Applied Financial Economics, 1998, **8**, 671–687

A Bayesian analysis of stock return volatility and trading volume

RONALD MAHIEU* and ROB BAUER§

* *Erasmus Universiteit Rotterdam, Econometric Institute, P.O. Box 1738, 3000 DR Rotterdam, Netherlands* and § *ABP Investments, Equity Department, P.O. Box 2889, 6401 DJ Heerlen, Netherlands*

The relationship between stock return volatility and trading volume is analysed by using the modified mixture model (MMM) framework proposed by Andersen (1996). This theory postulates that price changes and volumes are driven by a common latent information process, which is commonly interpreted as the volatility. Using GMM estimation Andersen finds that the persistence in this latent process falls when a bivariate model of returns and volume, i.e. the MMM, is estimated instead of a univariate model for returns. This empirical finding is inconsistent with the MMM. As opposed to Andersen's study we apply recently developed simulation techniques based on Markov Chain Monte Carlo (MCMC). A clear advantage of MCMC methods is that estimates of volatility are readily available for use in, for example, dynamic portfolio allocation and option pricing applications. Using Andersen's data for IBM we find that the persistence of volatility remains high in the bivariate case. This suggests that the choice of the estimation technique could be important in testing the validity of the MMM.

I. INTRODUCTION

The relationship between stock return volatility and trading volume has been a subject on the research agenda for a few decades now. The results of these studies can be used in various fields of financial economics, but its main applications are used in dynamic portfolio allocation and the pricing of options. A better understanding of the underlying stochastic volatility process enables us to use the insights of modern finance in a more sophisticated way. Clark (1973) started the discussion by presenting the intuitively appealing Mixture of Distributions Hypothesis (MDH). The MDH posits that stock returns and trading volumes are jointly dependent on the same underlying, latent information flow variable. Empirical studies by Epps and Epps (1976), Tauchen and Pitts (1983) and Harris (1986, 1987) largely confirmed the predictions of this hypothesis. Recent work by, for example, Lamoureux and Lastrapes (1994), Richardson and Smith (1994) and Liesenfeld (1998) revealed some shortcomings of the standard mixture hypothesis. In a direct test of the standard mixture model, Richardson and Smith (1994) state that linking price changes and trading volume to the same latent information flow via a bivariate conditional *normal* distribution may not be the correct specification. Lamoureux and Lastrapes (1994) estimate the time series behaviour of the mixing variable. They conclude that it does not account fully for the observed persistence in volatility.

In the last decade a substantial number of papers in the market microstructure literature focused on the link between return volatility and trading volume. Most models in this area of research assume that price movements are caused by the arrival of new information and the process that incorporates this information into market prices. Important variables in these models are trading volume, the number of trades and liquidity. Most work is devoted to explain the *intra-daily* relationship between volatility and volume. A first approach to merge the insights of the MDH with those of the market microstructure theory is an empirical model of the daily return–volume relationship developed by Andersen (1996). Andersen's model is explicitly motivated by the results of the market microstructure models by Glosten and Milgrom (1985), Kyle (1985) and Admati and Pfleiderer (1988, 1989). He combines several, important features of these models – for instance an asymmetric information structure and the presence of liquidity or noise traders

0960–3107

– with the MDH and the related concept of stochastic volatility. The resulting model, called the Modified Mixture Model (MMM), is estimated with a dynamic AR(1) stochastic volatility process for the latent rate of information arrival, as proposed by Andersen (1994).

In this paper we start by presenting the MMM as well as the main ideas of the underlying theoretical model of Glosten and Milgrom (1985). Andersen (1996) provides a testable version of this model. Subsequently, he tests this model on a heavily liquid stock (IBM) listed on the New York Stock Exchange (NYSE). In the estimation stage he uses the Generalized Method of Moments (GMM) technique to estimate several empirical specifications. One of the contributions of our study is that we use a different technique to estimate the empirical model. In particular, we construct an algorithm based on Markov Chain Monte Carlo (MCMC) simulation methods and Bayesian analysis. Jacquier *et al.* (1994) estimate a *univariate* stochastic volatility model using these techniques. We adapt and extend their ideas by estimating both a *univariate* model for stock returns and a *bivariate* model for stock returns and trading volume. The resulting bivariate specification is equivalent to the modified version of the mixture model as advocated by Andersen (1996). In order to be able to compare our results directly with Andersen's study we decided to use the same stock return and trading volume series (IBM) for our analysis.[1] We are particularly interested in whether we are able to confirm Andersen's result that the persistence parameter will decrease significantly in the MMM. A further, major advantage of our estimation method is that it enables us to study the latent information process. This is not possible when GMM or Simulated Maximum Likelihood methods are used.

The remainder of this paper is organized as follows. In order to give the modified mixture model more theoretical background we first present the original model developed by Glosten and Milgrom (1985) in Section II. In Section III we derive an empirical version of the bivariate model, which is similar to the specification used by Andersen (1996). Section IV contains a discussion on the MCMC estimation technique in greater detail. In Section V we present summary statistics and other characteristics of the IBM stock return and corresponding trading volume series. Section VI provides simulation results, using the MCMC algorithm, for the univariate model, in the spirit of Jacquier *et al.* (1994). We present histograms of the respective draws of parameters simulating a univariate SV model. In Section VII we show results for the modified mixture model. We compare our simulation results directly with those of Andersen (1996) and indirectly with those of other studies, see for instance Liesenfeld (1998), who estimates the same model using simulated maximum likelihood estimation (SML) and a German database. Finally, Section VIII contains an interpretation of results and some concluding comments.

II. THE THEORETICAL MODEL

The modified version of the MDH by Andersen (1996) is in fact based on a theoretical model by Glosten and Milgrom (1985), henceforth GM. Their model is particularly suitable as it is explicitly structured to explain the process of information arrival and assimilation that occurs shortly after a piece of relevant information enters the market. GM base their model in an environment where there is a single market for an asset with a random liquidation value. Information on the terminal value of this asset enters into the market and each market player could possibly receive this information in a different way. At every point in time three different groups of risk-neutral traders are active in the market for this asset: a specialist, informed traders and uninformed traders. GM assume at this point that investors arrive at the market sequentially in random order. In this setting informed traders obtain private signals regarding the true value of the asset. GM show that these private information arrivals eventually induce a dynamic learning process that results in prices that fully reflect the content of the information through the sequence of trades and transactions. This period is referred to as a so-called *price discovery* or *information assimilation* phase. This turbulent period is then followed by an equilibrium phase.

The dynamics of GM's model are best characterized by a marketplace where each piece of new information relevant for the terminal value of the asset leads to a period of price discovery followed by a temporary equilibrium. Because of the frequent arrival of new information, all agents revise their estimates of the terminal value on a continuous basis. This information is either gathered from public signals, observable by each market player in a similar fashion, or from private sources that are only received by traders with a significant information advantage. Next to that each trader is able to derive information from the sequence of transaction prices.

Andersen (1996) formalizes the dynamics of the model as follows: C_t is the common information set at time t and each trader's information set is denoted by φ_t. From the outline above it follows that φ_t consists of C_t plus a possible private information set, dependent on the group of investors the particular trader belongs to. The value the specialist assigns to the asset is the expected value conditional on his current information: $P_t = E[V|S_t]$, where V denotes the terminal value of the asset and S_t denotes the information set at time t of the specialist. GM assume that the specialist knows the

[1] Here we would like to thank Torben Andersen for providing us with the data set.

structure of traders in the market as well as the structure of new arrivals taking place. The specialist works under a zero profit constraint. GM and Andersen (1996) explicitly state that P_t will not be the quoted price as the specialist is able to observe whether the next agent buys or sells. This can be interpreted as an additional source of information. The specialist will quote prices of V conditional on S_t plus the additional buy or sell information in his order book. This implies that the specialist will never regret a trade *ex post*. Transaction prices then follow a martingale and therefore observed prices for V at time t are fair assessments of the future value of the asset.

Before we present the empirical version of this model in the next section we proceed, analogous to Andersen (1996), by assuming that uninformed traders arrive at the market according to a constant Poisson information arrival process with intensity m_0 per day. The presence of liquidity traders (m_0) is necessary in this context, see Kyle (1985), because they are able to provide the market with liquidity, thereby helping to circumvent the no-trading theorem of Milgrom and Stokey (1982).[2] Andersen (1996) and GM subsequently assume that these liquidity traders have inelastic demand and supply functions, which implies that such a trader will buy or sell with probability one half. In contrast, informed traders base their trading decision on the expected value of the asset conditional on their information set, φ_t. The information sets of informed investors are of course correlated but not necessarily identical, which leaves room for initial disagreement among informed traders. An important point is, however, that the value they and other traders assign to the asset converges during the price discovery phase. This implies that there is a direct relationship between the informativeness of private information signals and the arrival rate of traders with an information advantage. The theory of GM is in fact meant to describe the price and information arrival process within a day. Andersen (1996), however, translates their theory into a daily framework. Thereby he avoids a lot of complications associated with short-run dependencies in informed trading. The properties of the return and volume series at the daily level are driven largely by the number of information arrivals per day. In the next section we show Andersen's formal derivation of an empirical and testable version of the GM model.

