Bayesian Model Estimation and Selection for the Weekly Colombian Exchange Rate

Norberto Rodríguez Niño *

September 2000

Abstract

This document reviews and applies recently developed techniques for Bayesian estimation and model selection in the context of Time Series modeling for Stochastic Volatility. After the literature review on Generalized Conditional Autoregressive models, Stochastic Volatility models, and the relevant results on Markov Chain Monte Carlo methods (MCMC), an example applying such techniques is shown. The methodology is used with a series of Weekly Colombian-USA Exchange Rate on seven different models. The GARCH model, which uses Type-IV Pearson distribution, is favored for the selecting technique, Reversible Jump MCMC, over other models, including Stochastic Volatility Models with a Student-t distribution.

*This work is based on the Master Report presented to The University of Texas at Austin as a requirement for my degree of Master of Science in statistics. I would like to thank Dr. Edward George for his guidance, his advice, and his patience. However, any mistakes that remain are my own. Needless to say, I am very thankful to the Colombian Central Bank for its financial support during my studies at UT.

Contents

1 Introduction **3**

2 Background **4**
2.1 Generalized Autoregressive Conditional Heterocedastic Models. 4
2.2 State Space Models. 7
2.3 Markov Chain Monte Carlo Methods. 9
2.4 Model Selection 9
2.4.1 Markov Chains for Super-Models. 10
2.4.2 Markov Chains with Jumps. 12

3 Methodology **14**
3.1 The Data. 14
3.2 Estimation. 15
3.3 Model Selection. 18

4 Results **19**
4.1 Estimation. 19
4.2 Model Selection. 21

5 Conclusions. **31**

References **32**

1 Introduction

Two different and competing techniques are nowadays used for econometricians and statisticians to model volatility, as in return assets or exchange rates: One of them, Autoregressive Conditional Heteroscedastic (ARCH), its generalization, GARCH, and multiple extensions have proven to be very successful in modeling financial time-varying volatility series.

The competing alternative to GARCH models are Stochastic Volatility models, mainly treated in the frequentist framework, which have more "theoretical" background. They appear in the financial literature on option pricing as a generalization of the Black-Scholes model.

Usually the researcher faces the question of what model to use. Several alternatives have been proposed in the frequentist statistical framework to deal with this, ranging from R^2, the traditional and extensively used Akaike Information Criterion (AIC), PRESS statistic, and many others. In those alternatives the model residuals are obtained (usually observed minus adjusted or perhaps deviance residuals) and aggregated to form the measures of adequacy.

Unfortunately, 'classical' approaches to model choice are limited. The well-known standard Neyman-Pearson theory provides an 'optimal' test through the likelihood ratio for the limited case of the comparison of two distinct models. More generally, the likelihood ratio test enables a choice between models only in the nested case, where there is an unambiguous null hypothesis. Selection is based upon an asymptotic χ^2 approximation, which usually is poor for small sample sizes. Frequentist theory does not offer much for model selection of non-nested models, which are not rare in practice. (See Piorier, 1995, and Gelfand, 1995, for references). This documents uses MCMC, an intuitive, computationally easy-to-implement, and inexpensive Bayesian alternative, to decide among suitable models.

This document is organized as follows: Section 2 presents a review of GARCH models and Stochastic Volatility models, especially those models to be used later in this paper; then it moves on to MCMC methods and Bayesian model selection techniques. Section 3 deals with how to use MCMC to implement estimation of GARCH models, as Vrontos et al. (2000) suggest, and how to estimate SV models. Section 4 deals with the results and section 5 presents the main conclusions, some suggestions and limitations.

2 Background

This section presents a short review of some statistical and econometric models proposed in the literature to model time-varying volatility series. After that, Bayesian techniques for model estimation and selection are briefly presented.

2.1 Generalized Autoregressive Conditional Heterocedastic Models.

Time series models, traditionally fitted in practice, are suppose to have constant variance, but when working with high frequency time series that is seldom the case, as can be seen in Figure 2. Autoregressive Conditional Heterocedastic (ARCH) models have been proposed in the literature to deal with this problem, and those in the spirit of Engel (1982) could be tried; this is

$$y_t = \epsilon_t \text{ or } y_t = c + \epsilon_t \text{ or } y_t = c_t + \epsilon_t \tag{1}$$

as some model for the levels of the observed time series, with c_t some function of lags of y_t or other exogenous variables, or even an ARMA term (see Box and Jenkins, 1976); where $\epsilon_t \sim N(0, h_t^2)$ and $h_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \epsilon_{t-i}^2$, for $\alpha_i \geq 0, i = 0, 1, \cdots, r$. Hence, the conditional variances are thought of as a function of the square of the previous observational residuals. In the original and simple case that normality is assumed, the likelihood is given by:

$$l(y|\alpha_0, \cdots, \alpha_r, h_0^2) = \prod_{t=1}^{T} (2\pi h_t^2)^{1/2} exp(-\epsilon_t^2/2h_t^2) \tag{2}$$

when, in return financial time series analysis, the usual most important unknown to estimate and forecast is the volatility, h_t^2.

A Generalized ARCH (GARCH), that usually results in more parsimonious representations, as Bollerslev (1986) proposed, assumes that the conditional variances follow an ARMA process, thus:

$$h_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j h_{t-j}^2 \tag{3}$$

with the analogous likelihood, as in the ARCH case, if normality for ϵ_t is assumed. With restrictions $\alpha_i \geq 0, i = 0, 1, \cdots, r$, and $\beta_j \geq 0, i = 1, 2, \cdots, s$,

to guarantee that $h_t^2 >= 0$, for all t, and $\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j < 1$, in order to assure stationarity in variance. Weaker restrictions can be required, in practice, though (See Nelson and Cao, 1992).

Another extension of (3) is using a Student-t distribution, with n degrees of freedom to account for the heavier tails of the distribution of the error process $\{\epsilon_t\}$, as was introduced by Bollerslev (1987), and by Baillie and Bollerslev (1989). The likelihood then is

$$l(y|\alpha_0, \cdots, \alpha_r, \beta_1, \cdots, \beta_s, n, h_0^2) = \prod_{t=1}^{T} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(n/2)[\pi\ n\ h_t^2]^{1/2}} \left(1 + \frac{\epsilon_t^2}{n h_t^2}\right)^{-(n+1)/2} \tag{4}$$

An Exponential GARCH (EGARCH), is introduced by Nelson (1991) to avoid imposing restrictions on α_i, β_j in (3) with

$$ln(h_t^2) = \alpha_0 + \sum_{i=1}^{r} \varphi_i(\left|\frac{\epsilon_{t-i}}{h_{t-i}}\right| - E\left|\frac{\epsilon_{t-i}}{h_{t-i}}\right|) + \alpha_i \frac{\epsilon_{t-i}}{h_{t-i}} + \sum_{j=1}^{s} \beta_j ln(h_{t-j}^2) \tag{5}$$

for the conditional variance of $\{\epsilon_t\}$, so $h_t^2 > 0$, for all t, and $E\left|\frac{\epsilon_{t-i}}{h_{t-i}}\right| = \frac{\Gamma(2/\nu)}{[\Gamma(1/\nu)\Gamma(3/\nu)]^{1/2}}$.

Additionally, assuming that ϵ_t follows a Generalized Gaussian Distribution, GED, then the likelihood is:

$$l(y|\alpha_0, \alpha_i, \varphi_i, \beta_j, \nu, h_0^2) = \prod_{t=1}^{T} c\ h_t^{-1}\ exp(-0.5\left|\frac{\epsilon_t}{\lambda h_t}\right|^{\nu}) \tag{6}$$

where $c = \frac{\nu}{\lambda 2^{(1+1/\nu)}\Gamma(\frac{1}{\nu})}$ and $\lambda = \left[2^{(-2/\nu)}\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^{1/2}$.

The use of Bernoulli-Mixtures of two normal distributions, proposed by Ball and Torous (1983), was successfully implemented by Vlaar and Palm (1993):

$$h_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j h_{t-j}^2$$

for the conditional variances. They used an MA(1) term to model the changes in levels of several exchange rates, as:

$$\epsilon_t = y_t - \phi_0 - \lambda\nu - \theta_1 \epsilon_{t-1} \tag{7}$$

and each ϵ_t is distributed as a mixture of two normals, hence

$$\epsilon_t \sim (1-\lambda)N(-\lambda\nu, h_t^2) + \lambda N((1-\lambda)\nu, h_t^2 + \delta^2)$$

where λ is the jump intensity, ν is the expectation in the jump size, and δ^2 the expected change in variance. This representation is useful and intuitive for economies with target or bands for their exchange rates, like Colombia or the European Economic Union. The MA term, $\theta_1\epsilon_{t-1}$, is explained as allowing for mean reversion. Therefore, the likelihood is expressed as:

$$l(y|\phi_0, \theta_1, \alpha_i, \beta_j, \lambda, \delta, \nu, h_0^2) = \prod_{t=1}^{T}(2\pi)^{-1/2}\left[\frac{1-\lambda}{h_t}exp\left(-\frac{(\epsilon_t+\lambda\nu)^2}{2h_t^2}\right) + \frac{\lambda}{(h_t^2+\delta^2)^{1/2}}exp\left(-\frac{(\epsilon_t-(1-\lambda)\nu)^2}{2(h_t^2+\delta^2)}\right)\right] \quad (8)$$

More recently, Bera and Premaratne (2000) propose the use of the Pearson Type-IV Family Distributions in order to model skewness and leptokurtosis that are larger than usual. For a GARCH(1,1), the log-likelihood is:

$$l(y|\Theta) = \sum_{t=1}^{T} l_t(y|\Theta)$$

$$l_t(y|\Theta) = -\ln(\alpha_0) - 0.5\ln(h_t) - \ln(C) + \delta\arctan\left(\frac{y_t-\phi_0-\phi_1 y_{t-1}-\mu h_t}{\alpha_0 h_t}\right) - \left(\frac{r+2}{2}\right)\ln\left[1+\left(\frac{y_t-\phi_0-\phi_1 y_{t-1}-\mu h_t}{\alpha_0 h_t}\right)^2\right] \quad (9)$$

with

$$C = \int_0^{\pi} sin^r\psi\, exp\{-\delta\,\psi\}\, d\psi$$

and $\Theta = (\alpha_0, \alpha_1, \beta_1, r, \delta, \mu, \phi_0, \phi_1)$. They used

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$$

and

$$h_t^2 = 1 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1}^2$$

for their empirical application, but, of course, the model on y_t can be extended.

