JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **15**: 671–696 (2000)

DAILY EXCHANGE RATE BEHAVIOUR AND HEDGING OF CURRENCY RISK

CHARLES S. BOS,[a]* RONALD J. MAHIEU[b] and HERMAN K. VAN DIJK[a]
[a]*Econometrics Institute and Tinbergen, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR, Rotterdam, The Netherlands*
[b]*Rotterdam School of Management, Erasmus University Rotterdam, The Netherlands*

SUMMARY

We construct models which enable a decision maker to analyse 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 and utilities from several alternative hedge strategies. The results indicate that modelling time-varying features of exchange rate returns may lead to improved hedge behaviour within currency overlay management. Copyright © 2000 John Wiley & Sons, Ltd.

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. For example, when a corporation sells its goods abroad it incurs foreign exchange rate exposure at the time it wants to repatriate the proceeds of the sales. Another large group of companies with foreign currency exposure are internationally operating investors, like banks, pension funds, and insurance companies. The currency exposures arise from the investment strategies that these institutions follow. For example, when a 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. The approach where currency hedging decisions are made independently from underlying investment decisions, is called 'currency overlay management' in the finance industry. Note that this approach may lead to suboptimal decisions from a fund's perspective as a currency overlay strategy ignores the diversifying characteristics that currencies may have. Continuing the example, when the investor perceives the risk of the Japanese yen depreciating too large, he may decrease his holdings of Japanese stocks. However, by applying currency overlay management the investor tries to manage his Japanese yen currency exposure irrespective of the amount of wealth invested in Japanese stocks. A major reason for investors to separate the currency and portfolio decisions is to obtain increased transparency of the investment strategy.

*Correspondence to: Charles Bos, Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR, Rotterdam, The Netherlands. e-mail: cbos@few.eur.nl

Received 30 September 1999
Revised 31 August 2000

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 options. In this paper we focus on hedging with forward contracts only.

To illustrate the practical importance of currency overlay management one may distinguish two special cases. First, the decision maker does not hedge at all. The return on the currency overlay strategy is then equal to the return on the exchange rate. Second, the decision maker hedges the currency risk completely. Now, the return is equal to the difference between the interest rates of the home country and that of the foreign country. A practical example is the case of a German firm with US investments. In the period 1998–1999, the cumulative return on the Dmark/US dollar (DEM/USD) exchange rate was more than 8%, while the cumulative difference between the two interest rates was less than minus 3%. Thus, the decision to hedge or not to hedge relates to a difference in cumulative return in two years of approximately 11%. Since multinational corporations and large institutional investors deal with substantial foreign currency exposures that may involve hundreds of millions of dollars, the specification of an effective strategy for foreign exchange rate management is an important topic.

In this paper we analyse the risk and return properties of currency overlay strategies using time series models that describe prominent features of daily exchange rate data. Our contribution focuses on three issues. First, we introduce a class of models which describes some major features of the data: local trends in the level or varying means in the return, time-varying volatility in the second moment of the return, and leptokurtosis of the returns. We integrate models for the analysis of varying means, varying variances, and heavy-tailed distributions. Then we obtain a flexible general framework which enables us to study the effects and relevance of different model specifications for hedging decisions. The topics that we investigate in this respect are unit roots versus persistent but stationary behaviour in expected returns, heavy-tailed distributions, and different ways to model conditional volatility. Second, for inference and decision analysis we make extensive use of Bayesian methods based on Markov chain Monte Carlo (MCMC) simulation. Third, in the decision analysis we investigate the payoff and utility from an optimal strategy using alternative models and corresponding results 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 and Migon (1985). In Section 4 we discuss our Bayesian methods, see e.g. Smith and Roberts (1993) and Chib and Greenberg (1995). In the recent literature these methods have been successfully applied for studying separately the pattern of varying means (see

Carter and Kohn, 1994; Koop and van Dijk, 2000) and the pattern of varying volatilities (see Kim, Shephard and Chib, 1998). Results are presented in Section 5 using the DEM/USD 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 the Appendix.

2. CURRENCY HEDGING

As noted in the introduction, we concentrate on effective 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 $s_{t+1} \equiv \ln(S_{t+1}/S_t)$, with S_t the exchange rate itself. Let $F_{t,\tau}$ be the current value of a forward contract with maturity date τ. By covered interest rate parity it is equal to

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

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

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

$$r_{t+1} \equiv \ln\left(\frac{(1-H_t)S_{t+1} + H_t F_t}{S_t}\right)$$

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

$$\exp(r_{t+1}) = (1 - H_t)\exp(s_{t+1}) + H_t \exp\left(r^h_t - r^f_t\right) \tag{1}$$

It is seen that the exponent of the return is a weighted average of the exponents of the exchange rate return s_{t+1} and the difference between the home and foreign risk-free interest rates. Note that when we set the hedge ratio H_t to zero, the return on the currency overlay part is equal to the return on the exchange rate only. On the other hand, if we set the hedge ratio to one, only

[1] See Solnik (2000) for a comprehensive review of covered interest rate parity.
[2] We have also checked our results with arithmetic returns. The results changed somewhat. We are indebted to a referee for bringing up this point.

the interest rate differential has an impact, whereas changes in the currency do not affect the return on the currency overlay.[3]

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

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

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

$$\max_{o \leq H_t \leq 1} \mathsf{E}_{s_{t+1|t}} U(W_{t+1}) = \max_{0 \leq H_t \leq 1} \mathsf{E}_{s_{t+1|t}} \left[\frac{\left(\exp\left(r_{t+1}(s_{t+1}, H_t, r_t^h, r_t^f)\right)\right)^\gamma - 1}{\gamma} \right] \tag{2}$$

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

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

$$\int_{-\mathrm{VaR}}^{\infty} f(r_{t+1|t})\, \mathrm{d}r_{t+1} = 1 - \alpha \tag{3}$$

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

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

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

$$\mathrm{Sh} = \frac{\mathsf{E}_{s_{t+1|t}}(r_{t+1})}{\sqrt{\mathsf{Var}_{s_{t+1|t}}(r_{t+1})}} \tag{4}$$

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

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

$$s_t = \mu_t + \epsilon_t \qquad \epsilon_t \sim \text{i.i.d.}(0, \sigma^2_{\epsilon,t}) \tag{5}$$

$$\mu_t = \rho\mu_{t-1} + \eta_t \qquad \eta_t \sim \mathcal{N}(0, \sigma^2_\eta) \qquad t = 1, \ldots, T \tag{6}$$

The unobserved mean component μ_t is an autoregressive process with disturbances η_t and autoregressive parameter ρ. This model, which we label the Generalized Local Level (GLL) model, is supposed to pick up the periods of rising or falling exchange rate levels.[4] 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 model is known as the Local Level (LL) model (see Harvey, 1989, p. 45). Given $\sigma^2_\eta > 0$, the LL model implies that the logarithm of the exchange rates follows an I(2) process. We expect that, when estimating this I(2) model on our data, the variance of η_t is small compared to the variance of ϵ_t, such that the I(1) behaviour of ln S_t overwhelms the I(2) effects. One can also take the limit case $\sigma^2_\eta = 0$, $\rho = 1$, which is White Noise (WN) around a fixed mean μ. Though extremely simple, it is a basic model in many financial market models.[5]

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

[5] This white noise model is the only model we consider with non-zero unconditional expectation for s_t. In Table II, Section 5.2, on the posterior density of the parameters, it is found that there is no strong evidence for a non-zero mean.

