac{\nu+1}{2}}, \\ \mathsf{E}(x) &= \mu \quad (\nu>1), \qquad \mathsf{var}(x) = \alpha^2\frac{\nu}{\nu-2} \quad (\nu>2), \qquad \mathsf{mode}(x) = \mu. \end{aligned}$$

iii. The Inverted Gamma density:

$$f_{\text{IG}}(x|\alpha,\beta) = [\Gamma(\alpha)]^{-1}\,\beta^{-\alpha}x^{-(\alpha+1)}\exp\left(-\frac{1}{x\beta}\right),$$
$$\mathsf{E}(x) = \frac{1}{\beta(\alpha-1)} \quad (\nu>2), \qquad \mathsf{var}(x) = \frac{1}{\beta^2(\alpha-1)^2(\alpha-2)} \quad (\nu>4),$$
$$\mathsf{mode}(x) = \frac{1}{\sqrt{\beta(\alpha+1)}}.$$

iv. The Inverted Gamma-1 density:

$$f_{\text{IG-1}}(x|\alpha,\beta) = [\Gamma(\alpha)]^{-1}\,\beta^{-\alpha}x^{-(2\alpha+1)}\exp\left(-\frac{1}{x^2\beta}\right),$$
$$\mathsf{E}(x) = \frac{1}{\sqrt{\beta}}\frac{\Gamma\left(\alpha-\frac{1}{2}\right)}{\Gamma(\alpha)} \quad (\nu>1), \qquad \mathsf{var}(x) = \frac{1}{\beta(\alpha-1)} - \mathsf{E}(x)^2 \quad (\nu>2),$$
$$\mathsf{mode}(x) = \frac{1}{\sqrt{\beta(\alpha+\frac{1}{2})}}.$$

v. The Cauchy density:

$$f_C(x) = B\left(\frac{1}{2},\frac{1}{2}\right)\frac{1}{1+x^2},$$
$$\mathsf{E}(x) \text{ does not exist}, \qquad \mathsf{var}(x) \text{ does not exist} \qquad \mathsf{mode}(x) = 0.$$

Chapter 5

Daily Hedging of Currency Risk

5.1 Introduction

The Bayesian approach to analysing econometric models which was described in chapter 4 can be applied to about every conceivable model, but the difference between the Bayesian and the classical methodology is most important when parameter uncertainty plays an influential role. This is certainly the case when we investigate the exchange rate behaviour of currencies on a daily basis: The returns on the exchange rate tend to look very much like pure random noise, though over the months some trending behaviour seems to occur. Combined with a complete decision framework, in order to decide if the risk involved in an investment denoted in a foreign currency should be hedged, exchange rate models are a very interesting field for Bayesian research.

In this chapter, which is an extension of the paper by Bos, Mahieu and Van Dijk (2000*a*) and Bos, Mahieu and Van Dijk (2000*b*), we analyse the risk and return properties of currency overlay strategies using time series models that describe prominent features of daily exchange rate data. Our contribution focuses on three issues. First, we introduce a class of models which describes some major features of the data: local trends in the level or varying means in the return, time varying volatility in the second moment of the return, and leptokurtosis of the returns. We integrate models for the analysis of varying means, varying variances, and heavy tailed distributions. Then we obtain a flexible general framework which enables us to study the effects and relevance of different model specifications for hedging decisions. The topics that we investigate in this respect are unit roots versus persistent but stationary behaviour in expected returns, heavy tailed distributions, and different ways to model conditional volatility. Second, for inference and decision analysis we make extensive use of the Bayesian methods based on Markov chain Monte Carlo (MCMC) simulation, as they were introduced in chapter 4. Third, in the decision analysis we investigate the payoff and utility derived from an optimal strategy using alternative models and corresponding results from alternative strategies for some selected models.

The outline of the chapter is as follows. In section 5.2 we introduce further the concept of hedging currency risk and present our procedure for executing the currency overlay strategy. In section 5.3 we present some time series models for describing daily exchange rate returns. We introduce a state space model for the time varying mean which

is augmented with a Generalized Autoregressive Conditional Heteroskedastic (GARCH) or a Stochastic Volatility (SV) model for a time varying variance and further augmented with a Student-t model for the disturbances for extreme observations. State space (or structural time series) models are nowadays widely used for describing time varying structures, see e.g. Harvey (1989) or West, Harrison and Migon (1985). In section 5.4 we discuss our Bayesian methods, see e.g. Smith and Roberts (1993) and Chib and Greenberg (1995). In the recent literature these methods have been successfully applied for studying separately the pattern of time varying means (see Carter and Kohn 1994, Koop and Van Dijk 2000) and the pattern of varying volatilities (see Kim, Shephard and Chib 1998). Results are presented in section 5.7 using the DM/USD and Yen/USD daily exchange rate series for the period January 1982 until December 2000. Some concluding remarks are given in section 5.8. Conditional densities used in MCMC sampling from the posterior distribution are summarized in table 5.11 in appendix 5.A.

5.2 Currency hedging

When investing abroad, international firms face the decision whether or not to hedge the risk of a depreciation of the foreign currency compared to the home currency. For example, when a corporation sells its goods abroad it incurs foreign exchange rate exposure at the time it wants to repatriate the proceeds of the sales. Another large group of companies with foreign currency exposure are internationally operating investors, like banks, pension funds, and insurance companies. The currency exposures arise from the investment strategies that these institutions follow. For example, when a U.S. dollar-based investor decides to diversify into Japanese stocks he runs the risk of a Japanese yen depreciation. Although the portfolio allocation decision could also depend on the risk and return characteristics of foreign currencies, in practice these two decisions, i.e. the investment and the hedging decision, are often separated. The approach where currency hedging decisions are made independently from underlying investment decisions, is called 'currency overlay management' in the finance industry. Note that this approach may lead to suboptimal decisions from a fund's perspective as a currency overlay strategy ignores the diversifying characteristics that currencies may have. Continuing the example, when the investor perceives the risk of the Japanese yen depreciating too large, he may decrease his holdings of Japanese stocks. However, by applying currency overlay management the investor tries to manage his Japanese yen currency exposure irrespective of the amount of wealth invested in Japanese stocks. A major reason for investors to separate the currency and portfolio decisions is to obtain increased transparency of the investment strategy.

When considering currency overlay management, relevant economic variables are the exchange rates and the values of the instruments used for hedging the exposures. A common instrument to hedge foreign currency exposure is the forward contract, which gives the investor the right (and the obligation) to convert the foreign currency exposure from one currency to another for a fixed rate at a specified point in time in the future. From covered interest rate parity we know that the forward exchange rate can be calculated from the current spot exchange rate and the difference between the short term interest rates in the home and foreign country, respectively. Other instruments may be considered

as well, notably foreign currency options. In this paper we focus on hedging with forward contracts only.

To illustrate the practical importance of currency overlay management one may distinguish two special cases. First, the decision-maker does not hedge at all. The return on the currency overlay strategy is then equal to the return on the exchange rate. Second, the decision-maker hedges the currency risk completely. Now, the return is equal to the difference between the interest rates of the home country and that of the foreign country. A practical example is the case of a German firm with U.S. investments. In the period 1998-2000, the U.S. Dollar rose almost sixteen percent in DM terms, while the cumulative difference between the two interest rates was around minus 4.7 percent. Thus, the decision to hedge or not to hedge relates to a difference in cumulative return for those three years of approximately twenty percent. Since multinational corporations and large institutional investors deal with substantial foreign currency exposures that may involve hundreds of millions of dollars, the specification of an effective strategy for foreign exchange rate management is an important topic.

After describing the importance of hedging of currency risk, let us introduce the setting that we investigate in this paper. Let s_{t+1} be the exchange rate return over the time interval $[t, t+1]$, defined as $s_{t+1} \equiv \log(S_{t+1}/S_t)$, with S_t the exchange rate itself. Let $F_{t,\tau}$ be the current value of a forward contract with maturity date τ. By covered interest rate parity it is equal to

$$F_{t,\tau} = S_t \exp\left(r^h_{t,\tau} - r^f_{t,\tau}\right), \tag{5.1}$$

with $r^h_{t,\tau}$ and $r^f_{t,\tau}$ the home and foreign risk-free interest rates with maturity τ, respectively.[1] With respect to the specific value of τ we note that in our empirical analysis we use interest rates with a 30-day maturity, implying that we have 30-day forward rates. A forward contract, which may have a remaining lifetime of less than 30 days, can be neutralised by taking an opposite forward position. As a consequence, a synthetic one-day forward contract is created. This approach is common in actual applications of currency hedging.

Define H_t as the fraction of the underlying exposure that is hedged with (synthetic, one-day) forward contracts. We refer to this variable as the hedge ratio. At time t we have an exposure of S_t. Note that the forward contract does not provide any cash flows at time t. At time $t+1$ we have a cash flow of $(1-H_t)S_{t+1} + H_t F_t$, dropping the subscript τ. The first part is the fraction of the exposure that we did not hedge, and the second part refers to the pay-out of the forward contract at time $t+1$. The continuously compounded return[2] is given as

$$r_{t+1} \equiv \log\left(\frac{(1-H_t)S_{t+1} + H_t F_t}{S_t}\right). \tag{5.2}$$

In our empirical work we make use of the exponent of the continuously compounded

[1] See Solnik (2000) for a comprehensive review of covered interest rate parity.

[2] We have also checked our results with arithmetic returns. The results changed somewhat. We are indebted to a referee for bringing up this point.

return,

$$\exp(r_{t+1}) = (1 - H_t)\exp(s_{t+1}) + H_t \exp\left(r_t^h - r_t^f\right). \tag{5.3}$$

It is seen that the exponent of the return is a weighted average of the exponents of the exchange rate return s_{t+1} and the difference between the home and foreign risk free interest rates. Note that when we set the hedge ratio H_t to zero, the return on the currency overlay part is equal to the return on the exchange rate only. On the other hand, if we set the hedge ratio equal to one, only the interest rate differential has an impact, whereas changes in the currency do not affect the return on the currency overlay.[3]

Given a time series model, to be introduced in the next section, which describes exchange rate behaviour, and given all data information up to time t, the currency manager wants to determine the hedge ratio that applies to the next period. In order to perform this task he is assumed to specify an objective function that captures his risk and return attitudes towards foreign currencies over some future time horizon. We assume that the investor has a standard power utility function with constant relative risk aversion

$$U(W_t) = \frac{W_t^\gamma - 1}{\gamma}, \qquad \gamma < 1. \tag{5.4}$$

The parameter γ describes the level of risk aversion and needs to be specified by the currency manager. The lower γ, the more risk averse the manager is. In the empirical analysis we present results for several values of γ. The variable W_t represents the wealth that the investor obtains by executing the currency overlay strategy. Wealth changes as a result of the hedging strategy only. The value of next period's wealth is given by $W_{t+1} = W_t \exp(r_{t+1})$. We assume that the currency manager follows a myopic strategy, i.e. he makes a hedging decision for the next period only, irrespective of possible states of the world after that period. In that case we can normalize W_t to one, without loss of generality. The problem that the currency manager needs to solve can be stated as

$$\max_{0 \le H_t \le 1} \mathsf{E}_{s_{t+1|t}} U(W_{t+1}) = \max_{0 \le H_t \le 1} \mathsf{E}_{s_{t+1|t}} \left[\frac{\left(\exp\left(r_{t+1}(s_{t+1}, H_t, r_t^h, r_t^f)\right)\right)^\gamma - 1}{\gamma} \right], \tag{5.5}$$

with $\mathsf{E}_{s_{t+1|t}}$ a conditional expectations operator, taken with respect to the predictive density of tomorrow's return s_{t+1}, $p(s_{t+1|t})$, given the information available at time t. In the optimization equation (5.5) we have inserted definition (5.3) for the return on the currency strategy.

In the empirical part of this paper the main focus is on the results of hedging decisions based on optimization of the power utility function. We briefly compare the results to

[3]The hedge ratio is restricted to values between 0 and 1. The reason for this is that our prime focus lies on currency overlay management for investors that have large, relatively static, portfolios of foreign securities. These investors are generally not interested in taking currency positions that exceed the value of their underlying securities. Indeed, for corporations that have frequently changing cash flows denoted in foreign currencies, other ranges for hedge ratios might be appropriate. We leave this as a topic for further research.

those obtained from hedging decisions based on Value-at-Risk (VaR), and decisions based on the Sharpe ratio.

Decision rules based on the VaR concept may be motivated as follows. A currency manager wants to control the risk of depreciation of foreign currencies. A popular measure for downside risk, advocated by financial regulatory institutions, is Value-at-Risk. VaR measures the maximum loss that is expected over a fixed horizon with a prespecified confidence probability. In our case we define the one-period VaR as

$$\int_{-\mathrm{VaR}}^{\infty} f(r_{t+1|t}) d\, r_{t+1} = 1 - \alpha, \tag{5.6}$$

with $1-\alpha$ the confidence probability, with α typically ranging from 1% to 10%. The choice of confidence level is motivated by the risk attitude of the investor in relation to the horizon over which the VaR is calculated, see Jorion (1997). The currency manager decides to hedge his currency exposure when the estimated VaR falls above a prespecified limit risk he is willing to take.

Another popular measure for the relation between expected return and risk is the Sharpe ratio, which compares the expected return with the standard deviation of the returns. The Sharpe ratio is given as

$$\mathrm{Sh} = \frac{\mathsf{E}_{s_{t+1|t}}(r_{t+1})}{\sqrt{\mathsf{var}_{s_{t+1|t}}(r_{t+1})}}, \tag{5.7}$$

with $\mathsf{var}_{s_{t+1|t}}(r_{t+1})$ the predictive variance of the return r_{t+1}. As in the case of Value-at-Risk, the investor makes a decision to hedge by comparing the value of the Sharpe ratio with a certain prespecified limit. If the Sharpe ratio is higher than this limit, no hedging is required, and vice versa.

5.3 Time series models for exchange rate returns

Many models have been suggested for describing time series properties of exchange rates (see e.g. LeBaron 1999). In this paper we concentrate on models that describe prominent data features of floating daily exchange rates. First, exchange rates may exhibit local trend behaviour. For several months for instance, a successive decline or successive appreciation of the exchange rate may occur. This implies a varying mean behaviour of the exchange rate return s_t. We model this by the state space model

$$s_t = c + \mu_t + \epsilon_t, \qquad \epsilon_t \sim \text{i.i.d.}(0, \sigma^2_{\epsilon,t}), \tag{5.8}$$

$$\mu_t = \rho\mu_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}\left(0, \sigma^2_\eta\right), \qquad t = 1, .., T. \tag{5.9}$$

Next to the local trend we μ_t we also allow for a non-zero mean return c, corresponding to a global trend in the exchange rate.[4] The local trend is the unobserved mean component μ_t. It is an autoregressive process with disturbances η_t and autoregressive parameter ρ.

[4]In most situations, there is no basis for assuming the exchange rate to trend in the same direction over many years. In subsequent sections the parameter c is dropped in most models.

This model, which we label the Generalized Local Level (GLL) model, is supposed to pick up the periods of rising or falling exchange rate levels.[5] The disturbances η_t are assumed to be independently and identically normally distributed with constant variance σ_η^2. The autoregressive model incorporates as a limiting case the fully integrated mean return model, when $\rho = 1$. This model is known as the Local Level (LL) model, see Harvey (1989, p. 45). Given $\sigma_\eta^2 > 0$, the LL model implies that the logarithm of the exchange rates $\log S_t$ follows an I(2) process. We expect that, when estimating this I(2) model on our data, the variance of η_t is small compared to the variance of ϵ_t, such that the I(1) behaviour of $\log S_t$ overwhelms the I(2) effects. One can also take the limit case where μ_t disappears, i.e. using $\sigma_\eta^2 = \rho = \mu_t = 0$, which is White Noise (WN) around a fixed mean c. Though extremely simple, it is a basic model in many financial market models.

The second main feature of financial series concerns the variance structure. Several model specifications have been suggested to account for periods of lower and higher variance in the data. See e.g. Bollerslev (1986), Engle (1982), Engle (1995), Nelson (1990) or Taylor (1994). Conditioning on the information available at time $t-1$ (indicated by the subscript $t|t-1$), we write

$$\epsilon_{t|t-1} \sim \mathcal{N}\left(0, \sigma_{\epsilon,t}^2\right). \tag{5.10}$$

For the variance structure, we distinguish between three cases. First, the simplest option is to ignore the time dependence of volatility altogether. Then the model is written as

$$\sigma_{\epsilon,t}^2 = \sigma_\epsilon^2, \tag{5.11}$$

in which case a standard (Gaussian, homoskedastic) state space model results.

More flexibility is obtained when a GARCH disturbance process is allowed for. The variance $\sigma_{\epsilon,t}^2$ of the observation equation (5.8) varies over time according to

$$\begin{aligned} \sigma_{\epsilon,t}^2 &= \sigma_\epsilon^2 h_t, \\ h_t &= \delta h_{t-1} + \omega + \alpha(\mathsf{E}\epsilon_{t-1})^2/\sigma_\epsilon^2, \\ \delta \geq 0, &\quad \alpha \geq 0, \quad \delta + \alpha < 1, \quad \omega \equiv 1 - \delta - \alpha. \end{aligned} \tag{5.12}$$

Note that we use one of the approximations of Harvey, Ruiz and Sentana (1992) in the GARCH equation. As the state μ_t of the process is not observed, we also do not observe the precise value of ϵ_{t-1}. Instead, we use $\mathsf{E}\epsilon_{t-1} = s_t - c - \mathsf{E}\mu_t$ with expectations taken conditional upon all information available at time t.[6] The restrictions on the parameters are sufficient to ensure strict positiveness of $\sigma_{\epsilon,t}^2$ and the existence of a finite value for

[5] Theoretically the interest rate differential should be introduced as the expectation of s_t, as the uncovered interest rate parity (UIP) prescribes. However, empirically the UIP does not hold when using high frequency exchange rate data. The interest rate differential will be introduced later in the evaluation of the returns.

[6] There exist several alternative operationalizations of the GARCH equation (5.12). The option we use here is conceptually simplest. Other options are to include the disturbance ϵ_t in the state of the state space model (leading to a model which is no longer conditionally Gaussian), or to use output of the Kalman filtering equations to compute $\mathsf{E}(\epsilon_{t-1}^2)$ and use this instead of $(\mathsf{E}\epsilon_{t-1})^2$. This last option is expected to stay slightly closer to the original GARCH model than the choice we use in the following.

the unconditional expectation $\mathsf{E}(\sigma^2_{\epsilon,t}) = \sigma^2_\epsilon$ or equivalently $\mathsf{E}(h_t) = 1$ (see Kleibergen and Van Dijk 1993).

The third choice for the variance process is to allow for Stochastic Volatility (SV, see Jacquier, Polson and Rossi 1994). The variance of the disturbances in the observation equation evolves according to

$$\begin{aligned} \sigma^2_{\epsilon,t} &= \sigma^2_\epsilon \exp(h_t), \\ h_t &= \phi h_{t-1} + \xi_t, \\ \xi_t &\sim \mathcal{N}\left(0, \sigma^2_\xi\right). \end{aligned} \tag{5.13}$$

The parameter σ^2_ϵ governs the overall variance of the process, with deviations modelled through h_t. Aguilar, Huerta, Prado and West (1999) propose a dynamic factor model with stochastic volatility which is similar to the one proposed here.

A third feature of financial time series is that the histograms of the returns exhibit heavier tails than the normal density, even after correcting for the time varying volatility. To model this, we replace equation (5.10) by

$$\epsilon_{t|t-1} \sim t\left(0, \sqrt{\frac{\nu-2}{\nu}}\sigma_{\epsilon,t}, \nu\right), \qquad \nu > 2, \tag{5.14}$$

where t indicates the Student-t density, with expectation 0, variance $\sigma^2_{\epsilon,t}$ and ν degrees of freedom (see appendix 4.C for the notation used).

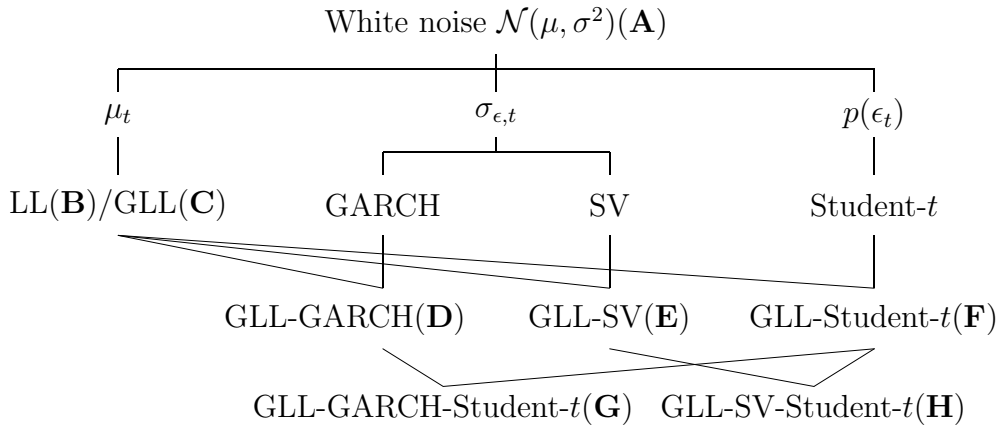

Figure 5.1: Hierarchy of models

Figure 5.1 summarizes the models that are used in subsequent sections. The basic model is the White Noise (WN) model, with normally distributed returns around a mean $\mu = c$. Then there are three directions of generalization: time dependence of the mean μ_t, time dependence of the variance $\sigma^2_{\epsilon,t}$, or the shape of the density of the innovations ϵ_t. More specifically, the third line in the figure indicates the models that we consider. Note that the Local Level (LL) model is a special case of the Generalized Local Level (GLL) model, with $\rho = 1$. The GLL is combined with the three generalizations (GARCH, SV and Student-t), such that a broad range of competing models is obtained. When the GLL

model is combined with both the Student-t elements and either the GARCH or the SV component, a further generalization is found. These two models are indicated in the fifth line of the figure. The models are indicated by the letters **A**-**H** in the figure and in text and tables in subsequent sections.

5.4 Bayesian inference and decision

5.4.1 Prior structure

Inference and decision analysis is performed within a Bayesian framework. In table 5.1 we present the priors on the parameters of the models that are used. We make use of proper priors which are expected to be weakly informative compared to the information in the likelihood. Given proper priors, we can compute marginal likelihoods in order to compare alternative models. Conjugate priors are used for all parameters, except parameters δ, α and ν. This facilitates the computations. Hyperparameters are chosen such that little information is put in the priors.

The first parameter encountered in the models is the global mean return c. It is expected a priori to be zero, or not much larger. A normal prior with standard deviation $\sigma_0 = 0.02$ will do perfectly well.

The autoregressive parameter ρ of the unobserved mean process μ_t is crucial in the analysis. It governs the amount of predictability in the series (together with the ratio of the variances in observation and transition equations (5.8) and (5.9), and, if included, the global mean c). Given the fact that trends in exchange rates may last for several months, we deem a large value of ρ in the unit interval a priori more plausible than a small value. As an intermediate position between a strongly informative and an uninformative prior, we choose a normal prior density with mean 0.8 and a rather large standard deviation of 0.2.[7] More information is available on the variance process in series like the one at hand. Therefore, the choice of prior for the AR parameter ϕ in the SV process is less influential. Again, a normal prior is used, now with mean 0.5 and standard deviation 0.3.

Table 5.1: Description of priors used

Parameter	Prior	Hyper-parameters	Used in model
c	$\mathcal{N}(\mu_0, \sigma_0^2)$	$\mu_0 = 0, \sigma_0 = 0.02$	**A**
σ_ϵ	IG-1$(\alpha_\epsilon, \beta_\epsilon)$	$\alpha_\epsilon = 2.5, \beta_\epsilon = 4/3$	**A**, **B**, **C**, **D**, **E**, **F**, **G**, **H**
ρ	$\mathcal{N}(\mu_\rho, \sigma_\rho^2)$	$\mu_\rho = 0.8, \sigma_\rho = 0.2$	**C**, **D**, **E**, **F**, **G**, **H**
σ_η	IG-1$(\alpha_\eta, \beta_\eta)$	$\alpha_\eta = 2.25, \beta_\eta = 100$	**B**, **C**, **D**, **E**, **F**, **G**, **H**
δ, α	Uniform at stationary region		**D**, **G**
ϕ	$\mathcal{N}\left(\mu_\phi, \sigma_\phi^2\right)$	$\mu_\phi = 0.5, \sigma_\phi = 0.3$	**E**, **H**
σ_ξ	IG-1(α_ξ, β_ξ)	$\alpha_\xi = 2.5, \beta_\xi = 4/3$	**E**, **H**
ν	Truncated Cauchy, $\nu > 2$		**F**, **G**, **H**

[7] Note that we did not restrict $\rho \in [0, 1]$. Other priors, including a uniform prior between 0 and 1, were used. Results were similar to the results presented here.

