# A CLASS OF MODELS FOR NON-NORMAL TIME SERIES

BY G. J. JANACEK AND A. L. SWIFT

*University of East Anglia*

*First version received May 1988*

**Abstract**. We propose a model for non-normal times series in which we regard the series as an instantaneous transformation of an underlying 'generating' series which is normal. We describe a procedure which simultaneously estimates both the transformation to normality and the time series structure of the underlying series. This model has several advantages; in particular several alternative types of forecast can easily be calculated. The relative merits of these forecasts are considered.

**Keywords**. Non-normal; ARIMA; Hermite; generalized beta.

## 1. INTRODUCTION

The techniques available for estimating the structure of stationary Gaussian time series are well developed. However, major difficulties can arise in the modelling process for a non-Gaussian series.

This has prompted some interest in modelling non-normal series and testing for normality (Epps, 1987). The approaches available depended to some extent on the departures from the normal assumptions that are suspected by the investigator. Thus Lawrance and Lewis (1985) devise a model structure of the autoregressive moving-average (ARMA) type with a non-Gaussian marginal distribution. Martin (1981, 1983) gives a 'robust' approach which seems most plausible when dealing with contaminated data, while Hopwood *et al.* (1982) and others have suggested transformations of the Box–Cox type as an integral part of the estimation procedure. We note in passing that transformations have long been a part of the Box–Jenkins methodology (Box and Jenkins, 1976) but are usually invoked to stabilize variances or ensure stationarity rather than to cope with non-normality. A recent interesting paper by Kitagawa (1987) employs the marginal distribution in an explicit way via the Chapman–Kolmogorov equations used to derive the Kalman filter.

Our view is that there are considerable advantages in regarding a non-Gaussian time series as an instantaneous transformation of a Gaussian series which can be handled by standard 'Box–Jenkins' or equivalent techniques. It also seems to us to be an advantage to take into account explicitly any information available about the marginal distribution of the series. To this end we have devised a procedure for simultaneously estimating both the transformation required and the time series structure of the underlying Gaussian series. This model also ensures that we are dealing with a linear

0143-9782/90/01 019–13 $02.50/0 

normal process. A series with an arbitrary marginal need not be linear and in consequence standard methods may not work. A major advantage of this approach is that we can then use the techniques and the large amounts of software already available for the estimation and forecasting of normal models.

## 2. TRANSFORMATIONS

Suppose that we have a stationary time series $\{Y_t\}$ where the marginal density has a continuous probability density function (p.d.f) $f(y)$. We propose a model where $\{Y_t\}$ is generated by a non-linear instantaneous transformation $T(z)$ of an underlying Gaussian series $\{Z_t\}$. We shall always assume that the series $\{Z_t\}$ has a standard normal marginal distribution and that all joint distributions are normal. We know from standard theory that for any $f(y)$ and corresponding cumulative distribution function (c.d.f.) $F(y)$, $F(Y)$ is uniform on (0, 1). In addition, $\Phi(z)$, where $\Phi$ is the standard normal c.d.f., is uniform on the same interval. Thus for any time period $t$ we can identify

$$F(Y_t) = \Phi(Z_t)$$

which gives

$$Y_t = F^{-1}\{\Phi(Z_t)\} = T(Z_t),$$

and our non-linear transformation $T$ is identified. In this way we regard our original series $\{Y_t\}$ as a transform of an underlying (normal) series $\{Z_t\}$. A schematic representation is given in Figure 1. For most distributions of interest this transformation and its inverse are reasonably well behaved. Note that, whilst for any $\{Y_t\}$ there is an inverse transformation to a series with normal *marginals*, our *model assumption* is that the $\{Z_t\}$ series is also Gaussian.