But according to Nagahara (1999), (9) should be

$$l_t(y|\Theta) = -\ln(\alpha_0) - 0.5\ln(h_t) - \ln(C) + (r+2)\delta \arctan\left(\frac{y_t - \phi_0 - \phi_1 y_{t-1} - \mu h_t}{\alpha_0 h_t}\right) - \left(\frac{r+2}{2}\right)\ln\left[1 + \left(\frac{y_t - \phi_0 - \phi_1 y_{t-1} - \mu h_t}{\alpha_0 h_t}\right)^2\right] \quad (10)$$

with

$$C = \exp[\pi(r+2)\delta/2]\int_0^{\pi} sin^r\psi\, exp[-\delta\,(r+2)\psi]\, d\psi$$

When information arrives at random order and data refers to close-to-close periods, Hsieh (1989) showed that the use of a normal-lognormal mixture distribution improves the fit over other GARCH alternatives. That distribution is not tried here because it requires the computation of a defined integral over $\Re$; one must rely on high numerical integration, however, to provide a suitable solution to that problem.

2.2 State Space Models.

A different alternative in time-series analysis is an State-Space (S-S), model, as West and Harrison (1997) state, they extend and update the seminal paper by Harrison and Stevens (1976). The model is typically represented by two equations:

$$ObservationEquation: \;\; Y_t = F_t'\theta_t + \nu_t, \quad \nu_t \sim N(0, V_t), \quad (11)$$

$$SystemEquation: \;\; \theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim \; N(0, W_t) \quad (12)$$

where y_t is the observed (sometimes latent) variable ; θ_t is a vector of unknown parameters, which follows the first order Markov process (12); ν_t is a vector of unobserved and uncorrelated stochastic error terms; G_t is a matrix of known coefficients; and w_t is an unobserved stochastic term, generally assumed uncorrelated with ν_t.

It is worth noting that any data series for which there exists a natural ordering of observation, fits into the dynamic framework, so the time series not need be equally spaced, and missing data problems can easily be handled

in this context. Pole et al. (1994) present applied methodology, as well as multiple examples.

Usually, and without much loss, V_t can be considered constant, and working in terms of the precision $\phi = V^{-1}$, it is possible to get estimations of W_t. The likelihood for the S-S model is given below:

$$l(y|\Theta) = \prod_{t=1}^{T} l(y|\theta_t) \propto \prod_{t=1}^{T} (2\pi V_t)^{-1} exp\left(-\frac{(y_t - F_t'\theta_t)^2}{2V_t}\right) \tag{13}$$

which is the density of a normal with mean $F_t'\theta_t$ and variance V_t. Prior probabilities can be set up on θ_0 and a fully Bayesian analysis of the State-Space model can be run.

A non-linear (non-gaussian) S-S model can be set up as follows, (See Harvey et al., 1994 for details):

$$y_t = \epsilon_t \exp(h_t/2) \tag{14}$$

as the non-linear observation equation, and

$$h_t = \gamma + \phi h_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \tag{15}$$

as the system equation, where $h_t = ln(\sigma_t^2)$.

This model, which is known in the financial and econometric literature as the Stochastic Volatility model (SV for short) can be transformed to get a linear observational equation, as

$$\ln(y_t^2) = h_t + \ln(\epsilon_t^2) \tag{16}$$

where y_t is the mean corrected return at time t; h_t is the log-volatility at time t, which is assumed to follow a stationary process, with h_0 drawn from a stationary distribution; ϵ_t and η_t are uncorrelated standard normal white noise shocks; ϕ is the persistence in the volatility (when for stationarity restriction $|\phi| <1$); and σ_η^2 is the volatility of the log-volatility.

The likelihood, assuming normal distribution, can be obtained by using a Kalman Filter (See Jaquier et al., 1994, Ap. B.1). The use of a ν degrees of freedom Student-t distribution on ϵ has been considered, and the Kalman Filter needs some minor modifications[1] (See Ruiz, 1994).

[1] E.g., the use of digamma and trigamma functions; Abramowitz and Stegun (1967) offer computational details.

2.3 Markov Chain Monte Carlo Methods.

When doing fully Bayesian analysis of complex or high-dimensionality models, the researcher usually faces the problem of non-conjugacy, meaning that non-exact analytical posterior distribution can be achieved. This leads to the necessity of using simulation approaches. Direct simulation is often impossible, due to the complicated mathematical form of the posterior distribution in many applied models. Because of that, an exponential rise in the interest and application of Markov Chain Monte Carlo (referred to by its acronym, MCMC) as a tool for numerical computation of complex integrals, particularly in Bayesian analysis, has emerged.

The key to Markov Chain simulation is to create a Markov process whose stationary distribution is a specified $\pi(\theta|y)$ and to run the simulation long enough that the distribution of the current draws is close enough to the stationary distribution. Once the simulation algorithm has been implemented, it should be iterated until convergence has been reached, or, if convergence is painfully slow, the algorithm should be altered. Hence, the study of MCMC has seen a corresponding interest in the convergence properties of the resultant chains, which may be assessed through a suite of diagnostics borrowed from diverse areas such as time series, exploratory data analysis (EDA), and central limit theory.

The most widely used Markov Chain Monte Carlo methods are the Gibbs Sampler and the group of Metropolis-Hastings algorithms and a good description of them can be found in Gamerman (1997) or Gelman et al. (1995), among many others.

2.4 Model Selection

> *The most damaging comment on the standard practice of choosing a single model, and then proceeding conditional on it, is that the research's uncertainty is understated.*
> — Piorier (1995, p. 605)

For frequentists the model selection problem reduces to choosing one from a set of M models. This is usually the main aim of the analysis, and is done according to some model selection criterion, as stated above. Bootstrap methods have been used for model selection (See Maddala and Li, 1996, sec 5, for references). For another perspective see Poskitt and Tremayne (1983).

Essentially, two alternative approaches in the Bayesian context are presented. The first was introduced by Carlin and Chib (1995) and considers all

models in a formation, called here a supermodel. The Markov Chain simulation scheme for this supermodel is presented below. The second approach presents sophisticated simulation techniques using Markov chain with jumps between the different models; it is referred to as Reversible Jumping, and it was introduced by Green (1995).

It will be assumed throughout this section that y is observed and it can be described according to a model M_j with parameters θ_j of dimension d_j, taking values in a parameter space $\Theta_j \subset \Re^{d_j}, j = 1, 2, \ldots, M$. The value of M could be ∞ as, for instance, when considering countable classes of models. m serves the purpose of indicating a specific model.

Assume for the moment that the posterior distribution $\pi(\theta, j)$, the joint distribution of the super-parameter and the model indicator are to be obtained. However, the main interest in inference is to obtain the posterior distribution of $\theta_j | m = j,\ j = 1, 2, \ldots, M$. These distributions respectively provide the posterior inference within each of the models and the posterior probabilities of the models. The supermodel approach provides a sample from this more general, perhaps unnecessary posterior distribution whereas the approach with jumps only provides samples from $\theta_j | m = j,\ j = 1, 2, \ldots, M$, and m. The presence of common parameters does not pose any problem here.

2.4.1 Markov Chains for Super-Models.

The joint distribution of all random quantities is given by

$$\pi(y, \theta, j) = \pi(y|\theta, j)\pi(\theta|j)\pi_j \tag{17}$$

where j is the value of m and $\pi_j = P(m = j)$. Given that $m = j$, the distribution of y depends on θ only through θ_j, or mathematically,

$$\pi(y|\theta, j) = \pi(y|\theta_j, j) \tag{18}$$

Assume also that the θ_j are conditionally independent, given the value of m. Hence,

$$\pi(\theta|j) = \prod_{i=1}^{M} \pi(\theta_i|j)$$

Note that the prior distribution $\pi(\theta_i|j)$, for $i \neq j$ does not make much sense. It specifies the distribution of the parameters of model i, conditioned on the fact that this is not the true model. Carlin and Chib (1995) refer to these as pseudo-prior or linking distributions. Due to the conditional independence

(18), these priors do not interfere in the expressions of the marginal predictive densities for each model. Nevertheless, they are relevant for the construction of the chain and must be specified.