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

$$\epsilon_{t|t-1} \sim \mathcal{N}(0, \sigma^2_{\epsilon,t}) \tag{7}$$

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

$$\sigma^2_{\epsilon,t} = \sigma^2_\epsilon \tag{8}$$

in which case a standard state space model results. More flexibility is obtained when a GARCH disturbance process is allowed for. The variance $\sigma^2_{\epsilon,t}$ of the observation equation (5) varies over time according to

$$\begin{aligned}
\sigma^2_{\epsilon,t} &= \sigma^2_\epsilon h_t \\
h_t &= \delta h_{t-1} + \omega + \alpha \epsilon^2_{t-1}/\sigma^2_\epsilon \\
\delta \geq 0, &\quad \alpha \geq 0, \quad \delta + \alpha < 1, \quad \omega \equiv 1 - \delta - \alpha
\end{aligned} \tag{9}$$

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

A second family of disturbance processes for ϵ_t with time-varying variance follows from the Stochastic Volatility (SV) process (see Jacquier, Polson and Rossi, 1994). The variance of the disturbances in the observation equation evolves according to

$$\begin{aligned}
\sigma^2_{\epsilon,t} &= \exp(h_t) \\
h_t &= \mu_h + \phi(h_{t-1} - \mu_h) + \xi_t \qquad 0 \leq \phi < 1 \\
\xi_t &\sim \mathcal{N}(0, \sigma^2_\xi)
\end{aligned} \tag{10}$$

A third feature of financial time series is that the histograms of the series exhibit heavier tails than the normal density, even after correcting for the time-varying volatility. To model this, we replace equation (7) by

$$\epsilon_{t|t-1} \sim t\Big(0, (\nu - 2)\sigma^2_{\epsilon,t}, 1, \nu\Big) \qquad \nu > 2 \tag{11}$$

where t indicates the Student-t density, with expectation 0, variance $\sigma^2_{\epsilon,t}$ and ν degrees of freedom.

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

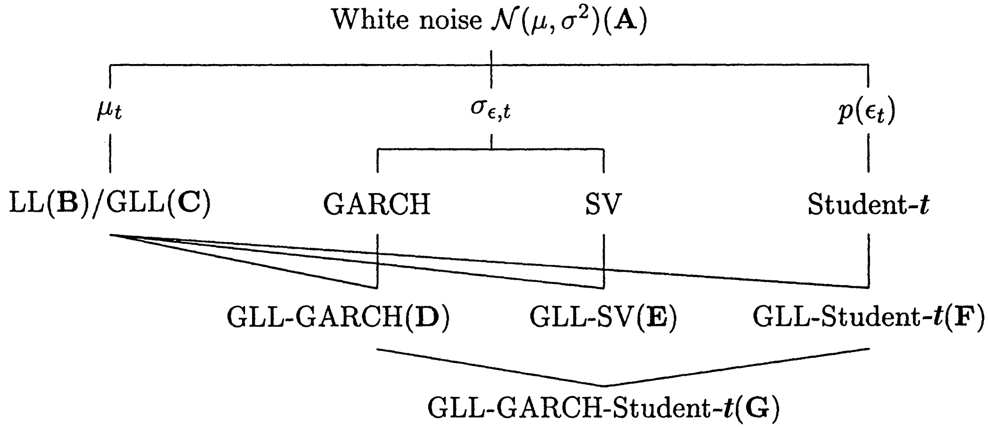

Figure 1. Hierarchy of models

most general model in the fifth line results. The models are indicated by the letters **A–G** in the figure and in text and tables in subsequent sections.

4. BAYESIAN INFERENCE AND DECISION ANALYSIS

4.1 Prior Structure

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

The autoregressive parameter ρ of the unobserved mean process μ_t is crucial in the analysis. It

Table I. Description of priors used

Parameter	Prior	Hyper-parameters	Used in model
μ	$\mathcal{N}(\mu_0, \sigma_0^2)$	$\mu_0 = 0$, $\sigma_0 = 0.02$	**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	$\mathcal{N}(\mu_\rho, \sigma_\rho^2)$	$\mu_\rho = 0.8$, $\sigma_\rho = 0.2$	**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	$\mathcal{N}(\mu_\phi, \sigma_\phi^2)$	$\mu_\phi = 0.5$, $\sigma_\phi = 0.3$	**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**

governs the amount of predictability in the series (together with the ratio of the variances in observation and transition equations (5) and (6)). Given the fact that trends in exchange rates may last for several months, we deem a large value of ρ in the unit interval *a priori* more plausible than a small value. As an intermediate position between a strongly informative and an uninformative prior, we choose a normal prior density with mean 0.8 and a rather large standard deviation of 0.2.[6] 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 priors for the variance parameters are all inverted gamma (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 σ^2_ϵ, σ^2_η and σ^2_ξ respectively. In Bauwens and 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 and α are bounded by the stationarity condition to be positive and smaller than 1 in sum. On the stationarity region, we assume a uniform prior.

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 density of the parameters, $p(\theta\,|\,\text{data})$, is obtained by multiplying the corresponding prior density with the likelihood function. Though the shape of this posterior might be highly non-normal, a general adaptive independent Metropolis–Hastings (MH) sampler (see Carter and Kohn, 1996; Chib and Greenberg, 1995; Koop and 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-GARCH-Student-*t* 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 the Appendix 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 for model M is defined as

$$m(M) = \int L(\text{data}; \theta, M)\pi(\theta|M)\text{d}\theta \tag{12}$$

and may be computed using Bayes' rule as

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

$$m(M) = \frac{L(\text{data}; \theta, M)\pi(\theta|M)}{p(\theta|\text{data}, M)} \tag{13}$$

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

In 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 directly 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 and 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 densities (see the Appendix). Chib (1995) describes a procedure to calculate the marginal likelihood in this case. In Section 5.3 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 the Appendix. In cases where a Metropolis–Hastings step was applied within a Gibbs chain, 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 $p(s_{t+1|t})$ of tomorrow's returns on the exchange rate s_{t+1}, given all available information. The conditional density $p(s_{t+1|t}|\theta)$, given the vector of parameters θ, is easily derived. The unconditional predictive density follows by marginalization with respect to θ,

$$p(s_{t+1|t}) = \int_{\theta\in\Theta} p(s_{t+1|t}|\theta)p(\theta|s_t, s_{t-1}, \ldots, s_1)\mathrm{d}\theta \tag{14}$$

see e.g. Geweke (1989) and Barberis (2000). Marginalization is done with respect to the posterior density of $\theta|s_t, s_{t-1}, \ldots, s_1$. On-line modelling and prediction requires that one re-estimates the posterior of the parameters for every day in the evaluation period. However, for computational reasons we refrain from doing this and use only N drawings $\theta^{(1)}, \ldots, \theta^{(N)}$ from the posterior of $\theta|s_T, \ldots, s_1$, with $s_T, \ldots, s_1$ the observations from the estimation sample ($T < t$). When the estimation sample is large compared to the evaluation sample, this approximation gives, under standard regularity conditions, a sufficient level of accuracy. The integral in equation (14) is approximated using

$$p(s_{t+1|t}) \approx \frac{1}{N}\sum_{i=1}^{N} p(s_{t+1|t}, \theta^{(i)}) \tag{15}$$