Before looking at specific cases we investigate the relationship between the correlation structure of the $\{Y_t\}$ and $\{Z_t\}$ processes. This can be done using the work of Granger and Newbold (1976), and we shall use their techniques extensively in what follows. The key idea is to view the transformation as an

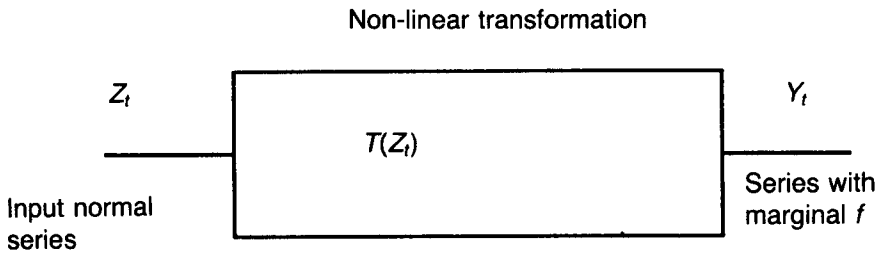


FIGURE 1.

orthogonal expansion in terms of hermite polynomials, i.e.

$$T(x) = \sum_{j=0}^{\infty} \alpha_j H_j(x) \tag{1}$$

where the $\alpha_n$ are constants and $H_n(x)$ is the Hermite polynomial of order $n$. We use the definition

$$H_n(x) = \frac{(-1)^n \varphi^{(n)}(x)}{\varphi(x)} \tag{2}$$

where $\varphi(x)$ is the p.d.f. of the standard normal distribution. Thus

$$H_0(x) = 1,\ H_1(x) = x,\ H_2(x) = x^2 - 1,\ H_3(x) = x^3 - 3x, \ldots .$$

In statistical terms the Hermite polynomials are orthogonal with respect to the standard normal p.d.f., i.e. when $U$ is standard normal

$$E\{H_n(U)H_m(U)\} = \begin{cases} 0 & (n \neq m) \\ n! & (n = m). \end{cases} \tag{3}$$

In addition

$$E\{H_n(U)\} = 0 \qquad \text{for all } n \neq 0,$$

and if $U$ and $V$ are bivariate normal with zero means, unit variances and correlation $\rho$, then

$$E\{H_n(U) \mid V = v\} = \rho^n H_n(v) \tag{4}$$

while

$$E\{H_n(U)H_m(V)\} = \begin{cases} 0 & (n \neq m) \\ \rho^n n! & (n = m) \end{cases} \tag{5}$$

Details of the Hermite polymials can be found in Erdelyi *et al.* (1953) or Granger and Newbold (1976) whose notation is used above. Definitions vary, but we feel that this notation is the natural one for use with expectations.

If a function $T(x)$ has a Hermite expansion

$$T(x) = \sum_{j=0}^{\infty} \alpha_j\ H_j(x),$$

then using orthogonality (see (3)) we can deduce that

$$\int_{-\infty}^{+\infty} \phi(x) T(x)\ H_j(x)\ dx = \alpha_j\ j!$$

Therefore, given $T(x)$, we can evaluate the integral above using a suitable numerical technique (e.g. Simpson's rule). This avoids the numerical problems of using the derivative form given by Granger and Newbold. Note that the recurrence

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$$

can be used to improve accuracy.

From (1) and the orthogonality relation (5) we find for our original series

$$E(Y_t) = E\{T(Z_t)\} = E\{\Sigma\alpha_j H_j(Z_t)\} = \alpha_0.$$

The autocovariance of the original $\{Y_t\}$ series at lag $k$ is

$$\begin{aligned} R_k &= E(Y_t Y_{t+k}) \\ &= E\{\Sigma\alpha_n H_n(Z_t)\Sigma\alpha_m H_m(Z_{t+k})\} \\ &= \Sigma\alpha_j{}^2 j!\rho_k{}^j \end{aligned} \tag{6}$$

while

$$E(Z_t Y_{t-k}) = \alpha_1\rho_k \tag{7}$$

where $\rho_k$ is the autocorrelation (equal to the autocovariance here) at lag $k$ of the normal series $\{Z_t\}$.