It follows from the above specification that

$$\pi(y, \theta, j) = \pi(y|\theta_j, j) \prod_{j=1}^{M} \pi(\theta_i|j)\pi_j$$

which is proportional to the joint posterior distribution of θ and m. A natural blocking is formed by grouping each model's parameters and m. The full conditional distributions for $\theta_1, \theta_2, \cdots, \theta_M$ and m are obtained as follows:

- For block $\theta_j,\ j = 1, \ldots, M$,

$$p_j(\theta_j) \propto \begin{cases} \pi(y|\theta_j, j)\pi(\theta_j|j), & \text{for } m = j \\ \pi(\theta_i|i), & \text{for } m = i \neq j \end{cases}$$

- For block m

$$p_M(j) = k^{-1}\pi(y|\theta_j, j) \prod_{j=1}^{M} \pi(\theta_i|j)\pi_j,\ j = 1, \ldots, M$$

  that is a discrete distribution with proportionality constant

$$k = \sum_{l=1}^{M} \pi(y|\theta_l, l) \prod_{i=1}^{M} \pi(\theta_i|l)\pi_l$$

m can always be sampled directly because it has a discrete distribution. Direct sampling from blocks θ_j will depend on the conjugacy structure for model $m = j$ and the form of the pseudo prior distribution. When direct sampling for some of the θ_j's is not possible, Metropolis-Hastings steps may be used.

The above scheme satisfies the conditions of a conventional Markov Chain and therefore converges to the target distribution given by the posterior. Comparison between models is based on the marginal posterior distribution of m, $p(j),\ j = 1, \ldots, M$. These probabilities are estimated by the proportion of values of m equal to j in the sample of size n.

The pseudo prior distributions must be carefully chosen, as they affect the rate of convergence of the chain. Carlin and Chib (1995) recommend

the use of simple standard approximations based on univariate estimates obtained from pilot chains. The same authors suggest using fairly vague prior distributions, but it is well-known that when using this practice on models with different dimensions the Bayes factors turn out to be very sensitive. So, this prior setting may need further justification to satisfy potential users.

Finally, this approach is not applicable to the case of countable number of models under consideration. Hence, the number of practical and theoretical difficulties of this approach suggest it should be used with care. See Gamerman (1997), where more details can be found.

2.4.2 Markov Chains with Jumps.

Green (1995) introduced a reversible-jump MCMC strategy for generating from the joint posterior $\pi(m, \theta_m|y)$, based on the standard Metropolis-Hastings approach. The reversible-jump MCMC was also applied by Richardson and Green (1997) for an analysis of univariate normal mixture; by Nobile and Green (2000), for factorial experiments using mixture modeling; and Dellaportas and Forester (1999), for analysis of contingency tables. During reversible-jump MCMC sampling, the constructed Markov Chain moves within and between models, so that the limiting proportion of visits to a given model is the required $\pi(m|y)$.

In general, suppose that the current state of the Markov Chain at time t is (m, θ_m), where θ_m has dimension $d(\theta_m)$ and a move is proposed at time $t+1$ to a new model m' with probability $j(m, m')$ and corresponding parameter vector $\theta'_{m'}$. Then, a vector u is generated from a specified proposal density $q(u|\theta_m, m, m')$, and $(\theta'_{m'}, u') = g_{m,m'}(\theta_m, u)$ is set for a specified invertible function $g_{m,m'}$ such that $g_{m,m'} = g^{-1}_{m,m'}$. Note that $d(\theta_m) + d(u) = d(\theta'_{m'}) + d(u')$. Green (1995) showed that, if the new move is accepted as the next realization of the Markov Chain with probability $\alpha = min\{1, r\}$, where

$$r = \frac{\pi(y|m', \theta'_{m'})\pi(\theta'_{m'}|m')\pi(m')j(m', m)q(u'|\theta'_{m'}, m', m)}{\pi(y|m, \theta_m)\pi(\theta_m|m)\pi(m)j(m, m')q(u|\theta_m, m, m')}|J| \qquad (19)$$

with $J = \partial(\theta'_{m'}, u')/\partial(\theta_m, u)$ denoting the Jacobian of the transformation, then the chain satisfies the condition of detailed balance and has the required limiting distribution $\pi(m, \theta_m|y)$. The condition of detailed balance requires that the equilibrium probability of moving from a state (m, θ_m) to $(m', \theta'_{m'})$ equals that of moving from $(m', \theta'_{m'})$ to (m, θ_m); for details, see Green (1995).

To implement the reversible-jump MCMC, the probabilities $j(m, m')$ need to be specified for every proposed move, as well as the proposal

distributions $q(u|\theta_m, m, m')$, $q(u'|\theta'_{m'}, m', m)$ and the function $g_{m,m'}$. These choices do not affect the results in terms of models selected but may affect crucially the convergence rate of the Markov Chain. For the probability $j(m, m')$, one non-informative alternative is $j(m, m') = (M-1)^{-1}$, for all $m,\ m' \in M$, when at each state of the chain a move from one model to other one is always proposed.

Vrontos et al. (2000) proposed a modification of Green's technique, which they have successfully implemented in a series of experiments with GARCH and EGARCH models; this is described as follows: First, they suggest that all the parameters of the proposed model be generated from a proposal distribution. Consequently, $(\theta'_{m'}, u') = (u, \theta_m)$ with $d(\theta_m) = d(u')$ and $d(\theta'_{m'}) = d(u)$, $q(u|\theta_m, m, m') = q(u|m')$, $q(u'|\theta'_{m'}, m', m) = q(u'|m)$, and the Jacobian in (19) is 1. In this case, the probability of acceptance of the new move as the next realization of the Markov chain is given by $\alpha = min\{1, r\}$, where

$$r = \frac{\pi(y|m', \theta'_{m'})\pi(\theta'_{m'}|m')\pi(m')j(m', m)q(u'|m)}{\pi(y|m, \theta_m)\pi(\theta_m|m)\pi(m)j(m, m')q(u|m')} \tag{20}$$

The proposal densities $q(u|m')$ and $q(u|m')$ can be chosen by investigation of a "pilot run." They start the chain from the best available starting values (e.g., the maximum likelihood estimates when available) and simulate the "within-model" Markov Chain many times to obtain approximate marginal posterior means and covariance matrices for each model parameter vector. These estimates are then used to construct proposal densities $q(u|m')$ and $q(u'|m)$, taken as multivariate normal densities.

3 Methodology

3.1 The Data.

In order to illustrate the estimation and selecting methodologies, the weekly observations of the US/Colombian (on spot) exchange rate are used. Let er_t represent the Fridays exchange rates, running from October 21, 1991, through December 29 of 1999; that made $T = 428$ observations. The daily exchange rate corresponds to the weight average of trading, selling and buying, of U.S. dollars in the open market. In the event of the market being closed on a Friday, the observation on the previous Thursday was used.

Figure 1 shows the raw data; from this and with the help of a unit root test, (See Enders, 1995), it is easy to conclude that the exchange rate has a unit root, so, as usual, one works with the first difference of the natural logarithm of exchange rates, returns, also known as continuously compounded rates of return, ($r_t = ln(er_t) - ln(er_{t-1})$ for $t = 1, \ldots, T$), whose representation is shown in Fig. 2.

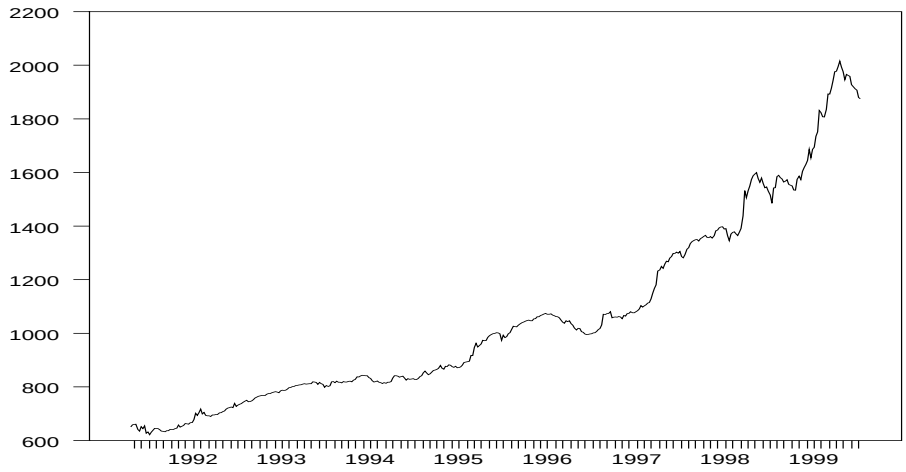

Figure 1: Weekly Colombian Exchange Rate.

From Figure 3, which shows the histogram of returns, it is noteworthy that the skewness and kurtosis coefficients for r_t, which are 0.8581 and 5.6872, respectively, are both significantly positive and much larger than common, thus showing asymmetry and leptokurtosis. As pointed out by Vlaar and

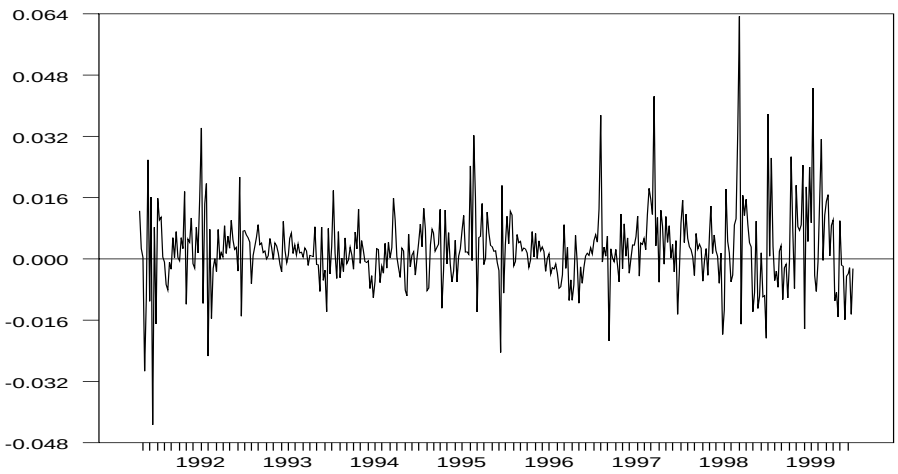

Figure 2: Weekly Return from Colombian Exchange Rate.

