RM-76-75

INFERENCE ROBUSTNESS OF ARIMA MODELS UNDER NON-NORMALITY - SPECIAL APPLICATION TO STOCK PRICE DATA

Johannes Ledolter

October 1976

Preface

An important class of stochastic models for describing time series is the class of autoregressive integrated moving average (ARIMA) models. This class provides a range of models, stationary and non-stationary, that adequately represent many of the time series met in practice.

At IIASA this class of models has been used to describe the time dependence of observations and to predict future observations from past data. Previous publications in the System and Decision Sciences area on this topic, for example, include a comparison of forecasts from ARIMA models and forecasts derived from exponential smoothing. Furthermore it is shown how ARIMA models can be used in modelling hydrologic sequences.

A common assumption in ARIMA models is the normality of the distribution of the errors (shocks which drive the system). In this paper it is investigated whether this assumption is critical or whether inference and prediction of ARIMA models is robust with respect to non-normality of the error distribution.

Inference Robustness of ARIMA Models under Non-normality - Special Application to Stock Price Data

Abstract:

Wold's [16] decomposition theorem states that every weakly stationary stochastic process can be decomposed into orthogonal shocks. For practical reasons, however, it is desirable to employ models which use parameters parsimoniously. Box and Jenkins [3] show how parsimony can be achieved by representing the linear process in terms of a small number of autoregressive and moving average terms (ARIMA-models). The Gaussian hypothesis assumes that the shocks follow a normal distribution with fixed mean and variance. In this case the process is characterized by first and second order moments. The normality assumption seems to be reasonable for many kinds of series. However, it was pointed out by Kendall [8], Mandelbrot [10,11,12], and Fama [6] that particularly for stock price data the distribution of the shocks appears leptokurtic.

In this paper we investigate the sensitivity of ARIMA models to non-normality of the distribution of the shocks. We suppose that the distribution function of the shocks is a member of the symmetric exponential power family, which includes the normal as well as leptokurtic and platikurtic distributions. A Bayesian approach is adopted and the inference robustness of ARIMA models with respect to

i) the estimation of parameters

ii) the forecasts of future observations

is discussed.

1. Statistical models for stock price series

An early contribution to the theory of stock prices was made by Bachelier [1]. He suggested the random walk model with normally distributed errors as a possible stochastic model for stock price series (Gaussian random walk hypothesis).

Empirical studies of stock price data show that successive differences of stock prices are nearly independent, thus confirming the random walk hypothesis [8,10]. However it is pointed out that the Gaussian hypothesis is subject to some doubt, since the distribution of the error terms appears leptokurtic.

This led Mandelbrot [10,11,12] to adopt the stable Paretian random walk hypothesis, where it is assumed that differences of stock prices follow a stable distribution with characteristic exponent $1 \leq \alpha \leq 2$. The stable Paretian random walk hypothesis has important implications for data analysis, since whenever $\alpha < 2$ the variance is infinite and the sample standard deviation which is used to measure risk becomes meaningless. Some doubt of this hypothesis is expressed by other authors. Hsu, Miller and Wichern [7] point out that the stable Paretian random walk hypothesis does not agree with many stock price series observed in practice.

In the literature on economic stock price series various other characterizations have been put forward. Press [15] considers a mixture of normal distributions with different variances and Praetz [14] suggests a scaled t distribution to explain the leptokurtic distribution of the error terms. Miller, Wichern and Hsu [13], instead of characterizing the errors by leptokurtic distributions, relax the stationarity assumption of the model. Changes in the parameters of the model over time can lead to leptokurtic distributions, an aspect which is further discussed in Ledolter [9].

In this paper, however, we consider the consequences of a different hypothesis. We assume the usual form of the ARIMA model with constant parameters, but allow the possibility of a symmetric, but not necessarily normal error distribution. Instead of assuming it to be a stable distribution we assume that it is from the class of exponential power distributions.

2. ARIMA time series model with shocks from the family of symmetric exponential power distributions:

We consider the linear filter model

$$z_t = \psi(B)a_t \quad , \tag{2.1}$$

where

i) B is the backshift operator; $B^m z_t = z_{t-m}$,

ii) $\psi(B) = \dfrac{\theta_q(B)}{\phi_p(B)(1-B)^d} = \dfrac{1-\theta_1 B-\ldots-\theta_q B^q}{(1-\phi_1 B-\ldots-\phi_p B^p)(1-B)^d}$, and

iii) a_t are independent drawings (shocks) from the family of symmetric exponential power distributions with probability density

$$p(a) = \omega(\beta)\sigma^{-1} \exp\left\{- \frac{c(\beta)}{\sigma^{2/1+\beta}} \left|a\right|^{\frac{2}{1+\beta}}\right\} , \qquad (2.2)$$

with

$$\omega(\beta) = \frac{\left\{\Gamma\left(\frac{3}{2}(1+\beta)\right)\right\}^{1/2}}{(1+\beta)\left\{\Gamma\left(\frac{1}{2}(1+\beta)\right)\right\}^{3/2}} ,$$

$$c(\beta) = \left\{\frac{\Gamma\left(\frac{3}{2}(1+\beta)\right)}{\Gamma\left(\frac{1}{2}(1+\beta)\right)}\right\}^{\frac{1}{1+\beta}} .$$

The quantity $\sigma > 0$ is the standard deviation of the population and $\beta(-1<\beta\leq 1)$ is a measure of kurtosis indicating the extent of non-normality of the parent distribution of the shocks. If $\beta = 0$ the shocks are normally distributed. $\beta > 0$ will result in a leptokurtic and $\beta < 0$ in a platikurtic distribution. This family ranges from the uniform (β approaching -1) to the double exponential distribution ($\beta = 1$).

In time series analysis the usual assumption is that shocks come from a normal distribution ($\beta=0$). In this paper we characterize the error distribution through one additional parameter, β, thus broadening the model. Proceeding this way has proved useful in studies of what has been called inference robustness by Box and Tiao [4,5]. Inference robustness is concerned with how inferences are affected when assumptions about the underlying distribution are changed.

