Autoregressive moving average models with t and hyperbolic innovations

December 10, 1998

Wolfgang Polasek*
Institute of Statistics and Econometrics
4051 Basel
University of Basel
Switzerland

Jeffrey Pai**
Warren Centre for Actuarial Studies and Research
University of Manitoba
Canada

Abstract

Bayesian analysis of autoregressive moving average models is discussed with different distributional assumptions for the noise process. The proposed Gaussian white noise is a mixture distribution where the hyper-parameter follows two different hyper-distributions to incorporate, the t distribution and the hyperbolic distribution. The Metropolis-Hastings and Gibbs sampler are used for the Bayesian estimation. Model selection is a also discussed through the use of the predictive distribution, pseudo-marginal likelihoods and pseudo-Bayes factors. The models are estimated for daily and monthly exchange rates (SFr./US$).

`Keywords`: ARMA processes, leptokurtic returns, t and hyperbolic distributions, MCMC methods, CPO plots, pseudo-marginal likelihood.

`Acknowledgment`: We thank Lei Ren for the computation of the models which are provided in the BASEL package. More information can be obtained from *http://www.unibas.ch/iso/basel/basel.html*.

1 Introduction

Recent research in empirical finance has shown renewed interests in the distributional properties of returns of securities since the pricing of derivatives depends heavily on distributional assumptions. The phenomenon of leptokurtic distributions of "speculative prices" has been known since Mandelbrot (1963) and Fama (1965), and econometric approaches to model leptokurtosis have produced only moderate success. The use of ARCH (autoregressive conditional heteroskdastic) models which were introduced by Engle (1982) is popular for modeling volatility clusters, but it could not resolve the puzzle of the 'right' distributional assumptions.
Many empirical studies for financial time series have shown that the usual ('normal') Gaussian distributional assumptions are not adequate. For financial time series it is reasonable to assume that the underlying distribution has heavier tails, for example a t distribution with low degrees of freedom. The hyperbolic distribution was suggested by Barndorff-Nielsen et al. (1985) as an alternative distribution when there is considerably more mass around the origin and in the tails than in the standard normal distribution. Engle (1980) expressed this kind of behavior as leptokurtosis which is very important in modeling financial data. Eberlein and Keller (1995) applied the class of hyperbolic distributions and improved the fitness to the empirical returns. Recently, Barndorff-Nielsen (1997) has suggested to use the normal inverse Gaussian distribution for stochastic processes in finance and stochastic volatility modeling.
Both, the t distribution and the hyperbolic distribution can be considered as a mixture of normal distributions. These results are conceptually straightforward in Bayesian analysis. Consider a normal distribution with variance δ^2 for the error term ϵ, i.e.

$$\epsilon \sim N(0, \delta^2).$$

Using δ^2 as a hyper-parameter in a scale mixture of normal distributions, we can get the t distribution and the hyperbolic distribution by mixing (see Appendix C). From the Bayesian point of view, the idea of mixing is formulated in a hierarchical model and the posterior distribution of all the parameters of the model will be numerically simulated by Markov chain Monte Carlo (MCMC) methods.
In Section 2 we start with the likelihood for the normal ARMA(p,q) model

and we derive all the complete conditional densities for the parameters of the t-ARMA and the *hyp*-ARMA model. We specify hyper-distribution on the variance of the Gaussian white noise distribution to arrive at t distribution and hyperbolic distribution.
Model choice and MCMC methods like the Gibbs sampler and the Metropolis-Hastings algorithm are explained in Section 3. The predictive density is discussed in Section 4. We study the CPO (conditional predictive ordinate) plots and the pseudo Bayes factors to select the best model. Since we have assumed for some parameters a non-informative prior distribution, we use the approach of Gelfand and Dey (1994) to calculate pseudo Bayes factors. In Section 5 we illustrate our models with exchange rate data (SFr./US$ from June 1973 to May 1984). In Section 6 we finish with some conclusions.
In the appendix we discuss the generalized inverse Gaussian (GIG) distribution, the Bessel function and the hyperbolic distribution.

2 ARMA Processes

Autoregressive moving average models constitute a broad class of parsimonious time series processes which are useful in describing a wide variety of time series (Box and Jenkins, 1976). The stochastic process $\{z_t\}$, $t \in \mathcal{Z}$ is said to be an ARMA(p,q) process with mean τ if it is generated by

$$\phi(B)(z_t - \tau) = \theta(B)\epsilon_t, \tag{1}$$

where B is the backshift operator $Bz_t = z_{t-1}$ and $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are lag polynomials in B of degrees p and q, respectively, and ϵ_t are independent and identically distributed random variables with mean 0 and variance δ^2.

2.1 The likelihood for the ARMA(p,q) model

Consider mean centered observations $y_t = z_t - \tau$, then the ARMA(p,q) process in (1) can be written as

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \quad t = 1, \ldots, T.$$

Let $f_n(z_n|Z_{n-1}, \Psi_0)$ be the conditional density of the observation z_n given all the observed time series values in Z_{n-1} up to time $n-1$ and the initial values Ψ_0. We denote the likelihood for n observations $Z_n = (z_1, \cdots, z_n)$ by $f(Z_n;\Psi)$ where the parameters are

$$\Psi = (\Phi, \Theta, \tau, \delta^2) \tag{2}$$

with $\Phi = (\phi_1, \cdots, \phi_p)$ and $\Theta = (\theta_1, \cdots, \theta_q)$. Our analysis is based on the conditional likelihood function where $Z_{[0]} = (z_{1-p}, \cdots, z_0)$ and $E_{[0]} = (\epsilon_{1-q}, \cdots, \epsilon_0)$ are the starting values. The initial values and parameters are given by $\Psi_0 = (\Psi, E_{[0]}, Z_{[0]})$ and the likelihood function under normality (Gaussian distribution) can be written as

$$\begin{aligned} f(Z_n|\Psi_0) &= f_1(z_1|\Psi_0) f_2(z_2|Z_1, \Psi_0) \cdots f_n(z_n|Z_{n-1}, \Psi_0) \\ &= (2\pi\delta^2)^{-\frac{n}{2}} exp[-\frac{\sum(y_t - \tau_t)^2}{2\delta^2}] \end{aligned} \tag{3}$$

where the conditional means of the process are

$$\tau_t = \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i (y_{t-i} - \tau_{t-i}) + \tau. \tag{4}$$

For the prior distribution we assume independent blocks, i.e.

$$\pi(\Psi) = \pi(\Phi)\pi(\Theta)\pi(\delta^2)\pi(\tau), \tag{5}$$

and we assume a diffuse prior for the location parameters $\pi(\Phi) \propto 1$, $\pi(\Theta) \propto 1$, $\pi(\tau) \propto 1$, and a conjugate inverse gamma (IG) distribution for the scale parameter δ^2

$$\pi(\delta^2) = \frac{1}{\Gamma(\alpha_\delta)\beta_\delta^{\alpha_\delta}(\delta^2)^{\alpha_\delta+1}} exp(-\frac{1}{\delta^2\beta_\delta}).$$

The posterior distribution for $\mathbf{\Psi}$ is proportional to the product of the likelihood (3) and the prior (5):

$$\begin{aligned} \pi(\Psi|Z_n, E_{[0]}, Z_{[0]}) &\propto f(Z_n|\Psi, E_{[0]}, Z_{[0]})\pi(\Psi) \\ &\propto (\delta^2)^{-\frac{n}{2}} exp[-\frac{\sum(y_t - \tau_t)^2}{2\delta^2}]\pi(\Psi). \end{aligned} \tag{6}$$

To apply the Gibbs-Metropolis sampler, we derive from the posterior distribution the so-called complete conditional distribution for blocks of parameters.

The complete conditional densities (c.c.d.) for the ARMA(p,q) parameters τ, Θ, and Φ is via (6) proportional to

$$\pi(\Phi|\Psi\setminus\Phi, Z_n) \propto \exp[-\frac{\sum(y_t-\tau_t)^2}{2\delta^2}], \tag{7}$$

$$\pi(\Theta|\Psi\setminus\Theta, Z_n) \propto \exp[-\frac{\sum(y_t-\tau_t)^2}{2\delta^2}], \tag{8}$$

$$\pi(\tau|\Psi\setminus\tau, Z_n) \propto \exp[-\frac{\sum(y_t-\tau_t)^2}{2\delta^2}]. \tag{9}$$

The c.c.d. for the residual variance δ^2 is again an inverse gamma distribution

$$\pi(\delta^2|\Psi\setminus\delta^2, Z_n) = IG(\alpha_\delta+\frac{n}{2}, \frac{2\beta_\delta}{2+\beta_\delta\sum(y_t-\tau_t)^2}), \tag{10}$$

where $\Psi\setminus\Phi$ denotes the set of parameters in Ψ without Φ, and similar for the other parameters in above c.c.d.'s.
As alternative we could specify an informative (normal) prior for the ARMA(p,q) parameters in (5), i.e.

$$\pi(\Phi,\Theta,\tau) = N_p[\mu_\Phi,\Sigma_\Phi]N_q[\mu_\Theta,\Sigma_\Theta]N[\mu_\tau,\sigma_\tau^2], \tag{11}$$

where $\mu_\Phi, \Sigma_\Phi, \mu_\Theta, \Sigma_\Theta, \mu_\tau$, and σ_τ^2 are known hyper-parameters. For the informative case the c.c.d.'s in (7), (8), and (9) have to be multiplied by the normal distributions in (11):

$$\pi(\Phi|\Psi\setminus\Phi, Z_n) \propto \exp\{-\frac{\sum(y_t-\tau_t)^2}{2\delta^2}\}\exp\{-\frac{1}{2}(\Phi-\mu_\Phi)'\Sigma_\Phi(\Phi-\mu_\Phi)\},$$

$$\pi(\Theta|\Psi\setminus\Theta, Z_n) \propto \exp\{-\frac{\sum(y_t-\tau_t)^2}{2\delta^2}\}\exp\{-\frac{1}{2}(\Theta-\mu_\Theta)'\Sigma_\Theta(\Theta-\mu_\Theta)\},$$

$$\pi(\tau|\Psi\setminus\tau, Z_n) \propto \exp\{-\frac{\sum(y_t-\tau_t)^2}{2\delta^2}\}\exp\{-\frac{1}{2}(\tau-\mu_\tau)^2/\sigma_\tau^2\}.$$

2.1.1 The ARMA(p,q) model with t distributed innovations

Suppose that, given the parameter σ^2, $v\sigma^2/\delta^2$ has a χ^2 (or gamma) prior distribution with v degrees of freedom, i.e.,

$$\pi(\delta^2|\sigma^2) = \frac{(v\sigma)^{v/2}\exp[-\frac{v\sigma^2}{2\delta^2}]}{\Gamma(v/2)2^{v/2}(\delta)^{v/2+1}}.$$

By integrating δ^2 over the product of the normal distribution and the χ^2 distribution, we will get the innovations ϵ/σ as a t distribution with v degrees of freedom (see Appendix C.1). The c.c.d.'s of τ, Φ, Θ are the same as in (7) to (9) of the previous subsection and to estimate the t distribution with ν degrees of freedom we have to extend the parameter vector in (2) to $\Psi = (\Theta, \Phi, \tau, \delta^2, \sigma^2)$ and the remaining c.c.d.'s are

$$(\delta^2|\Psi \setminus \delta^2, Z_n, \Psi_0) \sim IG(\frac{n+v}{2}, \frac{2}{v\sigma \sum(y_t - \tau_t)^2}),$$

$$(\sigma^2|\Psi \setminus \sigma^2, Z_n, \Psi_0) \sim Ga(\frac{v/2 + \alpha_\sigma}{2}, \frac{2\delta^2\beta_\sigma}{v\beta_\sigma + 2\delta^2}),$$

where α_σ and β_σ are the known prior hyper-parameters for σ^2. If the degrees of freedom ν of the t distribution is not known then we can extend the parameters set to $\tilde{\Psi} = (\Psi, \nu)$ and sample ν from a discrete set of integers.