The priors for the parameters governing the standard deviations are all Inverted Gamma-1 (see appendix 4.C, or Bauwens, Lubrano and Richard 1999) distributions. The hyperparameters are chosen based on similar series, with expectation of 0.5, 0.008 and 0.5 for $\sigma_\epsilon^2, \sigma_\eta^2$ and σ_ξ^2 respectively. In Bauwens and Lubrano (1998) it is proven that a prior for the degrees-of-freedom parameter ν with too heavy tails (e.g. $\pi(\nu) \propto 1$ or $\pi(\nu) \propto 1/\nu$) can ruin the properness of the posterior. The truncated Cauchy prior used here ensures that these problems do not occur.

The GARCH parameters δ and α are bounded by the stationarity condition to be positive and smaller than 1 in sum. On the stationarity region, we assume a uniform prior.

5.4.2 Constructing a posterior sample

In chapter 4 the basic methodology of constructing a posterior sample was explained. Models **A**-**C**, and also model **D** when the approximation of using $\mathsf{E}\epsilon_{t-1}$ instead of ϵ_{t-1} itself in the GARCH equation is used, allow using the Metropolis-Hastings sampling method, as in the example with Gaussian disturbances in section 4.5. For models **E**-**H**, the GLL-Stochastic Volatility, GLL-Student-t, GLL-GARCH-Student-t and the GLL-Stochastic Volatility-Student-t models, we need to apply a data augmentation scheme to obtain conditional normality and include the unobserved variables into the state. Appendix 5.A describes a Gibbs sampling scheme as in Kim et al. (1998) that can be applied on these models.

In following sections, results for all models are based on outcomes of the Gibbs sampler. Using the Metropolis-Hastings sampler for models **A**-**D** would take less computational effort, but the results of the hedging exercise are the same.

5.4.3 Evaluating the marginal likelihood

In order to judge the fit of the models to the data, the marginal likelihood of each of the models may be calculated. The marginal likelihood m for model M is defined as

$$m(M) = \int \mathcal{L}(\text{data}; \theta, M)\pi(\theta|M)d\theta \tag{5.15}$$

and may be computed using Bayes' rule as

$$m(M) = \frac{\mathcal{L}(\text{data}; \theta, M)\pi(\theta|M)}{p(\theta|\text{data}, M)}. \tag{5.16}$$

In this equation, $p(\theta|\text{data}, M)$ is the posterior density of model M evaluated at the location indicated by the vector of parameters θ, and $L(\text{data}; \theta, M)$ and $\pi(\theta|M)$ are the likelihood and prior, respectively (see e.g. Gelfand and Smith 1990).

In section 4.4 it is explained how the marginal likelihood can be calculated both in situations where the likelihood function is available in closed form (here: for models **A**-**C**, and also **D** with the approximation mentioned before), and when data augmentation is needed to get a tractable model. In section 5.6.2 the results for models **A**-**D** are calculated using both the methods of Kass and Raftery (1995) and Chib (1995), to judge

the accuracy and comparability of the approximation methods. For models **E-H**, only the Gibbs results are reported.

The method of Chib was uses the conditional densities as described in appendix 5.A. In some cases a Metropolis-Hastings step was applied within the Gibbs chain, as the full conditional density could not easily be sampled from. In such a case, it is also not easily possible to evaluate the conditional density $p(\theta_i|\theta_1,..,\theta_{i-1},\theta_{i+1},..,\theta_k)$, including the integrating constant. For this evaluation a LaPlace approximation is used.

5.4.4 Predictive analysis

The decision whether to hedge or not is based on the unconditional predictive density $p(s_{t+1|t})$ of tomorrow's returns on the exchange rate s_{t+1}, given all available information. The conditional density $p(s_{t+1|t}|\theta)$, given the vector of parameters θ, is easily derived. The unconditional predictive density follows by marginalization with respect to θ,

$$p(s_{t+1|t}) = \int_{\theta\in\Theta} p(s_{t+1|t}|\theta)p(\theta|s_t,s_{t-1},..,s_1)\,d\theta, \tag{5.17}$$

see e.g. Geweke (1989) and Barberis (2000). Marginalization is done with respect to the posterior density of $\theta|s_t,s_{t-1},..,s_1$. On-line modelling and prediction requires that the posterior of the parameters is reestimated for every day in the evaluation period. However, for computational reasons we refrain from doing this and use only N drawings $\theta^{(1)},..,\theta^{(N)}$ from the posterior of $\theta|s_T,..,s_1$, with $s_T,..,s_1$ the observations from the estimation sample ($T<t$). When the estimation sample is large compared to the evaluation sample, this approximation gives, under standard regularity conditions, a sufficient level of accuracy. The integral in (5.17) is approximated using

$$p(s_{t+1|t}) \approx \frac{1}{N}\sum_{i=1}^{N} p(s_{t+1|t},\theta^{(i)}) \tag{5.18}$$

at a fine grid of possible values s_{t+1}. The resulting predictive density is used for the decision analysis.

5.4.5 Decision analysis

The investor optimizes the expected utility, with respect to the predictive density for the exchange rate returns. We numerically solve

$$\begin{aligned} H_t &= \arg\max_{H_t} \mathsf{E}_{s_{t+1|t}} U(W_{t+1}) = \\ &= \arg\max_{H_t} \int_{s_{t+1}} \frac{\exp\left(r_{t+1}(H_t,s_{t+1},r^h,r^f)\right)^\gamma - 1}{\gamma} p(s_{t+1|t})\,ds_{t+1}, \end{aligned} \tag{5.19}$$

see equation (5.5). Optimal hedge ratios are computed using a grid search for every day in the evaluation period.

In section 5.2, two other decision strategies were presented. For the Value-at-Risk (VaR), we evaluate for each day what the 5% VaR is according to the model at hand.

The investor should decide if the VaR is acceptable for him, or that he deems the risk too high. For reasons of comparison, we specify the VaR cut-off level in such a way that the average hedge ratio corresponds to the average hedge ratio found when fully optimizing the utility function.

The final strategy was based on the Sharpe ratio, measuring the expected return the investor could get for one unit extra of variance. If expected return is higher than a cut-off level, one chooses not to hedge. In the other case, full hedging is chosen. Again, the cut-off level is calibrated to a level leading to comparable hedging results with the fully optimized case.

5.5 Exchange rate data and the interest rate

5.5.1 Stylized facts

Our data set consists of daily observations on the DMark/US Dollar (DM/USD) and the Yen/US Dollar (Yen/USD) exchange rate for the period January 1, 1982 until December 29, 2000 which gives a total of 4,956 observations. For the same period we have the 1-month Eurocurrency interest rates for the German DMark, the Japanese Yen and the US Dollar.[8] Of this data set, the first 16 years are used in constructing a sample from the posterior density. The years 1998, 1999 and 2000 are only used in computing the optimal hedge decision.

In the upper panel of figure 5.2 the time series are presented in levels (on the left) and in first differences of the logarithms (on the right) for the whole period. In the levels one may observe the changing trend which implies a changing mean in the exchange rate returns. The autocorrelation functions of both returns and squared returns (in the lower panels) exhibit patterns frequently found in high frequency financial return data. As for the returns, it is seen that there is no clear serial correlation pattern, corroborating the widely held view that financial return series are unpredictable. However, the local trends in the levels of the exchange rates may prove useful for practical currency overlay strategies. The phenomenon of local trends is, at a longer horizon, similar to the data feature of long swings in the dollar as observed by Engel and Hamilton (1990). We note that we use a state space model while these authors use a Markov switching process for describing exchange rate returns over longer periods.

The squared returns show a clear pattern. The slowly decaying autocorrelation has prompted many researchers to develop models for describing time varying volatilities. Note that the slow decay of the autocorrelation of the squared returns is more smooth for the DM/USD data than for the Yen/USD squared returns.

Figure 5.3 shows the time series of German, Japanese and U.S. interest rates. The maturity of the interest rates is 30 days. Compared to the exchange rate, the interest rates are much less volatile. Additionally, these series and the difference between the two (which is used in the hedging decision) are very persistent. Note that in the hedging

[8]Source: DATASTREAM, series `DMARKER/USDOLLR`, `JAPAYEN/USDOLLR`, `ECWGM1M`, `ECJAP1M` and `ECUSD1M` for the daily DM/USD and Yen/USD exchange rates and German, Japanese and U.S. 1-month Eurocurrency middle interest rates, respectively.

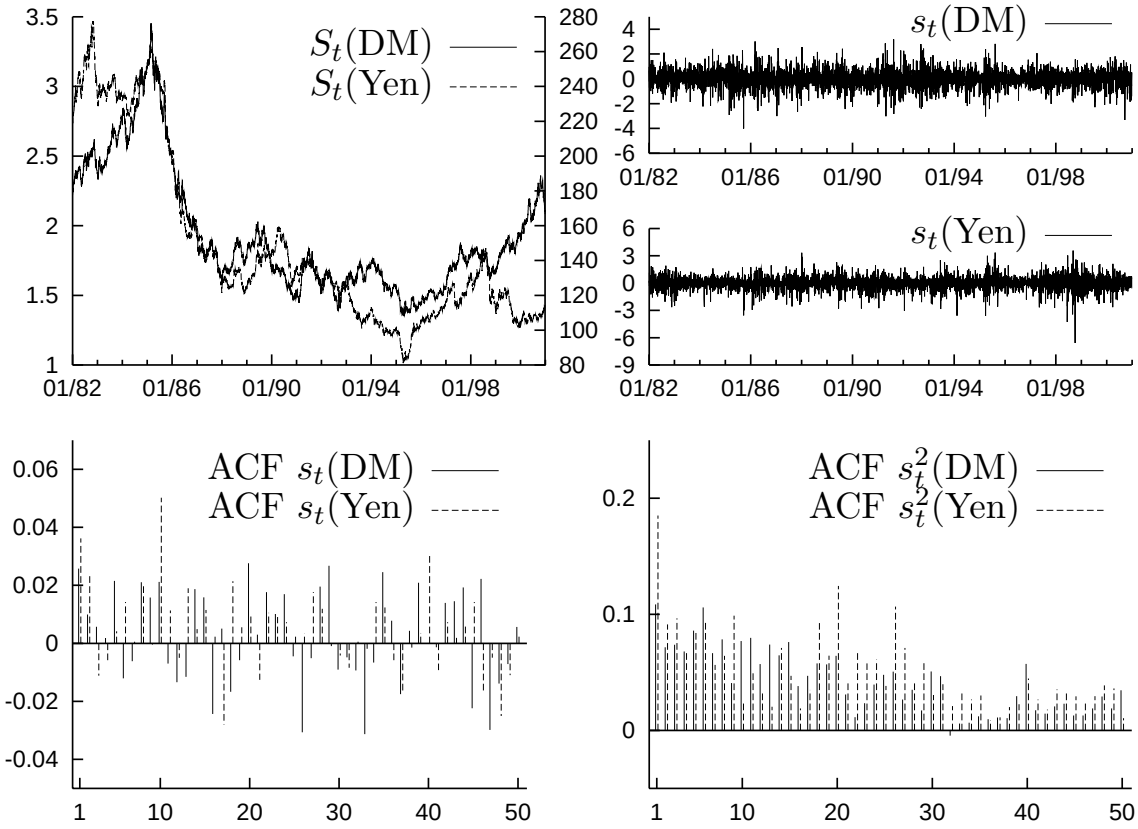

Figure 5.2: DM/USD and Yen/USD exchange rates, January 1, 1982 until December 29, 2000. Panels contain data in levels (top left), in percentage returns (top right), and the autocorrelation function of the returns and the squared returns (bottom).

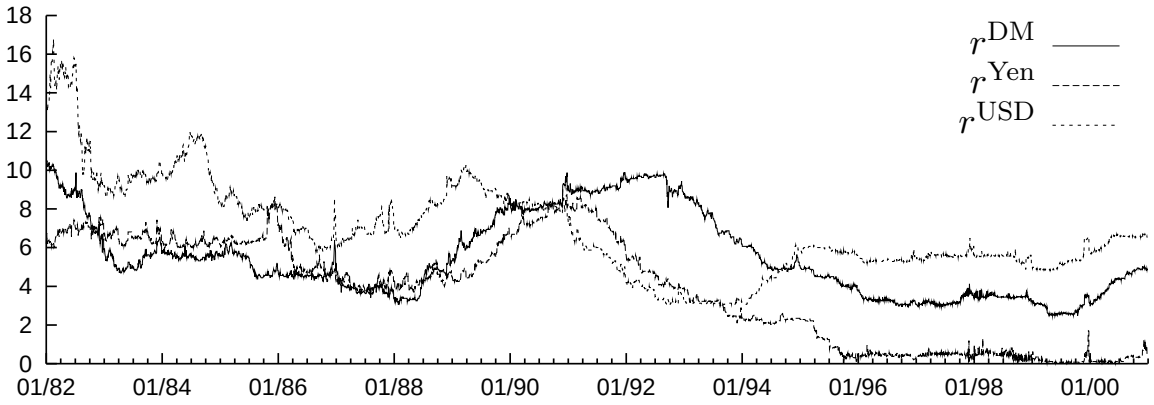

Figure 5.3: 1-month Eurocurrency interest rates for the German DMark, Japanese Yen and the US Dollar, in yearly percentages

decision we transformed the series to daily interest rates, see the remark in section 5.2.

5.5.2 Data in the evaluation period

After estimating the posterior densities of the model parameters over the period 1/1/1982–31/12/1997, in section 5.7 the evaluation of the hedging decision is described. In this section, data over the period 1/1/1998–29/12/2000 is used. The risk manager of the firm tries to decide between running the exchange rate risk, or to hedge this risk resulting in getting the interest rate differential $\Delta r(\mathrm{H/F}) = r^{\mathrm{H}} - r^{\mathrm{F}}$ between the home and foreign countries as a return. As this period is of special interest, figure 5.4 displays the cumulative returns over this period of the DM/USD and Yen/USD exchange rates, and the cumulative interest rate differentials over the same period.

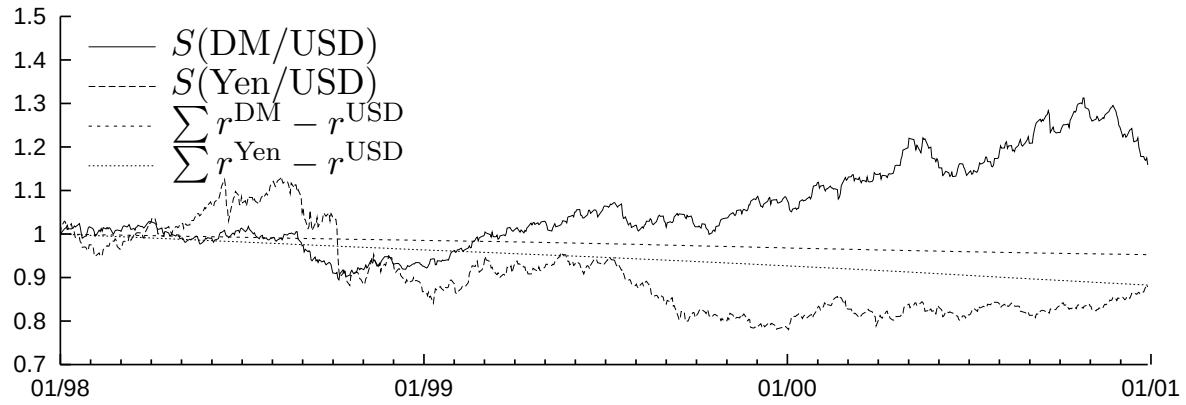

Figure 5.4: Cumulative returns on the DM/USD and Yen/USD exchange rates, and cumulative DM/USD and Yen/USD interest rate differentials, over the evaluation period 1/1/1998–29/12/2000

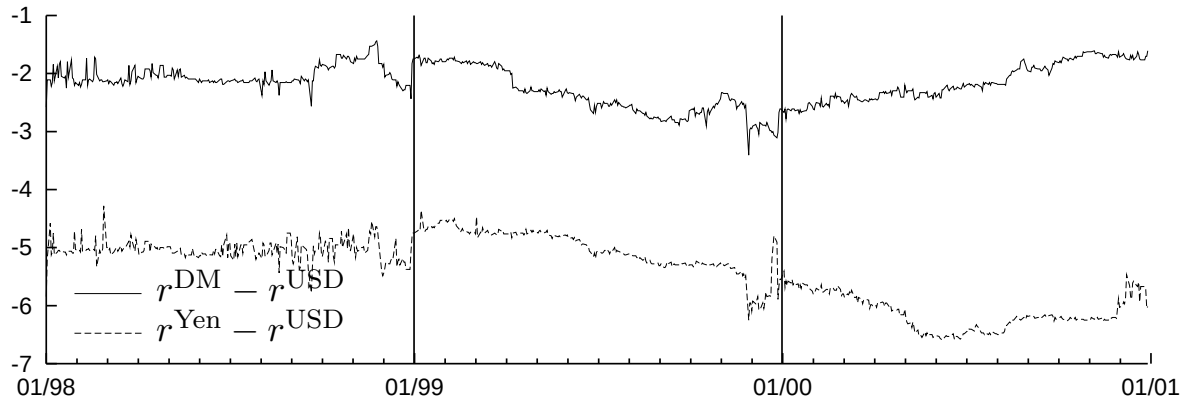

Figure 5.5: The interest rate differentials over the evaluation period

For Japan, one large drop in the exchange rate in October 1998 springs to the eye, with a second volatile period in June of the same year. Furthermore, the Yen slowly

depreciates, coincidently ending up at a cumulative loss equal to the cumulative interest rate differential. The interest rate differential between Japan and the U.S. is around 6% (on a yearly basis), indicating that hedging the exchange rate risk is only sensible if the expected loss on a certain day is higher than this 6%.[9]

The exchange rate between Germany and the U.S. promises to make a more interesting case for evaluating the hedging decision: After a year with hardly any movement for nine months and short periods with sharp losses and (less sudden) gains, the second year seems to deliver an almost steady rise of the value of the Dollar. In 2000, more fluctuation is seen, with higher volatility than before. The interest rate differential is small in size compared to the return on the exchange rate; a good hedging strategy should work reasonably well in each of these periods of relative calm, steady growth and erroneous appreciations and depreciations.

In the hedging decision, a choice is made between the (uncertain) return on the exchange rate and the (certain, at the time of taking the decision) interest rate differential. This difference between the interest rates is drawn in figure 5.5. The changes from one calendar year to the other are clearly influential for the interest rate differential, whatever the reason might be. The differential between U.S. and Japanese interest rates is huge, at 5–6.5%: This differential acts as a clear incentive either to buy dollars, or in our case not to hedge the Yen against depreciation of the dollar. In section 5.7 we will indeed establish that it is hard to find periods with enough evidence to take a sure loss of 5% on a yearly basis by hedging the Yen/USD currency risk, or vice versa to run the risk of a Yen depreciation expressed in dollar terms when a sure 5% return can be made by hedging this risk.

5.6 Convergence of MCMC and posterior results

5.6.1 The posterior distribution

For all the models **A**-**H** the Gibbs sampler was used. In our implementation we ran the sampler for a burn-in period of 50,000 iterations, and continued for another 500,000 iterations for constructing a sample. As high correlation is to be expected in a Gibbs chain, we use only one out of every 50 drawings, resulting in a final sample of 10,000 elements. As data we used the first 16 years of the data set, leaving the years 1998–2000 for evaluating the hedging decision. Separate runs of the Gibbs sampler were run for the DM/USD and Yen/USD data, and for each of the models **A**-**H**.

The correlation in a Gibbs chain with a larger number of parameters, when they are sampled in multiple blocks, can be quite high (see Kim et al. 1998). Already with a Local Level model, when the ratio between the variance parameters is not close to 1, strong correlation can result. This problem is aggravated by the inclusion of second order parameters in e.g. the stochastic volatility equation. Figure 5.6 shows the autocorrelation function of the drawings for the GLL-Stochastic Volatility model, sampling over the data

[9]The argument made here is a slight simplification, as the complete utility function has to be taken into account. However, in section 5.7 we find that indeed there is not much use in hedging the Yen/USD exchange rate, and vice versa that there is not much sense in running the risk of the USD/Yen exchange rate with such a (positive) interest rate differential to be gained at no risk.

concerning the DMark/US Dollar exchange rate. It is seen that only after about 30 drawings, correlation dies out.[10]

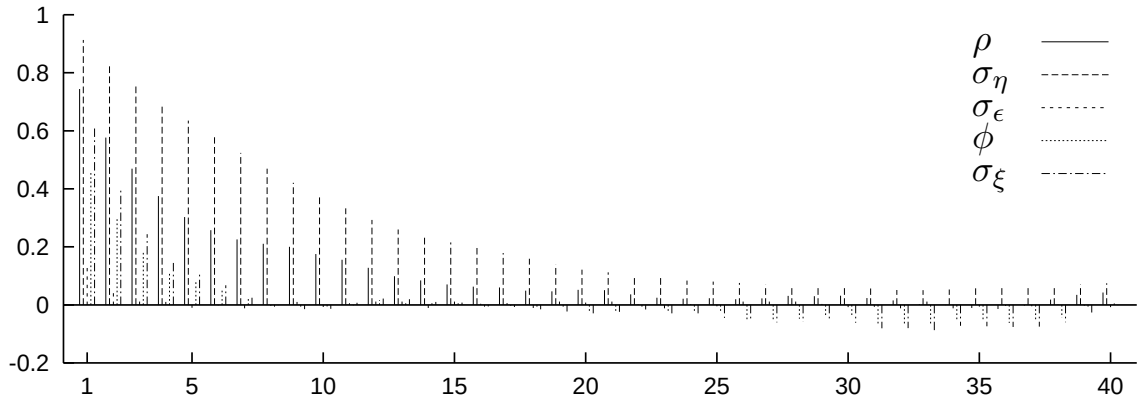

Figure 5.6: Autocorrelation function of drawings from the parameters of the GLL-Stochastic Volatility model, for the DM/USD data

The correlation in the sample influences the amount of information available in the posterior. A measure of the effective size of the posterior is the relative numerical efficiency (RNE, see Geweke (1992) and section 4.2.7). We calculated both the direct variance of the posterior, and compared it with a correlation-consistent estimate of the variance. Using the Newey-West variance estimator (Newey and West 1987), adjusting for correlation with lags up to 30 periods, we find values for the RNE of mostly around 30% for most parameters and models, with a minimum of around 10% for the parameter σ_ϵ in case of the GLL-GARCH-Student-t model on Yen/USD data. For this model, some extra attention for the (joint) sampling of parameters σ_ϵ, δ and α could help in lowering correlation and increasing the RNE, see also the discussion on improving sample quality with the Gibbs sampler in section 4.5.4.

The main characteristics of the posteriors are summarized in tables 5.2a and 5.2b for the data on the German/US Dollar exchange rate and in tables 5.3a and 5.3b for the model using Yen/US Dollar data. For each model and for each parameter, the mean, standard deviation (in parentheses), mode (on the second line) and the bounds of the 95% highest posterior density region[11] (between square brackets) are reported. The last rows of the tables indicate two implementations of the signal-to-noise ratio. The first of these two, labelled S/N, is calculated as the ratio between the unconditional variance of the signal μ_t, $\frac{\sigma_\eta^2}{1-\rho^2}$, and of the noise ϵ_t. The second is the measure $q_H = \sigma_\eta^2/\sigma_\epsilon^2$, which is the signal-to-noise ratio as defined by Harvey (1989, p. 68) in the case of the (non-stationary) Local Level model.[12]

[10]Note that the correlation only dies out after 30 drawings in the remaining chain, which resulted from skipping 49 out of 50 drawings of the original chain. The original chain therefore indeed had very strong correlation.

[11]All 95% HPD regions were continuous.

[12]By construction the S/N ratio is zero for the White Noise model, and infinite for the Local Level model.

Table 5.2a: Posterior results for the DM/USD exchange rate returns

Parameter	WN		LL		GLL	
$c \times 100$	-0.29	(0.93)				
	-0.34	[-2.20,1.45]				
ρ					0.71	(0.13)
					0.77	[0.45,0.92]
$\sigma_\eta \times 10$			0.23	(0.02)	0.65	(0.19)
			0.23	[0.19,0.28]	0.56	[0.32,1.03]
σ_ϵ	0.68	(0.01)	0.67	(0.01)	0.67	(0.01)
	0.68	[0.66,0.69]	0.67	[0.66,0.69]	0.67	[0.65,0.69]
S/N $\times$ 100	0		∞		2.14	
$q_H \times$ 100	0		0.12		1.02	

Note: For each parameter the mean and standard deviation (between parentheses) are reported on the first row. The second row contains the posterior mode and the 95% highest posterior density region, between brackets. S/N and q_H are two implementations of the signal-to-noise ratio.

Concerning the posteriors for the models using the DM/USD data the following remarks can be made. From table 5.2a it is seen that the posteriors of the two parameters of the White Noise model are very tight, with the mean and standard deviation centered at the corresponding moments of the dataset. Also the LL model, which is sparsely parameterized, results in tight posteriors, with a parameter σ_η governing the variance of the varying mean process sampled at a value of 0.023. The standard deviation of the observation disturbance, σ_ϵ, is 30-fold larger at 0.67. Note that the variance of the signal μ_t is 0 for the WN model, and infinite for the I(1) process in the LL model.