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

$$\begin{aligned} H_t &= \underset{H_t}{\arg\max}\ \mathsf{E}_{s_{t+1|t}} U(W_{t+1}) \\ &= \underset{H_t}{\arg\max} \int_{s_{t+1}} \frac{\exp\big(r_{t+1}(H_t, s_{t+1}, r^h, r^f)\big)^{\gamma} - 1}{\gamma} p(s_{t+1|t})\, \mathrm{d}s_{t+1} \end{aligned} \tag{16}$$

(see equation (2)). Optimal hedge ratios are computed using a grid search 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 5% VaR is according to the model at hand. The investor should decide if the VaR is acceptable for him, or that he deems the risk too high. For reasons of comparison, we 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 1 January 1982 until 31 December 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.[7]

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 (in the lower panels) exhibit patterns frequently found in high-frequency financial return data. As for the returns, it is seen that there is no clear serial correlation pattern, corroborating the widely held view that financial return series are unpredictable. However, the local trends in the levels of the exchange rates may prove useful for practical currency overlay strategies. The phenomenon of local trends is, at a longer horizon, similar to the data feature of long swings in the dollar as observed by Engel and Hamilton (1990). We note that we use a state space model while these authors use a Markov switching process for describing exchange rate returns over longer periods.

The squared returns show a clear pattern. The slowly decaying autocorrelation has prompted many researchers to develop models for describing time varying volatilities.

[7] Source: DATASTREAM, series `DMARKER/USDOLLR`, `ECWGM1M`, `ECUSD1M` for the daily DEM/USD exchange rate and German and US 1-month Eurocurrency middle interest rates, respectively.

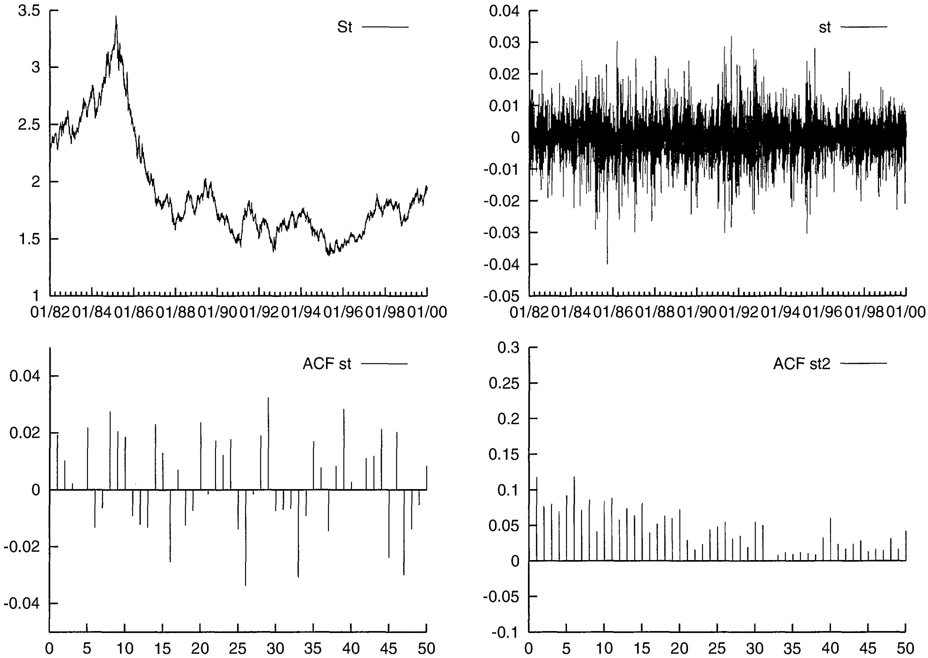

Figure 2. DEM/USD exchange rate, 1 January 1982 until 31 December 1999. Panels contain data in levels (top-left), in returns (top-right), and the autocorrelation function of the returns and the squared returns (bottom)

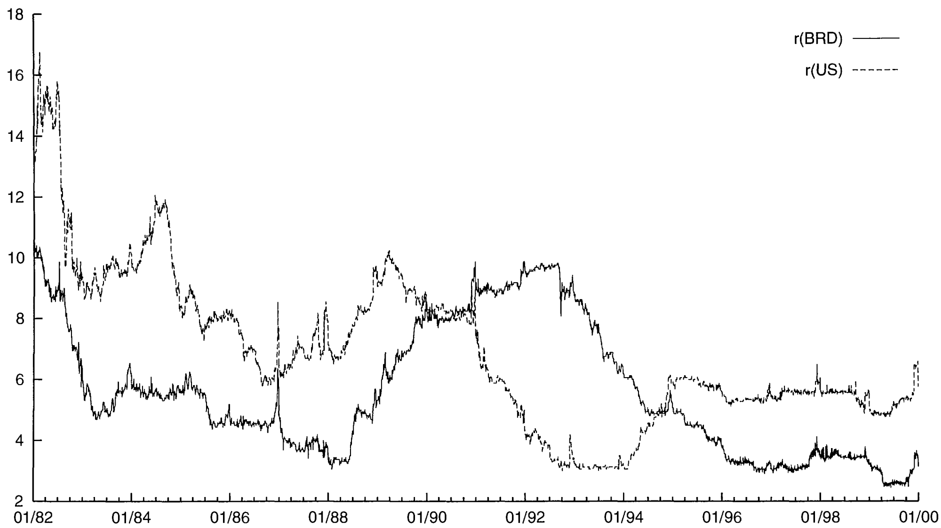

Figure 3. One-month Eurocurrency interest rates for the German Dmark and the US dollar, in yearly percentages

Figure 3 shows the time series of both US 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 and the difference between the two (which is used in the hedging decision) are very persistent. We note that in the hedging decision we transformed the series to daily interest rates.

5.2 Convergence of MCMC and Posterior Results

For models **A–D** the Metropolis–Hastings sampling algorithm was used.[8] After three initial runs of the MH sampler (with 500, 2000 and 10,000 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 200,000 drawings was accepted. From every 20 drawings, only 1 was saved, in order to lower correlation in the posterior sample. Acceptance rates were 98%, 93%, 67% and 61%, respectively. This corresponds to final sample sizes of 10,147, 10,682, 14,787 and 16,387.

The models with Student-*t* disturbances or Stochastic Volatility components did not allow for direct implementation of the MH sampler. The Gibbs sampler we used was run for a burn-in period of 50,000 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-Stochastic Volatility model; it is seen that only after about 30 drawings, correlation dies out.[9]

The correlation in the sample influences the amount of information available in the posterior. A measure of the effective size of the posterior is the relative numerical efficiency (RNE) see (Geweke, 1992). We calculated both the direct variance of the posterior, and compared it with a correlation-consistent estimate of the variance. Using the Newey–West variance estimator (Newey and West, 1987), adjusting for correlation with lags up to 4% of the size of the sample, we find values for the RNE 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-GARCH-Student-*t* model, the sample from the Markov chain of 10,000 dependent drawings roughly corresponds to a sample of 1000 independent drawings from the posterior.

The main characteristics of the posteriors are summarized in Tables II and III. For each model and for each parameter, the mean, standard deviation (in parentheses), mode (on the second line) and the bounds of the 95% highest posterior density region (between square brackets) are reported.[10]

[8] 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 and Doornik, 1999) was used extensively.
[9] Figure 4, and also Table III and Figure 5, report results for σ_ϵ, the unconditional standard deviation of the disturbance process. In the GLL-SV model, this parameter is not used. For comparison, results for σ_ϵ are constructed from the sample of μ_H, ϕ and σ_ξ.
[10] All 95% HPD regions were continuous.