2.1.2 The ARMA(p,q) model with hyperbolic innovations

Suppose that δ^2 has a generalized inverse Gaussian (GIG) distribution (see Appendix A) of the form

$$\delta^2 \sim \frac{1}{\xi^{-2} - 1} GIG(\lambda = 1, \chi = \Psi = \xi^{-2} - 1). \tag{12}$$

Then the marginal distribution of ϵ has a hyperbolic distribution (Barndorff-Nielsen et al., 1985). Now the parameters to be estimated for the *hyp*-ARMA(p,q) model (1) are defined by $\Psi = (\Theta, \Phi, \tau, \delta^2, \sigma^2, \xi)$ and the MCMC algorithm takes the following form. The c.c.d.'s for τ, Φ, Θ are given by (7), (8) and (9), of the previous section, respectively.
The c.c.d.'s for δ^2, σ^2 and ξ can be found to be

$$\pi(\delta^2|\Psi \setminus \delta^2, Z_n, \Psi_0) \propto (\delta)^{-n} \exp[-\frac{\sum(y_t - \tau_t)^2}{2\delta^2} - \frac{\delta\sigma}{2} - \frac{(\xi^{-2} - 1)^2}{2\delta^2\sigma^2}], \tag{13}$$

and

$$\pi(\sigma^2|\Psi \setminus \sigma^2, Z_n, \Psi_0) \propto (\sigma)^{\alpha_\sigma} \exp[-\frac{\delta\sigma}{2} - \frac{(\xi^{-2} - 1)^2}{2\delta^2\sigma^2} - \frac{\sigma}{\beta_\sigma}], \tag{14}$$

$$\pi(\xi|\Psi \setminus \xi, Z_n, \Psi_0) \propto \frac{\exp[-\frac{(\xi^{-2}-1)^2}{2\delta^2\sigma^2}]}{(\xi^{-2} - 1)K_1(\xi^{-2} - 1)}. \tag{15}$$

We use for the c.c.d. of ϕ and θ a normal distribution which is obtained by an iteration proposal given by

$$\phi \sim N[\phi_{old}, \sigma^2_\phi], \quad \theta \sim N[\theta_{old}, \sigma^2_\theta].$$

The mean $\phi_{old} = 0.4$, $\theta_{old} = 0.0$, and the variance $\sigma^2_\phi = 0.01$, $\sigma^2_\theta = 0.01$ are obtained iteratively from Metropolis runs (e.g. 3 times thousand runs). We calculate the mean and variance and repeat the iterative proposal until a satisfactory acceptance/rejection ratio is obtained.

3 Sampling based methods

3.1 Gibbs sampling

Bayesian inference proceeds by obtaining marginal posterior distributions of the components of Ψ as well as summary features of these distributions. The Metropolis-Hastings algorithm is attractive in Bayesian inference since the marginal or joint posterior distributions are always available up to a multiplicative constant. The Gibbs sampler was introduced by Gelfand and Smith (1990) as a tool for carrying out Bayesian calculations and is a Markovian updating scheme that proceeds as follows. Suppose we partition Ψ into k components $(\Psi_1, \Psi_2, \ldots, \Psi_k)$. Let $\pi_i = \pi(\Psi_i | \{\Psi_j\}_{j \neq i}, Z_n)$ denote the complete conditional density (c.c.d.) of Ψ. Given an arbitrary set of starting values $(\Psi_1^{(0)}, \Psi_2^{(0)}, \ldots, \Psi_k^{(0)})$, we draw

$$\begin{aligned} \Psi_1^{(1)} &\sim \pi_1(\Psi_1 | \Psi_2^{(0)}, \ldots, \Psi_k^{(0)}, Z_n), \\ \Psi_2^{(1)} &\sim \pi_2(\Psi_2 | \Psi_1^{(1)}, \Psi_3^{(0)}, \ldots, \Psi_k^{(0)}, Z_n), \end{aligned}$$

and so on, up to

$$\Psi_k^{(1)} \sim \pi_k(\Psi_k | \Psi_1^{(1)}, \Psi_2^{(1)}, \ldots, \Psi_{k-1}^{(1)}, Z_n),$$

completing one iteration. Thus, each variable is visited in the natural order and a cycle in this scheme requires k random variate generations. After i such iterations we would arrive at $(\Psi_1^{(i)}, \Psi_2^{(i)}, \ldots, \Psi_k^{(i)})$. Under very mild regularity conditions (see Gelfand and Smith (1990) or Tierney (1994)), this Markov chain converges to a stationary distribution for large i, i.e. $(\Psi_1^{(i)}, \ldots, \Psi_k^{(i)})$ and follows a distribution which is approximately the required posterior distribution $\pi(\Psi_1, \ldots, \Psi_k | Z_n)$. Importantly, each complete conditional distribution is proportional to the joint density $f(Z_n; \Psi) \cdot \pi(\Psi)$.

3.2 Metropolis-Hastings Algorithm

The Markov chain Monte Carlo using the Metropolis-Hastings (M-H) algorithm is a general method for the simulation of stochastic processes having probability densities known up to a constant of proportionality. The Metropolis algorithm was originally developed by Metropolis et al. (1953) and has been widely used by physicists since. Hastings (1970) refined the Metropolis algorithm and introduced to statisticians in Monte Carlo estimates. The idea of the M-H algorithm is to use a general transition kernel $P(u, v)$ (based on $P(v|u)$) and the property of reversibility to get the stationary distribution associated with $P(u, v)$. In the continuous case the property $\int P(u, v)dv = 1$ holds, while in the discrete case $P(u, v)$ is called a transition matrix and the summation over the discrete states yields $\sum P(u, v) = 1$. The property of reversibility means that

$$\pi(u)P(u, v) = \pi(v)P(v, u)$$

and the equilibrium distribution associated with $P(u, v)$ is $\pi(u)$. The M-H algorithm is the Markov chain Monte Carlo method which uses this result to generate samples from the desired distribution. Let $P(u, v) = \alpha(u, v)q(u, v)$ where

$$\alpha(u, v) = min(1, \frac{\pi(v)q(v, u)}{\pi(u)q(u, v)}),$$

and $q(u, v)$ is any transition kernel on the appropriate space. The $P(u, v)$ satisfies the reversibility condition as follows. Suppose $\pi(v)q(v, u) < \pi(u)q(u, v)$,

$$\begin{aligned} \pi(u)P(u, v) &= \pi(u)\frac{\pi(v)q(v, u)}{\pi(u)q(u, v)}q(u, v) \\ &= \pi(v)q(v, u), \end{aligned}$$

and

$$\begin{aligned} \pi(v)P(v, u) &= \pi(v)\frac{\pi(u)q(u, v)}{\pi(v)q(v, u)}q(v, u) \\ &= \pi(v)q(v, u). \end{aligned}$$

Similarly, the reversibility condition is satisfied if $\pi(v)q(v,u) > \pi(u)q(u,v)$. There are several possibilities for the choices of $q(u,v)$ [e.g., Metropolis et al. 1953), Hastings (1970), Müller (1994), Tanner (1993), Chib and Greenberg (1993), Tierney (1994)]. One possibility is to let $q(u,v) = v - u$ (Metropolis et al. 1953) where $z = v - u$ is a random increment. Let z be a symmetric univariate (or multivariate) distribution (e.g. normal, t or uniform) centered at the origin. The acceptance probability $\alpha(u,v)$ is in this case

$$\alpha(u,v) = min(1, \frac{\pi(v)}{\pi(u)}).$$

$q(u,v)$ is called the probing distribution which lines up the target distribution. This choice is also called the "random walk" or "dependent" chain. Another choice of $q(u,v)$ is called the "independent" chain [Hastings (1970), Tierney (1994)]. In this choice $q(u,v) = q(v)$, ignoring the current state, behaves like an importance sampling density. The first choice need to specify only the spread while the second choice needs both the location and the spread of the probing distribution. The probability that the process does not move from u is given by

$$r(u) = 1 - \int \alpha(u,v)q(u,v)dv.$$

4 Prediction and Model Choice

4.1 The conditional predictive ordinate (CPO)

In the following section we adopt the model choice approach of Gelfand and Dey (1994) for the time series models. If $z = (z_1, \ldots, z_n)$ denotes the observed sample then $z_{(r)}$ denotes the observed data set with r-th observation deleted, i.e. $z_{(r)} = (z_1, \ldots, z_{r-1}, z_{r+1}, \ldots, z_n)$. We call $f(z_r|z_{(r)})$ the cross-validation density and if this density is evaluated at the observed data it is called the conditional predictive ordinate (CPO)

$$CPO_r = f(Z = z_r|z_{(r)}). \tag{16}$$

The product over all CPO_s is called pseudo-predictive density or pseudo-marginal likelihoods $f(z)$ for non-informative priors:

$$PsML = \prod_{r=1}^{n} CPO_r = \prod_{r=1}^{n} f(z_r|z_{(r)}). \tag{17}$$

The ratio of the pseudo-marginal likelihoods for 2 models, say I and J, is called pseudo-Bayes factor (PsBF)

$$PsBF = \frac{PsML(I)}{PsML(J)} = \frac{f(z_r|z_{(r)}, I)}{f(z_r|z_{(r)}, J)}. \tag{18}$$

In the case of time series observations we use the notation

$$Z_{[n]} = (z_{1-p}, \ldots, z_n) \quad and \quad E_{[n]} = (\epsilon_{1-q}, \ldots, \epsilon_n). \tag{19}$$

The cross-validation density is simply the one step-ahead predictive density $f(z_n|Z_{[n]})$ and the pseudo-marginal likelihood is given by

$$PsML = \prod_{r=1}^{n} f(z_r|Z_{[r-1]}). \tag{20}$$

For two models I and J the pseudo-Bayes factor is given by

$$PsBF = \frac{PsML(I)}{PsML(J)}. \tag{21}$$

In the next section we show how the pseudo-marginal likelihood PsML can be calculated from the MCMC sample $(\Psi_1, \ldots, \Psi_L)$ of size L.

4.1.1 MCMC and pseudo-Bayes factors

In a Bayesian analysis, prediction for a future observation z_F proceeds via the predictive density, $f(z_F|Z_n)$ where Z_n is the observed data of size n. In practice, the predictive density may be obtained in two ways. We can treat the future unknown observation, z_F, as unknown parameters in the Gibbs sampler. Alternatively, the predictive density can be approximated from the MCMC sample of the posterior distribution as following:

$$\begin{aligned} f(z_F|Z_n) &= \int f(z_F|Z_n, \Psi)\pi(\Psi|Z_n)d\Psi \\ &\approx \frac{1}{L}\sum_{j=1}^{L} f(z_F|Z_n, \Psi_j), \end{aligned} \tag{22}$$

where Ψ_j, $j = 1, \cdots, L$ are the output of the Gibbs sampler. To obtain a sample of predictions from the density (22), for each Ψ_j we draw $z_{F,j}$ from

$f(z_F|Z_n, \Psi_j)$ which is simply the well-known conditional likelihood.
Using notation (19), the forecasting predictive density $f(z_t|Z_{[t-1]})$ can be written as

$$f(z_t|Z_{[t-1]}) = \frac{\int f(Z_{[t]}|\Psi)\pi(\Psi)d\Psi}{\int f(Z_{[t-1]}|\Psi)\pi(\Psi)d\Psi}. \tag{23}$$

After some transformations the conditional predictive density in (23) can be obtained from the MCMC output in the following way:

$$\begin{aligned} f(z_t|Z_{[t-1]}) &= \frac{\int \frac{1}{\Pi_{k=t+1}^n f(z_k|Z_{[k-1]},\Psi)}\pi(\Psi|Z_n)d\Psi}{\int \frac{1}{\Pi_{k=t}^n f(z_k|Z_{[k-1]},\Psi)}\pi(\Psi|Z_n)d\Psi} \\ &\simeq \frac{\sum_{j=1}^L \frac{1}{\Pi_{k=t+1}^n f(z_k|Z_{[k-1]},\Psi_j)}}{\sum_{j=1}^L \frac{1}{\Pi_{k=t}^n f(z_k|Z_{[k-1]},\Psi_j)}}. \end{aligned} \tag{24}$$

This follows from the decomposition

$$f(Z_n|\Psi) = f(Z_{[t-1]}|\Psi)\prod_{k=t}^{n} f(z_k|Z_{[k-1]}, \Psi).$$

4.2 The CPO Plot

For the model choice we use the so-called conditional predictive ordinate (CPO) plot. Let d_t be the density in (24) evaluated at the observed data. In comparing two models, the one with the larger ordinates d_t is the one more likely to have produced the z_t we observed. Hence, for say two models J and I, the ratio $d_t(J)/d_t(I)$ indicates relative support of the observed z_t for model J or model I. We can plot these ratios or log ratios versus t to obtain a so-called CPO ratio plot. In fact, if we aggregate over t and compute $\Pi_{t=1}^n d_t(J)/d_t(I)$ we obtain the pseudo-Bayes factor (PsBF) which, in the case of improper priors, has been suggested by Gelfand and Dey (1994) as an alternative model choice criterion to the Bayes factor (BF).
For model J the ordinate of the posterior density is calculated in following way

$$\begin{aligned} d_t(J) &= f(z_t|Z_{[t]}, J) \\ &\simeq \frac{\sum_{l=1}^L \Pi_{j=t+1}^n [f(z_j|Z_{[j-1]}, \Psi_l, J)]^{-1}}{\sum_{l=1}^L \Pi_{j=t}^n [f(z_j|Z_{[j-1]}, \Psi_l, J)]^{-1}}. \end{aligned} \tag{25}$$

