Econometric Institute Report EI2000-25/A

Daily Exchange Rate Behaviour and Hedging of Currency Risk

Charles S. Bos *
Econometric and Tinbergen Institutes, Erasmus University Rotterdam

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

Herman K. van Dijk
Econometric and Tinbergen Institutes, Erasmus University Rotterdam

August 20, 1999
July 4, 2000, revised

Abstract

We construct models which enable a decision-maker to analyze the implications of typical time series patterns of daily exchange rates 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 strategies. Consequently, we can make a distinction between statistical relevance of model specifications, and the economic consequences from a risk management point of view. We compute payoffs from several alternative hedge strategies. These payoffs indicate that modelling time-varying features of exchange rate returns may lead to improved hedge behaviour within currency overlay management.

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

*Correspondence to Charles Bos, Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands. Email: cbos@few.eur.nl, URL: http://www2.tinbergen.nl/~cbos/.
We thank three anonymous referees, the co-editor, Rob Engle, Christopher Gilbert, Frank de Jong, Michel Lubrano, Alan Timmerman and participants of seminars at UC-San Diego, UCLA, UC-Irvine, EC^2 in Madrid, ISBA-2000, and at universities in Stockholm, Marseilles and Amsterdam for several remarks which led to substantial improvements on an earlier version of this paper. The authors retain full responsibility for remaining errors. This is a revised version of Econometric Institute Report EI-9936/A.

1 Introduction

When investing abroad, international firms naturally face the decision whether or not to hedge the risk of a depreciation of the foreign currency compared to the home currency. In many cases the decision whether or not to hedge currency risk is made independently from the underlying investment decision. In the finance industry this approach to currency hedging is called 'currency overlay management', which is especially relevant for internationally operating corporations. For example, when such 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 exposed to foreign currency exposure are internationally operating investors, like banks, pension funds, and insurance companies. For these investors the currency exposures arise from the investment strategies that these institutions follow. For example, when a US dollar-based investor decides to diversify into Japanese stocks he runs the risk of the Japanese yen to depreciate. Although the portfolio allocation decision could also depend on the risk and return characteristics of foreign currencies, in practice these two decisions are often separated. In that case currency overlay management is used to manage the international exposures arising from the investments in all kinds of securities (stocks, bonds, cash, derivatives, among others). Note that this 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 Japanese yen depreciating too large, he may decrease his holdings of Japanese stocks in the first place. 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 exchange rate, which gives the investor the right (and the obligation) to convert the foreign currency exposure from one currency to another for a fixed rate somewhere 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 derivatives. In this paper we will 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 proportional to the return on the exchange rate. Second, the decision-maker hedges the currency risk completely. Now, the return is proportional to the difference between the short-term interest rates of the home country and that of the foreign country. A practical example is the case of a German firm with US-investments. In the period 1998-1999, the cumulative return on the DMark/US dollar (DEM/USD) exchange rate was almost eight percent, while the cumulative difference between the two interest rates was less than minus three percent. Thus, the decision to hedge or not to hedge relates to a difference in cumulative return in two years of approximately eleven 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 a tactical strategy for foreign exchange rate management, like currency overlay management, is an important topic.

In this paper we analyze 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 fat tails of a histogram 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 Bayesian methods based on Markov chain Monte Carlo (MCMC) simulation. Third, in the decision analysis we investigate the payoff from an optimal strategy using alternative models and payoffs from alternative strategies for some selected models.

The outline of the paper is as follows. In section 2 we introduce our procedure for executing the currency overlay strategy. In section 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 & Migon (1985). In section 4 we discuss our Bayesian methods, see e.g. Smith & Roberts (1993) and Chib & Greenberg (1995). In the recent literature these methods have been successfully applied for studying separately the pattern of varying means (see Carter & Kohn 1994, Koop & van Dijk 2000) and the pattern of varying volatilities (see Kim, Shephard & Chib 1998). Results are presented in section 5 using the DEM/USD dollar daily exchange rate series for the period January 1982 until December 1999. Some concluding remarks are given in section 6. Conditional densities used in MCMC sampling from the posterior are summarized in appendix A.

2 Currency hedging

As noted in the introduction, we concentrate on tactical strategies for exchange rate management. The setting that we investigate in this paper can be described as follows. Let s_{t+1} be the exchange rate return over the time interval $[t, t+1]$, defined as $s_{t+1} \equiv 100 \ln(S_{t+1}/S_t)$. Let $F_{t,\tau}$ be the current value of a forward contract with maturity date τ, which by covered interest rate parity is equal to

$$F_{t,\tau} = S_t \frac{1 + r^h_{t,\tau}}{1 + r^f_{t,\tau}},$$

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:

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

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 to one, only the interest rate differential has an impact, whereas changes in the currency do not affect the return on the currency overlay.[1] In the empirical part of our paper we set τ equal to thirty days, i.e. we consider 30-day forward contracts when hedging currency exposures. However, we allow the investor to change his hedge position every day. The reason why we do this is to investigate the robustness of our hedging strategies under the condition that the investor is most flexible to adjust his hedge position. Another reason to do this is that the exchange rate is much more volatile than the interest rate differential.[2]

Given a time series model that captures exchange rate behaviour and all information up to time t, the currency manager wants to determine the hedge ratio that should apply to the next period. In order to perform this task he needs to specify an objective function that captures his risk and return attitudes towards foreign currencies over some future time horizon. We will assume that the investor has a standard power utility function:

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

The variable W_t represents the wealth that the investor is obtaining 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 $W_{t+1} = W_t(1+r_{t+1})$. 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

$$\max_{0 \le H_t \le 1} \mathsf{E}_{s_{t+1}|\mathcal{I}_t} U(W_{t+1}|W_t) = \max_{0 \le H_t \le 1} \mathsf{E}_{s_{t+1}|\mathcal{I}_t} \left[\frac{\left(1 + r_{t+1}(s_{t+1}, H_t, r^h, r^f)\right)^\gamma - 1}{\gamma} \right],$$

with $\mathsf{E}_{s_{t+1}|\mathcal{I}_t}$ a conditional expectations operator, taken with respect to the predictive density of tomorrow's return s_{t+1}, given all available information $\mathcal{I}_t$.

In the empirical part of this paper we compare the hedging decisions based on optimization of a power utility function with hedging decisions based on Value-at-Risk (VaR), and decisions based on the Sharpe ratio. Comparison of optimal decisions with the results obtained from more pragmatic decision rules may give useful insight into issues like the robustness of the optimal strategy.

Decision rules based 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

[1]The hedge ratio is restricted to lie between 0 and 1. We do not allow currency overlay managers to use the currency position as an investment in its own right, going short or long. This is a topic for further research.

[2]The fact that we hedge currency exposures with contracts that have a longer maturity than the horizon at which the decisions are taken implies that the investor may need to reverse the forward contracts before they mature.

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

$$\int_{\mathrm{VaR}}^{\infty} f(r_{t+1}|\mathcal{I}_t)dr_{t+1} = 1 - \alpha, \tag{2}$$

with $1 - \alpha$ the confidence probability, which typically ranges from 90% to 99%. 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). In the application we concentrate on results for the 95% VaR. When the VaR is calculated, the investor can evaluate if it is higher than a certain limit, and choose to hedge if this is the case.

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

$$\mathrm{Sh} = \frac{\mathsf{E}_{s_{t+1}|\mathcal{I}_t}(r_{t+1})}{\sqrt{\mathsf{Var}_{s_{t+1}|\mathcal{I}_t}(r_{t+1})}}, \tag{3}$$

with $\mathsf{Var}_{s_{t+1}|\mathcal{I}_t}(r_{t+1})$ 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 –which indicates a relatively high return compared to the risk– no hedging is required, and vice versa.

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

$$s_t = \mu_t + \epsilon_t \qquad \epsilon_t \sim \text{i.i.d.}(0, \sigma^2_{\epsilon,t}) \tag{4}$$

$$\mu_t = \rho\mu_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma^2_\eta) \qquad t = 1, .., T, \tag{5}$$

The unobserved mean component μ_t is an autoregressive process with disturbances η_t and autoregressive parameter $\rho, 0 \leq \rho < 1$. This model is supposed to pick up the periods of rising or falling exchange rate levels[3]. The disturbances η_t are assumed to be independently and identically normally distributed with constant variance σ^2_η. The autoregressive 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 expect that in practice, the variance of η_t is so small that the I(1) behaviour of S_t overwhelms the I(2)

[3]Theoretically the interest rate differential should be introduces as the expectation of s_t, as the uncovered interest rate parity (UIP) prescribes. Preliminary analysis indicated that the scale of the UIP was small relative to the exchange rate return. The interest rate differential will be introduced later in the evaluation of the returns.

effects. One can also take the limit case $\sigma_\eta^2 = 0, \rho = 1$. 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.