More interesting in table 5.2a are the posteriors for the GLL model. The density of the observation standard deviation σ_ϵ hardly changes, but there is more movement in the mean process, indicated by the larger σ_η. Both parameters ρ and σ_η have a mode not very close to the mean, indicating skewness of the posterior densities. The signal-to-noise ratio is low at 0.0214. This corresponds with the findings of very little autocorrelation in the series, as seen from bottom-left panel of figure 5.2.

The skewness of the posterior of ρ and σ_η can also be observed for other models. In figure 5.7 the marginal posteriors of the parameters of the GLL-Stochastic Volatility model are plotted, together with the priors and the 95% HPD regions. Apart from the skewness of parameters ρ and σ_η, it is seen that the posteriors are somewhat more concentrated than the priors. The HPD region for the parameter ρ is wide, especially in view of the fact that the data set used in sampling comprises almost 4,200 datapoints. We note that $\rho = 0$ (WN) and $\rho = 1$ (LL) are not in the HPD interval, for all the models where ρ is not fixed.

The contrast between the posterior of ρ and of the GARCH parameters (both in the GLL-GARCH and the GLL-GARCH-Student-t model) is large. Both δ and α are estimated quite precisely, with tight and almost symmetric posterior densities. A similar effect is found for the parameters ϕ and σ_ξ in the GLL-SV model, which are also empirically well identified. Including the Student-t disturbances in the GARCH model does

Table 5.2b: Posterior results for the DM/USD exchange rate returns (continued)

Parameter	GLL-GARCH		GLL-SV		GLL-Student-t		GLL-GARCH-Student-t		GLL-SV-Student-t	
ρ	0.84	(0.07)	0.72	(0.13)	0.67	(0.14)	0.79	(0.11)	0.66	(0.14)
	0.87	[0.70,0.94]	0.79	[0.47,0.92]	0.74	[0.40,0.91]	0.85	[0.56,0.95]	0.73	[0.38,0.90]
$\sigma_\eta \times 10$	0.61	(0.14)	0.59	(0.16)	0.55	(0.13)	0.52	(0.11)	0.57	(0.15)
	0.57	[0.35,0.90]	0.54	[0.32,0.90]	0.50	[0.32,0.81]	0.49	[0.32,0.75]	0.52	[0.31,0.88]
σ_ϵ	0.67	(0.02)	0.60	(0.02)	0.69	(0.02)	0.74	(0.05)	0.61	(0.02)
	0.67	[0.63,0.72]	0.59	[0.56,0.63]	0.69	[0.66,0.73]	0.73	[0.65,0.85]	0.61	[0.57,0.66]
δ	0.90	(0.01)					0.92	(0.01)		
	0.90	[0.88,0.92]					0.92	[0.90,0.93]		
$\alpha \times 10$	0.67	(0.07)					0.65	(0.07)		
	0.66	[0.53,0.82]					0.64	[0.51,0.79]		
ν					4.30	(0.33)	5.20	(0.42)	8.48	(1.56)
					4.22	[3.66,4.97]	5.11	[4.39,6.05]	7.84	[5.83,11.67]
ϕ			0.92	(0.02)					0.95	(0.01)
			0.92	[0.89,0.95]					0.95	[0.92,0.96]
σ_ξ			0.29	(0.03)					0.23	(0.02)
			0.28	[0.23,0.35]					0.22	[0.19,0.27]
S/N $\times$ 100	3.07		2.30		1.32		1.55		1.78	
$q_H \times$ 100	0.87		1.06		0.66		0.53		0.94	

Note: See table 5.2a for an explanation of the entries in the table.

not alter the posterior of the GARCH parameters δ, α greatly. Only the standard deviations σ_η and σ_ϵ change, as the Student-t disturbance takes up part of the tails of the density. The resulting change in the S/N ratio is interesting: Due to the heavy tails of the Student-t density in the GLL-GARCH-Student-t model, the S/N ratio is only 0.0155, which is small compared to the value of 0.0307 for the GLL-GARCH model. A similarly small value of the S/N ratio is found for the GLL-Student-t model. The inclusion of Student-t disturbances into the Stochastic Volatility model seems to be less supported by the data than their combination with the GARCH model. Where the pure GLL-Student-t model has $\nu = 4.3$ degrees-of-freedom, which increases to $\nu = 5.2$ for the GARCH model, the combination GLL-SV-Student-t leads to a posterior mean of $\nu = 8.5$ and a 95% HPD region stretching out to $\nu = 11.7$. Though such a value of ν does not yet indicate normal disturbances, it also does not imply the heavy tails of a $t(4)$ density. A side effect of including the Student-tailed disturbance density with the Stochastic Volatility model is that the correlation structure of the mean process is slightly less clear: This model has the widest HPD region for ρ, and lowest posterior mean and mode.

The returns on the Yen/USD exchange rate are rather similar in distribution to the DM/USD returns. The standard deviation σ_ϵ takes on values of around 0.65 as before. More interesting are the different results for the parameter ρ: For the GLL model more correlation between successive states is found than before, with a posterior mode for ρ of 0.89 even to the right of the prior mode of 0.8 (see table 5.1). The same effect occurs with the GLL-GARCH model, but for the other models on the Yen/USD data, lower values of

Table 5.3a: Posterior results for the Yen/USD exchange rate returns

Parameter	WN		LL		GLL	
$c \times 100$	0.08	(0.90)				
	0.11	[-1.65,1.88]				
ρ					0.82	(0.10)
					0.89	[0.61,0.95]
$\sigma_\eta \times 10$			0.24	(0.02)	0.60	(0.17)
			0.24	[0.20,0.29]	0.53	[0.32,0.95]
σ_ϵ	0.65	(0.01)	0.64	(0.01)	0.64	(0.01)
	0.65	[0.63,0.66]	0.64	[0.63,0.66]	0.64	[0.62,0.65]
S/N $\times$ 100	0		∞		3.13	
$q_H \times$ 100	0		0.15		0.97	

Note: See table 5.2a for an explanation of the entries in the table.

Table 5.3b: Posterior results for the Yen/USD exchange rate returns (continued)

Parameter	GLL-GARCH		GLL-SV		GLL-Student-t		GLL-GARCH-Student-t		GLL-SV-Student-t	
ρ	0.88	(0.05)	0.57	(0.17)	0.70	(0.16)	0.70	(0.16)	0.55	(0.16)
	0.90	[0.78,0.96]	0.55	[0.26,0.89]	0.85	[0.38,0.96]	0.84	[0.39,0.95]	0.56	[0.24,0.86]
$\sigma_\eta \times 10$	0.54	(0.11)	0.47	(0.10)	0.48	(0.11)	0.46	(0.09)	0.49	(0.11)
	0.51	[0.35,0.77]	0.44	[0.29,0.68]	0.44	[0.30,0.70]	0.41	[0.30,0.64]	0.44	[0.30,0.72]
σ_ϵ	0.64	(0.02)	0.53	(0.01)	0.69	(0.02)	0.78	(0.07)	0.58	(0.02)
	0.64	[0.61,0.68]	0.52	[0.50,0.55]	0.68	[0.64,0.73]	0.76	[0.65,0.92]	0.59	[0.54,0.63]
δ	0.90	(0.01)					0.92	(0.01)		
	0.90	[0.87,0.92]					0.92	[0.90,0.94]		
$\alpha \times 10$	0.62	(0.08)					0.60	(0.08)		
	0.62	[0.47,0.77]					0.59	[0.45,0.76]		
ν					3.50	(0.23)	4.00	(0.24)	5.80	(0.73)
					3.49	[3.05,3.99]	3.95	[3.54,4.48]	5.65	[4.45,7.23]
ϕ			0.72	(0.06)					0.92	(0.02)
			0.73	[0.61,0.83]					0.92	[0.88,0.95]
σ_ξ			0.65	(0.07)					0.29	(0.03)
			0.64	[0.50,0.80]					0.28	[0.22,0.36]
S/N $\times$ 100	3.45		1.37		1.20		0.86		1.14	
$q_H \times$ 100	0.74		0.85		0.52		0.37		0.74	

Note: See table 5.2a for an explanation of the entries in the table.

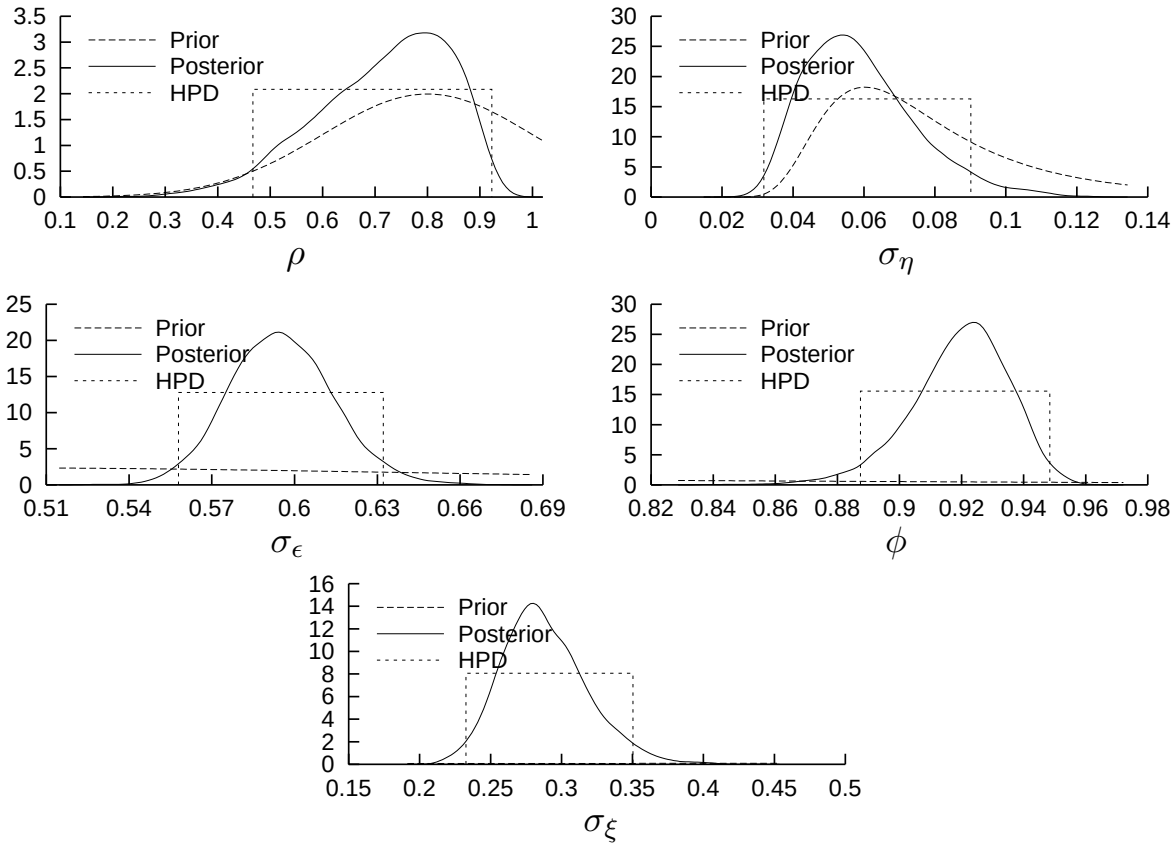

Figure 5.7: Prior, posterior and HPD region of the parameters in the GLL-Stochastic Volatility model, for the DM/USD exchange rate returns

ρ are found than in the case of the DM/USD data. The findings on ρ translate to results for the signal-to-noise ratio S/N. For the GLL and GLL-GARCH models the Japanese data appears to give a clearer signal of where the exchange rate is going. For the other models, even less signal is found than in the case of the exchange rates between Germany and the United States.

Comparing results for the models with varying variance, there is a striking difference between posterior densities for the GARCH parameters and for the SV parameters. Where the GARCH parameters δ and α are tightly estimated, for the SV parameters ϕ and σ_ξ wide HPD regions are found. Also the correlation parameter δ takes on a value of 0.9, indicating strong GARCH effect, whereas the SV model ends up with a parameter $\phi = 0.72$, with a large standard deviation σ_ξ of 0.65. In order to check whether indeed the Gibbs sampling chain has converged, figure 5.8 depicts the successive drawings in the chain for the parameter ϕ of the GLL-SV model on the Yen/USD data, together with the running mean and the mean of a moving window of the last 100 sampled parameter values. From the plot, no problems with convergence are apparent. This plot is typical for results for other parameters of this model, and also of the other models.

Including Student-t disturbances with the SV model helps getting a more precise estimate of the parameter ϕ. Note that the degrees-of-freedom parameter is $\nu \approx 3.5, 4.0$ and 5.8 for the models **F**-**H** with the Yen/USD data, comparing to values of $\nu \approx 4.3, 5.2$

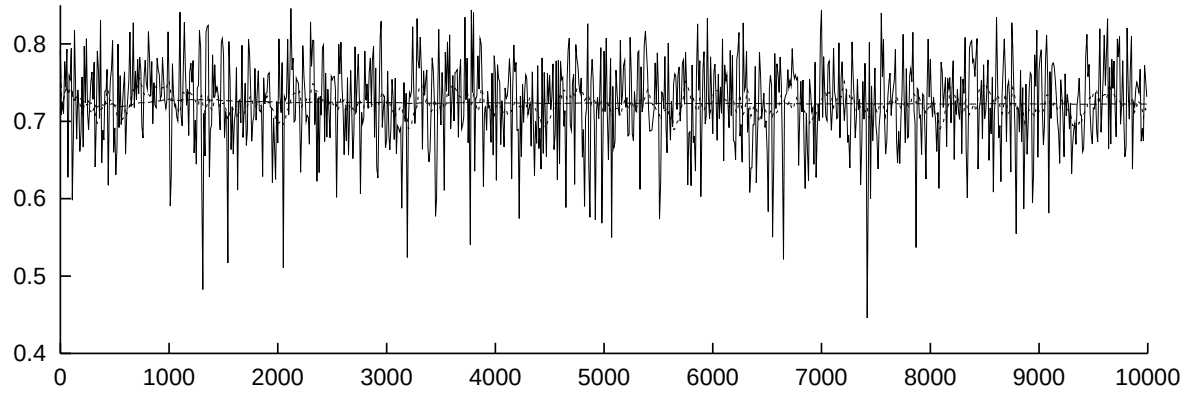

Figure 5.8: Successive drawings of the parameter ϕ for the GLL-Stochastic Volatility model for Yen/USD returns, with the running mean and the mean of a moving window of the last 100 sampled parameter values

and 8.5 for the DM/USD data. This indicates that the Yen/USD returns exhibit heavier tails than the DM/USD returns.

5.6.2 Marginal likelihood

For models **A**-**D**, both the methods based on the posterior kernel function of section 4.4.2 as the methods using the Gibbs' conditional densities (section 4.4.2 can be used for computing the marginal likelihood. Models **E**-**H** need data augmentation to get to a tractable model, and therefore do not allow for application of the kernel methods. Table 5.4 reports the marginal loglikelihoods (ML) for the models **A**-**H**, both for the DM/USD (first two columns) and for the Yen/USD returns (columns three and four). In columns one and three the results using the kernel method for models **A**-**D** are given, calculated using the LaPlace approximation to the posterior kernel at the location of the posterior mode.

Comparing the results for the first three models, we see that the Local Level model for exchange rate returns is not favoured by the data. Between models **A** and **C** there is not much of a difference. Even though the 95%-HPD region of ρ for the Generalized Local Level model does not include 0, which would imply that the GLL model would simplify to the White Noise model **A**, the marginal likelihoods are very close. Based on the data, there is no clear reason to prefer one model above the other.

The modelling steps on the varying variance structure (allowing for GARCH or SV in the GLL model) lead to a substantial improvement in the marginal likelihood over the more basic WN or GLL models. For the DM/USD exchange rate, the data seems to be best modelled using a Stochastic Volatility framework. The GLL-Student-t model is also quite an improvement over the WN and GLL models, with a gain in ML of 160 points, getting close to the GLL-GARCH model. When the GARCH and Student-t blocks are combined in model **G**, the resulting marginal loglikelihood of -4040 is close to the value found for the GLL-SV model, -4033. Including a Student-t disturbance distribution with the GLL-SV model is indicated as not supported by the data, as the ML value drops

Table 5.4: Marginal loglikelihoods

		DM/USD returns		Yen/USD returns	
Model		LaPlace	Gibbs	LaPlace	Gibbs
A	WN	−4306.12	−4306.43	−4116.70	−4117.01
B	LL	−4347.04	−4346.83	−4148.55	−4148.45
C	GLL	−4306.07	−4305.81	−4114.57	−4114.24
D	GLL-GARCH	−4139.37	−4140.34	−3981.39	−3982.34
E	GLL-SV		−4032.85		−4248.57
F	GLL-Student-t		−4146.75		−3876.23
G	GLL-GARCH-Student-t		−4040.37		−3779.72
H	GLL-SV-Student-t		−4037.84		−3844.35

Note: Reported are the marginal loglikelihoods calculated using the LaPlace method at the location of the posterior mode (columns 1 and 3, only available for models **A**-**D**) and using the Gibbs' conditional densities method.

5 points. This corresponds to the findings from the posterior density for ν in model **H** (section 5.6.1), where the 95%-HPD region ranged from 5.8 to 11.7 as opposed to a range of 3.7 to 5.0 for the GLL-Student-t model.

For the Yen/USD data, the correlation structure of the second moment of the data is not so strong, and the Student-t model has the best ML of models **D**-**F**. Combining it with the GARCH structure for the variance leads to a further improvement to a marginal loglikelihood of -3780. The Stochastic Volatility model does not fit this data set well, it performs even worse than the basic White Noise, Local Level or Generalized Local Level models. Combining Student-t disturbances with Stochastic Volatility behaviour is better than models **A**-**C**, the GLL-SV-Student-t model ends up according to the ML score as second best behind the model combining GARCH and Student-t disturbances.

5.6.3 Predictive density of the models

The hedging decision taken by the investor is based solely on the information contained in the predictive density $p(s_{t+1|t})$ corresponding to the model. We can distinguish two elements which are of importance for the predictive power of a model.

On the one hand, an investor would like to see strong predictive power in the first moment, in order to know what tomorrow's return will be. Predictive power in this sense is lacking in the WN model (as tomorrow's predictive return equals the global mean c of the series), and strong for the LL model, predicting tomorrow's return with today's μ_t, which follows a random walk. The GLL-based models take up positions in between. Their signal-to-noise ratios S/N (see tables 5.2a-5.3b) provide an indication of the predictive power we can expect.

On the other hand the uncertainty of the prediction is important. If the model is able to distinguish between tranquil periods and periods of larger fluctuations, the investor can take measures accordingly. On this aspect, the GARCH and SV models are expected to behave better.

It is instructive to look at the predictive densities in detail. Figure 5.9 contains the predictive density of the GLL-SV model on the DM/USD returns, which had the

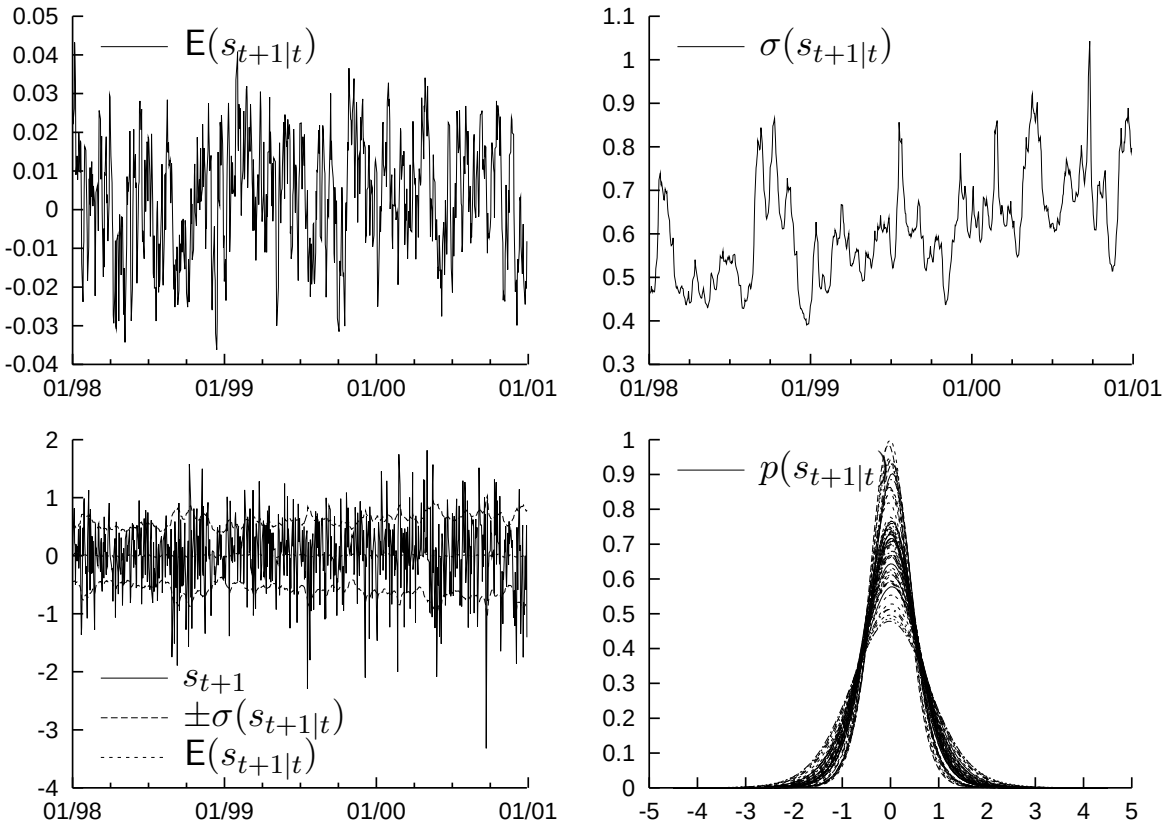

Figure 5.9: Predictive mean, standard deviation, together with observations and the full density (top left to bottom right) for the GLL-SV model on DM/USD returns, in percentage changes, over the period 1/1/98-29/12/00

best marginal loglikelihood score for this data set. Using the Yen/USD data, the GLL-GARCH-Student-t model was preferred by the marginal likelihood criterion. It results in a predictive density as in figure 5.11.

For the German data, the following remarks can be made. The top-left panel of figure 5.9 displays the mean $\mathsf{E}(s_{t+1|t})$ of the predictive density $p(s_{t+1|t})$. In our models $\mathsf{E}(s_{t+1|t})$ equals the prediction of the unobserved state μ_{t+1}. On average, the mean prediction is around zero, but with clear distinctions from period to period. Around September 1998, the continuing decline in the exchange rate (refer back to figure 5.4 for a plot of the exchange rate returns in the evaluation period) has its effect on the predictive mean, whereas in most months in 1999 $\mathsf{E}(s_{t+1|t})$ is positive. Throughout the year 2000, the prediction still is positive on the average, though with slightly less conviction than in 1999. Note that on the axis, the changes are indicated as daily percentages; though the changes from day to day are noticeable, the predictions show a maximum daily change of 0.04%.

In the top-right panel of figure 5.9, the standard deviation of the prediction is given. Around September 1998, when the predicted change in exchange rate becomes negative, the standard deviation jumps up. From that moment onwards the volatility remains high, until January 1999, where the stochastic volatility component indicates that the variance

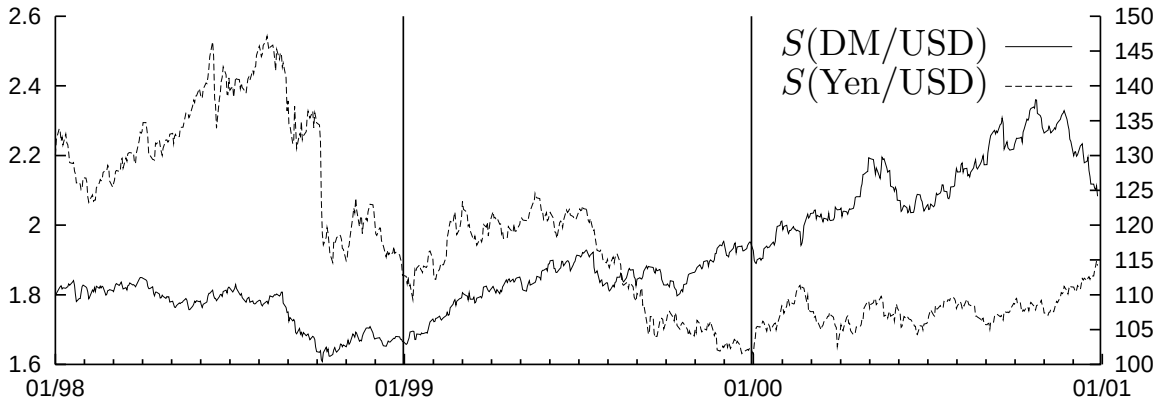

Figure 5.10: Exchange rates over the evaluation period 1/1/98-29/12/00

of the series diminishes again to the levels of mid-1998. Throughout 1999, the volatility does not change much, but the model indicates that the year 2000 was definitely riskier than previous years, as the volatility has a higher base-value and reaches its peak in this year 2000. The jumps in the standard deviation only occur in models **D**, **E**, **G** and **H**, which allow for GARCH or Stochastic Volatility. For the other models the standard deviation is constant.

