FRONTIERS OF ECONOMICS AND GLOBALIZATION VOLUME 3

# FORECASTING IN THE PRESENCE OF STRUCTURAL BREAKS AND MODEL UNCERTAINTY

*Edited by*

David E. Rapach

*Saint Louis University, Saint Louis, MO, USA*

Mark E. Wohar

*University of Nebraska at Omaha, Omaha, NE, USA*

Emerald

United Kingdom – North America – Japan – India – Malaysia – China

ISOQAR
QUALITY ASSURED
ISO 9001

UKAS
QUALITY MANAGEMENT
026

Certificate number ........1985.........

Awarded in recognition of
Emerald's production
department's adherence to
quality systems and processes
when preparing scholarly
journals for print

INVESTOR IN PEOPLE

CHAPTER 12

# Modeling Foreign Exchange Rates with Jumps

John M. Maheu[a] and Thomas H. McCurdy[b]

[a]Department of Economics, University of Toronto, Canada
E-mail address: jmaheu@chass.utoronto.ca
[b]Joseph L. Rotman School of Management, University of Toronto, Associated Fellow,
CIRANO, Canada
E-mail address: tmccurdy@rotman.utoronto.ca

## Abstract

We propose a new discrete-time model of returns in which jumps capture persistence in the conditional variance and higher-order moments. Jump arrival is governed by a heterogeneous Poisson process. The intensity is directed by a latent stochastic autoregressive process, while the jump-size distribution allows for conditional heteroskedasticity. Model evaluation focuses on the dynamics of the conditional distribution of returns using density and variance forecasts. Predictive likelihoods provide a period-by-period comparison of the performance of our heterogeneous jump model relative to conventional SV and GARCH models. Furthermore, in contrast to previous studies on the importance of jumps, we utilize realized volatility to assess out-of-sample variance forecasts.

**Keywords:** Jump clustering, jump dynamics, MCMC, predictive likelihood, realized volatility, Bayesian model averaging

**JEL classifications:** C22, C11, G1

## 1. Introduction

Measuring and forecasting the distribution of returns is important for many problems in finance. Pricing of financial securities, risk management decisions, and portfolio allocations all depend on the distributional features of returns. The most predictable component of the conditional distribution is the structure in the variance. As a result, a vast literature has sprung from the ARCH model of Engle (1982) and the stochastic volatility (SV) approach of Taylor (1986).

Current research has documented the importance of jump dynamics in combination with autoregressive volatility for modeling returns. Examples of this work

include Andersen *et al.* (2002), Bates (2000), Chernov *et al.* (2003), Chib *et al.* (2002), Eraker *et al.* (2003), Jorion (1988), and Maheu and McCurdy (2004), among others. Jumps provide a useful addition to SV models by explaining occasional, large abrupt moves in financial markets, but they are generally not used to capture volatility clustering. As a result, jumps account for neglected structure, usually tail dynamics, that autoregressive SV cannot capture. Additional SV factors is another possible solution. For example, Chernov *et al.* (2003) find that a multifactor loglinear SV specification is equivalent to an affine class of SV with jumps for equity data.

All financial data are measured in discrete time which suggests that jumps provide a natural framework to model price moves. Nevertheless, surprisingly few applications exclude an autoregressive SV component in order to focus on the potential performance of the jump specification to capture return dynamics. Recent research, including Das (2002), Lin *et al.* (1999), Johannes *et al.* (1999), and Oomen (2006), suggests that jumps alone could provide a good specification for financial returns. As in Maheu and McCurdy (2004), those applications all feature some form of dependence in the arrival rate of jumps. For example, Johannes *et al.* (1999) allow jump arrival to depend on past jumps and the absolute value of returns. Indeed, the success of a Jump model will depend on whether the specification can explain dynamics of the conditional distribution, in particular, volatility clustering. However, the performance of existing models in this regard is unclear. For example, can models with only jump dynamics produce good volatility forecasts? Are they competitive with standard volatility models? The purpose of this chapter is to investigate these questions.

This chapter proposes a new discrete-time model of returns in which jumps capture persistence in the conditional variance as well as time-variation in higher-order moments. The jump intensity is directed by a latent stochastic autoregressive process. Therefore, jump arrivals can cluster. We also allow the jump-size distribution to be conditionally heteroskedastic. Larger jumps occur during volatile periods and smaller ones during quiet periods. In contrast to GARCH and SV models, our heterogeneous Jump model allows periods of homoskedasticity and periods of heteroskedasticity.

We estimate the model using Markov chain Monte Carlo (MCMC) methods, and follow Johannes *et al.* (1999) in treating unobserved state variables such as jump times and jump sizes as parameters. Our MCMC estimation approach generates estimates of these quantities which incorporate parameter uncertainty.

Model evaluation focuses on the dynamics of the conditional distribution of returns. Using Bayesian simulation methods, we estimate predictive likelihoods for models as suggested by Geweke (1994). This is the relevant measure of out-of-sample predictive content of a model (Geweke and Whiteman, 2006). This allows for period-by-period comparisons of the Jump model with the benchmark SV model and is particularly useful in identifying influential observations. From these calculations, we are able to report the cumulative model probability as

a function of time, which provides insight into the performance of each model through time.

In addition to conditional density evaluation, we compare forecasts of volatility associated with each model, as well as the combined model average forecast. Andersen and Bollerslev (1998) and Andersen *et al.* (2001) show that the sum-of-squared intraday returns is an efficient and consistent estimate of ex post daily volatility. As in Andersen *et al.* (2001) and Maheu and McCurdy (2002), we compute this realized volatility estimate based on 5-minute bilateral exchange rate data from the Yen–US Dollar (JPY–USD) foreign exchange market. The out-of-sample forecasts of volatility from our alternative models are assessed using these realized volatility measures.

Our empirical results show that the process governing jump arrival is highly persistent. The conditional jump probability (intensity) varies widely over time, taking on values of less than 0.05 to values in excess of 0.90. This implies that jumps will tend to cluster which can also be seen from the estimated jump times. The jump-size variance is a positive function of the market's volatility as measured by last period's absolute value of daily returns. Therefore realized jumps will tend to be larger in a volatile market.

Based on the predictive Bayes factor, our jump specification is favored over both SV and GARCH alternatives. Although the evidence in favor of the Jump model varies considerably over the sample, it gradually strengthens from the perspective of the cumulative model probability. However, the SV model is competitive. Therefore, when we proceed to evaluating out-of-sample variance forecasts relative to realized volatility, we also include a model average variance forecast.

The main difference in the variance forecasts from the alternative models is that during low volatility periods the Jump model produces a constant variance forecast due to a low probability of jumps. On the other hand, the SV and GARCH model forecasts tend to be too volatile during these calm periods. In other words, periods of low volatility may be better represented by the constant variance generated by the Jump model. Another important feature of our model is that we can immediately move from high volatility to low volatility levels avoiding the slow decay built into SV and GARCH models. In summary, our proposed heterogeneous Jump model can account for volatility persistence and is very competitive with existing models of volatility.

The chapter is organized as follows. The next section presents a heterogeneous Jump model for foreign exchange returns, while Section 3 briefly discusses benchmark SV and GARCH specifications used for comparison purposes. Section 4 considers Bayesian estimation of the Jump model. Estimation of the predictive likelihood for model comparison is reviewed in Section 5, while model forecasts are explained in Section 6. Data sources are found in Section 7, results in Section 8, and conclusions in Section 9. Appendix A contains detailed calculations for the estimation algorithms.

## 2. Basic jump model

We begin with a brief review of the Press (1967) Jump model. Consider the following jump diffusion:

$$\frac{dP(t)}{P(t)} = \alpha \, dt + \sigma \, dW(t) + \exp(\xi) \, dq(t), \tag{2.1}$$

where $P(t)$ is the security price, $dW(t)$ is a standard Wiener process increment, $\exp(\xi)$ is the percent change in the share price from a jump, and $dq(t)$ is a Poisson counter ($dq(t) = 1$ for a jump and otherwise 0) with intensity parameter $\lambda$. From the solution to this differential equation the daily return evolves according to

$$r_t = \mu \, dt + \sigma z_t + \sum_{k=q(t-1)}^{q(t)} \xi_k, \tag{2.2}$$

where $r_t = \log(P(t)/P(t-1))$, and $\mu = \alpha - \sigma^2/2$. Although this model can produce constant conditional and unconditional skewness and kurtosis, it cannot capture volatility clustering. There is a large literature which augments this basic model with dynamics for $\sigma^2$. The purpose of this chapter is to investigate additional dynamics in the jump intensity and the jump-size distribution. Equation (2.2) motivates our starting point for exploring the importance of jump dynamics, which we describe next.

### 2.1. Heterogeneous jump parameterization

This subsection proposes a new discrete-time model which can capture the autoregressive pattern in the conditional variance of returns by allowing jumps to arrive according to a heterogeneous Poisson process. Our parameterization includes a latent autoregressive structure for the jump intensity, as well as a conditionally heteroskedastic variance for the jump-size distribution. The mean of the jump-size distribution can be significantly different from zero, allowing the specification to capture a skewed distribution of returns.

The parameterization of the heterogeneous Jump model is as follows:

$$r_t = \mu + \sigma z_t + J_t \xi_t, \quad z_t \sim N(0, 1), \tag{2.3}$$

$$\xi_t \sim N(\mu_\xi, \sigma_{\xi,t}^2), \qquad J_t \in \{0, 1\}, \tag{2.4}$$

$$P(J_t = 1 | \omega_t) = \lambda_t \quad \text{and} \quad P(J_t = 0 | \omega_t) = 1 - \lambda_t, \tag{2.5}$$

$$\lambda_t = \frac{\exp(\omega_t)}{1 + \exp(\omega_t)}, \tag{2.6}$$

$$\omega_t = \gamma_0 + \gamma_1 \omega_{t-1} + u_t, \quad u_t \sim N(0, 1), \ |\gamma_1| < 1, \tag{2.7}$$

$$\sigma_{\xi,t}^2 = \eta_0 + \eta_1 X_{t-1}, \tag{2.8}$$

where $r_t$ denotes daily returns; $t = 1, \ldots, T$; $\mu$ is the mean of the returns conditional on no jump; and $\xi_t$ is the jump size which follows a conditional normal distribution.