For $\beta \neq 0$ the distribution of z_t in (2.1) is complicated. However it is easily shown that for any given stationary process the kurtosis of z_t is given by

$$\gamma_2(z) = \frac{E(z^4)}{[E(z^2)]^2} - 3 = \frac{\sum_{j=0}^{\infty} \psi_j^4}{[\sum_{j=0}^{\infty}\psi_j^2]^2} \gamma_2(a)$$

where, as shown by Box [2],

$$\gamma_2(a) = \frac{\Gamma(\frac{5}{2}(1+\beta))\,\Gamma(\frac{1}{2}(1+\beta))}{[\Gamma(\frac{3}{2}(1+\beta))]^2} - 3 \quad .$$

3. Parameter estimation for autoregressive models of order p with shocks from the family of symmetric exponential power distributions

We consider the process

$$(1-\phi_1 B-\ldots-\phi_p B^p)\, z_t = a_t \quad , \tag{3.1}$$

where z_t is a stationary difference of original observations with $Ez_t = 0$. To assure stationarity the roots of $(1-\phi_1 B-\ldots-\phi_p B^p) = 0$ are assumed to lie outside the unit circle [3]. Furthermore the a_t are assumed independent with distribution given in (2.2). Thus,

$$p(a_{p+1},\ldots,a_n) = [\omega(\beta)]^{(n-p)}\sigma^{-(n-p)}\exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\sum_{t=p+1}^{n}\left|a_t\right|^{\frac{2}{1+\beta}}\right\} \quad .$$

Transforming

$$z_t - \phi_1 z_{t-1}-\ldots-\phi_p z_{t-p} = a_t \qquad p+1 \leq t \leq n$$

and treating $\underset{\sim}{z}_p' = (z_1,z_2,\ldots,z_p)$ as given, we derive the density of $\underset{\sim}{z}' = (z_{p+1},\ldots,z_n)$ to be

$$p(\underset{\sim}{z}|\sigma,\beta,\underset{\sim}{\phi},\underset{\sim}{z}_p) = [\omega(\beta)]^{(n-p)}\sigma^{-(n-p)}\exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\sum_{t=p+1}^{n}\left|z_t-\phi_1 z_{t-1}-\ldots-\phi_p z_{t-p}\right|^{\frac{2}{1+\beta}}\right\} \tag{3.2}$$

First we derive the posterior distribution of the parameters σ and $\underset{\sim}{\phi}' = (\phi_1,\ldots,\phi_p)$ for a specific parent distribution of the shocks, thus considering β fixed. For given β and fixed starting values $\underset{\sim}{z}_p$, the likelihood function of $(\sigma,\underset{\sim}{\phi})$ is given by

$$\ell(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta) \propto \sigma^{-(n-p)} \exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=p+1}^{n} |z_t - \phi_1 z_{t-1} - \ldots - \phi_p z_{t-p}|^{\frac{2}{1+\beta}}\right\} \tag{3.3}$$

Bayes formula states that the posterior distribution of $(\sigma,\underset{\sim}{\phi})$ is

$$p(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta) \propto p(\sigma,\underset{\sim}{\phi})\,\ell(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta)$$

where $\ell(\sigma,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta)$ is the likelihood given in (3.3) and $p(\sigma,\underset{\sim}{\phi})$ is a chosen prior.

An analysis would usually be required in circumstances where little was assumed to be known about the parameters a priori. The question of choosing a prior so as to be "non-informative" has been the subject of considerable research, speculation and arguments. In particular in cases where it is applicable one can use Jeffreys' principle to derive a non-informative prior distribution. According to this rule, the prior distribution is chosen proportional to the square root of Fisher's information matrix (see for example [4]). For the case $\beta < 0$ it is easily shown [9] that Jeffreys' principle leads to a prior distribution of the form

$$p(\sigma,\underset{\sim}{\phi}) = p(\sigma)p(\underset{\sim}{\phi}) \propto \sigma^{-1}|P_p|^{\frac{1}{2}} , \tag{3.4}$$

where P_p is a $p \times p$ autocorrelation matrix with elements

$$\rho_{|i-j|} = \frac{E(z_{t-i}z_{t-j})}{E(z_t^2)} \qquad 1 \le i,\ j \le p .$$

For autoregressive parameters well within the stationarity region the prior $p(\underset{\sim}{\phi})$ appears sufficiently flat compared with the likelihood and thus can be considered constant. In the following we therefore use the approximation

$$p(\sigma,\underset{\sim}{\phi}) \propto \sigma^{-1} \quad . \tag{3.5}$$

Combining the above prior with the likelihood in (3.3) we derive

$$p(\sigma,\underset{\sim}{\phi}\,|\,\underset{\sim}{z},\underset{\sim}{z}_p,\beta) \propto \sigma^{-(n-p+1)} \exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\, S(\underset{\sim}{\phi};\beta)\right\} \quad , \tag{3.6}$$

where

$$S(\underset{\sim}{\phi};\beta) = \sum_{t=p+1}^{n} |z_t - \phi_1 z_{t-1} - \ldots - \phi_p z_{t-p}|^{\frac{2}{1+\beta}} \quad .$$

Integration over σ gives

$$p(\underset{\sim}{\phi}\,|\,\underset{\sim}{z},\underset{\sim}{z}_p,\beta) \propto \{S(\underset{\sim}{\phi};\beta)\}^{-\frac{n-p}{2}(1+\beta)} \quad . \tag{3.7}$$

From (3.7) it is clear that the posterior distribution of $\underset{\sim}{\phi}$ depends heavily on the value of β. However, this does not necessarily mean that for a given body of data the inferences will be imprecise. Considering β as a random variable it will itself possess a posterior distribution. It is often the case in time series work that the number of observations is rather large. Thus some rather precise information about β can be supplied by the data and may be incorporated in the analysis.