The $l-th$ sample of the parameter vector is $\Psi_l = (\psi_{1l}, \ldots, \psi_{nl})$. For model I the ordinate of the posterior density is

$$\begin{aligned} d_t(I) &= f(z_t|Z_{[t]}, I) \\ &\simeq \frac{\sum_{l=1}^{L} \Pi_{j=t+1}^{n} [f(z_j|Z_{[j-1]}, \Psi_l)]^{-1}}{\sum_{l=1}^{L} \Pi_{j=t}^{n} [f(z_j|Z_{[j-1]}, \Psi_l)]^{-1}}. \end{aligned}$$

Therefore, the pseudo-Bayes factor for model J vs. model I is

$$PsBF = \Pi_{t=1}^{n} \frac{d_t(J)}{d_t(I)}, \tag{26}$$

where n is the sample size.
The pseudo-marginal likelihood of the ARMA(p,q) model and an MCMC sample of the posterior distribution is given by

$$PsML(p,q) = \prod_{t=1}^{T} d_t(p,q), \tag{27}$$

where d_t are the CPO's from the model with p and q parameters. We choose the orders p and q for which the PsML is the largest.
The log pseudo-marginal likelihood is given by

$$logPsML = \sum_{t=1}^{T} log(d_t).$$

5 Examples

We consider the AR(p) model with

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \quad t = 1, ..., n, \tag{28}$$

$$\epsilon_t \sim N(0, \delta^2),$$

where y_t is the observation at time t.
We present two examples to illustrate our models. The first example analyses a data set of monthly exchange rates for the Swiss Frank/Index for the period June 1973 to May 1984 (shown in Fig. 3) using AR(1) models. The time series are analyzed in differenced form, giving 150 observations of monthly

exchange rates. The second example estimates AR(1) models for daily exchange rates of the Swiss Frank/US \$ in 1987 (shown in Fig. 11). The time series is given in differenced form, giving 246 observations of daily exchange rates.

We ran the results for the normal distribution shown in Fig. 4 and Fig. 12, the t distribution with priors $v = 4$, $\alpha_\sigma = 1$, and $\beta_\sigma = 1000$ shown in Fig. 5 and Fig. 13, and the hyperbolic distribution with priors $\alpha_\sigma = 1$, and $\beta_\sigma = 1000$ shown in Fig. 6 and Fig. 14.

The means and standard errors of the parameters are shown in Table 2 for the SFr./Index, and in Table 3 for the SFr./US\$. Particularly the histograms of the SFr./Index and SFr./US\$ are compared with normal, hyperbolic distribution in Fig. 17 and 18 respectively.

For the model selection, we have calculated the pseudo-Bayes factors based on the conditional predictive ordinates (CPO) for all time points ($t = 1, ..., T$) AR(1) models are shown in Table 1.1 for the SFr./Index and SFr./US\$ and in Table 1.2 we compare the pseudo-Bayes factors in (26) for up to order $p = 4$ and the t and hyp-ARMA(1,1) model. For all these models the hyperbolic AR model turns out to give the best fit. In Table 1.3 we have calculated the pseudo-marginal likelihood (27) for selection of the order of the order of ARMA(p,q) process. The ARMA(1,1) model is selected (from the range $1 \leq p \leq 4$ and $1 \leq q \leq 4$) better than the t_4-ARMA(1,1) model.

The CPO plots for SFr./Index normal, t_4 and hyperbolic distribution are shown in Fig. 7, 8, 9, and 10. From these figures it is seen that the medians of the CPO plots are very similar for t innovations and hyperbolic models. The medians of the CPO plots for t innovations and hyperbolic distribution are larger than those for normal distribution.

The CPO plots of the SFr./US\$ are shown in Fig. 15. Again we see from these figures that the medians of the CPO plots are very similar. The medians of the CPO plots for the hyperbolic are larger than those for normal distribution and t innovations. The median of the CPO plots for hyperbolic AR(1) model is larger than those for the AR(2), AR(3) and AR(4) model.

6 Conclusions

IThis paper has developed a MCMC technique to estimate ARMA(p,q) models with t_v and hyperbolic distributed innovations. We have derived the

complete conditional distributions for simulation of the posterior distribution by introducing an mixture model for the scale parameters of the normal distribution. The convergence of the Gibbs sampler for monthly and daily exchange rate data (SFr./US$ from June 1973 until May 1984) was achieved very quickly.
For the model choice we have used conditional predictive ordinates (CPO's) and the pseudo-marginal likelihood (PsML) criterion which was suggested by Gelfand and Dey (1994). Additionally, pseudo-Bayes factors can be computed to choose between competing models. From the analysis we see that the *hyp*-ARMA(1,1) model dominates the normal or t distributed ARMA models. This shows that the exchange rate data favor a leptokurtic distributed time series model, like the *hyp*-ARMA(p,q) model. The leptokurtic effect can be seen also for the predictive distribution.
Using the MCMC approach we can obtain the small sample distribution for all the parameters and using the pseudo-marginal likelihood criterion we can do test and model choices. Further studies will show how the *hyp*-ARMA model can be extended to GARCH models and multivariate models (VAR).

A Appendix: The *GIG* distribution

A distribution which is not widely known in econometrics is the generalized inverse Gaussian distribution $GIG(x)$ which has the density

$$f(x|\lambda,\psi,\chi) = \frac{(\psi/\chi)^{\lambda/2}x^{\lambda-1}}{2K_\lambda(\sqrt{(\chi\psi)})}\exp[-\frac{1}{2}(\chi x^{-1}+\psi x)] \quad for \quad x \geq 0 \qquad (29)$$

where $K_\lambda(\cdot)$ (see Appendix B) is the modified Bessel function of the third kind with index λ. The domain of variation of the three parameter GIG distribution is

$$\begin{aligned} \chi > 0, \quad \psi \geq 0 \quad if \quad \lambda < 0; \\ \chi > 0, \quad \psi > 0 \quad if \quad \lambda = 0; \\ \chi \geq 0, \quad \psi > 0 \quad if \quad \lambda > 0. \end{aligned}$$

Lemma A1: Properties of the GIG distribution.
Let $GIG(\cdot,\cdot,\cdot)$ and $Ga(\cdot)$ denote GIG and gamma distribution, respectively,

and let all random variables be independent. Then we have the following distributional equivalences (see Devroye 1986, p. 479):

(a)

$$GIG(\lambda,\psi,\chi) = \frac{1}{c}GIG(\lambda,\frac{\psi}{c},\chi c) \quad for \quad all \quad c>0.$$

(b)

$$GIG(\lambda,\psi,\chi) = \sqrt{\frac{\chi}{\psi}}GIG(\lambda,\sqrt{\psi\chi},\sqrt{\psi\chi}),$$

(c)

$$GIG(\lambda,\psi,\psi) = GIG(-\lambda,\psi,\psi) + \frac{2}{\psi}Ga(\lambda),$$

(d)

$$GIG(\lambda,\psi,\chi) = \frac{1}{GIG(-\lambda,\chi,\psi)}.$$

Furthermore the mode is given by

$$Mode = \frac{1}{\sqrt{(\frac{\lambda-1}{\psi})^2+1}} - \frac{\lambda-1}{\psi}.$$

Special cases of the GIG distribution are the gamma distribution ($\chi = 0$), the reciprocal gamma distribution ($\psi = 0$), the inverse Gaussian distribution ($\lambda = -\frac{1}{2}$), and the inverse Gaussian or random walk distribution ($\lambda = \frac{1}{2}$).
The general hyperbolic density function is a four-parametric density given by

$$f(x|\phi,\gamma,\delta,\mu) = \frac{\sqrt{\phi\gamma}}{\delta(\phi+\gamma)K_1(\delta\sqrt{(\phi\gamma)}}exp\{-\frac{1}{2}[(\phi+\gamma)\sqrt{\delta^2+(x-\mu)^2}+(\phi-\gamma)(x-\mu)]\}, \tag{30}$$

where $K_1(\cdot)$ is the modified Bessel function of third kind (see Appendix B). If X is general hyperbolic distributed with the parameters μ, ϕ, δ, and γ, i.e. $X \sim Hyp(\mu,\phi,\delta,\gamma)$ then the first two moments are

$$E(X) = \mu + \frac{\delta(\phi-\gamma)}{2\sqrt{\phi\gamma}},$$

$$Var(X) = \delta.$$

Consider the symmetric centered case of the 4-parametric hyperbolic distribution $Hyp(\mu, \phi, \delta, \gamma)$ that is, $\mu = 0$, $\phi = \gamma$, and

$$\xi = (1 + \delta\sqrt{\phi\gamma})^{-1/2}.$$

Then the hyperbolic density (8) can be written as a 2-parametric hyperbolic density

$$f(x|\xi, \delta^2) = \frac{1}{2\delta K_1(\xi^{-2} - 1)} exp\{-\frac{(\xi^{-2} - 1)\sqrt{\delta^2 + x^2}}{\delta}\}, \tag{31}$$

where $| \, \xi \, | < 1$ and $0 < \delta^2 < \infty$.
In Fig. 1, the standard normal density function is compared with the hyperbolic density function (30) for the parameter values with $\xi = 0.5$, $\xi = 0.6$ and $\xi = 0.7$ (for $\delta = 1$).
In Fig. 2, the hyperbolic density function in (31) for $\xi = 0.5$ and $\delta = 1.0$ ($\mu = 0.0$, $\delta = 1.0$, $\phi = 3.0$ and $\gamma = 3.0$) is compared with the hyperbolic density function (30) with the different parameter values μ, δ, ϕ and γ.

B Appendix: Bessel function of the third kind

The Bessel function of the third kind with index 1 is given by

$$K_1(x) = [ln(\frac{x}{2}) + \gamma]I_1(x) + \frac{1}{x} - \frac{1}{2}\sum_{k=0}^{\infty}\frac{(x/2)^{2k+1}}{k!(k+1)!}[\phi(k) + \phi(k+1)]$$

with

$$I_1(x) = \frac{x}{2} + \frac{x^3}{2^2 \cdot 4} + \frac{x^5}{2^2 \cdot 4^2 \cdot 6} + \cdots,$$

$$\phi(k) = 1 + \frac{1}{2} + \cdots + \frac{1}{k}, \quad \phi(0) = 0,$$

and $\gamma = 0.5772156$ is the Euler constant.
The general Bessel function with index λ is defined by

$$K_\lambda(x) = \frac{1}{2}\int_{-\infty}^{\infty} cosh(\lambda u)e^{-x cosh(u)} du.$$

When ($\lambda = \frac{1}{2}$) or ($\lambda = -\frac{1}{2}$), we find

$$K_{\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}} \sin x, \quad and \quad K_{-\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}} \cos x,$$

$$K_{-\frac{1}{2}}(\sqrt{\psi\gamma}) = \sqrt{\frac{2}{\pi(\psi\gamma)^{\frac{1}{2}}}} \cos \sqrt{\psi\gamma}.$$