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. Engle (1982), Bollerslev (1986), Nelson (1990), Engle (1995) or Taylor (1994). Conditioning on all information $\mathcal{I}_{t-1}$ available at time t, we write

$$\epsilon_t | \mathcal{I}_{t-1} \sim \mathcal{N}(0, \sigma_{\epsilon,t}^2). \tag{6}$$

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

$$\sigma_{\epsilon,t}^2 = \sigma_\epsilon^2 \tag{7}$$

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

$$\begin{aligned} \sigma_{\epsilon,t}^2 &= \sigma_\epsilon^2 h_t \\ h_t &= \delta h_{t-1} + \omega + \alpha \epsilon_{t-1}^2 / \sigma_\epsilon^2 \\ \delta \geq 0, &\quad \alpha \geq 0, \quad \delta + \alpha < 1 \quad \omega \equiv 1 - \delta - \alpha. \end{aligned} \tag{8}$$

The restrictions on the parameters are sufficient to ensure strict positiveness of $\sigma_{\epsilon,t}^2$ and the existence of a finite value for the unconditional expectation $\mathsf{E}(\sigma_{\epsilon,t}^2) = \sigma_\epsilon^2$ or equivalently $\mathsf{E}(h_t) = 1$ (see Kleibergen & van Dijk 1993).

A second family of disturbance processes for ϵ_t with time varying variance is a stochastic volatility process (see Jacquier, Polson & Rossi 1994). The variance of the disturbances in the observation equation (4) follows the stochastic process

$$\begin{aligned} \sigma_{\epsilon,t}^2 &= \exp(h_t) \\ h_t &= \mu_h + \phi(h_{t-1} - \mu_h) + \xi_t \qquad 0 \leq \phi < 1 \\ \xi_t &\sim \mathcal{N}(0, \sigma_\xi^2) \end{aligned} \tag{9}$$

A third feature of financial time series is that the histograms of the series exhibit heavier tails than the normal density. To model this, we replace equation (6) by

$$\epsilon_t | \mathcal{I}_{t-1} \sim t(0, (\nu - 2)\sigma_{\epsilon,t}^2, 1, \nu), \qquad \nu > 2 \tag{10}$$

where t indicates the Student t density.[4]

Figure 1 summarizes the models that are used in subsequent sections. The basic model is the one based on normality of the returns. Then there are three directions of generalization: time dependence of the mean μ_t, time dependence of the variance $\sigma_{\epsilon,t}^2$, 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. The Generalized Local Level (GLL) model is combined with the three generalizations, such that a broad range of competing models is found. When the GLL model is combined with both the GARCH and the Student t elements, a most general model in the fifth line results. The models are indicated by the letters **A**-**G** in the figure and in tables in subsequent sections.

[4]The assumption of Student t distributed disturbances may also be explained as a generalization allowing for time variation of the variances $\sigma_{\epsilon,t}^2$.

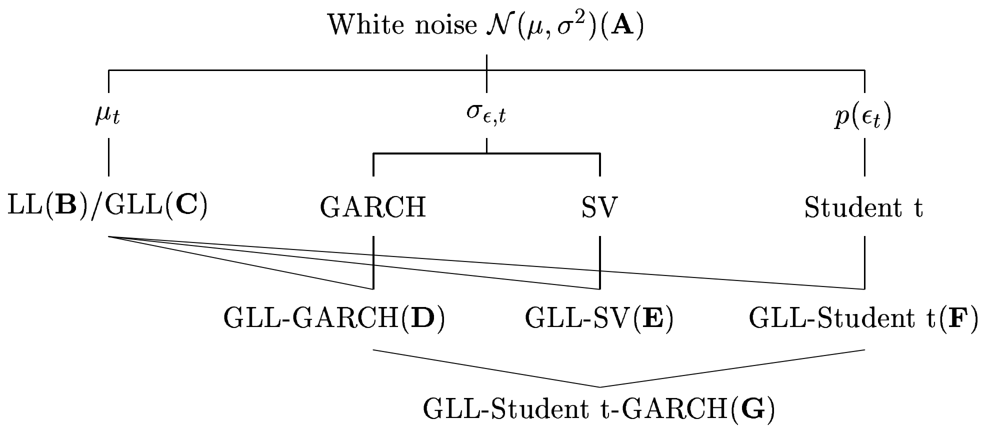

Figure 1: Hierarchy of models

4 Bayesian inference and decision

4.1 Prior structure

Inference and decision analysis is performed within a Bayesian framework. In table 1 we present the priors on the parameters of the models that are used. As we want to be able to compare the marginal likelihood of the models (see section 4.3), we are bound from using uninformative priors. Analysis is greatly simplified when the choice is further limited to the set of conjugate priors. The mean parameters μ in the White Noise model and μ_h in the volatility process of the SV model therefore get a normal prior. Hyperparmaeters are chosen such that effectively no information is put into these priors.

The autoregression 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 parameters for the signal-to-noise ratio). Given the fact that trends in exchange rates may last for several months, we deem a large value of ρ 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 distribution with mean 0.8 and a rather large standard deviation of 0.2.[5] 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.

The variance parameters are all given inverted gamma prior (see e.g. Poirier (1995), p. 111) 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 & Lubrano (1998) it is proven how a prior for the degrees-of-freedom parameter ν with too heavy tails can ruin the properness of the posterior. The truncated Cauchy prior used here ensures that these problems do not occur.

The GARCH parameters δ, 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]Other priors, including a uniform prior between 0 and 1, were used. Results were similar to the results presented in this paper.

Table 1: Description of priors used

Parameter	Prior	Hyper-parameters	Used in model
μ	$\mathcal{N}(\mu_0, \sigma_0^2)$	$\mu_0 = 0, \sigma_0 = 2/100$	**A**
σ_ϵ^2	$\text{IG}(\alpha_\epsilon, \beta_\epsilon)$	$\alpha_\epsilon = 2.5, \beta_\epsilon = 4/3$	**A**, **B**, **C**, **D**, **F**, **G**
ρ	$\mathcal{N}(\mu_\rho, \sigma_\rho^2)$	$\mu_\rho = .8, \sigma_\rho = .2$	**C**, **D**, **E**, **F**, **G**
σ_η^2	$\text{IG}(\alpha_\eta, \beta_\eta)$	$\alpha_\eta = 2.25, \beta_\eta = 100$	**B**, C, **D**, **E**, **F**, **G**
δ, α	Uniform at stationary region		**D**, **G**
μ_h	$\mathcal{N}(\mu_{\mu_h}, \sigma_{\mu_h}^2)$	$\mu_{\mu_h} = -1, \sigma_{\mu_h} = 1$	**E**
ϕ	$\mathcal{N}(\mu_\phi, \sigma_\phi^2)$	$\mu_\phi = .5, \sigma_\phi = .3$	**E**
σ_ξ^2	$\text{IG}(\alpha_\xi, \beta_\xi)$	$\alpha_\xi = 2.5, \beta_\xi = 4/3$	**E**
ν	Truncated Cauchy, $\nu > 2$	-	**F**, **G**

4.2 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), p. 104 and further. The posterior distribution of the parameters, $P(\theta|Y)$, 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 2000) 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 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**-**G**, the GLL-Stochastic Volatility, GLL-Student t and GLL-Student t-GARCH models, we apply a data augmentation scheme to obtain conditional normality and include the unobserved variables into the state. We make use of a Gibbs sampling scheme as in Kim et al. (1998). See Appendix A for further details.

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 is defined as

$$m = \int L(Y; \theta, M)\pi(\theta|M)\partial\theta \tag{11}$$

and may be computed as

$$m = \frac{L(Y; \theta, M)\pi(\theta|M)}{P(\theta|Y, M)} \tag{12}$$

with $P(\theta|Y, M)$ equal to the posterior density of a model M evaluated at the location indicated by the vector of parameters θ, and $L(Y; \theta, M)$ and $\pi(\theta|M)$ the likelihood and prior, respectively (see e.g. Gelfand & Smith 1990).