Since there appears to be no reason why β should depend on σ and $\underset{\sim}{\phi}$ a priori, we assume that the prior of $(\sigma,\underset{\sim}{\phi},\beta)$ is given by

$$p(\sigma,\underset{\sim}{\phi},\beta) = p(\sigma,\underset{\sim}{\phi})p(\beta) \propto \sigma^{-1}p(\beta) \quad . \tag{3.8}$$

Box and Tiao [5] introduce the concept of a reference prior for β. This is usually, but not necessarily, taken to be a uniform prior and is intended, as its name implies, for reference purposes. It has the property that if the data are viewed in the light of some other prior distribution the new posterior distribution could be readily obtained by using the reference prior. Using the prior in (3.8) we derive

$$p(\sigma,\underset{\sim}{\phi},\beta|\underset{\sim}{z},\underset{\sim}{z}_p) \propto p(\beta)\,[\omega(\beta)]^{n-p}\,\sigma^{-(n-p+1)}\exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\,S(\underset{\sim}{\phi};\beta)\right\} \quad . \tag{3.9}$$

Integrating over σ

$$p(\beta,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p) \propto p(\beta)\,\frac{\Gamma(1+\frac{n-p}{2}(1+\beta))}{[\Gamma(1+\frac{1}{2}(1+\beta))]^{n-p}}\left\{S(\underset{\sim}{\phi};\beta)\right\}^{-\frac{n-p}{2}(1+\beta)} \quad . \tag{3.10}$$

Furthermore

$$p(\beta|\underset{\sim}{z},\underset{\sim}{z}_p) \propto \int_{SR} p(\beta,\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p)\,d\underset{\sim}{\phi} \tag{3.11}$$

where the stationarity region

$$SR = \left\{\underset{\sim}{\phi}:(1-\phi_1B-\ldots-\phi_pB^p) = 0 \ \ \text{has roots outside the unit circle}\right\} \quad .$$

It is not possible to obtain a closed form expression for (3.11). However for low order autoregressive processes the integral can be evaluated numerically.

The posterior distribution $p(\beta|\underset{\sim}{z},\underset{\sim}{z}_p)$ serves as weight function in deriving the posterior distribution of $\underset{\sim}{\phi}$.

$$p(\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p) = \int_{-1}^{+1} p(\underset{\sim}{\phi}|\underset{\sim}{z},\underset{\sim}{z}_p,\beta)\,p(\beta|\underset{\sim}{z},\underset{\sim}{z}_p)\,d\beta \quad . \tag{3.12}$$

4. Parameter estimation for moving average models of order q with shocks from the family of symmetric exponential power distributions.

We consider the invertible model

$$z_t = (1 - \theta_1 B - \ldots - \theta_q B^q)\,a_t \quad , \tag{4.1}$$

where z_t is a stationary difference with $Ez_t = 0$ and where a_t are independent drawings from the distribution in (2.2). To assure invertibility we assume that the roots of $1-\theta_1 B-\ldots-\theta_q B^q = 0$ lie outside the unit circle [3]. Again,

$$p(a_1,a_2,\ldots,a_n|\sigma,\beta) = [\omega(\beta)]^n \sigma^{-n} \exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}} \sum_{t=1}^{n} |a_t|^{\frac{2}{1+\beta}}\right\} \quad .$$

Transforming

$$z_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \qquad 1 \le t \le n \quad ,$$

and treating the starting values $\underset{\sim}{a}_*' = (a_0, a_{-1}, \ldots, a_{-(q-1)})$ as nuisance parameters, we get

$$p(\underset{\sim}{z}|\sigma,\beta,\underset{\sim}{\theta},\underset{\sim}{a}_*) = [\omega(\beta)]^n \sigma^{-n} \exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}} S(\underset{\sim}{\theta};\beta,\underset{\sim}{a}_*)\right\} \tag{4.2}$$

where

$$S(\underset{\sim}{\theta};\beta,\underset{\sim}{a}_*) = \sum_{t=1}^{n} \left|\sum_{j=0}^{t-1} \pi_j z_{t-j} + \pi_t a_0 + (\theta_2 \pi_{t-1} + \ldots + \theta_q \pi_{t-(q-1)}) a_{-1} + \ldots + \theta_q \pi_{t-1} a_{-(q-1)}\right|^{\frac{2}{1+\beta}}$$

$\pi_0 = -1$ and the π_j weights $(j \geq 1)$ are the coefficients in the expansion

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = (1 - \theta_1 B - \ldots - \theta_q B^q)^{-1} \quad .$$

The likelihood function for (4.1) is given by

$$\ell(\sigma, \beta, \underset{\sim}{\theta}, \underset{\sim}{a}_* | \underset{\sim}{z}) \propto [\omega(\beta)]^n \sigma^{-n} \exp\left\{- \frac{c(\beta)}{\sigma^{2/1+\beta}} S(\underset{\sim}{\theta}; \beta, \underset{\sim}{a}_*)\right\} \quad (4.3)$$

Combining the non-informative prior distribution

$$p(\sigma, \underset{\sim}{\theta}, \underset{\sim}{a}_*) \propto \sigma^{-1}$$

with the likelihood in (4.3) we derive for given β

$$p(\sigma, \underset{\sim}{\theta}, \underset{\sim}{a}_* | \underset{\sim}{z}, \beta) \propto \sigma^{-(n+1)} \exp\left\{- \frac{c(\beta)}{\sigma^{2/1+\beta}} S(\underset{\sim}{\theta}; \beta, \underset{\sim}{a}_*)\right\} \quad . \quad (4.4)$$

Integrating over σ

$$p(\underset{\sim}{\theta}, \underset{\sim}{a}_* | \underset{\sim}{z}, \beta) \propto \{S(\underset{\sim}{\theta}; \beta, \underset{\sim}{a}_*)\}^{-\frac{n}{2}(1+\beta)} \quad . \quad (4.5)$$

Since for any invertible process the π-weights decrease fairly rapidly, the choice of the starting values $\underset{\sim}{a}_*$ will not be critical. A sensible approximation, and one which is most convenient in practice, is

$$p(\underset{\sim}{\theta} | \underset{\sim}{z}, \beta) = \int p(\underset{\sim}{\theta}, \underset{\sim}{a}_* | \underset{\sim}{z}, \beta) d\underset{\sim}{a}_*$$

$$\simeq p(\underset{\sim}{\theta}, \underset{\sim}{a}_* = \underset{\sim}{0} | \underset{\sim}{z}, \beta) \propto \{S(\underset{\sim}{\theta}; \beta)\}^{-\frac{n}{2}(1+\beta)} \quad (4.6)$$

where

$$S(\underset{\sim}{\theta}; \beta) = S(\underset{\sim}{\theta}; \beta, \underset{\sim}{a}_* = \underset{\sim}{0}) = \sum_{t=1}^{n} \left| \sum_{j=0}^{t-1} \pi_j z_{t-j} \right|^{\frac{2}{1+\beta}} \quad .$$