C Appendix: Mixture distributions

C.1 t innovations as a normal-gamma mixture

Lemma C1: The t distribution is a scale mixture of a normal and a gamma (χ^2)-distribution.
Consider a centered normal distribution

$$\epsilon|\delta^2 \sim N(0, \delta^2),$$

and the precision δ^{-2} is χ^2-distributed

$$\delta^{-2}|\sigma^2 \sim \frac{1}{v\sigma^2}\chi^2_v,$$

then the standardized variable ϵ/σ is t distributed with v degrees of freedom:

$$\frac{\epsilon}{\sigma}|\delta \sim t_v.$$

Proof: Since the χ^2 distribution is a special case of the gamma distribution we have

$$\delta_0 = \frac{v\sigma^2}{\delta^2} \sim Ga(v/2, 1/2),$$

$$\pi(\delta_0) = \frac{(\delta_0)^{v/2-1} exp\{-\frac{\delta_0}{2}\}}{\Gamma(v/2)2^{v/2}},$$

$$\pi(\delta^2|\sigma^2) = \frac{(v\sigma^2)^{v/2} exp\{-\frac{v\sigma^2}{2\delta^2}\}}{\Gamma(v/2)2^{v/2}(\delta^2)^{v/2+1}},$$

and integrating out δ^2 yields

$$\begin{aligned}
\pi(\epsilon|\sigma^2) &= \int_{\delta^2} \pi(\epsilon|\delta^2)\pi(\delta^2|\sigma^2)d\delta^2 \\
&= \int_{\delta^2} \frac{1}{\sqrt{2\pi\delta^2}} exp\{-\frac{\epsilon^2}{2\delta^2}\} \frac{(v\sigma^2)^{v/2} exp\{-\frac{v\sigma^2}{2\delta^2}\}}{\Gamma(v/2)2^{v/2}(\delta^2)^{v/2+1}} d\delta^2 \\
&= \frac{\Gamma(\frac{v+1}{2})(1+\frac{\epsilon^2}{v\sigma^2})^{-\frac{v+1}{2}}}{\sigma^2\Gamma(v/2)\sqrt{\pi v}}.
\end{aligned}$$

Consider the density function of the standardized variable $\epsilon_0 = \epsilon/\sigma$,

$$\begin{aligned}
\pi(\frac{\epsilon_0}{\sigma}) &= \sigma\frac{\Gamma(\frac{v+1}{2})(1+\frac{\epsilon^2}{v\sigma^2})^{-\frac{v+1}{2}}}{\sigma\Gamma(v/2)\sqrt{\pi v}} \\
&= \frac{\Gamma(\frac{v+1}{2})(1+\frac{\epsilon_0^2}{v\sigma^2})^{-\frac{v+1}{2}}}{\Gamma(v/2)\sqrt{\pi v}},
\end{aligned}$$

then this is clearly a t distribution with v degrees of freedom

$$\frac{\epsilon}{\sigma}|\delta^2 \sim t_v.$$

C.2 Hyperbolic distribution

Lemma C2: The hyperbolic distribution can be represented as a mixture of the normal and the general inverse Gaussian (GIG) distribution.
We consider the symmetric hyperbolic distribution (31), so we have to deal only with the location parameter ξ and the scale parameter δ^2.
Consider the central normal distribution

$$\epsilon|\delta^2 \sim N(0,\delta^2),$$

and assume a GIG distribution for δ^2

$$(\delta^2|\lambda = 1, \sigma^2, \xi) \sim \frac{\sigma}{\xi^{-2}-1} GIG(\lambda = 1, 1, \xi^{-2}-1)$$

then the mixture distribution is hyperbolic

$$\epsilon|\sigma, \xi \sim Hyp(\sigma, \xi).$$

Proof: Because the normal density is

$$f(\epsilon|\delta^2) = \frac{1}{\delta\sqrt{2\pi}} exp\{-\frac{\epsilon^2}{2\delta^2}\}$$

and the hyper-parameter δ^2 has the density

$$\pi(\delta^2|\sigma^2, \xi) = \frac{exp\{-\frac{1}{2}[\frac{\delta(\xi^{-2}-1)^2}{\sigma} + \frac{\sigma}{\delta}]\}}{2\sigma K_1(\xi^{-2}-1)} \cdot \frac{\xi^{-2}-1}{\sigma^2}$$

The 2-parameter hyperbolic distribution is obtained by integrating out δ^2 from the joint density

$$\int_{\delta^2} f(\epsilon|\delta^2)\pi(\delta^2|\sigma^2, \xi)d\delta^2 = \frac{exp\{-\frac{\xi^{-2}-1}{\sigma}\sqrt{\sigma^2+\epsilon^2}\}}{2\sigma K_1(\xi^{-2}-1)}.$$

References

[1] Bagpunar, J. (1988). Principles of random variate generation, Oxford Science Publications.

[2] Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size, Proceedings of the Royal London A, 353, 401-419.

[3] Barndorff-Nielsen, O. E. (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling, Scand. J. of Statistics 24, 1-14.

[4] Barndorff-Nielsen, O. E., Jensen, J. I. and Sorensen, M. (1989). Wind Shear and Hyperbolic Distributions, Boundary-Layer Meteorology 49. 417-434.

[5] Berger, J. O., Pericchi, C. R. (1996). The intrinsic Bayes factor for model selection and prediction, JASA 91, 109-122.

[6] Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control, Holden Day, San Fransisco.

[7] Chib, S. and Greenberg, E. (1994). Bayesian inference in regression models with ARMA(p,q) errors, Journal of Econometrics, 64, 183-206.

[8] Devroye, L. (1986). Non-uniform Random Variate Generations, Springer-Verlag, New York.

[9] Eberlein, E. and Keller, U. (1995). Hyperbolic Distributions in Finance, Bernoulli 1, 281-299.

[10] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, Econometrics 50, 987-1008.

[11] Fama, E. (1965). The behavior of stock market prices, J. of Business 38, 34-105.

[12] Gelfand, A. E. and Dey, D.K. (1994). Bayesian model choice asymptotics and exact calculations, JRSSB, 56, 501-514.

[13] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities, Journal of the American Statistical Association, 85:398-409.

[14] Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences, (with discussion), Statistical Science, 7(4), 457-511

[15] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:721-741.

[16] Geisser, S. (1980). Discussion of the paper by G.E.P. Box, JRSSA, 143, 416-417.

[17] Geyer, C. (1991). Markov chain Monte Carlo maximum Likelihood, Computing Science and Statistics, 156-163.

[18] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, biometrica, 57, 97-109.

[19] Mandelbrot, B. B. (1963). The variation of certain speculative prices, J. of Business 36, 394-410.

[20] Metropolis, N. Rosenbluth, A.W. Rosenbluth, M.N. Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machines,

[21] Müller, P. (1994). Metropolis posterior integration schemes, Stat. Computing., forthcoming.

[22] Norman, L. J. and Samuel, K. (1970). Sampling Techniques, Wiley series in probability and mathematical statistics, 149.

[23] Pai, J. and Ravishanker, (1996). ,Bayesian Modelling of ARFIMA Processes by MCMC Methods, J of Forecasting 15, 63-82.

[24] Tanner, M. and Wang, W. (1987). The calculation of posterior distributions by data augmentation, (with discussion), Journal of the American Statistical Association, 82:528-550.

[25] Tierney (1994). Markov chains for exploring posterior distributions, (with discussion), Ann. Statistics 22, 1701-1786.

	SFr./Index	SFr./US$
Hyp/Nor	717.75232	1.07906
$Hyp/t(4)$	69.84747	1.47597
$t(4)/Nor$	10.27599	0.731085

Table 1.1: Pseudo-Bayes factors (PsBF) of AR(1) models for SFr./Index and SFr./US$

	AR(2)	AR(3)	AR(4)	ARMA(1,1)
Hyp/Nor	7767.93	3630.67	8.7701	272.2072
$Hyp/t(4)$	1123.82	8172.67	7.7890	237.4144
$t(4)/Nor$	6.91207	0.4442	1.1259	1.1465

Table 1.2: Pseudo-Bayes factors (PsBF) of AR(2), AR(3), AR(4) and ARMA(1,1) models for SFr./US$

p	q	hyp-ARMA	t_1-ARMA	t_2-ARMA	t_3-ARMA	t_4-ARMA
1	0	-299.5067	-307.1074	-299.0349	-299.2931	-308.0917
1	1	-296.2734**	-302.2248*	-296.9265*	-297.3909*	-297.1681*
2	1	-296.3969	-306.1095	-298.4662	-298.5534	-297.7808
3	1	-299.3074	-307.5711	-298.6821	-298.6272	-300.7323
4	1	-300.6974	-305.7271	-299.5723	-299.9052	-301.6563
1	2	-298.1801	-306.5814	-300.3152	-300.6721	-297.3510
1	3	-299.4493	-307.4177	-301.2512	-301.7273	-298.9292
1	4	-299.1183	-306.6823	-299.6388	-300.5126	-298.0368
2	2	-303.2661	-305.5267	-300.8965	-301.3012	-319.9409
3	3	-301.3219	-304.6712	-301.6731	-302.6388	-317.6383
4	4	-307.2369	-306.0763	-302.7483	-303.7481	-326.7722
p	q	t_5-ARMA	t_6-ARMA	t_7-ARMA	t_8-ARMA	t_9-ARMA
1	1	-296.9873	-296.9852	-296.7322	-296.7311	-297.1494

Table 1.3: The log marginal likelihood (PsML) for SFr./US$

	Nor		$t(4)$		Hyp		Model
	Mean	St. error	Mean	St. error	Mean	St. error	
ϕ_1	0.40683	8.81e-04	0.40953	3.65e-04	0.40428	4.94e-04	$AR(1)$
δ	3.18054	2.55e-03	3.08906	2.41e-03	3.18421	2.56e-03	
σ			1.12084	3.81e-03	6.28449	3.72e-03	
ξ					0.57964	6.86e-04	
ϕ_1	0.48602	1.00e-03	0.48398	4.15e-04	0.48555	5.32e-04	$AR(2)$
ϕ_2	-0.18871	1.01e-03	-0.18908	4.05e-04	-0.19215	5.28e-04	
δ	3.16668	2.60e-03	3.00549	2.32e-03	3.07802	2.55e-03	
σ			1.12561	3.88e-03	6.31098	3.96e-03	
ξ					0.58340	7.18e-04	
ϕ_1	0.48580	1.02e-03	0.48497	4.17e-04	0.48661	5.38e-04	$AR(3)$
ϕ_2	-0.19785	1.18e-03	-0.19827	4.28e-04	-0.19884	5.88e-04	
ϕ_3	0.01776	1.04e-03	-0.01679	3.94e-03	-0.01866	5.55e-04	
δ	3.26154	2.99e-03	3.00013	2.21e-03	3.08530	2.40e-03	
σ			1.08220	3.73e-03	6.11120	3.54e-03	
ξ					0.58173	6.95e-04	
ϕ_1	0.48046	1.07e-03	0.48809	3.93e-04	0.49098	5.46e-04	$AR(4)$
ϕ_2	-0.17889	1.14e-03	-0.19393	4.34e-04	-0.19412	5.97e-04	
ϕ_3	-0.01167	1.21e-03	0.00108	4.31e-04	-0.00103	5.89e-04	
ϕ_4	0.03616	1.10e-03	0.03483	3.80e-04	0.03726	5.42e-04	
δ	3.35717	2.91e-03	3.01753	2.25e-03	3.08315	2.43e-03	
σ			1.08712	3.79e-03	6.49160	3.78e-03	
ξ					0.57391	7.02e-04	

Table 2: The means and standard errors of the parameters ϕ, δ, σ and ξ for SFr./Index

	Nor		$t(4)$		Hyp		Model
	Mean	St. error	Mean	St. error	Mean	St. error	
ϕ_1	0.00807	2.95e-06	0.00809	2.00e-06	0.01001	2.68e-04	$AR(1)$
δ	0.00013	4.80e-08	0.00012	4.52e-08	0.00012	4.42e-08	
σ			0.00006	2.60e-08	0.00028	1.05e-06	
ξ					0.57587	4.41e-04	
ϕ_1	0.00862	3.03e-06	0.00864	1.90e-06	0.00942	2.65e-04	$AR(2)$
ϕ_2	-0.06336	2.97e-06	-0.06335	2.25e-06	-0.06319	2.62e-04	
δ	0.00013	4.58e-08	0.00012	4.17e-08	0.00012	4.48e-08	
σ			0.00006	2.48e-07	0.00024	9.07e-07	
ξ					0.58712	4.29e-04	
ϕ_1	0.00296	2.86e-06	0.00292	1.97e-06	0.00087	2.68e-04	$AR(3)$
ϕ_2	-0.06266	3.05e-06	-0.06270	2.22e-06	-0.06292	2.64e-04	
ϕ_3	0.09176	3.07e-06	-0.09176	2.00e-06	-0.09280	2.63e-04	
δ	0.00013	4.60e-08	0.00012	4.34e-08	0.00012	4.70e-08	
σ			0.00006	2.56e-07	0.00022	8.33e-07	
ξ					0.59721	4.37e-04	
ϕ_1	0.00404	2.98e-06	0.00404	1.95e-06	0.00727	2.65e-04	$AR(4)$
ϕ_2	-0.17889	3.13e-06	-0.19393	2.27e-06	-0.05793	2.62e-04	
ϕ_3	-0.09177	3.14e-06	-0.09177	2.28e-06	-0.09486	2.66e-04	
ϕ_4	0.01238	3.10e-06	0.01237	2.05e-06	0.01203	2.56e-04	
δ	0.00013	4.65e-08	0.00012	4.33e-08	0.00012	4.61e-08	
σ			0.00006	2.52e-07	0.00025	8.63e-07	
ξ					0.57911	4.32e-04	