In the present setting, the normalizing constant of the posterior density is not known in closed form. Instead, we only have a sample from the posterior available. For the models

A-**D**, the likelihood can be evaluated, and therefore the integrating constant can be found by evaluating likelihood and prior in e.g. the posterior mean, and dividing it through by a kernel approximation to the posterior density in the same location (for details see Kass & Raftery 1995).

For models **E**-**G**, the likelihood function is only available as a high-dimensional integral over unobserved components, which are used in the Gibbs sampling algorithm to obtain tractable conditional distribution (see appendix A). Chib (1995) describes a procedure to calculate the marginal likelihood in this case. In section 5.2 the results for models **A**-**D** are calculated using both methods, to judge the accuracy and comparability of the approximation methods. For models **E**-**G**, only the Gibbs results are reported.

The method of Chib uses the conditional densities as described in appendix A. In cases where a Metropolis-Hastings step within Gibbs was applied, numerical integration was used to evaluate the necessary conditional posterior densities.

4.4 Predictive analysis

The decision whether to hedge or not is based on the unconditional predictive density $P(s_{t+1}|\mathcal{I}_t)$ of tomorrow's returns on the exchange rate s_{t+1}, given all available information $\mathcal{I}_t = \{s_1, .., s_t\}$. The conditional density $P(s_{t+1}|\mathcal{I}_t, \theta)$, given the vector of parameters θ, is easily derived. The unconditional predictive density follows by marginalization with respect to θ,

$$P(s_{t+1}|\mathcal{I}_t) = \int_{\theta\in\Theta} P(s_{t+1}|\mathcal{I}_t, \theta)P(\theta|\mathcal{I}_t)\partial\theta, \tag{13}$$

see e.g. Geweke (1989), Barberis (2000). Marginalization should be done with respect to the posterior distribution of $\theta|\mathcal{I}_t$. For computational reasons we do not reestimate the model for every day in our evaluation period, but use N drawings $\theta^{(1)}, .., \theta^{(N)}$ from the posterior of $\theta|\mathcal{I}_T$, with $\mathcal{I}_T$ the observations from the estimation sample $(T < t)$. When the estimation sample is large compared to the evaluation sample, this approximation gives a sufficient level of accuracy. The integral in (13) is approximated using

$$P(s_{t+1}|\mathcal{I}_t) \approx \frac{1}{N}\sum_{i=1}^{N} P(s_{t+1}|\mathcal{I}_t, \theta^{(i)}) \tag{14}$$

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

4.5 Decision analysis

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

$$\begin{aligned} H_t &= \arg\max_{H_t} \mathsf{E}_{s_{t+1}|\mathcal{I}_t} U(W_{t+1}|W_t) = \\ &= \arg\max_{H_t} \int_{s_{t+1}} \frac{(1 + r_{t+1}(H_t, s_{t+1}, r^h, r^f))^\gamma - 1}{\gamma} P(s_{t+1}|\mathcal{I}_t) ds_{t+1} \end{aligned} \tag{15}$$

Optimal hedge ratios are computed for every day in the evaluation period.

In section 2, two other decision strategies were presented. For the Value-at-Risk (VaR), we evaluate for each day what the 95% 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 fix a cut-off level for the VaR such that the average hedge ratio corresponds to the average hedge ratio found when fully optimizing the log-utility function, where $\gamma = 0$.

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 that 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 case $\gamma = 0$.

5 Hedging against DMark/US dollar currency risk

5.1 Stylized facts

Our data set consists of daily observations on the DMark/US dollar (DEM/USD) exchange rate for the period January 1, 1982 until December 31, 1999 which gives a total of 4695 observations. For this same period we have the 1-month Eurocurrency interest rates for the German DMark and the US Dollar.

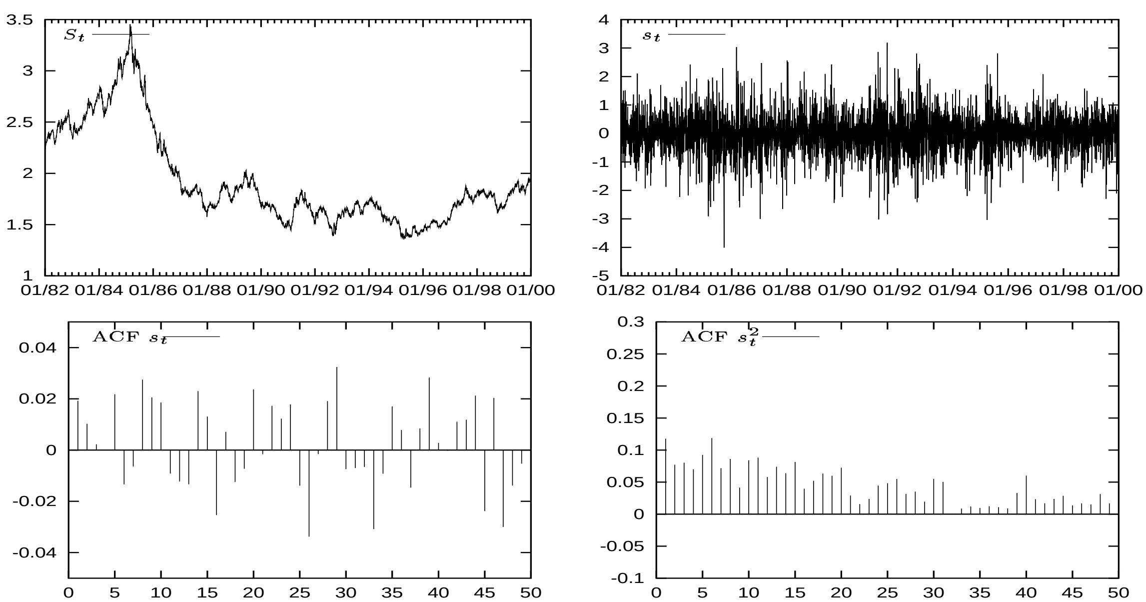

Figure 2: DEM /USD Exchange rate, January 1, 1982 until December 31, 1999. Panels contain data in levels (top left), in returns (top right), and the autocorrelation function of the returns and the squared returns (bottom)

In the upper panel of figure 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 exhibit patterns frequently found in high frequency financial return data. As for the returns, there is no clear serial correlation pattern, corroborating the widely held view that financial return series are unpredictable.

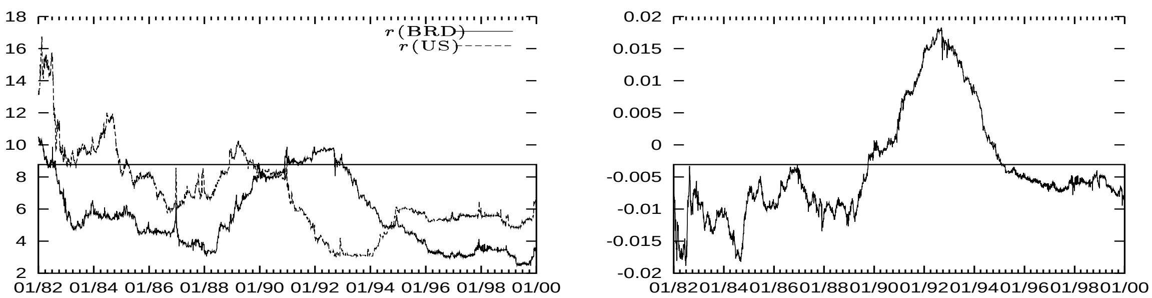

Figure 3: 1-month Eurocurrency interest rates for the German DMark and the US Dollar and the daily differential

However, the local trends in the levels of the exchange rates may prove useful for practical currency overlay strategies. The squared returns show a clear pattern. The slowly decaying autocorrelation has prompted many researchers to develop models for describing time-varying volatilities.

The first panel of figure 3 shows the time series of both U.S. and German 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 are very persistent. This feature can also be seen when we construct the interest rate differential in the second panel of figure 3. This latter series is used in the hedging decision.