Palm (1993), the skewness could be the result of the asymmetry in the movements of the parity adjustments, and a high kurtosis could result from a time varying-variance. These two results lead to considering those distributions different from the normal, whose use is unlikely to yield appropriate results.

Also, it is clear that the variance is not constant at all, which can be seen from the Lunjg-Box's autocorrelation statistic of the squared returns: $Q^2(12) = 28.32$ and $Q^2(24) = 44.86$ for lags 12 and 24, respectively.

3.2 Estimation.

Based on the theoretical justification of several models and taken in account the particular Colombian economy, seven different models were proposed for estimation and selection among them; They are described below.

With the aim to implement a fully Bayesian analysis, MCMCs were used to estimate the models [2], much of them are estimated by the first time from the Bayesian point of view, which is a new development of this work.

First, an ARMA(1,1) model with normal disturbances was fitted; this model could be tried in practice when the scholar ignores the fact that the variance is not constant. In this report, such a model is also used as the

[2]Maximum Likelihood Estimation (MLE) was tried, but problems getting precise estimations of the variance-covariance matrix, for some models, restricted its use.

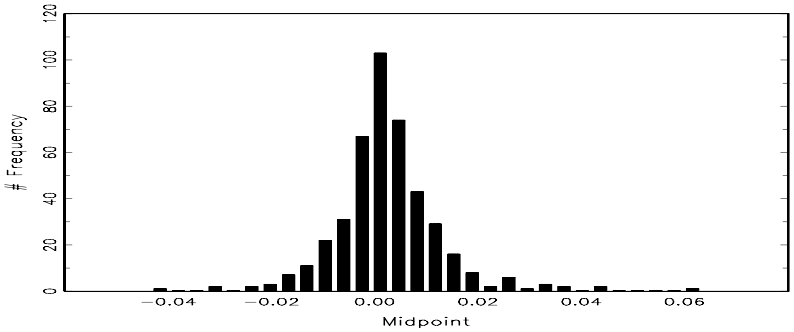

Figure 3: Histogram of return.

reference model, against which others are to be compared. Non-informative Beta(1,1) from -1 to 1 [3] priors were chosen for ϕ_1 and θ_1. A N(0,5) as prior for ϕ_0 was used; finally, a nearly non-informative but proper Inverse-Gamma(2.001,0.001), (So, Expected value and variance equal to 1/1000) was used for σ^2.[4]

Second, an ARMA(1,1) for the levels of return plus a GARCH(1,1) for the variance, so, c_t in (1) comes from an ARMA(1,1), and $V(\epsilon_t) = h_t^2$ from (3) with $r = 1$, and $s = 1$; still assuming normality, the likelihood is given by (2). Non-informative $U(-1, 1)$ prior for ϕ_1 and θ_1; $1/\alpha_0$ for α_0; a N(0,5) for ϕ_0; $U(0, 1)$ for α_1, and β_1, and in order to assure stationarity in (3), any proposal that did not fit $\alpha_1 + \beta_1 < 1$ is rejected. The same priors on ϕ_0 as before was used. From initial runs h_0^2 turned out to be statistically equal to zero, thus it is set to $\frac{\alpha_0}{(1-\alpha_1)}$, as Nelson and Cao (1992) propose for computational convenience.

Third, an ARMA(1,1) plus a GARCH(1,1) with Student-t distribution, and $\alpha_1 + \beta_1 >= 1$ are rejected in (3) for stationarity purposes; $n > 2$ in (4); h_0^2 was set as in before. Priors: $1/\alpha_0^2$ for α_0; $1/(n-2)^2$ on n, and α_1, β_1 as in the second model. ϕ_0, ϕ_1, and θ_1 were treated as in the first model.

Fourth, an ARMA(1,1) plus an EGARCH(1,1), (5) instead of (3); with GED distribution, therefore, the likelihood is given by (6). Priors: ϕ_0, ϕ_1,

[3] in the sense that if $X \sim Beta(1, 1)$ then $Y = 2X - 1 \sim$ Beta(1,1) from -1 to 1. See Jaquier et al. (1999)

[4] $1/\sigma^2$ as a prior for σ^2 was tried too, with similar results

and θ_1 as above; a N(0,1) was used as prior for α_0; N(0,5) for α_1, for β_1 an U(-1,1); φ_1 is assumed identically to zero; and U(0,80) for ν. Any proposal with $|\beta_1| > 1$ was rejected for stationarity purposes.

Fifth, an ARMA(1,1) plus an GARCH(1,1) with a Mixture of two normals as the error term distribution using (8) as the likelihood. Priors: ϕ_0, ϕ_1, and θ_1 as before. $1/\alpha_0^2$ for α_0; for α_1, and β_1, $U(0,1)$; N(0,10) for ν; $1/(\delta^2)^2$ for δ^2; finally, $U(0,1)$ for λ.

Sixth, an ARMA(1,1) plus an GARCH(1,1) with Type-IV Pearson distribution, hence, the likelihood is computed by using (10). Priors: Inverse-Gamma(2.001,0.0001) was used for α_0, $U(0,1)$ for α_1, and β_1; $(r-2)$ following an Inverse-Gamma(2.001,0.0001); N(0,1) for δ, and N(0,5) for μ. The distributions used previously for ϕ_0, ϕ_1, and θ_1 were used here too. The restriction $r > 3$ is imposed, so the first four moments from the Type-IV Pearson distribution exist. Again, any proposal with $\alpha_1 + \beta_1 >= 1$ is rejected. $\alpha_0 < 0.0001$ were rejected for computational reasons.

Seventh, and finally, a Stochastic Volatility model with a Student-t distribution on the error term, expressed in (16) and (15). Non-informative prior $U(-1,1)$ for ϕ, a $N(0,5)$ for γ; Inverse-Gamma(2.001,0.0001) for σ_η^2; and for ν a $U(5,80)$ was used as Jaquier et al. (1999) did to assure the t-Student has at least the first-four moments.

Although the use of single versus multiple (parallel) chains in MCMC is an open discussion, single chains for each model are used in this job because of time and computing resources limitations. For a recent discussion of this dilemma, the reader is referred to Mengersen et al. (1999), which contains points in favor of each alternative.

Raftery and Lewis' (1996) strategy was implemented here with constant c selected in such a way that the proportions of the proposals accepted were between 20 and 50%, as has become common practice. The update is element-by-element and in random order. However, when high correlations between parameters in conjunction with slow convergence were found the blocking update was implemented to improve convergence.

In every case, a final chain of 80,000 was run and then steps of 50 to 300 were taken to avoid large autocorrelations in the chains. In that way, first-order autocorrelations no larger than 0.55 were guaranteed.

Convergence of each chain is assessed by applying the Geweke's (1992) criterion, which null hypothesis is that stationarity has been reached, and the test-statistic is suppose to follow an standard normal distribution under H_0. This test is implemented by using CODA (See Best et al., 1997).

The mean-vector and Variance-covariance matrices are to be obtained in

order to feed or implement the RJMCMC, as explained later.

3.3 Model Selection.

The model selection exercise consists of applying the Reversible Jump MCMC algorithm and the posterior probabilities, running 200,000 iterations and showing the proportion of each 2,000 that model m (m=1, 2, ... 7) is selected. For checking stability, visual analysis is used. Although there are some fresh results about assessing convergence in RJMCMC, their value are not well-known yet, as mentioned by Brooks and Guidici (1999).

For all seven models the same priors mentioned in Section 3.2 are to be used. The proposal densities $q(u|m')$ and $q(u'|m)$ for each parameter were constructed by using the MCMC output of the separate model runs described above. These densities are taken as multivariate normals with mean vectors, consisting of the sample mean values and covariance matrix equal to the corresponding sample mean vector and covariance matrix of the parameters in each model.

4 Results

4.1 Estimation.

It took between 12 hours 24 minutes and 56 hours and 48 minutes, from the fastest to the slowest model, to made all the iterations, using a computer with a Pentium I 233 MHz processor, and 64 MB RAM, running under WINDOWS-98 Second Edition. Times reduce to one third using a Pentium III 700 MHz processor, with the same software and 192 MB RAM.

In the following presentation the return rates are expressed in 0-100 scale, which was used because of computational and presentational avenues; otherwise, models which use GARCH component get stuck, it seems because values for α_0 go so close to zero that the algorithm get overwhelmed.

Table 1: Geweke's Convergence z Scores for the Seven Models.