Table 3: The means standard errors of the parameters ϕ, δ, σ and ξ for SFr./US$

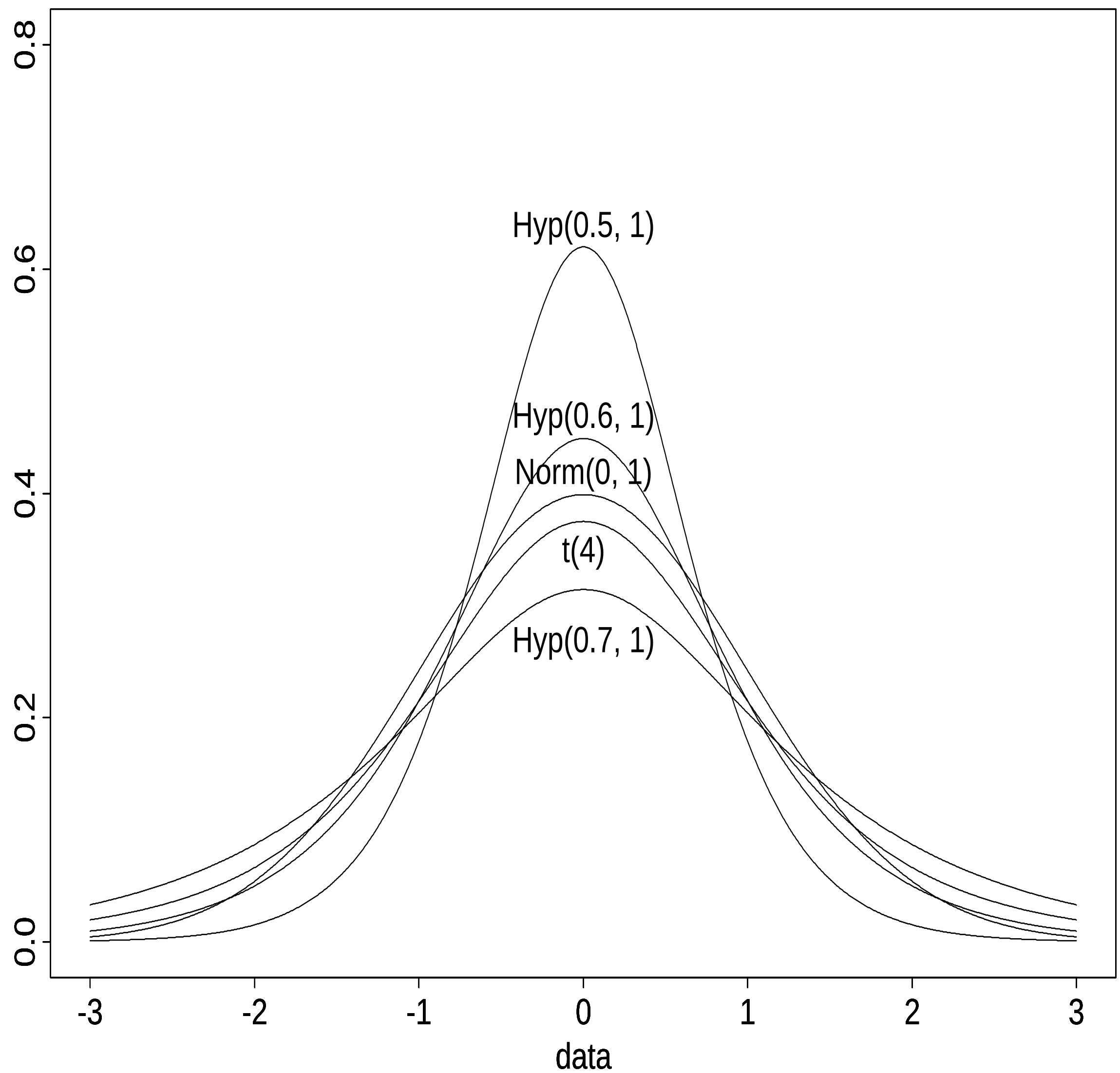


Figure 1: The standard normal density $N(0,1)$ compared with the hyperbolic density $Hyp(\xi, \delta)$ and $t(4)$.

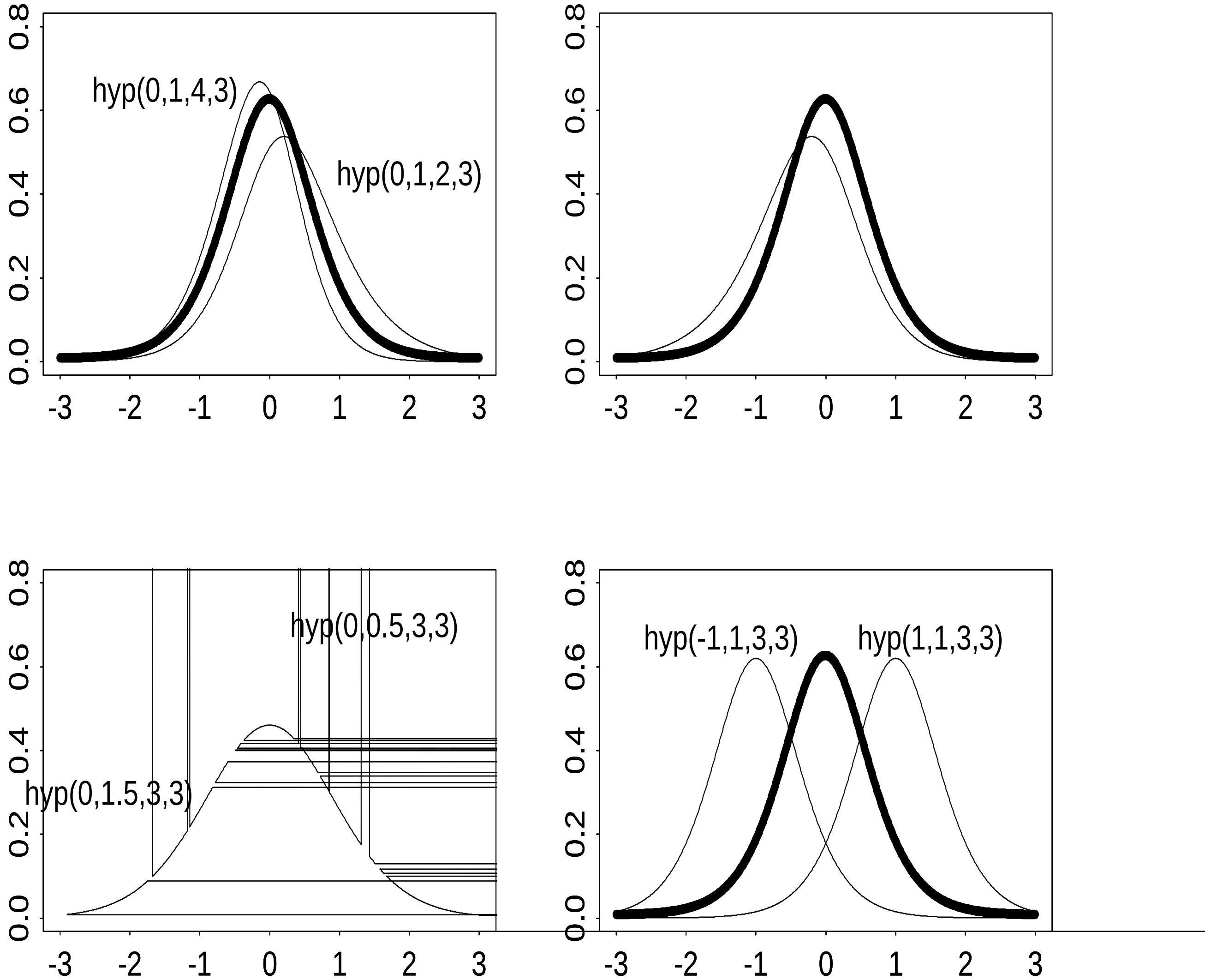


Figure 2: 4-parameters hyperbolic density function Hyp(μ,δ,ϕ,γ) compared with 2-parameters hyperbolic density (**solid line**) Hyp($\xi = 0.5$,$\delta = 1$)=Hyp($\mu = 0$,$\delta = 1$,$\phi = 3$, $\gamma = 3$).