5.2 Posterior results

For models **A**-**D** the Metropolis-Hastings sampling algorithm was used.[6] After three initial runs of the MH sampler (with 500, 2000 and 10000 drawings each) for improving the location and scale estimates for the normal candidate density, a final sample was collected. The sampling continued until a total of 100.000 drawings was accepted. From every 10 drawings, only 1 was saved, in order to lower correlation in the posterior sample. Acceptance rates were 98, 93, 67 and 61 percent, respectively. This corresponds to final sample sizes of 10147, 10682, 14787 and 16387.

The models with Student t disturbances or Stochastic Volatility components did not allow for implementation of the MH sampler. The Gibbs sampler we used was run for a burn-in period of 50000 iterations, and continued for another 500.000 iterations for constructing a sample. As higher correlation is to be expected in a Gibbs chain, we use only one out of every 50 drawings.

The correlation in a Gibbs chain with multiple blocks can be quite high (see Kim et al. 1998). Figure 4 shows the autocorrelation function of the drawings for the GLL-Student t-GARCH model; it is seen that only after about 30 drawings, correlation dies out.

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, see Geweke (1992). 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

[6] All results reported in this paper were obtained using programs written by the authors in Ox version 2.20, see Doornik (1999). For the filtering and smoothing of the state space models, SsfPack version 2.3 (see Koopman, Shephard & Doornik 1999) was used extensively.

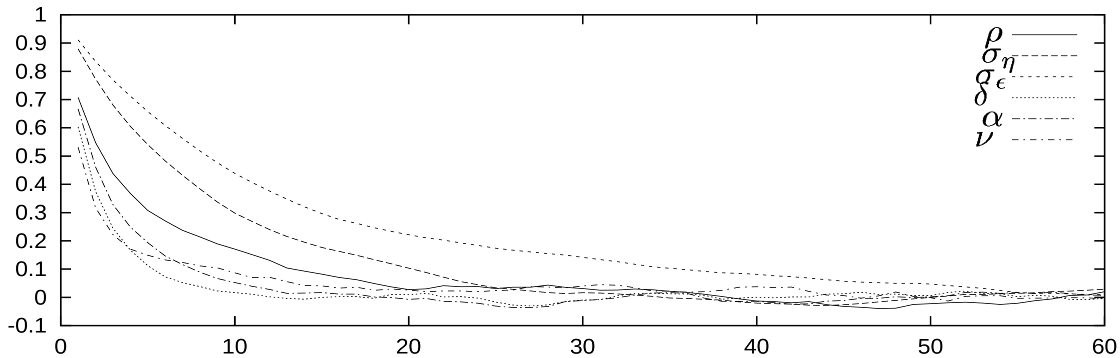

Figure 4: Autocorrelation function of draws from the parameters of the GLL-Student t-GARCH model

(Newey & West 1987), adjusting for correlation with lags up to 4% of size of the sample, we find values for the RNA of over 40% for the WN, LL and GLL models, of at least 25% for the GLL-GARCH model, and between 10 and 70% for models **E**-**G** where the Gibbs sampler was used. These numbers imply that in the worst case, for the GLL-Student t-GARCH model, the sample from the Markov chain of 10.000 dependent drawings roughly corresponds to a sample of 1.000 independent drawings from the posterior.

Table 2: Posterior results

Parameter	WN		LL		GLL	
$\mu \times 100$	-0.40	(0.92)				
	-0.39	[-2.26,1.35]				
ρ					0.69	(0.12)
					0.75	[0.44,0.90]
$\sigma_\eta \times 10$			0.24	(0.02)	0.70	(0.20)
			0.23	[0.19,0.28]	0.59	[0.36,1.10]
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]

In each cell of the table the posterior mean and standard deviation are given on the first line, with the mode and 95% highest posterior density region on the second line.

The main characteristics of the posteriors are summarized in tables 2-3. For each model and for each parameter, the mean, standard deviation, mode and the bounds of the 95% highest posterior density region are reported.[7]

The posterior of the two parameters of the White Noise model was 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 0.02. The standard deviation of the observation disturbance, σ_ϵ, is rather larger at 0.67. This corresponds a low signal-to-noise ratio. Therefore, the I(1) process of μ_t is not well distinguishable between the

[7] All 95%-HPD regions were continuous.

Table 3: Posterior results

Parameter	GLL-GARCH		GLL-SV		GLL-Student t		GLL-GARCH-Student t	
ρ	0.76	(0.10)	0.68	(0.14)	0.66	(0.14)	0.78	(0.11)
	0.83	[0.56,0.92]	0.77	[0.41,0.92]	0.72	[0.39,0.90]	0.85	[0.56,0.94]
$\sigma_\eta \times 10$	0.69	(0.20)	0.66	(0.21)	0.58	(0.14)	0.55	(0.12)
	0.60	[0.36,1.11]	0.54	[0.33,1.06]	0.53	[0.34,0.86]	0.50	[0.34,0.79]
σ_ϵ	0.66	(0.03)			0.70	(0.02)	0.78	(0.06)
	0.65	[0.61,0.71]			0.69	[0.67,0.73]	0.75	[0.67,0.92]
δ	0.90	(0.01)					0.92	(0.01)
	0.90	[0.88,0.93]					0.92	[0.90,0.93]
$\alpha \times 10$	0.65	(0.08)					0.66	(0.07)
	0.65	[0.51,0.80]					0.66	[0.52,0.80]
ν					4.22	(0.28)	4.83	(0.31)
					4.16	[3.67,4.78]	4.78	[4.23,5.46]
μ_h			-1.06	(0.06)				
			-1.07	[-1.18,-0.93]				
ϕ			0.92	(0.02)				
			0.93	[0.89,0.95]				
σ_ξ			0.28	(0.03)				
			0.28	[0.23,0.35]				

In each cell of the table the posterior mean and standard deviation are given on the first line, with the mode and 95% highest posterior density region on the second line.

I(0) noise.

More interesting are the posteriors for the GLL model. The distribution of the observation standard deviation 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 distributions. This same effect can also be observed for the posterior of ρ and σ_η for the other models. In figure 5 the marginal posteriors of the parameters of the GLL-Student t-GARCH model are plotted. Indeed, it is seen that the posteriors are rather more pronounced than the priors, and that the posterior distributions of ρ and σ_η are skewed.

The HPD region for the parameter ρ is wide, especially when one realizes that the dataset comprises almost 4500 datapoints. We note that $\rho = 0$ (WN) and $\rho = 1$ (LL) are not in the HPD interval.

The contrast with the GARCH parameters (both in the GLL-GARCH as in the GLL-Student t-GARCH model) is large. Both δ and α are estimated quite precisely, with tight and almost symmetric posteriors distributions. Including the Student t disturbances in the GARCH model does not alter the posterior of the GARCH parameters δ, α greatly. Only the standard deviations change, as the Student t disturbance takes up part of the variance. More interesting is the change in the posterior of σ_η: For the GLL-Student t-GARCH model, the standard deviation of the unobserved process μ_t is smaller than for the GLL-GARCH model. Note that the posteriors of both σ_η and σ_ϵ appear to be closer to the posteriors of the Student t model than of the GLL-GARCH model.