Model	ϕ_0	ϕ_1	θ_1	α_0	α_1	β_1	σ^2	n	
ARMA(1,1):Normal	-0.275	-0.601	0.118				-0.77		
ARMA(1,1)+									
GARCH(1,1):N	-.594	0.966	-1.13	-.002	0.752	-.186			
ARMA(1,1)+									
GARCH(1,1):t	-.948	0.624	-.374	-.439	-.494	0.667		-.873	
Model		ϕ_0	ϕ_1	θ_1	α_0	α_1	β_1	ν	
ARMA(1,1)+									
EGARCH(1,1):GED		0.079	-1.01	1.12	1.28	-1.42	-.581	-1.13	
Model	ϕ_0	ϕ_1	θ_1	α_0	α_1	β_1	λ	ν	δ^2
ARMA(1,1)+									
GARCH(1,1):MixN	2.43	2.40	-2.20	-2.90	-.276	1.08	1.57	-2.17	-2.39
Model	ϕ_0	ϕ_1	θ_1	α_0	α_1	β_1	r	δ	μ
ARMA(1,1)+									
GARCH(1,1):T-IV P	1.24	0.586	-.538	-0.092	0.801	-1.53	-1.66	-0.82	-.308
Model			ϕ	γ	ν	σ_η^2			
STOCHASTIC VOL.:t			1.70	-.713	-1.18	-1.52			

The Geweke's Convergence z Scores for the seven models are presented in Table 1, looking to the numbers it is clear that convergence has been reached for almost every parameter in all models.

The estimation results are presented in Table 2 with standard errors in parentheses. [5] Although, parameter transformations [6] were tried for some models, convergence was not improved, hence it use was discharged. From Table 2 it should be said that except for the non-inclusion of some parameters in most of the models, no additional work for the exclusion of non-significant parameters in any model was attempted because time limitations, and because it is not the main purpose of this work to improve every and/or one specific model.

Table 2: Estimation Results of the Seven Competing Models.

Model	ϕ_0	ϕ_1	θ_1	α_0	α_1	β_1	σ^2	n	
One	0.269	-.107	.112				0.103		
	(6.6e-3)	(0.022)	(0.010)				(3.1E-3)		
Two	0.295	-.586	0.564	0.203	0.490	0.398			
	(.0041)	(.0179)	(0.016)	(.0024)	(0.004)	(0.004)			
Three	0.153	0.151	-.075	0.119	0.623	0.525		3.20	
	(0.005)	(0.024)	(0.021)	(.008)	(0.01)	(.006)		(0.06)	
Model		ϕ_0	ϕ_1	θ_1	α_0	α_1	β_1	ν	
Four		0.045	0.578	-.441	-.484	0.419	0.431	0.864	
		(0.003)	(.032)	(.029)	(.009)	(0.007)	(0.009)	(0.005)	
Model	ϕ_0	ϕ_1	θ_1	α_0	α_1	β_1	λ	ν	δ^2
Five	0.256	-0.022	0.065	0.047	0.321	0.475	0.247	0.525	2.33
	(.013)	(.052)	(.044)	(.009)	(.011)	(.017)	(.015)	(.046)	(.182)
Model	ϕ_0	ϕ_1	θ_1	α_0	α_1	β_1	r	δ	μ
Six	-.079	-.0627	1.9e-3	0.111	2.9e-4	0.999	3.17	0.063	-6.1e-3
	(.009)	(.015)	(.013)	(3.8e-4)	(1.1e-5)	(1.8e-5)	(8.1e-3)	(4.3e-3)	(7.4e-4)
Model			ϕ	γ	ν	σ_η^2			
Seven			-.196	0.223	60.70	0.007			
			(0.023)	(0.109)	(0.53)	(.0016)			

[5] Such standard errors refer to the time-series estimates which are asymptotic, the square root of the spectral density estimate divided by the sample size.

[6] Like logarithm or $\phi' = ln(\frac{1+\phi}{1-\phi})$, when $|\phi| < 1$, hence $\phi' \in \Re$.

Fig. 4 [7] presents the resulting chains diagrams and the histograms of the posterior sample of the parameters of the ARMA(1,1) model. The shape of the posterior distribution of ϕ_0 and σ^2 parameters indicate asymmetry, hence deviation from normality. Fig. 5 does the same for GARCH-N Model, Fig. 6 does the same for GARCH-MixN Model, Fig. 7 does the same for EGARCH-GED Model, Fig. 8 does the same for GARCH-MixN Model, Fig. 9 for Model GARCH Type-IV Pearson, and Fig. 10 for SV-t model.

4.2 Model Selection.

The processing time for 80,000 iterations, using the same computer as for estimation, was 16 hours and 23 minutes, for all the seven models, a total of 47 parameters which use above 32 MB of disk-space. Note that, according to Fig 11, this is a conservative run length, and less than one-fourth of the run could be sufficient to achieve the same posterior distributions.

Table 3: Posterior Probabilities and Bayes Factors of Seven Competing Models.

Model	Distribution	*Posterior Prob.*	*Bayes Factor*
ARMA(1,1)	NORMAL	0.00001	1.000
ARMA(1,1)+GARCH(1,1)	NORMAL	0.00001	1.000
ARMA(1,1)+GARCH(1,1)	t	0.00003	3.000
ARMA(1,1)+EGARCH(1,1)	GED	0.00001	1.000
ARMA(1,1)+GARCH(1,1)	Mixt. Normal	0.00005	5.000
ARMA(1,1)+GARCH(1,1)	T.-IV Pearson	0.99990	199989.000
STOCHASTIC VOL	t	0.00001	1.000

The RJMCMC results and Bayes factor are displayed in Table 3; which shows the posterior probabilities and Bayes Factors for the seven models. The last column refers to the relative weight against the worst models, ARMA(1,1), SV, and ARMA+GARCH. According to this results it is very clear that model six, ARMA(1,1)+GARCH(1,1) with Type-IV Pearson distribution over perform the rest of them, with posterior probability 100%.

[7] In this graph, as well as in the rest in this Chapter, the next convention is used: T_1 refers to ϕ_0, T_2 to ϕ_1, T_3 to θ_1, and T_4 to σ^2. Hence, for example T_9 in Fig. 8 is used to note δ^2.

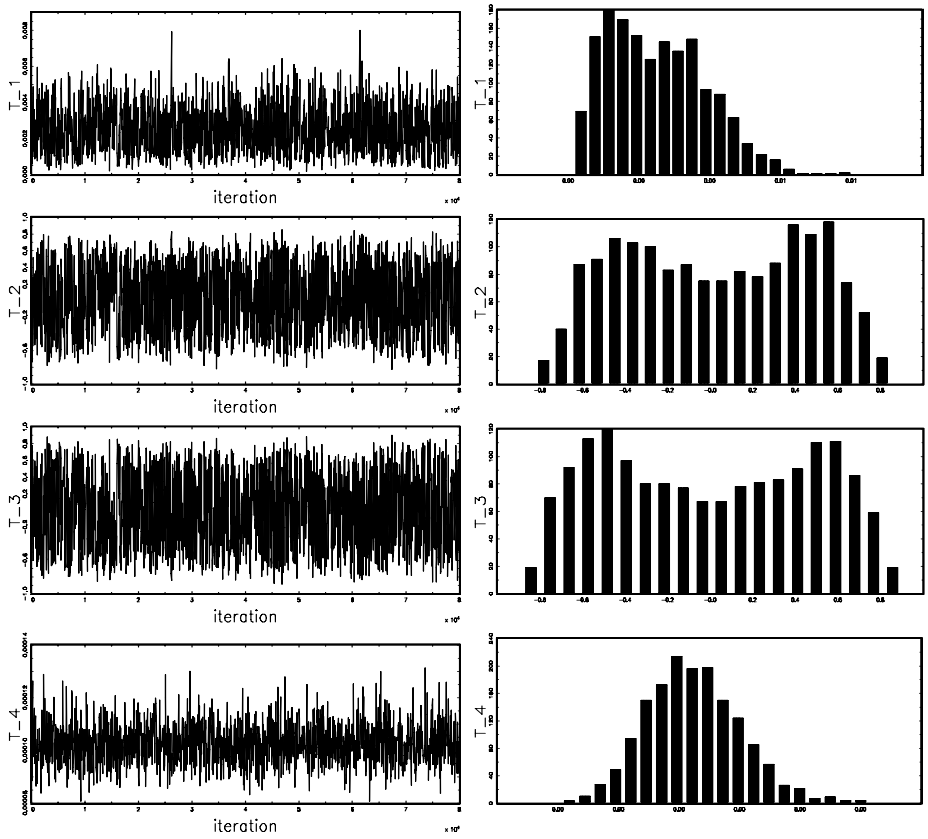

Figure 4: Convergence Diagrams and Histograms of the Posterior Sample for ARMA(1,1) Model

Figure 11 shows the convergence behavior of the chain. That figure illustrate the probability of each of the seven models across the sweeps calculated ergodically every 2,000 iterations. Note that the only model that is visited very often in the Reversible Jump MCMC algorithm is model six.

Next exercise consist in rerun the chain, this time with the six less probable models, that is to say all but model six. Analogous results are obtained, this time that outperforming model being model five, ARMA(1,1)+ GARCH(1,1) with Normal Mixture Distribution. Neither figure nor table are presented for this case.