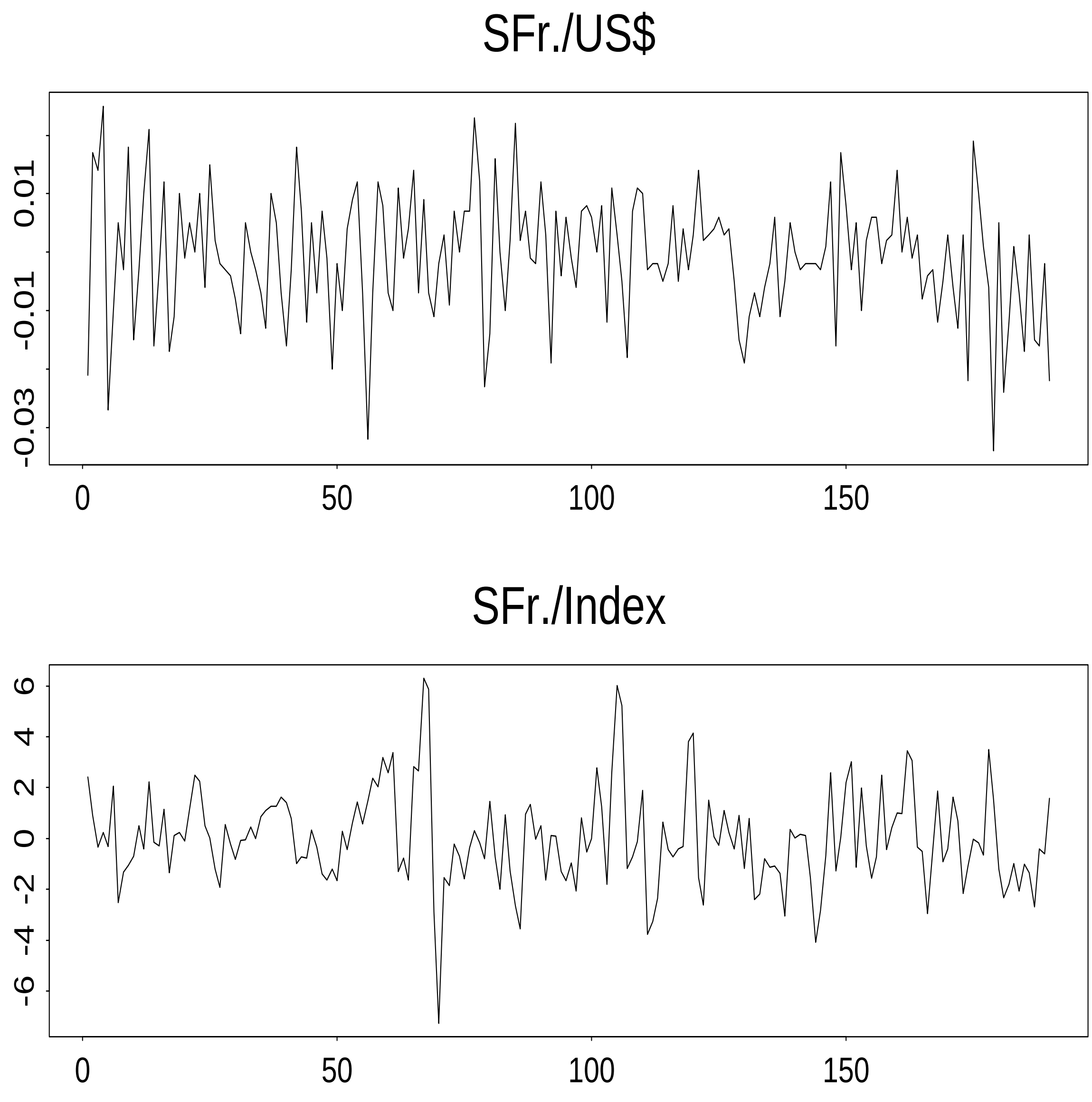


Figure 3: Monthly exchange rate for Swiss Franc/Index from Jan. 1973 to May 1984

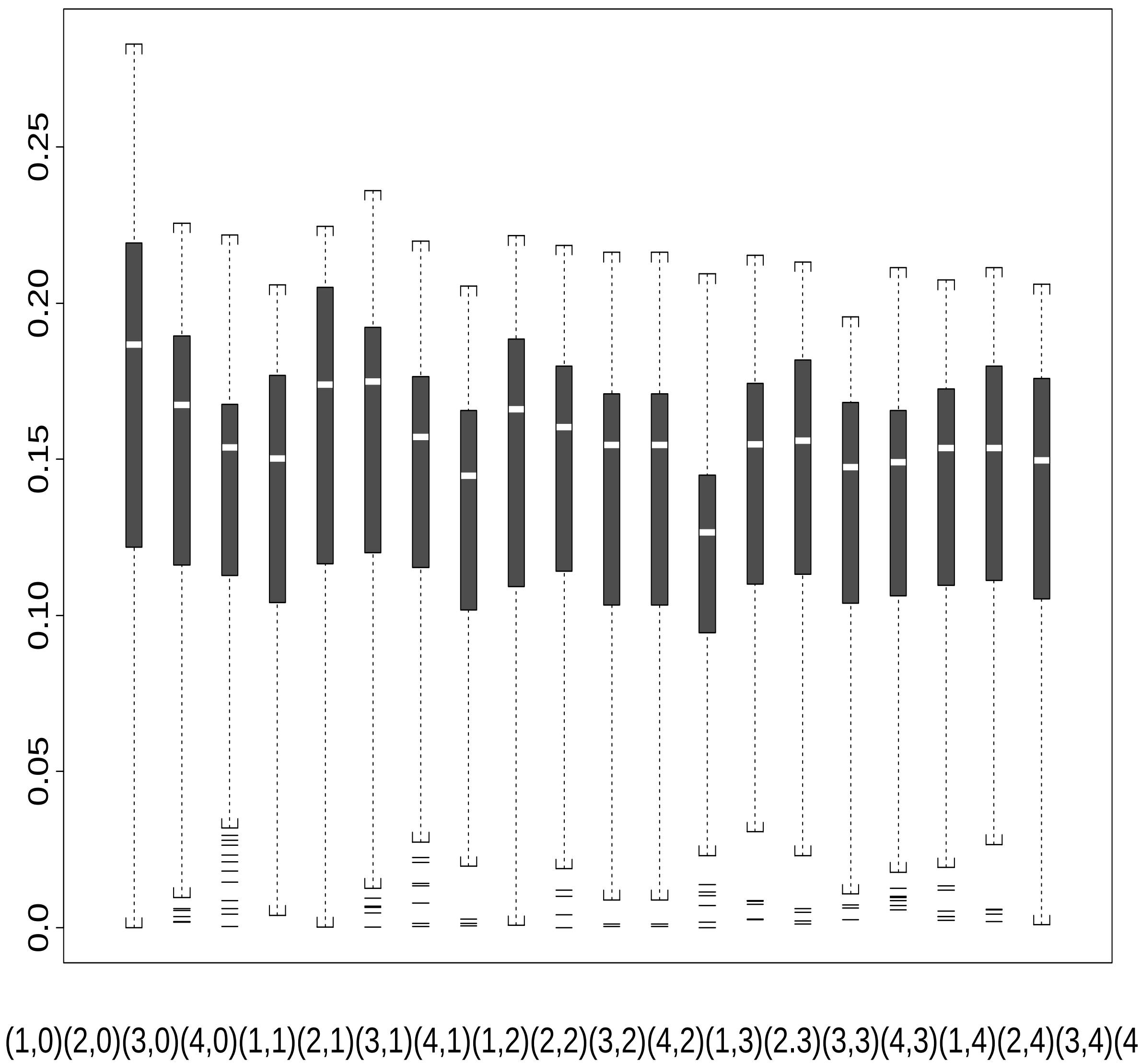


Figure 7: CPO plots of monthly rates(SFr./Index) for normal AR(p,q) models

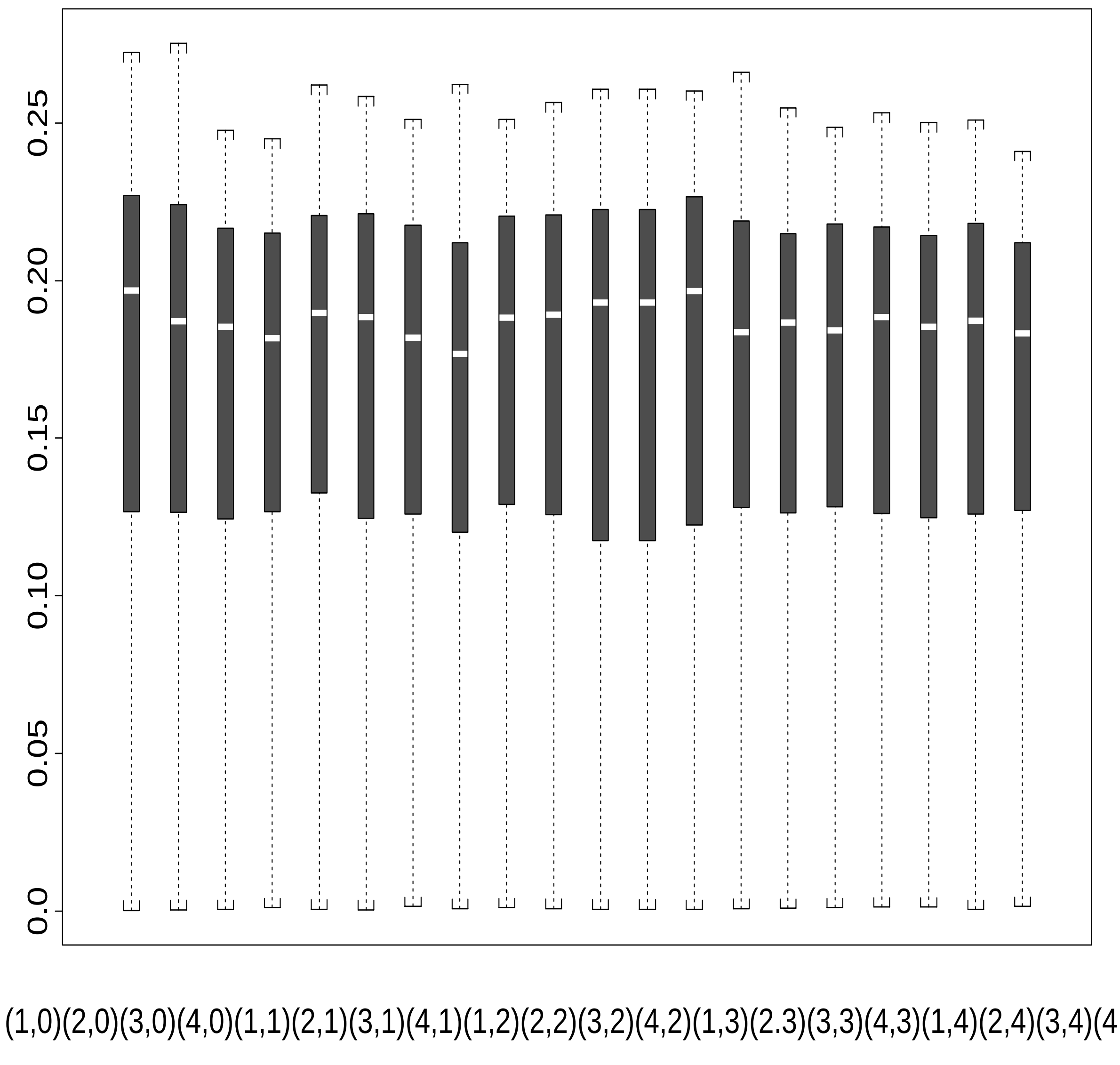


Figure 8: CPO plots of monthly rates(SFr./Index) for mixture normal-t innovations AR(p,q) models

