Econometric Institute Report EI-9936/A

Daily Exchange Rate Behaviour and Hedging of Currency Risk

Charles S. Bos *
Tinbergen Institute, Erasmus University Rotterdam

Ronald J. Mahieu
Rotterdam School of Management, Erasmus University Rotterdam

Herman K. van Dijk
Econometric Institute, Erasmus University Rotterdam

August 20, 1999

Abstract

Exchange rates typically exhibit time-varying patterns in both means and variances. The histograms of such series indicate heavy tails. In this paper we construct models which enable a decision-maker to analyze the implications of such time series patterns for currency risk management. Our approach is Bayesian where extensive use is made of Markov chain Monte Carlo methods. The effects of several model characteristics (unit roots, GARCH, stochastic volatility, heavy tailed disturbance densities) are investigated in relation to the hedging decision strategies. Consequently, we can make a distinction between statistical relevance of model specifications, and the economic consequences from a risk management point of view. The empirical results suggest that econometric modelling of heavy tails and time-varying means and variances pays off compared to a efficient markets model. The different ways to measure persistence and changing volatilities appear to strongly influence the hedging decision the investor faces.

JEL classification: C11, C15, C44, E47, G15
Keywords: Bayesian decision making, econometric modelling, exchange rates, risk management, forward contracts, stochastic volatility, GARCH

*Correspondence to Charles Bos, Tinbergen Institute, Erasmus University Rotterdam, Burg. Oudlaan 50, NL-3062 PA Rotterdam, The Netherlands. Email: cbos@few.eur.nl, URL: http://www2.tinbergen.nl/~cbos

1 Introduction

When investing abroad investors are naturally confronted with the problem whether or not to hedge their currency exposures. Currencies do add short-term volatility to the portfolio. Another time series feature of exchange rates is that periods of depreciation and appreciation alternate.[1] From an investor's viewpoint this 'local trend' behaviour is very interesting, as it provides opportunities to formulate an active currency hedge program, that could improve the efficiency of the international portfolio.

In this paper we concentrate on tactical strategies for exchange rate management. We assume that the correlations with the underlying exposures are zero. This means that the decision whether or not to hedge currency exposures is made independently from the underlying portfolio allocation. In the finance industry this approach to currency hedging is called 'currency overlay management' for obvious reasons. The hedging is performed through forward contracts. We are interested in examining the impacts of several univariate time series models on the hedging decision.

In order to make the currency risk hedge strategy operational we need to be explicit on the objective functions. We investigate the consequences of (1) a standard mean-variance utility function and (2) an objective function based on Value at Risk.[2] In our setup the variable of interest to the decision maker is the exchange rate. In Figure 1 we plotted the German DMark/US dollar daily exchange rate series for the period January 1982 until January 1999, which we use in the empirical part of this paper. By casual inspection of this series several prominent features show up. First, the exchange rate is either appreciating or depreciating locally. A second feature is that exchange rate volatility is clustered, i.e. periods with low volatility are followed with periods with much higher volatility. This can be seen more clearly from the lower panel in Figure 1, which presents the daily returns on the Dmark/US Dollar. Third, the time series and simple histogram analysis indicate that the distribution of returns is heavy-tailed.

We introduce a State Space model for the time-varying mean, which is augmented with a Generalized Conditional Heteroscedastic (GARCH), a Stochastic Volatility (SV) for the time-varying variance, and with a Student t model for the heavy tailed disturbances. State Space models are nowadays widely used for describing time varying patterns in economic series, see e.g. Harvey (1989) and the references cited there. For inference and decision purposes we apply Bayesian methods where extensive use is made of Markov chain Monte Carlo (MCMC, see Smith & Roberts 1993, Chib & Greenberg 1995). In the recent literature these methods have been succesfully applied for studying separately the pattern of varying means (see Carter & Kohn 1994, Koop & van Dijk 1999) and the pattern of varying volatilities (see Kim, Shephard & Chib 1998). In this paper we integrate the models for the analysis of varying means and varying variances, and the models of varying means 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.

The outline of this paper is as follows. In Section 2 we introduce the econometric models and our Bayesian estimation method using MCMC. In Section 3 we present the decision

[1] This observation has been successfully captured in switching regime models, which were first introduced by Engel & Hamilton (1990).

[2] See Jorion (1997) for more information on Value at Risk.

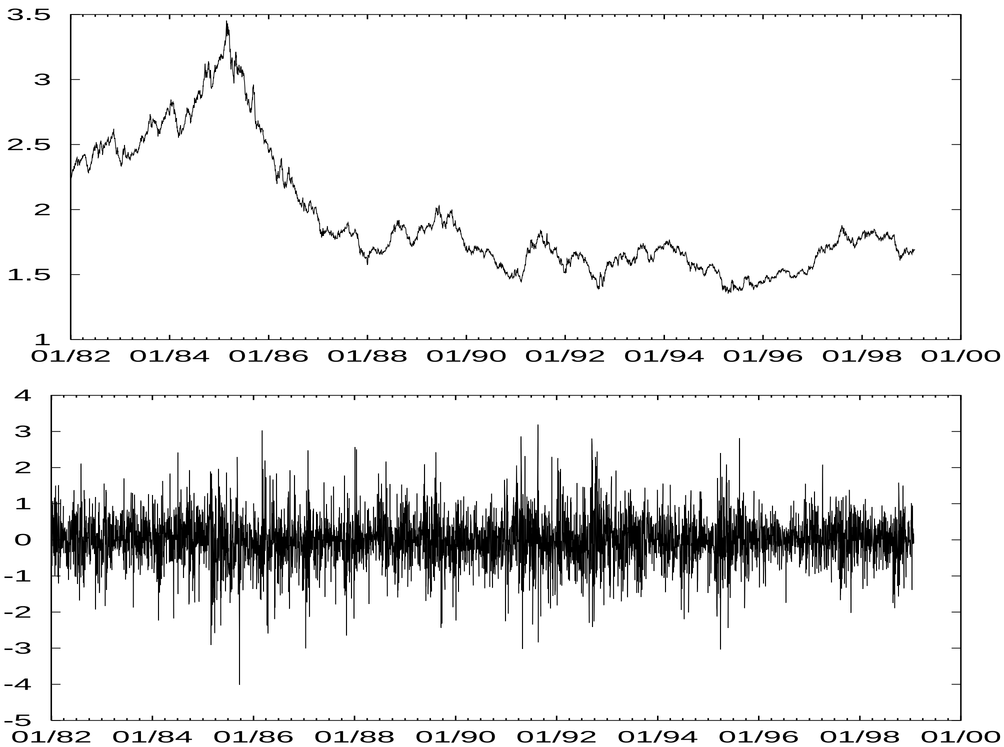

Figure 1: Daily exchange rate (upper panel) and returns (lower panel) of Deutschmark vs. US Dollar, 1/1/1982 until 26/1/1999

problems that we use in order to determine the actions an international investor can take with respect to the management of his currency exposures. Section 4 contains the results of applying the modelling techniques and hedging decisions on the German DMark/US Dollar exchange rate. Concluding remarks are given in Section 5.

2 Models for exchange rate returns

2.1 The models

Many models have been suggested for describing the time series properties of exchange rates, see e.g. Engel & Hamilton (1990). In this paper we concentrate on models that describe the most prominent data features of floating daily exchange rates. In Figure 1 we saw that the DMark / US Dollar exchange rate exhibits local trend behaviour. We model this feature by the following State Space model

$$s_t = \mu_t + \epsilon_t \tag{1}$$
$$\mu_t = \rho\mu_{t-1} + \eta_t, \qquad t = 1, .., T, \tag{2}$$

where s_t is the return on the exchange rate S_t, i.e. $s_t = 100 \times (\ln S_t - \ln S_{t-1})$. The unobserved mean component μ_t is modelled as an autoregressive process with disturbances η_t. Typically, the autoregressive parameter ρ would be close to one, signifying that the underlying state evolves slowly. This feature is hoped to pick up the periods of rising or falling prices. Intuitively, this equation states that the underlying return is persistent. The disturbances η_t are assumed to be independently and identically normally distributed with constant variance σ^2_η. The AR(1) model incorporates as a limiting case the fully integrated mean return model, when $\rho = 1$. This implies that the (log) level of the exchange rates follows an I(2) process. We note that in practice, the variance of η_t is so small that the I(1) behaviour of S_t overwhelms the I(2) effects. One can also take the limit case $\sigma^2_\eta = 0$. Then a model for s_t results which is white noise around a fixed mean μ. Though extremely simple, it is a basic model in many financial market models. More specifically, this model does not violate the efficient market hypothesis. [3]

The second main feature of financial series concerns the variance structure. Several model specifications have been suggested to account for the variance clustering in the data. See Bollerslev (1986), Nelson (1990), Engle (1995) or Taylor (1994). Conditioning on all information $\mathcal{I}_{t-1}$ available at time t, we write

$$\epsilon_t|\mathcal{I}_{t-1} \sim \mathcal{N}(0, \sigma^2_{\epsilon,t}). \tag{3}$$

The simplest model, ignoring the time dependance of volatility is written as

$$\sigma^2_{\epsilon,t} = \sigma^2_\epsilon \tag{4}$$

in which case a pure State Space model results. More flexibility is obtained when a GARCH disturbance process is allowed for, which is written as

$$\sigma^2_{\epsilon,t} = \delta\sigma^2_{\epsilon,t-1} + \omega + \alpha\epsilon^2_{t-1} \tag{5}$$

$$\delta \geq 0, \quad \omega > 0, \quad \alpha \geq 0, \quad \delta + \alpha < 1. \tag{6}$$

The restrictions on the parameters are sufficient to ensure strict positiveness of $\sigma^2_{\epsilon,t}$ and the existence of a finite value for the unconditional expectation $\mathsf{E}(\sigma^2_{\epsilon,t})$ (see Kleibergen & van Dijk 1993).

A second family of disturbance processes for ϵ_t results when we let the variances $\sigma^2_{\epsilon,t}$ change over time according to a stochastic volatility process. See, for example, Jacquier, Polson & Rossi (1994). The model used here is given as

$$\sigma^2_{\epsilon,t} = \exp(h_t) \tag{7}$$

$$h_t = \mu_h + \phi(h_{t-1} - \mu_h) + \xi_t \qquad 0 \leq \phi < 1 \tag{8}$$

$$\xi_t \sim \mathcal{N}(0, \sigma^2_\xi) \tag{9}$$

The GARCH and stochastic volatility models both allow for periods of lower or higher variance, where the variance process is correlated over time.

A third option is to assume the disturbance process ϵ_t in the observation equation has heavier tails than the normal density. We replace the assumption (3) with

$$\epsilon_t|\mathcal{I}_{t-1} \sim t(0, \sigma_\epsilon, \nu). \tag{10}$$

The expectation of ϵ_t still equals zero; the variance $\mathsf{Var}(\epsilon_t) = \sigma^2_\epsilon \nu/(\nu - 2), \nu > 2$.

[3] See also Fama (1991) for more details on the efficient markets hypothesis.

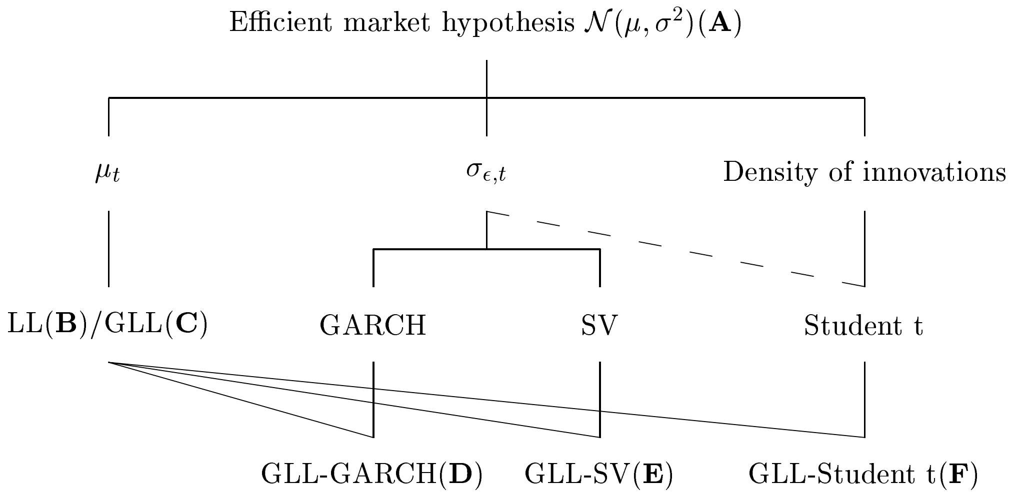

Figure 2: Relation between assumptions and models

Figure 2 summarizes the models that are evaluated in subsequent sections. The basic model is the one based on normality of the returns. Then there are three directions of generalization: At the mean of the process μ_t, at the variances σ_t^2, or at the shape of the density of the innovations ϵ_t.[4] More specifically, the third line in the figure indicate the models that we consider. The Generalized Local Level (GLL) model is combined with each of the other three generalizations, such that a broad range of competing models is found. The models are indicated by the letters **A**-**F** in the figure and are summarized in Table 1.