$J_t$ is an indicator that identifies when jumps occur. In particular, the set $\{\forall J_t \mid J_t = 1\}$ denotes jump times. $\lambda_t$ is the time-varying jump intensity or arrival process which is directed by the latent autoregressive process $\omega_t$. In this chapter, we allow for one jump per period, that is, $J_t = 0, 1$. The logistic function ensures that $\omega_t$ is mapped into a $(0, 1)$ interval for $\lambda_t$. Conditional on $\omega_t$, the probability of a jump follows a Bernoulli distribution. This specification could be extended to allow for more than one jump per period by using an ordered probit model for $P(J_t | \lambda_t)$. Besides capturing occasional large moves in returns, this specification can account for volatility clustering through persistence in $\omega_t$ and dependence in higher-order moments of the distribution. This model is a time-varying mixture of normals and therefore possess time-variation in all higher-order moments. One interpretation of $\omega_t$ is that it represents unobserved news flows into the market that cause trading activity.

The variance of the jump-size distribution, $\sigma_{\xi,t}^2$, is allowed to be a function of weakly exogenous regressors $X_{t-1}$. In this chapter, we consider $X_{t-1} = |r_{t-1}|$. This specification permits the jump-size variance to be sensitive to recent market volatility levels. For example, if $\eta_1 > 0$, jumps will tend to be larger (smaller) in volatile (quiet) markets.

Imposing the restrictions $\lambda_t = \lambda$, $\forall t$, and $\eta_1 = 0$ obtains a simple pure Jump model with an iid arrival of jumps and a homogeneous jump-size distribution (Press, 1967). Although not considered in this chapter, it is straightforward to allow additional dynamics such as a different jump-size distribution or time dependence in $\mu_\xi$, the mean of the jump-size distribution. In addition, other possibilities could be explored for $X_{t-1}$ such as the range.

## 3. Benchmark specifications

### 3.1. SV model

Our application is to foreign exchange rates, for which asymmetries and leverage effects are generally absent. Therefore, we consider a standard log-linear stochastic volatility (SV) model as a benchmark for comparison purposes. A large literature discusses estimation methods for this model; see, for example, Melino and Turnbull (1990), Danielsson and Richard (1993), Gourieroux *et al.* (1993), Jacquier *et al.* (1994), Andersen and Sorensen (1996), Gallant *et al.* (1997), Kim *et al.* (1998), Alizadeh *et al.* (2002), and Jacquier *et al.* (2004). Surveys of the SV literature include Taylor (1994), Ghysels *et al.* (1996), and Shephard (1996).

The discrete-time SV model is parameterized as

$$r_t = \mu + \exp(h_t/2) z_t, \quad z_t \sim N(0, 1), \tag{3.1}$$

$$h_t = \rho_0 + \rho_1 h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0, 1). \tag{3.2}$$

Given $h_t$, the conditional variance of returns is $\exp(h_t)$. Conditionally, the model produces fat tails. Our estimation approach imposes stationarity by rejecting any MCMC draws that violate $|\rho_1| < 1$.

Compared to GARCH models, estimation is more difficult due to the latent volatility which must be integrated out of the likelihood. Bayesian methods rely on Markov chain Monte Carlo (MCMC) sampling to estimate SV models. The properties of the estimator compare favorably with other approaches (Jacquier et al., 1994). It is straightforward to obtain smoothed estimates of volatility from MCMC output. In addition, these estimates of volatility take parameter uncertainty into account.

## 3.2. GARCH model

The second benchmark specification that we include is a standard GARCH(1, 1) model:

$$r_t = \mu + \epsilon_t, \quad \epsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1), \tag{3.3}$$

$$\sigma_t^2 = \kappa + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{3.4}$$

We restrict the model to be covariance stationary by restricting $\alpha + \beta < 1$ in estimation.

## 4. Posterior inference

Johannes and Polson (2006) provide an excellent overview of Bayesian methods for financial models including a simple Jump model. Eraker et al. (2003) and Johannes et al. (1999) discuss a data augmentation approach to deal with jump times and jump sizes.

From a Bayesian perspective, inference regarding parameters takes place through the posterior which incorporates both the prior and likelihood function. Let the history of data be denoted as $\Phi_t = \{r_1, \ldots, r_t\}$. In the case of our heterogeneous Jump model, we augment the parameters $\theta = \{\mu, \sigma^2, \mu_\xi, \eta, \gamma\}$ where $\eta = \{\eta_0, \eta_1\}$ and $\gamma = \{\gamma_0, \gamma_1\}$, with the unobserved state vectors $\omega = \{\omega_1, \ldots, \omega_T\}$, jump times $J = \{J_1, \ldots, J_T\}$, and jump sizes $\xi = \{\xi_1, \ldots, \xi_T\}$ and treat these as parameters. For the Jump model, Bayes rule gives us:

$$p(\theta, \omega, J, \xi | \Phi_T) \propto p(r | \theta, \omega, J, \xi) p(\omega, J, \xi | \theta) p(\theta) \tag{4.1}$$

where $r = \{r_1, \ldots, r_T\}$; $p(r | \theta, \omega, J, \xi)$ is the joint density of returns conditional on the state variables $\omega$, $J$, and $\xi$; $p(\omega, J, \xi | \theta)$ is the density of the state variables; and $p(\theta)$ is the prior. By treating jump times and jump sizes as parameters to sample we avoid the difficulty of integrating them out, but more importantly we can compute smoothed estimates of them from the posterior sample. In practice, analytical results are not available and we use MCMC methods to draw samples from the posterior. Surveys of MCMC methods include Geweke (1997), Robert and Casella (1999), and Chib (2001).

MCMC theory allows valid draws from the posterior to be obtained by sampling from a series of conditional distributions. It is often much easier to work with the conditional distributions. In the limit, draws converge to samples from the posterior. A simulation-consistent estimate of any function of the parameter vector can be constructed from sample averages. For instance, if we have $\{\mu^{(i)}\}_{i=1}^N$ draws of $\mu$ from the posterior, and assuming the integral of $g(\mu)$ with respect to the marginal posterior exists, we can estimate $E[g(\mu)]$ as $\frac{1}{N} \sum_{i=1}^N g(\mu^{(i)})$. For example, to compute $E(\mu)$, set $g(\mu) = \mu$; to compute $E(\mu^2)$ and the variance, set $g(\mu) = \mu^2$, etc. See Tierney (1994) for technical details. Further assumptions on the integrability of $g(\mu)^2$ permit consistent estimation of the asymptotic standard error of the estimate, based on conventional time series methods.

In the following we denote the vector $\theta$ excluding the $k$th element $\theta_k$ as $\theta_{-\theta_k}$, the subvector $\{\omega_t, \ldots, \omega_\tau\}$ as $\omega_{(t,\tau)}$ and $\omega$ excluding $\omega_{(t,\tau)}$ as $\omega_{-(t,\tau)}$. Sampling is based on Gibbs and Metropolis-Hasting (MH) routines. Draws from the posterior $\psi = \{\theta, \omega, J, \xi\}$ are obtained by cycling over the following steps:

1. Sample $\mu | \theta_{-\mu}, \omega, J, \xi, r$;
2. Sample $\sigma^2 | \theta_{-\sigma^2}, \omega, J, \xi, r$;
3. Sample $\mu_\xi | \theta_{-\mu_\xi}, \omega, J, \xi, r$;
4. Sample $\eta | \theta_{-\eta}, \omega, J, \xi, r$;
5. Sample blocks $\omega_{(t,\tau)} | \theta, \omega_{-(t,\tau)}, J, \xi, r, t = 1, \ldots, T$;
6. Sample $\gamma | \theta_{-\gamma}, \omega, J, \xi, r$;
7. Sample $\xi | \theta, \omega, J, r$;
8. Sample $J | \theta, \omega, \xi, r$;
9. Go to 1.

A pass through 1–8 provides a draw from the posterior. We repeat this several thousand times and collect these draws after an initial burn-in period. Note that the parameters $\eta$, $\gamma$, $J$, $\xi$, and $\omega$ are sampled as blocks which may contribute to better mixing of the MCMC output. Detailed steps of the algorithm are collected in Appendix A.

## 5. Model comparison

The key ingredient in Bayesian model comparison is the marginal likelihood, which can be used to form Bayes factors or model probabilities. However, a drawback of any statistical approach that summarizes a model's performance with a single number is understanding why and when a model performs well or poorly. Geweke (1994) suggests the use of a predictive likelihood decomposition of the marginal likelihood. Estimating the predictive likelihood allows us to compare models on an observation by observation basis. This may be useful in identifying influential observations or periods that make a large contribution to predictive Bayes factors. Applications of this idea include Gordon (1997) and Min and Zellner (1993).