III. THE EMPIRICAL MODEL

In the marketplace set by GM the market moves from one temporary equilibrium to the other, during and across trading days, in response to a large number of information arrivals each trading day. Using the same framework as Clark (1973) and Tauchen and Pitts (1983) and assuming a large number of daily information arrivals, Andersen (1996) presents a modified version of the mixture of distributions hypothesis. He starts with the following return specifications:

$$R_t | K_t \sim N(0, \sigma^2 K_t) \tag{1}$$

Here R_t denotes the daily return, which is the logarithmic difference in prices of two consecutive trading days. K_t is the intensity of information arrivals, measured relative to a benchmark of a fixed, large number of information arrivals and σ^2 is a constant scaling factor. This implies that returns are conditionally normal, but the variances of the returns reflect the intensity of the information. This result is similar to the more general theorem presented by Clark (1973). In this setting the dynamics of the return volatility is solely dependent on the properties of the information flow in the market, see also Lamoureux and Lastrapes (1990, 1994) for this result. In contrast to these and other related studies, Andersen (1996) assumes that daily trading volume (V_t) can be divided into an informed component (IV_t) and a noisy component (NV_t). In addition to that, he assumes that the noise trading part of trading volume is governed by a stochastic process with an arrival intensity of m_0 per day. This implies that the noise component of trading volume is directed by a time-invariant Poisson process $Po(m_0)$. Consequently, the systematic part of daily trading volume is due to the dynamics of the underlying process of information arrival. Using this result and the insights of the GM model in the previous section, Andersen (1996) presents a framework in which each informed trader on average makes only a few trades per day. Combined with the expressions for noise and informed components of trading volume this eventually leads to the following distribution for daily trading volume:[3]

$$V_t | K_t \sim Po(m_0 + m_1 K_t) \tag{2}$$

Volume is Poisson distributed, conditional on K_t. In this setting m_0 reflects the noise component of trading volume, i.e. liquidity trades, and m_1 is the factor of proportionality, or the informed component of trading volume, which is proportional to the information flow. Andersen (1996) shows that this factor determines how strongly daily trading volume fluctuates in response to the arrival of unexpected news. Moreover he shows in his formal derivation of the empirical model that m_1 consists of the product of two other factors: the maximum number of insiders that might obtain a private signal and the expected number of trades by an insider. These factors are, however, not identifiable and therefore we ignore these parameters. We proceed along the same path. Andersen normalizes the system by setting

[2] See also O'Hara (1995) for a detailed discussion on the no-trading theorem.
[3] The formal derivation and assumptions made are displayed in Andersen (1996, p. 176). Andersen argues that the Poisson approximation is more precise than a normal approximation.

σ in Equation 1 to 1, which leads to the following return equation:

$$R_t|K_t \sim N(0, K_t) \tag{3}$$

Here it becomes clear that the return volatility in Equation 3 is identical to the information flow and therefore the unobserved information process (K_t) can be interpreted as a stochastic volatility process, see Andersen (1994) and Taylor (1994). At this point we have to bear in mind that the trading volume series generally exhibits a strong time trend, especially when measured over a time period longer than a few years. Andersen (1996) shows for the IBM stock that there is indeed a trend present in the trading volume series. In order to be able to compare our results with those of Andersen (1996) we decided to apply the same detrending procedure. In this procedure the daily volume series are detrended by extracting a time trend (a_t) from the observed volume series:[4]

$$Vo_t = V_t a_t \tag{4}$$

where Vo_t denotes the observable volume series and V_t the stationary volume series. It should be pointed out here that we face a scaling problem here, see Andersen (1996) and Liesenfeld (1998). The estimated time trend, $\hat{a}_t$ will at best reproduce the true underlying trend up to a certain proportionality factor (c):

$$c\hat{a}_t = a_t \tag{5}$$

where c is an unknown positive constant. The detrended volume series then become:

$$\hat{V}_t = \frac{Vo_t}{\hat{a}_t} = cV_t \tag{6}$$

This leads to the following distribution for daily trading volume:

$$\hat{V}_t|K_t \sim c.\ Po(m_0 + m_1 K_t) \text{ with } m_0 \geqslant 0,\ c > 0 \tag{7}$$

The detrending procedure introduces a parameter c that can be estimated, see also Liesenfeld (1998). The detrended volume series in Equation 6 is given by a product of a Poisson distributed random variable V_t and a positive constant c. The resulting empirical specification is a combination of the return (Equation 3) and volume (Equation 7) equations. The difference with the standard bivariate mixture model for price changes and trading volume, see Tauchen and Pitts (1983), is in the specification of the volume equation. In our modified version of the mixture hypothesis there is a much stronger link with the market microstructure literature. This link is represented by assuming the presence of liquidity traders in the market, represented by the noisy part of trading volume. In addition, we assume a conditional Poisson distribution in the volume specification, instead of the conditional normal distribution. This explicitly respects the non-negativity constraint for trading volume. Note that the return specification is left unchanged. Thus, by accounting for an information independent part of trading volume, the MMM can be interpreted as a generalization of the standard bivariate Tauchen and Pitts (1983) mixture model. Because of the Poisson distribution, the density function is only defined for integer-valued random variables. This however cannot be ensured as V_t and c are real-valued scalars. Hence Liesenfeld (1998) concludes that Maximum Likelihood (ML) estimation using a Poisson distribution cannot be applied directly. If one, however, assumes that ($m_0 + m_1$) is large enough, the Poisson distribution can be approximated by a corresponding normal distribution:

$$\hat{V}_t|K_t \sim N(c[m_0 + m_1 K_t],\ c^2[m_0 + m_1 K_t]) \quad \text{with} \quad m_0 \geqslant 0,\ c > 0 \tag{8}$$

Liesenfeld (1996) uses this volume specification in a simulated maximum likelihood procedure, originally proposed by Danielsson and Richard (1993). Andersen (1996) estimates his empirical model with the Generalized Method of Moments (GMM). We will estimate the MMM with a different estimation technique based on Bayesian analysis. To estimate the model we first need to specify the stochastic process, which is assumed to govern the latent number of information arrivals (K_t). In line with Andersen's approach we propose a full dynamic presentation for the mixing variable.

Andersen (1996) makes a few remarks that are relevant in the selection of a dynamic representation of the information arrival process. Information on a company tends to be positively correlated: unexpected announcements tend to be followed by several other, often *related* news items. Consequently Andersen states, looking at the bulk of empirical research on volatility processes, that the dynamic process must display positive conditional dependencies. Andersen proposes two possible specifications: a standard SARV model and an exponential SARV specification, which is equivalent to a lognormal specification for information arrival. Bearing in mind the clustering of new arrivals, we specify a lognormal Stochastic Volatility model for the latent variable. This leads to the following specification for the empirical version of the modified mixture model:

$$y_t|h_t \sim N(0, \exp(h_t)) \tag{9}$$

$$V_t|h_t \sim Po(m_0 + m_1 \exp(h_t)) \tag{10}$$

$$h_t = \mu + \varphi h_{t-1} + \eta_t \qquad \eta_t \sim N(0, \sigma^2_\eta) \tag{11}$$

[4] In the empirical section we implement the detrending procedure using a nonparametric regression with a normal kernel. Here we would like to thank Torben Andersen for providing a copy of his GAUSS program.

where we introduced $h_t \equiv \ln(K_t)$ as the logarithm of the latent information process.[5]

IV. THE ESTIMATION PROCEDURE

In this section we describe issues relating to the alternative estimation method for the modified mixture model (Equations 9–11). In particular we propose to construct an algorithm based on Markov Chain Monte Carlo (MCMC) simulation techniques. All parameters in the model are estimated using Bayesian analysis. Note that the model without a trading volume, Equation 10 is exactly equal to the univariate Stochastic Volatility (SV) model, studied extensively in the literature, see Jacquier *et al.* (1994), Ruiz (1994), Andersen and Sørensen (1996) Mahieu and Schotman (1998) and Kim *et al.* (1998) among others. The techniques for estimating these simple SV models can be extended in order to deal with the estimation of the parameters and the latent information process in the bivariate modified mixture model of Andersen (1996). Andersen (1996) estimates the MMM using Hansen's (1982) Generalized Method of Moments (GMM). GMM is a relatively fast and robust method for estimating dynamic latent variable models. A major advantage of GMM is that the implications of the distributional assumptions of a model can be estimated adequately. A serious disadvantage of GMM is, however, that no estimate of the latent process itself is made.[6] For financial applications, such as the pricing and hedging of options, an estimate of the volatility process is required. Another paper investigating Andersen's (1996) model is Liesenfeld (1998). In that paper a Simulated Maximum Likelihood (SML) estimation procedure, advocated by Danielsson and Richard (1993), is applied. This approach also suffers from the disadvantage that the latent information process itself is not estimated.

From the previous paragraph it has become clear that a need exists for an estimation method of the MMM that gives us an estimate of the latent information process and the associated volatility process. In this paper we will build on the MCMC methods of Shephard and Pitt (1997). These methods allow a simultaneous estimate of both the latent variable and the unknown parameters. In the following we will sketch our estimation procedure. More details on the algorithm can be found in the Appendix. The parameters in the MMM can be collected in the vector $\theta \equiv (\mu, \varphi, \sigma_\eta^2, m_0, m_1)$. Note that we do not include the volume scaling parameter c in this vector, which is used in Andersen (1996). Our setup implies that estimates form m_0 and m_1 include the volume scaling.[7] This leads to the observation that we cannot directly test hypotheses on these parameters, see Andersen (1996). However, we are still able to measure the fractions of the average daily volume that are independent (m_0) and dependent (m_1) on the information flow, respectively.