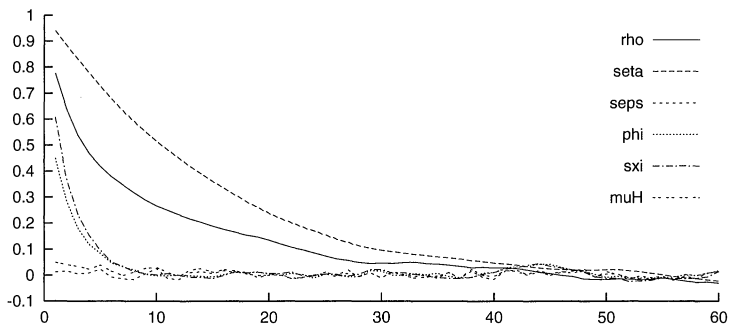

Figure 4. Autocorrelation function of draws from the parameters of the GLL-Stochastic Volatility model

Table II. Posterior results

Parameter	WN		LL		GLL	
$\mu \times 100$	−0.40	(0.92)				
	−0.39	[−2.26,1.35]				
ρ	1		1		0.69	(0.12)
					0.75	[0.44,0.90]
$\sigma_\eta \times 10$	0		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]
S/N × 100	0		∞		2.36	

Table III. Posterior results

Parameter	GLL-GARCH		GLL-SV		GLL-Student-t		GLL-GARCH-Student-t	
ρ	0.76	(0.10)	0.68	(0.14)	0.68	(0.14)	0.79	(0.11)
	0.83	[0.56,0.92]	0.77	[0.40,0.91]	0.75	[0.40,0.91]	0.86	[0.56,0.95]
$\sigma_\eta \times 10$	0.69	(0.20)	0.66	(0.21)	0.60	(0.16)	0.54	(0.12)
	0.60	[0.36,1.11]	0.54	[0.32,1.05]	0.53	[0.32,0.90]	0.49	[0.33,0.79]
σ_ϵ	0.66	(0.03)	0.68	(0.02)	0.67	(0.01)	0.78	(0.06)
	0.65	[0.61,0.71]	0.67	[0.63,0.72]	0.67	[0.66,0.69]	0.76	[0.67,0.90]
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]					0.64	[0.52,0.80]
ν					4.54	(0.22)	4.84	(0.31)
					4.48	[4.11,4.99]	4.82	[4.24,5.47]
μ_h			−1.06	(0.06)				
			−1.07	[−1.19,−0.93]				
ϕ			0.92	(0.02)				
			0.93	[0.89,0.95]				
σ_ξ			0.28	(0.03)				
			0.28	[0.23,0.34]				
S/N × 100	2.92		2.06		1.66		1.46	

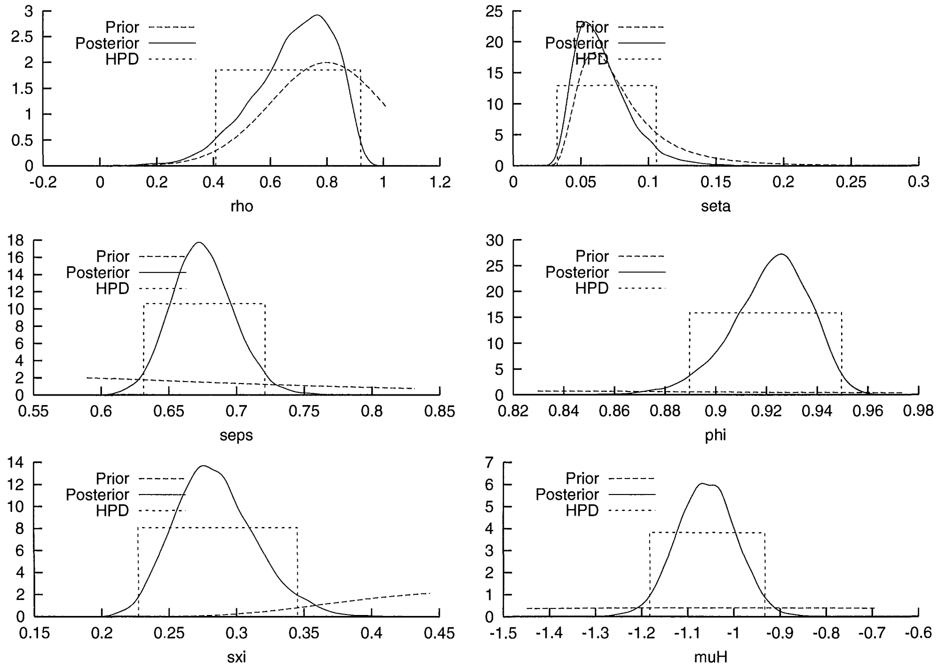

Figure 5. Prior, posterior and HPD region of the parameters in the GLL-Stochastic Volatility model

The last row of the tables indicates the signal-to-noise ratio (S/N), calculated as the ratio between the unconditional variance of the signal μ_t and the noise ϵ_t.[11]

The posteriors of the two parameters of the White Noise model were very tight, with the mean and standard deviation centred 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 0.024. The standard deviation of the observation disturbance, σ_ϵ, is rather larger at 0.67. Note that the variance of the signal μ_t is 0 for the WN model, and infinity for the I(1) process in the LL model.

More interesting are the posteriors for the GLL model. The density 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 densities. The signal-to-noise ratio is low at 0.0236. This corresponds with the findings of very little autocorrelation in the series, as seen from Figure 2.

The skewness of the posterior of ρ and σ_η can also be observed for other models. In Figure 5 the marginal posteriors of the parameters of the GLL-Stochastic Volatility model are plotted, together with the priors and the 95% HPD regions. Apart from the skewness of parameters ρ

[11] Note that Harvey (1989, p. 68) uses the definition of the signal-to-noise ratio $q = \sigma_\eta^2/\sigma_\epsilon^2$. This definition is commonly used with non-stationary models, when the variance of the signal μ_t is infinite. For our models, this statistic (times 100) takes the values 0.00, 0.13, 1.17, 1.22, 1.07, 0.74 and 0.51, respectively.

and σ_η, it is seen that the posteriors are somewhat more concentrated than the priors. The HPD region for the parameter ρ is wide, especially 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, for all the models where ρ is not fixed.

The contrast between the posterior of ρ and of the GARCH parameters (both in the GLL-GARCH and the GLL-GARCH-Student-*t* model) is large. Both δ and α are estimated quite precisely, with tight and almost symmetric posteriors densities. A similar effect is found for the parameters μ_H, ϕ and σ_ξ in the GLL-SV model, which are also empirically well identified. Including the Student-*t* disturbances in the GARCH model does not alter the posterior of the GARCH parameters δ, α greatly. Only the standard deviations σ_η and σ_ϵ change, as the Student-*t* disturbance takes up part of the variance. The resulting change in the S/N ratio is interesting: Due to the heavy tails of the Student-*t* density in the GLL-GARCH-Student-*t* model, the S/N ratio is only 0.0146, which is small compared to the value of 0.0292 for the GLL-GARCH model. A similarly small value of the S/N ratio is found for the GLL-Student-*t* model.