The bottom-left panel of the figure indicates the uncertainty involved in predicting tomorrow's appreciation or depreciation. In the graph we plotted the mean prediction from the top-left panel plus and minus one standard deviation, together with the actual exchange rate returns. From the graph we see that the predictions are very small compared to the actual returns. The bottom-right panel depicts the shape of the predictive densities $p(s_{t+1|t})$ for (a selection of) the days of the evaluation period. It is seen that the spread of the density changes considerably, the location hardly moves. For models **A**-**C** and **F**, the corresponding plot shows less variation over time as the variance is fixed.

Figure 5.11 shows the predictive density for the GLL-GARCH-Student-t model on the Yen/USD returns. The most striking element of these panels is the big increase in volatility in June and especially October 1998, connected to the large drops of 4.3 and 6.6% in the exchange rate in single days. For the remainder of the evaluation period, the volatility does not change much; with these plots in mind it is not so strange that the GLL-SV model without Student-t disturbances did not give tight posterior densities for the volatility parameters ϕ and ξ, see table 5.3b. It is more amazing that for the GLL-GARCH model tight posterior densities for δ and α were found.

The amount of predictability of the first moment is reflected in the scale of the y-axis of the first panel of figures 5.9 and 5.11. Both for the GLL-SV model on the German data as for the GLL-GARCH-Student-t model on Japanese data, a range of approximately $[-0.03\%, 0.04\%]$ is found. By definition, the White Noise model does not change its prediction, and thus has zero range. With the Yen/USD data, the Local Level model reacts most strongly to jumps in the exchange rate, leading to projections of tomorrow's returns ranging from -0.55% to +0.29%. The GARCH model claims to have quite some predictive power, resulting in a range for $\mathsf{E}(s_{t+1|t})$ of $[-0.12\%, 0.12\%]$. The other models show less extreme predictions, with the SV-based models not predicting a change larger

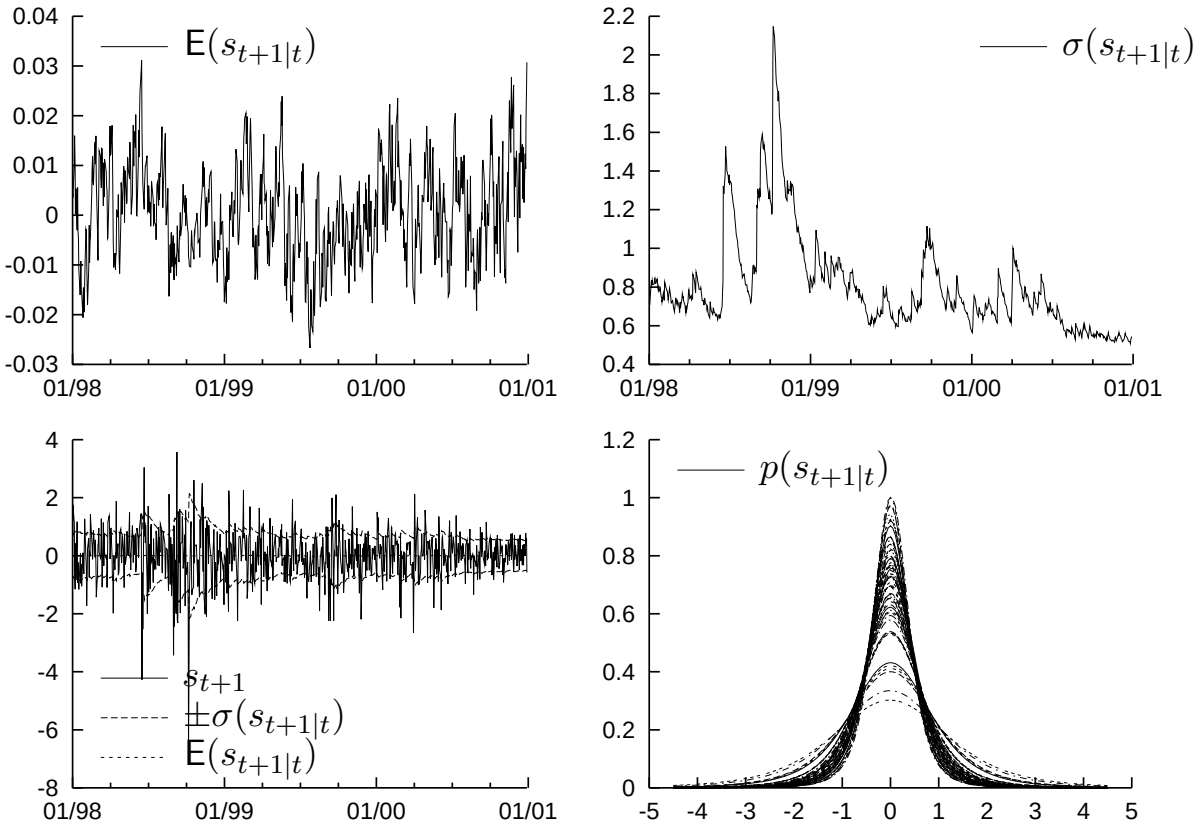

Figure 5.11: Predictive mean, standard deviation, together with observations and the full density (top left to bottom right) for the GLL-GARCH-Student-t model on Yen/USD returns, in percentage changes, over the period 1/1/98-29/12/00

than 0.011% in absolute value.

The German data leads to predictions with a maximum of $|\mathsf{E}(s_{t+1})| = 0.24$ for the LL model, returns of $\pm 0.1\%$ for the GARCH model and a range of $[-0.04\%, 0.05\%]$ for the other models.

In the present section, and in sections 5.5 and 5.6.1 we described results for the DM/USD and Yen/USD exchange rates. As the models we use (see section 5.3) are symmetric apart for the sign of the constant applied in the White Noise models, we can also use the predictive densities presented here to construct a hedging strategy for the USD/DM and USD/Yen data. In the following section we investigate whether the predictive densities provide sufficient information for constructing effective currency overlay strategies, for the four different situations of hedging the DM/USD, Yen/USD, USD/DM or Yen/USD exchange rates.

5.7 Hedging results

5.7.1 On the setup

After the description of the data and the interest rate differential over the evaluation period in section 5.5, the posterior density of the model parameters in section 5.6.1, and the computation of the predictive density of the exchange rate returns in section 5.6.3, we continue in the present section with describing the hedging decisions and resulting utility and returns.

We analyse the hedging decision for four situations:

i. A Germany-based investor with investments denoted in the USD currency (DM/USD for short, with results to be discussed in table 5.5)

ii. The inverse situation, of an American firm investing in Germany (USD/DM, table 5.6)

iii. Hedging a dollar currency risk from Japan (Yen/USD, table 5.7)

iv. And the inverse: Hedging against depreciation of the yen vis-a-vis the dollar (USD/Yen, table 5.8)

These four positions for which hedging can be used are evaluated over the period 1998-2000. Results in this chapter are reported for the whole three-year period and for each of the years separately. Numerical results are presented for a risk-tolerant investor, with a log-utility function (corresponding to a power utility function 5.4 with $\gamma = 0$) and for a risk-averse risk manager, with risk tolerance parameter $\gamma = -10$. Other positions of risk tolerance are discussed briefly, in section 5.7.5.

5.7.2 Naive hedging strategies

Before we continue with a description of the model based hedging decisions, we first concentrate on three naive hedging strategies. Obviously, the investor can choose to hedge nothing, with as a resulting return the return on the exchange rate. Alternatively, the risk can be hedged over the complete period, taking hedging position $H_t = 1, t = 1, \ldots, T$, with the cumulative interest rate differential as the result of this scheme. A third straightforward hedging scheme can be denoted as the Random Walk scheme: Hedge tomorrows return if today a depreciation occurred, i.e.

$$H_{t+1} = \begin{cases} 1 & \text{if } S_t < S_{t-1}, \\ 0 & \text{else.} \end{cases} \tag{5.20}$$

The returns from the first two schemes could be derived from figure 5.4 with the cumulative return of the exchange rate and the cumulative interest rate differential. In the first three rows of tables 5.5-5.8 numerical results are presented. For the complete three-year period, the cumulative return on the DM/USD exchange rate was 14.73% (fifth column, labelled r). A risk-tolerant investor values profits as much as losses, and therefore utility is equal

to the return at 14.73%.[13] When the manager does not tolerate such risks, we see in the first row of the second panel of table 5.5 that the utility of this scheme drops to -0.14: As the return of 14.73% comes with large shocks, with a larger penalty for sudden (large) depreciations, the utility derived from the no-hedge scheme is not high. Note that in table 5.6 the home and foreign countries are switched, such that the no-hedge case leads to a loss of 14.73%. The derived utility for the risk-intolerant investor was -29.27%.

Other columns in the table indicate the average hedge position $\overline{H}$, the number of times an extreme position of $H_t = 0$ or $H_t = 1$ is taken, and also the average switch in position $\overline{|\Delta H|}$ that is made. For the no-hedge and the full-hedge scheme these statistics are of little interest, as we can see that the full 782 days the same extreme position is taken, leading to an average change in position of zero. For the investor who is not taking any risks, hedging fully throughout the evaluation period, the value of the risk tolerance parameter γ does not make much of a difference for the derived utility. The daily interest rate differential is fairly small; Only for the USD/Yen case, in panel 2 of table 5.8, a utility of 11.73 is found compared to a return of 11.74%.

While for the DM/USD exchange rate the uncovered interest rate parity does not seem to hold over the evaluation period, for the Yen/USD exchange rate it seems to be valid approximately. By coincidence, for the three year period the return on the exchange rate, -12.97%, and the cumulative interest rate, -11.74%, are close. Breaking up the evaluation period in separate years (columns 7 and further in the tables) we see that this only occurs for the complete period.

The Random Walk hedging scheme can be expected to work if there is some positive correlation to be exploited. This scheme is nice for comparison with the model based schemes, but it does not have much practical importance: As the exchange rate is jumping up and down (see also the next section) almost every other day, the RW scheme changes its position often, on the average every 2.2 (DM/USD) or 2.0 (Yen/USD) days.[14] No firm would be willing to change its entire hedging position so often and so randomly. For simplicity we are assuming that the firm does not incur transaction costs. For large firms this is an acceptable approximation, but only as long as the position is not changing frequently.

The RW method of hedging is superior in returns and utility to the no-hedge strategy in two out of three years. Though the RW strategy manages to stay out of the market on some days of depreciation, the risk is still considerable, as can be seen from the large penalty on the utility when the risk tolerance parameter is changed from 0 to -10.

5.7.3 Variability of the hedging decision

The model-based strategies change the hedging position less often than the RW hedge strategy, see column 4 of the rows labelled WN - GLL-SV-Student-t in tables 5.5–5.8. For the risk-tolerant case, in the top panel of the tables, it is often optimal to choose a

[13] The table reports the sum of the continuously compounded returns as in (5.2). The three-year increase of the DM/USD rate of almost 16% mentioned in section 5.2 is retrieved by computing $\tilde{r} = \exp(r/100) \times 100$. This manner of presenting the results is chosen for comparison to the utilities, which are multiplied by 100.

[14] The average duration before the complete position is switched can be calculated as $D = 1/\overline{|\Delta H|}$.

Table 5.5: Hedging results for a German firm with exposures in the U.S. Dollar, DM/USD

Model	$\overline{H}$	$H=0$	$H=1$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U
$\gamma = 0$			1998–2000						1998				1999				2000	
No hedge	0.00	782	0	0.000	14.73	14.73	0.00	0.000	−7.65	−7.65	0.00	0.000	15.83	15.83	0.00	0.000	6.55	6.55
Full hedge	1.00	0	782	0.000	−4.75	−4.75	1.00	0.000	−1.47	−1.47	1.00	0.000	−1.72	−1.72	1.00	0.000	−1.55	−1.55
RW Hedge	0.45	431	351	0.452	26.38	26.38	0.50	0.515	−3.60	−3.60	0.42	0.427	11.17	11.17	0.42	0.413	18.81	18.81
WN	0.02	612	0	0.005	14.62	14.62	0.01	0.007	−8.35	−8.35	0.01	0.004	15.47	15.47	0.02	0.005	7.50	7.50
LL	0.32	530	244	0.057	18.92	18.92	0.54	0.085	−6.16	−6.16	0.21	0.042	12.68	12.68	0.20	0.044	12.40	12.40
GLL	0.27	532	176	0.168	21.23	21.23	0.32	0.193	−3.21	−3.21	0.20	0.135	12.79	12.79	0.30	0.174	11.65	11.65
GLLGA	0.36	486	256	0.157	15.68	15.68	0.45	0.156	−2.80	−2.80	0.26	0.150	10.24	10.24	0.35	0.165	8.24	8.24
GLLSV	0.28	533	182	0.175	21.54	21.54	0.37	0.207	−1.94	−1.94	0.19	0.139	13.04	13.04	0.26	0.180	10.45	10.45
GLLT	0.21	575	129	0.180	19.31	19.31	0.28	0.244	−4.82	−4.82	0.14	0.140	12.03	12.03	0.21	0.156	12.10	12.10
GLLGAT	0.27	548	185	0.154	18.82	18.82	0.39	0.180	−2.81	−2.81	0.18	0.148	10.31	10.31	0.24	0.136	11.32	11.32
GLLSVT	0.23	558	136	0.185	19.27	19.27	0.31	0.248	−3.26	−3.26	0.16	0.143	12.73	12.73	0.21	0.167	9.79	9.79
$\gamma = -10$			1998–2000						1998				1999				2000	
No hedge	0.00	782	0	0.000	14.73	−0.14	0.00	0.000	−7.65	−11.51	0.00	0.000	15.83	11.75	0.00	0.000	6.55	−0.38
Full hedge	1.00	0	782	0.000	−4.75	−4.75	1.00	0.000	−1.47	−1.48	1.00	0.000	−1.72	−1.72	1.00	0.000	−1.55	−1.55
RW Hedge	0.45	431	351	0.452	26.38	18.41	0.50	0.515	−3.60	−5.33	0.42	0.427	11.17	8.70	0.42	0.413	18.81	15.05
WN	0.89	0	0	0.003	−2.38	−2.56	0.90	0.005	−2.17	−2.21	0.88	0.003	0.21	0.15	0.89	0.002	−0.41	−0.49
LL	0.41	377	244	0.066	13.83	5.24	0.68	0.097	−7.97	−8.88	0.28	0.050	11.52	8.77	0.26	0.051	10.28	5.35
GLL	0.77	0	183	0.104	4.16	2.92	0.83	0.088	−2.58	−2.73	0.74	0.108	1.48	1.12	0.73	0.115	5.26	4.53
GLLGA	0.58	150	260	0.154	5.99	1.32	0.66	0.151	−3.93	−4.67	0.47	0.180	3.95	2.16	0.61	0.134	5.97	3.83
GLLSV	0.66	43	184	0.152	8.75	7.01	0.68	0.167	−1.80	−2.21	0.57	0.178	5.58	4.82	0.74	0.112	4.97	4.40
GLLT	0.76	0	132	0.105	3.86	2.72	0.82	0.098	−2.52	−2.67	0.73	0.111	1.73	1.34	0.74	0.107	4.65	4.05
GLLGAT	0.60	82	188	0.157	9.29	5.44	0.67	0.164	−3.29	−3.96	0.48	0.180	5.52	3.87	0.64	0.127	7.06	5.53
GLLSVT	0.70	16	139	0.141	7.07	5.71	0.71	0.164	−1.73	−2.05	0.63	0.162	4.78	4.16	0.77	0.096	4.02	3.60

Note: Reported are the average hedge ratio $\overline{H}$, the number of occurrences of no-hedging $H = 0$ or full-hedging $H = 1$ and the average change in hedging position, $\overline{|\Delta H|}$, together with the cumulative hedged currency return r (as a percentage) and corresponding cumulative utility (times 100). Results are reported for the entire evaluation period 1998–2000 and for the three years separately. The top panel report the results for a risk-tolerant, the bottom panel for a risk-averse investor.

Table 5.6: Hedging results for a U.S. firm with exposures in the German DMark, USD/DM

Model	$\overline{H}$	$H=0$	$H=1$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U
$\gamma = 0$			1998–2000					1998				1999				2000		
No hedge	0.00	782	0	0.000	−14.73	−14.73	0.00	0.000	7.65	7.65	0.00	0.000	−15.83	−15.83	0.00	0.000	−6.55	−6.55
Full hedge	1.00	0	782	0.000	4.75	4.75	1.00	0.000	1.47	1.47	1.00	0.000	1.72	1.72	1.00	0.000	1.55	1.55
RW Hedge	0.57	337	445	0.448	14.55	14.55	0.52	0.504	4.98	4.98	0.59	0.427	−2.94	−2.94	0.59	0.417	12.51	12.51
WN	0.98	0	612	0.005	4.64	4.64	0.99	0.007	0.78	0.78	0.99	0.004	1.36	1.36	0.98	0.005	2.50	2.50
LL	0.68	244	530	0.057	8.94	8.94	0.46	0.085	2.97	2.97	0.79	0.042	−1.43	−1.43	0.80	0.044	7.40	7.40
GLL	0.73	176	532	0.168	11.25	11.25	0.68	0.193	5.92	5.92	0.80	0.135	−1.32	−1.32	0.70	0.174	6.65	6.65
GLLGA	0.64	256	486	0.157	5.70	5.70	0.55	0.156	6.33	6.33	0.74	0.150	−3.87	−3.87	0.65	0.165	3.24	3.24
GLLSV	0.72	182	533	0.175	11.56	11.56	0.63	0.207	7.19	7.19	0.81	0.139	−1.07	−1.07	0.74	0.180	5.45	5.45
GLLT	0.79	129	575	0.180	9.33	9.33	0.72	0.244	4.31	4.31	0.86	0.140	−2.08	−2.08	0.79	0.156	7.11	7.11
GLLGAT	0.73	185	548	0.154	8.84	8.84	0.61	0.180	6.32	6.32	0.82	0.148	−3.80	−3.80	0.76	0.136	6.32	6.32
GLLSVT	0.77	136	558	0.185	9.29	9.29	0.69	0.248	5.87	5.87	0.84	0.143	−1.38	−1.38	0.79	0.167	4.79	4.79
$\gamma = -10$			1998–2000					1998				1999				2000		
No hedge	0.00	782	0	0.000	−14.73	−29.27	0.00	0.000	7.65	3.87	0.00	0.000	−15.83	−19.84	0.00	0.000	−6.55	−13.31
Full hedge	1.00	0	782	0.000	4.75	4.75	1.00	0.000	1.47	1.47	1.00	0.000	1.72	1.72	1.00	0.000	1.55	1.55
RW Hedge	0.57	337	445	0.448	14.55	8.26	0.52	0.504	4.98	3.12	0.59	0.427	−2.94	−4.53	0.59	0.417	12.51	9.67
WN	1.00	0	683	0.000	4.72	4.72	1.00	0.000	1.39	1.39	1.00	0.001	1.70	1.70	1.00	0.001	1.63	1.63
LL	0.78	106	531	0.052	9.09	6.10	0.62	0.079	3.61	2.05	0.86	0.042	−0.94	−1.41	0.84	0.035	6.43	5.46
GLL	0.93	0	534	0.053	6.57	6.21	0.93	0.059	1.87	1.74	0.96	0.034	1.29	1.25	0.91	0.064	3.41	3.22
GLLGA	0.82	43	487	0.096	10.46	8.44	0.73	0.134	5.24	4.21	0.87	0.080	0.56	0.18	0.85	0.073	4.66	4.04
GLLSV	0.92	4	536	0.068	4.40	4.17	0.86	0.117	1.37	1.22	0.95	0.046	0.97	0.92	0.95	0.040	2.07	2.02
GLLT	0.96	0	579	0.035	5.16	5.05	0.95	0.048	1.35	1.30	0.98	0.021	1.24	1.23	0.96	0.035	2.57	2.51
GLLGAT	0.91	3	552	0.070	5.29	4.78	0.85	0.114	1.57	1.32	0.93	0.054	0.43	0.32	0.94	0.040	3.30	3.15
GLLSVT	0.95	1	564	0.051	4.70	4.58	0.91	0.090	1.40	1.33	0.96	0.034	1.23	1.21	0.97	0.027	2.06	2.04

Note: See table 5.5 for an explanation of the entries in the table.

Table 5.7: Hedging results for a Japanese firm with exposure to the U.S. Dollar, Yen/USD

Model	$\overline{H}$	$H=0$	$H=1$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U
$\gamma=0$			1998–2000					1998				1999				2000		
No hedge	0.00	782	0	0.000	−12.97	−12.97	0.00	0.000	−14.21	−14.21	0.00	0.000	−9.71	−9.71	0.00	0.000	10.95	10.95
Full hedge	1.00	0	782	0.000	−11.74	−11.74	1.00	0.000	−3.66	−3.66	1.00	0.000	−3.66	−3.66	1.00	0.000	−4.42	−4.42
RW Hedge	0.51	387	395	0.501	6.06	6.06	0.51	0.477	7.67	7.67	0.53	0.512	−4.85	−4.85	0.47	0.514	3.24	3.24
WN	0.00	782	0	0.000	−12.97	−12.97	0.00	0.000	−14.21	−14.21	0.00	0.000	−9.71	−9.71	0.00	0.000	10.95	10.95
LL	0.43	440	335	0.059	3.57	3.57	0.43	0.028	4.47	4.47	0.61	0.056	−3.90	−3.90	0.26	0.093	3.00	3.00
GLL	0.31	527	224	0.155	−8.30	−8.30	0.36	0.161	−3.14	−3.14	0.39	0.186	−10.91	−10.91	0.18	0.119	5.76	5.76
GLLGA	0.33	495	237	0.136	−5.44	−5.44	0.33	0.109	0.25	0.25	0.42	0.179	−10.07	−10.07	0.24	0.120	4.37	4.37
GLLSV	0.00	777	0	0.001	−12.16	−12.16	0.00	0.002	−13.40	−13.40	0.00	0.001	−9.71	−9.71	0.00	0.000	10.95	10.95
GLLT	0.11	669	68	0.103	−6.69	−6.69	0.14	0.106	−1.94	−1.94	0.18	0.174	−13.62	−13.62	0.02	0.029	8.88	8.88
GLLGAT	0.06	695	24	0.063	−18.56	−18.56	0.05	0.053	−14.48	−14.48	0.12	0.119	−13.50	−13.50	0.01	0.014	9.42	9.42
GLLSVT	0.00	778	0	0.000	−12.87	−12.87	0.00	0.001	−14.11	−14.11	0.00	0.000	−9.71	−9.71	0.00	0.000	10.95	10.95
$\gamma=-10$			1998–2000					1998				1999				2000		
No hedge	0.00	782	0	0.000	−12.97	−43.14	0.00	0.000	−14.21	−31.52	0.00	0.000	−9.71	−17.53	0.00	0.000	10.95	5.91
Full hedge	1.00	0	782	0.000	−11.74	−11.74	1.00	0.000	−3.66	−3.66	1.00	0.000	−3.66	−3.66	1.00	0.000	−4.42	−4.42
RW Hedge	0.51	387	395	0.501	6.06	−5.04	0.51	0.477	7.67	2.63	0.53	0.512	−4.85	−8.87	0.47	0.514	3.24	1.20
WN	0.61	0	0	0.004	−10.98	−15.16	0.63	0.007	−7.16	−9.39	0.63	0.003	−6.02	−7.06	0.57	0.003	2.21	1.29
LL	0.52	303	335	0.075	−5.49	−15.32	0.47	0.048	0.04	−5.53	0.69	0.074	−6.26	−8.03	0.39	0.105	0.73	−1.76
GLL	0.56	114	224	0.166	−8.54	−16.10	0.55	0.175	−4.90	−9.11	0.66	0.159	−6.24	−8.00	0.46	0.164	2.60	1.01
GLLGA	0.60	149	240	0.114	−1.69	−7.59	0.64	0.091	−0.07	−2.95	0.73	0.116	−5.98	−7.18	0.43	0.135	4.36	2.55
GLLSV	0.65	0	0	0.075	−4.64	−6.57	0.73	0.054	−1.62	−2.26	0.70	0.072	−4.79	−5.30	0.52	0.097	1.77	0.99
GLLT	0.60	4	70	0.143	−4.46	−9.59	0.60	0.146	−1.05	−3.81	0.70	0.139	−4.79	−6.01	0.51	0.145	1.38	0.23
GLLGAT	0.65	19	24	0.098	−4.91	−7.87	0.74	0.061	−0.92	−2.03	0.75	0.092	−5.91	−6.53	0.47	0.140	1.93	0.69
GLLSVT	0.68	0	0	0.062	−5.03	−6.75	0.77	0.037	−1.89	−2.38	0.73	0.057	−5.04	−5.49	0.53	0.093	1.90	1.12

Note: See table 5.5 for an explanation of the entries in the table.