Consider a model with parameter vector $\Theta$. The predictive density for observation $y_{t+1}$ based on the information set $\Phi_t$ is

$$p(y_{t+1}|\Phi_t) = \int p(y_{t+1}|\Phi_t, \Theta)p(\Theta|\Phi_t)\,d\Theta, \tag{5.1}$$

where $p(\Theta|\Phi_t)$ is the posterior and $p(y_{t+1}|\Phi_t, \Theta)$ is the conditional distribution. When it is clear, we suppress conditioning on a model for notational convenience. Evaluating (5.1) at the realized $\tilde{y}_{t+1}$ gives the predictive likelihood,

$$\hat{p}_{t+1} = p(\tilde{y}_{t+1}|\Phi_t) = \int p(\tilde{y}_{t+1}|\Phi_t, \Theta)p(\Theta|\Phi_t)\,d\Theta, \tag{5.2}$$

which can be estimated from MCMC output. Models that have a larger predictive likelihood are preferred to ones with a smaller value, as they are more likely to have generated the data.

Geweke (1994) shows that the predictive likelihood for observations $\tilde{y}_u, \ldots, \tilde{y}_v, u < v$, can be decomposed as

$$p(\tilde{y}_u, \ldots, \tilde{y}_v|\Phi_{u-1}) = \int p(\tilde{y}_u, \ldots, \tilde{y}_v|\Phi_{u-1}, \Theta)p(\Theta|\Phi_{u-1})\,d\Theta$$

$$= \prod_{i=u}^{v} \hat{p}_i. \tag{5.3}$$

Therefore we can compute the predictive likelihood for $\tilde{y}_u, \ldots, \tilde{y}_v$ observation by observation. If $u = 1$ and $v = T$, then (5.3) provides a full decomposition of the marginal likelihood. In practice we will use a *training sample* that we condition on for all models. This initial sample of observations $1, 2, \ldots, u - 1$, combined with the likelihood and prior, forms a new prior, $p(\Theta|\Phi_{u-1})$, on which all calculations are based. If $u$ is large then $p(\Theta|\Phi_{u-1})$ will be dominated by the likelihood function, and the original prior $p(\Theta)$ will have a minimal contribution to model comparison exercises. Note that conditional on this training sample, the log predictive Bayes factor in favor of model $j$ versus $k$ for the data $\tilde{y}_u, \ldots, \tilde{y}_v$ is

$$\log B_{j,k} = \sum_{i=u}^{v} \log \frac{\hat{p}_i^j}{\hat{p}_i^k}, \tag{5.4}$$

where the predictive likelihood is now indexed by model $j$ and model $k$.

Model probabilities associated with the predictive likelihood are calculated as

$$p(M_i|\tilde{y}_u, \ldots, \tilde{y}_v, \Phi_{u-1})$$
$$= \frac{p(\tilde{y}_u, \ldots, \tilde{y}_v|M_i, \Phi_{u-1})p(M_i|\Phi_{u-1})}{\sum_{k=1}^{K} p(\tilde{y}_u, \ldots, \tilde{y}_v|M_k, \Phi_{u-1})p(M_k|\Phi_{u-1})}, \quad i = 1, \ldots, K, \tag{5.5}$$

where there are $K$ models and $M_k$ denotes model $k$. In all calculations equal prior model probabilities are used. At the end of the training sample we start

with an equal prior on the models: $p(M_i|\Phi_{u-1}) = P(M_i)$, with $P(\text{SV}) = P(\text{GARCH}) = P(\text{Jump}) = 1/3$. Models with a high predictive likelihood will be assigned a high model probability. Calculating (5.5) for each $v = u + 1$, $\ldots, T$ provides a cumulative assessment of the evidence for model $i$ as more observations are used. The model probabilities are used to calculate the model average in the out-of-sample period. For example, the Bayesian model average of $g(y_{v+1})$ using (5.5) is

$$E\big[g(y_{v+1})|\tilde{y}_u, \ldots, \tilde{y}_v, \Phi_{u-1}\big]$$
$$= \sum_{k=1}^{K} E\big[g(y_{v+1})|\tilde{y}_u, \ldots, \tilde{y}_v, \Phi_{u-1}, M_k\big]p(M_k|\tilde{y}_u, \ldots, \tilde{y}_v, \Phi_{u-1}). \tag{5.6}$$

In addition to the predictive likelihood which focuses on the whole density, we also compare models by one-step ahead out-of-sample variance forecasts. The predictive variance from (5.1) is computed by simulation methods and detailed below. The Bayesian model average for the predictive variance can be computed using (5.6) as in $E[y_{v+1}^2|\Phi_v] - E[y_{v+1}|\Phi_v]^2$. We report mean squared error (MSE), mean absolute error (MAE), and Mincer and Zarnowitz (1969) forecast regressions for each model's predictive variance against realized volatility. Realized volatility serves as our target ex post volatility measure and is discussed below.

### 5.1. Calculations

#### 5.1.1. Jump model

The predictive likelihood can be estimated from the MCMC output. For the Jump model, with $y_{t+1} = r_{t+1}$, and $\Theta = (\theta, \omega)$, the predictive likelihood is

$$\hat{p}_{t+1} = \int p(r_{t+1}|\theta, \omega_t, \Phi_t)p(\theta, \omega_t|\Phi_t)\,d\theta\,d\omega_t \tag{5.7}$$

$$= \int p(r_{t+1}|\theta, \omega_{t+1}, \Phi_t)p(\omega_{t+1}|\omega_t, \theta)$$
$$\times p(\theta, \omega_t|\Phi_t)\,d\theta\,d\omega_t\,d\omega_{t+1} \tag{5.8}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N}\frac{1}{R}\sum_{j=1}^{R}p\big(r_{t+1}|\theta^{(i)}, \omega_{t+1}^{(j)}, \omega_t^{(i)}, \Phi_t\big), \tag{5.9}$$

where $i$ denotes the $i$th draw from the posterior $p(\theta, \omega|\Phi_t)$, $i = 1, \ldots, N$, $j$ indexes simulated values of $\omega_{t+1}$, and

$$p\big(r_{t+1}|\theta^{(i)}, \omega_{t+1}^{(j)}, \omega_t^{(i)}, \Phi_t\big)$$
$$= \lambda_{t+1}^{(j)}\phi\big(r_{t+1}|\mu^{(i)} + \mu_\xi^{(i)}, \sigma^{2(i)} + \sigma_{J,t+1}^{2(i)}\big) \tag{5.10}$$
$$+ \big(1 - \lambda_{t+1}^{(j)}\big)\phi\big(r_{t+1}|\mu^{(i)}, \sigma^{2(i)}\big), \tag{5.11}$$

$$\lambda_{t+1}^{(j)} = \frac{\exp(\omega_{t+1}^{(j)})}{1 + \exp(\omega_{t+1}^{(j)})}, \tag{5.12}$$

$$\omega_{t+1}^{(j)} = \gamma_0^{(i)} + \gamma_1^{(i)}\omega_t^{(i)} + \epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0, 1). \tag{5.13}$$

$\phi(x|\mu, \sigma^2)$ denotes the normal density function evaluated at $x$ with mean $\mu$ and variance $\sigma^2$. The following steps summarize the estimation of $\hat{p}_t^{t+1}$:

1. Set model parameters to the $i$th draw from the posterior $\{\theta^{(i)}, \omega^{(i)}\}$.
2. Generate $j = 1, \dots, R$, values of $\omega_{t+1}^{(j)}$ according to (5.13) and calculate the average of $p(r_{t+1}|\theta^{(i)}, \omega_{t+1}^{(j)}, \omega_t^{(i)}, \Phi_t)$ for these values. Save the result.
3. If $i < N$ then set $i = i + 1$ and go to 1.
4. Calculate the average of the $N$ values obtained in step 2.

Standard errors for this estimate can be calculated as usual from the MCMC output. Note that the measure of accuracy should account for the fact that the posterior draws are from a Markov chain and display dependence. Following Geweke (1992), we calculate the long-run variance of the terms in 2 using the Newey–West variance–covariance estimator. The square-root of this divided by the sample size is the numerical standard error. The delta method is used to convert this to a standard error for $\log(\hat{p}_{t+1})$. From this a numerical standard error can be derived for the log of (5.3) assuming all estimates are independent.

### 5.1.2. SV model

Similar calculations are used for the SV model:

$$\hat{p}_{t+1} = \int p(r_{t+1}|\theta, h_t, \Phi_t)p(\theta, h_t|\Phi_t)\, d\theta\, dh_t \tag{5.14}$$

$$= \int p(r_{t+1}|\theta, h_{t+1}, \Phi_t)p(h_{t+1}|h_t, \theta)$$

$$\times p(\theta, h_t|\Phi_t)\, d\theta\, dh_t\, dh_{t+1} \tag{5.15}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N}\frac{1}{R}\sum_{j=1}^{R} p(r_{t+1}|\theta^{(i)}, h_{t+1}^{(j)}, h_t^{(i)}, \Phi_t), \tag{5.16}$$

where $\theta^{(i)} = \{\mu^{(i)}, \rho_0^{(i)}, \rho_1^{(i)}, \sigma_v^{2(i)}\}$. To summarize,

1. Set model parameters to the $i$th draw from the posterior $\{\theta^{(i)}, h^{(i)}\}$, given $\Phi_t$.
2. Generate $j = 1, \dots, R$, values of $h_{t+1}^{(j)}$ according to

$$h_{t+1}^{(j)} = \rho_0^{(i)} + \rho_1^{(i)}h_t^{(i)} + \sigma_v^{(i)}v_{t+1}, \quad v_{t+1} \sim N(0, 1), \tag{5.17}$$

and calculate the average of $p(r_{t+1}|\theta^i, h_{t+1}^{(j)}, h_t^{(i)}, \Phi_t) = \phi(r_{t+1}|\mu^{(i)}, \exp(h_{t+1}^{(j)}))$ for these values. Save the result.
3. If $i < N$ then set $i = i + 1$ and go to 1.
4. Calculate the average of the $N$ values obtained in step 2.