5.3 Marginal Likelihood

The marginal (log)likelihood has been calculated for each of the models (see Table IV). The kernel method is only used 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 marginal loglikelihoods indicate that the data provide 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 and the WN model. The modelling steps on the varying variance structure (allowing for GARCH or SV in the GLL model) lead to a substantial improvement in the marginal likelihood over the more basic WN or GLL models. The fixed variance Student-*t* and the GARCH extensions result in an improvement of the loglikelihood score of 144 and 163 points, respectively. Better is the combination of the two, with both varying variances and heavy-tailed disturbances. The GLL-SV model, which in flexibility is a close competitor to the GLL-GARCH-Student-*t* model, fits the data best, according to the marginal likelihood.

We summarize the findings of Sections 5.2 and 5.3 as follows:

(1) The parameter ρ has a 95% HPD interval that ranges from 0.4 until 0.95 over the different models. The values of $\rho = 0$ (White Noise) and $\rho = 1$ (LL) are outside the 95% HPD interval.

Table IV. Marginal loglikelihoods

			Relative to GLL	
Model	Kernel	Gibbs	Kernel	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		−4028.5		270.9
GLL-Student-*t*		−4155.6		143.8
GLL-GARCH-Student-*t*		−4043.4		256.0

(2) The parameters of the time-varying variance for the GARCH and SV models have highly concentrated posterior densities.
(3) The posterior of the degrees-of-freedom parameter ν indicates that the normal model is not supported by the data.
(4) The GLL-SV model has the highest marginal (log)likelihood. Further, the GLL-SV and the GLL-GARCH-Student-*t* models have a much better fit than the GLL-GARCH and GLL-Student-*t* models. These latter two models clearly outperform the GLL model, which in itself outperforms the LL and WN models. This ranking of models indicates that modelling time varying mean and variance, and fat tails contributes to a much better within-sample fit.

5.4 Predictive Density

The predictive density $p(s_{t+1|t})$ 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.

The case of the GLL-SV 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 $p(s_{t+1|t})$. 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

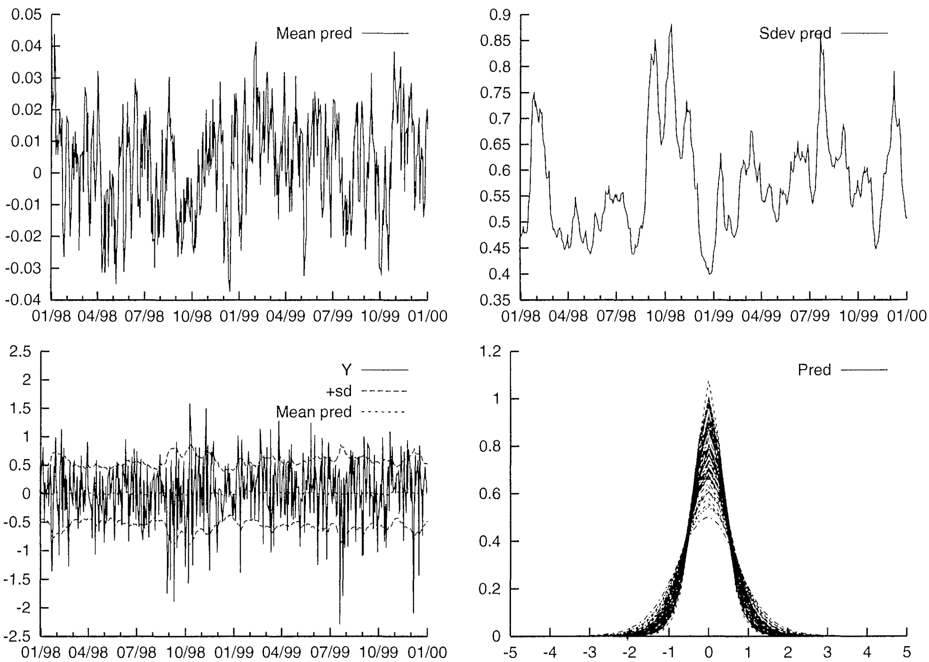

Figure 6. Predictive mean, standard deviation, together with observations and the full density over 1/1/98–31/12/99 (top-left to bottom-right) of the exchange rate return, for the GLL-SV model, in percentage changes

$\mathsf{E}(s_{t+1|t})$ is positive. On the axis the changes are indicated as daily percentages; though the changes from day to day are noticeable, they are of a size of as a maximum 0.04%.

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 becomes negative, the standard deviation jumps up. From that moment onwards the volatility remains high, until January 1999, where the stochastic volatility component indicates that the variance 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 $p(s_{t+1|t})$ for the days of the evaluation period. It is seen that the spread of the density changes considerably, the location hardly moves. For models **A–C** and **F**, the corresponding plot shows less variation over time as the variance is fixed. In the following we investigate whether the predictive densities provide information for constructing effective currency overlay strategies.

5.5 Benchmark Hedging Strategies

In the next three sections we present results on hedging strategies for currency risk management. When evaluating our results we focus on a number of criteria. First, we look at the risk and return characteristics of the different strategies and the different models. Second, we investigate the impact of the risk aversion parameter γ in case of hedging strategies based on the power utility function. Third, we investigate whether modelling is important or whether a naive benchmark strategy will do, and, when modelling pays off, which model one should choose. Fourth, we look into the issue of time variation in the hedge ratios H_t. In practice, a hedge strategy which has too much variation will not be attractive from a transaction cost perspective. Also operational risk may be too high for such strategies.

We start with results for three naive hedging strategies, which can be viewed as benchmark strategies against which we can set the results of the strategies based on time series models. The first strategy is the case for which the currency exposure is hedged at all times, i.e. $H_t = 1$ for all days in our evaluation period, 523 days within the period 1 January 1998 to 31 December 1999. Consequently, exchange rate risk is eliminated completely. The price that the investor pays for this strategy is the difference between home and foreign interest rates. The second strategy is the no hedge case, i.e. $H_t = 0$ for all days. The investor accepts all risks (and returns) on the foreign currency exposure. The third strategy is the random walk strategy, which sets the hedge ratio to one (zero) if the foreign currency depreciated (appreciated) in the previous period.

In Table V we list some characteristics of these benchmark strategies. The first column presents the average hedge ratio over the evaluation period. The second and third columns list the number of times the hedge ratios are either zero or one, respectively. Note that for the random walk strategy the hedge ratios are almost evenly distributed between $H = 0$ and $H = 1$. The fourth column gives the average absolute changes $\overline{|\Delta H|}$ in the hedge ratios for each strategy, which can be viewed as a measure for the variability of the hedge ratios. The fifth

Table V. Results of deterministic hedging strategies

Model	$\bar{H}$	$\#(H=0)$	$\#(H=1)$	$\overline{\lvert\Delta H\rvert}$	$\sum r_t$	Cumulative utility $\gamma=-10$	$\gamma=-2$	$\gamma=0$
Full hedging	1.00	0	523	0.000	−3.20	−3.20	−3.20	−3.20
No hedging	0.00	523	0	0.000	8.18	0.24	6.60	8.18
RW hedging	0.46	281	242	0.471	7.56	3.35	6.73	7.56

column, labelled $\sum r_t$, presents the cumulative returns[12] for each strategy. Over the 1998-1999 period the full hedge strategy would have returned a negative result of −3.20%, whereas the no hedge strategy would have yielded 8.18% as a result of a US dollar appreciation over this period. The cumulative return of the random walk strategy (7.56%) is very close to the no hedge strategy. In order to determine the impact of the riskiness of these strategies we have also computed the associated utilities for distinct values of the risk aversion parameter γ. These are presented in the last three columns of the table. For the full hedge strategy the utilities do not change since this strategy has no exchange rate risk, and the risk in the interest rate differential can be neglected. For $\gamma=-10$, -2 the riskiness of the no hedge strategy is reflected in the utility values. Only for the more risk-tolerant case ($\gamma=0$), the no hedge strategy has a higher utility than the random walk strategy.[13] Concluding we can say that the random walk strategy may be hard to beat. However, a disadvantage of the random walk strategy is that, due to the variability of exchange rate returns, the hedge position needs to be changed (too) often from unhedged to hedged and vice versa. In the next sections we turn our focus on evaluating hedging strategies based on the time series models that were presented in Section 3.