Table 5.8: Hedging results for a U.S. firm with exposure to the Japanese Yen, USD/Yen

Model	$\overline{H}$	$H=0$	$H=1$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U	$\overline{H}$	$\overline{\lvert\Delta H\rvert}$	r	U
$\gamma = 0$			1998–2000						1998				1999				2000	
No hedge	0.00	782	0	0.000	12.97	12.97	0.00	0.000	14.21	14.21	0.00	0.000	9.71	9.71	0.00	0.000	−10.95	−10.95
Full hedge	1.00	0	782	0.000	11.74	11.74	1.00	0.000	3.66	3.66	1.00	0.000	3.66	3.66	1.00	0.000	4.42	4.42
RW Hedge	0.51	386	396	0.497	31.89	31.89	0.50	0.481	26.54	26.54	0.48	0.512	8.66	8.66	0.54	0.502	−3.31	−3.31
WN	1.00	0	782	0.000	11.74	11.74	1.00	0.000	3.66	3.66	1.00	0.000	3.66	3.66	1.00	0.000	4.42	4.42
LL	0.57	335	440	0.059	28.28	28.28	0.57	0.028	22.34	22.34	0.39	0.056	9.47	9.47	0.74	0.093	−3.53	−3.53
GLL	0.69	224	527	0.155	16.41	16.41	0.64	0.161	14.73	14.73	0.61	0.186	2.46	2.46	0.82	0.119	−0.78	−0.78
GLLGA	0.67	237	495	0.136	19.27	19.27	0.67	0.109	18.12	18.12	0.58	0.179	3.30	3.30	0.76	0.120	−2.16	−2.16
GLLSV	1.00	0	777	0.001	12.55	12.55	1.00	0.002	4.47	4.47	1.00	0.001	3.66	3.66	1.00	0.000	4.42	4.42
GLLT	0.89	68	669	0.103	18.02	18.02	0.86	0.106	15.93	15.93	0.82	0.174	−0.25	−0.25	0.98	0.029	2.35	2.35
GLLGAT	0.94	24	695	0.063	6.15	6.15	0.95	0.053	3.39	3.39	0.88	0.119	−0.13	−0.13	0.99	0.014	2.89	2.89
GLLSVT	1.00	0	778	0.000	11.84	11.84	1.00	0.001	3.76	3.76	1.00	0.000	3.66	3.66	1.00	0.000	4.42	4.42
$\gamma = -10$			1998–2000						1998				1999				2000	
No hedge	0.00	782	0	0.000	12.97	−15.58	0.00	0.000	14.21	−1.61	0.00	0.000	9.71	1.97	0.00	0.000	−10.95	−15.94
Full hedge	1.00	0	782	0.000	11.74	11.73	1.00	0.000	3.66	3.66	1.00	0.000	3.66	3.65	1.00	0.000	4.42	4.41
RW Hedge	0.51	386	396	0.497	31.89	14.79	0.50	0.481	26.54	16.05	0.48	0.512	8.66	4.94	0.54	0.502	−3.31	−6.21
WN	1.00	0	782	0.000	11.74	11.73	1.00	0.000	3.66	3.66	1.00	0.000	3.66	3.65	1.00	0.000	4.42	4.41
LL	0.63	236	440	0.056	24.32	9.46	0.61	0.045	16.97	7.78	0.43	0.050	9.63	5.05	0.85	0.075	−2.28	−3.37
GLL	0.84	37	527	0.099	15.98	8.14	0.79	0.115	13.10	6.86	0.80	0.124	1.93	0.65	0.94	0.056	0.95	0.63
GLLGA	0.86	24	497	0.077	9.08	6.47	0.91	0.046	5.69	5.00	0.78	0.104	3.78	2.46	0.89	0.079	−0.39	−0.99
GLLSV	1.00	0	778	0.000	11.82	11.81	1.00	0.000	3.74	3.74	1.00	0.000	3.66	3.65	1.00	0.000	4.42	4.41
GLLT	0.98	0	670	0.019	11.22	11.06	0.98	0.022	4.41	4.29	0.97	0.033	2.59	2.56	1.00	0.003	4.22	4.21
GLLGAT	0.99	0	703	0.008	10.93	10.90	1.00	0.006	3.57	3.56	0.98	0.017	3.09	3.07	1.00	0.001	4.28	4.27
GLLSVT	1.00	0	782	0.000	11.74	11.73	1.00	0.000	3.66	3.66	1.00	0.000	3.66	3.65	1.00	0.000	4.42	4.41

Note: See table 5.5 for an explanation of the entries in the table.

boundary solution $H = 0$ or $H = 1$.[15,16] Whenever the model indicates that the return on the exchange rate is expected to be larger than the interest rate differential, the risk-tolerant investor chooses not to hedge. This effect is also apparent from figures 5.12–5.13, which display the optimal hedge decision through time for the RW strategy (top panel) and the model-based strategies. Figure 5.12 shows the results for the risk-tolerant investor ($\gamma = 0$) hedging the DM/USD exchange rate, while figure 5.13 displays the decisions for the Yen/USD case with $\gamma = -10$. The wild behaviour of the RW hedge decision is immediately apparent from the figures; other hedging decisions are clearly more consistent over time.

The hedge decision of the White Noise model is only guided by the interest rate differential, as the predictive density does not change over time. For the DM/USD case, only during a period at the end of 1998, and a single day in November 2000, the (negative) interest rate differential is close enough to zero that the model decides to hedge. On other days, hedging would lead to a sure negative return, and therefore no hedging is deemed optimal. For the Yen/USD data, the difference between the interest rates is between -5% and -6.5% on a yearly basis. Unless there is strong evidence of a possible depreciation, the models will not hedge. In table 5.7 it can be seen that for $\gamma = 0$ the WN, GLL-SV and GLL-SV-Student-t models the result is that the decision $H = 0$ is taken on (almost) all days, with returns and utilities equal to the unhedged returns and utilities. With $\gamma = -10$, the investor chooses more safety, and average hedging in figure 5.13 is higher, with the WN hedging ratio closely following the movements of the interest rate differential.

The other models show more variation in the optimal decisions. The Local Level model is quite extreme in its position: Once it chooses (not) to hedge, it continues to do so for a longer period of time. The GLL model showed smaller predictions $\mathsf{E}(s_{t+1|t})$ (see the last paragraphs of section 5.6.3), and therefore changes its mind more often about the optimal hedging decision. Modelling the variance of the process more elaborately through (combinations of) GARCH, SV, or Student-t processes helps in recognizing periods of better prospects from periods with a higher probability of depreciation. For the German data in figure 5.12 the GLL-GARCH, -SV, -GARCH-Student-t and -SV-Student-t all coincide in recognizing the earlier part of 1998 as a period of adverse risk, together with the months around January 1999 and a short stretch of time in October 1999. At the higher level of risk aversion of $\gamma = -10$ (figure with the hedging decisions not included, the second half of 2000 is also recognized as a period where the increased variance outweighs the possibility of making a slightly higher return; with $\gamma = -10$, the average hedging ratio goes up during the last months of 2000.

For Japan, with $\gamma = 0$ only the Local Level model delivers a signal strong enough to hedge over longer periods of time, while the other models keep switching their position. For a risk-averse investor, the penalty of the -5% interest rate differential connected to hedging the currency risk is not so much of a problem, and the average hedge ratio is higher than before. The second half of 1998 and of 1999 is recognized as a volatile period with high risk of depreciations by most models. Note how the GLL-SV model is not

[15] See also footnote 3 on page 104, concerning the restriction of $H \in [0, 1]$.

[16] Note how the hedge decision for $\gamma = 0$ for the DM/USD data is exactly opposite to the decision for the USD/DM data, and likewise for the data on the Japanese exchange rate. For $\gamma = 0$, the utility function is symmetric as well as the models.

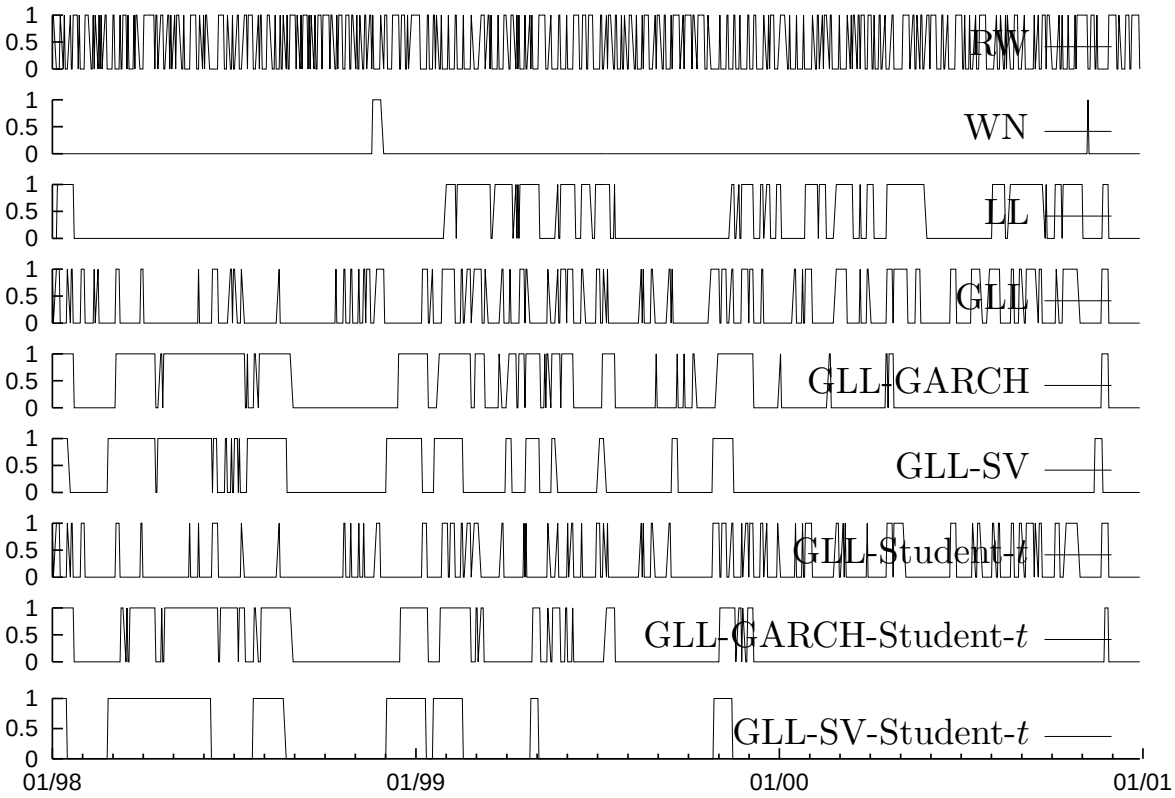

Figure 5.12: Hedging decisions through time for the DM/USD data, for $\gamma = 0$

taking much risk in the first years; only when the predicted volatility settles down to lower values, the GLL-SV model starts to hedge less (refer back to the second panel of figure 5.11 for the predictive standard deviation of the GLL-GARCH-Student-t model; the predictive standard deviation of the GLL-SV is similar).

5.7.4 Returns and utilities of model-based strategies

While it is interesting to see the variability of the hedging decisions over time, the most important result of the hedging strategy is the return, both financially, as the cumulative return of the strategy r and non-financially, as the cumulative utility U. These statistics are reported in tables 5.5–5.8 as well.

First look back at table 5.5 concerning the DM/USD data. As mentioned before, the three years in the evaluation period are rather different. This is seen e.g. from the fact that of the deterministic hedging strategies the full hedge strategy resulted in the best utility in the first year, followed in the second year by the no-hedge strategy while both are surpassed in the last year by the RW strategy. The deterministic outcomes fluctuate strongly throughout the years, and are therefore nice for comparison, but cannot be considered as serious strategies to be implemented by firms for covering their currency exposure.

From the model based strategies, over the complete period the SV model comes out first as a basis for providing hedging decisions, with utilities of 21.54 and 7.01 for the

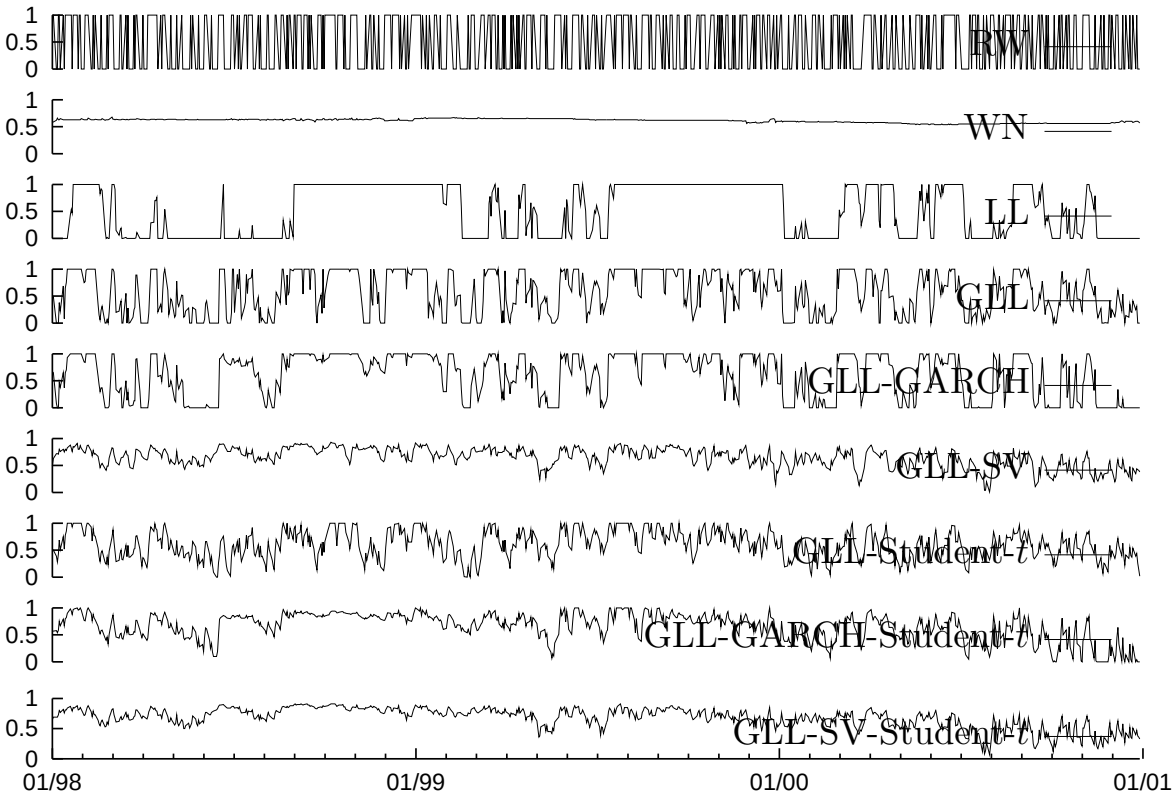

Figure 5.13: Hedging decisions through time for the Yen/USD data, for $\gamma = -10$

risk-tolerant and -intolerant case, respectively. For the risk-tolerant case, in both 1998 and 1999, its returns are close to the maximum returns of the deterministic strategies, whereas in 2000 the full hedge and the no hedge strategies are left behind. In this last year of exchange rates staggering wildly upwards, the RW strategy has a better return, as it does not care at all about risk. With the higher volatility during this period, the model based strategies tend to hedge more than in the previous period, missing out on some exchange rate appreciation. With $\gamma = -10$ the SV model is still best on average, with a 3-year utility of 7.01. Though best over the three years together, in each of the separate years it only ranks second or fourth.

The investor who faces the opposite hedging decision, hedging the DMark exchange rate from the United States, would be ex post best off using the SV model when he is risk-tolerant, or the GLL-GARCH model if he is more risk-averse. In the latter case, the GLL-SV model displays over time a more consistent behaviour than the GLL-GARCH model, with the consequence that the SV based hedging decision proves to be rather conservative, especially during the increase in the USD/DM exchange rate in September 1998, July 1999 and December 2000. On the other hand, the strategy based on the GLL-GARCH model is on-the-whole more aggressive than the other strategies: In the months mentioned, GLL-GARCH-based gains are larger, but also the losses can be considerable, e.g. at the end of March 2000 the GLL-GARCH strategy is the only one loosing almost 2% of the return over just 4 days.

In section 5.6.2 we saw that with the Yen/USD returns it was harder to get a good

fit to the data. The GLL-GARCH-Student-t model, with the highest marginal likelihood, displayed a mediocre signal-to-noise ratio of 0.86%. The model with second-best ML score resulted in a slightly higher S/N ratio, though with a parameter estimate of $\overline{\rho} = 0.55$ little predictive power of the first moment can be expected. Together with the high interest rate differential, a sensible hedging strategy may be hard to find.

By construction, the theoretical S/N ratio of the Local Level model is infinite. This model indeed displayed the largest range of prediction of tomorrow's exchange rate. When risk is not a problem, the LL model might be considered as the basis for the hedging decision. As the signal is rather strong here, this model leads to a hedging decision which switches little and manages to make the best out of the situation. Notice how on average the amount of hedging $\overline{H}$ by the RW strategy and by the strategy based on the LL model are rather similar, but the amount of switching the hedge position $\overline{|\Delta H|}$ definitely is not: When the LL model chooses to hedge, it continues to do so for on average 17 days, whereas the RW strategy keeps switching every other day.

In section 5.6.1 on the posterior distribution of the parameters, and in section 5.6.2 on the marginal likelihood, we found that the GARCH model fitted the model a lot better than the GLL-SV model. Indeed, with $\gamma = 0$ the SV model is hardly hedging at all, and the return is close to the unhedged return. The GARCH is second to the LL model over the complete evaluation period in attained utility. Ex post the GARCH model seems to have been rather careful over the year 2000, hedging on average 24% while the exchange rate appreciated over the year.

With $\gamma = -10$, results are harder to interpret. The SV model delivers hedging decisions which appear to be best, according to its utility, even though the model itself had very little explanatory power. The hedging decision for this model is driven mainly by the interest rate differential. Note that the GARCH model over 1998–2000 has a higher return than the SV model; this return however is penalized for the larger variation in the returns from day to day, resulting in a lower utility. The combination of a high interest rate differential and stronger risk aversion makes a strategy of doing as little as possible optimal.

The last setting considered is the situation of the U.S. investor with exposure in Japan, see table 5.8. In the risk-tolerant case, both SV models choose to hedge virtually all the time, as they judge the (in this situation, positive) interest rate differential attractive enough. If the investor is less risk-tolerant, there is even more incentive to hedge, and returns stay close to the cumulative return of the interest rate differential. Again, the LL model squeezes out enough of a signal to get some excess return over both the exchange rate return and the cumulative interest rate differential. In 1998, where there is a lot of variability in the exchange rate but also larger gains to be made, it hedges 57% on average. In the more tranquil period of 1999 this percentage is lowered and increased again over 2000, as in the last year a negative trending of the exchange rate is found. For the GARCH model the signal is weaker, leading to a more volatile hedging decision ($\overline{|\Delta H|} = 0.136$ instead of 0.059 for the LL model), but the model is still able to get a good return. The GLL-Student-t model follows closely. These effects can also be recognized in figure 5.14, where the cumulative utilities over the three years are plotted of the three naive and the Local Level and GLL-GARCH strategies.

With $\gamma = -10$, the LL model essentially takes the same decisions. However, this time

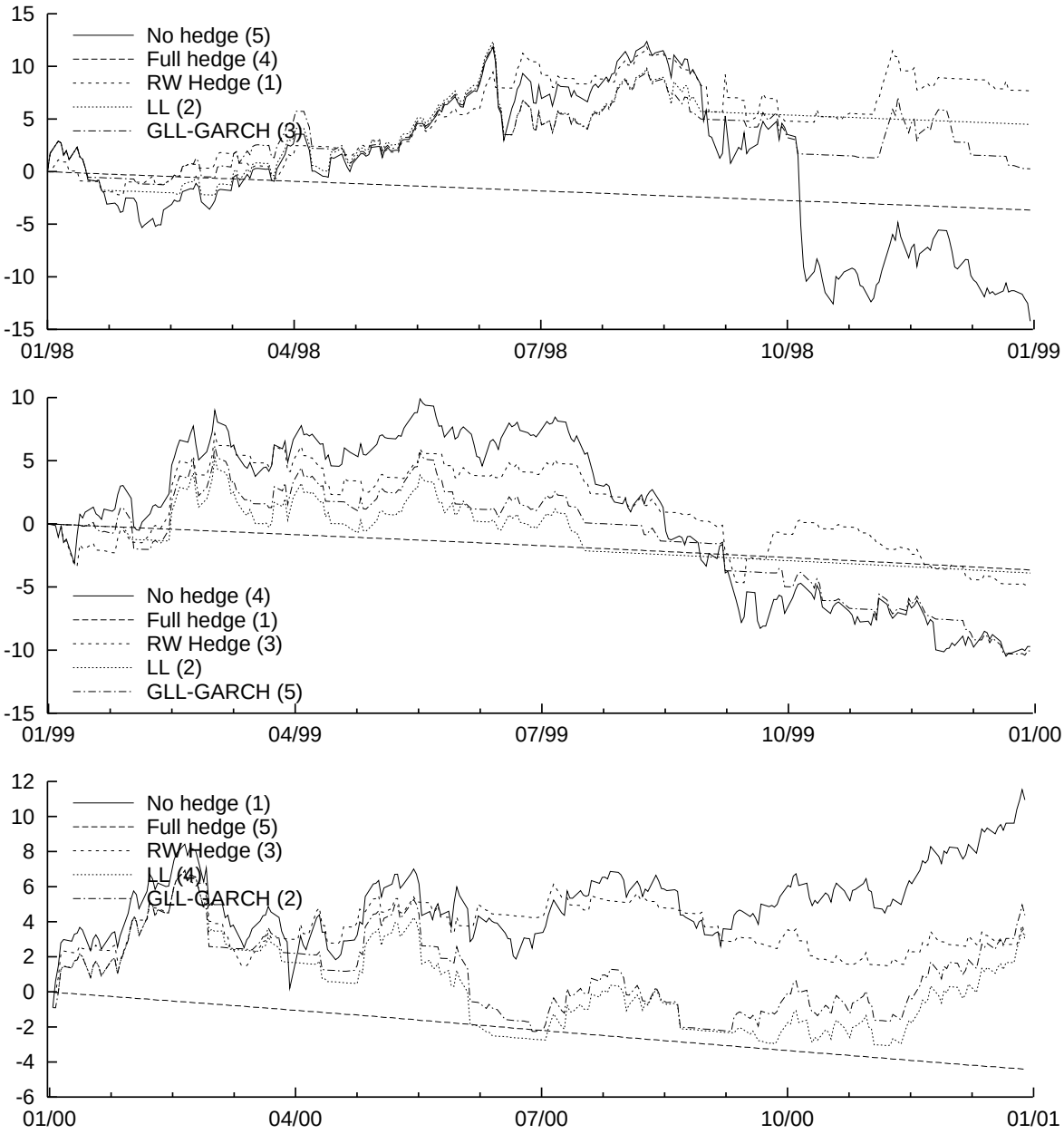

Figure 5.14: Cumulative utilities over the years 1998, 1999 and 2000, for the naive strategies and the strategies building on the Local Level and GLL-GARCH models, for the Yen/USD data with $\gamma = 0$. Indicated between parentheses is the ordering of the utilities at the end of the year.

the return of 24.3% is only valued at a cumulative utility of 9.5, as days with large gains are countered by other days with equally large losses, resulting in a negative impact on total utility. The GLL-GARCH is more conservative, resulting in smaller profits but a utility which is not much different. At $\gamma = -10$ the models which do not (or hardly) take any risk lead to the highest utility, (approximately) equal to the utility of full-hedge strategy.

5.7.5 Value-at-Risk and the Sharpe ratio

In sections 5.2 and 5.4.5 two alternative hedging strategies were presented, using the outcome of the model estimation but not going through a full utility optimization. Table 5.9 reports results for the USD/DM data. Results are given for the total evaluation period and split out per year. The first column in each panel copies results from table 5.6 for comparison, on the naive and utility-optimizing hedging strategies. Second and third columns report the cumulative utility for the Value-at-Risk and Sharpe strategies.