### 5.1.3. GARCH model

Unlike the other models, the GARCH specification does not have latent variables, which simplifies calculation of the predictive likelihood. If $\theta = \{\mu, \kappa, \alpha, \beta\}$, then

$$p_{t+1} = \int p(r_{t+1}|\theta, \Phi_t)p(\theta|\Phi_t)\, d\theta \tag{5.18}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} p(r_{t+1}|\theta^{(i)}, \Phi_t), \tag{5.19}$$

where $\theta^{(i)}$ is a draw from the posterior $p(\theta|\Phi_t)$ and $p(r_{t+1}|\theta^{(i)}, \Phi_t) = \phi(r_{t+1}|\mu^{(i)}, \sigma_{t+1}^{2(i)})$. Note that the full series of conditional variances must be computed starting at $t = 1$ to obtain $\sigma_{t+1}^{2(i)} = \kappa^{(i)} + \alpha^{(i)}\epsilon_t^2 + \beta^{(i)}\sigma_t^{2(i)}$, with $\epsilon_t = r_t - \mu^{(i)}$.

## 6. Volatility forecasts

Consider a generic model with parameter vector $\Theta$. Moments of $y_{t+1}$ (assuming they exist), based on time $t$ information, can be calculated from the predictive density as

$$E[y_{t+1}^s|\Phi_t] = \int y_{t+1}^s p(y_{t+1}|\Phi_t)\, dy_{t+1} \tag{6.1}$$

$$= \int y_{t+1}^s \int p(y_{t+1}|\Phi_t, \Theta)p(\Theta|\Phi_t)\, d\Theta\, dy_{t+1} \tag{6.2}$$

$$= \int E[y_{t+1}^s|\Phi_t, \Theta]p(\Theta|\Phi_t)\, d\Theta, \quad s = 1, 2, \dots, \tag{6.3}$$

which can be approximated from the MCMC output. Note that the posterior $p(\Theta|\Phi_t)$ was used in the last section for calculating the predictive likelihood. Therefore, very little additional computation is needed to obtain out-of-sample forecasts. Conditional variance forecasts are $E[y_{t+1}^2|\Phi_t] - E[y_{t+1}|\Phi_t]^2$.

### 6.1. Calculations

#### 6.1.1. Jump model

$$E[r_{t+1}^s|\Phi_t] = \int E[r_{t+1}^s|\Phi_t, \theta, \omega_t]p(\theta, \omega_t|\Phi_t)\, d\theta\, d\omega_t \tag{6.4}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} E[r_{t+1}^s|\Phi_t, \theta^{(i)}, \omega_t^{(i)}], \quad s = 1, 2, \tag{6.5}$$

where $\{\theta^{(i)}, \omega_t^{(i)}\}$ is the $i$th draw from the posterior distribution, $p(\theta, \omega|\Phi_t)$. For the variance we require the first two moments:

$$E\left[r_{t+1}|\Phi_t, \theta^i, \omega_t^i\right] = \mu^{(i)} + \mu_\xi^{(i)} E\left[J_{t+1}|\theta^{(i)}, \omega_t^{(i)}\right], \tag{6.6}$$

$$E\left[r_{t+1}^2|\Phi_t, \theta^{(i)}, \omega_t^{(i)}\right]$$

$$= \mu^{(i)2} + \sigma^{(i)2} + \left(\mu_\xi^{(i)2} + \sigma_J^{(i)2}\right) E\left[J_{t+1}|\theta^{(i)}, \omega_t^{(i)}\right] \tag{6.7}$$

$$+ 2\mu^{(i)}\mu_\xi^{(i)} E\left[J_{t+1}|\theta^{(i)}, \omega_t^{(i)}\right]. \tag{6.8}$$

These moments are substituted into (6.5) for each draw of $\{\theta^{(i)}, \omega_t^{(i)}\}$. Since $\lambda_{t+1}$ depends on $\omega_{t+1}$ we approximate each conditional expectation of $J_{t+1}$ as

$$E\left[J_{t+1}|\theta^{(i)}, \omega_t^{(i)}\right] \approx \frac{1}{R} \sum_{j=1}^{R} \frac{\exp(\omega_{t+1}^{(j)})}{1 + \exp(\omega_{t+1}^{(j)})}, \tag{6.9}$$

where $\omega_{t+1}^{(j)}$ is generated from

$$\omega_{t+1}^{(j)} = \gamma_0^{(i)} + \gamma_1^{(i)}\omega_t^{(i)} + \epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0,1), \; j = 1, \ldots, R. \tag{6.10}$$

### 6.1.2. SV model

For the SV model, we have

$$E\left[r_{t+1}^s|\Phi_t\right] = \int E\left[r_{t+1}^s|\Phi_t, \theta, h_t\right] p(\theta, h_t|\Phi_t) \, d\theta \, dh_t \tag{6.11}$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} E\left[r_{t+1}^s|\Phi_t, \theta^{(i)}, h_t^{(i)}\right], \quad s = 1, 2. \tag{6.12}$$

The conditional moments are

$$E\left[r_{t+1}|\Phi_t, \theta^{(i)}, h_t^{(i)}\right] = \mu^{(i)}, \tag{6.13}$$

$$E\left[r_{t+1}^2|\Phi_t, \theta^{(i)}, h_t^{(i)}\right] = \mu^{(i)2} + \exp\left(\rho_0^{(i)} + \rho_1^{(i)}h_t^{(i)} + \sigma_v^{(i)2}/2\right). \tag{6.14}$$

### 6.1.3. GARCH model

The results for the SV model can be used for the GARCH model, except we replace the conditional moments with

$$E\left[r_{t+1}|\Phi_t, \theta^{(i)}\right] = \mu^{(i)}, \tag{6.15}$$

$$E\left[r_{t+1}^2|\Phi_t, \theta^{(i)}\right] = \mu^{(i)2} + \sigma_{t+1}^{2(i)}. \tag{6.16}$$

## 7. Data

Five-minute intraday FX quote data (bid and ask) were obtained from Olsen and Associates for every day from 1986/12/16–2002/12/31 for the Japanese Yen–US Dollar (JPY–USD) exchange rate. With a few exceptions, the construction

of daily returns and realized volatility closely follow Andersen *et al.* (2001) and Maheu and McCurdy (2002).

Currencies trade 24 hours a day, 7 days a week. The raw data included a number of missing observations. Missing quotes on the 5-minute grid where linearly interpolated by Olsen and Associates from the nearest available quote. We use the midpoint of the bid and ask quote, giving us 1.8 million 5-minute price observations. A day was defined as beginning at 00:05 GMT and ending 24:00 GMT. Continuously compounded 5-minute returns (in percent, that is, scaled by 100) were constructed from the price data. Following Andersen *et al.* (2001), all weekends (Saturday and Sunday) were removed as well as the following slow trading days: December 24–26, 31 and January 1, 2. In addition, the moving holidays (Good Friday, Easter Monday, Memorial Day, July Fourth, Labor Day, Thanksgiving and the day after), as well as any days in which more than half (144) of the day's quotes were missing, were removed. The remaining 5-minute returns data were linearly filtered by an MA($q$) to remove high-frequency autocorrelation which may be due to the discrete nature of bid/ask quotes and market microstructure effects. We set $q = 4$ based on the sample autocorrelations of the 5-minute returns.

From these filtered data, daily realized volatilities were constructed as

$$\text{RV}_t = \sum_{j=1}^{288} r_{t,j}^2, \tag{7.1}$$

where $r_{t,j}$ is the $j$th 5-minute return in day $t$. Daily returns were constructed as the sum of the intraday 5-minute returns, that is, $r_t = \sum_{j=1}^{288} r_{t,j}$.

In a series of papers, Andersen, Bollerslev, Diebold and co-authors and Barndorff-Nielson and Shephard, among others, have documented various features of realized volatility and derived its asymptotic properties as the sampling frequency of prices increases over a fixed-time interval. Reviews of this large literature include Andersen *et al.* (2005) and Barndorff-Nielsen *et al.* (2004). For our purpose, we note that $\text{RV}_t$ serves as a natural ex post estimate of volatility and provides a superior measure to rank the forecasts of competing models. Therefore, we treat $\text{RV}_t$ as the ideal target for conditional variance forecasts from our alternative models.

Table 1 reports our summary statistics for daily returns and realized volatility.

## 8. Results

Our daily data for JPY–USD covers 1986/12/16 to 2002/12/31 (4001 observations). Table 2 contains full-sample model estimates. For each model a total of 90 000 MCMC iterations were performed. The first 10 000 draws were discarded to minimize the influence of starting values. Thus, $N = 80\,000$ samples from the posterior distribution are used to calculate posterior moments.

Figure 1 shows various features of the Jump model, while Figures 2 and 3 display characteristics of the out-of-sample performance for the SV and jump

**Table 1.**   *Summary statistics, JPY–USD*

| Statistic | $r_t$ | $RV_t$ |
|---|---|---|
| Mean | −4.566e−6 | 0.5992 |
| Variance | 0.6352 | 0.6264 |
| Skewness | −0.3504 | 21.6655 |
| Kurtosis | 8.0587 | 854.2149 |
| Minimum | −6.9768 | 0.0290 |
| Maximum | 5.9780 | 34.3872 |
| Obs | 4001 | 4001 |

*Note*: $r_t$ is percent log differences of daily spot exchange rates and $RV_t$ is realized volatility for the period 1986/12/16–2002/12/31.

specifications, for the JPY–USD data. Out-of-sample calculations condition on a training sample of the first 3000 observations. These observations provide accurate parameter estimates for all models and minimize parameter uncertainty. Therefore, predictive likelihood estimates and model forecasts, appearing in Tables 3 and 4, are based on the remaining 1001 observations. Predictive likelihood estimates are illustrated in Figure 2.