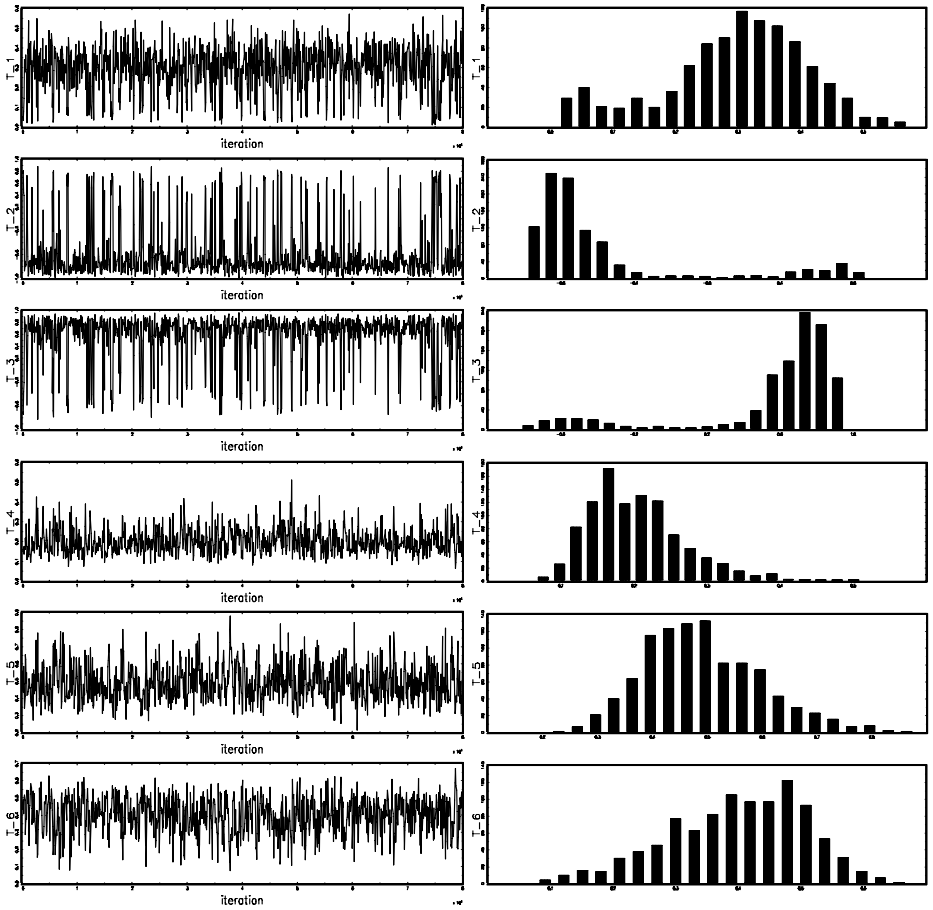

Figure 5: Convergence Diagrams and Histograms of the Posterior Sample for GARCH-Normal Model

Additionally, a chain of the same length was run with only the less probable five models. This time the favored model is ARMA(1,1)+GARCH(1,1) with Student-t distribution.

Finally, an exercise with models one, two, four, and seven was run. Figure 12 and Table 4 show the exercise result with only the four more improbable models. In this case the transition probabilities, $j(m, m')$, were taken as inversely proportional to the number of parameters on each model. This

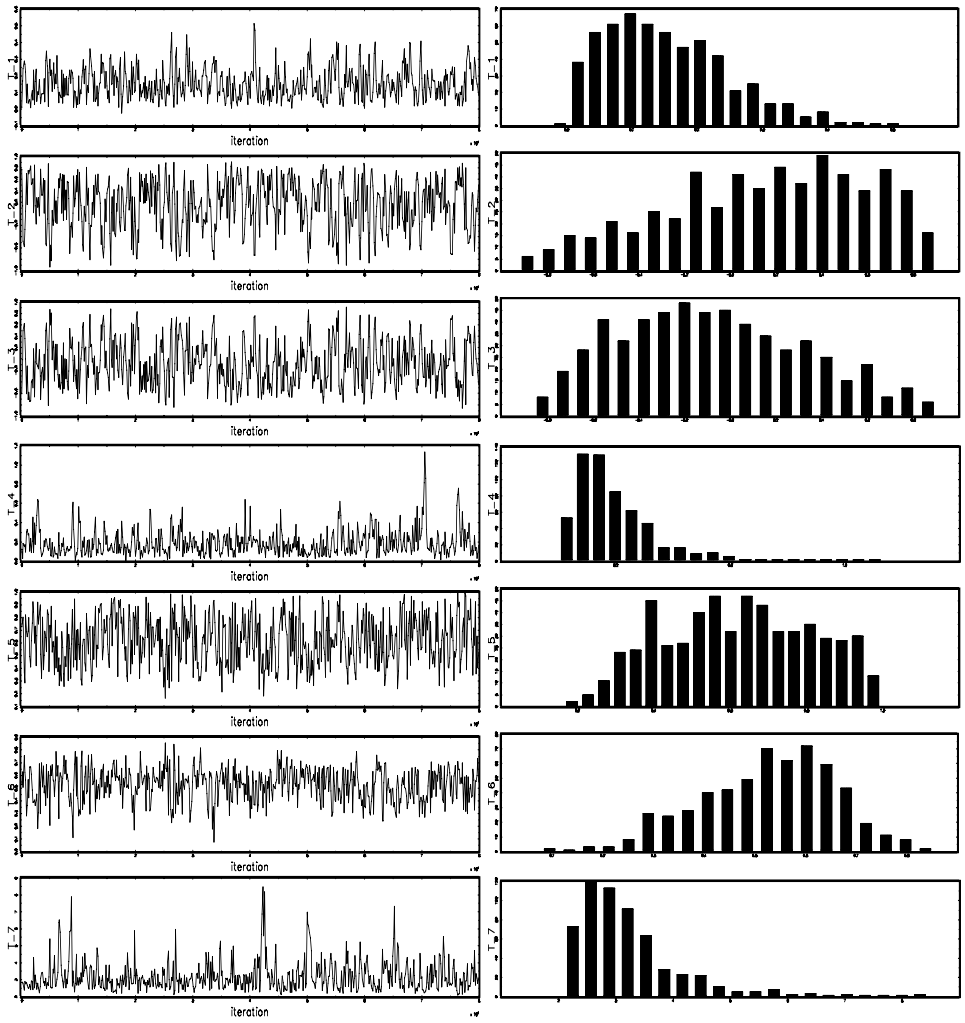

Figure 6: Convergence Diagrams and Histograms of the Posterior Sample for GARCH-t Model

exercise provide evidence that model four is *a posteriori* the fourth most probable after models six, five, and three, in that order.

Table 4: Posterior Probabilities and Bayes Factors of Four of the Seven Competing Models.

Model	Distribution	*Posterior Prob.*	*Bayes Factor*
ARMA(1,1)	NORMAL	0.0097	3.070
ARMA(1,1)+GARCH(1,1)	NORMAL	0.1796	56.754
ARMA(1,1)+EGARCH(1,1)	GED	0.8075	255.132
STOCHASTIC VOL	t	0.0032	1.000

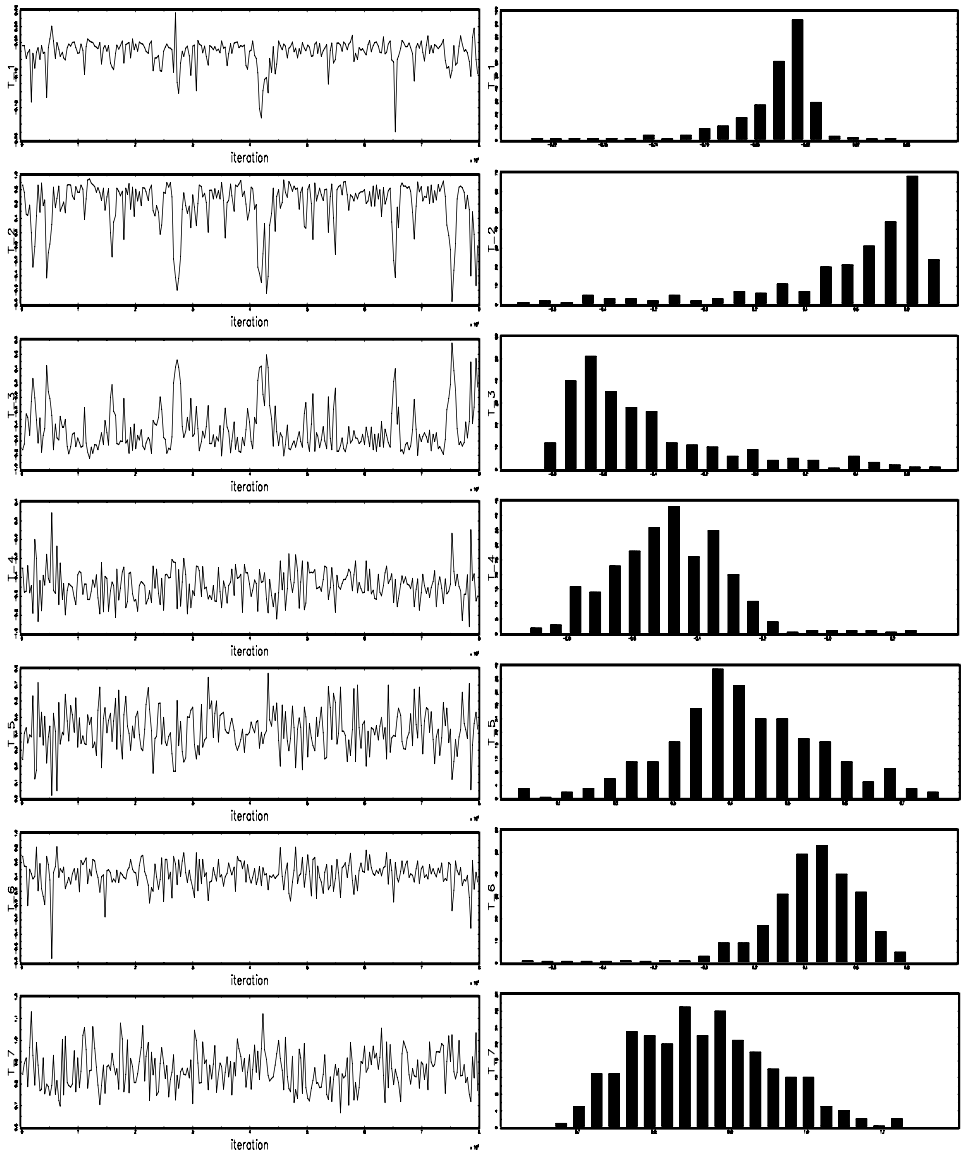

Figure 7: Convergence Diagrams and Histograms of the Posterior Sample for EGARCH-ged Model