Table 5.9: Cumulative utilities for alternative hedging strategies, for the USD/DM data

Model	U_{Opt}	U_{VaR}	U_{Sh}	U_{Opt}	U_{VaR}	U_{Sh}	U_{Opt}	U_{VaR}	U_{Sh}	U_{Opt}	U_{VaR}	U_{Sh}
$\gamma = 0$	1998–2000			1998			1999			2000		
No hedge	−14.73			7.65			−15.83			−6.55		
Full hedge	4.75			1.47			1.72			1.55		
RW Hedge	14.55			4.98			−2.94			12.51		
WN	4.64	7.78	7.78	0.78	−0.97	−0.97	1.36	1.72	1.72	2.50	7.03	7.03
LL	8.94	4.32	4.32	2.97	2.28	2.28	−1.43	3.38	3.38	7.40	−1.34	−1.34
GLL	11.25	−0.51	−0.51	5.92	7.25	7.25	−1.32	1.79	1.79	6.65	−9.55	−9.55
GLLGA	5.70	−2.02	−3.29	6.33	−2.33	5.50	−3.87	3.24	−3.15	3.24	−2.93	−5.64
GLLSV	11.56	19.21	−5.80	7.19	9.13	3.68	−1.07	6.69	−2.33	5.45	3.40	−7.16
GLLT	9.33	−3.79	−4.31	4.31	2.62	2.62	−2.08	2.18	1.66	7.11	−8.59	−8.59
GLLGAT	8.84	4.94	−3.59	6.32	−1.67	5.34	−3.80	2.85	−0.96	6.32	3.76	−7.97
GLLSVT	9.29	18.67	1.54	5.87	6.53	3.72	−1.38	6.85	0.14	4.79	5.29	−2.32
$\gamma = -10$	1998–2000			1998			1999			2000		
No hedge	−29.27			3.87			−19.84			−13.31		
Full hedge	4.75			1.47			1.72			1.55		
RW Hedge	8.26			3.12			−4.53			9.67		
WN	4.72	4.79	4.79	1.39	1.52	1.52	1.70	1.72	1.72	1.63	1.55	1.55
LL	6.10	2.45	2.45	2.05	2.23	2.23	−1.41	1.76	1.76	5.46	−1.54	−1.54
GLL	6.21	−1.94	−2.37	1.74	1.42	1.42	1.25	0.76	1.29	3.22	−4.12	−5.08
GLLGA	8.44	4.14	3.46	4.21	2.19	2.13	0.18	1.39	3.38	4.04	0.55	−2.05
GLLSV	4.17	13.29	1.28	1.22	2.11	1.05	0.92	3.55	0.25	2.02	7.62	−0.02
GLLT	5.05	4.19	5.67	1.30	1.22	1.22	1.23	1.97	1.97	2.51	0.99	2.47
GLLGAT	4.78	4.44	1.68	1.32	1.21	1.37	0.32	2.80	0.72	3.15	0.43	−0.41
GLLSVT	4.58	15.12	4.23	1.33	−0.68	1.43	1.21	1.72	1.59	2.04	14.08	1.22

Note: Reported are the utilities of the deterministic strategies, followed by utilities based on estimating the exchange rate models. Reported are U_{Opt}, the utility following from optimizing the utility function (from table 5.6) and results U_{VaR} and U_{Sh} from the Value-at-Risk and Sharpe decision strategy.

The results are ambiguous. For both SV models (which display best marginal likelihood scores on this data set), the VaR strategy outperforms over the three years the utility-optimizing strategy. With $\gamma = 0$, the gains are made in 1998 and 1999, with a performance not as good as the utility-based strategy in 2000. For the risk-averse case, the reverse is true: Clear outperformance in 2000, little difference in earlier years. Striking is also the difference in results between the GLL-SV and GLL-GARCH-Student-t models: Though the marginal likelihoods of the models are close, and the hedging results after utility optimization are similar, the VaR strategy yields 14.27 resp. 8.85 points higher utility for the GLL-SV model than for the GLL-GARCH-Student-t model.

For the DM/USD and Yen/USD data sets (results not reported), all VaR and Sharpe results over the complete period were either close to the optimal results, or far worse (e.g. the utility of the GLL-SV model on DM/USD data with a risk-averse investor dropped from 7.01 to -28.96 for the VaR strategy and -12.16 for the Sharpe strategy). For USD/Yen data, U_{VaR} results were better for both GARCH-models, and slightly better for both the SV models, compared to the utility-optimizing results. With $\gamma = 0$, the utility of the Sharpe strategy even surpassed the utility of the VaR hedging methodology.

The background of these odd results for the VaR and Sharpe decision framework is unclear. One possible explanation is our choice of calibrating the cut-off level of maximum Value-at-Risk and the limiting level of the Sharpe ratio in such a way that the average hedging percentage equalled the utility-optimizing result. This topic is left for further research.

5.7.6 Other viewpoints on the results

In the previous section, the main classification of results was according to the utility attained by the hedging strategies based on the different models. Though important, the utility is not the only statistic the applied manager looks at.[17] This section focuses on the influence of the value of the risk-aversion parameter γ on the results, it briefly investigates the maximum losses and gains which are incurred for each of the strategies, and compares the utility to the frequency and size of change of the hedging position.

To start with the influence of γ, figure 5.15 depicts the level of utility attained for $\gamma \in [-20, 0.5]$, for the deterministic hedging strategies and for the GLL-GARCH and the GLL-SV based strategies. The data used in this graph are for the DM/USD exchange rate. The top-left panel of the figure displays the results for all years together, whereas the other panels show the behaviour of the cumulative utility for different values of γ throughout the evaluation period.

The value of the risk aversion parameter does not influence the utility attained by the full-hedge strategy, as with this strategy there is no risk. The RW strategy is deterministic as well, and therefore attains the same return irrespective of the value of γ. The lower the risk aversion parameter, the lower also the utility of the RW strategy, as the actual risks are valued more negatively. A similar effect occurs with the model based strategies: Lower values of γ are connected with lower values of the utility. Again, we see how different the three years in the evaluation period have been: In 1998, it was hardly possible to do

[17] It is not even clear that each fund manager hás an explicit, unidimensional utility function to be maximized, taking the risks into account.

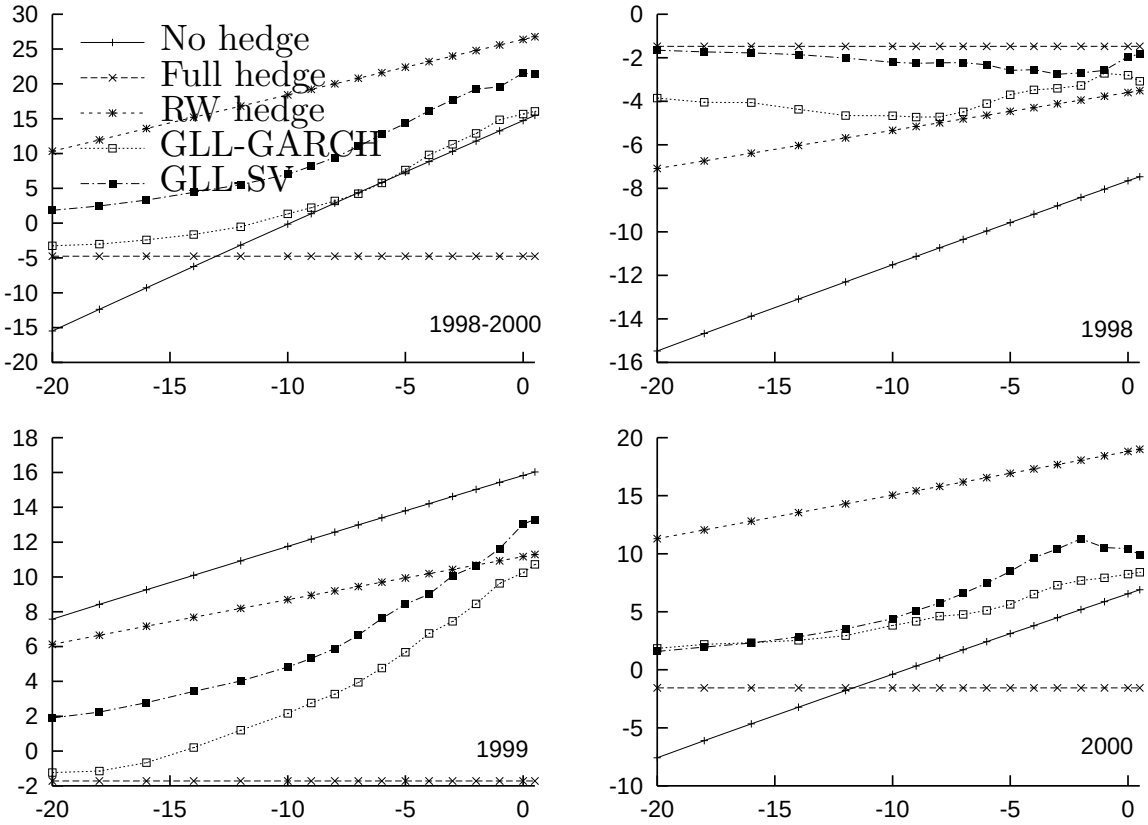

Figure 5.15: Cumulative utility attained throughout the evaluation period for a range of risk-tolerance parameters, using DM/USD data

better than to hedge fully. The model-based strategies, which indeed hedge considerably for low values of γ but only around 40% for $\gamma = 0$, reach a similar utility for all values of γ, not running into large losses over this year. In 1999, the exchange rate went up in a manner that the RW strategy could not improve on. If the investor is willing to take the risk, a similar return can be obtained using the models, but for a more risk-averse investor, naturally the returns get closer to the return of the fully hedged case. In 2000, the RW strategy ex post is proved to be most successful, but still the models reach higher utility levels than could be arrived at by betting only on exchange rate appreciation.

From these plots, the model based hedging strategies show to be a valid alternative to the naive hedging strategies, curbing the risks of years with larger depreciation, while not neglecting opportunities of possible gain from the exchange rate.

Finally, table 5.10 reports for the case of the risk-averse investor in the U.S. with investments in Germany, what the maximum losses and gains over a three month period have been (columns 1 and 2), for each of the naive and model based strategies. The last two columns report the cumulative utility of the strategy and the utility divided by the amount of changing the position, $\overline{|\Delta H|}$, that was needed. Again we see that the LL model is rather extreme, leading to a maximum loss of 5.5%, even larger than the maximum loss of the RW strategy. The other model-based strategies are well able to cover the down-side risk. Differences with the RW case are huge when the amount of switching position is

Table 5.10: Losses and gains for the USD/DM data, for $\gamma = -10$

Model	3M-L	3M-G	U	$U/\overline{\lvert\Delta H\rvert}$
No hedge	−12.85	12.86	−29.27	
Full hedge	0.31	0.49	4.75	
RW Hedge	−4.82	9.18	8.26	0.024
WN	0.24	0.49	4.72	12.076
LL	−5.46	9.02	6.10	0.150
GLL	−0.62	3.14	6.21	0.151
GLLGA	−2.39	7.20	8.44	0.112
GLLSV	−0.55	1.54	4.17	0.078
GLLT	−0.16	1.83	5.05	0.186
GLLGAT	−0.87	2.73	4.78	0.088
GLLSVT	−0.32	1.28	4.58	0.115

Note: Reported are the maximum losses and gains over a three month period, and the utility U, both in total and expressed as the utility attained per unit change of position, $U/\overline{|\Delta H|}$. Data concerns the USD/DM exchange rate with $\gamma = -10$ over the period 1998–2000.

taken into account. The utility of the RW strategy shrivels to a tiny 0.024 utility for each time the hedging position is changed. The model-based strategies have utilities of around 0.8-0.15 for each complete turnover of the hedging position. Therefore, when transaction costs come into play, as they do in practice when there is a frequent change in position, the RW strategy has to be discarded as not profitable.

5.8 Concluding remarks

During the past twenty years many models have been developed for the description of financial time series. Time varying variances are one of the most outstanding features of financial time series, and, as a consequence, much attention has been paid to modelling the variance of these series. However, many decision problems in finance depend on the full probability density of financial returns. In this paper we focused on currency overlay strategies for hedging foreign exchange rate exposure for an international investor. We investigated a wide range of competing models that describe the most prominent features of the exchange rates between Germany and the U.S. and Japan and the U.S.

Special attention has been given to describe the mean of exchange rate returns. The motivation for investigating models that integrate time varying means and variances springs from observing exchange rate time series. Besides the feature of time varying variances, there is some evidence that these series exhibit local trend behaviour, i.e. prolonged periods of exchange rate appreciation or depreciation. Capturing this feature may lead to better risk and return characteristics of hedging strategies. When estimating our models we use Bayesian estimation methods.

In section 5.7 we presented the results of various hedging strategies for four different currency positions. The returns and utilities of adopting naive or model-based hedging

strategies were analysed over the period 1998–2000. The character of the fluctuations of the DM/USD and Yen/USD exchange rates were found to be quite different. The exchange rate of the dollar with the DMark was best modelled using a GLL-Stochastic Volatility model, whereas in the case of the Yen/USD exchange rate the heavy tails of the Student-t density, possibly combined with a GARCH process for the varying variance, were most important.

For the DM/USD rate, modelling time varying features, and using a power utility objective function, pays off in terms of risk-adjusted returns for a moderately risk averse currency overlay manager. Ex post, in each of the years one of the deterministic hedging strategies (no hedge, full hedge or a Random Walk hedge) performed better, but ex ante there is no way to choose between these naive strategies. Hedging the Japanese yen against the dollar and vice versa proved more difficult. Among practicioners it is known what we found in sections 5.6.1-5.6.2 on the posterior densities and the marginal likelihoods: This exchange rate is hard to model. As the interest rate differential between the U.S. and Japan was large over the evaluation period, it either hardly made sense to hedge (for the Yen/USD case) or to take the risk of a depreciation when a risk-free return of 5-6% on a yearly basis could be gained (USD/Yen). A risk-tolerant currency overlay manager can choose to hedge according to the optimal decision based on the Local Level model: Though the model is not best at fitting the data according to its marginal likelihood, it magnifies any signal encountered in the data and hedges accordingly with good returns. A risk-averse investor in this market is better off adapting a strategy based on the GLL-GARCH or GLL-Student-t model.

Other findings in this chapter are that, first, for decreasing risk tolerance the decision to use one model or the other is less relevant, as in the limit of $\gamma \to -\infty$ all models choose to hedge all the time. Secondly, when modelling is worthwhile, it appears that there is not one model that is uniformly superior for all criteria. The marginal likelihood score of a model is however a good indicator of the model's hedging performance. Thirdly, using these models to construct predictive densities only to compute a hedging decision based on the Value-at-Risk or the Sharpe ratio is not useful, at least not in the present setup. The utilities attained by these strategies fluctuate strongly over the years, the models, and the exchange rates used. In the USD/DM case, the GLL-SV based VaR hedging strategy appears to perform better than the utility optimizing strategy, but in other cases both the VaR and Sharpe strategies tend to be worse. Fourth, care has to be taken with the interpretation of the results. Even though the evaluation period included three years quite distinct in behaviour of the exchange rates, results still can be strongly influenced by large shocks on a small number days. As returns of 1% or more are not uncommon, with extremes reaching exchange rate returns of 5% in a single day, an incorrect decision on the wrong day can spoil the final outcome of a hedging strategy.

The topic of integrating models for risk and return into a framework for financial decision making can be extended in several ways. First, the AR(1) structure that we applied in this paper for the unobserved time varying mean describes the local trend behaviour of the exchange rate levels, but other models may be investigated as well. For instance, a finite mixture model or the RiskMetrics model (see JP Morgan 1997) are obvious candidate models for comparison.

Secondly, the models could be extended with information from other economic vari-

ables. Within the exchange rate literature much attention has been given to the uncovered interest rate parity and/or the purchasing power parity as building blocks for predicting exchange rates. References to this field include Mark (1995), Bansal (1997), Bansal and Dahlquist (2000) and Evans and Lewis (1995).

Thirdly, the final hedging results can depend strongly on a few days with large absolute returns. The consequences of decision making may be investigated over longer periods, or comparing subperiods in greater detail. Results may be contrasted to simulation results, where the data generating process is known and the effect of changing the hedge strategy is more purely observed.

Fourth, one may perform the hedge decision for several currencies simultaneously. An obvious advantage of this approach is that hedging costs could become lower due to diversification. Crucial input for making hedge decisions in this way is the availability of multivariate time series models for exchange rate returns; work in this field has been done by Aguilar and West (2000). Another possibility is to incorporate the currency hedging decision in portfolio choice models. This approach steps away from the currency overlay principle that we pursued in this paper, and integrates the hedging decision into the international allocation problem. Bayesian references on portfolio choice include Jorion (1985), Jorion (1986), Geweke and Zhou (1996), McCulloch and Rossi (1990), McCulloch and Rossi (1991), and Kandel, McCulloch and Stambaugh (1995).

Fifth, the alternative hedging strategies using the Value-at-Risk and the Sharpe ratio need to be investigated further. What is the reason for the results differing so strongly between years and between exchange rates?

Finally, it is of interest to extend the decision framework and allow for currency options as an instrument in the decision process. One may allow for the hedging parameter to be outside the unit interval. Hence, managers may use currencies as an investment in their own right.

5.A Gibbs sampling with data augmentation

In this chapter a sample from the posterior distribution of the parameters was needed. For the smaller models, a Metropolis-Hastings algorithm could have been applied, but the choice was made to sample from all models using a Gibbs sampling procedure with data augmentation. Details on the Gibbs sampling algorithm in general are given in section 4.2.5 and the references mentioned there. The example in section 4.5 presents a Local Level model with Student-t errors, and clarifies on the specific sampling procedure for this model. In this appendix, we elaborate on the methods used to attain conditional normality in the SV model, and list the full conditional distributions for the parameters.

For reference, the basic model presented in section 5.3 is

$$s_t = c + \mu_t + \sigma_\epsilon \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, h_t), \tag{5.8′}$$

$$\mu_t = \rho \mu_{t-1} + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathcal{N}(0,1), \qquad t = 1, .., T. \tag{5.9′}$$

For ease of notation we introduce a varying variance component h_t in the observation equation (5.8′). By construction, h_t has expectation 1, such that the unconditional variance of the observation equation is σ_ϵ^2.

The mean return c is only used in conjunction with the White Noise model; for this model, the varying mean process if fixed at zero ($\mu_t = \rho = \sigma_\eta = 0$), and the variance is constant ($h_t = 1$). The LL and GLL models keep the variance constant as well. Model **B**, the LL model, incorporates a Random Walk mean process μ_t, obtained by setting $\rho = 1$. The GLL-based models allow ρ to take on other values. The parameter is not a priori limited to the range $[0, 1]$. With model **D**, the GARCH variance is approximated (see Harvey et al. 1992) using

$$h_t^{\text{GARCH}} = (1 - \delta - \alpha) + \delta h_{t-1}^{\text{GARCH}} + \alpha \left(\mathsf{E}\epsilon_{t-1}\right)^2. \tag{5A.1}$$

With pure GARCH models, without the state space component, the true lagged disturbance ϵ_{t-1} can be filled in instead of the filtered disturbance $\mathsf{E}\epsilon_{t-1}$ used here. The approximated GARCH variance is expected to be smoother than the pure GARCH variance, though the effect is small.

For the Stochastic Volatility model we use $h_t = \exp(h_t^{\text{SV}})$, with

$$h_t^{\text{SV}} = \phi h_{t-1}^{\text{SV}} + \sigma_\xi \xi_t, \qquad \xi_t \sim \mathcal{N}(0,1). \tag{5.13′}$$

The Student-t disturbances are included using data augmentation (see section 4.5.2), by sampling using $h_t = z_t^2, z_t \sim \text{IG-1}(\alpha = \nu/2, \beta = 2/(\nu - 2))$. The GLL-GARCH-Student-$t$ is a combination of models **D** and **F**, with $h_t = z_t^2 \times h_t^{\text{GARCH}}$. Likewise, the GLL-SV-Student-t model puts $h_t = z_t^2 \times h_t^{\text{SV}}$.

Most of the conditional densities can be derived in a straightforward manner, following the lines in section 4.5. All the sampling densities are reported in table 5.11. The only step which is less intuitive is the sampling of h_t^{SV} in models **E** and **H**.

For the SV model, the observation equation (5.8′) can be written as

$$s_t = \mu_t + \sigma_\epsilon \sqrt{\exp(h_t^{\text{SV}})}\epsilon_t \tag{5A.2}$$

which is clearly nonlinear in h_t^{SV}. In Harvey, Ruiz and Shephard (1994) the equation is linearized by taking squares and logarithms,

$$\log(s_t - \mu_t)^2 - \log \sigma_\epsilon^2 = h_t^{\text{SV}} + \log \epsilon_t^2. \tag{5A.3}$$

Write $y_t^* = \log(s_t - \mu_t)^2 - \log \sigma_\epsilon^2$ and $v_t = \log \epsilon_t^2 \sim f(v_t)$. The first way to operationalize the SV model, in Harvey et al. (1994), was to use quasi-maximum likelihood, approximating the density $f(v_t)$ using a normal density. In Kim et al. (1998) (and more recently in Chib, Nardari and Shephard 2000), in a Bayesian paper, the density is approximated more precisely using a mixture of 7 normal densities, i.e.

$$f(v_t) \approx \sum_{i=1}^{7} \mathcal{I}_{s_t=i} f_{\mathcal{N}}(v_t; \mu_{s_t}, \sigma_{s_t}^2). \tag{5A.4}$$

To further improve on this approximation, they use an importance sampling step to get rid of small discrepancies between $f(v_t)$ and the mixture. In the underlying chapter we used the mixture of normal densities without afterwards calculating the importance weights, as this would have meant a rather large calculational effort on top of a sampling scheme which already is computationally intensive, with little improvement expected from such effort.

Augmenting the sample space with the vector of indices s_t into the mixture density, normality is regained. To sample from these indices, the discrete density $p(s_t = i|y_t^*, h_t)$ is evaluated for $i = 1, .., 7$. Drawing from the discrete density is straightforward.

Afterwards, conditional on the s_t, the system

$$y_t^* = h_t^{\text{SV}} + v_t(s_t), \tag{5A.5}$$

$$h_t^{\text{SV}} = \phi h_{t-1}^{\text{SV}} + \sigma_\xi \xi_t \tag{5A.6}$$

is a Gaussian state space model along the lines of (5.8′)-(5.9′), and the simulation smoother (De Jong and Shephard 1995) can be used to sample a new vector of h_t^{SV}.

The full conditional posterior densities which are needed in the Gibbs sampling algorithm are given without derivation in table 5.11. For the GARCH parameters $\sigma_\epsilon, \delta, \alpha$ and for the degrees-of-freedom parameter ν no closed form expression of the conditional density is available. Therefore, we use in these steps a Metropolis-within-Gibbs sampler (see Koop and Van Dijk 2000, Zeger and Karim 1991, and section 4.3.1). Note that the priors in table 5.1 in section 5.4.1 have been applied.

Table 5.11: Conditional posterior densities

Parameter	In model	Full conditional density
c	**A**	$\mathcal{N}\left(\frac{\hat{c}\sigma_c^2+\mu_c\hat{\sigma}_c^2}{\sigma_c^2+\hat{\sigma}_c^2}, \frac{\hat{\sigma}_c^2\sigma_c^2}{\sigma_\rho^2+\hat{\sigma}_\rho^2}\right)$ with $\hat{\rho}$ and $\hat{\sigma}_\rho^2$ the least squares estimate of ρ with corresponding variance.
μ	**B**, **C**, **D**, **E**, **F**, **G**, **H**	Use the simulation smoother, see De Jong and Shephard (1995).
ρ	**C**, **D**, **E**, **F**, **G**, **H**	$\mathcal{N}\left(\frac{\hat{\rho}\sigma_\rho^2+\mu_\rho\hat{\sigma}_\rho^2}{\sigma_\rho^2+\hat{\sigma}_\rho^2}, \frac{\hat{\sigma}_\rho^2\sigma_\rho^2}{\sigma_\rho^2+\hat{\sigma}_\rho^2}\right)$ with $\hat{\rho}$ and $\hat{\sigma}_\rho^2$ the least squares estimate of ρ with corresponding variance.
σ_ϵ	**B**, **C**, **E**, **F**, **H**	IG-1 $\left(\alpha = \frac{T}{2}+\alpha_\epsilon, \beta = 2\Big/\left(\sum \frac{(y_t-\mu_t)^2}{h_t} + \frac{2}{\beta_\epsilon}\right)\right)$.
$\sigma_\epsilon, \delta, \alpha$	**D**, **G**	Use MH sampling. The conditional posterior is proportional to the likelihood from the Kalman filter equations and the prior.
σ_η	**B**, **C**, **D**, **E**, **F**, **G**, **H**	IG-1 $\left(\alpha = \frac{T}{2}+\alpha_\eta, \beta = 2\Big/\left(\sum(\mu_t-\rho\mu_{t-1})^2 + \frac{2}{\beta_\eta}\right)\right)$.
ϕ	**E**, **H**	$\mathcal{N}\left(\frac{\hat{\phi}\sigma_\phi^2+\mu_\phi\hat{\sigma}_\phi^2}{\sigma_\phi^2+\hat{\sigma}_\phi^2}, \frac{\hat{\sigma}_\phi^2\sigma_\phi^2}{\sigma_\phi^2+\hat{\sigma}_\phi^2}\right)$ with $\hat{\phi}$ and $\hat{\sigma}_\phi^2$ the least squares estimate of ϕ with corresponding variance.
σ_ξ	**E**, **H**	IG-1 $\left(\alpha = \frac{T}{2}+\alpha_\xi, \beta = 2\Big/\left(\sum\left(h_t^{\mathrm{SV}} - \phi h_{t-1}^{\mathrm{SV}}\right)^2 + \frac{2}{\beta_\xi}\right)\right)$.
s_t	**E**, **H**	The indices into the mixture in the distribution of $\ln\epsilon_t^2$ are discretely distributed.
$h_{t,SV}$	**E**, **H**	From the simulation smoother on the transformed model (5A.5)-(5A.6).
z_t	**F**, **G**, **H**	IG-1 $\left(\alpha = \frac{\nu+1}{2}, \beta = 2\Big/\left((\nu-2) + \frac{(y_t-\mu_t)^2}{\sigma_\epsilon^2 h_t}\right)\right)$.
ν	**F**, **G**, **H**	The posterior is not of a known form. It is proportional to $\prod_t$ IG-1 $\left(z_t; \alpha = \frac{\nu}{2}, \beta = \frac{2}{\nu-2}\right) \times$ Cauchy$(\nu; \mu = 0; s = 1)$. Apply a MH step to sample a new value of ν.