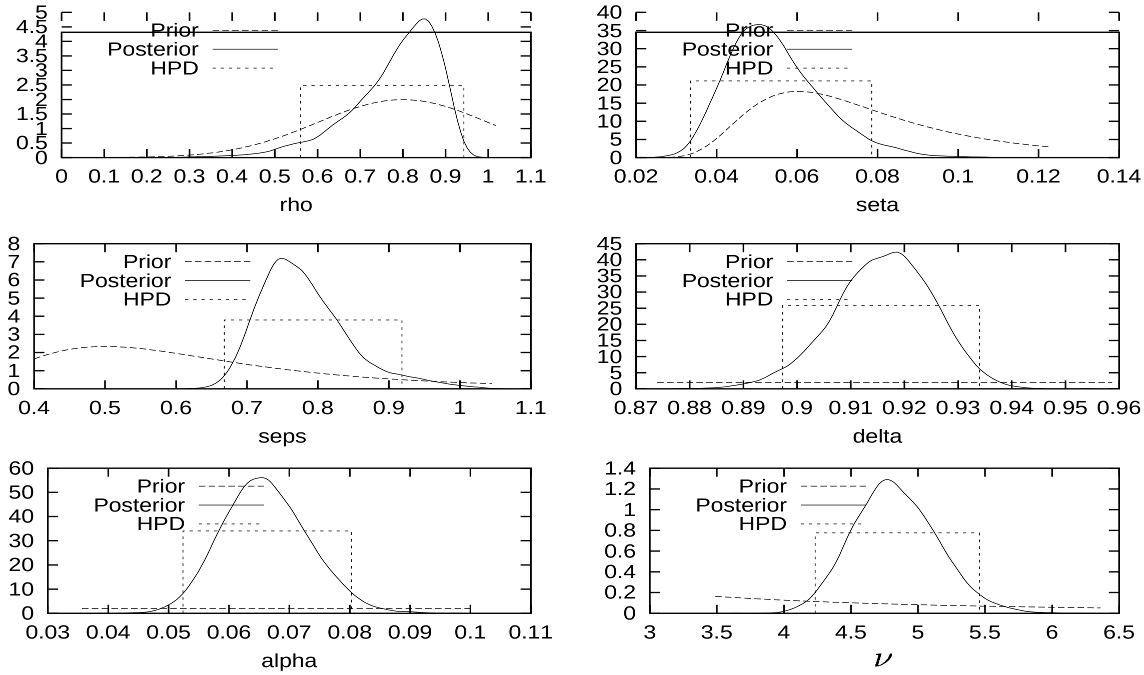

Figure 5: Prior, posterior and HPD region of the parameters in the Generalized Local Level-Student t-GARCH model

The degrees-of-freedom parameter ν in the last two models is estimated between 4 and 5. The mean and mode of ν for the GLL-Student t-GARCH model is not included in the HPD region of the model without the GARCH component; apparently the fact that the GLL-Student t model cannot yield changing volatility results in a smaller ν.

The GLL-SV model results in posterior estimates for ρ which are slightly lower than for the GLL-Student t-GARCH case. Again, the HPD interval does not include border values of $\rho = 0, 1$. The parameters of the stochastic volatility part are well-identified, like the parameters $\sigma_\epsilon, \delta, \alpha$ for the models with a GARCH component. The estimate of μ_h governs the unconditional variance. The standard deviation can roughly be approximated by $\mathbf{E}(\sigma_{\epsilon,t}) \approx \exp(\mu_h/2) = 0.6$.

5.3 Marginal likelihood

The marginal (log)likelihood has been calculated for each of the models (see table 4). The kernel method is only available for models **A**-**D**; for these models, the loglikelihoods calculated using the kernel approximation correspond well to the values found using the Gibbs' conditional densities approach.

The loglikelihoods indicate that the data does not provide a strong evidence in terms of gain in the likelihood function when the varying mean component is introduced (compare the results for the WN and the GLL models). The LL model is inferior to the GLL model. The modelling steps on the varying variance structure (allowing for GARCH or SV) lead to a substantial improvement in the marginal likelihood over the more basic WN or GLL models. The fixed variance Student-t and the GARCH extensions result in an improvement of the loglikelihood score of 142 and 163 points, respectively. Better is the combination of the

Table 4: Marginal loglikelihoods

Model	Kernel	Gibbs	Relative to GLL Kernel	Relative to GLL Gibbs
WN	−4306.9	−4305.9	−1.1	−6.5
LL	−4355.5	−4354.0	−49.7	−54.6
GLL	−4305.8	−4299.4	0	0
GLL-GARCH	−4138.8	−4136.2	167.0	163.2
GLL-SV	.	−4021.7	.	277.7
GLL-Student t	.	−4157.2	.	142.2
GLL-Student t-GARCH	.	−4051.1	.	248.3

two, with both varying variances and heavy tailed disturbances. The GLL-SV model, which in flexibility is a close competitor to the GLL-Student t-GARCH model, fits the data best, according to the marginal likelihood.

5.4 Predictive density

The predictive density $P(s_{t+1}|Y_t)$ summarizes all information on which the investor bases the decision whether to hedge or not. It is instructive to look at the implications of model assumptions for the possible shape and time variability of this density.

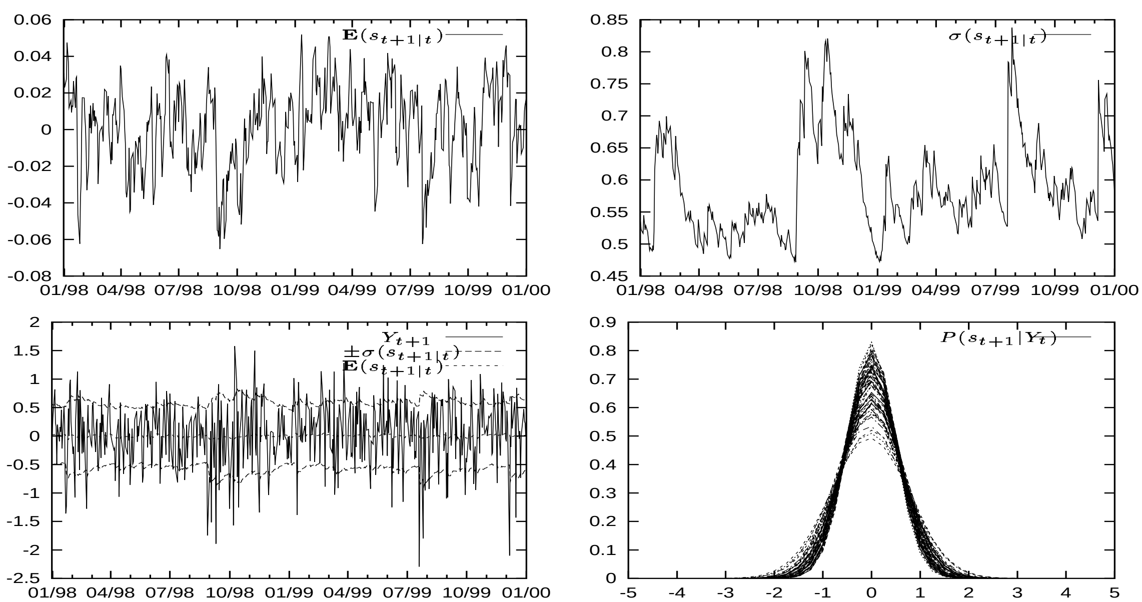

Figure 6: Predictive mean, standard deviation, together with observations and the full density over 1/1/98-31/12/99 (top left to bottom right) for the GLL-GARCH model

The case of the GLL-GARCH model entails the most important characteristics of our set of models. The top left panel of figure 6 displays the mean $\mathsf{E}(s_{t+1|t})$ of the predictive density $P(s_{t+1}|Y_t)$. 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, a continuing decline in the exchange rate is predicted, whereas in most months in 1999 $\mathbf{E}(s_{t+1|t})$ is positive. The scale is indicated in daily percentage changes; though the changes from day to day are noticeable, they are of a size of as a maximum 0.02%.

In the top right panel of the same figure, the standard deviation of the prediction is given. Around September 1998 where the predicted change in exchange rate became negative, the standard deviation jumped up. From that moment on the volatility remains high, until in January 1999, where the GARCH component indicates that the variance of the series diminishes again to the levels of mid-1998. These jumps in the standard deviation only occur in models **D**, **E** and **G**, 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 $P(s_{t+1}|Y_t)$ for the days of the evaluation period. Is it seen that the spread of the distribution changes considerably, the location hardly moves. For models **A**-**D** and **F**, the corresponding plot shows less variation over time as the variance is fixed. In the following we will investigate whether the predictive densities provide enough information for constructing practical currency overlay strategies.

5.5 Benchmark hedging strategies

In the following sections optimal hedging decisions and decisions based on measures of Value-at-Risk and the Sharpe ratio are presented. To have a benchmark against which to judge our results, table 5 contains the hedging results for three base cases.