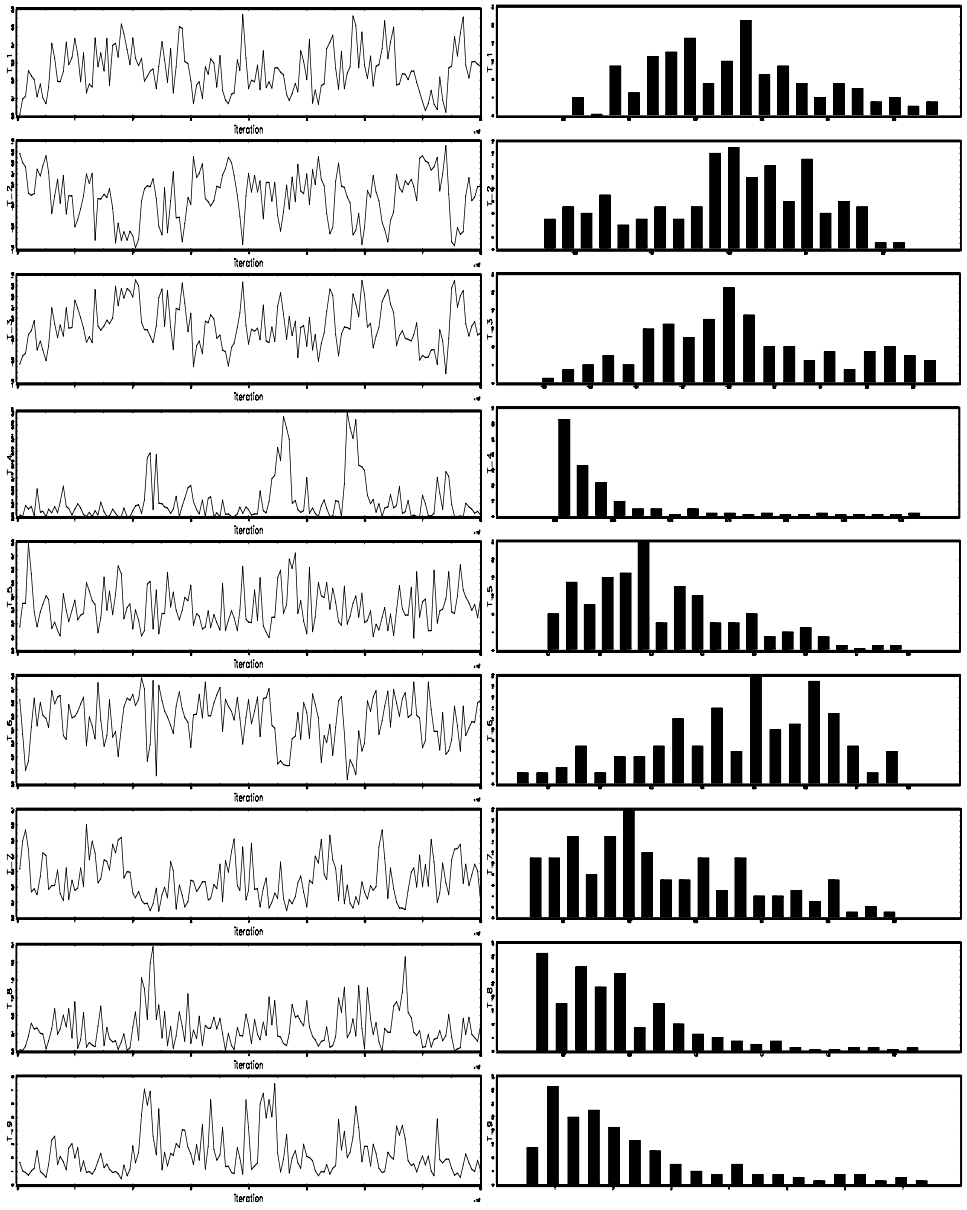

Figure 8: Convergence Diagrams and Histograms of the Posterior Sample for GARCH-MixN Model

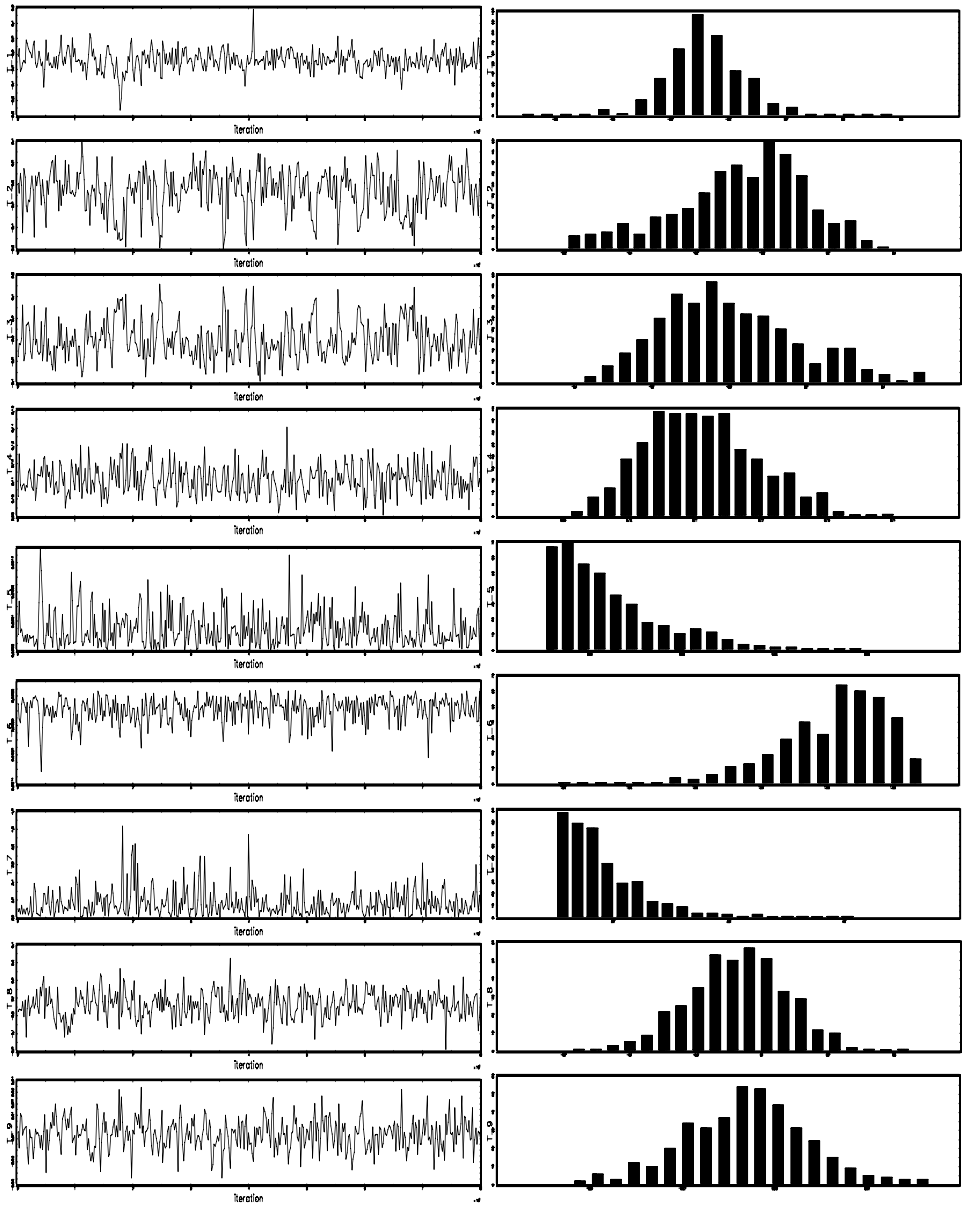

Figure 9: Convergence Diagrams and Histograms of the Posterior Sample for GARCH-T4P Model