Model	Description	Corresponding equations	Parameters
A	White noise (WN)	(1)-(4), $\rho = 1$, $\sigma_\eta^2 = 0$	μ, σ_ϵ
B	Local level (LL)	(1)-(4), $\rho = 1$	$\sigma_\epsilon, \sigma_\eta$
C	GLL	(1)-(4)	$\rho, \sigma_\epsilon, \sigma_\eta$
D	GLL-GARCH	(1) - (3), (5)	$\rho, \sigma_\eta, \delta, \omega, \alpha$
E	GLL-SV	(1) - (3), (7)-(9)	$\rho, \sigma_\eta, \mu_h, \phi, \sigma_\xi$
F	GLL-Stud	(1), (2), (10)	$\rho, \sigma_\epsilon, \sigma_\eta, \nu$

Table 1: Summary of models

2.2 Prior structure

The inference and decision analysis is performed within a Bayesian framework. In this section, we present a set of priors on the parameters of the models.

In model **A**, White Noise, μ is a parameter for the mean return. On a daily basis the mean return is rather small. No further strong information is available. Therefore, a flat prior [-0.1%, 0.1%] is used.

[4] Note that the assumption of student t distributed disturbances can also be explained as a generalization in the direction of the variances σ_t^2, as is explained in Appendix B.

For the other models, the varying mean equation (2) applies. The parameter ρ governs the persistence of the mean series μ_t. From the stylized fact on the evolution of exchange we expect a large value of ρ, which is larger than 0, representing the local trend behaviour of the exchange rate series. A Beta prior with parameters $\alpha_\rho = 10$ and $\beta_\rho = 3$, which has expectation 0.77 and standard deviation 0.113, incorporates this empirical prior information on expected values of ρ. The prior for σ_η^2 is the Inverted Gamma-2 distribution, which is the conjugate prior for the residual variances as in Kim et al. (1998). The prior is chosen such that it has a mean of 0.008 and a standard deviation of 0.016.

In models **A**, **B** and **C** the disturbances ϵ_t are normally distributed with a variance σ_ϵ^2. The prior on σ_ϵ^2 is also taken to be inverted Gamma-2, with parameters that imply a prior mean of 0.5 and a standard deviation of $\frac{1}{2}\sqrt{2}$, respectively.

In model **D**, where GARCH-type disturbances $\epsilon_t, t = 1, .., T$ are modelled, priors are needed for the parameters δ, ω and α. The unconditional expectation of the variance of such a process is

$$\mathsf{E}(\sigma_{\epsilon,t}^2) = \frac{\omega}{1-\delta-\alpha} \tag{11}$$

We choose to be non-informative on the unconditional precision, i.e.

$$\pi(\delta, \omega, \alpha) \propto 1/\mathsf{E}(\sigma_{\epsilon,t}^2) = \frac{1-\delta-\alpha}{\omega} \tag{12}$$

Note that the original restrictions (6) ensure the strict positiveness of this prior density.

The Stochastic Volatility model, model **E** in Table 1, contains extra parameters μ_H, ϕ and σ_ξ^2. These are the mean and the correlation coefficient of the varying variance equation (8) and the variance of the disturbances. A uniform prior for μ_h on the interval [-3, 0] is weakly informative. The expected value of $\sigma_{\epsilon,t}^2$ lies between roughly 0.2 and 1 with these choices. Likewise, ϕ is a parameter of similar meaning as ρ above, modelling the persistence of the stochastic variance process. Positive correlation is expected, and also reasonable persistence. We use a Beta prior with parameters $\alpha_\phi = 10$ and $\beta_\phi = 3$. The variance parameter is assumed again to come from an Inverted Gamma-2 distribution, with hyperparameters $s_\xi = 1.5$ and $\nu_\xi = 5$, as in the case of the σ_ϵ^2 parameter.

The last model, **F**, includes the degrees of freedom parameter ν. A priori we want to allow for a broad range of values ν. Preliminary investigation learned us that no posterior mass is found at values of $\nu < 6$ or $\nu > 17$. To reduce the amount of computations we limit the prior to the discrete range $\nu = 6, \ldots, 16$. Each of the values of ν has equal prior weight. The fixed variance factor σ_ϵ^2 has the same prior as in models **A**-**C**.

In Table 2 we have collected all the prior distributions. We note that our priors may be characterized as weakly informative.

2.3 Constructing a posterior sample

For models **A**-**D** it is possible to write the likelihood function in a convenient prediction-error form, see Harvey (1989). The posterior distribution is obtained by multiplying the corresponding prior distribution with the likelihood function. Though the shape of this posterior might be highly non-normal, a general adaptive independent Metropolis-Hastings sampler (see Carter & Kohn 1996, Chib & Greenberg 1995, Koop & van Dijk 1999) with a normal candidate works well for obtaining a set of simulated parameter vectors from the target density. An adaptive sampling scheme is used: Several rounds of the sampler are run, with an

Parameter	Prior	Hyper-parameters
μ	Uniform(L_μ, H_μ)	$L_\mu = -0.1, H_\mu = 0.1$
σ^2_ϵ	IG-2$(s_\epsilon, \nu_\epsilon)$	$s_\epsilon = 1.5, \nu_\epsilon = 5$
ρ	Beta$(\alpha_\rho, \beta_\rho)$	$\alpha_\rho = 10, \beta_\rho = 3$
σ^2_η	IG-2(s_η, ν_η)	$s_\eta = 0.02, \nu_\eta = 4.5$
δ, ω, α	'Non-informative' $\propto (1-\delta-\alpha)/\omega$	-
μ_H	Uniform(L_{μ_H}, H_{μ_H})	$L_{\mu_H} = -3, H_{\mu_H} = 0$
ϕ	Beta$(\alpha_\phi, \beta_\phi)$	$\alpha_\phi = 10, \beta_\phi = 3$
σ^2_ξ	IG-2(s_ξ, ν_ξ)	$s_\xi = 1.5, \nu_\xi = 5$
ν	Discrete uniform(L_ν, H_ν)	$L_\nu = 6, H_\nu = 16$

Table 2: Description of priors used

update of the estimate of the location and scale of the target density to be used in the normal candidate density. The sampler is started at the maximum likelihood estimates of the location and scale.

For models **E** and **F**, the Stochastic Volatility and Student t models, we found the posterior distributions by a data augmentation scheme in order to integrate out the unobserved state variables $\sigma^2_{\epsilon,t}$. We make use of a Gibbs sampling scheme as in the article by Kim et al. (1998). See Appendix B for more details.

3 Currency hedging

3.1 The hedging decision

As noted in the introduction, we concentrate on tactical strategies for exchange rate management. The setting may be summarized as follows. Let s_{t+1} be the exchange rate return in the time interval $[t, t+1]$, defined as $s_{t+1} = \ln(S_{t+1}/S_t)$. Let $F_{t,\tau}$ be the current value of a forward contract with maturity date τ, which by arbitrage restrictions is equal to $F_{t,\tau} = S_t \exp(r^h_{t,\tau} - r^f_{t,\tau})$, with $r^h_{t,\tau}$ and $r^f_{t,\tau}$, the home and foreign risk-free interest rates with maturity τ, respectively. Define H_t as the fraction of the underlying exposure that is hedged with forward contracts. We refer to this variable as the hedge ratio. Through the definition of a forward contract we find that the currency return, hedged with forward contracts, is a weighted average of exchange rate return s_{t+1} and the difference between the home and foreign risk free interest rates with weights equal to the hedge ratio:

$$r_{t+1} \equiv (1 - H_t)s_{t+1} + H_t(r^h_{t,\tau} - r^f_{t,\tau}) \tag{13}$$

Given a time series model that captures exchange rate behaviour and all information up to time t, the investor wants to determine the hedge ratio that should apply to the next periods. In order to perform this task the investor needs to specify an objective function that captures the risk and return attitudes over some future time horizon.

Consider the situation that the investor wants to determine the hedge ratio for the next period $[t, t+1]$. Assume that the investor has a simple mean-variance objective function. In order to determine the value of the objective function we need projections for expected returns and variances. Based on a time series model we can compute the predicted mean

$(\mathsf{E}(r_{t+1}|I_t))$ and its variance $(\mathsf{Var}(r_{t+1}|I_t))$.[5] The variable I_t represents the information set containing information up to and including time t. Introduce the risk preference parameter λ. This parameter needs to be set by the investor in order to express his attitude towards risk in relation to expected returns. The problem of finding the optimal hedge ratio now becomes

$$\max_{H_t} \left[\mathsf{E}(r_{t+1}|I_t) - \lambda \mathsf{Var}(r_{t+1}|I_t)\right]. \tag{14}$$

The optimal hedge ratio can easily be calculated as

$$\hat{H}_t(\lambda) = 1 - \frac{\mathsf{E}(s_{t+1}|I_t) - (r_t^h - r_t^f)}{2\lambda \mathsf{Var}(s_{t+1}|I_t)}, \tag{15}$$

with $\mathsf{E}(s_{t+1}|I_t)$ and $\mathsf{Var}(s_{t+1}|I_t)$ the predicted mean and variance of the exchange rate returns.

The mean-variance hedge ratios are very similar to hedge ratios computed with a standard power utility function, like the Constant Relative Risk Aversion. In order to see this, consider the power utility function defined over the investor's wealth W_t:

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

Wealth W_t is recursively defined as $W_{t+1} = \exp(r_{t+1})W_t$, with r_{t+1} the currency return on the exposures over the period $[t, t+1]$, defined in Equation (13). We suppose that the investor maximizes expected utility, conditional on information from the previous period. The investor needs to solve

$$\max_{H_t} \mathsf{E} U(W_{t+1}|W_t) = \max_{H_t} \mathsf{E}\left[\frac{\exp(\gamma r_{t+1}) W_t^\gamma}{\gamma}\right],$$

which, assuming normality of predicted returns, leads to the following optimal hedge ratios (see Appendix A)

$$\hat{H}_t(\gamma) = 1 - \frac{(r_t^h - r_t^f) - \mathsf{E}(s_{t+1}|I_t)}{2\gamma \mathsf{Var}(s_{t+1}|I_t)}.$$

If $\lambda = -\gamma/2$, the hedge ratios for the mean-variance and power utility functions are identical. Remember that the more negative γ the more risk-averse the investor is. This corresponds with a higher positive value of λ, signifying that the investor is more concerned about variance.

3.2 Incorporating the Value-at-Risk in the hedge decision

The mean-variance objective function is a very simplified, although convenient, way of looking at the risk and return trade-off that investors make. A popular measure, advocated by financial regulatory institutions, is the Value-at-Risk (VaR) of a portfolio. 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

$$\int_{\mathrm{VaR}}^{\infty} f(r_{t+1|t})\, dr_{t+1|t} = 1 - \alpha, \tag{16}$$

[5] We will also use the shorthand notation $r_{t+1|t}$ for conditional variables.

with $1-\alpha$ the confidence probability, which typically ranges from 90% to 99%.[6]

What we calculate using our models is the predictive density of the exchange rate returns $s_{t+1|t}$, conditional on a vector of parameters θ. In case of a normal predictive density, with expectation $\mathsf{E}(s_{t+1}|I_t)$ and variance $\mathsf{Var}(s_{t+1}|I_t)$, the VaR can be calculated analytically. Subsequently, the return r_{t+1} conditional on information up to time t is also normal. with mean $\mu_r(H) = \mathsf{E}(r_{t+1}|I_t) = (1-H_t)\mathsf{E}(s_{t+1}|I_t) + H_t(r_t^h - r_t^f)$ and $\sigma_r^2(H) = \mathsf{Var}(r_{t+1}|I_t) = (1-H_t)^2\mathsf{Var}(s_{t+1}|I_t)$. The value at risk can then be computed as

$$\mathrm{VaR}(H|\theta)_t = \sigma_r(H)q_{\mathcal{N}}(\alpha) + \mu_r(H) = (\sigma_s q_{\mathcal{N}}(\alpha) + \mu_s) + H((r_t^h - r_t^f) - (\sigma_s q_{\mathcal{N}}(\alpha) + \mu_s)), \tag{17}$$

where $q_{\mathcal{N}}(\alpha)$ is the α-th quantile computed of a standard normal distribution. If we have a sample of θ's from the posterior distribution, we can estimate the posterior density of the Value-at-Risk. Furthermore, it is sufficient to calculate the Value-at-Risk for just a single value of the hedge ratio, as the VaR is a linear function of H.

These calculations are based on the normality assumption that the mean and variance of the future return describe the predictive density completely. This assumption holds for models **A**-**D**(for the GLL-GARCH model it only holds one period ahead), but it holds only approximately for the GLL-SV and GLL-Student t models. In the subsequent sections the normal approximation to these models is used in making the final hedging decisions, which is a good second order approximation.