Treating β as random variable with prior $p(\beta)$

$$p(\sigma,\beta,\underset{\sim}{\theta}|\underset{\sim}{z}) \propto p(\beta)\,[\omega(\beta)]^{n}\sigma^{-(n+1)}\exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\,S(\underset{\sim}{\theta};\beta)\right\} \tag{4.7}$$

and

$$p(\beta,\underset{\sim}{\theta}|\underset{\sim}{z}) \propto p(\beta)\frac{\Gamma(1+\frac{n}{2}(1+\beta))}{[\Gamma(1+\frac{1}{2}(1+\beta))]^{n}}\,\{S(\underset{\sim}{\theta};\beta)\}^{-\frac{n}{2}(1+\beta)} \tag{4.8}$$

$$p(\beta|\underset{\sim}{z}) \propto \int_{IR} p(\beta,\underset{\sim}{\theta}|\underset{\sim}{z})\,d\underset{\sim}{\theta} \tag{4.9}$$

where the invertibility region IR = $\{\underset{\sim}{\theta}:(1-\theta_1B-\ldots-\theta_qB^q) = 0$ has all roots outside the unit circle$\}$

$$p(\underset{\sim}{\theta}|\underset{\sim}{z}) = \int_{-1}^{1} p(\underset{\sim}{\theta}|\underset{\sim}{z},\beta)\,p(\beta|\underset{\sim}{z})\,d\beta \quad . \tag{4.10}$$

5. Forecasting time series models with shocks from the family of symmetric exponential power distributions with special reference to the ARIMA (0,1,1) model.

In the following we discuss the integrated moving average model of the form

$$z_t - z_{t-1} = a_t - \theta a_{t-1} \quad . \tag{5.1}$$

This type of model is particularly important since many economic, business and engineering data behave according to this model. Furthermore, as pointed out in Section 1, stock price data follow a model of this kind in which the moving average parameter is close to zero.

Sample theory approach

Two approaches to forecasting can be distinguished. The first is a sample theory approach. The minimum mean square error (MMSE) forecast of a future observation $z_{n+\ell}$ is the conditional expectation of $z_{n+\ell}$ at time n. For any class of distributions with finite second order moments the distributional assumption about the shocks a_t are irrelevant for the derivation of the MMSE forecast. Forecasts, however, are of little value if they are not accompanied by some measure of their variability. The variance of the forecast error

$$e_n(\ell) = z_{n+\ell} - \hat{z}_n(\ell) = a_{n+\ell} + \sum_{j=1}^{\ell-1} \psi_j a_{n+\ell-j} \quad , \qquad (5.2)$$

provides such a measure, and is given by

$$V(e_n(\ell)) = \sigma_a^2 (1 + \sum_{j=1}^{\ell-1} \psi_j^2) \quad .$$

The distributional assumptions about the shocks a_t change the interpretation of the probability interval

$$\{\hat{z}_n(\ell) \pm \lambda[V(e_n(\ell))]^{\frac{1}{2}}\} \quad . \qquad (5.3)$$

If one is interested in one step ahead forecasts, the forecast error is given by $e_n(1) = a_{n+1}$. Thus $(1-\alpha)$ 100% probability limits for the future observation z_{n+1} are given by $\{\hat{z}_n(1) \pm \lambda\sigma_a\}$ where λ is chosen such that Prob $\{|a| > \lambda\sigma_a\} = \alpha$. For the case $\beta = 0$, the normal table provides λ corresponding to α. If $\beta > 0$ however, the distribution of the shocks a_t is leptokurtic and the normal probability limits will underestimate the risk of a realization in the extreme tails. For the platikurtic case ($\beta < 0$) the normal theory probability limits will overestimate this risk. Box and Tiao [5] show that the probability limits can be quite different for very small α (α<<.05); however for α = .05 they are not very sensitive to the choice of β.

For lead times $\ell \geq 2$ the distribution of the forecast errors $e_n(\ell)$ can readily be derived only in the case of a normal parent. For general β its distribution is complicated. Nevertheless some idea of the approach to normality of $e_n(\ell)$ can be obtained by considering the kurtosis of the forecast error

$$\gamma_2(e_n(\ell)) = \frac{1 + \sum_{j=1}^{\ell-1} \psi_j^4}{\left[1 + \sum_{j=1}^{\ell-1} \psi_j^2\right]^2} \gamma_2(a) \quad . \tag{5.4}$$

The kurtosis depends on

i) the non-normality parameter in the distribution of the shocks,

ii) the ψ-weights of the ARIMA model.

For the case of an ARIMA (0,1,1) model, $\psi_j = (1-\theta)$ for all $j \geq 1$ and

$$\gamma_2(e_n(\ell)) = \frac{1 + (\ell-1)(1-\theta)^4}{[1 + (\ell-1)(1-\theta)^2]^2} \gamma_2(a) \quad . \tag{5.5}$$

The above sample theory interpretation of forecasting has the drawback that it assumes that the values of the parameters are known. But parameters are estimated and parameter estimation errors are therefore present. A sample theory development which allows for errors in the parameters would be extremely difficult. However, some progress has been made by investigating how much the variance of the forecast errors increases if the parameters are estimated from the data [3].

<u>Bayesian approach</u>

Another approach to forecasting is a Bayesian one. This approach does provide a manageable way of incorporating estimation errors in the parameters. Treating the parameters in the ARIMA model as random variables, the predictive distribution of future observations can be derived.

We illustrate this approach for the one step ahead predictive distribution of the integrated first order moving average process.

For known parameters of the process the one-step-ahead predictive distribution of the integrated first order moving average process is given by

$$p(z_{n+1}|\sigma,\beta,\theta,\underset{\sim}{z}) = \omega(\beta)\,\sigma^{-1}\exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\,|z_{n+1}-\hat{z}_n(1)|^{\frac{2}{1+\beta}}\right\} \quad (5.6)$$

where

$$\hat{z}_n(1) = (1-\theta)\sum_{j=0}^{n-1}\theta^j z_{n-j} \quad .$$

In a Bayesian context the parameters σ,β,θ are considered random and its posterior distribution can be derived.