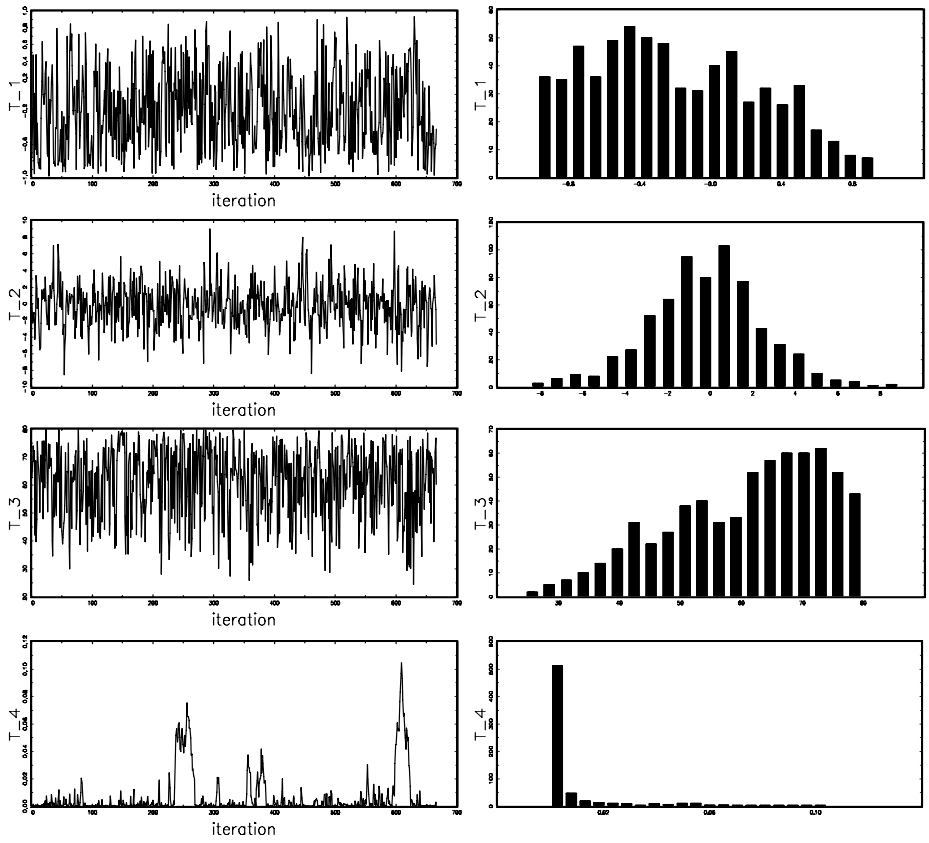

Figure 10: Convergence Diagrams and Histograms of the Posterior Sample for SV-t Model

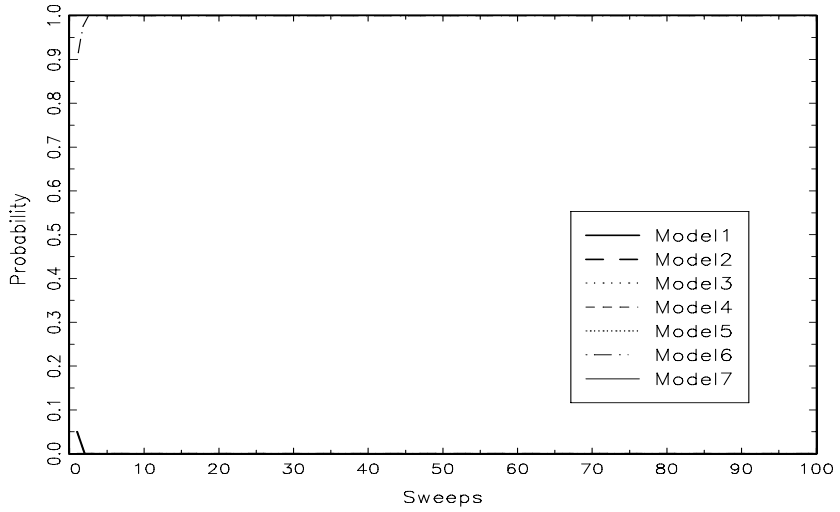

Figure 11: Convergence Behavior of the Seven Models

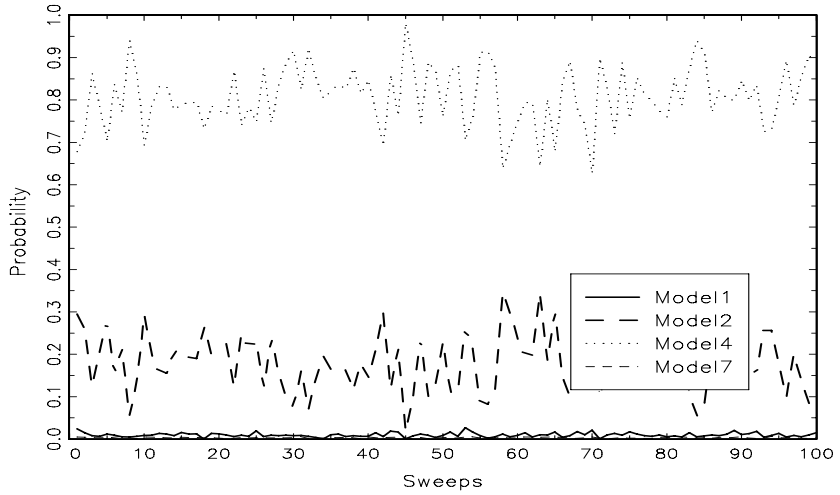

Figure 12: Convergence Behavior of FOUR of the Seven Models

5 Conclusions.

In this report the important issue of model estimation and model selection on time-varying volatility models was addressed, using a Bayesian approach and MCMC methods; this offers advantages over other competing alternatives.

The two more important approaches to time-varying volatility were considered (SV and GARCH). From the results Bera and Premaratne's GARCH were widely favored; after that the Normal Mixture is selected as the best. However, it is clear that checking work for assumptions should be done on each of the favored models.

It should be clear that all the results from models choices are conditional to the seven models initially selected; if other models are included or some are not, the results could change. Similar comments as those by George (1999) on Hoeting et al. (1999) apply here. In practice there will always be models left out, as GARCH models of high order or SV with many alternative distributions; unfortunately, the option of including many other models at the same time is highly limited by computational resources.

The robustness of the RJMCMC to different priors could be tested but much more computational time required discourages this practice.

When the main purpose of the model selection exercise is to forecast, work on Bayes model averaging will be easily implemented, once the RJMCMC has been run and results have been saved (See Hoeting, et al., 1999 or Clyde, 1999, and the specific GARCH and EGARCH case in Vrontos et al., 2000).

A fruitful avenue for future research would be the parsimonious incorporation of these features in multivariate models of stochastic volatility, see Jaquier et al. (1999).

As for the specific case of the Colombian exchange rate, the effect of exogenous shocks should be modeled with dummy variables, as Copeland and Wang (1994) did; such task could be the topic of forthcoming work to be reported elsewhere.

Finally, and no less important, more work on computational algorithm and randomness behavior faced when working with values extremely near zero, as described at the beginning of Section 4.1, is required by specialists.

It seems wise to end with a quote by G. E. P. Box, referenced by Piorier (1995, p. xi):

> *All models are wrong but some are useful.*

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. New York.: Dover Publications, Inc.

Baillie, R. T. and Bollerslev, T. (1989). The Message in Daily Exchange Rates: A Conditional-Variance Tale. *Journal of Business and Economic Statistics*, *7*(3), 297–305.

Ball, C. and Torous, W. (1983). A Simplified Jump Process for Common Stock Returns. *Journal of Financial and Quantitative Analysis*, *18*(1), 53–65.

Bera, A. and Premaratne, G. (2000). Modeling Asymmetry and Excess Kurtosis in Stock Return Data. Technical report, Department of Economics, University of Illinois U. C., Urbana-Champaign, IL.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics*, *69*, 542–547.

Brooks, S. P. and Guidici, P. (1999). Convergence Assessment for Reversible Jump MCMC Simulation. *Bayesian Statistics 6. (Bernardo, J. M., Berger, J. O., Dawid, A.P. and Smith, A.F.M., eds.). Oxford University Press*, 733–742.

Carlin, B. P. and Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society (Ser B)*, *57*(3), 473–484.

Copeland, L. S. and Wang, P. (1994). Estimation Daily Seasonality in Foreign Exchange Rate Changes. *Journal of Forecasting*, *13*(6), 519–528.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation. *Econometrica*,

50(4), 987–1007.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* New York: Chapman-Hall.

Gelman, Andrew, C., John, S., S., H., and Rubin, D. B. (1995). *Bayesian Data Analysis.* New York: Chapman-Hall/CRC.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, *82*(4), 711–732.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian Forecasting (with discussion). *Journal of the Royal Statistical Society (Ser B)*, *38*(1), 205–247.

Hoeting, J. A., Madigan, David, R., E., A., and Volinsky, C. (1999). Bayesian Modeling Averaging: A Tutorial. (with comments). *Statistical Science*, *14*(4), 382–417.

Hsieh, D. A. (1989). Modeling Heteroscedasticity in Daily Foreign-Exchange Rates. *Journal of Business and Economic Statistics*, *7*(3), 307–317.

Jaquier, E., Polson, N., and Rossi, P. (1999). Stochastic Volatility: Univariate and Multivariate Extensions. Technical report, Graduate School of Business, The University of Chicago, Chicago, IL.

Mengersen, K. L., Robert, C. P., and Chantal, G.-J. (1999). MCMC Convergence Diagnostics: A Reviewww. *Bayesian Statistics 6. (Bernardo, J. M., Berger, J. O., Dawid, A.P. and Smith, A.F.M., eds.). Oxford University Press*, 415–440.

Nagahara, Y. (1999). The PDF and CDF of Pearson Type IV Distribution and the ML Estimation of the Parameters. *Statistics and Probability Letters*, *43*, 251–264.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New approach. *Econometrica*, *59*(2), 347–370.

Nelson, D. B. and Cao, C. Q. (1992). Inequality Constraints in the Univariate GARCH Model. *Journal of Business and Economic Statistics*, *10*(2),

229–235.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: a Comparative Approach.* London, England: The MIT Press.

Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis.* New York: Chapman-Hall.

Poskitt, D. S. and Tremayne, A. R. (1983). On the Posterior Odds of Time Series Models. *Biometrika*, *70*, 157–162.

Vlaar, P. J. G. and Palm, F. C. (1993). The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps. *Journal of Business and Economic Statistics*, *11*(3), 351–360.

Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2000). Full Bayesian Inference for GARCH and EGARCH Models. *Journal of Business and Economic Statistics*, *18*(2), 187–198.

West, M. and Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models.* Second Ed. New York: Springer-Verlag.