Table 5: Results for deterministic hedging strategies

	Full hedging		No hedging		RW	
C ($\sigma(C)$)	-3.15	(0.02)	8.52	(12.51)	7.85	(9.10)
$U(C) \times 100$	-3.20		8.18		7.56	

The first column gives the results for the case where the investor chooses to hedge each day. Consequently, the investor pays the interest rate differential. Over the evaluation sample, 522 days within the period January 1, 1998–December 31, 1999, the cumulative difference between the interest rates results in a loss of 3.2%. The approximative standard deviation of the cumulative return, presented between parentheses, was 0.02%.[8]

[8]The cumulative return is calculated as

$$C = 100\left(\prod_{t=1}^{T}(1+\frac{r_t}{100}) - 1\right)$$

with r_t the realized return based on the realized exchange rate return, the hedge ratio and the interest rate differential. As an approximative standard deviation, the quantity $\sigma(C) = \sqrt{T\mathsf{Var}(r)}$ is calculated, with $\mathsf{Var}(r)$ the variance of the daily returns over the evaluation period.

The second column gives the results when no hedging is undertaken. Over the two years, the dollar appreciated 8.5%, with a standard deviation of 12.5%. We also added a Random Walk strategy, that hedges the exposure when the previous period's exchange rate was depreciating and vice versa. This strategy leads to a final return of 7.9%, with a lower variance, as this strategy naturally includes days on which the hedge ratio was equal to one.

5.6 Optimal hedging

We have evaluated currency overlay strategies for a currency manager with power utility, for three different values of the risk aversion parameter $\gamma, \gamma = 0, \gamma = -2$ and $\gamma = -10$. The first case coincides with a logarithmic utility, which is often used as a benchmark utility function. The other two values of γ coincide with a moderately risk averse investor, and a more risk averse investor.

For every day within the evaluation period we calculated optimal hedge ratios, using only information available up to each particular day in this period. In this way we closely mimic the historical hedging behaviour of a currency manager. In table 6 we present the results of this strategy.

Table 6: Hedging results

		γ=-10		γ=-2		γ=0	
WN	$\overline{H}$	0.91		0.68		0.11	
	$C\ (\sigma(C))$	−2.15	(1.12)	0.41	(4.12)	5.82	(11.22)
	$U(C) \times 100$	−2.43		0.40		5.66	
LL	$\overline{H}$	0.47		0.39		0.37	
	$C\ (\sigma(C))$	3.57	(8.49)	4.87	(9.55)	6.93	(9.81)
	$U(C) \times 100$	2.96		4.53		6.70	
GLL	$\overline{H}$	0.78		0.38		0.27	
	$C\ (\sigma(C))$	−1.05	(3.27)	6.63	(8.96)	9.37	(10.20)
	$U(C) \times 100$	−1.11		6.02		8.95	
GLL-GARCH	$\overline{H}$	0.61		0.38		0.34	
	$C\ (\sigma(C))$	−0.55	(6.35)	7.50	(9.23)	9.43	(9.73)
	$U(C) \times 100$	−0.57		6.73		9.01	
GLL-SV	$\overline{H}$	0.62		0.35		0.29	
	$C\ (\sigma(C))$	4.40	(4.88)	8.92	(9.20)	10.08	(10.11)
	$U(C) \times 100$	3.50		7.86		9.60	
GLL-Student t	$\overline{H}$	0.77		0.35		0.21	
	$C\ (\sigma(C))$	−0.74	(3.26)	5.55	(9.08)	8.03	(10.79)
	$U(C) \times 100$	−0.77		5.12		7.72	
GLL-Student t-GARCH	$\overline{H}$	0.58		0.34		0.28	
	$C\ (\sigma(C))$	2.26	(6.68)	8.59	(9.57)	7.55	(10.32)
	$U(C) \times 100$	2.00		7.60		7.27	

Indicated in the table are the average hedge ratio over the period 1/1/1998 until 31/1/1999, optimizing the utility function, for different values of risk aversion parameter γ. The cumulative return C is given in the second row together with the standard deviation of observed returns, $\sigma(C)$. $U(C)$ indicates the utility derived from the cumulative return.

The choice of the risk tolerance parameter γ is reflected in the average hedge ratios. For $\gamma = -10$, hedge ratios are higher than the hedge ratios computed for $\gamma = 0$. As was seen in section 5, over this period hedging nothing led to a larger return than hedging everything. This fact corresponds to the observation that in table 6 low average hedge ratios tend to lead to higher cumulative returns.

Apart from the average level of hedging, it is important that the models take the right decision at the right time. Compare e.g. the entries for $\gamma = -2$ for the GLL-SV and GLL-Student t models. In both cases $\overline{H} = 0.35$, but cumulative returns differ by 3.4%. As the Student t model assumes constant variance, the model is not able to adapt to periods of higher uncertainty by hedging more. The Stochastic Volatility model can take the changing variance into account in the hedging decision, and therefore manages to escape some of the days with negative returns.

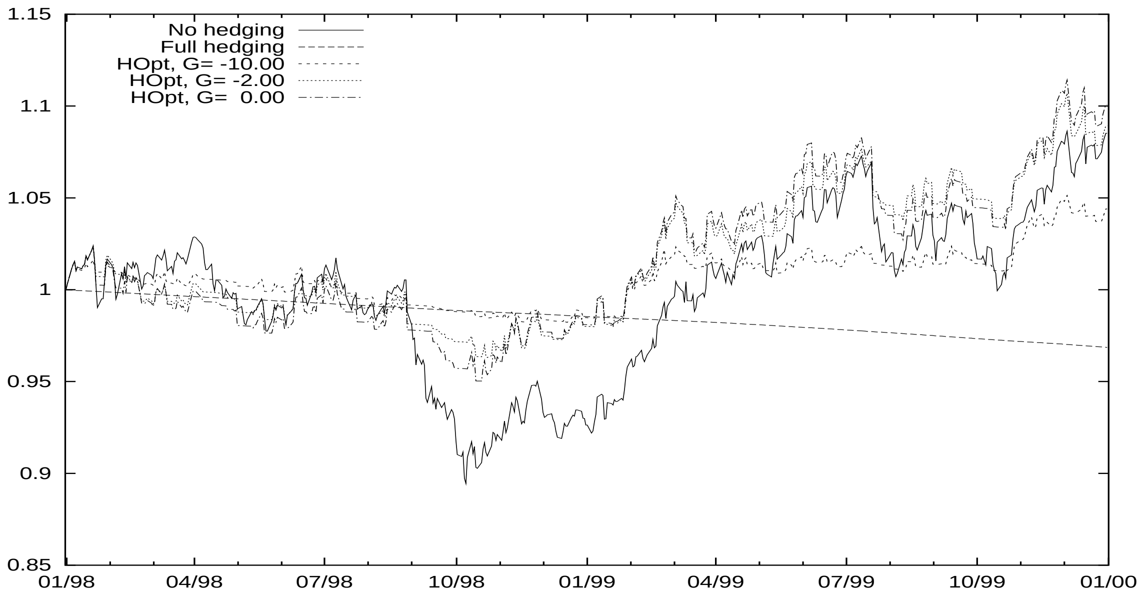

Figure 7: Returns from optimal hedging, for the GLL-Stochastic Volatility model

In figure 7 the evolution of the cumulative returns for the GLL-SV models are plotted. The straight line is the base case when all risk is hedged, the continuous line ending at 1.085 is the evolution of the exchange rate itself, obtained when no hedging is done. The dotted line, ending around 1.05, represents the results for a risk averse investor choosing $\gamma = -10$. It is seen that in the first year the cumulative return evolves almost along the full-hedging return line. The other γ's lead to losses on some days in the first period, but in August 1998 all strategies end up at the same level again. Until February 1999 the difference between the returns is clear: All the optimal strategies indicate to hedge almost all days in this period, and so avoid the losses on the unhedged position in this period. After two years, both the $\gamma = 0$ and the $\gamma = -2$ investor made gains compared to the exchange rate itself, and experienced less variation in their outcomes. The risk averse investor ($\gamma = -10$) hardly experienced any losses larger than the interest rate differential. With his downward risk covered, he still manages to pick up part of the rise in the exchange rate.

5.7 Alternative hedging strategies

In table 7 we have listed some results for alternative hedging strategies, notably the Value-at-Risk and Sharpe ratio strategies. For ease of comparison, the table replicates the results from table 6 for the case $\gamma = 0$, the risk tolerant investor optimizing his (log-)utility function.