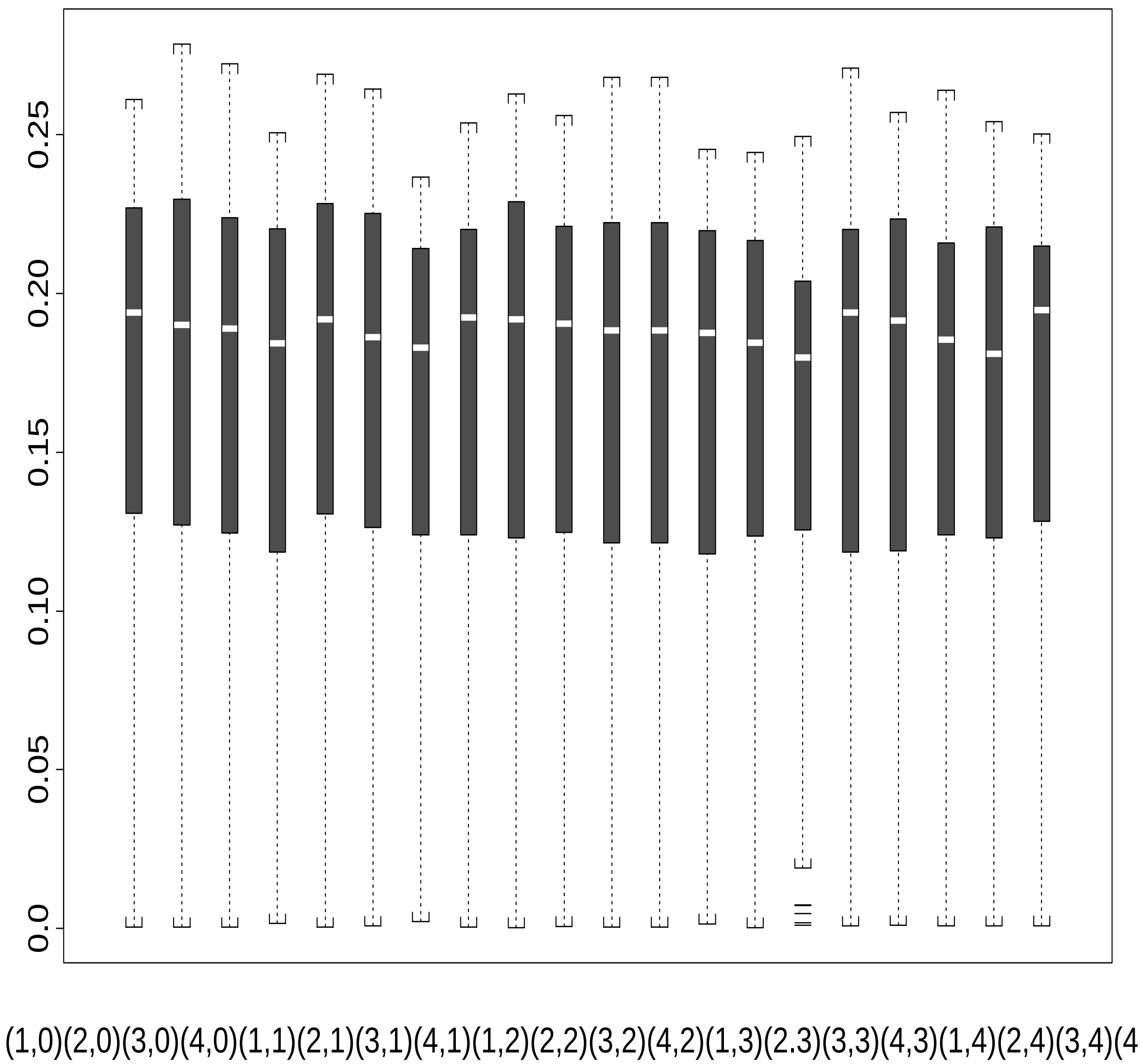


Figure 9: CPO plots of monthly rates(SFr./Index) for hyperbolic AR(p) models

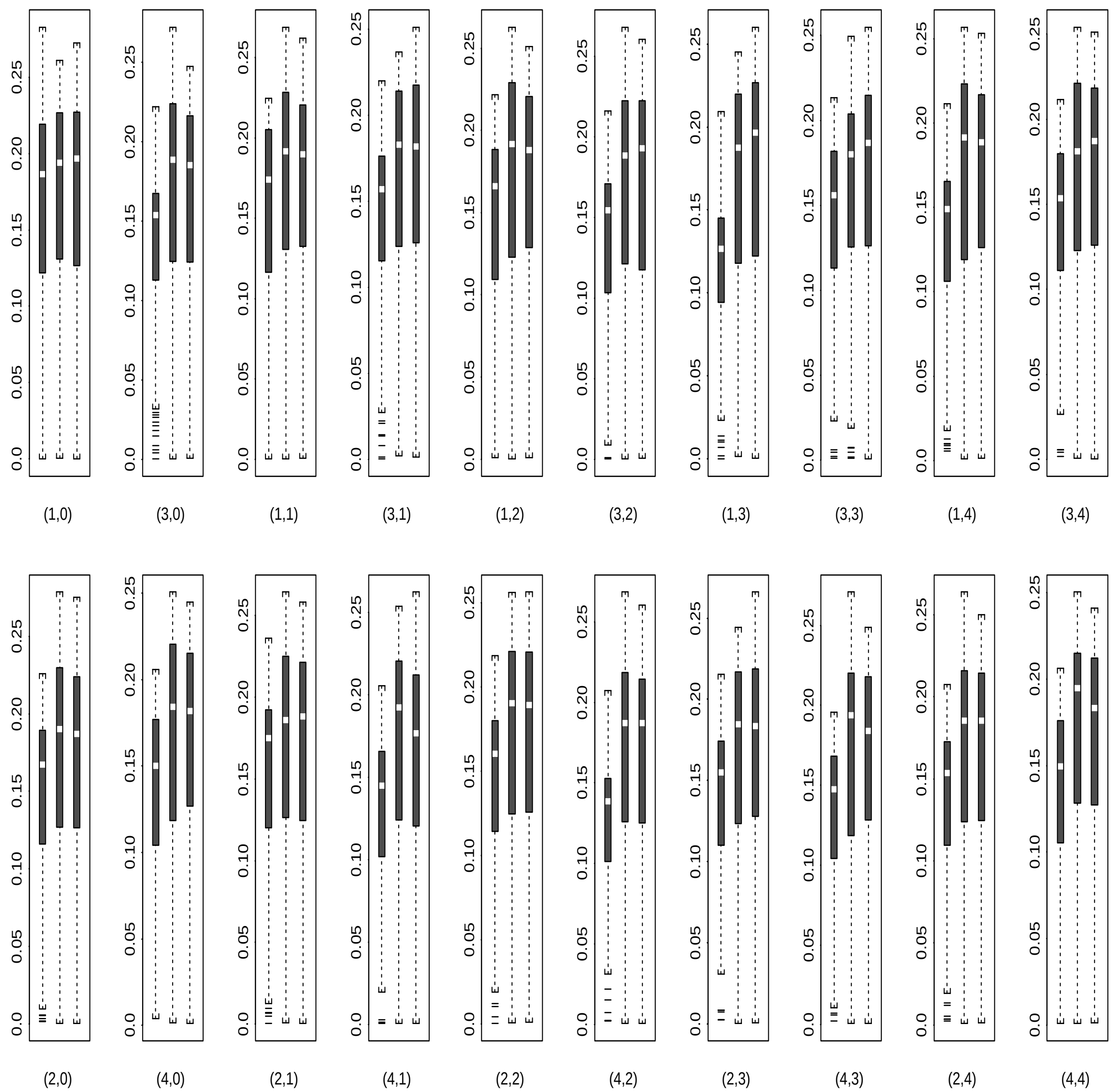


Figure 10: CPO plots of monthly rates(SFr./Index) for normal, t_4-AR(p) and hyperbolic AR(p) models