We apply a Bayesian estimation procedure for the parameters. We use both informative and non-informative priors. The constant μ in the latent variable transition equation has a non–informative prior distribution which leads to a normal posterior. For the transition parameter φ we chose a beta distribution prior on the interval $(-1,1)$, with a mean of 0.90 and a standard deviation of 0.10. The resulting posterior is non-conjugate, which led us to sample from it by using an accept–reject algorithm. The prior distribution for σ_η^2 is an inverse gamma distribution which again leads to an inverse gamma posterior. These priors are similar to the ones used in Shephard and Pitt (1997). The two parameters in the volume equation (Equation 10) were sampled from non-conjugate posterior distributions using accept–reject algorithms. The priors for m_0 and m_1 are a gamma and a normal distribution, respectively. The parameters in these prior distributions for m_0 and m_1 are chosen to have an implied mean and standard deviation that correspond to the GMM parameter estimates in Andersen (1996).

The MCMC algorithm that we employ cycles through six conditional distributions. The first five correspond with the parameters θ and the sixth with the latent variable. Let $\theta_{|i}$ be the parameter vector with the ith parameter deleted. We can then draw from the distributions:

$$f(\theta_i | Y_T, V_T, \hat{h}_T, \theta_{|i}) \tag{12}$$

with

$$Y_T \equiv \{y_t\}_{t=1}^T \qquad V_T \equiv \{V_t\}_{t=1}^T \qquad h_T \equiv \{h_t\}_{t=1}^T \tag{13}$$

where the last stage of the estimation procedure refers to the conditional distribution of the latent variables h_T:

$$f(h_T | Y_T, V_T, h_T, \theta_{|i}) \tag{14}$$

In each round of the MCMC algorithm a new element from the vector (θ, h_T) is drawn. Each draw replaces the old value for that element and the algorithm moves to the next

[5] Note that our latent information arrival process differs slightly from the SARV model Andersen (1996) primarily uses. The logarithmic specification remains close to the processes that have been proposed in theoretical financial models for volatility as, for example, in Hull and White (1987).

[6] Another problem with GMM is the possible inefficiency of the parameter estimates, resulting from choosing a particular set of moments, see Andersen and Sørensen (1996). Gallant and Tauchen (1996) fix this problem in their Efficient Method of Moment (EMM) estimation procedures. However, EMM does not provide us with an estimate of the latent variable either.

[7] This implicitly means that we estimate cm_0 and cm_1, respectively.

element. As Tierney (1994) has shown this algorithm converges to drawings from the joint distribution of the parameters and the latent variables, under mild conditions that the densities should be positive over their support. For common lengths of financial time series, draws from the conditional distributions can be prohibitive as the conditional distribution of the latent variables is proportional to a T-dimensional integral (see Jacquier *et al.* 1994) due to the nonlinear relation between the measurement Equations 9 and 10 and the transition Equation 11. Much attention has been given to this issue, as was described above. Here we follow the approach of Shephard and Pitt (1997) who propose to draw blocks of states instead of single states as in Jacquier *et al.* (1994). Within the MCMC chain the length of the blocks is determined by so-called knots that are randomly drawn from the states of the previous round of the MCMC. As opposed to single move samplers as in Jacquier *et al.* (1994), the multi move block samplers are quicker and show much less autocorrelation in successive draws from the chain. We use the pseudo Metropolis–Hastings algorithm of Tierney (1994) to evaluate each block of latent variables. Compared to the standard Metropolis–Hastings sampler (see Chib and Greenberg, 1995), this algorithm allows the blanket function to be a dominating distribution on parts of the support only. For all technical details of this sampler we refer to the appendix.

V. DATA DESCRIPTION AND SUMMARY STATISTICS

This section briefly describes the general features of the IBM stock return and volume series. As stated earlier we use the same data set as Andersen (1996), who uses a large sample of continuously compounded daily returns, corrected for dividends and stock splits. The investigation period starts 2 January 1973 and ends 23 December 1991, spanning 4693 return observations.[8] The closing prices were obtained from the *Standard & Poor's Daily Stock Price Guide*. The time series of returns and detrended volumes series we use for our empirical analysis are shown in Fig. 1. Corresponding histograms for the returns and volume series are provided in Fig. 2.[9] The sample return mean is very small and the corresponding variance of returns is much higher, see Table 1. Daily return series display the expected excess kurtosis. Excluding the volatile period of October 1987 would obviously yield a much lower kurtosis value. Skewness values are apparently dominated by the negative value in October 1987 too. Daily IBM stock returns are therefore not in line with a normal distribution. Finally, looking at the autocorrelation coefficients shows that the IBM returns series display the usual dependencies we find in higher order moments.

As noted earlier we have to detrend the volume series as the sample period is very long. Andersen (1996) provides evidence that there is a significant growth percentage in the IBM volume series. He describes two possible detrending procedures. We apply the first procedure where a nonparametric regression with a normal kernel is used.[10] In this method, the trend component that produces a *normal* volume series is estimated. Subsequently, the detrended series is obtained by simply dividing each trading figure with the corresponding *normal* volume for that day, which leads to an average of approximately one. This corresponds to a two-sided moving average with weights that decline as we move further from the trading day.

Summary statistics for the detrended trading volume series are also shown in Table 1. The mean of this series is near unity, which is consistent with the normalization rule used in the detrending procedure. Skewness measures are clearly positive and the kurtosis value is significantly greater than 3.[11] As expected, the lower lag autocorrelations for trading volume have very high values. In line with the standard mixture model of Harris (1987), trading volume displays a higher degree of autocorrelation than the return series.[12]

VI. UNIVARIATE SIMULATION RESULTS

The MCMC simulation algorithm, see Section IV, is first applied to the IBM stock return series. We start estimating both the parameters and the latent information process of

[8] Andersen (1996) deletes all observations between 24 December and 1 January. For reasons of comparison we follow his approach. The summary statistics are presented in Andersen (1996) too. We still decided to include the main statistics in Table 1 as they will be referred to in the empirical section. Andersen (1996) also presents a table with cross-correlations between squared returns and trading volume (not detrended) for several lags.

[9] In addition, we divided the full sample into various sub-samples. The results were equivalent to those reported in Andersen (1996). We choose to provide the main summary statistics here. Andersen also displays the autocorrelation functions up to 50 lags for returns, squared returns and absolute returns, respectively. These series are clearly not i.i.d. as the appropriate confidence bands are frequently violated.

[10] The other detrending method, based on a centred two year rolling sample mean, rendered the same results.

[11] The sub-sample analysis in Andersen (1996) shows that the higher kurtosis for returns is caused by the stock market crash in 1987. In the majority of the sub-samples the kurtosis of volume is higher than for returns.

[12] Andersen (1996) reports highly significant cross-correlations between return volatility and trading volume, which is also in line with the results of previous studies of Karpoff (1987).

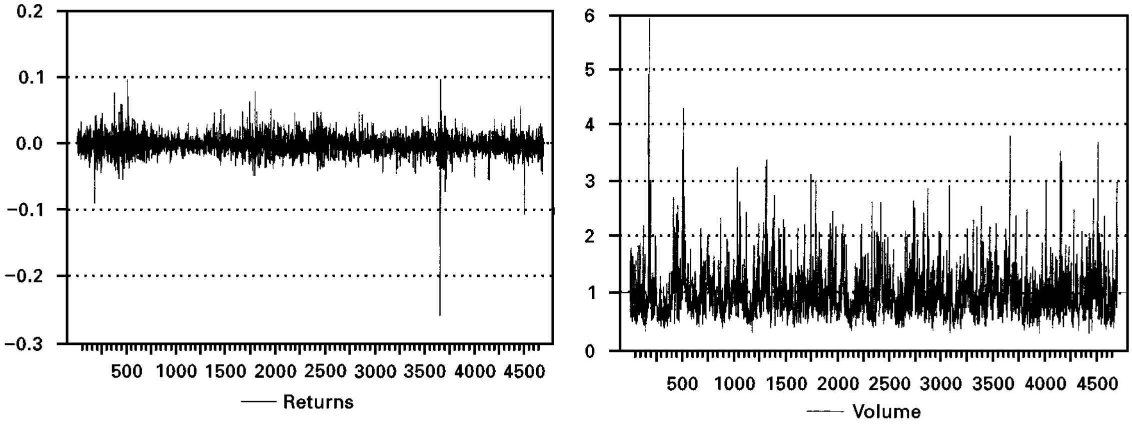

Fig. 1. *Time series of IBM stock returns and trading volumes (1973–1991)*

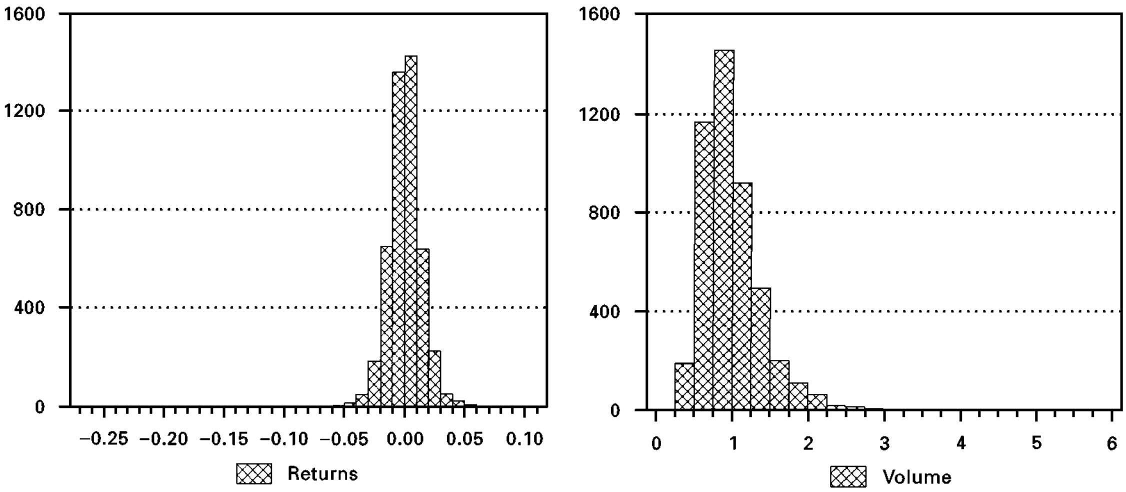

Fig. 2. *Histograms of IBM stock returns and trading volumes series (1973–1991)*

Table 1. *Summary statistics of the IBM stock return and detrended trading volume series*