Table 7: Results for alternative hedging strategies

		$\gamma = 0$		VaR		Sh	
WN	C $(\sigma(C))$	5.82	(11.22)	6.79	(11.56)	9.51	(11.82)
	$U(C) \times 100$	5.66		6.57		9.09	
LL	C $(\sigma(C))$	6.93	(9.81)	0.55	(10.01)	7.35	(9.85)
	$U(C) \times 100$	6.70		0.54		7.10	
GLL	C $(\sigma(C))$	9.37	(10.20)	15.32	(11.02)	11.35	(10.38)
	$U(C) \times 100$	8.95		14.25		10.75	
GLL-GARCH	C $(\sigma(C))$	9.43	(9.73)	6.83	(10.35)	12.32	(9.83)
	$U(C) \times 100$	9.01		6.61		11.62	
GLL-SV	C $(\sigma(C))$	10.08	(10.11)	1.18	(11.74)	4.82	(10.90)
	$U(C) \times 100$	9.60		1.17		4.71	
GLL-Student t	C $(\sigma(C))$	8.03	(10.79)	8.14	(11.47)	10.36	(10.86)
	$U(C) \times 100$	7.72		7.82		9.86	
GLL-Student t-GARCH	C $(\sigma(C))$	7.55	(10.32)	7.78	(10.56)	11.07	(10.45)
	$U(C) \times 100$	7.27		7.49		10.50	

Indicated in the table are the average hedge ratio over the period 1/1/1998 until 31/1/1999, for different hedging strategies. The first column copies the optimal hedging result for $\gamma = 0$ from table 6. Second and third columns describe results for the strategies base on the VaR and Sharpe ratio. The cumulative return C is given in the second row together with the standard deviation of observed returns, $\sigma(C)$. $U(C)$ indicates the utility derived from the cumulative return.

The evidence is mixed. Based on the final results, one would say that the VaR- and Sharpe ratio-based hedging strategies perform very well for models WN, GLL-GARCH, GLL-Student t, GLL-Student t-GARCH and especially for the GLL model. However, if one takes the evolution of the cumulative return into account (see figure 8), it is found that the higher return comes at the cost of a higher uncertainty of the returns. This is also seen from the higher standard deviation of the returns in table 7. For reasons of comparability we have set the critical VaR-level such that the subsequent average hedge ratio is equal to the average hedge ratio in the $\gamma = 0$ case. From table 7 we see that cumulative returns vary much more among the models. Especially the GLL model gives a high cumulative return (15.3%). Note however that the standard deviation of the cumulative returns is higher for the VaR strategies than it is for the optimal strategies with $\gamma = 0$.

In table 8 we present the fraction of realized returns not exceeding the Value-at-Risk returns. Casual inspection shows that the realizations compare favourably with the a priori confidence levels α.[9]

[9] It is well known that testing whether realized portfolio return numbers are significantly differ from what VaR predicts is very difficult. See for example Jorion (1997) and Kupiec (1995).

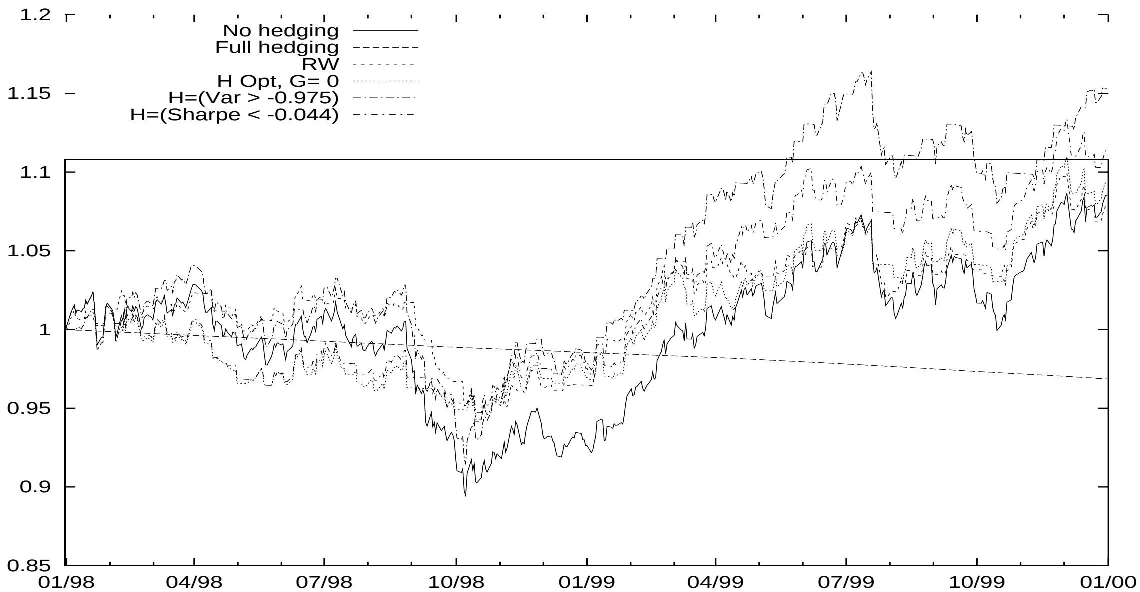

Figure 8: Returns from alternative hedging, for the GLL model

Table 8: Coverage probabilities of VaR

Model	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$
WN	0.952	0.967	0.985
LL	0.948	0.969	0.987
GLL	0.952	0.967	0.985
GLL-GARCH	0.927	0.960	0.973
GLL-SV	0.914	0.954	0.985
GLL-Student t	0.948	0.969	0.989
GLL-Student t-GARCH	0.925	0.964	0.987

For the strategy based on the Sharpe ratio we also find that the cumulative returns vary strongly. Again the standard deviation is higher than for the $\gamma = 0$ case.

In general we see that currency hedge strategies based on statistical models may add value with respect to naive strategies like the Random Walk, the fully hedged case, or the unhedged case. Results do vary among statistical models, and among the objective function the investor chooses.

6 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 put on 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 DMark/US Dollar exchange rate.

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.

We find that, even though the numerical value of the estimated and predicted mean returns is not large, it cannot be disregarded in practical hedging strategies. Strikingly, we find that models which are comparable when we use them to describe exchange rate returns, differ considerably when we evaluate them in the hedging strategies. From inspection of our results a general conclusion is: Modelling both time-varying means and time-varying variances leads to improved risk and return profiles within a currency overlay environment.

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. 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 variables. 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 & Dahlquist (1999) and Evans & Lewis (1995).

Thirdly, the final hedging results depend strongly on a few days with large absolute returns. The consequences of decision making may be investigated over longer periods, or comparing subperiods. 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. 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).

Finally, it is of interest to extend the decision framework and allow for options as an instrument in the decision process. Further, one may allow for the hedging parameter to be outside the unit interval. Hence, managers may use the currency position as an investment in its own right, going short or long.

A Gibbs sampling with data augmentation

To construct the sample from the posterior density in models **E**-**G**, direct application of the Metropolis-Hastings sampler is not possible as the likelihood function is not available in closed form.

In this appendix, a Gibbs method with data augmentation is described which attains conditional normality of the state space models. Given the conditional normality, the state space model can be handled using the standard Kalman filter and simulation equations (see Harvey 1989, de Jong & Shephard 1995), which simplifies the analysis.

The full set of equations for model **E**, the GLL-Stochastic Volatility model, reads

$$\begin{aligned} y_t &= \mu_t + \epsilon_t & \epsilon_t &\sim \mathcal{N}(0, \sigma^2_{\epsilon,t}) \\ \mu_t &= \rho\mu_{t-1} + \eta_t & \eta_t &\sim \mathcal{N}(0, \sigma^2_\eta) \\ \ln \sigma^2_{\epsilon,t} \equiv h_t &= \mu_h + \phi(h_{t-1} - \mu_h) + \xi_t & \xi_t &\sim \mathcal{N}(0, \sigma^2_\xi) \end{aligned}$$

for $t = 1, .., T$. Conditional on the values of the log-variance process h_t, the model is Gaussian. Following Kim et al. (1998), a linear process for the variance can be constructed by writing

$$y^*_t = \ln(y_t - \mu_t)^2 = h_t + z_t \qquad z_t = \ln(\epsilon^2_t) \tag{16}$$

$$h_t = \mu_h + \phi(h_{t-1} - \mu_h) + \xi_t \tag{17}$$

The non-normal disturbance process z_t can be approximated by a mixture of normal densities. This way, conditional on an index s_t indicating the element of the mixture, full conditional normality is regained and the Kalman equations can again be used. A more elaborate exposition is found in Kim et al. (1998) or Chib, Nardari & Shephard (1998).