The one step ahead predictive distribution (unconditional on the parameters) is thus given by

$$\begin{aligned} p(z_{n+1}|\underset{\sim}{z}) &= \int_{-1}^{1}\int_{-1}^{1}\int_{0}^{\infty} p(z_{n+1},\sigma,\beta,\theta|\underset{\sim}{z})\,d\sigma d\beta d\theta \\ &= \int_{-1}^{1}\int_{-1}^{1}\int_{0}^{\infty} p(z_{n+1}|\sigma,\beta,\theta,\underset{\sim}{z})\,p(\sigma,\beta,\theta|\underset{\sim}{z})\,d\sigma d\beta d\theta . \end{aligned} \quad (5.7)$$

Combining (5.6) with the posterior distribution of the parameters in (4.7) we derive

$$p(z_{n+1},\sigma,\beta,\theta|\underset{\sim}{z}) \propto p(\beta)\,[\omega(\beta)]^{n+1}\,\sigma^{-(n+2)}\exp\left\{-\frac{c(\beta)}{\sigma^{2/1+\beta}}\left[|z_{n+1}-\hat{z}_n(1)|^{\frac{2}{1+\beta}}+S(\theta;\beta)\right]\right\} \tag{5.8}$$

Integrating over σ and for fixed β

$$p(z_{n+1},\theta|\underset{\sim}{z},\beta) \propto [S(\theta;\beta)]^{-\frac{1}{2}(1+\beta)}\left\{1+\frac{|z_{n+1}-\hat{z}_n(1)|^{\frac{2}{1+\beta}}}{S(\theta;\beta)}\right\}^{-\frac{n+1}{2}(1+\beta)} p(\theta|\underset{\sim}{z},\beta) \tag{5.9}$$

where $p(\theta|\underset{\sim}{z},\beta)$ is given in (4.6)

The posterior predictive distribution for given β and θ is given by

$$p(z_{n+1}|\underset{\sim}{z},\beta,\theta) \propto \left\{1+\frac{|z_{n+1}-\hat{z}_n(1)|^{\frac{2}{1+\beta}}}{S(\theta;\beta)}\right\}^{-\frac{n+1}{2}(1+\beta)} \tag{5.10}$$

and

$$p(z_{n+1}|\underset{\sim}{z},\beta) = \int_{-1}^{1} p(z_{n+1}|\underset{\sim}{z},\beta,\theta)\,p(\theta|\underset{\sim}{z},\beta)\,d\theta\ . \tag{5.11}$$

It can be seen that for the case $\beta = 0$, $p(z_{n+1}|\underset{\sim}{z},\beta=0)$ is an average of t distributions, weighted by $p(\theta|z,\beta=0)$.

Equation (5.11) allows one to determine the sensitivity (or conversely the robustness) of the predictive distribution to changes in the assumption about the distribution of the β shocks a_t. It expresses how the predictive distribution varies for changing β.

$p(z_{n+1}|\underset{\sim}{z})$ is derived by averaging the conditional predictive distributions in (5.11). The posterior distribution of β given in (4.9) acts as a weighting function.

$$p(z_{n+1}|\underset{\sim}{z}) = \int_{-1}^{1} p(z_{n+1}|\underset{\sim}{z},\beta)p(\beta|\underset{\sim}{z})d\beta \quad . \tag{5.12}$$

6. Example and concluding remarks

To illustrate the above robustness study we analyze daily IBM common stock closing prices given in (3). It is shown that the model for this series is given by an integrated first order moving average process. Under the assumption of normal shocks the point estimates (means of the posterior distribution) are

$$\hat{\theta} = -.087 \qquad \hat{\sigma}_a^2 = 52.2 \quad .$$

The conditional posterior distributions $p(\theta|\underset{\sim}{z},\beta)$ are derived for various values of β. They are given in Figure 1 and show moderate sensitivity to changes in β.

The parameter β, considered as a random variable, possesses a posterior distribution. Assuming a uniform reference prior for β, the posterior distribution $p(\beta|\underset{\sim}{z})$ is derived and given in Figure 2. It shows very clearly that the error distribution is not normal (β=0); strong evidence for a leptokurtic error distribution is given.

The posterior distribution $p(\theta|\underset{\sim}{z})$ is compared to the posterior distribution of θ assuming normal distributed errors, $p(\theta|\underset{\sim}{z},\beta=0)$. Both are plotted in Figure 3.

The predictive distributions $p(z_{n+1}|\underset{\sim}{z},\beta)$ for various values of β are given in Figure 4. The modes of these distributions are shown to be insensitive to the choice of β; the shape of the distribution (uncertainty of forecast), however, changes considerably.

The predictive distributions $p(z_{n+1}|\underset{\sim}{z})$ and $p(z_{n+1}|\underset{\sim}{z},\beta=0)$ are compared in Figure 5. The difference in the shape of the predictive distributions is very marked and shows that the assumption of normality would result in quite a different 50% highest posterior density region for the forecast. 90% highest posterior density regions, however, are virtually the same, whereas the normality assumption would underestimate the risk in the extreme tails of the distribution. Summarizing, it can be said that in this example the pointforecast (mode of the posterior distribution) and 90% highest posterior density regions are not seriously affected by symmetric non-normality of the error distribution.