In fact (6) gives a one-to-one relationship between the autocovariances of the $\{Y_t\}$ series and those of the $\{Z_t\}$ which Sondhi (1983) uses as the basis of a simulation procedure (see also Janacek and Swift, 1987). For simulation purposes we regard $\{Z_t\}$ as the generating process and aim to tailor the autocorrelations of this series to achieve a $\{Y_t\}$ process with a desired autocorrelation structure. Suppose that we would like a realization of $\{Y_t\}$ where $\{Y_t\}$ has a specified marginal distribution, say with p.d.f. $f(y)$ and a specified autocovariance structure $\{R(k)\}$. We can deduce the required correlation structure of the normal process $\{Z_t\}$ transformation from (6) by solving for $\rho_k$ ($k = 1, 2, \ldots$) since the form of the non-linear transformation $T$ and hence the $\alpha_j$ are known. Given the correlations of the $\{Z_t\}$, $\{\rho_k\}$, we can then devise a linear filter to generate such a $\{Z_t\}$ from normally distributed white noise. The outline of the method is shown schematically in Figure 2. An implementation of this is described by Janacek and Swift (1987).

If $f(y)$ is known, we can write the likelihood of the $\{Y_t\}$ series modelling in this way as the likelihood of the Gaussian series $\{Z_t\}$ times the Jacobian of the transformation $Y_t = T(Z_t) = F^{-1}\{\Phi(Z_t)\}$ where the Jacobian is just

$$J = \Pi \frac{f(Y_t)}{\varphi(Z_t)}$$

The log likelihood is therefore

$$L(\Theta) = L_n(\Theta) + \log J$$

where $L_n(\Theta)$ is the log likelihood of the normal series and $\Theta$ is a vector of parameters.

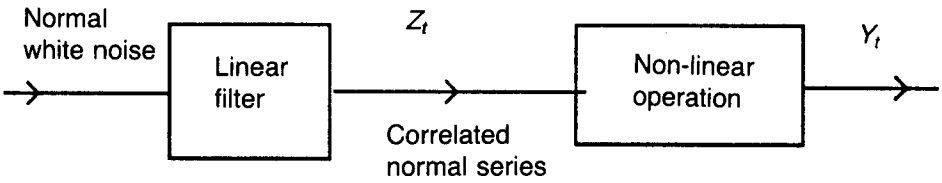


FIGURE 2.

We can then estimate the parameters of the normal $\{Z_t\}$ process using standard techniques. In what follows we assume that an ARMA model is appropriate.

## 3. FORECASTING

To investigate the role of the underlying normal process we look at the problem of forecasting following the lines of Granger and Newbold (1976) who extended the approach of Barrett and Lampard (1955). In what follows, 'optimal' will mean in the minimum mean square error (MSE) sense.

Suppose that $f_{n,h}$ is the optimum $h$-step-ahead predictor of $Z_t$ at time $n$. Granger and Newbold assume that this is normally distributed and standardize it. We have no need to make such an assumption as our underlying series is assumed to be Gaussian with zero mean and unit variance. Four possible predictors for $Y_{h+n}$ at time $n$ were suggested by Granger and Newbold. These can be listed as follows.

Forecast 1: the optimal predictor $g^1_{n,h}$.
Forecast 2: the 'naive' forecast $g^2_{n,h} = T(f_{n,h})$.
Forecast 3: $g^3_{n,h}$ the optimal forecast linear in $Z_t$.
Forecast 4: $g^4_{n,h}$ the optimal forecast linear in $\{Y_t\}$.