Table 2 contains estimates of the standard SV and GARCH models, as well as our heterogeneous jump specification. We report various features of the posterior distribution. Estimates of the SV specification in Table 2 show latent volatility to be persistent. All of the parameters are accurately estimated. The SV model implies an unconditional variance of 0.617 which is close to the value in Table 1. Similarly, $\alpha + \beta$ in the GARCH model is about 0.96, and the implied unconditional variance is 0.723.

Based on the posterior mean, the unconditional expectation for the autoregressive latent variable, $\omega_t$, driving jump arrival is −2.62. This implies that jumps are infrequent on average. For instance, the empirical average of $\lambda_t$ is 0.07. This is consistent with previous studies which combine SV with simple jumps. However, in our case, this jump probability shows clear time dependencies. From the model estimates, it is seen that the process governing jump arrival is very persistent. A 95% density interval for $\gamma_1$ is (0.937, 0.971).

The posterior mean of $\eta_1$ is 0.4532 which indicates that lagged absolute returns are important in affecting the jump-size variance. Notice that the jump-size variance is 3–4 times larger than the normal innovation variance $\sigma^2$.

Features of the Jump model are displayed in Figure 1. Panel (a) shows returns, panel (b) plots the inferred jump probability $\lambda_t$, panel (c) illustrates the jump size and panel (d) displays estimated jump times over the full sample. Clearly, this model identifies large moves in returns as jumps. Panels (c) and (d) suggest that some jumps are isolated events while others cluster and lead to more jumps and higher volatility. This is an important feature of the model.

Predictive likelihood estimates appear in Table 3. The estimates were obtained by re-estimating the posterior for each observation from 3001 on. For

**Table 2.**   *Model estimates*

A. *SV model*

$$r_t = \mu + \exp(h_t/2)z_t, \quad z_t \sim N(0, 1)$$
$$h_t = \rho_0 + \rho_1 h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0, 1), \ |\rho_1| < 1$$

| | Mean | St. dev. | Median | 95% density interval |
|---|---|---|---|---|
| $\mu$ | 0.0165 | 0.0105 | 0.0165 | (−0.0008, 0.0337) |
| $\rho_0$ | −0.0568 | 0.0125 | −0.0559 | (−0.0786, −0.0379) |
| $\rho_1$ | 0.9234 | 0.0148 | 0.9244 | (0.8974, 0.9460) |
| $\sigma_v^2$ | 0.0763 | 0.0160 | 0.0749 | (0.0524, 0.1043) |

B. *GARCH model*

$$r_t = \mu + \epsilon_t, \quad \epsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1)$$
$$\sigma_t^2 = \kappa + \alpha \epsilon_{t-1}^2 \beta \sigma_{t-1}^2$$

| | Mean | St. dev. | Median | 95% density interval |
|---|---|---|---|---|
| $\mu$ | 0.0083 | 0.0123 | 0.0083 | (−0.0140, 0.0338) |
| $\kappa$ | 0.0305 | 0.0064 | 0.0302 | (0.0184, 0.0431) |
| $\alpha$ | 0.1031 | 0.0143 | 0.1022 | (0.0784, 0.1342) |
| $\beta$ | 0.8547 | 0.0198 | 0.8555 | (0.8144, 0.8919) |

C. *Jump model*

$$r_t = \mu + \sigma z_t + J_t \xi_t, \quad z_t \sim N(0, 1)$$
$$\xi_t \sim N(\mu_\xi, \sigma_{\xi,t}^2), \quad J_t \in \{0, 1\}$$
$$P(J_t = 1|\omega_t) = \lambda_t \quad \text{and} \quad P(J_t = 0|\omega_t) = 1 - \lambda_t$$
$$\lambda_t = \frac{\exp(\omega_t)}{1 + \exp(\omega_t)}$$
$$\omega_t = \gamma_0 + \gamma_1 \omega_{t-1} + u_t, \quad u_t \sim N(0, 1), \ |\gamma_1| < 1$$
$$\sigma_{\xi,t}^2 = \eta_0 + \eta_1 X_{t-1}$$

| | Mean | St. dev. | Median | 95% density interval |
|---|---|---|---|---|
| $\mu$ | 0.0307 | 0.0125 | 0.0307 | (0.0103, 0.0513) |
| $\sigma^2$ | 0.3226 | 0.0159 | 0.3223 | (0.2968, 0.3491) |
| $\mu_\xi$ | −0.1272 | 0.0510 | −0.1269 | (−0.2115, −0.0440) |
| $\eta_0$ | 0.9089 | 0.1238 | 0.8961 | (0.7273, 1.1276) |
| $\eta_1$ | 0.4532 | 0.1153 | 0.4443 | (0.2793, 0.6599) |
| $\gamma_0$ | −0.1169 | 0.0305 | −0.1158 | (−0.1688, −0.0686) |
| $\gamma_1$ | 0.9554 | 0.0103 | 0.9562 | (0.9372, 0.9709) |

*Notes*: The data are percent log differences of daily spot exchange rates for the JPY–USD from 1986/12/16–2002/12/31. Tables for the alternative models report posterior mean, standard deviation, median and 95% probability density interval for each parameter estimate.

each run, we collected $N = 20\,000$ (after discarding the first $10\,000$) samples from the posterior simulator and set $R = 100$. Kass and Raftery (1995) recommend considering twice the logarithm of the Bayes factor for model comparison, as it has the same scaling as the likelihood ratio statistic. They suggest the following interpretation of $2 \log \text{BF}$: 0 to 2 not worth more than a bare

### (a) Returns



### (b) Jump Probability
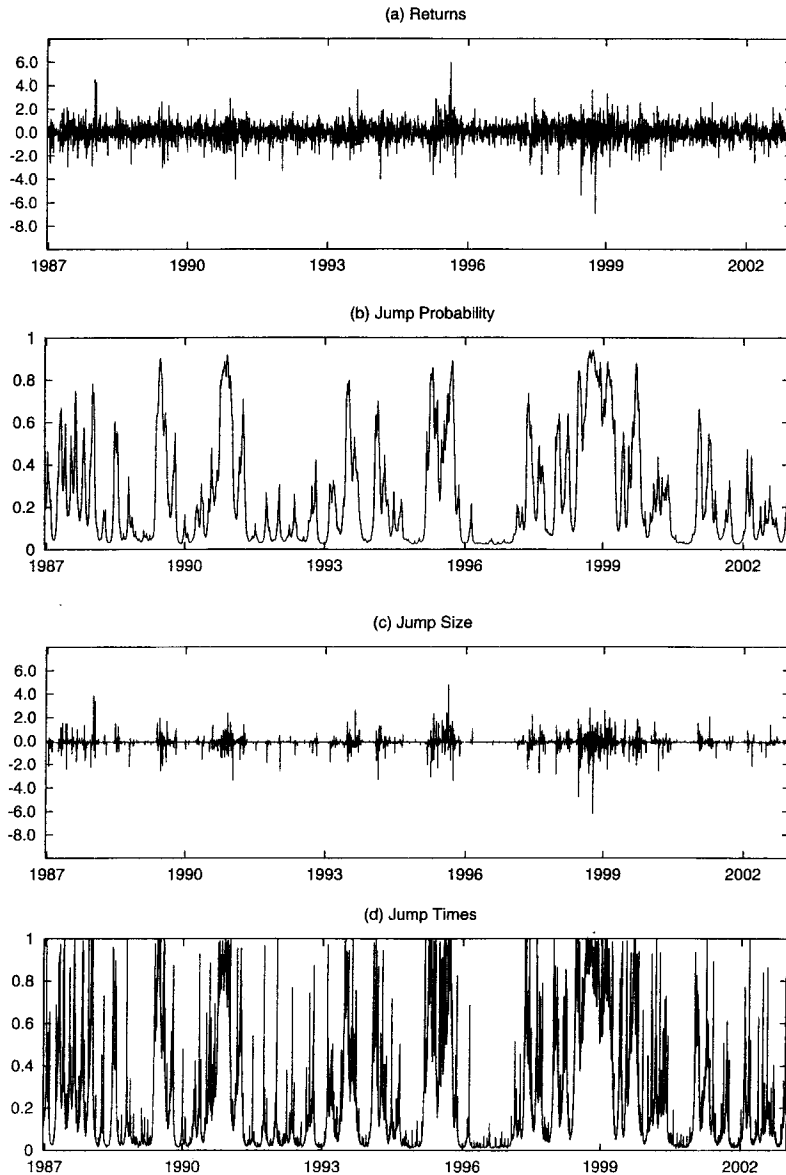


### (c) Jump Size



### (d) Jump Times



*Fig. 1.    Jump model.*

mention, 2 to 6 positive, 6 to 10 strong, and greater than 10 as very strong. Based on this, the evidence in favor of the Jump model is positive. For example, $2 \log \text{BF} = 2(-1103.760 + 1105.4758) = 3.431$ in favor of the Jump model
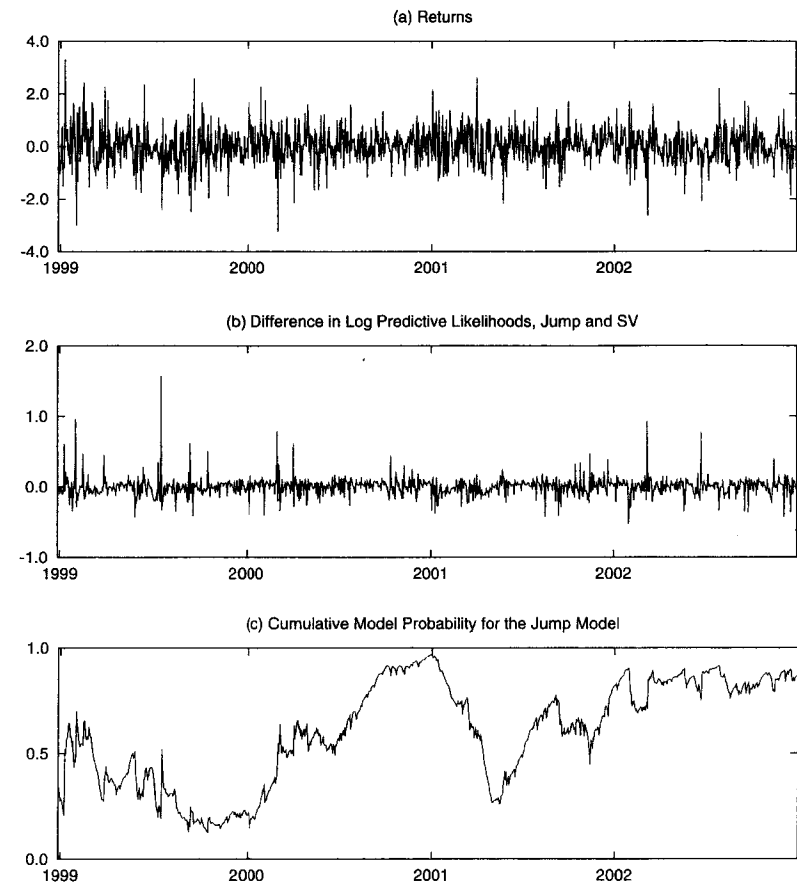
### (a) Returns