	Mean (0.10^2)	St. Dev.	Maximum	Minimum	Skewness	Kurtosis
Returns	1.51	1.46	10.05	− 26.09	− 1.04	27.82
Volumes	0.99	0.41	5.93	0.22	2.10	13.00
	AC(1)	AC(2)	AC(3)	AC(4)	AC(5)	AC(10)
Returns	− 0.031	− 0.008	0.005	− 0.035	0.025	− 0.022
Volumes	0.521	0.315	0.249	0.234	0.223	0.128

Number of observations (T) is 4693; AC(p) denotes the return autocorrelation with lag p; The standard error of the return autocorrelations $(1/T)^{0.5}$ is equal to 0.015.

the underlying, univariate stochastic volatility model without an additional measurement equation for trading volume. In Fig. 3 histograms of the draws of each of the three parameters (μ, φ, and σ_η^2, respectively) in this specification are shown together with the corresponding sequence of draws of these parameters.

For each parameter we present simulation results for 25 000 consecutive draws. Convergence of the MCMC chain

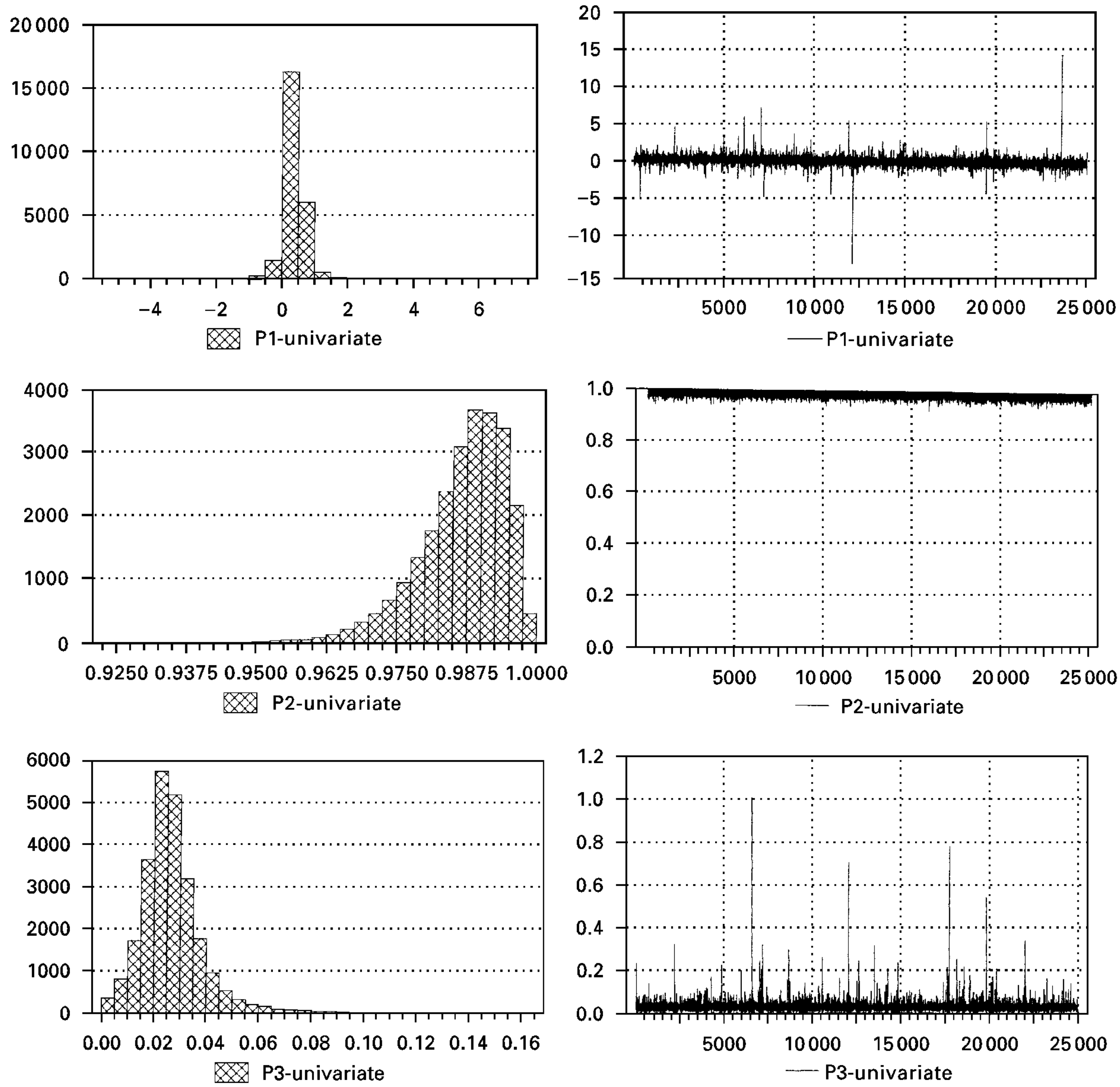

Fig. 3. *Histograms of simulated parameters and sequence of draws: univariate SV model*
P1 = μ; P2 = φ; P3 = σ^2_η.

is checked by comparing the draws of several chains simultaneously.[13] The marginal distributions in the histograms are in line with our expectations. Table 2 provides summary statistics on these distributions. The constant in the transition equation, μ (P1-Univariate), has a distribution that is severely positively skewed. Furthermore it is clear that this distribution has fat tails. The large span between minimum and maximum values for μ confirms the existence of a non-normal distribution for the draws of this parameter. The associated Monte Carlo standard error is relatively large. We can therefore conclude that the simulated values for this parameter are very unstable in the univariate model. As expected, the volatility persistence parameter, φ (P2-Univariate), is close to unity in the univariate case. The distribution of this parameter is negatively skewed with a mean of approximately 0.99. The Monte Carlo standard errors are now very small relative to the mean value for this parameter. In addition, the sequence of draws in Fig. 3 shows that almost every value of the persistence parameter is above 0.95. This result is consistent

[13] For an extensive review of current algorithms for checking MCMC output we refer to Cowles and Carlin (1996).

Table 2. *Summary statistics univariate simulation results*

	μ	φ	σ_η^2
Mean	0.382	0.987	0.027
MC std. error	0.359	0.008	0.018
Minimum	− 13.499	0.924	0.000
Maximum	15.078	0.999	1.006
Skewness	1.140	− 1.166	16.63
Kurtosis	230.401	5.084	679.226
Correlations			
μ	1.000		
φ	0.013	1.000	
σ_η^2	− 0.082	− 0.105	1.000
AC(1)	0.064	0.000	− 0.348
AC(2)	0.027	− 0.001	0.354
AC(3)	− 0.011	0.006	− 0.197
AC(4)	0.005	− 0.007	0.189
AC(5)	− 0.015	0.000	− 0.123
AC(10)	− 0.007	0.009	0.034
AC(20)	0.003	− 0.008	0.001
Number of draws	25 000	25 000	25 000

MC std. error denotes the Monte Carlo standard errors of the simulated parameters, AC(p) denotes the autocorrelation with lag p.

with Andersen (1996), who estimates a univariate SARV-model on the same series, using GMM estimation. Finally, estimates for σ_η^2 (P3-Univariate) indicate that we have quite a lot of outliers in our simulations for this parameter too. This is also indicated by the corresponding skewness and kurtosis values and the sequence of draws in Fig. 3.

Looking at the correlations between the parameters in Table 2, we can conclude that these are generally low. Moreover, autocorrelation coefficients are negligible for μ and φ. For σ_η^2 we find that these estimates are high for lower lags, but the coefficients decrease quickly for longer lag lengths. This result is also found in the study of Shephard and Pitt (1997). In the next section we discuss the simulation results for the bivariate, modified mixture model of Andersen (1996). In particular, we are interested in whether the bivariate parameter estimates will have the same distributional properties as those in the univariate case. Another point of interest is whether we are able to confirm Andersen (1996) and Liesenfeld (1998) with respect to the decrease of the measure of persistence in volatility (φ).