The second forecast $g^2_n = T(f_{n,h})$ is just a transformation of the optimal linear predictor of $Z_{n+h}$. As such it is intuitively attractive and has been referred to by Granger and Newbold as the 'naive' predictor. The third forecast $g^3_{n,h}$, is the one which is linear in $Z_t$, i.e. is of the form $\Sigma_{j=0}^{\infty} c_j Z_{n-j}$ with the coefficients $c_j$ chosen to minimize the MSE. As

$$E(Z_t Y_{t-k}) = \alpha_1 \rho_k,$$

it is possible to show (Whittle, 1983, section 3.7) that

$$g^3_{n,h} = \alpha_0 + \alpha_1 f_{n,h}$$

which only requires the first two terms of the Hermite expansion of $T(f_{n,h})$. The $h$-step-ahead MSE $V^i(h)$ for each forecast type $i$ ($i = 1, 2, 3$) can be calculated and is set out below:

$$V^1(h) = \sum_{j=1}^{M} \alpha_j{}^2 j!(1 - A^{2j})$$

$$V^2(h) = V^1(h) + \sum_{j=0}^{M} A^{2j}(j!) - 1 \left\{ \sum_{k=1}^{(M-j)/2} \alpha_{j+2k} \frac{(j+2k)!}{k!}(-0.5B^2)^k \right\}^2$$

$$V^3(h) = \sum_{j=2}^{M} \alpha_j{}^2 j! + \alpha_1{}^2 S^2(h)$$

where $A = \{1 - S^2(h)\}^{1/2}$, $B^2 = 1 - A^2$ and $S^2(h)$ is the variance of the conditional distribution of $Z_{n+h}$ at time $n$ ($Z_{n+h}$ is normal with mean $f_{n,h}$). In

common with Granger and Newbold we shall usually assume that $T(x) = \Sigma_{j=0}^{M} \alpha_j H_j(x)$ for suitable $M$.

For our models, the optimal predictor $g_{n,h}^{1}$ cannot be calculated directly as its structure is unknown.

The last of the forecast functions $g_{n,h}^{4}$ is more feasible. We could perhaps use the least-squares method on the non-normal series and for long series hope for asymptotic convergence of estimators or use a spectral factorization technique (e.g. Wilson, 1969; Bhansali, 1974). There are three problems with this approach. Firstly, we must expect that for these methods to work we need a reasonable amount of data. Secondly, we have no idea of the form of the distribution of the predictors and will in consequence have difficulty in producing prediction intervals. Thirdly, there is the problem of assessing the adequacy of the fit of any model used in prediction, which is a real problem for non-normal series.

Forecasts 2 and 3 are both functions of $f_{n,h}$ which is easily calculated if $Z_t$ is a normal ARMA process. We would expect the naive forecast to be preferable as it makes use of the entire non-linear transformation. This is borne out as it can be shown that it is the median of the conditional distribution of $Y_{n+h}$ at time $n$ (Granger and Newbold, 1976) and, as such, is optimal if loss is proportional to the absolute value of the forecast error. Unfortunately, however, we have been unable to show analytically that the MSE of the naive forecast is less that that of the third forecast.

We can make the following general observations.

(1) It is clear that knowledge of the $\alpha$ coefficients is essential to forecasting. Happily they are easy enough to compute.

(2) We can compute the expected MSEs to give some guide as to the lower bound of MSE that we might expect.

(3) When using forecasts 2 and 3 prediction limits for the $Y_t$ forecasts are just the appropriate function of the corresponding prediction (optimal) limits for the $Z_t$.

(4) What is more *all* the usual techniques in the catalogue of those available in the study of normal series can be used. For example, there is no problem about diagnostics since we can study both the model fitted to the $Z_t$ series and the transformation to the observed $\{Y_t\}$.

Thus if we assume that the marginal distribution is known we can produce forecasts and prediction intervals based on the underlying normal series. Naturally, we do not know the form of the marginal but we would claim that in common with most statistical practices we can determine a broad family of distributions and reduce the problem to one of estimation.

## 4. PROCEDURE AND EXAMPLES

We have not yet considered how the parameters of our model are to be estimated from raw data. The parameters of our model are of two types: (1)

those of the class of marginal distributions assumed and (2) those of the underlying ARMA model. For any set of model parameters we can calculate the likelihood as follows.