### (b) Difference in Log Predictive Likelihoods, Jump and SV



### (c) Cumulative Model Probability for the Jump Model



*Fig. 2.    Returns, predictive likelihoods, and cumulative model probabilities.*

versus the SV model. On the other hand, the GARCH model is dominated by the other specifications. Model probabilities based on the predictive likelihood estimates are calculated as in (5.5). The final cumulative model probability in favor of the jump specification is 0.85 while it is 0.15 for the SV model.

Differences in the log predictive likelihood in favor of the Jump model compared to the SV model are found in panel (b) of Figure 2. A positive (negative) value occurs when an observation is more likely under the Jump (SV) model. There are several influential observations favoring the Jump model that appear to be high volatility episodes. These influential observations show up as spikes in panel (b). The cumulative probability for the Jump model is plotted in panel (c) of Figure 2. It is interesting to note that there are several upward trending periods in panel (c) which are not from tail occurrences in returns, for instance from
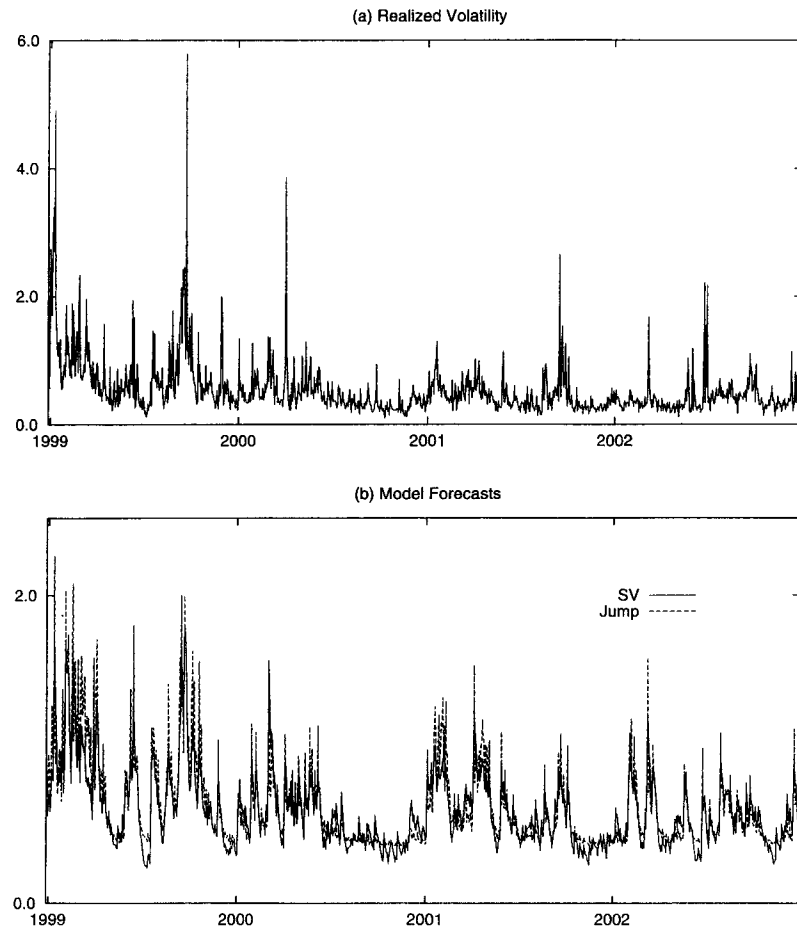
**(a) Realized Volatility**



**(b) Model Forecasts**



*Fig. 3. Volatility and model forecasts.*

the middle of 2000 to 2001. The SV model appears to perform best at the start of the sample, 1999–2000.

Although Bayes factors are an assessment of all distributional features, it is interesting to also evaluate volatility forecasts. Out-of-sample forecasts for all models and the model average are assessed using realized volatility. Table 3 contains Mincer and Zarnowitz (1969) forecast regressions of realized volatility regressed on a model's conditional variance forecast. We report regression $R^2$, mean squared error (MSE) and mean absolute error (MAE). Although the GARCH model does poorly in accounting for the predictive density of returns, it performs very well in one-period-ahead forecasts of realized volatility.

*Table 3. Log predictive likelihoods*

| Model | JPY–USD |
|---|---|
| GARCH | −1117.6374 |
|  | (0.0565) |
| SV | −1105.4758 |
|  | (0.2153) |
| Jump | −1103.7603 |
|  | (0.3982) |

*Notes*: This table reports estimates of log predictive likelihoods for observations 3001–4001. Numerical standard errors appear in parentheses.

*Table 4. Out-of-sample forecast performance for RV*

|  | $a$ | $b$ | $R^2$ | MSE | MAE |
|---|---|---|---|---|---|
| GARCH | −0.0244 | 0.9413 | 0.2471 | 0.1619 | 0.2443 |
|  | (0.0337) | 0.0520 |  |  |  |
| SV | 0.0024 | 0.8899 | 0.2345 | 0.1657 | 0.2505 |
|  | (0.0334) | (0.0509) |  |  |  |
| Jump | 0.0477 | 0.7829 | 0.2509 | 0.1694 | 0.2582 |
|  | (0.0298) | (0.0428) |  |  |  |
| Model average | 0.0141 | 0.8587 | 0.2467 | 0.1649 | 0.2511 |
|  | (0.0318) | (0.0475) |  |  |  |

*Notes*: This table reports Mincer and Zarnowitz (1969) forecast regressions of

$$RV_t = a + b\mathrm{Var}_{t-1}(r_t) + \mathrm{error}_t,$$

where $\mathrm{Var}_{t-1}(r_t)$ is a model forecast of the one-period-ahead conditional variance based on time $t-1$ information and $RV_t$ is realized volatility for day $t$. Standard errors appear in parentheses. $R^2$ is the coefficient of determination. MSE and MAE are the mean squared error and mean absolute error, respectively, for $(RV_t - \mathrm{Var}_{t-1}(r_t))$. The number of out-of-sample forecasts is 1001.

The Jump model has the largest $R^2$ although the differences in the models is small.

Panel (a) of Figure 3 displays realized volatility for the out-of-sample period. Panel (b) plots the model forecasts for SV and the Jump model. Notice that the model forecasts are much less variable than realized volatility. During high volatility periods both models produce similar forecasts, but during low periods the Jump model's variance is lower. During these times the Jump variance is essentially flat, a time-series pattern similar to that of realized volatility, while the SV forecast appears to be trending upward. This is due to the fact that the Jump model can have periods of homoskedasticity while the SV model always has a time-varying variance.

The differences in the volatility forecasts become clear when we compare Figure 3(b) to Figure 2(c). Periods when the Jump model's variance is low are exactly when the probability for the Jump model is increasing (just before 2001 and around 2002). This suggests that state-dependent volatility clustering may be important, that is, periods of normal homoskedasticity along with periods of high volatility that clusters. The Jump model is able to capture these dynamics relatively well.

Overall the Jump model is quite competitive with a simple parameterization of SV and GARCH. Further, out-of-sample predictive likelihoods reveal periods when the heteroskedastic jump structure appears to do a better job at capturing the dynamics of realized volatility.

There are several potential extensions to our heterogeneous Jump model. For example, we could consider a fat-tailed distribution for return innovations and for jump-size innovations. We have allowed the conditional variance of the jump-size distribution to be a function of the absolute value of lagged returns to reflect recent volatility levels. Other conditioning variables that could be used are the range statistic or realized volatility.

## 9. Conclusions

This chapter proposes a new discrete-time model of returns in which jumps capture persistence and time variation in the conditional variance and higher-order moments of returns. Time-varying jump arrival is governed by a latent autoregressive process and the jump-size distribution is allowed to be conditionally heteroskedastic with a non-zero mean.

We discuss a Bayesian approach to estimation and advocate the use of predictive likelihoods to compare alternative models. We also compare the out-of-sample volatility forecasts against realized volatility. In particular, using JPY–USD exchange rate returns, we compare forecasts from our heterogeneous Jump model, a conventional stochastic volatility model, a GARCH model and the model average.

Our results indicate that the heterogeneous Jump model effectively captures volatility persistence through jump clustering. In addition, we find that the jump-size variance is heteroskedastic and increasing in volatile markets. Overall, this model is very competitive with the stochastic volatility specification. The state dependence in jumps allows it to perform well in both quiet periods and in volatile periods when tail realizations occur. We conclude that the modeling of jump dynamics to describe market returns is a fruitful avenue for future research.