5.6 Optimal Hedging

In this section we investigate the properties of currency hedge strategies for the different time series models based on the power utility function. In Table VI we present our results. Each panel in the table corresponds with a particular value of γ. The first thing to note is the sensitivity of the results for the choice of γ. Not surprisingly, for each model the average hedge ratio increases when the risk aversion increases (in the limit, for our case when $\gamma < -50$, the investor hedges fully). Also, the number of times the hedge ratio is equal to zero increases when the investor becomes less risk averse.[14] The variation of the hedge ratio $\overline{\lvert\Delta H\rvert}$ is considerably lower than in the random walk case. When comparing the results in Table VI with the benchmark results from Section 5.5 we find that for a very risk-averse investor ($\gamma=-10$) it holds that all model strategies have lower returns and utilities than the corresponding results for a random walk strategy. Note that the GLL-SV model is a close second in utility, even though the return is 3% lower. Also note that for this investor the model strategies always beat the full hedge benchmark case. This implies that modelling exchange rate behaviour is worth while for a risk-averse investor with a full hedge benchmark. For $\gamma=0$ or $\gamma=-2$ one can beat the random walk

[12] The returns in Tables V–VII and in Figure 8 are expressed in percentages. Utilities are multiplied by 100.
[13] Note that the utility levels for $\gamma=0$ are equal to the cumulative return.
[14] The hedge ratio may take values between zero and one. The table only reports the number of times the hedge ratio is exactly equal to zero or one.

Table VI. Results for optimal hedging strategies

Model	$\bar{H}$	#$(H=0)$	#$(H=1)$	$\overline{\lvert\Delta H\rvert}$	$\sum r_t$	$\sum U_t$
			$\gamma=-10$			
WN	0.91	0	0	0.004	−2.18	−2.24
LL	0.47	206	191	0.074	3.51	−0.16
GLL	0.78	0	119	0.105	−1.05	−1.59
GLL-GARCH	0.61	65	162	0.179	−0.56	−2.62
GLL-SV	0.62	46	135	0.187	4.29	3.10
GLL-Student-*t*	0.75	0	99	0.122	−0.50	−1.17
GLL-GARCH-Student-*t*	0.58	66	135	0.176	2.31	−0.01
			$\gamma=-2$			
WN	0.68	0	0	0.012	0.40	0.23
LL	0.39	309	191	0.066	4.75	3.83
GLL	0.38	205	118	0.210	6.42	5.61
GLL-GARCH	0.38	281	161	0.185	7.23	6.37
GLL-SV	0.35	293	134	0.209	8.54	7.69
GLL-Student-*t*	0.35	240	99	0.230	5.66	4.80
GLL-GARCH-Student-*t*	0.34	301	134	0.177	8.35	7.42
			$\gamma=0$			
WN	0.11	195	0	0.023	5.66	5.66
LL	0.37	325	191	0.064	6.70	6.70
GLL	0.27	360	117	0.175	8.95	8.95
GLL-GARCH	0.34	337	161	0.173	9.01	9.01
GLL-SV	0.29	353	133	0.188	9.60	9.60
GLL-Student-*t*	0.23	374	99	0.196	7.72	7.72
GLL-GARCH-Student-*t*	0.28	366	134	0.170	7.40	7.40

strategy in terms of returns and utilities. For $\gamma=-2$ this holds for a small subset, notably the GLL-SV and GLL-GARCH-Student-*t* models, whereas for $\gamma=0$ almost all models based on the Generalized Local Level specification beat the random walk strategy.

From a currency overlay management perspective modelling exchange rate returns becomes more important for investors who are less risk averse. Zooming in on the results for $\gamma=0$ we see that the GLL-SV model comes out best. Variants of the GLL with GARCH and/or Student-*t* turn out second. Note however that the differences among models are small, at the end of the evaluation sample (but also see Section 5.7, where the results of various strategies during the evaluation sample are compared). Concluding, one may note that modelling time-varying features of exchange rate series pays off when the investor's risk appetite is high. The choice among models that satisfy these criteria is less important.

In practice model choice may be influenced by the number of forward transactions that have to be done in order to implement a hedging strategy. When the hedge position needs to be adjusted frequently, more transaction costs may have to be paid to the counterparty. Furthermore, management of the exposures is more vulnerable to operational risk. In Figure 7 we have plotted the hedge ratios over our 523-day evaluation period. The upper panel shows the hedge ratios from the Random Walk strategy. They are quite erratic as a result of the exchange rate variability. Not surprisingly, the White Noise hedge ratios (second panel) are quite stable. The hedge ratios of the models based on time varying means are strikingly similar. This partly

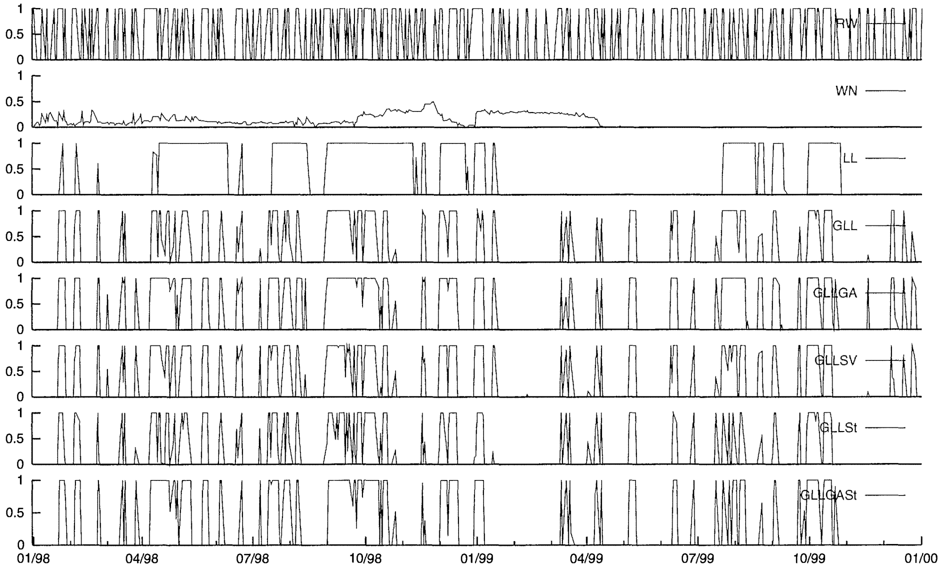

Figure 7. Hedging decisions through time, for $\gamma = 0$

Table VII. Results for alternative hedging strategies, $\gamma = 0$

	Optimal		VaR		Sharpe	
Model	H	$\sum r_t$	$\sum r_t$	Limit	$\sum r_t$	Limit
WN	0.11	5.66	−1.74	−1.10	−1.74	−1.32
LL	0.37	6.70	2.00	−1.06	2.13	5.88
GLL	0.27	8.95	12.81	−1.09	12.81	0.39
GLL-GARCH	0.34	9.01	6.74	−0.87	10.32	1.07
GLL-SV	0.29	9.60	1.17	−0.81	6.23	1.13
GLL-Student-*t*	0.23	7.72	10.02	−1.01	10.66	0.78
GLL-GARCH-Student-*t*	0.28	7.40	7.95	−0.82	8.35	2.12

explains why the performances of these models are sometimes close to each other. This observation corroborates the statement that the precise functional form of the model is less important within the class of Generalized Local Level models.