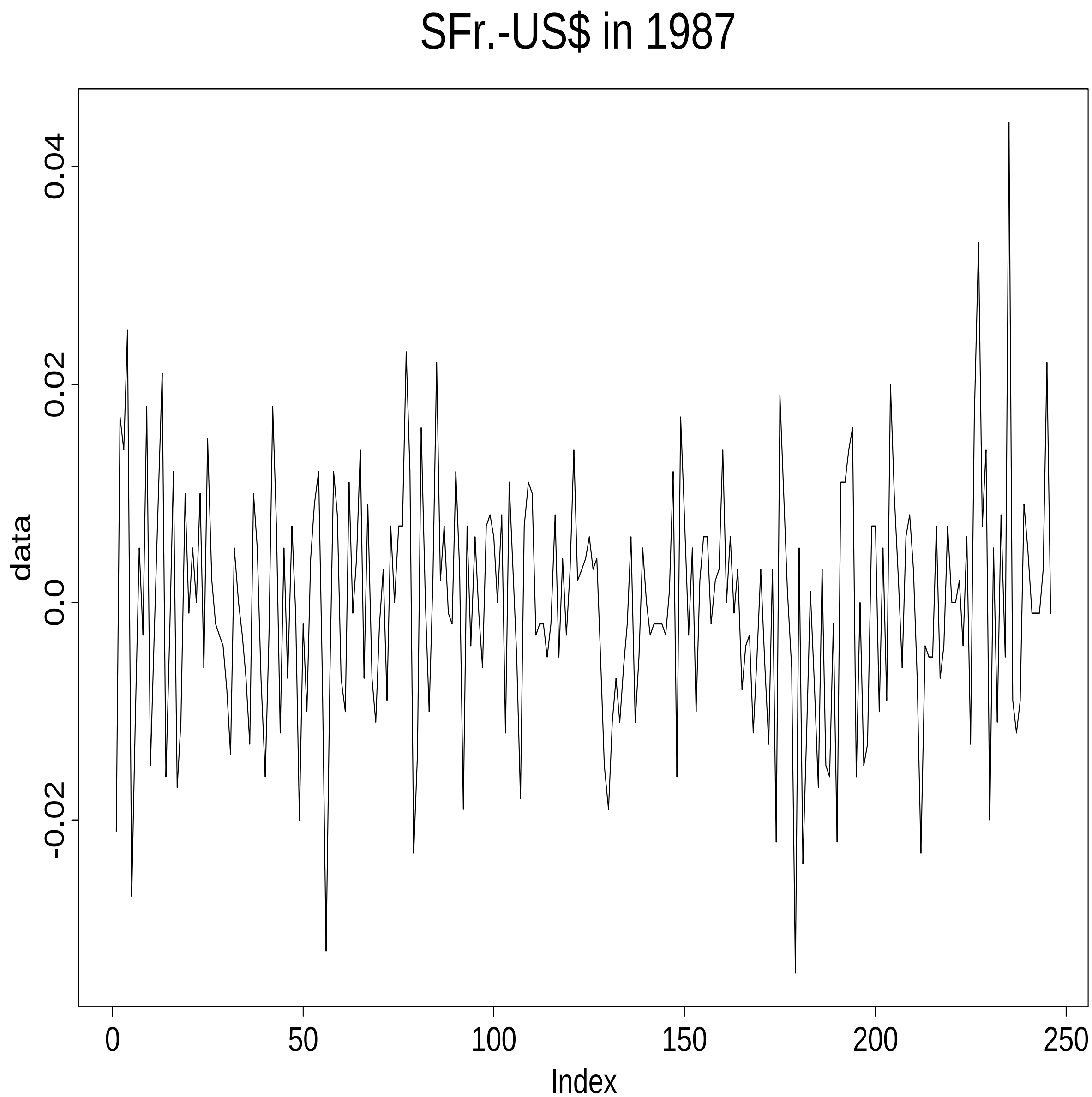


Figure 11: Daily exchange rate for SFr./US$ in 1987

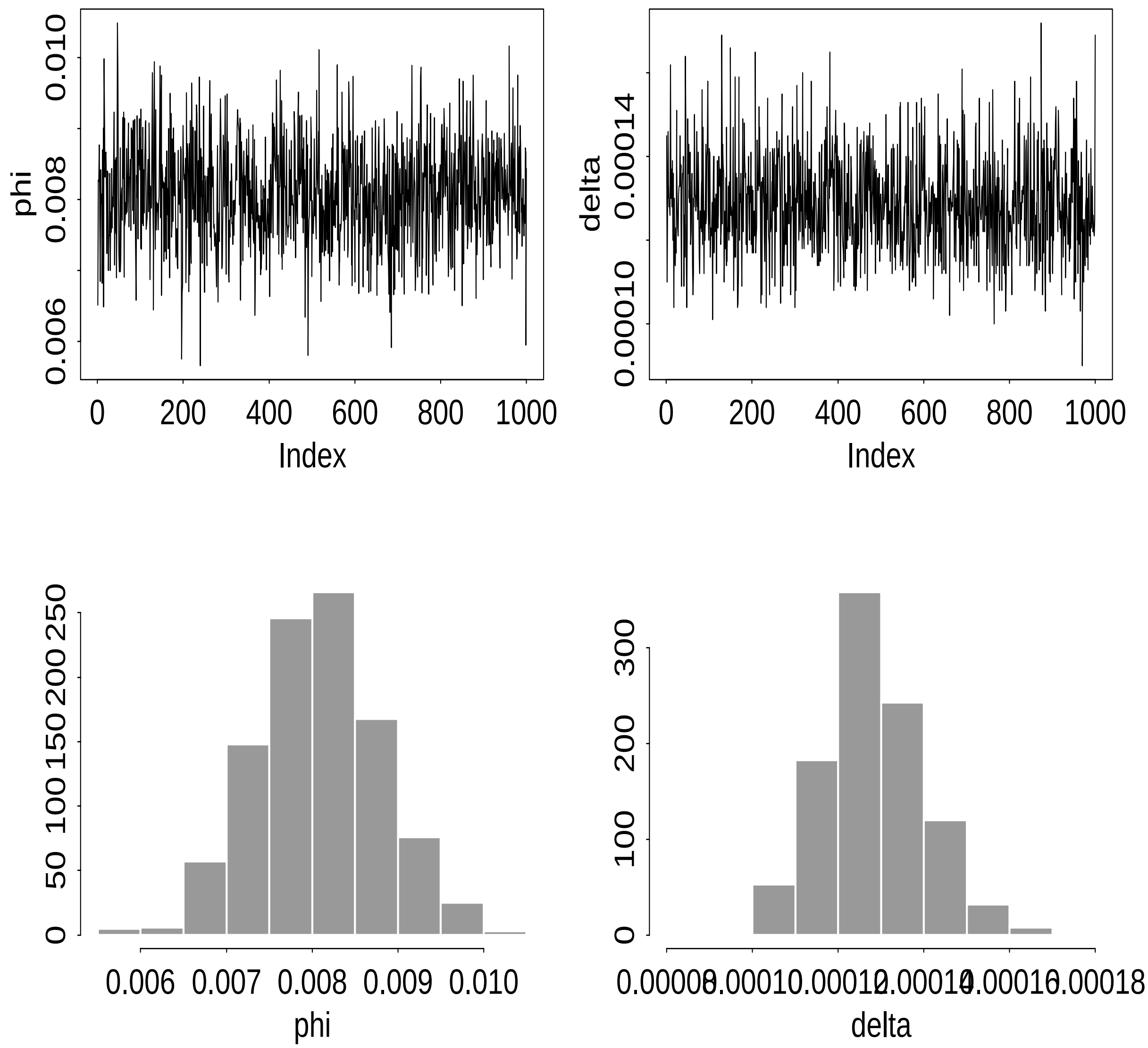


Figure 12: AR(1) model of daily exchange rates(SFr./US\$) for $N(0, \delta^2)$

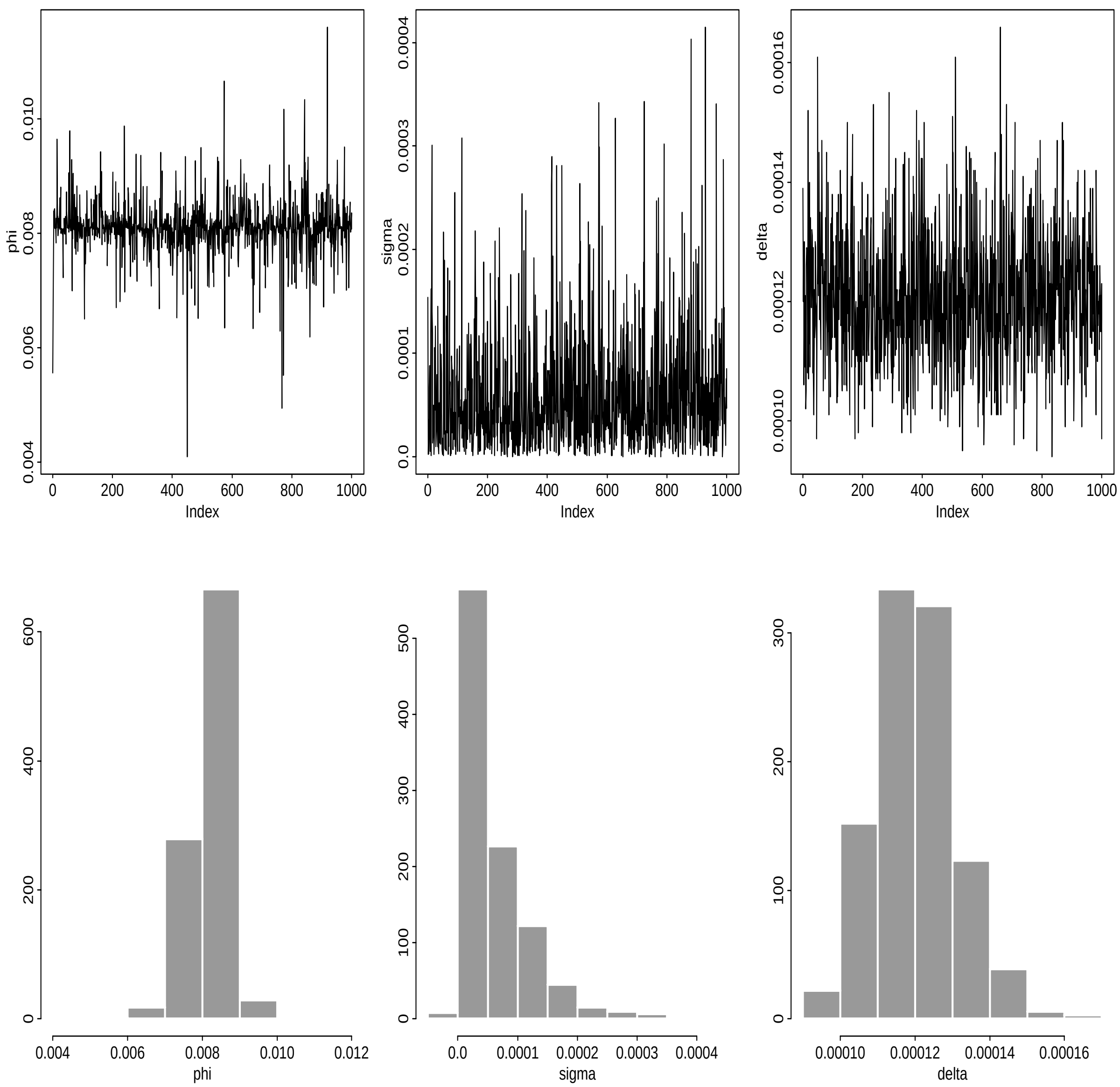


Figure 13: The Gibbs output of the t_4-AR(1) model of daily exchange rates(SFr./US\$)

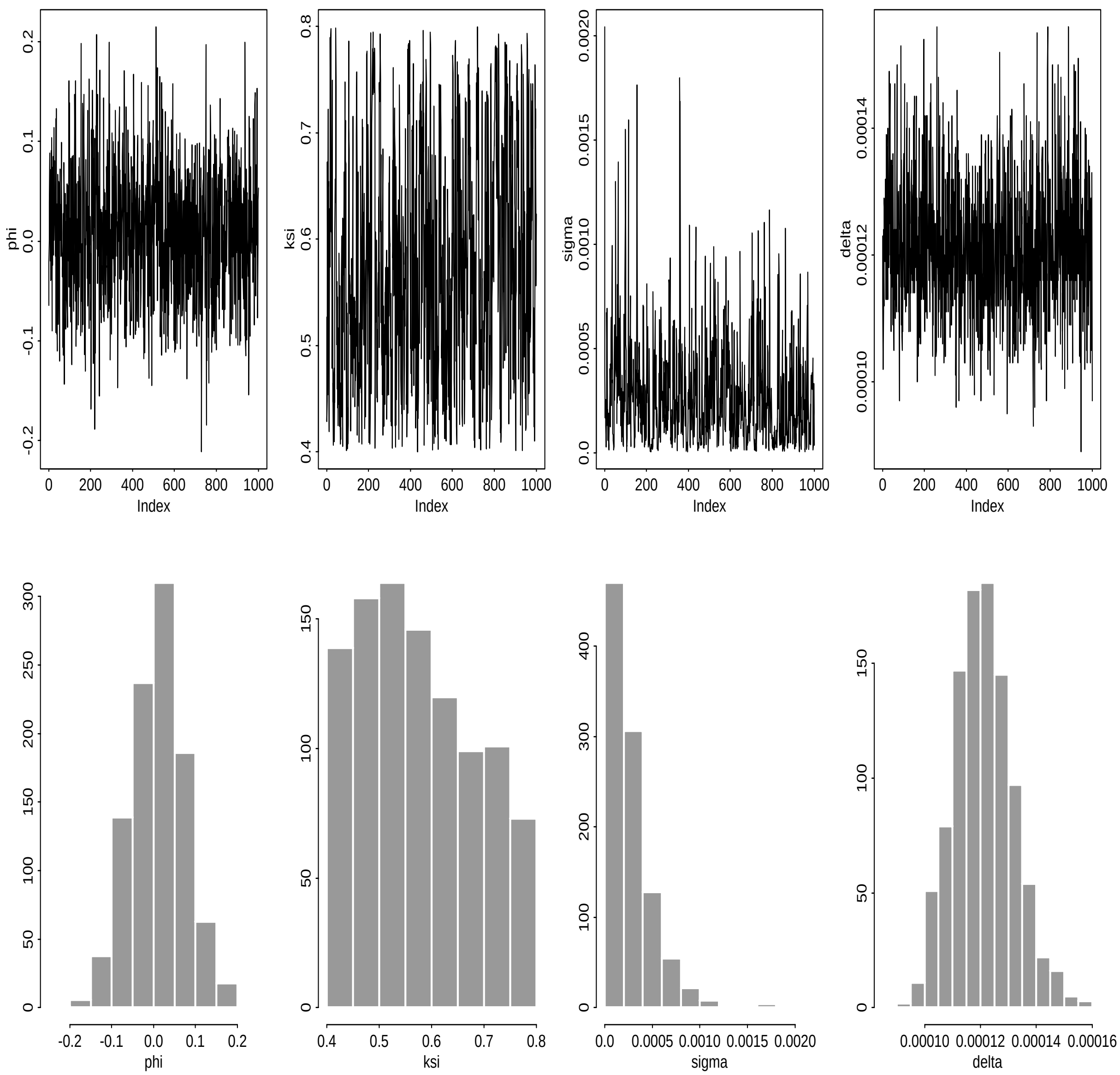


Figure 14: The Gibbs output of the *hyp*-AR(1) model of daily exchange rates(SFr./US$)

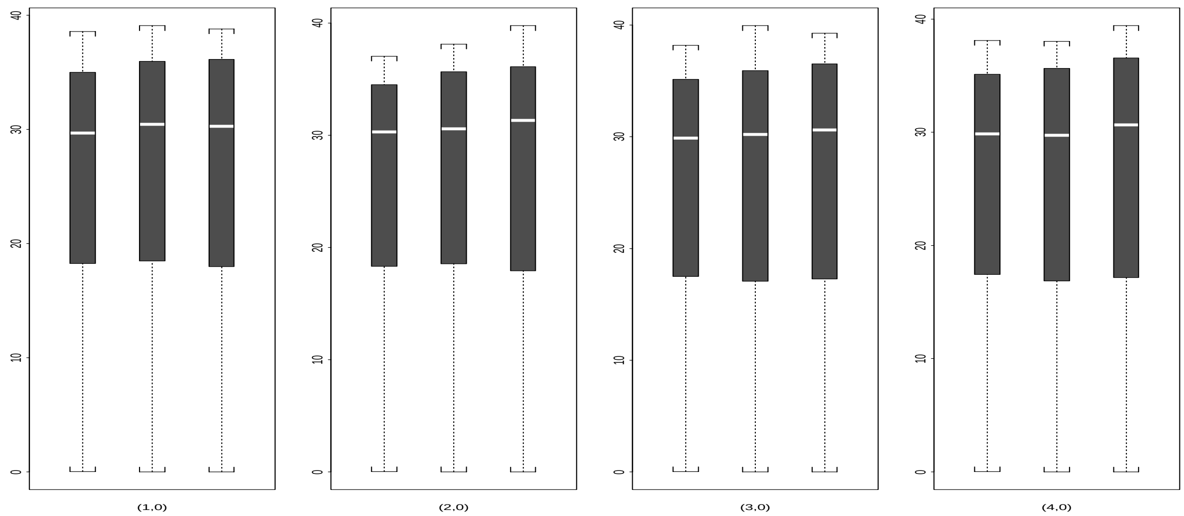


Figure 15: CPO plots of daily exchange rates(SFr./US$) for normal, t_4 and hyperbolic AR(p) models for $p = 1, \ldots, 4$

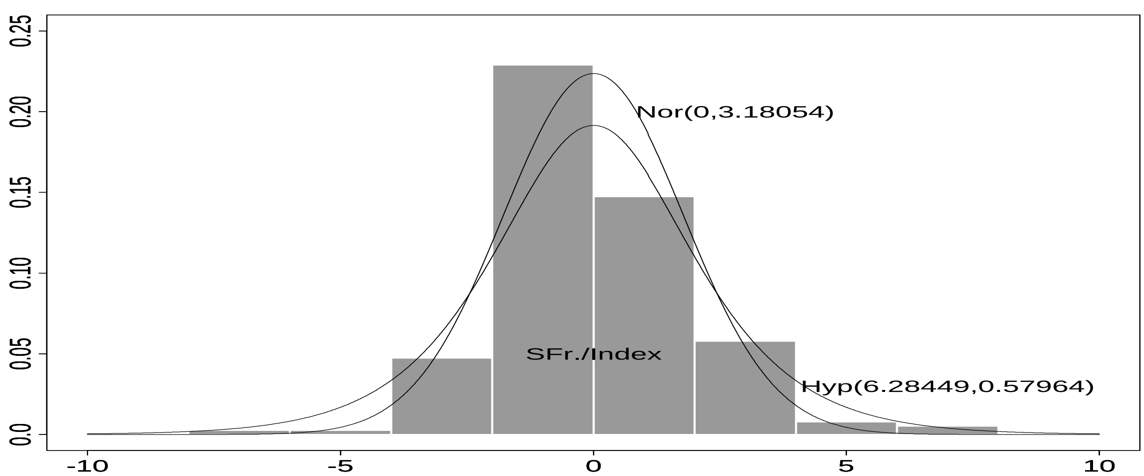
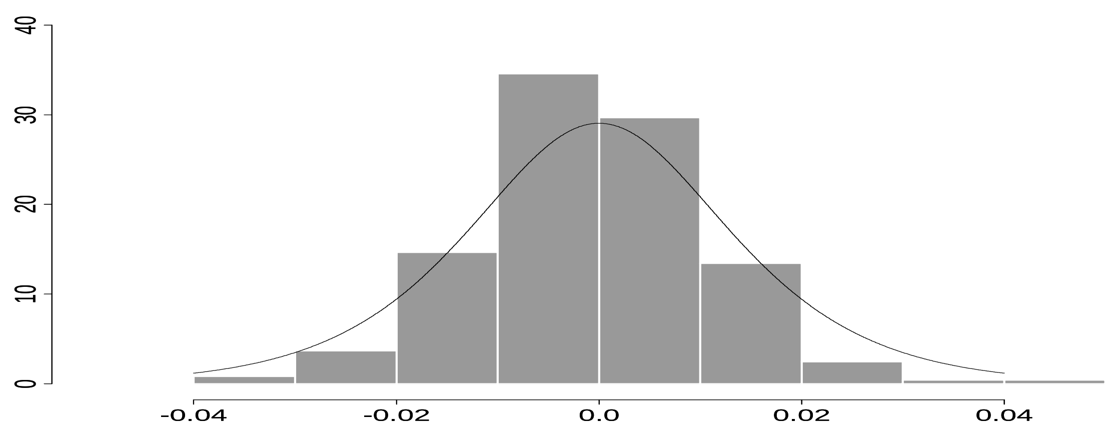


Figure 6: The histogram of SFr./Index compared with Nor(0,δ) and Hyp(σ,ξ)


0
10
20
30
40
-0.04
-0.02
0.0
0.02
0.04