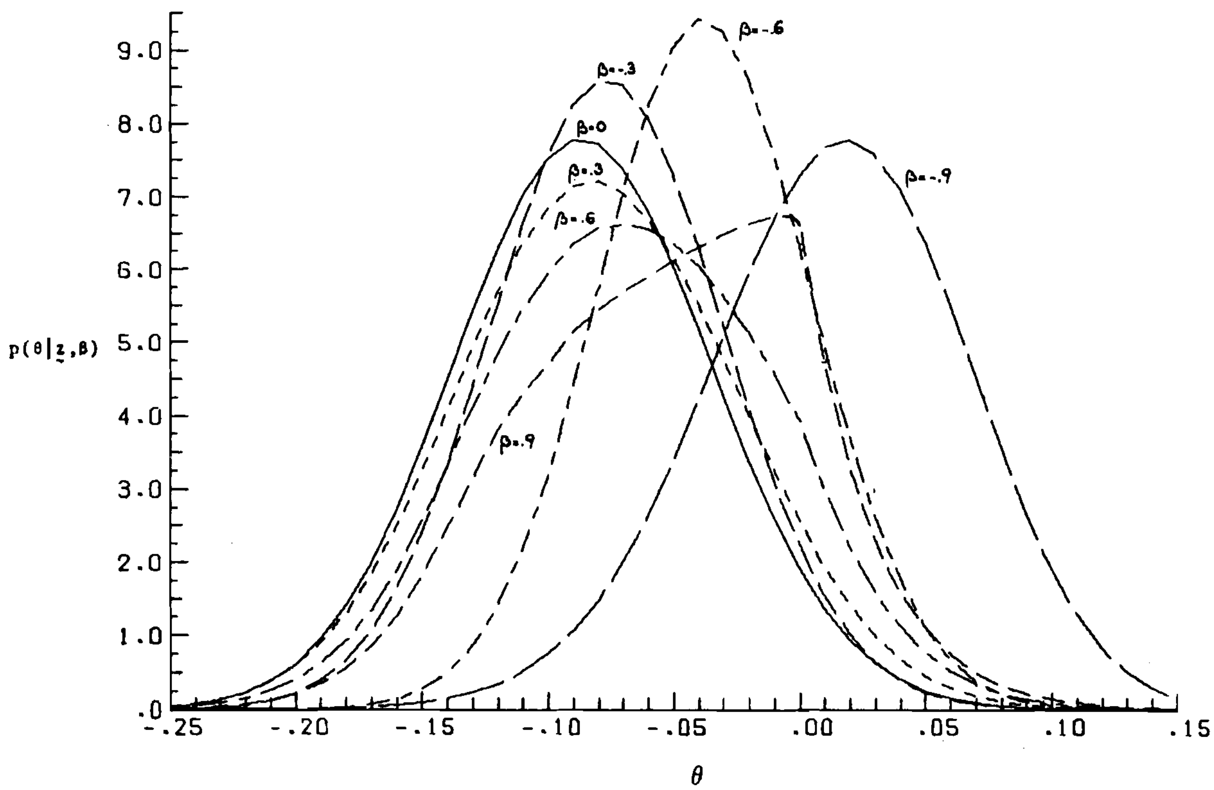

Figure 1. Posterior distribution of θ for various fixed β

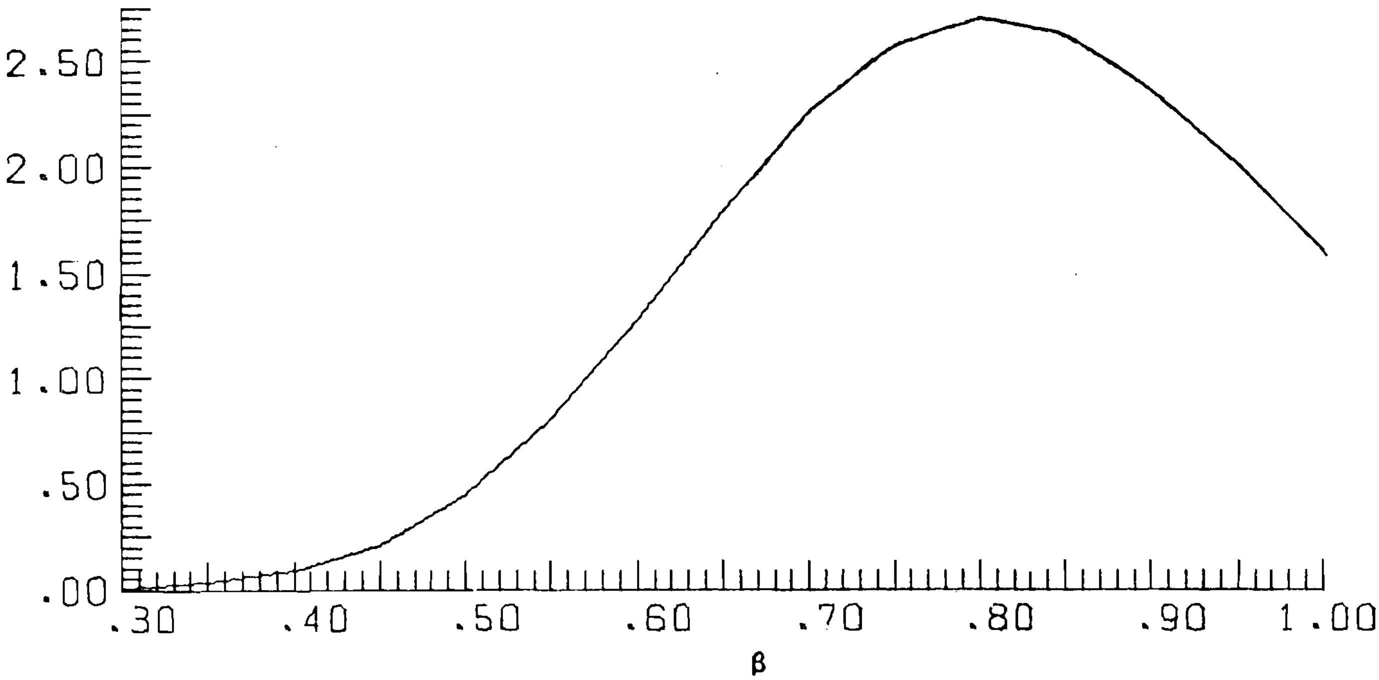

Figure 2. Posterior distribution of β assuming uniform reference prior.