## Acknowledgements

## Appendix A

Below we provide the details of the posterior simulator for the Heterogeneous Jump specification and the SV model.

### A.1 Jump model

For steps 1–3 and 6 we use standard conjugate results for the linear regression model (see Koop, 2003).

1. $\mu | \theta_{-\mu}, \omega, J, \xi, r$. If the prior is normal, $p(\mu) \sim N(a, A^{-1})$,

$$p(\mu | \theta_{-\mu}, \omega, J, \xi, r) \propto p(r | \theta, \omega, J, \xi) p(\mu) \qquad (A.1)$$

$$\sim N(m, V^{-1}), \qquad (A.2)$$

where $V = \sigma^{-2} T + A$ and $m = V^{-1}(\sigma^{-2} \sum_{t=1}^{T}(r_t - \xi_t J_t) + Aa)$.

2. $\sigma^2 | \theta_{-\sigma^2}, \omega, J, \xi, r$. With an inverse gamma prior, $\sigma^2 \sim IG(v_{\sigma^2}/2, s_{\sigma^2}/2)$, we have

$$p(\sigma^2 | \theta_{-\sigma^2}, \omega, J, \xi, r)$$

$$\propto p(r | \theta, \omega, J, \xi) p(\sigma^2) \qquad (A.3)$$

$$\sim IG\left(\frac{T + v_{\sigma^2}}{2}, \frac{\sum_{t=1}^{T}(r_t - \xi_t J_t - \mu)^2 + s_{\sigma^2}}{2}\right). \qquad (A.4)$$

3. $\mu_\xi | \theta_{-\mu_\xi}, \omega, J, \xi, r$. If $p(\mu_\xi) \sim N(b, B^{-1})$

$$p(\mu_\xi | \theta_{-\mu_\xi}, \omega, J, \xi, r) \propto p(\xi | \theta) p(\mu_\xi) \qquad (A.5)$$

$$\sim N(m, V^{-1}), \qquad (A.6)$$

where $V = \sigma_J^{-2} T + B$ and $m = V^{-1}(\sigma_J^{-2} \sum_{t=1}^{T} \xi_t + Bb)$.

4. $\eta | \theta_{-\eta}, \omega, J, \xi, r$. We use independent inverse gamma priors for both parameters, $p(\eta_0) \sim IG(v_{\eta_0}/2, s_{\eta_0}/2)$, $p(\eta_1) \sim IG(v_{\eta_1}/2, s_{\eta_1}/2)$. This ensures that the jump-size variance is always positive. The target distribution is

$$p(\eta | \theta_{-\eta}, \omega, J, \xi, r) \propto p(\xi | \theta) p(\eta_0) p(\eta_1)$$

$$\propto p(\eta_0) p(\eta_1) \prod_{t=1}^{T} (\eta_0 + \eta_1 X_{t-1})^{-1/2}$$

$$\times \exp\left(-\frac{1}{2} \frac{(\xi_t - \mu_\xi)^2}{(\eta_0 + \eta_1 X_{t-1})}\right), \qquad (A.7)$$

which is a nonstandard distribution. We use a MH algorithm to sample this parameter using a random walk proposal. The proposal distribution

$q(x|\eta^{i-1})$, is a fat-tailed mixture of normals where the covariance matrix is calibrated so approximately 50% of candidate draws are accepted. The mixture specification is the same as what is used in step 5 below. If $\eta^i \sim q(x|\eta^{i-1})$ is a draw from the proposal distribution, it is accepted with probability

$$\min\left\{\frac{p(\eta^i|\theta_{-\eta}, \omega, J, \xi, r)}{p(\eta^{i-1}|\theta_{-\eta}, \omega, J, \xi, r)}, 1\right\} \tag{A.8}$$

and otherwise $\eta^i = \eta^{i-1}$.

5. $\omega_{(t,\tau)}|\theta, \omega_{-(t,\tau)}, J, \xi, r, t < \tau, t = 1, \dots, T$. The conditional posterior is

$$p(\omega_{(t,\tau)}|\theta, \omega_{-(t,\tau)}, J)$$

$$\propto p(\tilde{J}_{(t,\tau)}|\omega_{(t,\tau)})p(\omega_{(t,\tau)}|\omega_{-(t,\tau)}, \theta) \tag{A.9}$$

$$\propto p(\omega_{\tau+1}|\omega_\tau, \theta)\prod_{i=t}^{\tau}p(J_t|\omega_t)p(\omega_t|\omega_{t-1}, \theta), \tag{A.10}$$

which is a nonstandard distribution. To improve the mixing properties of the MCMC output we adapted the blocking procedure that Fleming and Kirby (2003) use for SV models. Specifically, we approximate the first-difference of $\omega_t$ as a constant over the interval $(t, \tau)$, and sample a block using an independent Metropolis routine. If $\omega_t$ is strongly autocorrelated this provides a good proposal density. The proposal distribution is a fat-tailed multivariate mixture of normals,[1]

$$q(x|\omega_{-(t,\tau)}) \sim \begin{cases} N(m, V) & \text{with probability } p, \\ N(m, 10V) & \text{with probability } 1-p, \end{cases} \tag{A.11}$$

where $p = 0.9$. The mean vector and variance–covariance matrix are

$$m_l = \omega_{t-1} + \frac{l}{k+1}(\omega_{t-1} + \omega_{\tau+1}), \quad l = 1, \dots, k, \tag{A.12}$$

$$V_{lm} = \min(l, m) - \frac{lm}{k+1}, \quad l = 1, \dots, k, \; m = 1, \dots, k, \tag{A.13}$$

where $k = \tau - t + 1$ is the block length and is chosen randomly from a Poisson distribution with parameter 15. $V^{-1}$ has a very convenient tridiagonal form in which $V_{ii}^{-1} = 2$ and $V_{ij}^{-1} = -1, 1 \leqslant i \leqslant k, j = i-1, i+1$, and otherwise 0. A new draw $\omega_{(t,\tau)}^{(i)}$ from $q(x|\omega_{-(t,\tau)}^{(i-1)})$, is accepted with probability

$$\min\left\{\frac{p(\omega_{(t,\tau)}^{(i)}|\theta, \omega_{-(t,\tau)}^{(i-1)}, J)/q(\omega_{(t,\tau)}^{(i)}|\omega_{-(t,\tau)}^{(i-1)})}{p(\omega_{(t,\tau)}^{(i-1)}|\theta, \omega_{-(t,\tau)}^{(i-1)}, J)/q(\omega_{(t,\tau)}^{(i-1)}|\omega_{-(t,\tau)}^{(i-1)})}, 1\right\} \tag{A.14}$$

and otherwise $\omega_{(t,\tau)}^{(i)} = \omega_{(t,\tau)}^{(i-1)}$.

---

[1] In practice, a fat-tailed proposal was found to be important when comparing the accuracy of the block version with a single move sampler.

6. $\gamma|\theta_{-\gamma}, \omega, J, \xi, r$. The conjugate prior for $\gamma$ is a bivariate normal $p(\gamma) \sim N(m, V)$ which results in a conditional distribution that is also normal. We reject any draw which does not satisfy $|\gamma_1| < 1$.

7. $\xi|\theta, \omega, J, r$. The jump size can be sampled in a block by using the conditional independence of each $\xi_t$. The conditional distribution of $\xi_t$ is

$$p(\xi_t|\theta, \omega, J, r) \propto p(r_t|\theta, \omega_t, J_t, \xi_t)p(\xi_t|\theta) \sim N(c, C^{-1}), \tag{A.15}$$

where $C = \sigma^{-2}J_t + \sigma_{\xi,t}^{-2}$ and $c = C^{-1}(\sigma^{-2}J_t(r_t - \mu) + \sigma_{\xi,t}^{-2}\mu_\xi)$.

8. $J|\theta, \omega, \xi, r$. $J_t$ is a Bernoulli random variable with parameter $\lambda_t$. To find the probability of $J_t = 0, 1$ note that

$$p(J_t = 0|\theta, \omega, \xi, r) \propto p(r_t|\theta, J_t, \xi_t)p(J_t = 0|\omega)$$

$$\propto \exp\left(-0.5\sigma^{-2}(r_t - \mu)^2\right)(1 - \lambda_t)$$

$$p(J_t = 1|\theta, \omega, \xi, r) \propto p(r_t|\theta, J_t, \xi_t)p(J_t = 1|\omega)$$

$$\propto \exp\left(-0.5\sigma^{-2}(r_t - \mu - \xi_t)^2\right)\lambda_t,$$

which allows calculation of the normalizing constant and hence a draw of $J_t$.

9. Go to 1.

The priors used for this model are independent $\mu \sim N(0, 1000)$, $\mu_\xi \sim N(0, 100)$, $\sigma^2 \sim IG(3, 0.02)$, $\eta_0 \sim IG(2.5, 1)$, $\eta_1 \sim IG(2.5, 1)$, $\gamma_0 \sim N(0, 100)$ and $\gamma_1 \sim N(0, 100)I_{|\gamma_1|<1}$. The priors selected for this model are for the most part non-informative. Note that $\omega_t$ is restricted to be stationary. The priors on $\eta_0$ and $\eta_1$ reflect a reasonable range for the conditional variance of the jump-size distribution. For example, the probability that these parameters lie in the interval $(0.1, 2)$ is approximately 0.96.

### A.2 SV model

For the SV model we obtain draws from the full posterior iterating on the following steps:

Sample $\mu \sim p(\mu|\rho_0, \rho_1, \sigma_v^2, r, h)$
Sample $\rho_0, \rho_1 \sim p(\rho_0, \rho_1|\mu, \sigma_v^2, r, h)$
Sample $\sigma_v^2 \sim p(\sigma_v^2|\mu, \rho_0, \rho_1, r, h)$
Sample $h_t \sim p(h_t|h_{-t}, \mu, \rho_0, \rho_1, \sigma_v^2, r), t = 1, \dots, T$.