Chapter 6

Conclusions

6.1 General remarks

Analysing time series can be hard, as empirical series hardly ever fit straight into a standard class of models. Knowledge of the theory behind time series is a prerequisite for attaining fruitful results. Knowledge however is a necessary, not a sufficient condition for good research. There are too many different ways of approaching an empirical problem to allow a researcher to try them all, it is more of an art that is needed for making the right modelling decision at the right occasion.

This thesis intends to combine the art of choice with the knowledge of the advanced methods of time series analysis. It presents results for models on inflation and exchange rates. In both cases, a large toolbox containing fractionally integrated models, state space models, advanced Bayesian sampling methods and elaborate evaluation of estimation results and of the consequences for optimal decision making was needed.

The following sections describe briefly the conclusions that are drawn in this thesis. Section 6.2 summarizes the findings for the modelling of the inflation rate series in chapters 2 and 3. In section 6.3 results for the chapter on Bayesian simulation methods are described. A final section of the thesis, section 6.4, looks back at chapter 5, on hedging the exchange rate.

The end of the thesis definitely does not signify that research on these topics has come to an end. Each of the sections describes some possibilities for future research on improving the models and the methods.

6.2 Inflation rates

Chapters 2 and 3 look into the estimation and prediction of inflation rate series. Inflation is the key example in the literature of a series with long memory characteristics: Even after 40 months, the effect of a shock has not yet died out, a clear indication of fractional integration.[1]

[1]Such an autocorrelation function can partly be mimicked using a very high AR coefficient, partly offset by a large negative MA root. However, with inflation the fit of ARIMA models improved much by allowing for fractional integration.

The first of the two chapters investigates the matter of fractional integration of the inflation in G7 countries more profoundly. It was known beforehand that a number of mean shifts may spuriously lead to concluding that a series is non-stationary. Similarly a break in the series can increase the evidence for fractional integration. This effect is indeed found in a simulation exercise. The Wald and Lagrange Multiplier tests can be used for testing the significance of the breaks. We find in the simulation that the Lagrange Multiplier test has power for detecting a level shift, though some size distortion occurs.

For the G7 countries, we estimate the ARFIMA model including zero, two or four level shifts around the oil crises. Inclusion of the level shifts is significant for Canada, France, Italy, the United Kingdom and the United States. Inflation rates in Japan and Germany do not seem to react in the same way, or at the same time, to the oil crises. Two shifts, at 1973:07 and 1982:07, suffice to take out the largest change in the mean inflation, adding two extra shifts for the period in between the two oil crises does not alter results strongly. Allowing for the mean shifts lowers the amount of fractional integration found in the data, leading to an insignificant parameter d for Canada and Japan. For other countries, the degree of integration decreases considerably.

Chapter 3 looks into the effect of the degree of integration on U.S. core inflation forecasts. For short horizons, predicting not more than 3 months ahead, the difference between short memory ARMA(1,1), non-stationary ARIMA(1,1,1) and fractionally integrated ARFIMA(1,d,1) models is small. At longer horizons, the effect of the degree of integration on the width of the forecast intervals becomes more important. The ARIMA(1,1,1) model delivers a confidence interval for the prediction which is too wide. The results for the ARFIMA(0,d,0) model are more realistic.

The models make use of explanatory variables to improve on the forecasts. It is found that for short horizons, the forecast precision is slightly improved by using the short term interest rate as a leading indicator for inflation. For longer horizons, none of the examined variables contains clear predictive power for the inflation rate.

In the recursive estimation of the parameters of the models we find a robust, constant estimate of d for the ARFIMA(0,d,0), with and without explanatory variables. For the ARMA(1,1) and ARIMA(1,1,1) models the estimates for the ARMA parameters ϕ and θ change considerably when the sample is enlarged from 1960:04-1984:01 to 1960:04-1999:11. For all the models we find an indication that the variance of the residuals in the later part of the sample was lower than in the months before January 1984. Downweighting earlier observations, which essentially corresponds to changing the variance, improves the results for the tests of correct unconditional coverage of the forecast intervals.

In the discussion of chapters 2 and 3 we touched upon the following points open for improvement:

i. **Endogenous timing:** It is claimed in chapter 2 that fixing breakdates exogenously gives a sufficient approximation to the timing of the moment the mean of inflation shifted. Would an endogenous selection of breakpoints, or even a smooth transition mechanism for moving between regimes, lead to further lowering of the degree of integration? Preliminary results, by ourselves and by other authors (Hsu 2000), give the impression that the analysis in chapter 2 is quite close to being 'optimal', in the sense that the data does not contain much more information for selecting jointly the

timing, the size, and the shift-mechanism of the breaks, together with estimating the degree of integration.

ii. **Endogenous weighting:** Similarly, chapter 3 uses an exogenous weighting scheme for downweighting past observations. For the research question of the chapter, 'What model delivers reasonable prediction intervals for future inflation?', the precise weighting scheme did not matter much. Further analysis on this topic is however needed.

6.3 Bayesian simulation methods

While chapters 2 and 3 have a classical statistical background, in chapters 4 and 5 the switch to Bayesian methods is made. Chapter 4 lays the groundwork for chapter 5 by reviewing a range of sampling methods, clearly describing the links between the methods. Section 4.3.4 presents the newly devised Adaptive Polar Sampling (APS) method (Bauwens et al. 2000). On the example in section 4.5 the APS algorithm delivers a sample with very little correlation, though this comes at the cost of a computing time which was higher than for the Metropolis-Hastings algorithm, of the same order as the Gibbs sampling algorithm.

Section 4.4 summarizes a collection of methods for computing the Bayesian marginal likelihood of a model. This marginal likelihood is the basis for comparisons between two models using the posterior odds. The different methods are compared for their accuracy in the example in section 4.5. For the model considered, with a relatively low number of parameters, the simple method based on comparing the LaPlace or kernel-smoothed approximation of the posterior density to the height of the posterior kernel works well for computing the marginal likelihood. The chapter provides quantitative indications of the accuracy of the computational methods, opposed to the comparisons available only in qualitative terms in the literature until now.

The theory of sampling methods has matured tremendously over the last decade. With many methods available, the researcher can pick at will a simulation method which is convenient for the problem at hand. Chapter 4, and especially section 4.5, intend to provide a comparison between the algorithms on the basis of their ease of use, their efficiency and their accuracy. Improvement can be made on many fields. In connection to this chapter, two interesting paths for continuing research are:

i. **Efficiency:** In section 4.3.4 the Adaptive Polar Sampler is introduced. In the example at the end of the chapter it delivers a sample of the posterior density of low correlation, though at the cost of a relatively large computational effort. Research continues on models where APS has a comparative advantage over e.g. Metropolis-Hastings. This is expected to be the case especially for multimodal posterior densities or models with parameter vectors of a moderately large dimension.

   A different line of research is to combine the one-dimensional numerical quadrature of APS with $n-1$ dimensional importance sampling, instead of the Metropolis-Hastings step which is used at present. Such a combination, called the APIS algorithm, may lead to slightly higher efficiency.

ii. **Accuracy:** Marginal likelihood calculations are no longer as exotic as they used to be about five years ago. For the low dimensional model in the example in section 4.5, using LaPlace or kernel approximations leads to quick, simple and quite precise marginal likelihood computations. A further comparison of the behaviour of the computational methods on more elaborate models is interesting.

6.4 Hedging currency risk

Chapter 5 is the most elaborate chapter of this thesis. It uses the Bayesian methodology of chapter 4 to construct optimal hedging decisions for currency exposure under (parameter and other) uncertainty. The hedging decisions are taken by optimizing a power utility function, using as input the predictive density of tomorrow's exchange rate return and the difference between the home and the foreign interest rates. Predictions of the future exchange rate are made on the basis of eight models. These eight models each capture a different combination of the aspects of returns on exchange rates, such as the time variation in the trending of the value of the currency, the variation in volatility, and the fatness of the tails of the disturbances. Parameters in the models are estimated using Bayesian sampling methods.

All this work is done in order to hedge the currency risk of a large investor, with exposure to risk due to investments denoted in a foreign currency. With the posterior density of the parameters we optimize the expected utility over possible hedging decisions, ranging from not hedging at all to hedging the currency exposure fully.

The two extreme hedging positions have their respective returns and utilities. When the investor chooses never to hedge, he hands over his fate to the prevailing winds at the sea of exchange rates; with the other extreme, not taking any risk, he obtains a return equal to the interest rate differential between the home and foreign countries. Ex post, at the end of a period (e.g. a year), it is easy to see which strategy would have worked better, but ex ante it is of course unknown if the home currency will appreciate or depreciate.

Model-based decisions are given in the chapter for the DM/USD and the Yen/USD exchange rates, and vice versa for the USD/DM and USD/Yen rates. The returns of the DM rate with respect to the dollar are best modelled using a Generalized Local Level model with Stochastic Volatility (GLL-SV). The most important characteristic of these exchange rate returns is the predictability of the variability of the returns over time, with some extra gain made by including local trending behaviour of the exchange rate. Japanese Yen rates are harder to predict. Marginal likelihoods (section 5.6.2) indicate that the heavy tails of a Student-t density for the disturbances are most important, with a further improvement for GARCH-type varying variance. Using the Student-t density implies that the variance of the predictive density is constant. Therefore the predictive density cannot discriminate between periods of low and high risk, and model-based hedging decisions for the Yen/USD currency are expected to add less value for an investor than using the utility-optimizing hedging strategy on the DM/USD exchange rate returns.

The analysis of the resulting returns and utilities of the hedging decisions in section 5.7 confirms the results of the marginal likelihood computations. Over the years, with the

GLL-SV model for DM/USD (and also, but less convincingly so, for USD/DM[2]) exchange rate, consistent hedging decisions are taken. The model chooses to hedge in periods of increased risk and downward trending exchange rates. The Yen/USD exchange rate is found to be harder to model. The model which fits the data best according to the marginal likelihood criterion displayes a signal-to-noise ratio of S/N= 0.86%, whereas in the case of the DM/USD data the size of the variance of the signal is 2.3% of the variance of the noise. Furthermore, the interest rate differential between the U.S. and Japan was large (around 6% on a yearly basis) during the evaluation period, leading to a strong incentive to either hedge completely or not at all, depending on the direction of the currency exposure. A risk-tolerant investor can choose to hedge according to decisions based on the Local Level model. This model is not best according to the marginal likelihood, but has by construction the largest (theoretically infinitely large) signal-to-noise ratio, and therefore is one of the few models which takes a decision to go against the interest rate differential on a larger number of days. For a risk averse investor, a model like the GLL-SV leads to a higher utility over the three years. When a U.S. based investment firm is considering to hedge exposure in Japan, it is almost always better to hedge, as the interest rate differential gives a risk-free return of 6% per year.

The details of the analysis are provided for both a risk-tolerant (with risk-tolerance parameter $\gamma = 0$) as for a risk-averse ($\gamma = -10$) investor. For other values of the risk-tolerance parameter only graphical results on the cumulative utility are presented. The value of γ influences the final utilities and returns strongly, even more than the precise choice of the model. For an extremely risk-averse investor ($\gamma \to -\infty$), results tend to the results derived from the deterministic strategy of hedging continuously. For values of the parameter γ closer to zero, investors have the incentive to take more risk, resulting in hedging results being more similar between models as well. For intermediate values of γ, the power of the models for supporting a hedging decision are best appreciated.

Chapter 5 also considers alternative hedging strategies. A 'Random Walk' strategy, hedging tomorrow's return whenever the exchange rate is going down, leads to reasonable returns but excessive changes in the hedging position. Strategies based on a limiting Value-at-Risk or on the Sharpe ratio between expected return and the predicted standard deviation of the return did not lead to stable utilities throughout the years of the evaluation period for the models at hand.

The chapter intends to answer the most elaborate set of research questions in this thesis, searching for a good model for exchange rate returns, simulating from the posterior densities of the models, computing the marginal likelihood for model comparison, evaluating the predictive densities, optimizing a utility function over the hedging decision and judging the results for four different currency positions. While answering these questions, many others are left unanswered. The list below presents an incomplete set of open roads for a continuation of research concerning chapter 5.

i. **Limiting risk:** The hedging strategies based on Value-at-Risk and on the Sharpe ratio lead to erratic results over the years of the evaluation period. This seems to be caused by our method of calibrating the limiting risk. A better method of choosing

[2]Note that the models we use are symmetric, but the utility function (for a risk tolerance parameter $\gamma \neq 0$) and the resulting hedging decision are not.

the limits has to be introduced.

ii. **Multivariate analysis:** The models used for the exchange rate returns are univariate. Multivariate (factor) models could strengthen the signal, possibly leading to better hedging decisions. Multivariate models also allow for evaluating the utility attached to a decision concerning a portfolio of investment options, instead of the univariate decision of choosing the optimal hedging ratio.

iii. **Updating:** At present we do not update the posterior density of the parameters of the exchange rate return model as new observations become available. When the evaluation period is short, this is of little consequence. With an evaluation period of three years, a slight modification of the results can be expected if the posterior sample is updated at e.g. the end of each year.

iv. **GARCH:** In the GARCH models as described in section 5.3 we use the expectation of the residuals (Harvey et al. 1992) for updating the variance instead of the true residuals, as in the original framework of Engle (1982). The precise implication of this change needs further research, though preliminary results indicate that the difference has little influence on the outcomes of the hedging decisions.

Bibliography

Adenstedt, R. K. (1974), 'On large-sample estimation for the mean of a stationary random sequence', *Annals of Statistics* **2**, 1095–1107.

Aguilar, O., Huerta, G., Prado, R. and West, M. (1999), Bayesian inference on latent structure in time series, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds, 'Proceedings of the 6th International Meeting on Bayesian Statistics, Valencia, Spain', Oxford University Press.

Aguilar, O. and West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business & Economic Statistics* **18**(3), 338–357.

Andrews, D. W. K. (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica* **61**(4), 821–856.

Bai, J. (1997), 'Estimation of a change point in multiple regression models', *Review of Economics and Statistics* **79**(4), 551–562.

Bai, J. and Perron, P. (1998), 'Testing for and estimation of multiple structural changes', *Econometrica* **66**(1), 47–79.

Baillie, R. T. (1996), 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics* **73**, 5–59.

Baillie, R. T., Chung, C.-F. and Tieslau, M. A. (1996), 'Analysing inflation by the fractionally integrated ARFIMA-GARCH model', *Journal of Applied Econometrics* **11**, 23–40.

Ball, L. and Mankiw, N. G. (1995), 'Relative-price changes as aggregate supply shocks', *Quarterly Journal of Economics* **110**(1), 161–193.

Bansal, R. (1997), 'An exploration of the forward premium puzzle in currency markets', *Review of Financial Studies* **10**, 369–403.

Bansal, R. and Dahlquist, M. (2000), 'The forward premium puzzle: Different tales from developed and emerging economies', *Journal of International Economics* **51**(1), 115–144.

Barberis, N. (2000), 'Investing for the long run when returns are predictable', *Journal of Finance* **55**, 225–264.

Bauwens, L., Bos, C. S. and Van Dijk, H. K. (1999), Adaptive polar sampling with an application to a Bayes measure of Value-at-Risk, Tinbergen Discussion Paper TI 99-082/4, Tinbergen Institute.

Bauwens, L., Bos, C. S. and Van Dijk, H. K. (2000), Adaptive polar sampling: A new MCMC method for ill-behaved posterior surfaces, *in* W. Jansen and J. G. Bethlehem, eds, 'Proceedings in Computational Statistics 2000', Statistics Netherlands, pp. 13–14.

Bauwens, L. and Lubrano, M. (1998), 'Bayesian inference on GARCH models using the Gibbs sampler', *Econometrics Journal* pp. C23–C46.

Bauwens, L., Lubrano, M. and Richard, J.-F. (1999), *Bayesian Inference in Dynamic Econometric Models*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Bayes, T. (1763), An essay towards solving a problem in the doctrine of chance, *in* E. S. Pearson and M. G. Kendall, eds, 'Studies in the History of Statistics and Probablity', Griffen, London.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327.

Bos, C. S., Franses, P. H. and Ooms, M. (1999), 'Long memory and level shifts: Reanalyzing inflation rates', *Empirical Economics* **24**, 427–449.

Bos, C. S., Franses, P. H. and Ooms, M. (2001), 'Inflation, forecast intervals and long memory regression models', *International Journal of Forecasting* **18**.

Bos, C. S., Mahieu, R. J. and Van Dijk, H. K. (2000*a*), 'Daily exchange rate behaviour and hedging of currency risk', *Journal of Applied Econometrics* **15**(6), 671–696.

Bos, C. S., Mahieu, R. J. and Van Dijk, H. K. (2000*b*), On the variation of hedging decisions in daily currency risk management, *in* 'Bayesian Methods with Applications to Science, Policy, and Official Statistics', International Society for Bayesian Analysis, pp. 31–40.

Bos, P. C. (1969), The Functions of Money in Equilibrium and Disequilibrium, PhD thesis, Vrije Universiteit Amsterdam.

Box, G. E. P. and Jenkins, G. M. (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994), *Time Series Analysis, Forecasting and Control*, 3rd edn, Prentice Hall, Englewood Cliffs, NJ.

Breidt, F. J., Crato, N. and de Lima, P. (1998), 'On the detection and estimation of long memory in stochastic volatility', *Journal of Econometrics* **83**(1-2), 325–348.

Brodsky, J. and Hurvich, C. M. (1999), 'Multi-step forecasting for long-memory processes', *Journal of Forecasting* **18**, 59–75.

Carlin, B. P., Polson, N. G. and Stoffer, D. S. (1992), 'A Monte Carlo approach to nonnormal and nonlinear state-space modeling', *Journal of the American Statistical Association* **87**(418), 493–500.

Carter, C. K. and Kohn, R. (1994), 'On Gibbs sampling for state space models', *Biometrika* **81**(3), 541–553.

Casella, G. and George, E. (1992), 'Explaining the Gibbs sampler', *The American Statistician* **46**(3), 167–174.

Chen, M.-H. and Schmeiser, B. (1998), 'Toward black-box sampling: A random-direction interior-point Markov chain approach', *Journal of Computational & Graphical Statistics* **7**(1), 1–23.

Cheung, Y.-W. (1993), 'Tests for fractional integration: A Monte Carlo investigation', *Journal of Time Series Analysis* **14**(5), 331–345.

Chib, S. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association* **90**(432), 1313–1321.

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

Chib, S. and Greenberg, E. (1996), 'Markov chain Monte Carlo simulation methods in econometrics', *Econometric Theory* **12**(3), 409–431.

Chib, S., Nardari, F. and Shephard, N. (1998), Markov chain Monte Carlo methods for generalized stochastic volatility models, Nuffield College working paper 1998-W21, University of Oxford.

Chib, S., Nardari, F. and Shephard, N. (2000), Markov chain Monte Carlo methods for generalized stochastic volatility models. Unpublished manuscript, revision of Chib, Nardari and Shephard (1998).

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**(4), 841–862.

Courant, R. (1954), *Differential and Integral Calculus*, Vol. II, 2 edn, Blackie, London.

Davies, R. B. and Harte, D. S. (1987), 'Tests for Hurst effect', *Biometrika* **74**, 95–102.

De Jong, P. (1989), 'Smoothing and interpolation with the state-space model', *Journal of the American Statistical Association* **84**(408), 1082–1088.

De Jong, P. and Shephard, N. (1995), 'The simulation smoother for time series models', *Biometrika* **82**(2), 339–350.

Dickey, J. (1971), 'The weighted likelihood ratio, linear hypotheses on normal location parameters', *Annals of Statistics* **42**, 204–224.

Doornik, J. A. (1999), *Object-Oriented Matrix Programming using Ox*, 3rd edn, Timberlake Consultants Ltd, London. See `http://www.nuff.ox.ac.uk/Users/Doornik`.

Doornik, J. A. and Hendry, D. F. (2001), *Econometric Modelling using PcGive*, Vol. III, London: Timberlake Consultants Ltd.

Doornik, J. A. and Ooms, M. (1999), A package for estimating, forecasting and simulating Arfima models: Arfima package 1.0 for Ox, Technical report, Nuffield College, Oxford, UK.

Durbin, J. and Koopman, S. J. (2000), 'Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives', *Journal of the Royal Statistical Society, Series B* **62**(1), 3–56.

Engel, C. and Hamilton, J. D. (1990), 'Long swings in the dollar: Are they in the data and do markets know it?', *American Economic Review* **4**, 689–714.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations', *Econometrica* **50**, 987–1008.

Engle, R. F. (1995), *ARCH: Selected Readings*, Advanced Texts in Econometrics, Oxford: Oxford University Press.

Evans, M. D. D. and Lewis, K. K. (1995), 'Do long-term swings in the dollar affect estimates of the risk premia?', *Review of Financial Studies* **8**(3), 709–742.

Franses, P. H. (1998), *Time Series Models for Business and Economic Forecasting*, Cambridge University Press.

Frühwirth-Schnatter, S. (1994), 'Data augmentation and dynamic linear models', *Journal of Time Series Analysis* **15**(2), 183–202.

Gali, J. and Gertler, M. (2000), Inflation dynamics: A structural econometric analysis, Technical Report 7551, National Bureau of Economic Research.

Gelfand, A. E. and Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**(410), 398–409.

Geman, S. and Geman, D. (1984), 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, 721–741.

Geweke, J. (1989), 'Exact predictive densities for linear models with ARCH disturbances', *Journal of Econometrics* **40**, 63–86.

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 4: Proceedings of the Fourth Valencia International Meeting', Oxford: Clarendon Press, pp. 169–193.

Geweke, J. (1999), 'Using simulation methods for Bayesian econometric models: Inference, development, and communication', *Econometric Reviews* **18**(1), 1–73.

Geweke, J. and Zhou, G. (1996), 'Measuring the pricing error of the arbitrage pricing theory', *Review of Financial Studies* **9**(2), 557–588.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J., eds (1996), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

Gilks, W. R., Roberts, G. O. and George, E. I. (1994), 'Adaptive direction sampling', *The Statistician* **43**, 179–189.

Gradshteyn, I. S. and Ryzhik, I. M. (1965), *Tables of Integrals, Series and Products*, Academic Press, London.

Granger, C. W. J. and Joyeux, R. (1980), 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis* **1**, 15–29.

Green, P. J. (1995), 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika* **82**, 711–732.

Greene, W. H. (1990), *Econometric Analysis*, second edn, Prentice Hall, New York.

Hammersley, J. M. and Handscomb, D. C. (1964), *Monte Carlo Methods*, Methuen & Co, London.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A. C., Ruiz, E. and Sentana, E. (1992), 'Unobserved component time series with ARCH disturbances', *Journal of Econometrics* **52**, 129–157.

Harvey, A. C., Ruiz, E. and Shephard, N. (1994), 'Multivariate stochastic variance models', *Review of Economic Studies* **61**, 247–264.

Harvey, D. I., Leybourne, S. J. and Newbold, P. (1998), 'Tests for forecast encompassing', *Journal of Business & Economic Statistics* **16**, 254–259.

Harvey, D. I. and Newbold, P. (2000), 'Tests for multiple forecast encompassing', *Journal of Applied Econometrics* **15**(5), 471–482.

Hassler, U. and Wolters, J. (1995), 'Long memory in inflation rates: International evidence', *Journal of Business & Economic Statistics* **13**(1), 37–46.

Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**, 97–109.

Hauser, M. A. (1999), 'Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study', *Journal of Statistical Planning and Inference* **80**(1–2), 229–256.

Hidalgo, J. and Robinson, P. M. (1996), 'Testing for structural change in a long-memory environment', *Journal of Econometrics* **70**(1), 159–174.

Hobert, J. P. and Casella, G. (1996), 'The effect of improper priors on Gibbs sampling in hierarchical linear mixed models', *Journal of the American Statistical Association* **91**(436), 1461–1473.

Hooker, M. A. (1999), Are oil shocks inflationary? Asymmetric and nonlinear specifications versus changes in regime, Technical Report 1999-65, Federal Reserve Board.

Hosking, J. R. M. (1981), 'Fractional differencing', *Biometrika* **68**(1), 165–176.

Hsu, C.-C. (2000), Long memory or structural change: Testing method and empirical examination. Unpublished manuscript.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994), 'Bayesian analysis of stochastic volatility models', *Journal of Business & Economic Statistics* **12**, 371–417.

Jorion, P. (1985), 'International portfolio diversification with estimation risk', *Journal of Business* **58**(3), 259–278.

Jorion, P. (1986), 'Bayes-Stein estimation for portfolio analysis', *Journal of Financial and Quantitative Analysis* **21**, 297–292.

Jorion, P. (1997), *Value at Risk: The New Benchmark for Controlling Market Risk*, New York: McGraw-Hill.