4 Hedging against the D-Mark/US dollar currency risk

4.1 Stylized facts

Our data set consists of daily observations on the D-mark/US dollar exchange rate for the period January 1, 1982 until January 26, 1999 which gives a total of 4454 observations. For this same period we have the 1-month Eurcurrency interest rates for the German DMark and the US Dollar.

In the upper panel of Figure 3 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 can observe the changing trend which implies a changing mean in the exchange rate returns. The volatile behaviour of the series makes it difficult to recognize the changing mean in the graph that shows the exchange rate returns. The left panel at the bottom of this figure depicts the average return over a three month period. Here, the changing mean return is immediately apparent. The last panel graphs the evolution of the three period mean absolute return. The mean absolute return is an indicator for the volatility in the series. Periods of high volatility alternate with periods of more stable behaviour.

A subperiod analysis confirms these findings. Figure 4 shows the graphs corresponding to the ones in the previous figure, for the years 1992 and 1993. A one month average was taken for the two graphs in the bottom panels. The depreciation of the dollar over the period April 1992-September 1992 is followed by several months of appreciation. The changing mean return is modelled through the autoregressive process on μ_t, in equation (2). The absolute values of the returns show clearly the clustering of volatility. The GARCH or Stochastic

[6] 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).

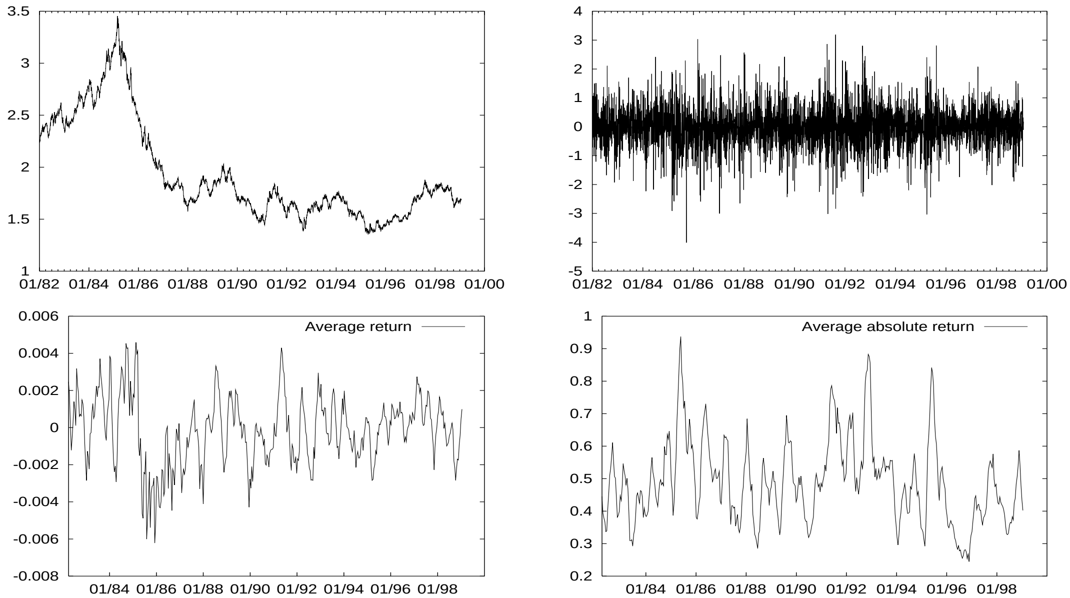

Figure 3: Exchange rate between German DMark and US Dollar, January 1, 1982 until January 26, 1999. Panels contain data in levels (top left), in returns (top right), the three month average return (bottom left) and the three month average absolute return (bottom right)

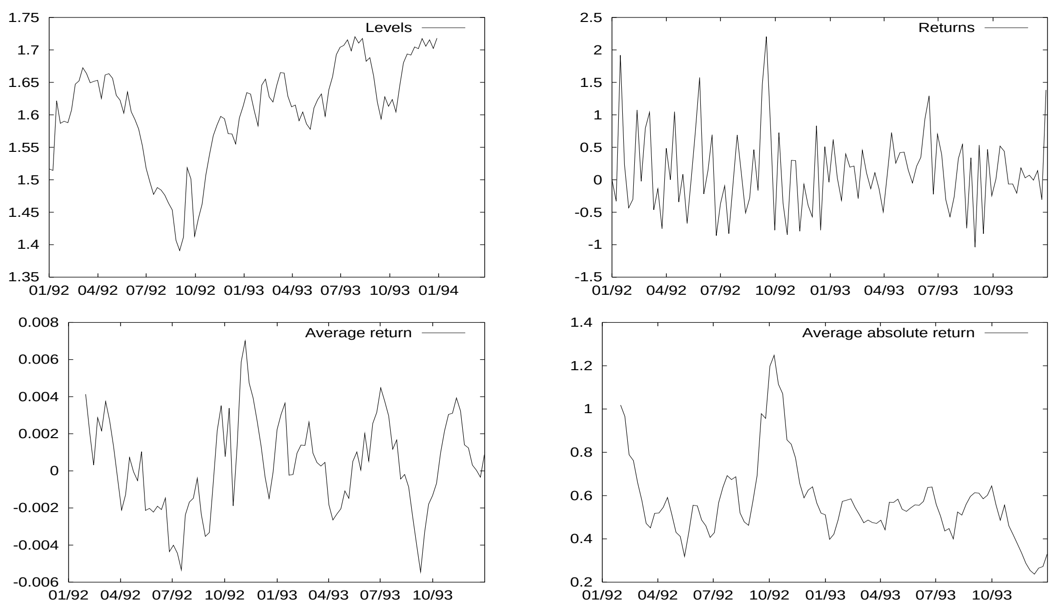

Figure 4: Exchange rate between German DMark and US Dollar, January 1, 1992 until December 31, 1993. Panels contain data in levels (top left), in returns (top right), the one month average return (bottom left) and the one month average absolute return (bottom right)

Volatility process on the disturbances is aimed at modelling this stylized fact. Note that the data in levels is nonstationary. After taking first differences, a stationary series results.

4.2 Posterior results

In this section posterior results are presented for the different models.

The final round of the Metropolis-Hastings sampler applied on the models **B**-**D** continued until a sample of 10000 accepted drawings had been accumulated, after an initial period in which 1000 accepted drawings were disregarded to get rid of a possible initial effect. Table 3 summarizes the acceptance rates during the four rounds of the adaptive MH sampler and also reports the size of the final sample.[7] Acceptance rates are all at least 0.65 in the final round, which is rather good. Models **E** and **F**, incorporating the Stochastic Volatility and the Student t disturbances, were sampled using a Gibbs sampling scheme. Convergence of the sampler is evaluated in this case using the autocorrelation function of the sampled parameter vectors. From a series of preliminary runs it was found that a final Gibbs run of 250.000 iterations was sufficient for the analysis. Of these iterations, every 25th vector of sampled parameters was used. Retaining all sampled parameter vectors was not useful as there was strong serial correlation in the original posterior sample.

		Acceptance rate				
	Model	1	2	3	4	Sample size
Burn in		0	0	0	1000	
Acc. dr.		500	1000	5000	10000	
B	LL	0.91	0.91	0.93	0.93	10753
C	GLL	0.61	0.70	0.70	0.72	13889
D	GLL-GARCH	0.55	0.59	0.57	0.65	15385

Table 3: Acceptance rates over the rounds, for the models White Noise, Local Level, Generalized Local Level and GLL-GARCH

Figure 5 plots a histogram of the sampled values of the parameters μ (left panel) and σ_ϵ (right) in the White Noise model, against the priors that were used.[8] In this pure White Noise model, the posterior mean of the parameter controlling the average daily return on the DM/US exchange rate is approximately -0.005. This estimate approximates the empirical mean closely. However, the highest posterior density region is rather wide, and includes zero. From an informal Bayesian statistical viewpoint, including a mean appreciation in the model might not be significant. A conclusion on its decision-theoretical relevance is postponed until Section 4.4. The posterior density of the standard deviation is strongly concentrated compared to the prior. This indicates the high information content of the data on the variance.

The second model, the Local Level model, allows for a non-zero standard deviation σ_η in the transition equation of the mean process. In Figure 6 it is seen that the posterior of σ_η shifts away from the prior distribution. We note that little correlation between sampled parameter vectors was found.

[7]Note that the posterior for μ and σ^2 for case **A**can be determined analytically as a normal-inverted gamma density. Posterior results for this model were also computed with a Metroplis-Hastings sampler, however.

[8]For the purpose of the illustration, the prior is scaled up such that it integrates to one over the range of support of the posterior.

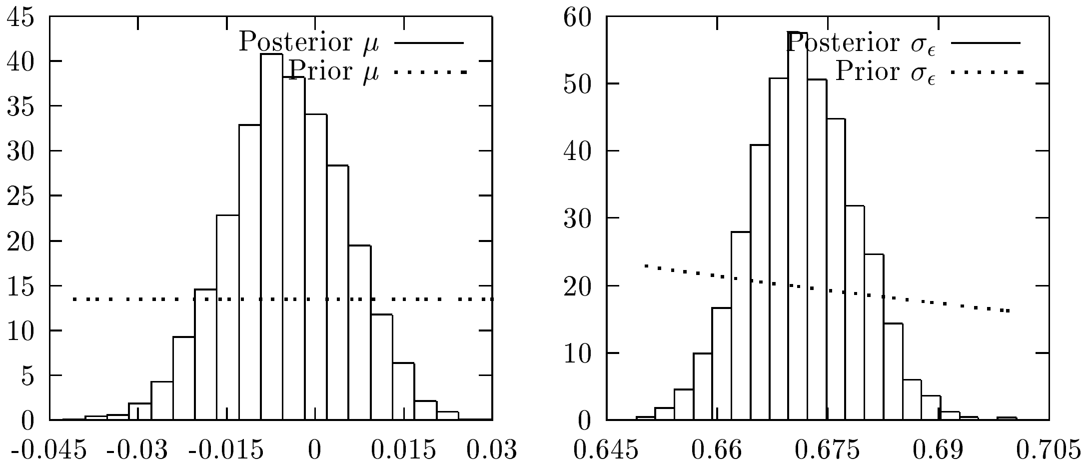

Figure 5: Marginal posterior density, model **A**, White Noise, for μ (left) and σ_ϵ (right)

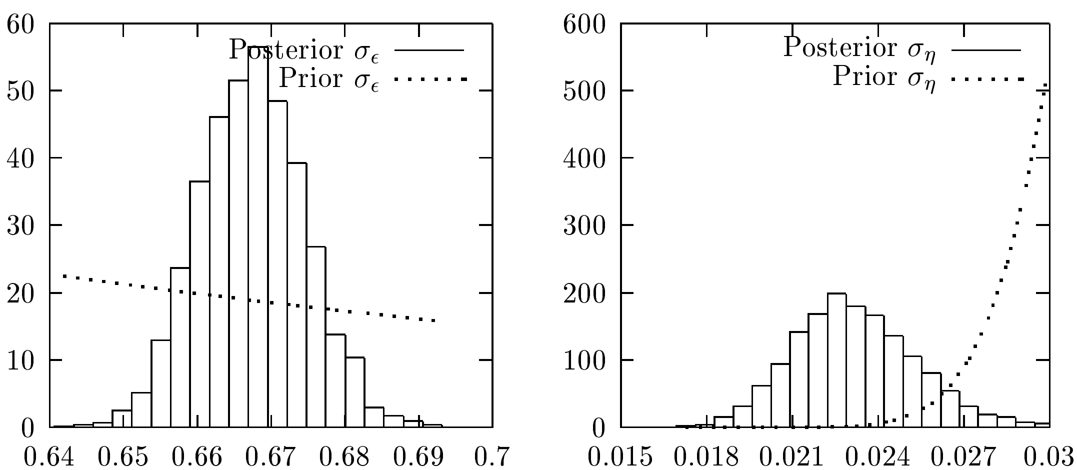

Figure 6: Marginal posterior density, model **B**, Local Level, for σ_ϵ (left) and σ_η (right)

The Generalized Local Level model introduces a parameter ρ in the μ_t process.The posterior density of σ_ϵ does not change strongly in comparison to the previous results on the White Noise and Local Level model (see Figure 7, second panel). The histograms of both the autocorrelation parameter ρ in the mean series and of its standard deviation, σ_η, mimic closely the shape of the prior density. Apparently there is little extra information in the likelihood. The posterior of σ_η is, however, much more plausible in the GLL model than in the LL model. This result is due to the presence of the autoregressive parameter ρ in the GLL model. Again, the autocorrelation in the posterior sample is low.