For models **F** and **G**, the problem lies in the Student t density. Write the model **G**, the GLL-Student t-GARCH model as

$$\begin{aligned} y_t &= \mu_t + \epsilon_t & \epsilon_t &\sim t(0, \frac{\nu - 2}{\nu} h_t \sigma^2_\epsilon, 1, \nu) \\ \mu_t &= \rho\mu_{t-1} + \eta_t & \eta_t &\sim \mathcal{N}(0, \sigma^2_\eta) \\ h_t &= \delta h_{t-1} + \omega + \alpha\epsilon^2_{t-1}/\sigma^2_\epsilon & \omega &\equiv 1 - \delta - \alpha \end{aligned}$$

for $t = 1, .., T$. Note that $\mathsf{Var}(\epsilon_t) = h_t$ and that the unconditional variance of ϵ is $\mathsf{E}(h_t) = 1$.

Instead of using the Student t density as it is, we write the density of ϵ_t as a normal density with a random variance,

$$\epsilon_t | \nu \sim \mathcal{N}(0, h_t z_t \sigma^2_\epsilon) \tag{18}$$

with a prior density for $z_t|\nu$ given as

$$\pi(z_t|\nu) \sim \text{IG}(\alpha = \frac{\nu}{2}, \beta = \frac{2}{\nu - 2}) \tag{19}$$

It is straightforward to derive that, after integrating out the mixing parameter z_t indeed the Student t density with ν degrees of freedom for ϵ_t results.

The full conditional posterior densities which are needed in the Gibbs sampling algorithm are given without derivation in table 9. For the GARCH parameter $\sigma^2_\epsilon, \delta, \alpha$ and for the degrees of freedom parameters ν no closed form expression of the conditional density is available.

Table 9: Conditional posterior densities

Parameter	In model	Full conditional distribution
μ	**E**, **F**, **G**	Use the simulation smoother, see de Jong & Shephard (1995)
ρ	**E**, **F**, **G**	$\mathcal{N}\left(\frac{\hat{\rho}\sigma_\rho^2+\mu_\rho\hat{\sigma}_\rho^2}{\sigma_\rho^2+\hat{\sigma}_\rho^2}, \frac{\hat{\sigma}_\rho^2\sigma_\rho^2}{\sigma_\rho^2+\hat{\sigma}_\rho^2}\right)$ with $\hat{\rho}$ and $\hat{\sigma}_\rho^2$ the least squares estimate of ρ with corresponding variance
σ_η^2	**E**, **F**, **G**	$\text{IG}\left(\alpha = \frac{T}{2} + \alpha_\eta, \beta = 2 \Big/ \left(\sum(\mu_t - \rho\mu_{t-1})^2 + \frac{2}{\beta_\eta}\right)\right)$
μ_h	**E**	$\mathcal{N}(\hat{\mu}, \hat{\sigma}_\mu^2)$ with $\hat{\mu} = \hat{\sigma}_\mu^2 \left(\frac{1-\phi^2}{\sigma_\xi^2}h_0 + \frac{1-\phi}{\sigma_\xi^2}\sum(h_t - \phi h_{t-1})\right)$ and $\hat{\sigma}_\mu^2 = \sigma_\xi^2 / \left((T-1)(1-\phi)^2 + (1-\phi^2)\right)$
ϕ	**E**	$\mathcal{N}\left(\frac{\hat{\phi}\sigma_\phi^2+\mu_\phi\hat{\sigma}_\phi^2}{\sigma_\phi^2+\hat{\sigma}_\phi^2}, \frac{\hat{\sigma}_\phi^2\sigma_\phi^2}{\sigma_\phi^2+\hat{\sigma}_\phi^2}\right)$ with $\hat{\phi}$ and $\hat{\sigma}_\phi^2$ the least squares estimate of ϕ with corresponding variance
σ_ξ^2	**E**	$\text{IG}\left(\alpha = \frac{T}{2} + \alpha_\xi, \beta = 2 \Big/ \left(\sum\left((h_t - \mu_h) - \phi(h_{t-1} - \mu_h)\right)^2 + \frac{2}{\beta_\xi}\right)\right)$
s_t	**E**	The indices into the mixture in the distribution of $\ln \epsilon_t^2$ are discretely distributed
$\sigma_\epsilon^2, \delta, \alpha$	**G**	Use MH sampling. The conditional posterior is proportional to the likelihood from the Kalman filter equations and the prior.
z_t	**F**, **G**	$\text{IG}\left(\alpha = \frac{\nu+1}{2}, \beta = \frac{2}{(\nu-2)+(y_t-\mu_t)^2 \big/ (\sigma_\epsilon^2 h_t)}\right)$
ν	**F**, **G**	The posterior is not of a known form. It is proportional to $\prod_t \text{IG}\left(z_t; \alpha = \frac{\nu}{2}, \beta = \frac{2}{\nu-2}\right) \times \text{Cauchy}(\nu; \mu = 0; s = 1)$. Apply a MH step to sample a new value of ν.

Therefore, we use in these steps a Metropolis-within-Gibbs sampler (see Zeger & Karim 1991). Note that the priors in table 1 in section 4.1 have been applied.

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.

Barberis, N. (2000), 'Investing for the long run when returns are predictable', *Journal of Finance* **55**, 225–264.

Bauwens, L. & Lubrano, M. (1998), 'Bayesian inference on GARCH models using the Gibbs sampler', *The Econometrics Journal* pp. C23–C46.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327.

Carter, C. K. & Kohn, R. (1994), 'On Gibbs sampling for state space models', *Biometrika* **81**(3), 541–553.

Carter, C. K. & Kohn, R. (1996), 'Markov chain Monte Carlo in conditionally Gaussian state space models', *Biometrika* **83**(3), 589–601.

Chib, S. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association* **90**(432), 1313–1321.

Chib, S. & Greenberg, E. (1995), 'Understanding the Metropolis-Hastings algorithm', *The American Statistician* **49**(4), 327–335.

Chib, S., Nardari, F. & Shephard, N. (1998), Markov chain Monte Carlo methods for generalized stochastic volatility models. Unpublished manuscript.

de Jong, P. & Shephard, N. (1995), 'The simulation smoother for time series models', *Biometrika* **82**(2), 339–350.

Doornik, J. A. (1999), *Object-Oriented Matrix Programming using Ox*, 3rd edn, London: Timberlake Consultants Ltd. See `http://www.nuff.ox.ac.uk/Users/Doornik`.

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 [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**(3), 709–742.

Gelfand, A. E. & Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**(410), 398–409.

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 & A. F. M. Smith, eds, 'Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting', Oxford: Clarendon Press, pp. 169–193.

Geweke, J. & Zhou, G. (1996), 'Measuring the pricing error of the arbitrage pricing theory', *Review of Financial Studies* **9**(2), 557–588.

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

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

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

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. (2000), 'Testing for integration using evolving trend and seasonal models: A Bayesian approach', *Forthcoming in Journal of Econometrics* .

Koopman, S. J., Shephard, N. & Doornik, J. A. (1999), 'Statistical algorithms for models in state space using SsfPack 2.2', *Econometrics Journal* **2**, 107–160.

Kupiec, P. (1995), 'Techniques for verifying the accuracy of risk measurement models', *Journal of Derivatives* **2**, 73–84.

LeBaron, B. (1999), 'Technical trading rule profitability and foreign exchange intervention', *Journal of International Economics* **49**, 125–143.

Mark, N. C. (1995), 'Exchange rates and fundamental: Evidence on long-horizon predictability', *American Economic Review* **85**(1), 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.

Newey, W. K. & West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**, 703–708.

Poirier, D. J. (1995), *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Cambridge [etc.].

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**(1), 3–24.

Taylor, S. J. (1994), 'Modeling stochastic volatility: A review and comparative study', *Mathematical Finance* **4**(2), 183–204.

West, M., Harrison, P. J. & Migon, H. S. (1985), 'Dynamic generalized linear models and Bayesian forecasting', *Journal of the American Statistical Association* **80**(389), 73–98.

Zeger, S. L. & Karim, M. R. (1991), 'Generalized linear models with random effects: A Gibbs sampling approach', *Journal of the American Statistical Association* **86**, 79–86.