VII. BIVARIATE SIMULATION RESULTS

The next step is to apply the MCMC simulation algorithm to the IBM return and trading volume series. We start by estimating both the parameters and the latent information process of the underlying, bivariate stochastic volatility model, which includes an additional measurement equation for trading volume. In Fig. 4 histograms of the draws of each of the five parameters (μ, φ, σ_η^2, m_0 and m_1, respectively) are shown together with the corresponding sequence of draws of these parameters. Again, we present results for 25 000 consecutive draws. A first, quick glance at the summary statistics presented in Table 3, tells us that the distributions of these parameters have different features when compared to the univariate case. The constant in the transition equation, μ (P1-Bivariate), has a slightly lower mean, but the standard error is much lower than in the univariate setting. The displayed minimum and maximum values are in line with this phenomenon. The most interesting result is that the volatility persistence parameter, φ (P2-Bivariate), is still close to one in the bivariate model simulations. In fact it does not even decrease in value. This is in sharp contrast with the results in Andersen (1996) and Liesenfeld (1998), who find substantially lower persistence in volatility in the case of the modified mixture model.[14]

Andersen (1996) investigates whether the reduction of the persistence parameter could be caused by the choice of the estimation procedure. He concludes that the significant reduction in the estimated volatility persistence cannot be explained by differences between estimation methods. In particular, he shows that *univariate* GARCH and SV models for the IBM return series, estimated by both GMM and Maximum Likelihood, have a high volatility persistence (0.99). The general structure of the modified mixture model could be an alternative explanation of the significant drop in persistence. Using the MCMC simulation approach we show that the bivariate mixture model still has the same, high persistence level. This indicates that the choice of the estimation procedure may indeed be a very important factor in this issue. The mean value for σ_η^2 does not change very much either relative to the univariate approach, but again the Monte Carlo standard error decreases substantially in the bivariate model specification.

As mentioned earlier, m_0 (P4-Bivariate) measures the fraction of daily trading volume independent of the underlying latent information process: $\exp(h_t)$. On average we find that about 82.2% of daily trading volume is unrelated to the information flow.[15] The fraction of daily trading volume in IBM stocks that is directly influenced by the latent

[14] On the other hand, Foster and Viswanathan (1995) obtain the result that the extension of the univariate model into bivariate models with trading volume possibly leads to a reduction in the volatility persistence.

[15] It should be noted that the detrended volume series has a mean of unity, which is obtained through the detrending procedure described in the previous section.

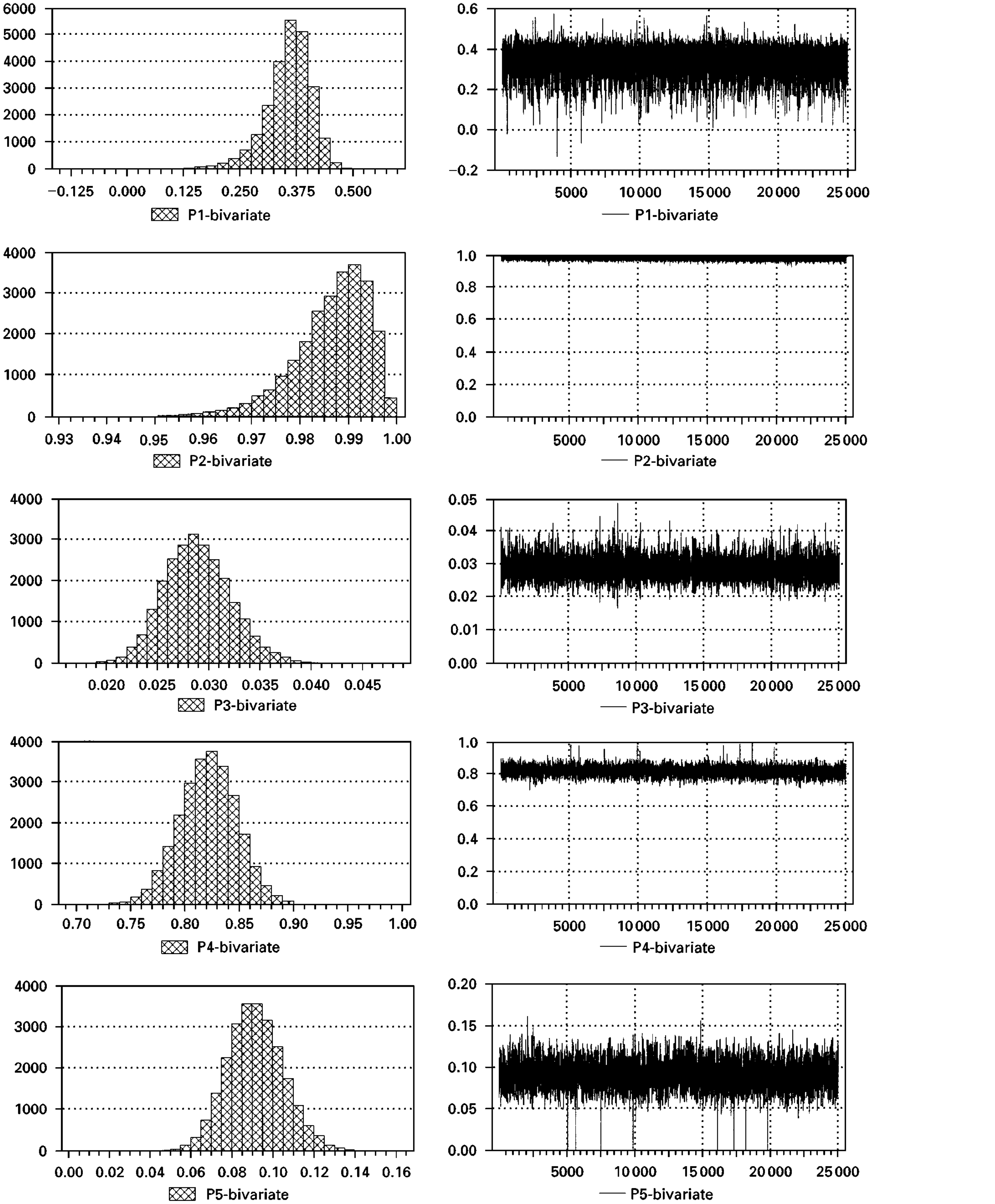

Fig. 4. *Histograms of simulated parameters and sequence of draws: modified mixture model.* P1 = μ; P2 = φ; P3 = σ_η^2; P4 = m_0; P5 = m_1.

Table 3. *Summary statistics bivariate simulation results*

	μ	φ	σ^2_η	m_0	m_1
Mean	0.356	0.987	0.029	0.822	0.092
MC std. error	0.054	0.008	0.003	0.027	0.014
Minimum	− 0.131	0.931	0.017	0.697	0.000
Maximum	0.580	0.999	0.049	0.996	0.163
Skewness	− 1.136	− 1.130	0.379	0.048	0.069
Kurtosis	6.236	4.844	3.503	3.541	3.795
Correlations					
μ	1.000				
φ	− 0.006	1.000			
σ^2_η	0.097	− 0.010	1.000		
m_0	− 0.145	− 0.007	0.002	1.000	
m_1	− 0.104	0.002	0.017	− 0.784	1.000
AC(1)	0.046	0.016	− 0.975	0.658	0.698
AC(2)	0.010	0.005	0.926	0.498	0.527
AC(3)	− 0.011	− 0.001	− 0.878	0.377	0.398
AC(4)	0.007	− 0.001	0.833	0.281	0.299
AC(5)	− 0.012	− 0.006	− 0.789	0.207	0.218
AC(10)	− 0.001	0.009	0.599	0.055	0.049
AC(20)	0.004	0.003	0.341	− 0.002	− 0.008
Number of draws	25 000	25 000	25 000	25 000	25 000

MC std. error denotes the Monte Carlo standard errors of the simulated parameters, AC(p) denotes the autocorrelation with lag p.