As described in Section 2, for a given marginal distribution and parameters the transformation $y = T(z)$ is implicit and so the underlying series $Z_t = T^{-1}\{Y_t\}$ can be calculated. The likelihood of the parameters of the underlying ARMA model can be obtained by using existing techniques. We used the efficient state space algorithm given by Gardner *et al.* (1980). As described in Section 2, the Jacobian of the transformation is easily calculated and when added to the ARMA likelihood gives the likelihood that the original data will assume our model structure.

We propose a two-stage estimation procedure, each stage being repeated alternately until the parameter estimates converge. At each stage we calculate the maximum likelihood estimates (MLEs) of one type of parameter (marginal or ARMA) conditional on fixed values of the other type. We retain each set of MLEs as the fixed parameters in the next stage to be executed.

In stage 1 we estimate the parameters of our class of marginal distributions. These parameters then provide the fixed parameters for the succeeding stage 2. In stage 2 the parameters of the class of marginal probability distributions are fixed and the ARMA parameters are estimated. These ARMA estimates provide the fixed parameters for the succeeding stage 1. Note that the ARMA noise variance is not a 'free' parameter here as the underlying series, by assumption, has unit variance. As the algorithum of Gardner *et al.* calculates the terms of the log likelihood which are not dependent on the noise variance, it is easily adapted to calculate the likelihood of a model with variance constrained to unity.

We have found ARMA(2, 2) or ARMA(3, 3) models quite adequate for demonstrating the effectiveness of our model, as larger ARMA orders usually give only small increases in the log likelihood. We used a NAG minimization routine E04JBF to maximize the likelihood function in both stages. This calculates numerical derivatives from user-supplied differencing intervals.

We suggest starting the two-stage procedure by executing stage 1 with the ARMA parameters fixed to zero. This gives the MLEs of the parameters of the original distribution under the assumption that the data are independent. The resulting log likelihood then provides a useful measure for comparison with our final maximum log likelihood. Another such measure is the maximum log likelihood obtained by fitting a linear (Gaussian) model to the data.

Akaike (1974) (see also Kitagawa 1987, rejoinder) justified the use of the maximum log likelihood as a means of comparison between models from different families. His automatic information criterion (AIC) of model fitting is the maximum log likelihood less a correction of the number of parameters in the model. For non-normal data the improvement in log likelihood obtained by modelling both the autocorrelations and the marginal p.d.f. is so pronounced that such a correction is of minor importance.

This procedure has been applied to a variety of simulated data sets to give very promising results. The results obtained from two such sets are described below.

We simulated 200 data items using the underlying ARMA model

$$Z_t + 0.8Z_{t-1} + 0.2Z_{t-2} = \varepsilon_t + 0.3\varepsilon_t.$$

where the noise $\varepsilon_t$ was normal with variance 0.77294 to ensure that $Z_t$ had unit variance. This data set $\{Z_t\}$ was duly transformed to a beta(2, 6) series, using the transformation $T(z)$ described earlier.

Similarly, 200 observations of an exponentially distributed series with parameter 0.5 (and therefore a mean of 2) were generated from the underlying ARMA model

$$Z_t - 0.3y_{t-1} + 0.2y_{t-2} = \varepsilon_t - 0.9\varepsilon_{t-1} + 0.18\varepsilon_{t-2}$$

where the variance of $\varepsilon_t$ was 0.71068. The autocorrelation functions histograms of these series are shown in Figures 3–6. The histogram of the exponential series is clearly non-normal, but the beta series could possibly be mistaken for normal.