5.7 Alternative Hedging Strategies

In Table VII 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 VI for the case $\gamma = 0$. The limiting levels of the acceptable VaR and Sharpe values (reported in columns 4 and 6) have been chosen such that, *ex post*, the average hedge ratio corresponds to the value found for the strategy optimizing the utility. The returns (which equal

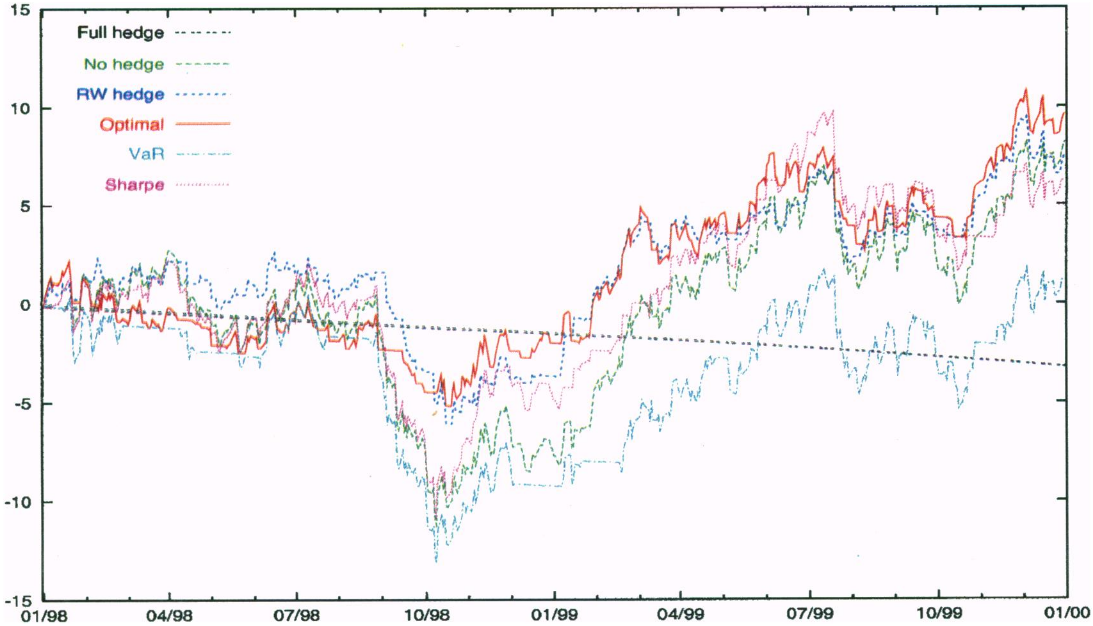

Figure 8. Utilities/returns for the GLL-Stochastic Volatility model, for $\gamma = 0$

Table VIII. Coverage probabilities of VaR

Model	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
WN	0.036 (0.13)	0.023 (0.76)	0.013 (0.46)
LL	0.040 (0.29)	0.025 (0.98)	0.011 (0.74)
GLL	0.036 (0.13)	0.023 (0.76)	0.013 (0.46)
GLL-GARCH	0.061 (0.26)	0.034 (0.19)	0.019 (0.06)
GLL-SV	0.050 (0.98)	0.023 (0.76)	0.008 (0.57)
GLL-Student-*t*	0.048 (0.82)	0.023 (0.76)	0.011 (0.74)
GLL-GARCH-Student-*t*	0.065 (0.13)	0.031 (0.43)	0.013 (0.46)

the utilities, as $\gamma = 0$) are high for models that have a GLL component. Apparently, both criteria favour a strong signal of the time-varying mean. The VaR and Sharpe ratio objective functions focus, however, on particular aspects of the distribution of exchange rate returns. VaR concentrates on the left tail of the distribution and Sharpe focuses exclusively on mean and volatility. From our results we may conclude that this property can lead to performance losses, measured in risk-adjusted returns, on currency overlay strategies.

In Plate 1 the evolution of the cumulative utilities for different strategies for the GLL-SV model are plotted. In the plot, all utilities have been calculated using a risk tolerance of $\gamma = 0$, the case where (cumulative) utility equals the (cumulative) return. The straight black line is the base case where all risk is hedged, the green line ending at 8.2% is the evolution of the exchange rate itself, obtained when no hedging is applied. The red line represents the case where optimal hedging (given $\gamma = 0$) is used. In the first nine months, the model is rather careful, and the return stays close to the fully hedged return. This is also the case during the months September and October 1998, where no huge losses are incurred. Other strategies, especially the zero hedge case, the Sharpe (purple) and the VaR (light blue) strategy run into a loss of around 12% over those months. The RW (dark blue) limits the drop to around 8%. The loss for the 'optimal' strategy is contained within 3.5%. In periods of appreciation of the exchange rate, the different strategies have similar returns, as little hedging takes place (see also Figure 7).

At the end of the sample, the cumulative utilities of the different strategies are similar. From the results in Tables V and VI we found that from a risk and return perspective the GLL-SV model does slightly better than the benchmark strategies, also for other values of γ.

In Table VIII we present the fraction of realized returns not exceeding the Value-at-Risk returns, together with the test on the unconditional coverage probability as in Christoffersen (1998) (*p*-values, between parentheses). All models pass the test for these confidence levels.

6. CONCLUSIONS

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 DEM/USD 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.

The empirical results which we find for the DEM/USD exchange rate over the period 1998-1999 are summarized as follows. First, modelling time-varying features, and using a power utility objective function, pays off in terms of risk-adjusted returns for a moderately risk-averse currency overlay manager. Second, modelling becomes less valuable when risk aversion increases. Simple random walk strategies outperform our optimal strategies that are based on time series models. But the time variation of hedge ratios in the random walk strategy may be prohibitive in actual implementation of strategies. Third, when modelling is worthwhile it appears that there is not one model that is uniformly superior for all criteria. However the GLL-SV and the GLL-GARCH-Student-*t* are close competitors. Fourth, for some time series models strategies based on VaR and Sharpe objective functions have better results. However, care has to be taken since these objective functions focus on distinctive parts of the distribution of exchange rate returns only. From the period September-October 1998 we infer that modelling exchange rate returns and using utility analysis is especially important in periods of high risk of depreciation.

Our overall conclusions for practical currency management are:

(1) It is more important to choose the level of risk behaviour and the class of criterion functions in combination with a specific model than to endlessly fight on specific functional forms of time series models for exchange rate returns.
(2) Modelling the time-varying mean and variance features of exchange rate returns in an integrated framework appears worth while, in particular in periods with large decreases in exchange rates.

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.