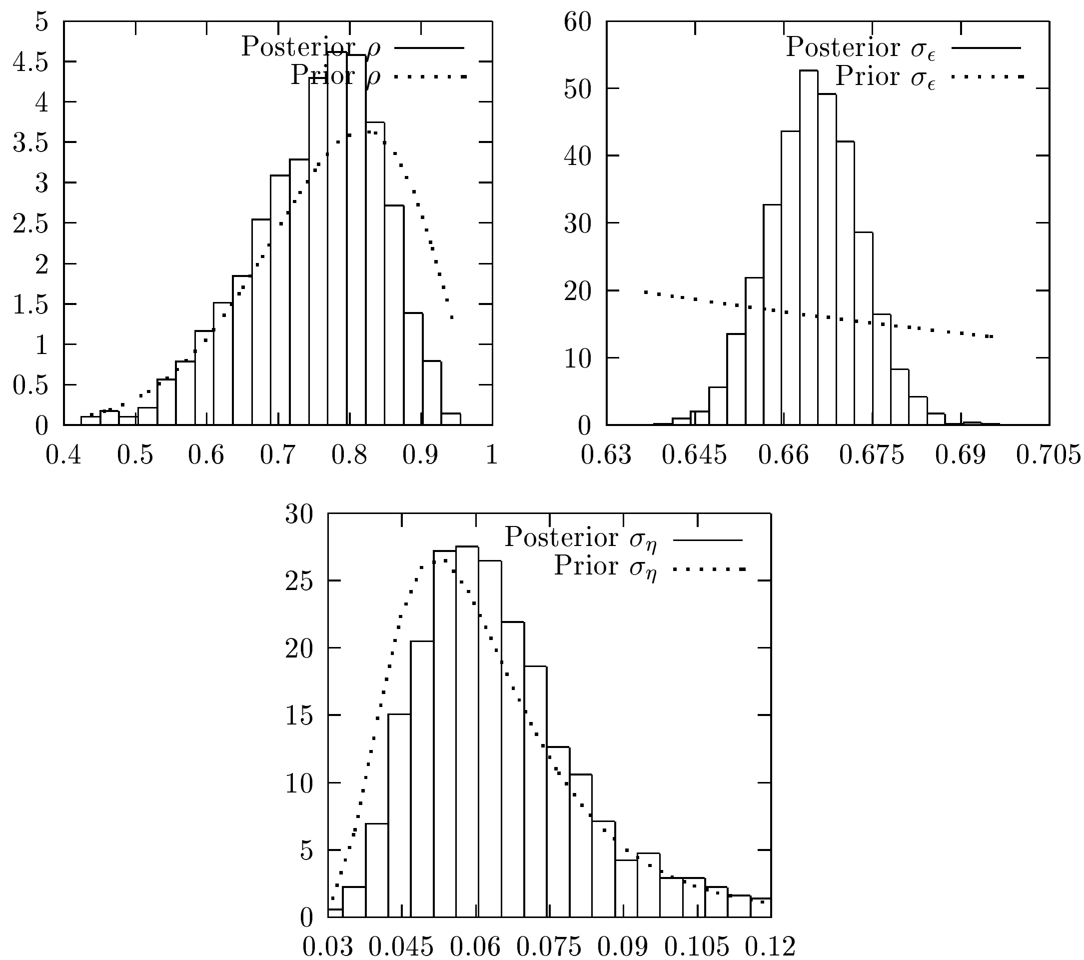

Figure 7: Marginal posterior density, model **C**, Generalized Local Level, for ρ (top left), σ_ϵ (top right) and σ_η (bottom)

In Figure 8 the assumption of homoskedastic disturbances on the observation equation is relaxed, allowing for a GARCH process to model the changing volatility (model **D**). The first two panels, on ρ and σ_η of the transition equation (2) are similar to the corresponding plots in Figure 7, though slightly stronger correlation is found. Note that the plots on the GARCH parameters δ, ω and α do not contain the curve of the prior density: The non-informative joint prior on δ and α does not allow for a straightforward calculation of the marginal prior.

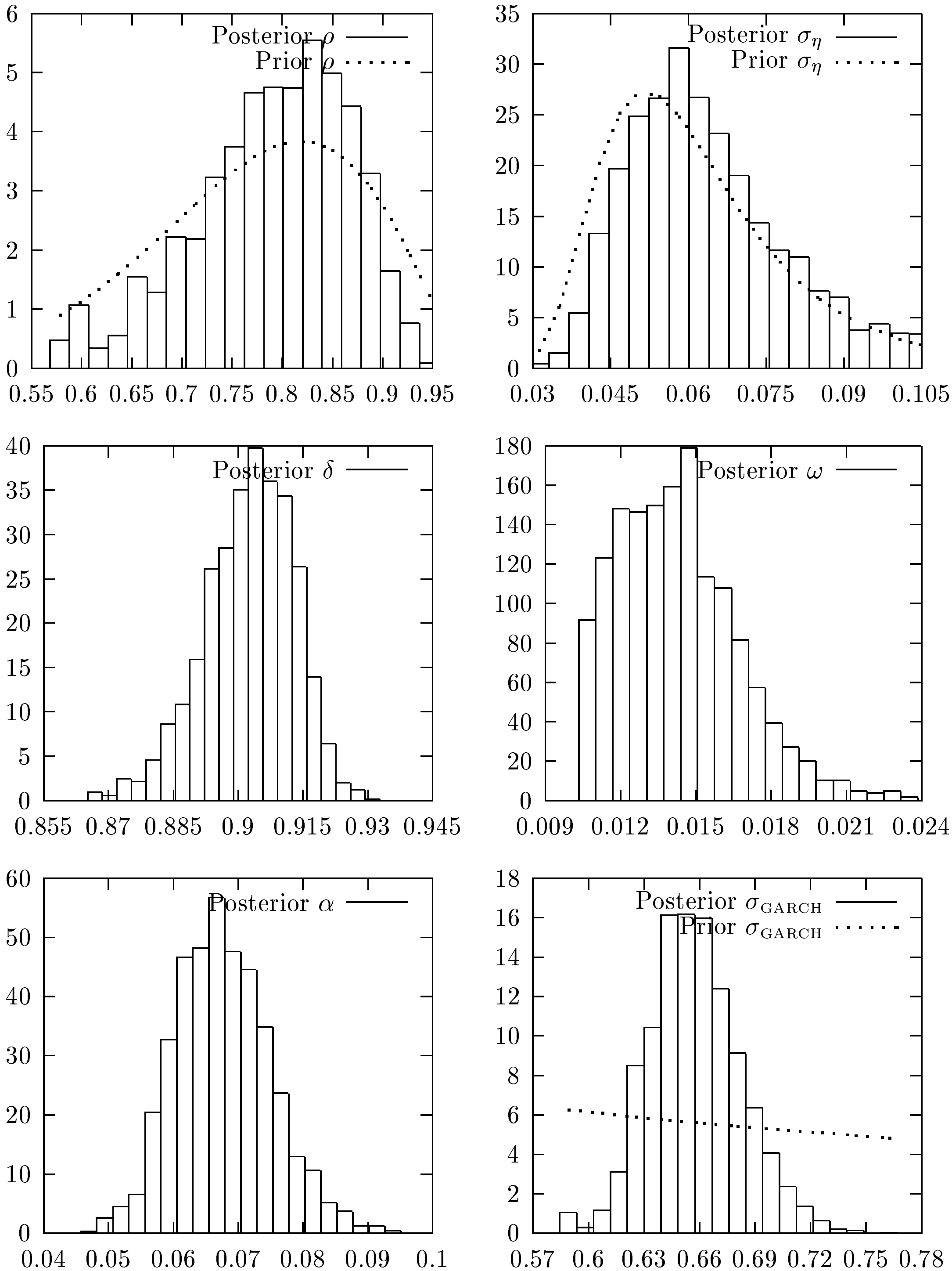

Figure 8: Marginal posterior density, model **D**, Generalized Local Level-GARCH, for ρ (top left), σ_η (top right), δ (middle left), ω (middle right), α (bottom) and the unconditional standard deviation σ_{GARCH} (bottom right)

Instead, in the last panel of the figure, the sample information in δ, ω and α is combined into the implied unconditional standard deviation $\sigma_{\text{GARCH}}(\delta, \omega, \alpha) = \sqrt{\omega/(1-\delta-\alpha)}$. The histogram of this standard deviation is compared with its prior in the bottom right panel of Figure 8. Some extra correlation is found in the posterior sample of the parameters of this model, but after about 20 periods the autocorrelation is negligible.

As was described in Section 2.2, sampling the Generalized Local Level-Stochastic Volatility model requires the use of Gibbs sampling. The Gibbs sampler was run for 251.000 iterations. The first 1000 were skipped to allow for a (short) burn-in period. Afterwards, only every twentyfifth sampled vector of parameters was recorded (see Geweke 1999). The resulting sample of 10000 parameters was used in the histograms in Figure 9.

In the two upper panels, it is seen that the parameter ρ as well as the standard deviation of the disturbance σ_η are smaller. The mean series reverts more quickly back to the no-information level of $\mu_t = 0$, and it does not divert from this level too much as the unconditional variance of μ_t, $\sigma_\mu^2 = \sigma_\eta^2/(1-\rho^2)$ is smaller than in the previous models. The posterior mode of μ_h, the mean level of h_t which governs the unconditional expected volatility level, implies an expected standard deviation of the observation equation of 0.59. For comparison with the plot on the GARCH model the unconditional standard deviation of the Stochastic Volatility model this standard deviation $\sigma_{\text{SV}}(\mu_H) \approx \exp(\mu_H/2)$[9] is added in the bottom right panel of the figure. The correlation between successive values of the volatility, governed by parameter ϕ in equation (8), is found to be even larger than its a priori mean of 0.8.

Finally we consider the Generalized Local Level-Student t model (model **F**). Again, the Gibbs sampling scheme was used to obtain the posterior in Figure 10. The posterior on ρ, indicates a rather persistent mean. The standard deviation parameter σ_ϵ is estimated as below the values found with the GLL model. We note, however, that this parameter σ_ϵ only governs part of the variance in the observation equation; the variance depends also on the value for the degrees of freedom parameter ν, in the fourth panel. A Student t density with 10 degrees of freedom fits the data best. It is of interest that, even though the prior information on the ν parameter was weak, such a clear posterior is found.[10] Only in model **E**, the GLL-SV model, the standard deviation of η was lower. The fifth panel depicts the standard deviation of the disturbance process in the observation equation, calculated as $\sigma_{\text{Stud}} = \sigma_\epsilon \sqrt{\nu/(\nu-2)}$. This model indicates a higher variance in the observation equation, and is more certain about this variance than in the GLL-GARCH and GLL-SV cases, as can be seen from the smaller range [0.69, 0.74] instead of [0.6, 0.71] or [.55, .64].

In the Gibbs sampling algorithms the method of data augmentation is used, to sample the states of μ_t and of h_t, and of an auxiliary indicator variable s_t (see Casella & George (1992) on the general idea of data augmentation, and Kim et al. (1998) for details on the need for sampling h_t and s_t; a similar reasoning holds for the sampling of the state μ_t, see e.g. Koop & van Dijk (1999)). Figure 11 plots the autocorrelation of the sampled parameter values.[11] These graphs indicate that the present sample is big enough for convergence to have taken place.

[9]Note that this calculation of the standard deviation is not entirely correct, as the expectation operation is applied on a non-linear transformation.

[10]In preliminary runs, a wider range of values for ν was allowed. This did not change the outcome. For reasons of computational efficiency, in the final run the bounds on ν were more restrictive.

[11]Note that only every 25th sampled vector of parameters was saved. Autocorrelation between parameters 40 elements apart in the sample that was saved indicates autocorrelation between parameters $40 \times 25 = 1000$ elements apart in the original sample.