Table I shown the log likelihoods of our model and of independent data. A beta marginal distribution and an exponential marginal distribution were assumed for the simulated beta series and exponential series respectively. The log likelihood of the Gaussian linear model is also shown. As we would expect, the log likelihood of our fitted model is much greater than both of these. However, our objective in using our model, which reflects both the autocorrelation structure and the probability distribution of the data, is not merely to obtain a better 'fit', but also to be able to forecast with greater precision. Table II shows the theoretical unconditional MSEs of the four types of forecast for one-to-ten steps ahead calculated as in Section 3, assuming that our fitted parameters give the true model.

| Midpoint | Count | |
|---|---|---|
| 0.05 | 15 | *************** |
| 0.10 | 23 | *********************** |
| 0.15 | 27 | *************************** |
| 0.20 | 33 | ********************************* |
| 0.25 | 27 | *************************** |
| 0.30 | 23 | *********************** |
| 0.35 | 20 | ******************** |
| 0.40 | 11 | *********** |
| 0.45 | 7 | ******* |
| 0.50 | 8 | ******** |
| 0.55 | 3 | *** |
| 0.60 | 1 | * |
| 0.65 | 0 | |
| 0.70 | 2 | ** |

FIGURE 3. Histogram of the beta (2, 6) series.

```
                 -1.0  -0.8  -0.6  -0.4  -0.2   0.0   0.2   0.4   0.6   0.8   1.0
                  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 1    -0.445                  XXXXXXXXXXXX
 2     0.190                        XXXXXX
 3    -0.091                      XXX
 4    -0.028                       XX
 5     0.002                        X
 6    -0.060                       XX
 7    -0.019                        X
 8     0.097                       XXX
 9    -0.039                       XX
10     0.008                        X
11     0.051                        XX
12    -0.061                      XXX
13     0.047                        XX
14    -0.071                      XXX
15     0.017                        X
16     0.004                        X
17    -0.080                      XXX
18     0.066                        XXX
19    -0.055                       XX
```

FIGURE 4. Autocorrelations of the beta series.

```
Midpoint      Count
       0         50  **************************************************
       1         57  *************************
       2         29  ***************
       3         22  ***********
       4         21  ***********
       5          7  ****
       6          7  ****
       7          4  **
       8          0
       9          2  *
      10          0
      11          1  *
```

FIGURE 5. Histogram of the exponential series.

For both simulated series at least one of forecast 2 (naive forecast) or forecast 3 gave MSEs close to forecast 1 (the optimal forecast). The MSEs of forecast 4 are larger, indicating that, for non-normal data with the autocorrelation structure implied by our fitted models, a linear model (not necessarily normal) will not forecast as well as ours. It should also be emphasized that the MSEs of forecast 4 are merely lower bounds and cannot necessarily by attained by a linear model.

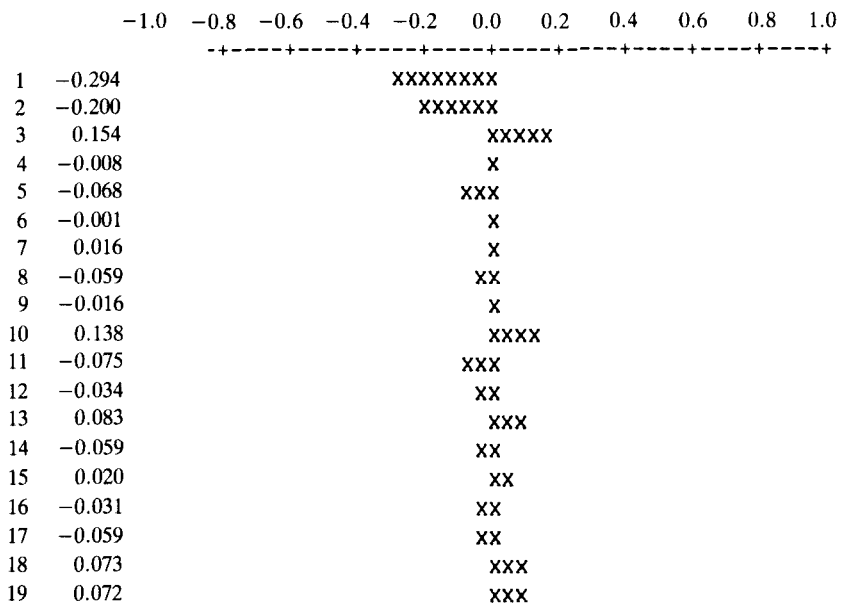


FIGURE 6. Autocorrelations of the exponential series.