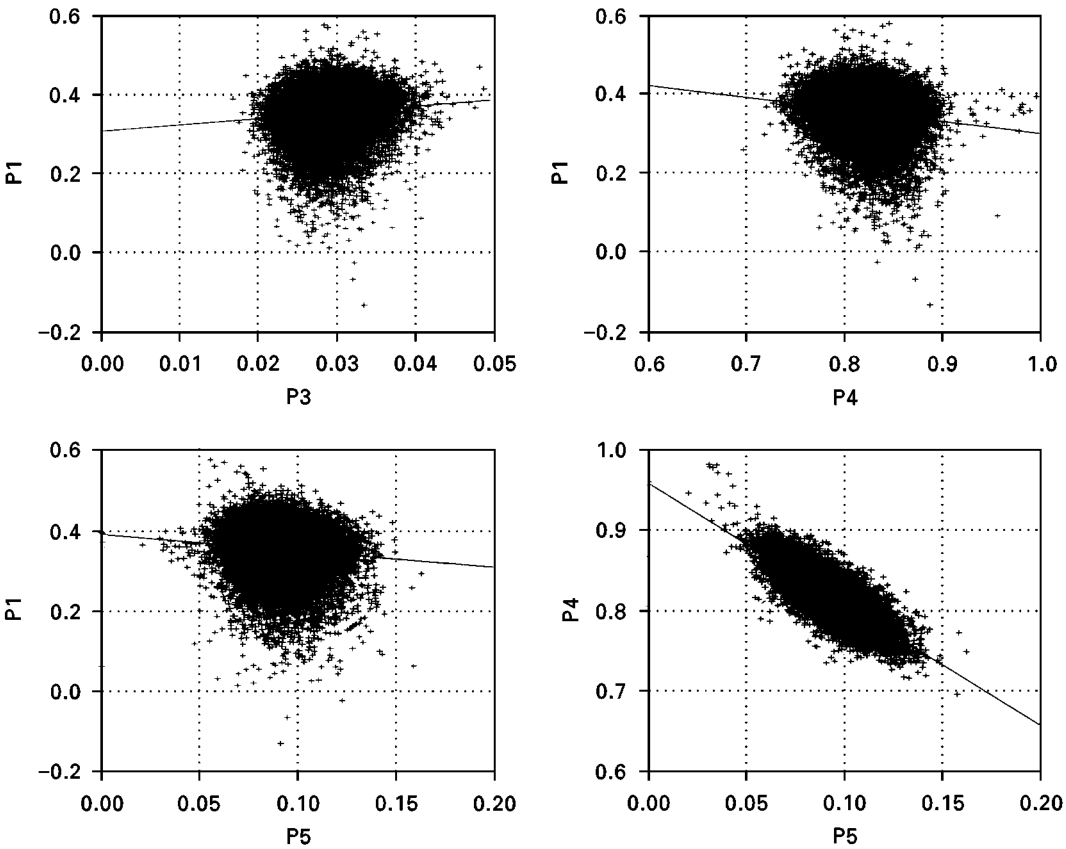

Fig. 5. *Scatter diagrams bivariate model*

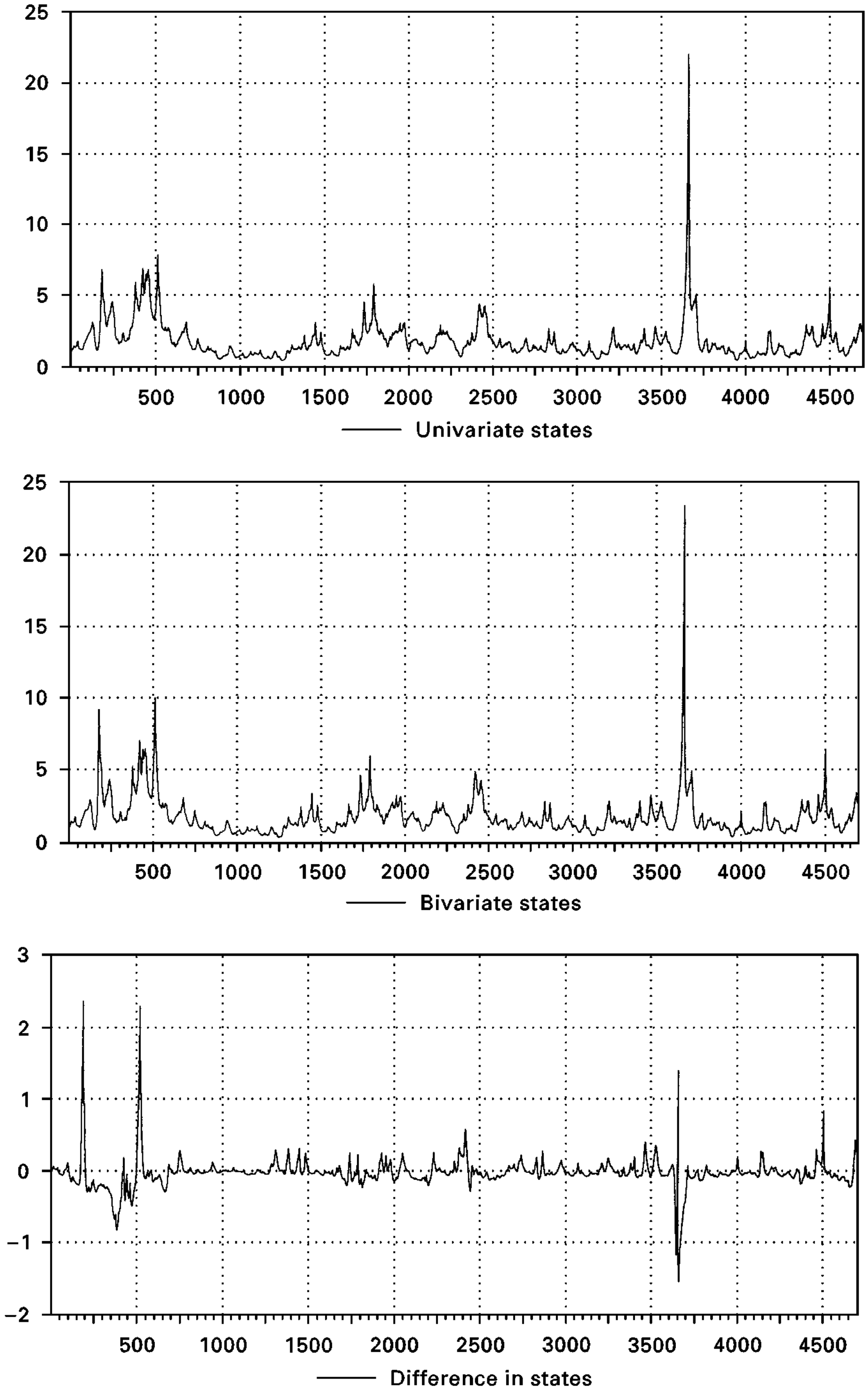

Fig. 6. *Estimated volatility series; univariate model, bivariate model and difference*

information process, is measured by m_1 (P5-Bivariate). The marginal distribution of m_1 has a mean of 9.2%. These fractions differ from the values found by Andersen. The percentage of information-insensitive trading he finds, is much lower (65.0%) and the information related part of trading volume is larger in Andersen's case (17.1%).

The correlation between the simulated bivariate parameter values, see also Table 3, is significant in four cases now. Particularly, μ is weakly correlated with all other parameters except φ. The strongest negative correlation is, however, present between m_0 and m_1 : 0.784. This may be expected because of the structure of the modified mixture model. We expect high values for m_1 whenever there is little information-insensitive trading and vice versa. In order to check whether these correlations are caused by outliers, we present scatter diagrams for these four cases. In the lower right panel of Fig. 5 we plot the series of the two correlated volume parameters and the corresponding regression line. We can clearly see that there is a strong negative correlation between m_0 and m_1. The other diagrams are rather inconclusive. The summary statistics in Table 3 furthermore show that the estimated autocorrelations for the parameters of the bivariate model are still low for μ and φ. For the other three parameters we find high coefficients for lower lag lengths. For m_0 and m_1 these values quickly decrease, but it takes more than 50 lags to obtain a low value for the autocorrelation coefficients of σ_η^2. Possibly this problem can be solved by a multi-move block procedure for the structural parameters, proposed in Pitt and Shephard (1998). Sub-sample analysis, however, shows that the marginal distributions of σ_η^2 are very similar.

As mentioned in the introduction, one of the advantages of the MCMC simulation procedure is that it enables us to study the latent information process, which can be very useful in several areas of finance. In the upper panel of Fig. 6 we present the estimated volatility series for the univariate model specification:

$$\exp(h_t) = \frac{1}{M}\sum_{i=1}^{M} \exp(h_t^{(i)}) \qquad t = 1, \dots, T \tag{15}$$

with $h_t^{(i)}$ a time-t draw from the MCMC chain in iteration i. The middle panel displays the volatility series for the bivariate mixture model. From a practical point of view we are mainly interested in whether we find a significant difference between the two estimated volatility series. Therefore we plot the difference between the two series in the lower panel of Fig. 6. This graph shows us that these series do not differ very much in most periods, but in some volatile periods, e.g. the oil crisis in 1973 and the stock market crash of 1987, we observe differences in squared percentages of more than 2%. This can have a substantial impact on the valuation of, for instance, derivative instruments and several strategic or tactical asset allocation topics.

Summary statistics are presented in Table 4.[16] We can see that the means of both series do not differ very much.

Table 4. *Summary statistics univariate and bivariate volatility series*

	Univariate	Bivariate
Mean	1.838	1.823
Std. Dev.	1.340	1.353
Minimum	0.415	0.376
Maximum	22.190	23.607
Skewness	4.782	4.927
Kurtosis	48.001	51.573

VIII. INTERPRETATION OF RESULTS AND CONCLUDING COMMENTS

We have studied the joint distribution of daily returns and trading volumes. The contemporaneous relationship between the two variables is derived from a market microstructure model in which the presence of liquidity traders and asymmetric information structures are the main features. The resulting specification is consistent with the mixture of distributions hypothesis documented in earlier papers. Analogous to Andersen (1996), the standard mixture model is modified by specifying a modified volume equation. In this setup trading volume is Poisson distributed, which implies that the modified mixture model explicitly accounts for the presence of liquidity traders by assuming that part of daily trading volume is unrelated to the latent information flow (noise trading) and that part is directly linked to the unobservable information process. The resulting bivariate system is governed by a random mixing variable representing the information flow or stochastic volatility variable. The lognormal stochastic volatility process is modelled as an AR(1) process. We make the following contributions to this discussion started by Andersen. We apply a different estimation procedure: a Markov Chain Monte Carlo based on Bayesian analysis. In contrast to the GMM approach of Andersen (1996), the MCMC method has the clear advantage that we are able to produce an estimate of the latent information process. This estimate can be used in several areas of modern finance.

The simulation results give reasons to believe that the discussion on this issue has only just started. Simulation results of the univariate stochastic volatility model confirm Andersen's result that the persistence parameter is close to unity for the liquid IBM stock return series. Monte Carlo standard errors are, however, rather large for the other two parameters, which indicates that the results are relatively unstable for the univariate model. Results for the bivariate mixture model are more robust in the sense that the marginal distributions of the simulated parameters are much less skewed and kurtotic. The most important result of this paper is that the persistence in volatility does not decrease in the bivariate model. Andersen (1996) and Liesenfeld

[16] These results are consistent with earlier findings of Richardson and Smith (1994) who find that the information flow tends to exhibit positive skewness and large kurtosis.