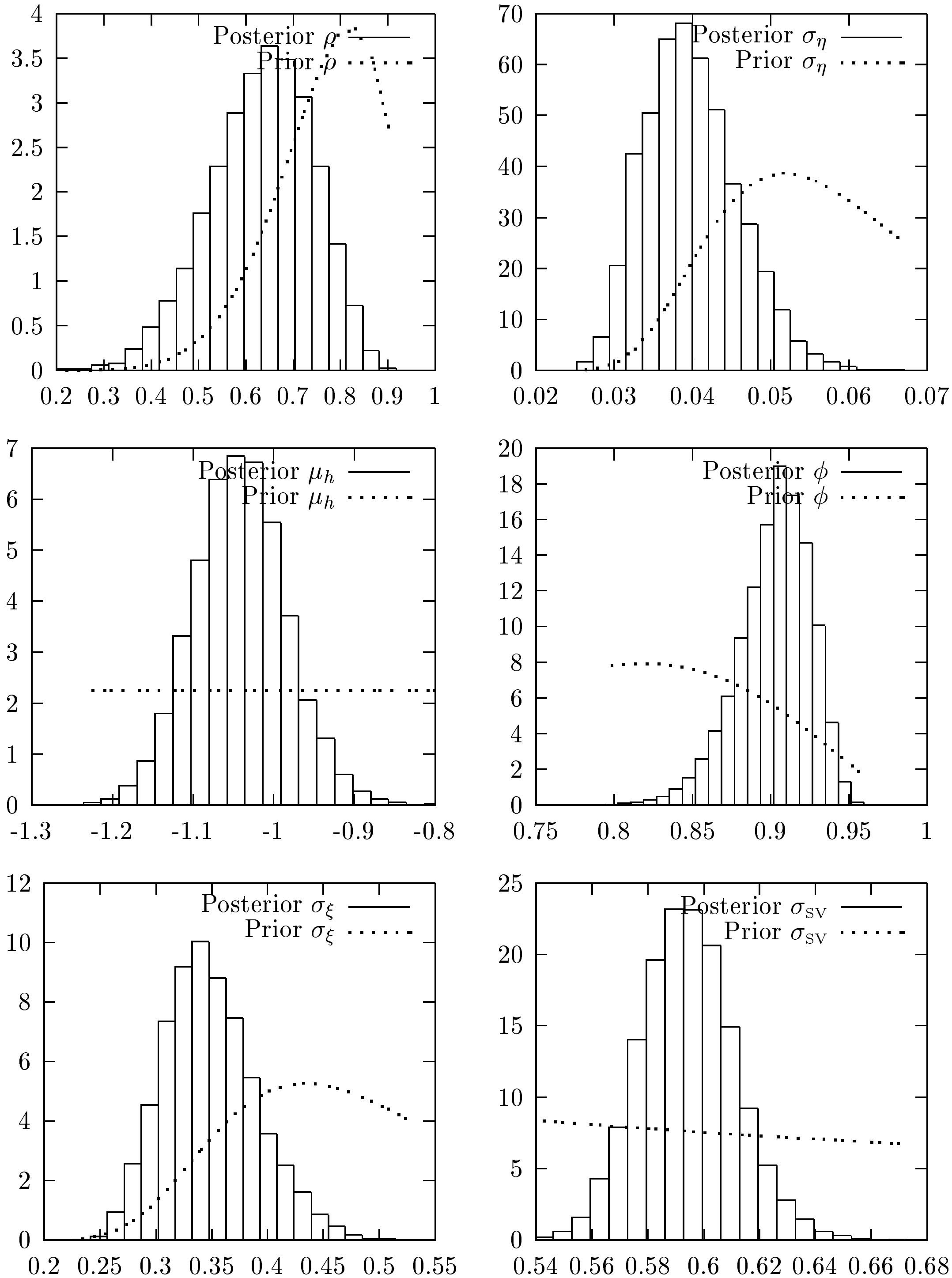

Figure 9: Marginal posterior density, model **E**, Generalized Local Level-Stochastic Volatility, for ρ (top left), σ_η (top right), μ_h (middle left), ϕ (middle right), σ_ξ (bottom left) and the unconditional standard deviation σ_{SV} (bottom right)

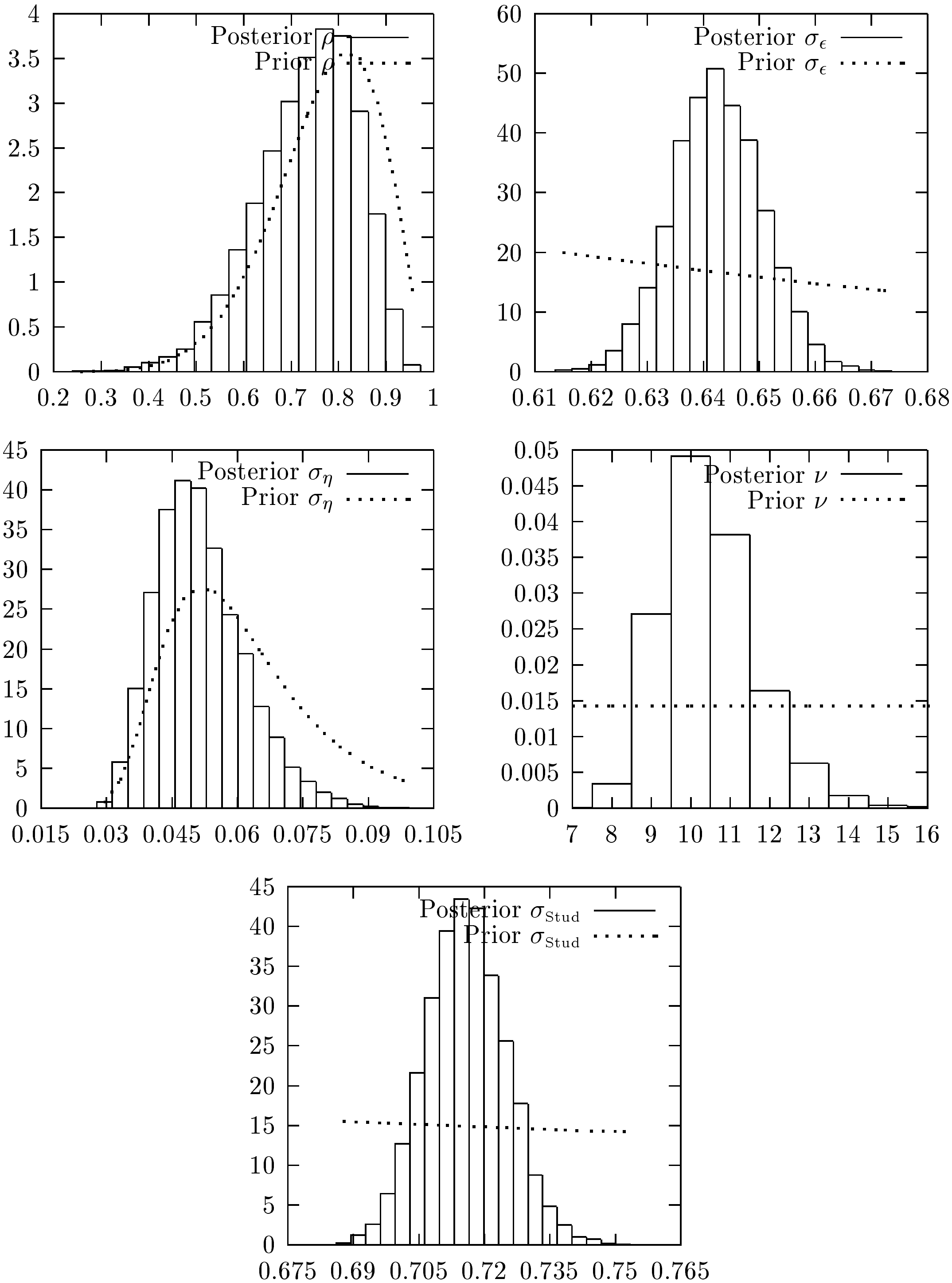

Figure 10: Marginal posterior density, model **F**, Generalized Local Level-Student t, for ρ (top left), σ_ϵ (top right), σ_η (middle left), ν (middle right), and the unconditional standard deviation σ_{St} (bottom)

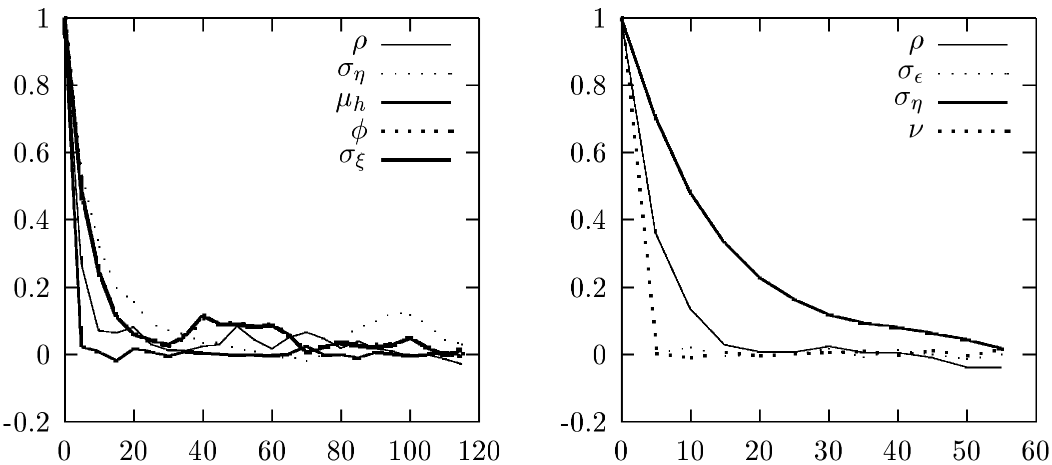

Figure 11: Posterior sample correlation, models **E**, GLL-SV, (left) and **F**, GLL-Student t

4.3 Predictive density

The investor in our setup is confronted on a daily basis with the problem of making a prediction for next day's return on the market, on which to base his investment decision. In this section we compare the different predictions that result for each of the four models that were proposed in Section 2. The one-step ahead prediction densities are calculated for each of the vectors of parameters in the posterior sample. We repeated this exercise for each of the days in the last month of the data set (observations 1/1/1999-26/1/1999).

When the returns on the exchange rate are modelled as white noise (model **A**), The predictive mean and variance are constant. The three graphs in the upper row in Figures 12-13 show the median[12] of (1) the predicted means (left panel), (2) variance (middle panel) and (3) the mean with confidence bounds (right panel). On the x-axis the days in January 1999, for which the prediction has been calculated, are given. These graphs are important in order to study the consequences for the hedging decisions.

Model **B**, the pure Local Level model, has by construction a prediction of the first moment which follows a random walk and a constant predictive variance, see the second row of Figure 12.

In model **C**, the predicted variance is still constant. The prediction of s_t has however less memory than model **B**, see the third row of graphs in Figure 12.

The GLL-GARCH model, with graphs in the first row of Figure 13, allows for changes in the predicted variance. As common to GARCH explorations, the variance can undergo sharp upward swings as a result of unanticipated shocks. Afterwards, the higher volatility slowly declines.

Model **E** incorporates a stochastic volatility process instead of the GARCH effect in the second moment. This results in a series with smoother variance changes, as the large shock can be "anticipated" in an increasing variance some days ahead. Note that the predicted variance is higher than for the GLL-GARCH model, with less predictability (the predicted s_t's are closer to zero) in the first moment. This increased uncertainty makes the GLL-SV

[12]Each of the parameter vectors in the posterior sample leads to a value for the predicted mean and variance. Of these, the medians have been used in constructing the plots.

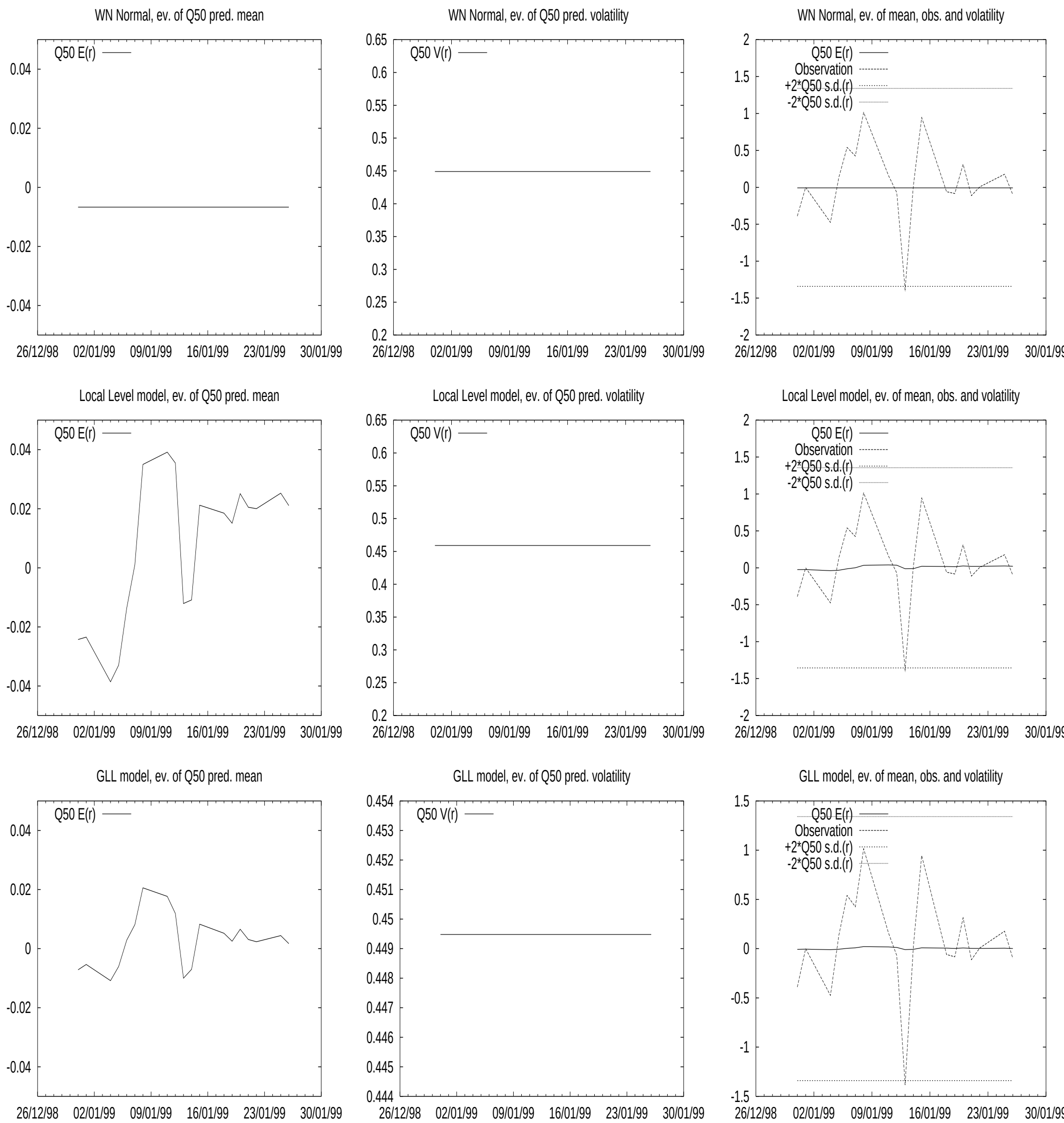

Figure 12: Medians of predicted means (left column), variances (middle) and mean with confidence bounds of two standard deviations for the White Noise, Local Level model, and Generalized Local Level model (top to bottom, one model per row)