JP Morgan (1997), RiskMetrics, Technical document, `http://www.riskmetrics.com`.

Kandel, S., McCulloch, R. and Stambaugh, R. (1995), 'Bayesian inference and portfolio efficiency', *Review of Financial Studies* **8**, 1–53.

Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Kass, R. E., Tierney, L. and Raftery, A. E. (1990), The validity of posterior expansions based on Laplace's method, *in* S. Geisser, ed., 'Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard', North Holland, Amsterdam, pp. 473–488.

Kim, S., Shephard, N. and Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **64**, 361–393.

Kleibergen, F. R. and Van Dijk, H. K. (1993), 'Non-stationarity in GARCH models: A Bayesian analysis', *Journal of Applied Econometrics* **8**, S41–S61.

Kloek, T. and Van Dijk, H. K. (1978), 'Bayesian estimates of equation system parameters: An application of integration by Monte Carlo', *Econometrica* **46**, 1–20.

Koop, G. and Van Dijk, H. K. (2000), 'Testing for integration using evolving trend and seasonal models: A Bayesian approach', *Journal of Econometrics* **97**(2), 261–291.

Koopman, S. J., Shephard, N. and Doornik, J. A. (1999), 'Statistical algorithms for models in state space using SsfPack 2.2', *Econometrics Journal* **2**, 107–160.

Lamoureux, C. and Lastrapes, W. (1990), 'Persistence-in-variance, structural change and the GARCH model', *Journal of Business & Economic Statistics* **8**, 225–234.

LeBaron, B. (1999), 'Technical trading rule profitability and foreign exchange intervention', *Journal of International Economics* **49**, 125–143.

Liu, J. S. and Sabatti, C. (1998), 'Simulated sintering: Markov Chain Monte Carlo with spaces of varying dimensions', *Proceedings of the 6th International Meeting on Bayesian Statistics, Valencia, Spain* .

Marinari, E. and Parisi, G. (1992), 'Simulated tempering: A new Monte Carlo scheme', *Europhysics Letters* **19**(6), 451–458.

Mark, N. C. (1995), 'Exchange rates and fundamentals: Evidence on long-horizon predictability', *American Economic Review* **85**(1), 201–218.

McCulloch, R. E. and Rossi, P. (1992), 'Bayes factors for nonlinear hypotheses and likelihood distributions', *Biometrika* **79**, 663–676.

McCulloch, R. and Rossi, P. (1990), 'Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory', *Journal of Financial Economics* **28**, 7–38.

McCulloch, R. and Rossi, P. (1991), 'A Bayesian approach to testing the arbitrage pricing theory', *Journal of Econometrics* **49**, 141–168.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953), 'Equations of state calculations by fast computing machines', *Journal of Chemical Physics* **21**, 1087–1091.

Nelson, C. R. and Plosser, C. I. (1982), 'Trends and random walks in macroeconomic time series; some evidence and implications', *Journal of Monetary Economics* **10**, 139–162.

Nelson, D. B. (1990), 'Stationarity and persistence in the GARCH(1, 1) model', *Econometric Theory* **6**, 318–334.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**, 703–708.

Newton, M. A. and Raftery, A. E. (1994), 'Approximate Bayesian inference by the weighted likelihood bootstrap', *Journal of the Royal Statistical Society, Series B* **3**, 3–48.

Ooms, M. (1996), Long memory and seasonality in US consumer price inflation: An empirical investigation at varying levels of aggregation. Unpublished manuscript.

Ooms, M. and Hassler, U. (1997), 'On the effect of seasonal adjustment on the log-periodogram regression', *Economics Letters* **56**, 135–141.

Perron, P. (1989), 'The great crash, the oil price shock, and the unit root hypothesis', *Econometrica* **57**(6), 1361–1401.

Perron, P. and Vogelsang, T. J. (1992), 'Testing for a unit root in a time series with a changing mean: Corrections and extensions', *Journal of Business & Economic Statistics* **10**(4), 467–469.

Richardson, S. and Green, P. J. (1997), 'On Bayesian analysis of mixtures with an unknown number of components', *Journal of the Royal Statistical Society, Series B* **59**(4), 731–792.

Ritter, C. and Tanner, M. A. (1992), 'Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler', *Journal of the American Statistical Association* **87**(419), 861–868.

Rubin, D. B. (1987), 'Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm', *Journal of the American Statistical Association* **82**, 543–546.

Satagopan, J. M. and Newton, M. A. (2000), Easy estimation of normalizing constants and Bayes factors from posterior simulation: Stabilizing the harmonic mean estimator. Unpublished manuscript.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, New York.

Smith, A. F. M. and Roberts, G. O. (1993), 'Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods', *Journal of the Royal Statistical Society, Series B* **55**(1), 3–24.

Solnik, B. H. (2000), *International Investments*, The Addison-Wesley Series in Finance, 4th edn, Addison-Wesley, Reading, Mass.

Sowell, F. (1992), 'Maximum likelihood estimation of stationary univariate fractionally integrated time series models', *Journal of Econometrics* **53**, 165–188.

Stock, J. H. and Watson, M. W. (1999), 'Forecasting inflation', *Journal of Monetary Economics* **44**, 293–335.

Tanner, M. A. and Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association* **82**, 528–550.

Taylor, S. J. (1994), 'Modeling stochastic volatility: A review and comparative study', *Mathematical Finance* **4**(2), 183–204.

Tierney, L. (1994), 'Markov chains for exploring posterior distributions', *Annals of Statistics* **22**, 1701–1762.

Van Dijk, D. J. C. (1999), Smooth Transition Models: Extensions and Outlier Robust Inference, PhD thesis, Tinbergen Institute, Erasmus University Rotterdam. TI 200.

Van Dijk, H. K. and Kloek, T. (1980), 'Further experience in Bayesian analysis using Monte Carlo integration', *Journal of Econometrics* **14**, 307–328.

Van Dijk, H. K., Kloek, T. and Boender, C. G. E. (1985), 'Posterior moments computed by mixed integration', *Journal of Econometrics* **29**, 3–18.

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**(430), 614–618.

West, K. D. (2001), 'Tests for forecast encompassing when forecasts depend on estimated regression parameters', *Journal of Business & Economic Statistics* **19**, 29–33.

West, M., Harrison, P. J. and Migon, H. S. (1985), 'Dynamic generalized linear models and Bayesian forecasting', *Journal of the American Statistical Association* **80**(389), 73–98.

Yu, B. and Mykland, P. (1998), 'Looking at Markov samplers through cusum path plots: A simple diagnostic idea', *Statistics and Computing* **8**(3), 275–286.

Zeger, S. L. and Karim, M. R. (1991), 'Generalized linear models with random effects: A Gibbs sampling approach', *Journal of the American Statistical Association* **86**, 79–86.

Nederlandse samenvatting (Summary in Dutch)

Inleiding

Eén van de onderwerpen binnen de Econometrie is het zoeken naar de structuur die schuil gaat achter macroeconomische reeksen. Dit proefschrift gaat in op twee van dergelijke reeksen, namelijk inflatie en wisselkoers reeksen. Uiteraard zijn dergelijke reeksen al vele malen geanalyseerd, met wisselend resultaat. In dit proefschrift wordt gepoogd op nieuwe wijze de data te analyseren, waarbij speciaal aandacht is voor modellen met tijdsvariërende parameters. Uit de gereedschapskist van de econometrie worden diverse technieken tevoorschijn gehaald die de beste (lees: meest gedetailleerde, waarheidsgetrouwe) beschrijving beloven te leveren van de structuren die aan de inflatie en wisselkoers reeksen ten grondslag liggen.

Hierna worden de diverse hoofdstukken, met hun conclusies, kort beschreven.

Inflatie

Hoofdstuk 2, gebaseerd op Bos, Franses en Ooms (1999), beschrijft een model voor de inflatie in de G7 landen (Canada, Frankrijk, Duitsland, Italië, Japan, Verenigd Koninkrijk en de Verenigde Staten). Onder economen en econometristen zijn de meningen verdeeld of de inflatie (en ook andere) reeksen wel of niet stationair zijn rondom een vast gemiddelde. Niet-stationariteit van een reeks wil zeggen dat de reeks geen gemiddelde heeft maar alle kanten uit kan lopen (zie bijvoorbeeld Nelson en Plosser (1982), waar deze discussie over het detecteren van zogenoemde eenheidswortels begonnen is).

Een tussenweg die veel helpt bij het modelleren van inflatie is het toelaten van fractionele integratie (FI). Deze fractionele integratie wordt bereikt door in het AutoRegressive Integrated Moving Average (ARIMA) model van Box et al. (1994), dat geschreven kan worden als

$$\begin{aligned} \Phi(L)(1-L)^d y_t &= \Theta(L)\epsilon_t, \\ \Phi(L) &= 1 - \phi_1 L - \cdots - \phi_p L^p, \\ \Theta(L) &= 1 + \theta_1 L + \cdots + \theta_q L^q, \\ \epsilon_t &\sim i.i.d.(0, \sigma_\epsilon^2) \end{aligned} \tag{1.1'}$$

toe te laten dat de parameter d, die de mate van integratie modelleert, ook niet-gehele waarden aanneemt (zie de introductie, sectie 1.2 en hoofdstukken 2 en 3 voor een nadere

uitleg van het ARFIMA model). Het resulterende ARFIMA model wordt ook een lang-geheugen model genoemd, aangezien het effect van een schok in de reeks bij deze modellen slechts zeer langzaam wegebt. Figuur 3.1 op pagina 33 bijvoorbeeld toont in het middelste paneel de autocorrelatie van de U.S. core inflatie; te zien valt hoe de correlatie slechts langzaam uitdooft.

In inflatie reeksen zijn er enkele waarnemingen die achteraf een zeer langdurig effect gehad blijken te hebben, namelijk de waarnemingen rondom het begin en einde van de oliecrises. Hoofdstuk 2 onderzoekt, eerst in een simulatie, wat het effect is van een veranderend gemiddelde van de reeks op de schatting van de parameter d, die de graad van integratie weergeeft. Vervolgens wordt voor de maandelijkse inflatie reeksen van de G7 landen een variërend gemiddelde toegelaten in het model. Uit de simulatie blijkt dat het veronachtzamen van een verandering in het gemiddelde een te hoge schatting voor d oplevert. Voor de G7 landen vinden we dat de mate van integratie sterk daalt, zo sterk dat bij Canada en Japan er geen significante fractionele integratie meer overblijft ($\hat{d} \approx 0$). Om dit effect te bereiken zijn in principe twee breuken in het gemiddelde, in juli 1973 en juli 1982, voldoende. Het blijkt vrij weinig uit te maken voor de parameter d om twee extra veranderingen van het gemiddelde, tussen de twee oliecrises in, in juli 1976 en januari 1979, toe te staan.

Het daaropvolgende hoofdstuk, hoofdstuk 3, bouwt voort op deze resultaten (gepubliceerd als Bos, Franses en Ooms 2001). Het schatten van modellen voor inflatie op zichzelf kan een interessante exercitie zijn, maar voor een beleidsmaker wordt het pas echt van belang zodra er een voorspelling voor de toekomstige inflatie uit voortkomt. Nu blijkt dat de keuze voor een stationair model met $d = 0$, een geïntegreerd/niet-stationair model met $d = 1$ of een fractioneel geïntegreerd model met een geschatte $\hat{d}$ veel uitmaakt voor de onzekerheid van de voorspelling. Deze voorspelonzekerheid komt tot uitdrukking in de breedte van het voorspelinterval. Vooral als er meer dan enkele maanden, tot bijvoorbeeld twee jaar vooruit voorspeld wordt, heeft de parameter d een grote invloed op de breedte van het voorspelinterval. Hoe groter de waarde voor d, des te breder is het voorspelinterval; voor het geïntegreerde ARIMA(1,1,1) model worden deze intervallen tè groot geschat.

Het bepalen van voorspelintervallen voor Amerikaanse basisinflatie (i.e., inflatie exclusief het effect van veranderingen in voedsel- en energieprijzen) wordt uitgevoerd voor een reeks van modellen. Zowel kort-geheugen ARMA(1,1), geïntegreerde ARIMA(1,1,1) en lang-geheugen ARFIMA(1,d,1) modellen passeren de revue. Daarbij worden diverse verklarende variabelen als de korte rente, de werkloosheid en/of het verschil tussen de lange en korte rente opgenomen (de modellen worden dan aangeduid als ARFIMAX modellen, met de X voor *eXplanatory*, voor de verklarende variabele). Op korte termijn, bij een voorspelling tot ongeveer 3/6 maanden vooruit, dragen de verklarende variabelen bij aan de precisie van de voorspelling. Op langere termijn blijkt dit effect niet sterk of zelfs afwezig te zijn.

Uit de schatting blijkt dat de geschatte residuele variantie van de modellen veel groter, tot een factor 2.7 keer groter aan toe, te zijn dan de werkelijke variantie van de residuen over de evaluatie periode 1984:01-1999:11. Tijdens het regime van de centrale bank-presidenten Volcker en Greenspan fluctueert de inflatie duidelijk minder dan tijdens de oliecrises, en ook iets minder dan in de jaren zestig. In secties 3.3 en 3.4 wordt

gecorrigeerd voor de variërende variantie, door drie verschillende variantieregimes te veronderstellen. Waarnemingen uit voorbije perioden worden minder zwaar meegewogen, zodat deze minder invloed op de schattingsresultaten zullen hebben. Het effect van deze aanpassing is dat de residuele variantie inderdaad beter geschat wordt, alhoewel vooral het ARFIMA(X)(1,1,1) model nog steeds een te groot voorspelinterval oplevert bij langere-termijn voorspellingen.

Bayesiaanse simulatietechnieken

Hoofdstuk 2 en 3 gebruiken de klassieke benadering van de statistiek, waarbij er van uit wordt gegaan dat iedere parameter in het model een zekere ware waarde heeft. Deze waarde wordt vervolgens via de methode van maximale aannemelijkheid geschat. In hoofdstuk 5 (zie onder) is de onzekerheid van de parameters groter dan in hoofdstukken 2 en 3. Hierdoor is ook de onnauwkeurigheid bij het schatten van de parameters van belang voor de uitkomsten. In zo'n geval is de Bayesiaanse statistiek beter geschikt voor de analyse dan de klassieke statistiek.

In de Bayesiaanse benadering van de statistiek is iedere parameter onzeker, en heeft bijgevolg een kansverdeling. Voordat er gegevens tot onze beschikking staan wordt deze kansverdeling de prior- of vóór-verdeling genoemd. De prior-verdeling omschrijft onze kennis aangaande de parameter die we vooraf, uit de structuur van het model of uit eerdere soortgelijke onderzoeken hebben. Na het waarnemen kan de prior-verdeling aangepast worden tot de posterior- of na-verdeling.

Deze beknopte inleiding brengt ons tot het onderwerp van hoofdstuk 4: de Bayesiaanse simulatietechnieken. Voor slechts enkele modellen is het mogelijk om analytisch de na-verdeling te bepalen op grond van de vóór-verdeling en de data. In andere gevallen zullen numerieke technieken gebruikt moeten worden. De simulatietechnieken die in dit hoofdstuk besproken worden dienen om de verdeling van de parameters te bepalen. Ook omvat het hoofdstuk een onderdeel (sectie 4.4) waarin het concept van de marginale aannemelijkheid (*marginal likelihood*) wordt uitgelegd. Deze marginale aannemelijkheid, in het hoofdstuk aangeduid als logaritme met log-m, kan dienen om modellen te vergelijken wat betreft hun aanpassing aan de data. Het berekenen van de log-m is in veel gevallen een lastig karwei. Diverse berekeningsmethoden zoals deze bekend, maar niet veel gebruikt, zijn, worden besproken en vergeleken in sectie 4.4.

Het laatste deel van het hoofdstuk past de diverse simulatietechnieken toe op een eenvoudig model, dat in uitgebreide vorm in hoofdstuk 5 weer op zal duiken. In het voorbeeld worden diverse eigenschappen van de technieken, zoals correlatie tussen de trekkingen en duur van de berekeningen, becommentarieerd. De marginale aannemelijkheid wordt op vele manieren berekend, voor twee varianten van het model, en de numerieke precisie van de berekeningsmethoden wordt geëvalueerd. Eenvoudige berekeningsmethoden van de marginale aannemelijkheid op basis van een LaPlace of kernel benadering van de na-verdeling blijken een goede precisie op te leveren.

Wisselkoers risico

Het laatste toegepaste hoofdstuk van dit proefschrift gaat in op het afdekken van wisselkoers risico. Het uitgangspunt van hoofdstuk 5 is een grote investeerder met een belang in het buitenland, dat al naar gelang de veranderingen in de wisselkoers meer of minder waard kan worden. Zo werd over de periode 1/1/1998–31/12/2000 de dollar bijna 16% meer waard ten opzichte van de Duitse mark. Deze waardestijging was echter opgebouwd uit een daling van 7.4% in het eerste jaar en stijgingen van 17.1 en 6.8% in de laatste twee jaren, wat laat zien dat deze waardevermeerdering bepaald niet risicoloos was.

Het kan aantrekkelijk zijn om het wisselkoers risico af te dekken. Een gangbaar instrument hiervoor is het *forward contract*, waarbij, kort gezegd, de onzekere koersverandering ingewisseld wordt voor het zekere verschil tussen de rentestanden in de beide landen. Het onderliggende hoofdstuk, dat deels is gebaseerd op het artikel van Bos, Mahieu en Van Dijk (2000*a*), introduceert *state space* (of toestands-ruimte) modellen die zo goed mogelijk de trendmatige bewegingen in de trend en de variantie van de wisselkoersen proberen te volgen en voorspellen. Aan de hand van deze modellen, acht in getal, wordt een voorspelverdeling voor de wisselkoers één dag vooruit bepaald, die als invoer dient voor een optimalisatie proces waarbij een *power* nutsfunctie geoptimaliseerd wordt.

Deze analyse wordt uitgevoerd op de wisselkoersen tussen de Duitse mark en de dollar, en tussen de Japanse yen en de dollar (in beide richtingen, zowel voor de DM/USD en Yen/USD als voor de USD/DM en USD/Yen koersrisico's). Uit de bepalingen van de marginale aannemelijkheid blijkt een model met variërend gemiddelde en stochastisch variërende variantie het beste te voldoen voor de wisselkoers tussen DMark en dollar. Voor de yen/dollar reeks blijkt het lastiger te zijn een goed model te vinden; het belangrijkste is om rekening te houden met uitbijters door het opnemen van een Student-t verdeling voor de storingsterm. Daarnaast helpen het variërend gemiddelde en een GARCH proces voor de variantie ook om een betere aanpassing van het model aan de data te krijgen.

De resultaten zijn het sterkst voor de wisselkoers tussen de mark en de dollar. Deze reeks levert een optimaal model op met meer voorspellende kracht dan het model voor de yen/dollar data. De modellen die het best scoorden op de marginale aannemelijkheid blijken ook goed te werken bij het afdekken van de koersrisico's, alhoewel voor de yen/dollar en dollar/yen analyse blijkt dat een risicomanager die vrij tolerant is ten opzichte van risico ook kan kiezen een zogenoemd *Local Level* model te hanteren, omdat dit model iets beter in staat is een signaal te filteren uit de wisselkoers data.

De modelresultaten worden vergeleken met resultaten van drie deterministische strategiën, te weten:

i. Dek het risico nooit af,

ii. Dek continu af,

iii. Dek het risico voor de dag van morgen af als er vandaag een koersdaling plaats vond.

In ieder van de jaren uit de evaluatie periode 1998-2000 blijkt één van deze regels een hoger nut op te leveren dan de model-gebaseerde strategiën. Echter, van te voren is met geen mogelijkheid te voorspellen wèlke deterministische strategie het beste zal werken.

Er kan geconcludeerd worden dat de modellen met variërend gemiddelde en variërende variantie inderdaad helpen om het risico te beperken, en om tegelijkertijd niet te veel van het mogelijke rendement van koersstijgingen te missen.

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

200. D.J.C. VAN DIJK, *Smooth transition models: Extensions and outlier robust inference.*

201. E. KAPER, *Panel effects in consumer research. Statistical models for underreporting.*

202. K. SADIRAJ, *Albania: Transition to a market economy.*

203. A. RANASINGHE, *A pearl of great price: The free education system in Sri Lanka.*

204. P.J. VAN DER SLUIS, *Estimation and inference with the efficient method of moments: With applications to stochastic volatility models and option pricing.*

205. P.G.W. ALDERS, *Family ties in an aging world. Causes and consequences of fertility shifts.*

206. P. VAN HASSELT, *Dynamics of price formation in financial markets.*

207. K. VERWEIRE, *Performance consequences of financial conglomeration with an empirical analysis in Belgium and the Netherlands.*

208. P.W.T. GHIJSEN, *Labour and technology in Japan. An analysis of labour adjustment and technological change.*

209. E. DRISSEN, *Government decisions on income redistribution and public production. A political-economic general equilibrium approach.*

210. J. SPREEUW, *Heterogeneity of hazard rates in insurance.*

211. G.T. POST, *Finding the frontier: Methodological advances in data envelopment analysis.*

212. L.D. MEIJERS, *Ruimtelijke netwerken van de zakelijke dienstverlening.*

213. R.P. PLASMEIJER, *Maintenance optimisation techniques for the preservation of highways.*

214. J. TUINSTRA, *Price dynamics in equilibrium models.*

215. P.A. GROENENDIJK, *Essays on exchange rate dynamics.*

216. S.M. DE BRUYN, *Economic growth and the environment: An empirical analysis.*

217. Y. SCHIPPER, *Market structure and environmental costs in aviation. A welfare analysis of European air transport reform.*

218. E. VAN GAMEREN, *The internal economics of firms. An investigation into the labour mobility within firms.*

219. A.J. DUR, *Political institutions and economic policy choice.*

220. B.E. BAARSMA, *Monetary valuation of environmental goods: Alternatives to contingent valuation.*

221. E.C. VAN DER SLUIS-DEN DIKKEN, *Management learning and development: The role of learning opportunities and learning behavior in management development and career success.*

222. A.J.H. PELS, *Airport economics and policy: Efficiency, competition and interaction with airlines.*

223. B. VAN DER KLAAUW, *Unemployment duration determinants and policy evaluation.*

224. F. MEDDA, *The assembled city: Analyses of multiple urban dimensions.*

225. A.F. TIEMAN, *Evolutionary game theory and equilibrium selection.*

226. R.R.P. KOUWENBERG, *Dynamic asset liability management.*

227. J.S. SIDHU, *Organization mission, business domain orientation and performance: A conceptual and empirical inquiry.*

228. G. ROMIJN, *Economic dynamics of Dutch construction.*

229. M.C. VERSANTVOORT, *Analysing labour supply in a life style perspective.*

230. J.J.J. GROEN, *Testing multi-country exchange rate models.*

231. C.F.A. VAN WESENBEECK, *How to deal with imperfect competition: introducing game-theoretical concepts in general equilibrium model of international trade.*

232. M.L. NDOEN, *Migrants and entrepreneurial activities in peripheral Indonesia. A socioeconomic model of profit-seeking behaviour.*

233. L.A. GROGAN, *Labour market transitions of individuals in eastern and western Europe.*

234. E.G. VAN DE MORTEL, *An institutional approach to transition processes.*

235. P.H. VAN OIJEN, *Essays on corporate governance.*

236. H.M.M. VAN GOOR, *Banken en industriefinanciering in de 19e eeuw. De relatie tussen Mees en Stork, Van den Bergh gaat naar Engeland.*

237. F.R.M. PORTRAIT, *Long-term care services for the Dutch elderly. An investigation into the process of utilization.*

238. M. VAN DE VELDEN, *Topics in correspondence analysis.*

239. G. DRAISMA, *Parametric and semi-parametric methods in extreme value theory.*

240. I.F.C. MULDER, *Soil degradation in Benin: Farmers' perceptions and responses.*

241. A.W. SALY, *Corporate entrepreneurship. Antecedents and consequences of entrepreneurship in large established firms.*

242. S. VAN VELZEN, *Supplements to the economics of household behavior.*

243. R.A. VAN DER GOOT, *High performance linda using a class library.*

244. E. KUIPER, *The most valuable of all Capital. A gender reading of economic texts.*

245. P. KLIJNSMIT, *Voluntary corporate governance disclosures; An empirical investigation of UK practices.*

246. P.J.G. TANG, *Essays on economic growth and imperfect markets.*

247. H. HOOGEVEEN, *Risk and insurance in rural Zimbabwe.*

248. A.J. VAN DER VLIST, *Residential mobility and commuting.*

249. G.E. BIJWAARD, *Rank estimation of duration models.*

250. A.B. BERKELAAR, *Strategic asset allocation and asset pricing.*

251. B.J. VAN PRAAG, *Earnings management; Empirical evidence on value relevance and income smoothing.*

252. E. PEEK, *Discretion in financial reporting and properties of analysts' earnings forecasts.*

253. N. JONKER, *Job performance and career prospects of auditors.*

254. M.J.G. BUN, *Accurate statistical analysis in dynamic panel data models.*

255. P.C. VERHOEF, *Analyzing customer relationships: Linking attitudes and marketing instruments to customer behavior.*