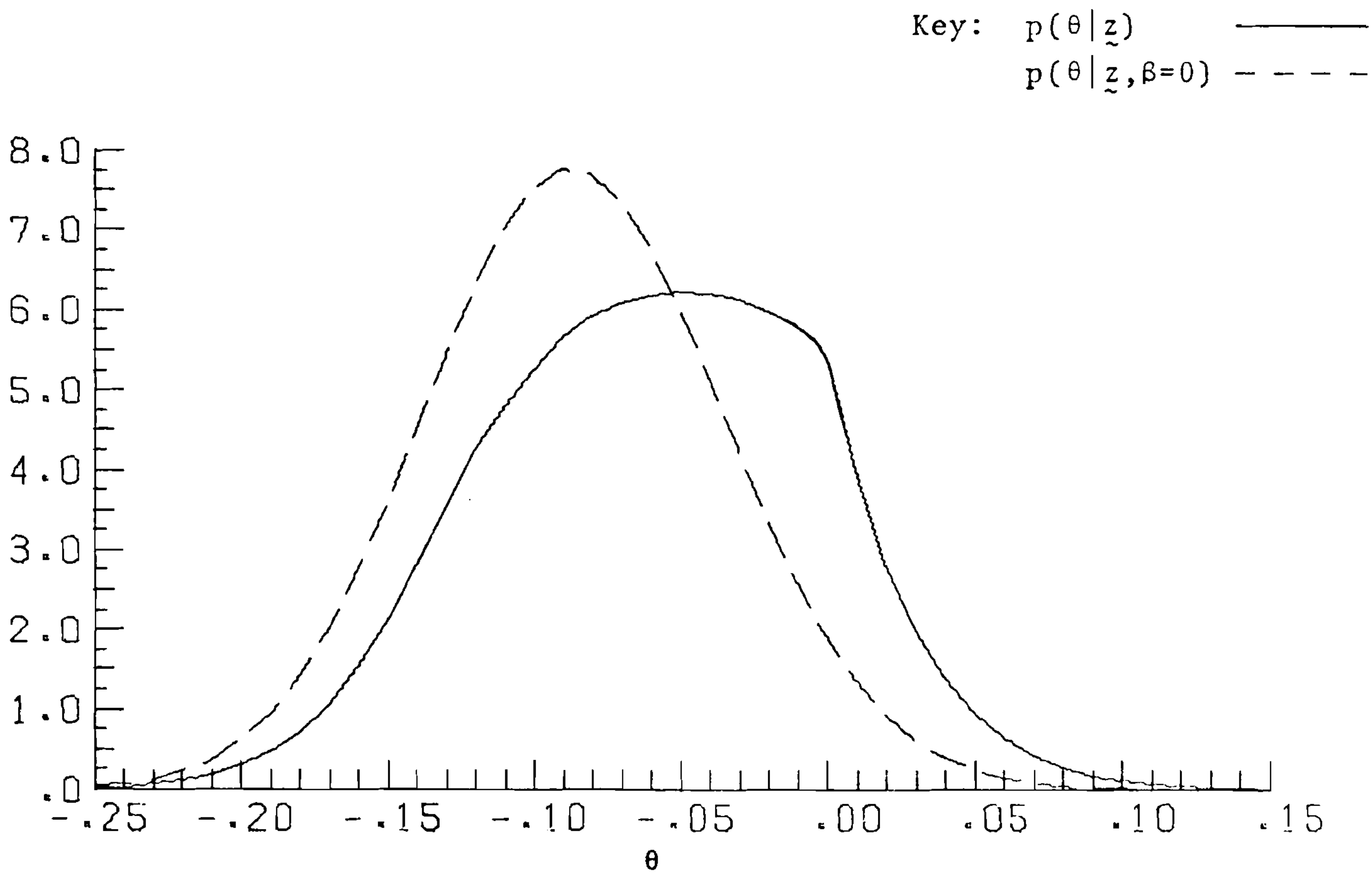

Figure 3. Posterior distribution of θ, $p(\theta/\underset{\sim}{z})$, and the posterior distribution of θ assuming normal distribution shocks, $p(\theta/\underset{\sim}{z}, \beta = 0)$.

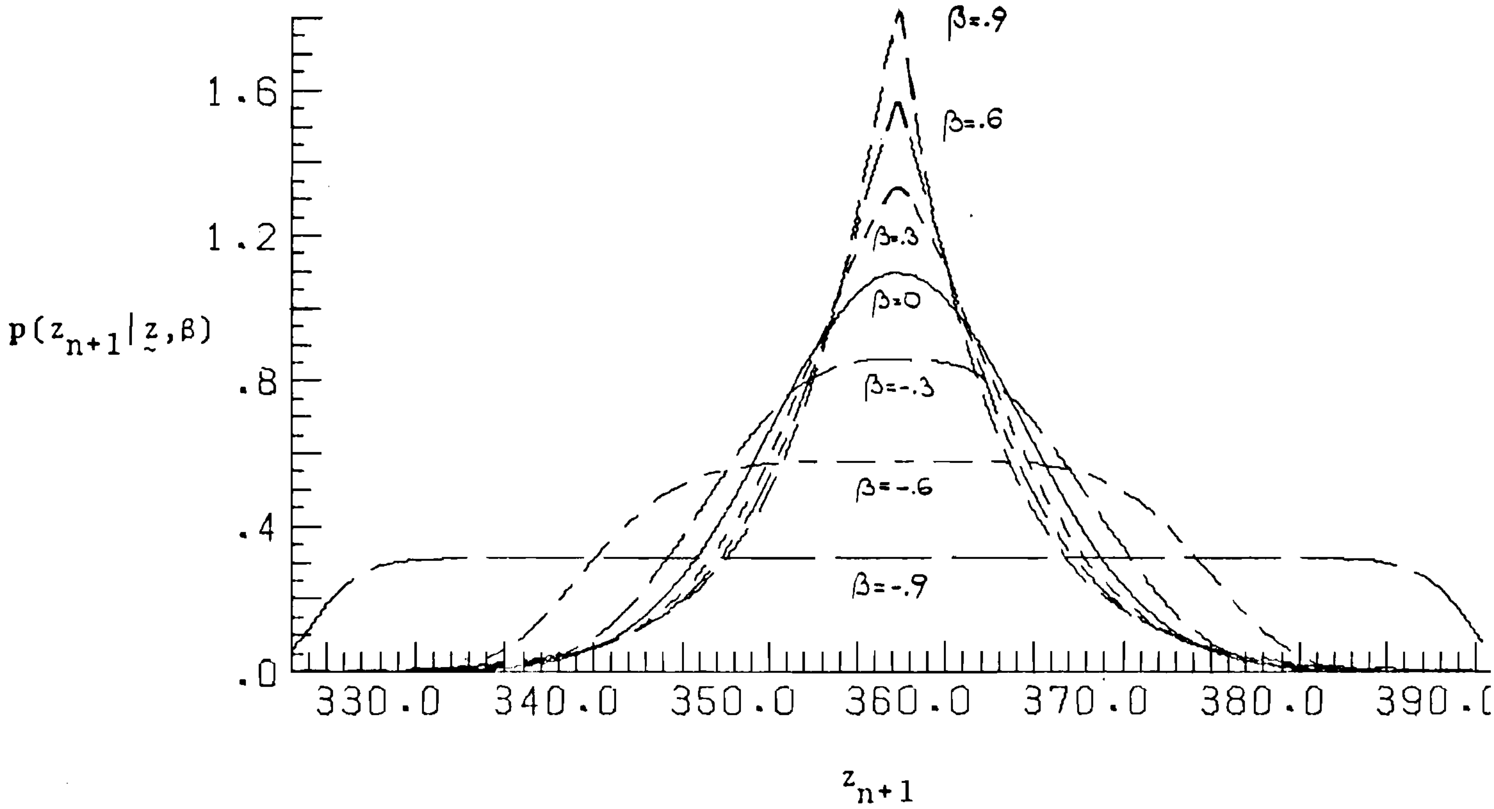

Figure 4. Posterior one step ahead predictive distribution for various fixed β: n = 369.