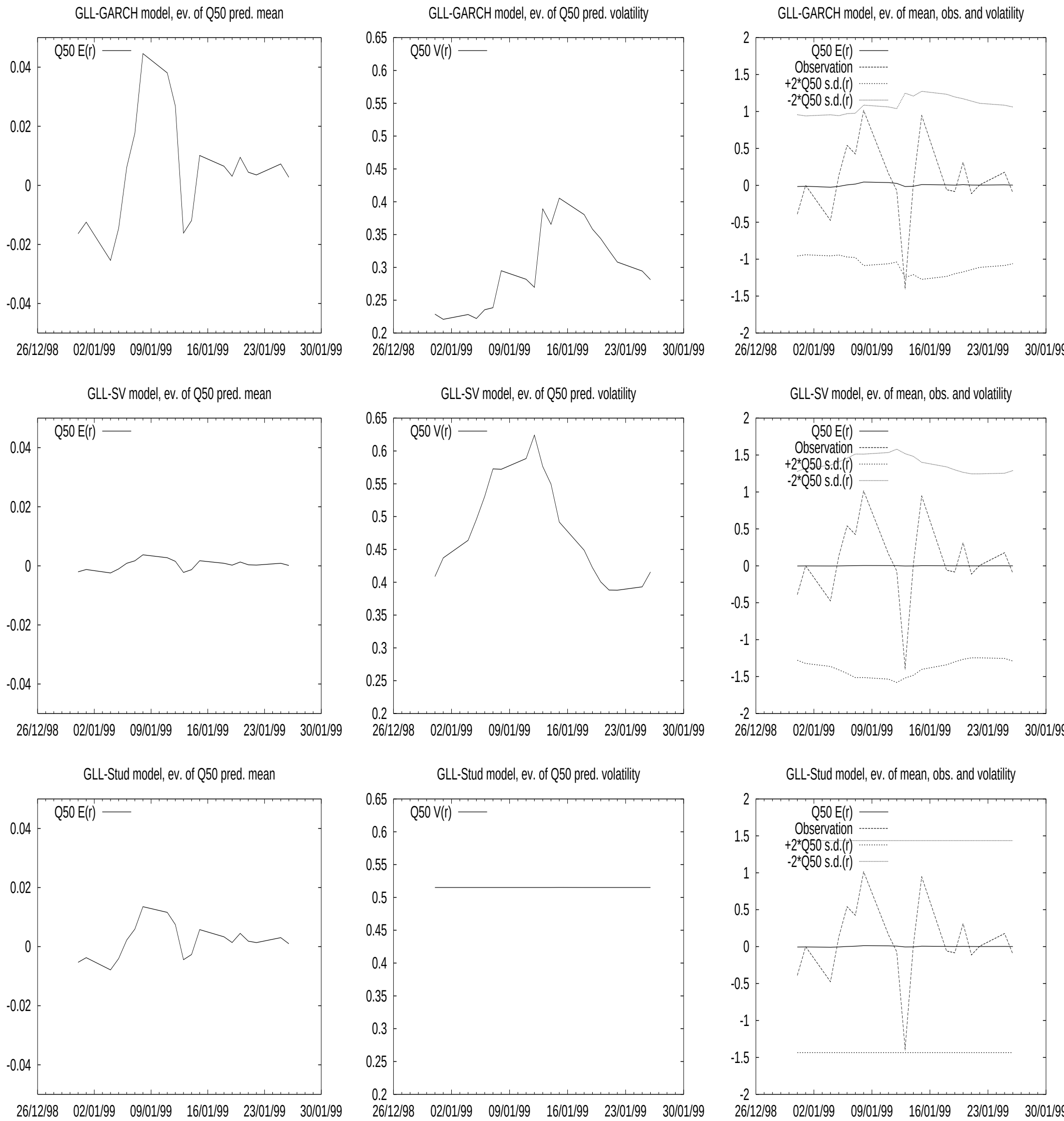

Figure 13: Medians of predicted means (left column), variances (middle) and mean with confidence bounds of two standard deviations for the GLL-GARCH model, GLL-SV and GLL-Stud model (top to bottom, one model per row)

model more in agreement with the White Noise model than with the GLL-GARCH model.

Incorporating Student t disturbances in the observation equation, as in model **F**, implies that no correlation is modelled in the second moment of the data generating process. In a pure linear state space model with normal disturbances (like models **B** and **C**) the variances of the predicted state and observation converge to a steady state (see second and third row of Figure 12).[13] Notice that this model produces the highest variance of all models, except for a period of 7 days where the stochastic volatility of model **E** is estimated higher. Due to the higher estimated variance, a shock in an observation exerts less influence on the estimated underlying mean. The resulting predictive return is more stable than for the GLL or GLL-GARCH case, but it is more informative than for the GLL-SV model.

4.4 Optimal hedging decisions

In this section we investigate the influence of the model specification on the optimal hedge decisions.

The risk tolerance λ of the investor is a parameter in the decision process. Low values of the tolerance indicate that the investor does not mind risk, higher values signify that the investor demands stronger positive expected return before being willing to take the risk. In the calculations, the value of the risk tolerance was varied between 0.001 and 0.1.

For all available trading days in 1999 in the sample the one step ahead prediction of exchange rate return and variance have been calculated, based on the posterior sample for each of the models **A**-**F**. The interest rate differential $r^h_{t,\tau} - r^f_{t,\tau}$ ranges from -0.0052 to -0.0048 over the period 1/1/1999-26/1/1999.

Two sets of graphs are displayed, in Figures 14 and 15. In Figure 14 the hedging decision is made for the last day in the data set, January 26, 1999. The panels show the optimal hedge ratio for different values of the risk tolerance, for each of the models. The horizontal lines indicate the interquartile range, whereas the curve connects the medians of the optimal hedge ratio as calculated for all sampled vectors of parameters in the models. In Figure 15 the medians of the optimal hedge ratios are drawn for all of the trading days between January 1 and January 26, 1999.

In the upper left panel of the first figure, results for the White Noise model are graphed. Since the predictive mean are constant, the results in this panel are driven by the interest rate differential at January 26. In case the sampled parameter μ is smaller than the differential, running risk is of no use and the investment should be fully hedged. As the error bars reach out to a hedge ratio of one, at least 25% of the sampled μ's is smaller than the difference between home and foreign interest rates (see also the upper left panel of Figure 5, the histogram of the posterior sample of μ's). When turning to the first panel in Figure 15, we find that there is hardly a difference between hedging decisions made on each of the trading days in the last month of the data set. As the hedging decision is based solely on the value of the parameters, with no influence for the history of the series, the only effect of changing the date of the decision lies in the changed interest rate differential.

The second panel in both figures covers results for the Local Level model. As this model allows for a varying mean return, it introduces more variety in the hedging decisions. In Section 4.2 it was recognized that the integrated mean process did not fit the data well. This translates to even wider interquartile ranges on the hedging decisions. From the plots on the

[13] Also see Harvey (1989) for more details.

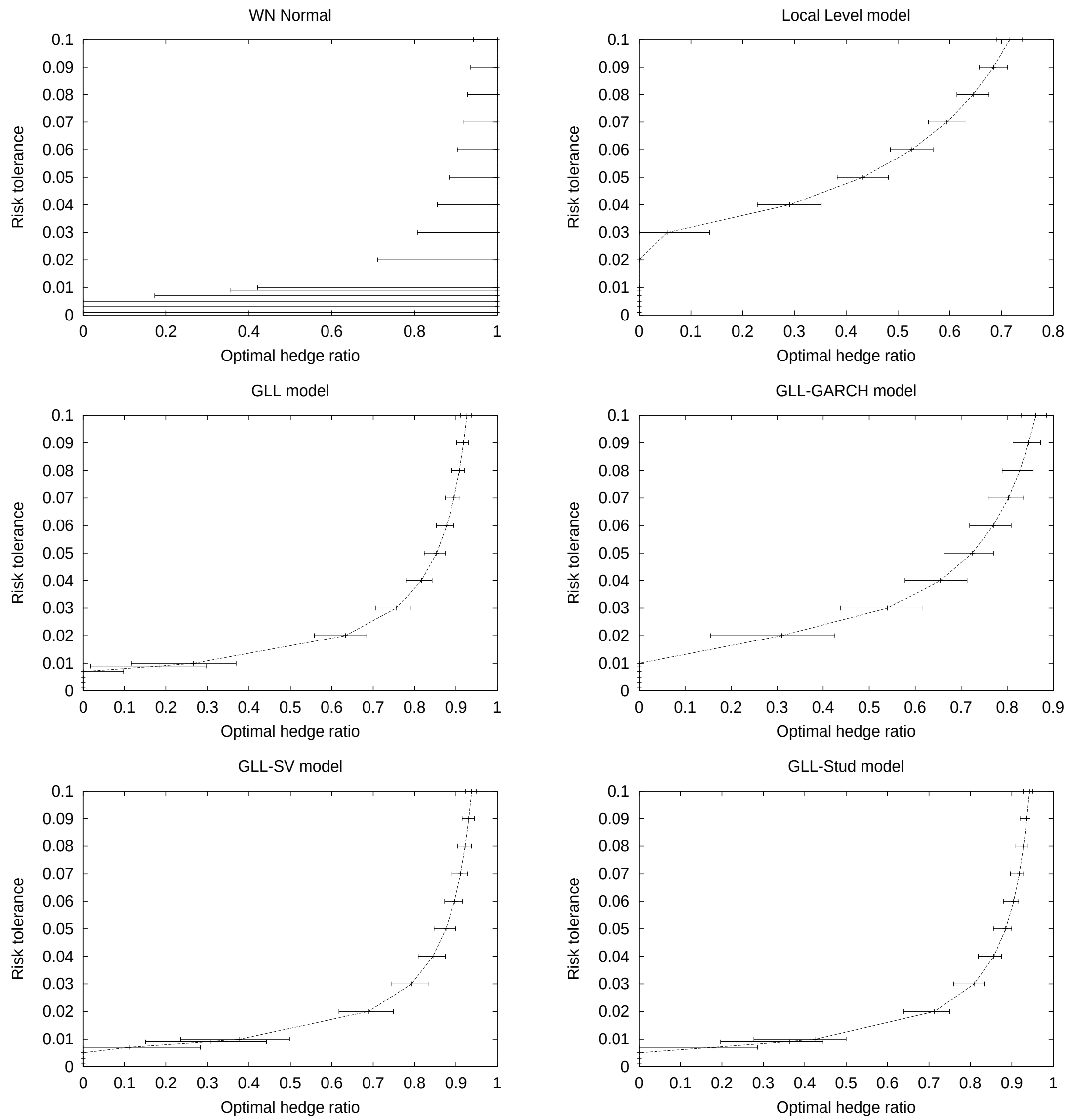

Figure 14: Optimal hedging decision made on January 26, 1999, for the White Noise, Local Level, Generalized Local Level, GLL-GARCH, GLL-SV and GLL-Stud model