The priors are independent $\mu \sim N(0, 100)$, $\rho_0 \sim N(0, 100)$, $\rho_1 \sim N(0, 100) \times I_{|\rho_1|<1}$ and $\sigma_v^2 \sim IG(5/2, 0.1/2)$.

1. $\mu|\rho_0, \rho_1, \sigma_v, r, h$. Note that since $h$ is known the above model can be rewritten as

$$r_t \exp(-h_t/2) = \mu \exp(-h_t/2) + v_t, \quad t = 1, \dots, T. \tag{A.16}$$

This is a regression model $y_t = x_t\mu + v_t$, where $y_t = r_t \exp(-h_t/2)$, $x_t = \exp(-h_t/2)$ and the innovation term has variance 1. Hence we can apply Gibbs sampling results for the linear regression model.

2. $\rho_0, \rho_1 | \mu, \sigma_v, r, h$. The conjugate prior for $\rho_0, \rho_1$ is a bivariate normal which results in a conditional distribution that is also normal. We reject any draw which does not satisfy $|\rho_1| < 1$.

3. $\sigma_v^2 | \mu, \rho_0, \rho_1, r, h$. The inverse Gamma prior results in an inverse Gamma posterior.

4. $h_t | h_{-t}, \mu, \rho_0, \rho_1, \sigma_v^2, r, t = 1, \ldots, T$. We use the single move sampler of Kim *et al.* (1998) except that we use a MH step as opposed to accept/reject sampling. Here the proposal density is $N(\tilde{\mu}_t, \sigma^2)$, with $\tilde{\mu}_t = \mu_t + \frac{\sigma^2}{2}(r_t^2 \exp(-\mu_t) - 1)$, $\mu_t = \frac{\rho_0(1-\rho_1)+\rho_1(h_{t-1}+h_{t+1})}{1+\rho_1^2}$ and $\sigma^2 = \frac{\sigma_v^2}{1+\rho_1^2}$. Therefore sample $h' \sim N(\tilde{\mu}_t, \sigma^2)$, which has density $q(\cdot | h_{t-1}, h_{t+1}, \omega)$ and accept this with probability

$$\min\left\{1, \frac{p(h'_t | h_{t-1}, h_{t+1}, \omega)/q(h' | h_{t-1}, h_{t+1}, \omega)}{p(h_t | h_{t-1}, h_{t+1}, \omega)/q(h | h_{t-1}, h_{t+1}, \omega)}\right\} \qquad (A.17)$$

and otherwise reject.

## A.3 GARCH model

If $\theta$ denotes the GARCH model parameters then we use a multivariate random walk proposal in a MH step. Specifically, a new proposal is obtained as $\theta' = \theta + V$, where $V$ is a fat-tailed mixture of two multivariate normal densities. This proposal is accepted with probability $\min\{1, p(\theta' | \Phi_T)/p(\theta | \Phi_T)\}$; otherwise the new value of the chain is set to the previous draw $\theta$. The first time the model is estimated we calibrate the covariance of $V$ to be proportional to the posterior covariance matrix using output from a single-move random walk. Thereafter, we use the joint sampler with this covariance matrix. When a new observation becomes available, we use the previous posterior draws to calibrate the covariance matrix in the next round of estimation. This approach provides very efficient sampling and automatic updates to $V$. We collect 10 000 draws for posterior inference. The priors are $\mu \sim N(0, 100)$, $\kappa \sim N(0, 100)$, $\alpha \sim N(0, 100)$, $\beta \sim N(0, 100)$, $\kappa > 0$, $\alpha \geqslant 0$, $\beta \geqslant 0$ and $\alpha + \beta < 1$.

## *References*

Alizadeh, S., Brandt, M., Diebold, F.X. (2002), Range-based estimation of stochastic volatility models. *Journal of Finance* 57, 1047–1092.

Andersen, T.G., Sorensen, B. (1996), GMM estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Business & Economic Statistics* 14, 328–352.

Andersen, T.G., Bollerslev, T. (1998), Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39 (4), 885–905.

Andersen, T.G., Benzoni, L., Lund, J. (2002), An empirical investigation of continuous-time equity return models. *Journal of Finance* 62, 1239–1284.

Andersen, T.G., Bollerslev, T., Diebold, F.X. (2005), Parametric and nonparametric volatility measurement. In: Ait-Sahalia, Y., Hansen, L.P. (Eds.), *Handbook of Financial Econometrics*. North Holland, Amsterdam.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (2001), The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.

Barndorff-Nielsen, O.E., Graversen, S.E., Shephard, N. (2004), Power variation and stochastic volatility: a review and some new results. *Journal of Applied Probability* 41A, 133–143.

Bates, D.S. (2000), Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics* 94, 181–238.

Chernov, M., Gallant, R.A., Ghysels, E., Tauchen, G. (2003), Alternative models for stock price dynamics. *Journal of Econometrics* 116, 225–257.

Chib, S. (2001), Markov chain Monte Carlo methods: computation and inference. In: Heckman, X., Leamer, X. (Eds.), *Handbook of Econometrics*. Elsevier Science.

Chib, S., Nardari, F., Shephard, N. (2002), Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108, 281–316.

Danielsson, J., Richard, J.F. (1993), Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics* 8, S153–S173.

Das, S.R. (2002), The surprise element: jumps in interest rate. *Journal of Econometrics* 106, 27–65.

Engle, R.F. (1982), Autoregressive conditional heteroskedasticity with estimates of the UK inflation. *Econometrica* 50, 987–1008.

Eraker, B., Johannes, M.S., Polson, N.G. (2003), The impact of jumps in volatility and returns. *Journal of Finance* 58 (3), 1269–1300.

Fleming, J., Kirby, C. (2003), A closer look at the relationship between GARCH and stochastic autoregressive volatility. *Journal of Financial Econometrics* 1, 365–419.

Gallant, A.R., Hsieh, D., Tauchen, G. (1997), Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* 81, 159–192.

Geweke, J. (1992), Evaluating the accuracy of sampling based approaches to the calculation of posterior moments. In: Bernardo, X., Berger, X., Dawid, X., Smith, X. (Eds.), Bayesian Statistics, vol. 4. Clarendon Press, Oxford.

Geweke, J. (1994), Bayesian comparison of econometric models. Federal Reserve Bank of Minneapolis, Working Paper 532.

Geweke, J. (1997), Using simulation methods for Bayesian econometric models: Inference, development, and communication. *Econometric Reviews* 18 (1), 1–73.

Geweke, J., Whiteman, C. (2006), Bayesian forecasting. In: Elliot, G., Granger, C., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*. Elsevier, Amsterdam.

Ghysels, E., Harvey, A.C., Renault, E. (1996), *Stochastic Volatility*. Handbook of Statistics, vol. 14. Elsevier Science (Chapter 5).

Gordon, S. (1997), Stochastic trends, deterministic trends, and business cycle turning points. *Journal of Applied Econometrics* 12, 411–434.

Gourieroux, C., Monfort, A., Renault, E. (1993), Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.

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

Jacquier, E., Polson, N.G., Rossi, P.E. (2004), Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122, 185–212.

Johannes, M., Polson, N. (2006), MCMC methods for financial econometrics. In: Ait-Sahalia, Y., Hansen, L. (Eds.), *Handbook of Financial Econometrics*. Elsevier.

Johannes, M., Kumar, R., Polson, N.G. (1999), State dependent jump models: how do US equity indices jump? Working Paper, Graduate School of Business, University of Chicago.

Jorion, P. (1988), On jump processes in the foreign exchange and stock markets. *Review of Financial Studies* 1 (4), 427–445.

Kass, R.E., Raftery, A.E. (1995), Bayes factors. *Journal of the American Statistical Association* 90 (420), 773–795.

Kim, S., Shephard, N., Chib, S. (1998), Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.

Koop, G. (2003), *Bayesian Econometrics*. Wiley, Chichester, England.

Lin, S.J., Knight, J., Satchell, S. (1999), Modelling intra-day equity prices and volatility using information arrivals – a comparative study of different choices of information proxies. In: Lequeux, P. (Ed.), *Financial Markets Tick by Tick*. Wiley.

Maheu, J.M., McCurdy, T.H. (2002), Nonlinear features of FX realized volatility. *Review of Economics and Statistics* 84 (4), 668–681.

Maheu, J.M., McCurdy, T.H. (2004), News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance* 59 (2).

Melino, A., Turnbull, M. (1990), Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239–265.

Min, C., Zellner, A. (1993), Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics* 56, 89–118.

Mincer, J., Zarnowitz, V. (1969), The evaluation of economic forecasts and expectations. In: Mincer, X. (Ed.), *Economic Forecasts and Expectations*. National Bureau of Economic Research, New York.

Oomen, R.C.A. (2006), Properties of realized variance under alternative sampling schemes. *Journal of Business & Economic Statistics* 24, 219–237.

Press, S.J. (1967), A compound events model for security prices. *Journal of Business* 40, 317–335.

Robert, C.P., Casella, G. (1999), *Monte Carlo Statistical Methods*. Springer, New York.

Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility. In: Cox, D.V.H.D.R., Barndorff-Nielsen, O.E. (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*. Chapman and Hall.

Taylor, S.J. (1986), *Modeling Financial Time Series*. John Wiley.

Taylor, S.L. (1994), Modelling stochastic volatility. *Mathematical Finance* 4, 183–204.

Tierney, L. (1994), Markov chains for exploring posterior distributions. *The Annals of Statistics* 21, 1701–1762.