(1998) find that the persistence in volatility drops significantly when the univariate specification is extended into a bivariate specification with trading volume. Andersen (1996) argues that this might be caused by types of information arrival processes that have a different impact on volume and return volatility persistence. News releases and periodic events such as macro-economic announcements induce heavy trading volumes, but have only a short-lived effect on volatility. Failing to control for this difference could bias the estimation results. We, however, think that the choice of the estimation procedure also effects results. Using the same return and volume series (IBM) and a specification similar to that of Andersen (1996), we still find a high persistence in volatility in the bivariate case. Furthermore we find that a smaller part of daily trading volume is directly related to the unobservable information process.

The modified mixture model has proven to be a very fruitful area for further research as the results clearly indicate that trading volume can be an important variable in understanding the latent information and volatility process. This indicates that a bivariate framework may be the path to follow in new projects. Another interesting area for further research is to discriminate between various types of information that enter financial markets.

ACKNOWLEDGEMENTS

We would like to thank Torben Andersen, Roman Liesenfeld and seminar participants at the Erasmus University Rotterdam for giving valuable comments. The usual disclaimer applies.

REFERENCES

Admati, A. and Pfleiderer, P. (1988) A theory of Intraday patterns: volume and price variability, *Review of Financial Studies*, **1**, 3–40.

Admati, A. (1989) Divide and conquer: a theory of intraday and day-of-the-week mean effects, *Review of Financial Studies*, **2**, 189–223.

Andersen, T. (1994) Stochastic autoregressive volatility: a framework for volatility modeling, *Mathematical Finance*, **4**, 75–102.

Andersen, T. (1996) Return volatility and trading volume: an information flow interpretation of stochastic volatility, *Journal of Finance*, **51**, 169–204.

Andersen, T. and Sørensen, B. (1996) GMM estimation of a stochastic volatility model: a Monte Carlo study, *Journal of Business and Economics Statistics*, **14**, 328–52.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis–Hastings Algorithm, *American Statistician*, **49**, 327–35.

Clark, P. (1973) The subordinated stochastic process model with finite variance for speculative prices, *Econometrica*, **41**, 135–55.

Cowles, M. and Carlin, B. (1996) Markov Chain Monte Carlo convergence diagnostics: a comparative review, *Journal of American Statistical Association*, **91**, 883–904.

Danielsson, J. and Richard, J.-F. (1993) Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics*, **8**, 153–73.

DeJong, P. and Shephard, N. (1995) Efficient sampling from the smoothing density in time series models, *Biometrika*, **82**, 339–50.

Epps, T. and Epps, M. (1976) The stochastic dependence in security price changes and transaction volumes: implications for the mixture-of-distribution hypothesis, *Econometrica*, **44**, 305–21.

Foster, F. and Viswanathan, S. (1995) Can speculative trading explain the volume–volatility relation, *Journal of Business and Economics Statistics*, **13**, 379–96.

Gallant, A. and Tauchen, G. (1996) Which moments to match?, *Econometric Theory*, **12**, 657–81.

Geweke, J. (1994) Monte Carlo Simulation and Numerical Integration, working paper 526, Federal Reserve Bank of Minneapolis.

Glosten, L. and Milgrom, P. (1985) Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics*, **14**, 71–100.

Hansen, L. (1982) Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029–54.

Harris, L. (1986) Cross-securities tests of the mixture of distribution hypothesis, *Journal of Financial and Quantitative Analysis*, **21**, 39–46.

Harris, L. (1987) Transaction data tests of the mixture of distributions hypothesis, *Journal of Financial and Quantitative Analysis*, **22**, 127–41.

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

Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities, *Journal of Finance*, **42**, 281–300.

Jacquier, E., Polson, N. and Rossi, P. (1994) Bayesian analysis of stochastic volatility models, *Journal of Economics and Business Statistics*, **12**, 371–417, with discussion.

Karpoff, J. (1987) The relation between price changes and trading volume: a survey, *Journal of Financial and Quantitative Analysis*, **22**, 109–26.

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

Koopman, S. (1993) Disturbance smoother for state space models, *Biometrika*, **80**, 117–26.

Kyle, A. (1985) Continuous auctions and insider trading, *Econometrica*, **53**, 1315–35.

Lamoureux, C. and Lastrapes, W. (1990) Heteroskedasticity in stock return data: volume versus GARCH effects, *Journal of Finance*, **45**, 487–98.

Lamoureux, C. (1994) Endogenous trading volume and momentum in stock-return volatility, *Journal of Business and Economic Statistics*, **12**, 253–60.

Liesenfeld, R. (1998) Dynamic bivariate mixture models: modelling the behaviour of prices and trading volume, *Journal of Business and Economics Statistics*, **16**, 101–9.

Mahieu, R. and Schotman, P. (1998) An empirical application of stochastic volatility models, *Journal of Applied Econometrics*, forthcoming.

Milgrom, P. and Stokey, N. (1982) Information, trade, and common knowledge, *Journal of Economic Theory*, **26**, 17–27.

O'Hara, M. (1995) *Market Microstructure Theory*. Blackwell Business.

Pitt, M. and Shephard, N. (1998) Analytic convergence rates and parameterization issues for the Gibbs sampler applied to state space models, *Journal of Time Series Analysis*, forthcoming.

Richardson, M. and Smith, T. (1994) A direct test of the mixture of distributions hypothesis: measuring the daily flow information, *Journal of Financial and Quantitative Analysis*, **29**, 101–16.
Ruiz, E. (1994) Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of of Econometrics*, **63**, 289–306.
Shephard, N. (1994) Partial non-Gaussian state space, *Biometrika*, **81**, 115–31.
Shephard, N. and Pitt, M. (1997) Likelihood analysis of non-Gaussian measurement time series, *Biometrika*, **84**, 653–67.
Tauchen, G. and Pitts, M. (1983) The price variability–volume relationship on speculative markets, *Econometrica*, **51**, 485–505.
Taylor, S. (1994) Modeling stochastic volatility: a review and comparative study, *Mathematical Finance*, **4**, 183–204.
Tierney, L. (1994) Markov chains for exploring posterior distribution, *Annals of Statistics*, **21**, 1701–62.

APPENDIX

In this appendix we explain some technical details about our estimation technique. The basic aspects of the multi-move block sampler from Shephard and Pitt (1995) are described first. The parameters of the models are estimated by Bayesian techniques. The prior distributions and the resulting posterior distributions are presented in the second part of this appendix.

The block sampler

Basically, the block sampler of Shephard and Pitt (1997) draws samples from a multivariate distribution by importance sampling. This method is applicable to a wide range of models, including the non-Gaussian state space models described in the text. One of the main advantages of this sampler is that it is considerably faster than a sampler that draws from univariate distributions sequentially. Now consider the SV model with a bivariate measurement equation that we analyse in the text:

$$\begin{aligned} y_t | h_t &\sim N(0, \exp(h_t)) \\ V_t | h_t &\sim Po(m_0 + m_1 \exp(h_t)) \qquad \text{(A1)} \\ h_t = \mu + \varphi h_{t-1} + \eta_t & \qquad \eta_t \sim N(0, \sigma^2_\eta) \end{aligned}$$

If we now let $Y_T \equiv \{y_1, y_2, \ldots, y_T\}$ and $V_T \equiv \{V_1, V_2, \ldots, V_T\}$ and write down the marginal likelihood function for the parameters θ, we eventually obtain:

$$\log f(Y_T, V_T; \theta) \propto \int_{h_1} \cdots \int_{h_T} f(Y_T, V_T | H_T; \theta) f(H_T; \theta)\, \mathrm{d}h_1 \, \mathrm{d}h_T \qquad \text{(A2)}$$

The parameter vector θ containing the relevant hyper-parameters can only be estimated by simulation since we need to integrate out the latent variables or states $H_T \equiv \{h_1, h_2, \ldots, h_T\}$ from the likelihood function. In the special case of Gaussian models we could perform this integration analytically by the Kalman filter.[17] In order to be able to perform inference on either of the parameters or the states in the SV model, we need to solve the integration problem. As analytical methods are impossible we resort to simulating the likelihood value. Suppose we can draw M vectors $(H_T^{(1)}, \ldots, H_T^{(M)})$ from the distribution $f(H_T; \theta)$. An estimate of the integral is then:

$$\frac{1}{M} \sum_{k=1}^{M} f(Y_T, V_T | H_T^{(k)}; \theta)$$

The number M can be set in such a way that the Monte Carlo standard error is smaller than a pre-specified number, see Geweke (1994). The drawing of states from the highly dimensional density function $f(H_T; \theta)$ seems to be a daunting task at first. Recently, a number of techniques have been developed that are able to tackle this problem. One of the most widely investigated ways is to use simulation methods based on Markov Chain Monte Carlo (MCMC). The MCMC methods basically break down highly dimensional problems into smaller, more tractable problems. By combining a series of solutions to the latter we are able to construct an answer to our initial problem. In particular, for our SV model we set up a MCMC for estimating both the states and the parameters. More specifically, we use the Gibbs sampler, which cycles through a series of conditional distributions. Tierney (1994) shows that draws from the conditional distributions converge to draws from the multivariate density. The Bayesian Gibbs sampler in our case can be represented by the following six steps:

1. Set $i = 1$.
2. Get starting values for the parameters $\theta^{(i)}$ and the states $H_T^{(i)}$.
3. For each of the P parameters in $\theta^{(i)}$ draw a new value from the conditional distributions:
$$f(\theta_j^{(i+1)} | Y_T, V_T, H_T; \{\theta_p^{(i+1)}\}_{p=1}^{j-1}, \{\theta_p^{(i)}\}_{p=j+1}^{P}) \quad j = 1, \ldots, P$$
4. Draw $H_T^{(i+1)}$ from $f(H_T^{(i+1)} | Y_T, V_T, \theta^{(i+1)})$.
5. Set $i = i + 1$.
6. Go to 3.