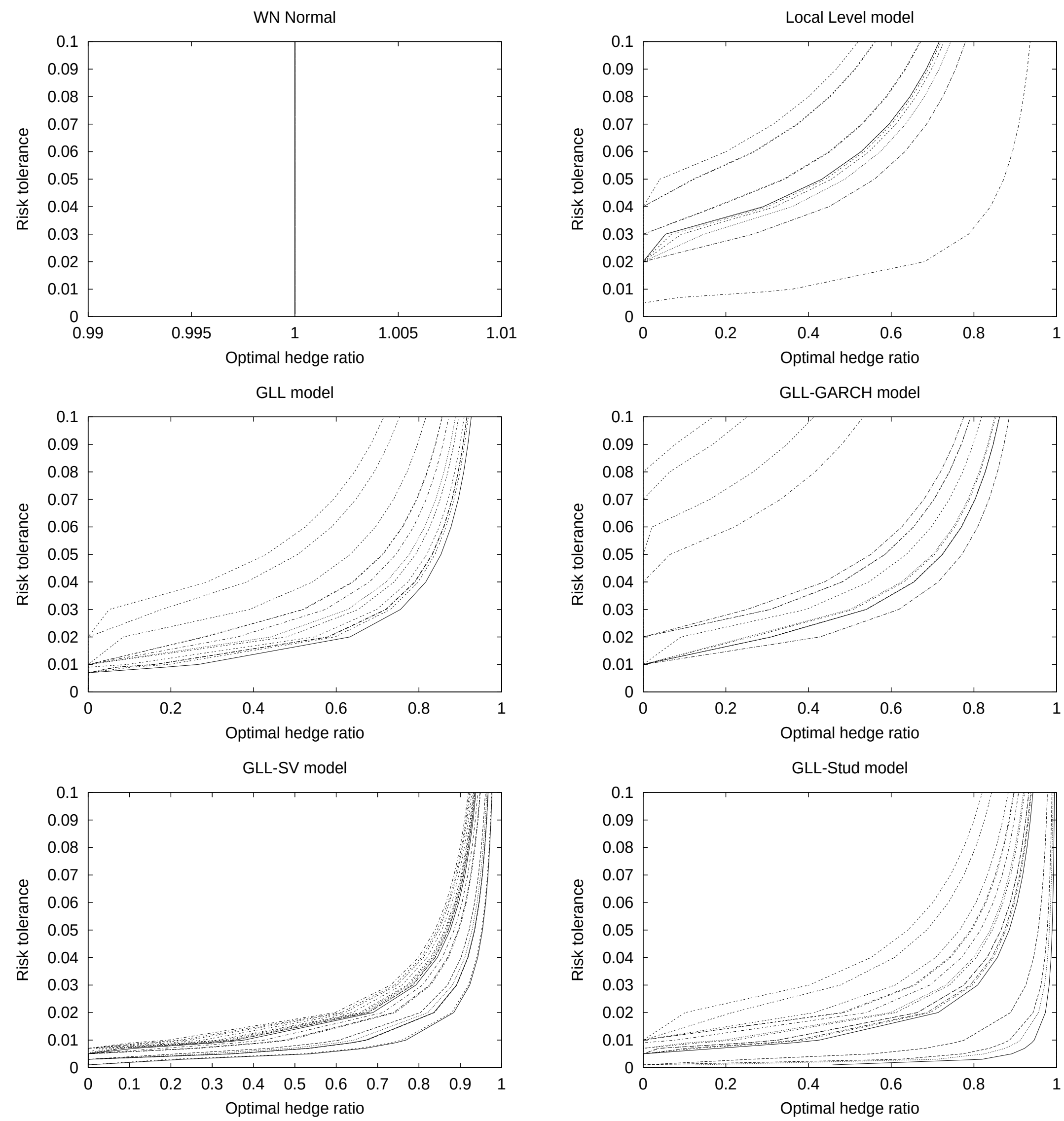

Figure 15: Optimal hedging decision made between January 1, 1999 and January 26, 1999, for the White Noise, Local Level, Generalized Local Level, GLL-GARCH, GLL-SV and GLL-Student t model

predictive expected return we saw that $\mathsf{E}(s_{t+1})$ changes over the days. This results in slightly different hedging decisions over the investigated period (see the second panel in Figure 15).

The Generalized Local Level model had a smaller range of predicted means (compare the left panels of the second and third rows in Figure 12) than the Local Level model. Therefore, also the hedge ratios change less over time (Figure 15, third panel). On average less hedging is needed according to this model, when compared with the white noise model. The GLL hedge ratios are higher than the hedge ratios for the Local Level model.

Using a GARCH component for the disturbances in the observation equation (1) led to more flexibility in predicted mean returns (again, left panels of Figures 12 and 13). The predicted volatility is varying, but over the investigated period lower than with the previous models. This leads to even more pronounced hedging decisions: As the investor claims to be relatively sure of tomorrows expected return, he does not hedge the currency risk.

Sampling the parameters for the GLL-SV model, and calculating the predicted returns and variances led to a predicted mean series with little movement over time. Also the estimated uncertainty was generally higher. On the last day of the sample, the predicted return and variance approximately equal the same quantities for the GLL model **C**. Therefore, also the hedging decision for that day is very similar to the decision made for that model, as can be concluded from panels 3 and 5 in Figure 14. In Figure 15 the difference between the decisions for the GLL-SV and pure GLL model can be recognized: As the GLL-SV model tends to predict a smaller return and a larger variance, or in short a more risky investment, it advises to hedge more than is the case for the GLL model. In the present setting the estimated predictability of the returns is not strong enough to make it worthwhile for the investor to run the risk; the optimal hedging curves in Figure 15 quickly tend to one already for small values of the risk tolerance.

Model **F** assumes a heavy tailed density for the disturbances in the observation equation. The hedging decision on the last day of the sample mimicks the results for the GLL-SV model for the different choices of the risk tolerance parameter. However, there are days that the predicted positive return on the exchange rate is such that the fully hedged position is left in order to take advantage.

4.5 Evaluating the Value-at-Risk

In Section 3.2 we presented the Value-at-Risk (VaR) as a linear function of the hedge ratio. This means that the optimal hedge ratio will be either zero or one. From a practical risk management viewpoint it is still very valuable to report VaR measures, see for example Jorion (1997). In this section we present results for one fixed value ($H = 0.5$) of the hedge ratio. Figure 16 plots the posterior distribution of the VaR with a 95% confidence level for models **A**-**F** on the last day of the prediction period, January 26, 1999.[14] Notice that the scale on the x-axis is equal for models **A**, **B**, **C** and **F**; the GLL-GARCH and GLL-SV model results in different locations for the posterior VaR.

The results for the White Noise, Local Level and Generalized Local Level model are similar. Apparently the mean process is not strong enough on itself, without accounting for varying volatility or heavy tails, to change results in the tail of the predictive density. The VaR is a characteristic measuring the thickness of the tail of the predictive density of the returns.

[14]Other days give results very similar to this last day; only the location of the posterior is changed slightly.

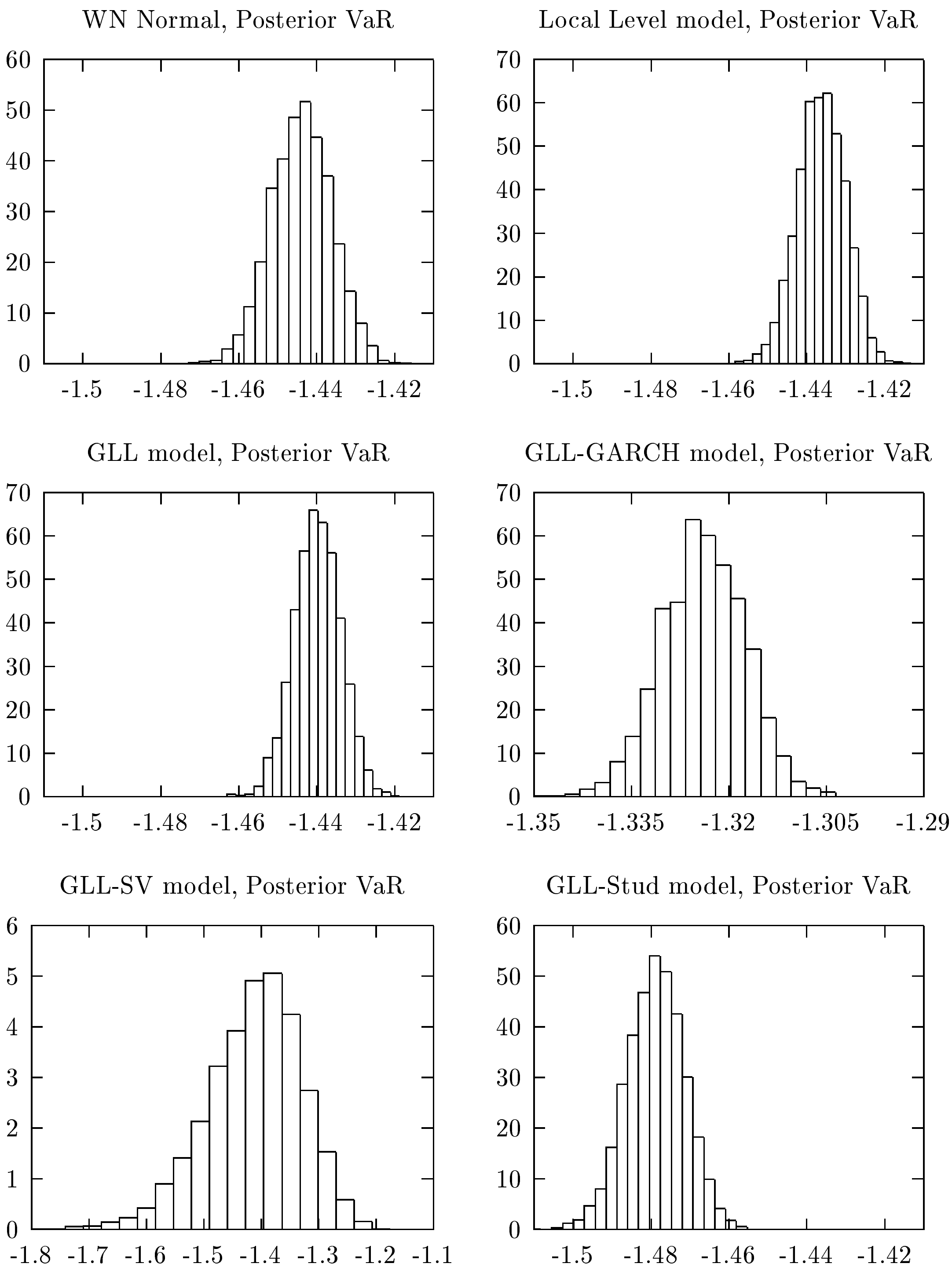

Figure 16: Posterior distribution of the Value-at-Risk for the WN, LL, GLL, GLL-GARCH, GLL-SV and GLL-Student t models (left to right, top to bottom)

The extensions of the GLL model result in different conclusions. Of the GLL-GARCH, GLL-SV and GLL-Stud the GLL-SV model has the largest VaR range in the posterior distribution of the VaR. The VaR range in the GLL-GARCH and GLL-Stud models is approximately equal (notice again the different scaling). The GARCH model generates higher expected returns for the last day in the data set. Losses of more than 1.34% are deemed very unlikely. The difference stems from different predictions for the variance. While the Student t model takes possible heavy shocks into account, resulting in a predicted variance of 0.51, the GARCH model is quite positive with a variance of 0.3.

5 Concluding remarks

During the past twenty years many models have been developed for the description of financial time series. Changes in variance are one of the most outstanding features of financial time series, and, as a consequence, much attention has been put on modelling the variance of these series. Many financial decision problems, however, depend on the full probability density of financial returns. In this paper we focused on hedging foreign exchange rate risk for an international investor and investigated a set of competing models that describe the most prominent features of the DMark /US Dollar exchange rate. Special attention has been given to describe the mean of the exchange rate returns. It was found that including the varying mean process resulted in predictions of the mean with an order of magnitude which corresponds to the order of magnitude of the interest rate differential. Even though the numerical value of the mean process is not large, it cannot be disregarded in the hedging decision, where the relative size of the mean process versus the interest rate differential is at play.

The predicted volatility is, of course, a key element in the hedging decision. Especially in the combination of the Generalized Local Level-GARCH model a relatively small variance was forecasted over the period 1/1/99-26/1/99. The implication of this low predicted volatility is a hedging position which changes strongly throughout January 1999. Further research is needed to check whether this result is robust or very specific for our data and/or model. The GLL-Student t model was more conservative in its estimation of the future risk. This resulted in a more risk-averse position being taken in the hedging decision, and was also reflected by a higher estimated loss in the section on Value-at-Risk.

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 describes the local trend behaviour, but other models may be investigated. Secondly, the models could be extended with information from other economic variables. Within the exchange rate literature much attention has been given to UIP and/or PPP as building blocks for predicting exchange rates. References to this field include Mark (1995), Bansal (1997), Bansal & Dahlquist (1999), and Evans & Lewis (1995).

Thirdly, the consequences of decision making over longer periods could be investigated. In the hedging decision we focussed attention on deciding whether or not to hedge over a one day horizon. Over longer horizons the estimation of the varying mean process gets a larger weight in the results of the inference.