Second, 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 and Dahlquist (2000) and Evans and Lewis (1995).

Third, 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 and Zhou (1996), McCulloch and Rossi (1990), McCulloch and Rossi (1991), and Kandel, McCulloch and 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 currencies as an investment in their own right.

APPENDIX: 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 trivial as the likelihood function is only available as a multivariate integral.

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 and 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 $t = 1, \ldots, T$. 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

$$y^*_t = \ln(y_t - \mu_t)^2 = h_t + z_t \qquad z_t = \ln(\epsilon^2_t) \tag{17}$$

$$h_t = \mu_h + \phi(h_{t-1} - \mu_h) + \xi_t \tag{18}$$

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 and Shephard (1998).

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

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

Table IX. Conditional posterior densities

Parameter	In model	Full conditional density
μ	**E, F, G**	Use the simulation smoother, see de Jong and 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**	$\mathrm{IG}\left(\alpha = \frac{T}{2} + \alpha_\eta, \beta = 2/\left(\sum(\mu_t - \rho\mu_{t-1})^2 + \frac{2}{\beta_\eta}\right)\right)$
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**	$\mathrm{IG}\left(\alpha = \frac{T}{2} + \alpha_\xi, \beta = 2/\left(\sum((h_t - \mu_h) - \phi(h_{t-1} - \mu_h))^2 + \frac{2}{\beta_\xi}\right)\right)$
s_t	**E**	The indices into the mixtures 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**	$\mathrm{IG}\left(\alpha = \frac{\nu+1}{2}, \beta = \frac{2}{(\nu-2)+(y_t-\mu_t)^2/(\sigma_\epsilon^2 h_t)}\right)$
ν	**F, G**	The posterior is not of a known form. It is proportional to $\prod_t \mathrm{IG}\left(z_t; \alpha = \frac{\nu}{2}, \beta = \frac{2}{\nu-2}\right) \times \mathrm{Cauchy}\,(\nu; \mu = 0; s = 1)$. Apply a MH step to sample a new value of ν

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

We obtain the Student-t density for the disturbances ϵ_t as the marginal of the normal-inverted gamma density,

$$\epsilon_t, z_t|\nu \sim \mathcal{N}(0, h_t z_t \sigma_\epsilon^2) \times \mathrm{IG}\left(\alpha = \frac{\nu}{2}, \beta = \frac{2}{\nu - 2}\right) \tag{19}$$

where the marginalization takes place with respect to the mixing parameter z_t. It is straightforward to derive that the marginal density $p(\epsilon_t|\nu) = \int_z p(\epsilon_t, z|\nu)\,\mathrm{d}z$ is indeed the Student-t density with ν degrees of freedom (see e.g. Bauwens, Lubrano and Richard, 1999, theorem A.7).

The full conditional posterior densities which are needed in the Gibbs sampling algorithm are given without derivation in Table IX. For the GARCH parameters σ_ϵ^2, δ, α and for the degrees-of-freedom parameter ν no closed form expression of the conditional density is available. Therefore, we use in these steps a Metropolis-within-Gibbs sampler (see Koop and van Dijk, 2000; Zeger and Karim, 1991). Note that the priors in Table I in Section 4.1 have been applied.

ACKNOWLEDGEMENTS

We thank three anonymous referees, the co-editor, Rob Engle, Christopher Gilbert, Frank de

Jong, Michel Lubrano, Allan Timmermann 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 comments which led to substantial improvements on earlier versions of this paper. The authors retain full responsibility for remaining errors.

REFERENCES

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

Bansal, R. and M. Dahlquist (2000), 'The forward premium puzzle: different tales from developed and emerging economies', *Journal of International Economics*, **51(1)**, 115–144.

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

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

Bauwens, L., M. Lubrano and J.-F. Richard (1999), *Bayesian Inference in Dynamic Econometric Models*, Advanced Texts in Econometrics. Oxford University Press, Oxford.

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

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

Carter, C. K. and R. Kohn (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. and E. Greenberg (1995), 'Understanding the Metropolis–Hastings algorithm', *The American Statistician*, **49(4)**, 327–335.

Chib, S., F. Nardari and N. Shephard (1998), Markov chain Monte Carlo methods for generalized stochastic volatility models. Nuffield College working paper 1998-W21, University of Oxford.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review*, **39(4)**, 841–862.

De Jong, P. and N. Shephard (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). Timberlake Consultants Ltd, London. See http://www.nuff.ox.ac.uk/Users/Doornik.

Engel, C. and J. D. Hamilton (1990), 'Long swings in the dollar: Are they in the data and do markets know it?', *American Economic Review*, **4**, 689–714.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations', *Econometrica*, **50**, 987–1008.

Engle, R. F. (1995), *ARCH: Selected Readings*, Advanced Texts in Econometrics. Oxford University Press, Oxford.

Evans, M. D. D. and K. K. Lewis (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. and A. F. M. Smith (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 and A. F. M. Smith (Eds), *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, Clarendon Press, Oxford, pp. 169–193.

Geweke, J. and G. Zhou (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., N. G. Polson and P. E. Rossi (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*, McGraw-Hill, New York.
Kandel, S., R. McCulloch and R. Stambaugh (1995), 'Bayesian inference and portfolio efficiency', *Review of Financial Studies*, **8**, 1–53.
Kass, R. E. and A. E. Raftery (1995), 'Bayes factors', *Journal of the American Statistical Association*, **90(430)**, 773–795.
Kim, S., N. Shephard and S. Chib (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies*, **64**, 361–393.
Kleibergen, F. R. and H. K. van Dijk (1993), 'Non-stationarity in GARCH models: A Bayesian analysis', *Journal of Applied Econometrics*, **8**, S41–S61.
Koop, G. and H. K. van Dijk (2000), 'Testing for integration using evolving trend and seasonal models: A Bayesian approach', *Journal of Econometrics*, **97(2)**, 261–291.
Koopman, S. J., N. Shephard and J. A. Doornik (1999), 'Statistical algorithms for models in state space using SsfPack 2.2', *Econometrics Journal*, **2**, 107–160.
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. and P. Rossi (1990), 'Posterior, predictive, and utility-bases approaches to testing the arbitrage pricing theory', *Journal of Financial Economics*, **28**, 7–38.
McCulloch, R. and P. Rossi (1991), 'A Bayesian approach to testing the arbitrage pricing theory', *Journal of Econometrics*, **49**, 141–168.
Morgan, J. P. (1997), 'RiskMetrics, Technical document, `http://www.riskmetrics.com`
Nelson, D. B. (1990), 'Stationarity and persistence in the GARCH(1,1) model', *Econometric Theory*, **6**, 318–334.
Newey, W. K. and K. D. West (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.
Smith, A. F. M. and G. O. Roberts (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.
Solnik, B. H. (2000), *International Investments*, The Addison-Wesley Series in Finance, (4th edn), Addison-Wesley, Reading, MA.
Taylor, S. J. (1994), 'Modeling stochastic volatility: A review and comparative study', *Mathematical Finance*, **4(2)**, 183–204.
West, M., P. J. Harrison and H. S. Migon (1985), 'Dynamic generalized linear models and Bayesian forecasting', *Journal of the American Statistical Association*, **80(389)**, 73–98.
Zeger, S. L. and M. R. Karim (1991), 'Generalized linear models with random effects: A Gibbs sampling approach', *Journal of the American Statistical Association*, **86**, 9–86.