The conditional distributions of the parameters in step 3 are described later in this appendix.

Here we direct our attention to drawing from the conditional distribution of the states H_T. Drawing the vector H_T could then be split up by drawing from T univariate conditional densities $f(h_t | Y_T, V_T, \theta, H_{Tt})$ where H_{Tt} is the state vector with the tth state deleted. This is the approach taken by Jacquier *et al.* (1994). Shephard (1994) argues that

[17] See Harvey (1989) for an extensive overview of applications of the Kalman filter in econometrics.

for state space models these single-move samplers are slower with respect to multi-move samplers, in which a whole vector is sampled at once. Shephard and Pitt (1997) further refine the methods of Shephard (1994).[18]

As Shephard and Pitt (1997) note, direct sampling of the states H_T might suffer from the fact that this distribution is highly degenerate.[19] Consequently, attention is directed towards sampling from the conditional distribution of the errors $\{\eta_1, \eta_2, \ldots, \eta_T\}$ in the state transition Equation A1. For expositional purposes we redefine the state transition equation by:

$$h_t = \varphi h_{t-1} + \sigma_\eta \eta_t \qquad \eta_t \sim N(0,1)$$

As the joint distribution of the errors is highly dimensional, sampling from this distribution is performed in blocks. These blocks are determined by stochastic knots, i.e. the end points change in every round of the general Gibbs sequence described above. The distribution from which we sample is then:

$$\log f(\eta_{t-1}, \ldots, \eta_{t+k-1} \mid h_{t-1}, h_{t+k+1}, y_t, \ldots, y_{t+k}, V_t, \ldots, V_{t+k}) \tag{A3}$$

This conditional distribution is approximated by a multivariate Gaussian density that can be obtained by expanding expression (A3) about an initial point:

$$(\eta_{t-1}, \ldots, \eta_{t+k-1} \mid \hat{h}_t, \ldots, \hat{h}_{T+k})$$

For this reason we rewrite expression (A3) as:

$$\begin{aligned}
&\log f(\eta_{t-1}, \ldots, \eta_{t+k-1} \mid h_{t-1}, h_{t+k+1}, y_t, \ldots, y_{t+k}, V_t, \ldots, V_{t+k}) \\
&\propto \log f(\eta_{t-1}, \ldots, \eta_{t+k-1} \mid h_{t-1}, h_{t+k+1}) \\
&\quad + \log f(y_t, \ldots, y_{t+k}, V_t, \ldots, V_{t+k} \mid h_{t-1}, h_{t+k+1}) \\
&\quad = -\frac{1}{2}\sum_{j=0}^{k} \eta_{t+j-1}^2 + \sum_{s=t}^{t+k} \log f(y_s, V_s \mid h_s) \\
&\quad \approx -\frac{1}{2}\sum_{j=0}^{k} \eta_{t+j-1}^2 + \sum_{s=t}^{t+k} \log f(y_s, V_s \mid \hat{h}_s) \\
&\qquad + (h_s - \hat{h}_s)\frac{\partial \log f(y_s, V_s \mid \hat{h}_s)}{\partial \hat{h}_s} \\
&\qquad + \frac{1}{2}(h_s - \hat{h}_s)^2 \frac{\partial^2 \log f(y_s, V_s \mid \hat{h}_s)}{\partial \hat{h}_s^2}
\end{aligned}$$

This awkward looking density can be calculated using the Gaussian state space model with the following measurement and transition equation:

$$\hat{h}_s - \left(\frac{\partial^2 \log f(y_s, V_s \mid \hat{h}_s)}{\partial \hat{h}_s^2}\right)^{-1}\left(\frac{\partial \log f(y_s, V_s \mid \hat{h}_s)}{\partial \hat{h}_s}\right) = h_s + \varepsilon_s,$$

$$h_s = \varphi h_{s-1} + \sigma_\eta \eta_t \qquad \eta_t \sim N(0,1)$$

The error term ε_s has a normal distribution with mean zero and variance:

$$\mathrm{Var}(\varepsilon_s) = -\left(\frac{\partial^2 \log f(y_s, V_s \mid \hat{h}_s)}{\partial \hat{h}_s^2}\right)^{-1}$$

As this model is Gaussian, the simulation smoother from de Jong and Shephard (1995) can be used to draw from the required proposal density. Subsequently, draws from the smoother can be used in a Metropolis–Hastings accept–reject framework. The states around which the second-order expansion is made, are obtained by iterating the moment smoother of Koopman (1993) to the mode of the density (A3). Shephard and Pitt (1997) show that convergence to this mode occurs quickly.

Priors and Posteriors

In drawing the parameters from the conditional distributions we make frequent use of the pseudo-dominating Metropolis–Hastings algorithm (pseudo-MH) sampler, see Tierney (1994). This sampler is based on the general acceptance–rejection principle, but differs in the sense that the blanket density function does not need to be dominant for the total support of the distribution we want to draw from. In the following we describe the prior and resulting posterior distributions for each of the parameters.

μ

As in Pitt and Shephard (1998) we assume a non-informative prior distribution for μ:

$$f(\mu) \propto c$$

This leads to the normal posterior with the following mean and variance:

$$\mathrm{E}[\mu \mid h_T; \theta_{\backslash\mu}] = \frac{h_1 + \sum_{t=2}^{T}(h_t - \varphi h_{t-1})}{(1-\varphi)\left(T - 1 + \frac{1}{(1+\varphi)^2}\right)}$$

$$\mathrm{Var}[\mu \mid h_T; \theta_{\backslash\mu}] = \frac{\sigma_\eta^2}{(1-\varphi)^2\left(T - 1 + \frac{1}{(1+\varphi)^2}\right)}$$

φ

For the persistence parameter in the latent process H_T we specify a prior on $(\varphi + 1)/2$. When we allow $-1 < \varphi < 1$

[18] In particular, the multi-move samplers in Shephard (1994) suffered from serial correlation between successive draws H_T.
[19] For example, this might occur in an SV model with very high persistence.

this assures that $0 < (\varphi + 1)/2 < 1$. The prior distribution is a beta distribution with parameters δ_1 and δ_2. This leads to the following posterior for the persistence parameter:

$$f(\varphi | h_T, \sigma_\eta^2) \propto \left(\frac{\varphi + 1}{2}\right)^{\delta_1 - 1} \left(\frac{1 - \varphi}{2}\right)^{\delta_2 - 1}$$
$$\times \exp\left(- \frac{1}{2\sigma_\eta^2} \left[(1 - \varphi^2)\left(h_1 - \frac{\mu}{(1 + \varphi)}\right)^2 + \sum_{t=2}^{T} (h_t - \mu(1 - \varphi) - \varphi h_{t-1})^2 \right]\right.$$

Sampling from this distribution is performed using the pseudo-MH with a gamma type blanket function.

σ_η

Here we use an inverse gamma prior with parameters S_0 and p as in Shephard and Pitt (1997):

$$f(\sigma_\eta^2 | \varphi) \propto \sigma_\eta^{2(-p/2+1)} \exp\left(- \frac{S_0}{2\sigma_\eta^2}\right)$$

This leads to the following conjugate posterior:

$$f(\sigma_\eta^2 | H_T, \varphi) \propto \sigma_\eta^{2(-(T+p)/2+1)} \exp\left(- \frac{1}{2\sigma_\eta^2} \left[S_0 + (1 - \varphi^2) \times \left(h_1 - \frac{\mu}{(1 + \varphi)}\right)^2 + \sum_{t=2}^{T} (h_t - \mu(1 - \varphi)^2 - \varphi h_{t-1})^2 \right] \right)$$

m_0

We specify a gamma prior for m_0 with parameters γ_1 and γ_2. This leads to the non-conjugate posterior:

$$f(m_0 | V_T, m_1) \propto \exp\left(- (\gamma_2 + T) m_0 + \sum_{t=1}^{T} \ln\left[m_0^{(\gamma - 1)/T} (m_0 + m_1 h_t)^{V_t} \right] \right)$$

We sample again with the pseudo-MH with a normal distribution as blanket. The mean and variance parameters of the blanket function were obtained by a second–order Taylor expansion of the argument in the exponent.

m_1

The prior for the informed component of trading volume is normal with mean μ and variance v^2. The posterior is again non-conjugate:

$$f(m_1 | V_T, m_0) \propto \frac{1}{v} \exp\left(- \frac{1}{2} \frac{(m_1 - \mu)^2}{v^2} - m_1 \sum_{t=1}^{T} h_t \right) \times \prod_{t=1}^{T} (m_0 + m_1 h_t)^{V_t}$$

Sampling from this distribution is performed by pseudo-MH using a normal blanket function obtained in a similar way as for m_0.