Fourth, simple mean-variance utility functions correspond to a utility function with Constant Relative Risk Aversion. This hold for the White Noise, Local Level, Generalized Local Level models, and also for the GLL-GARCH model one period ahead, but not for the GLL-

Stochastic Volatility and GLL-Student t models. In these cases, the hedging decision based on the mean-variance utility function is only on approximation to the decision made using the CRRA utility function.

Finally, one may perform the hedge decision for several currencies simultaneously. 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. 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 & Zhou (1996), McCulloch & Rossi (1990), McCulloch & Rossi (1991), and Kandel, McCulloch & Stambaugh (1995).

A A different utility function: Constant Relative Risk Aversion

For ease of notation we rewrite the objective function of an investor with a CRRA utility function as

$$\begin{aligned}
\max_H \mathsf{E}(U(H)) &= \max_H \mathsf{E}\left[\frac{1}{\gamma}\exp(\gamma R(H))W^\gamma\right] && \\
R(H) &= (1-H)s + H\Delta r && \text{Hedged return} \\
s &\sim \mathcal{N}(\mu, \sigma^2) && \text{Return on exchange rate} \\
\Delta r &= r^h - r^f && \text{Interest rate differential}
\end{aligned}$$

The first order condition, assuming that we can interchange the expectation and differentiating operation, is given by

$$\begin{aligned}
\frac{d\mathsf{E}(U(H))}{dH} &= \mathsf{E}\left[\frac{dU(H)}{dH}\right] = \mathsf{E}\left[\frac{dU(H)}{dR} \times \frac{dR}{dH}\right] \\
&= \mathsf{E}\left[\frac{1}{\gamma}e^{\gamma R(H)}\,\gamma W^\gamma \times (-s+\Delta r)\right] \\
&= \mathsf{E}\left[e^{\gamma((1-H)s+H\Delta r)}\,W^\gamma \times (-s+\Delta r)\right] \\
&= W^\gamma\, e^{\gamma H \Delta r}\,\mathsf{E}\left[e^{\gamma(1-H)s} \times (-s+\Delta r)\right] = 0 \\
\Leftrightarrow 0 &= \mathsf{E}\left[(\Delta r - s)e^{\gamma(1-H)s}\right] \equiv \mathsf{E}[(\Delta r - s)\, g(H)]. \qquad (18)
\end{aligned}$$

The last line follows because $W > 0, \gamma < 1, 0 < H < 1$ and Δr is bounded away from $-\infty$. Consequently, elements before the expectation sign cannot become zero. Note that we have not made use of the normality assumption on the expected exchange rate returns yet. For arbitrary distributions the hedge ratios can be found by numerically solving $0 =$

$\mathsf{E}[(\Delta r - s)g(H)]$. Now assume that normality holds. Then it is possible to obtain analytical solutions for the hedge ratio.

First note that the expectation is taken over two terms, we have $\mathsf{E}[(\Delta r - s)\ g(H)] = \Delta r \mathsf{E}_s(g(H)) - \mathsf{E}_s(sg(H))$, where the expectation is with respect to the normally distributed s. We start with the first part, write $c = \gamma(1-H)$, and exclude the multiplication by the constant Δr. We obtain

$$\begin{aligned}
\mathsf{E}_s(g(H)) &= \int e^{cs} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(s^2 - 2\mu s + \mu^2)} ds \\
&= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(s^2 - 2\mu s + \mu^2 - 2\sigma^2 cs)} ds \\
&= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left((s-(\mu+\sigma^2 c))^2 - (2\mu\sigma^2 c + \sigma^4 c^2)\right)} ds \\
&= e^{\mu c + \frac{1}{2}\sigma^2 c^2} \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left(s-(\mu+\sigma^2 c)\right)^2} ds \\
&= \exp\left(\mu c + \frac{1}{2}\sigma^2 c^2\right)
\end{aligned}$$

In a similar fashion, the second term $\mathsf{E}_s(sg(H))$ can be calculated.

$$\begin{aligned}
\mathsf{E}_s(sg(H)) = \ldots &= e^{\mu c + \frac{1}{2}\sigma^2 c^2} \int s\ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left(s-(\mu+\sigma^2 c)\right)^2} ds \\
&= \exp\left(\mu c + \frac{1}{2}\sigma^2 c^2\right) \times (\mu + \sigma^2 c)
\end{aligned}$$

Combining terms, we get for equation (18)

$$\begin{aligned}
\mathsf{E}[(\Delta r - s)\ g(H)] &= \Delta r\ \exp\left(\mu c + \frac{1}{2}\sigma^2 c^2\right) - (\mu + \sigma^2 c)\ \exp\left(\mu c + \frac{1}{2}\sigma^2 c^2\right) \\
&= \left(\Delta r - (\mu + \sigma^2\gamma(1-H))\right) \exp\left(\mu\gamma(1-H) + \frac{1}{2}\sigma^2\gamma^2(1-H)^2\right) = 0 \\
&\Leftrightarrow \Delta r = \mu + \sigma^2\gamma(1-H) \qquad \Leftrightarrow 1 - H = \frac{\Delta r - \mu}{\gamma\sigma^2} \\
&\Leftrightarrow H = 1 - \frac{\Delta r - \mu}{\gamma\sigma^2}
\end{aligned}$$

B Sampling from the posterior density

B.1 Models A-D: Metropolis-Hastings

In Section 2.3 a very brief description of the sampling methods used for the construction of a sample from the posterior density is given. In this appendix, further details are presented.

For models **A-D** the prediction-error decomposition of the likelihood function can be written down. For the White Noise model, this decomposition is immediate, for the Local Level and Generalized Local Level a thorough explanation is given in e.g. Harvey (1989), and for GARCH element in the GLL-GARCH model one can consult Bollerslev (1986). Given the likelihood of the model $\mathcal{L}(Y;\theta)$ for a set of parameters θ, and given the prior $\pi(\theta)$ as specified in Section 2.2, the posterior density is

$$p(\theta|Y) \propto \mathcal{L}(Y;\theta)\pi(\theta)$$

As we are not able to sample directly from this posterior distribution, as it is not a standard distribution, we have to take resort into other sampling methods. For the first four models, we apply the Metropolis-Hastings algorithm using an independent normal candidate density. In short, given an estimate of the location μ and scale Σ of the vector of parameters θ, we sample $\theta^* \sim \mathcal{N}(\mu, \Sigma)$. When the algorithm reached a certain $\theta^{(i)}$ after iteration i, we calculate the acceptance probability $\alpha_{\text{MH}}(\theta^{(i)}, \theta^*)$,

$$\alpha_{\text{MH}}(\theta^{(i)}, \theta^*) = \min \left[\frac{p(\theta^*) f_{\mathcal{N}}(\theta^{(i)})}{p(\theta^{(i)}) f_{\mathcal{N}}(\theta^*)}, 1 \right].$$

With this probability α the candidate vector θ^* is accepted as our new $\theta^{(i+1)}$, else we set $\theta^{(i+1)} = \theta^{(i)}$.

We start with the algorithm using an estimate of the mode of the posterior as the location, and calculate a local approximation to the scale of the density. After a round of sampling of a relatively low number of θ's, we update our estimates of μ and Σ and start anew, sampling a larger number of vectors. Four rounds are executed, where the last contains 5000 accepted drawing from the posterior.

This method is explained in various articles and textbooks. Basic references are Carter & Kohn (1996) and Chib & Greenberg (1995).

B.2 Models E and F: Gibbs sampling with data augmentation

In model **E**, the GLL-Stochastic Volatility model, the variance process h_t is not observable, and therefore it is not possible to write down the likelihood function of the model. However, given the variances the MH sampler of the previous section could be used. Calculations in model **F** are not trivial since the distribution of the disturbances is Student t. It is, however, possible to write the Student t density as an uncountable mixture of normal densities with an Inverted Gamma as the mixing density. Conditioning on a variance gives normality, so that basic sampling methods can be used.

The Gibbs sampling scheme that we apply is an extension of the scheme in Kim et al. (1998). Our models differ from the models in Kim et al. (1998) since we have also included a process for the mean μ_t. The Gibbs sampling scheme is augmented with the following steps

1. Sample a new vector of $\mu_1, .., \mu_T$ conditionally on h, ρ, σ_η^2 and the data. As we condition on the variance elements h, a linear state space model with normal disturbances results. We sample μ using the simulation smoother of de Jong & Shephard (1995).

2. The parameter ρ can be sampled conditionally on the values of the process μ and the variance of the transition equation, σ_η^2. The posterior of ρ would be a normal density with least squares estimates of mean and variance if the prior would be uninformative. With our Beta$(\alpha_\rho, \beta_\rho)$ prior, a slight change in posterior occurs. We sample a candidate ρ^* from the approximating normal density and accept or reject it according to the Metropolis-Hastings acceptance probability. This way, the prior is accounted for correctly.

3. Conditionally on μ and ρ, the variance σ_η^2, and given the IG-2$(\nu_\eta/2, s_\eta/2)$ prior, the posterior can be shown to be IG-2 with parameters $(s_\eta + \sum(\mu_t - \rho\mu_{t-1})^2$ and $\nu_\eta + T)$. This posterior can be sampled from directly.

The posterior for ν in the Student t model **F** is discrete. The posterior is $p(\nu|z,..) \propto t((y_t - \mu_t)/\sigma_\epsilon|\nu,..)\pi(\nu)$ which is not hard as the prior for ν has a discrete support, over a limited number of values of ν.

References

Bansal, R. (1997), 'An exploration of the forward premium puzzle in currency markets', *Review of Financial Studies* **10**, 369–403.

Bansal, R. & Dahlquist, M. (1999), The forward premium puzzle: Different tales from developed and emerging economies. Unpublished manuscript.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Carter, C. K. & Kohn, R. (1994), 'On Gibbs sampling for state space models', *Biometrika* **81**, 541–553.

Carter, C. K. & Kohn, R. (1996), 'Markov chain Monte Carlo in conditionally Gaussian state space models', *Biometrika* **83**, 589–601.

Casella, G. & George, E. (1992), 'Explaining the Gibbs sampler', *The American Statistician* **46**, 167–174.

Chib, S. & Greenberg, E. (1995), 'Understanding the Metropolis-Hastings algorithm', *The American Statistician* **49**, 327–335.

de Jong, P. & Shephard, N. (1995), 'The simulation smoother for time series models', *Biometrika* **82**, 339–350.

Engel, C. & 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. (1995), *Arch: Selected Readings*, Advanced Texts in Econometrics, Oxford [etc.]: Oxford University Press.

Evans, M. D. D. & Lewis, K. K. (1995), 'Do long-term swings in the dollar affect estimates of the risk premia?', *Review of Financial Studies* **8**, 709–742.

Fama, E. F. (1991), 'Efficient capital markets: Ii', *The Journal of Finance* **46**, 1575–1617.

Geweke, J. (1999), Using simulation methods for Bayesian econometric models: Inference, development, and communication. Forthcoming in Econometric Reviews.

Geweke, J. & Zhou, G. (1996), 'Measuring the pricing error of the arbitrage pricing theory', *Review of Financial Studies* **9**, 557–588.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994), 'Bayesian analysis of stochastic volatility models', *Journal of Business and Economic Statistics* **12**, 371–417.

Jorion, P. (1985), 'International portfolio diversification with estimation risk', *Journal of Business* **58**, 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*, McGraw-Hill.

Kandel, S., McCulloch, R. & Stambaugh, R. (1995), 'Bayesian inference and portfolio efficiency', *Review of Financial Studies* **8**, 1–53.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **64**, 361–393.

Kleibergen, F. R. & van Dijk, H. K. (1993), 'Non-stationarity in GARCH models: A Bayesian analysis', *Journal of Applied Econometrics* **8**, S41–S61.

Koop, G. & van Dijk, H. K. (1999), Testing for integration using evolving trend and seasonal models: A Bayesian approach. Unpublished manuscript.

Mark, N. C. (1995), 'Exchange rates and fundamental: Evidence on long-horizon predictability', *American Economic Review* **85**, 201–218.

McCulloch, R. & Rossi, P. (1990), 'Posterior, predictive, and utility-bases approaches to testing the arbitrage pricing theory', *Journal of Financial Economics* **28**, 7–38.

McCulloch, R. & Rossi, P. (1991), 'A Bayesian approach to testing the arbitrage pricing theory', *Journal of Econometrics* **49**, 141–168.

Nelson, D. B. (1990), 'Stationarity and persistence in the GARCH(1, 1) model', *Econometric Theory* **6**, 318–334.

Smith, A. F. M. & Roberts, G. O. (1993), 'Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods', *Journal of the Royal Statistical Society B* **55**, 3–24.

Taylor, S. J. (1994), 'Modeling stochastic volatility: A review and comparative study', *Mathematical Finance* **4**, 183–204.