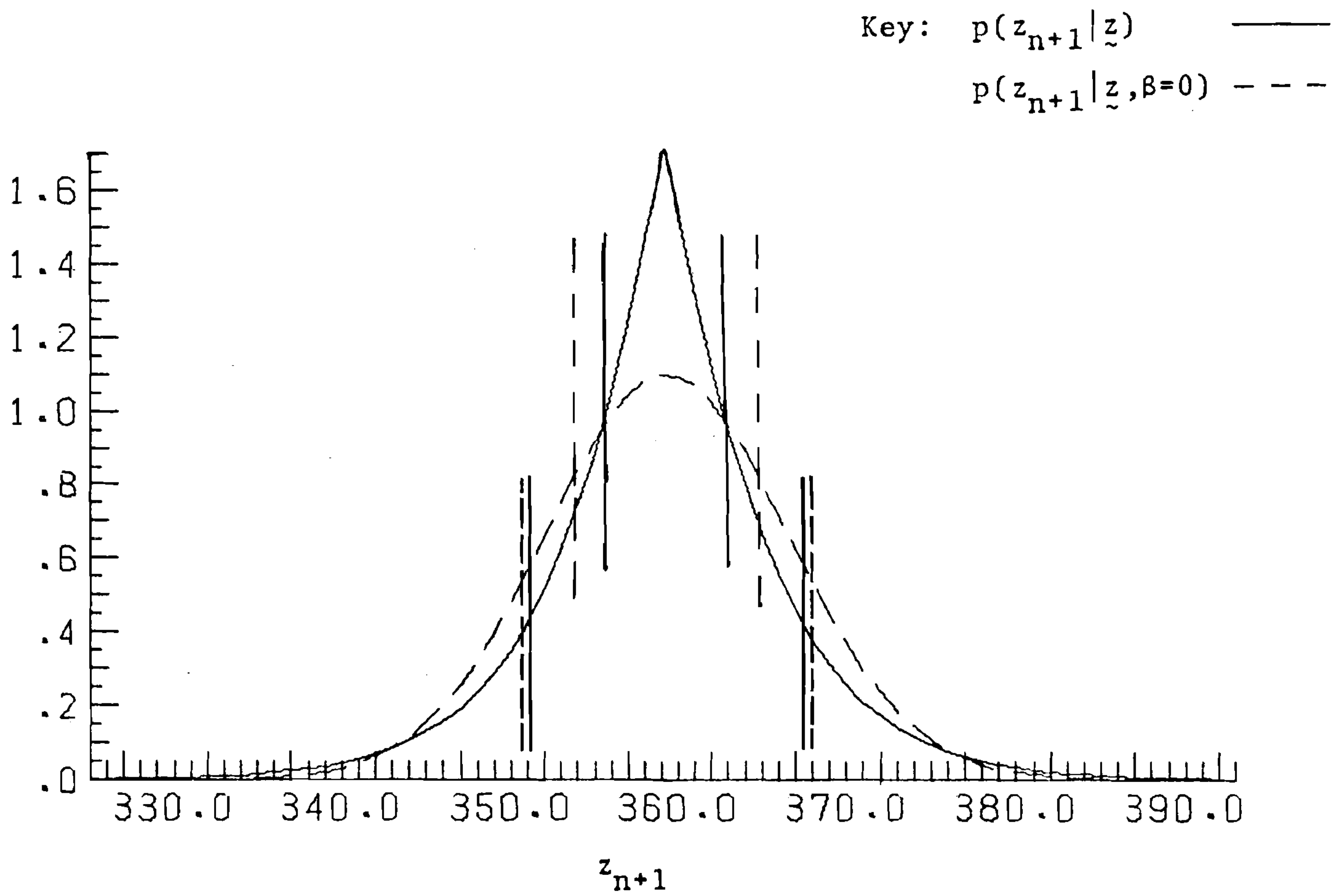

Figure 5. Posterior one step ahead predictive distribution $p(z_{n+1}/\underset{\sim}{z})$ compared with $p(z_{n+1}/\underset{\sim}{z},\beta = 0)$, the posterior one step ahead predictive distribution assuming normal distributed shocks, and their 50% (90%) probability limits.

References

[1] Bachelier, L., "The Theory of Speculation," (1900), reprinted in Cootner, P.H. (ed.), "The Random Character of Stock Market Prices," M.I.T. Press, Massachusetts, (1964).

[2] Box, G.E.P., "A Note on Regions for Test of Kurtosis," Biometrika, 40, 465-469, (1953).

[3] Box, G.E.P. and Jenkins, G.M., "Time Series Analysis, Forecasting and Control," Holden-Day; San Francisco, (1970).

[4] Box, G.E.P. and Tiao, G.C., "A Further Look at Robustness via Bayes theorem," Biometrika, 49, 419-433, (1962).

[5] Box, G.E.P. and Tiao, G.C., "Bayesian Inference in Statistical Analysis," Addison-Wesley; Reading, Massachusetts, (1973).

[6] Fama, E.F., "The Behavior of Stock Market Prices," Journal of Business, 38, 34-105, (1965).

[7] Hsu, D., Miller R.B., and Wichern D.W., "On the Paretian Stable Behavior of Stock Market Prices," JASA, 69, 108-113, (1974).

[8] Kendall, M.G., "The Analysis of Economic Time Series, Part I: Prices," JRSS, Series A, 116, 11-34, (1953).

[9] Ledolter, J., "Topics in Time Series Analysis," Ph.D. thesis, Department of Statistics, University of Wisconsin-Madison, (1975).

[10] Mandelbrot, B., "The Variation of Certain Speculative Prices," Journal of Business, 36, 394-419, (1963).

[11] Mandelbrot, B., "The Variation of Some Other Speculative Prices," Journal of Business, 40, 393-413, (1967).

[12] Mandelbrot, B. and Taylor, H.M., "On the Distribution of Stock Price Difference," Operations Research, 15, 1057-1062, (1967).

[13] Miller, R.B., Wichern, D.W. and Hsu, D., "Changes of Variance in Time Series Models," Technical Report No. 284, Department of Statistics, (1971).

[14] Praetz, P.D., "The Distribution of Share Price Changes," Journal of Business, 45, 49-55, (1972).

[15] Press, S.J., "A Compound Events Model for Security Prices," Journal of Business, 40, 317-335, (1967).

[16] Wold, H.O., "A Study in the Analysis of Stationary Time Series," (1938), Almquist and Wicksell; Uppsala, 2nd ed. (1954).