TABLE I

| | Beta(2 6) | exp(1/2) |
|---|---|---|
| Final log likelihood (our model) | 162.03 | −306.82 |
| Log likelihood (independent observations) | 136.57 | −344.66 |
| Log likelihood (linear Gaussian model) | 146.74 | −401.13 |
| Estimate of prediction error variance of linear Gaussian model | 0.0135 | 3.228 |
| Final marginal parameters | 2.471<br>7.5389 | 0.495 |

In these and other simulations we have performed it is not obvious that either of the two forecasts considered (forecasts 2 and 3) consistently perform better than the other in MSE terms, and so it seems reasonable to assume that this depends strongly on the parameters of the model.

As an initial study of the relative merits of the four forecast types we calculated ratios of their one-step MSEs for an underlying AR(1) process and an exponential marginal distribution. As it seemed likely that the degree of autocorrelation in the underlying model would influence the MSEs, we performed these calculations for a range of autoregressive (AR) parameters (see table III). Choice of the exponential parameter $\lambda$ does not influence the

TABLE II

FORECASTING MEAN SQUARE ERRORS OF THE BETA (2, 6) AND EXPONENTIAL (1/2) MODELS ONE TO TEN

| $h$ | $V^1(h)$ | $V^2(h)$ | $V^3(h)$ | $V^4 9h)$ |
|---|---|---|---|---|
| 1 | 0.0131546 | 0.0131582 | 0.0131803 | 0.0134029 |
| 2 | 0.0161437 | 0.0161453 | 0.0161447 | 0.0161568 |
| 3 | 0.0166769 | 0.0166775 | 0.0166770 | 0.0166857 |
| 4 | 0.0168485 | 0.0168487 | 0.0168468 | 0.0168495 |
| 5 | 0.0168741 | 0.0168742 | 0.0168741 | 0.0168746 |
| 6 | 0.0168840 | 0.0168845 | 0.0168845 | 0.0168846 |
| 7 | 0.0168856 | 0.0168856 | 0.0168856 | 0.0168857 |
| 8 | 0.0168863 | 0.0168863 | 0.0168863 | 0.0168863 |
| 9 | 0.0168863 | 0.0168863 | 0.0168863 | 0.0168863 |
| 10 | 0.0168864 | 0.0168864 | 0.0168864 | 0.0168864 |
| *Exponential (1/2) model* | | | | |
| 1 | 3.010857 | 3.013533 | 3.076663 | 3.566214 |
| 2 | 3.875442 | 3.876931 | 3.877964 | 3.937892 |
| 3 | 4.049227 | 4.049500 | 4.049268 | 4.056077 |
| 4 | 4.073643 | 4.073649 | 4.073643 | 4.073810 |
| 5 | 4.074094 | 4.074094 | 4.074094 | 4.074095 |
| 6 | 4.074198 | 4.074098 | 4.094098 | 4.074099 |

TABLE III

RELATIVE MEAN SQUARES ERROR OF ONE STEP FORECASTS

| | AR coefficient | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.9 | −0.7 | −0.5 | −0.3 | −0.1 | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| $V^2/V^3$ | 0.6466 | 0.9285 | 0.9868 | 0.9990 | 1.0001 | 1.0001 | 0.9990 | 0.9868 | 0.9285 | 0.6466 |
| $V^2/V^4$ | 0.9996 | 0.9884 | 0.9886 | 0.9933 | 0.9989 | 0.9983 | 0.9774 | 0.9141 | 0.7892 | 0.6151 |
| $V^3/V^4$ | 1.5460 | 1.0645 | 1.0018 | 0.9943 | 0.9988 | 0.99820 | 0.9784 | 0.9263 | 0.8499 | 0.9514 |
| $V^2/V^1$ | 1.0004 | 1.0008 | 1.0009 | 1.0005 | 1.0001 | 1.0001 | 1.0001 | 1.0009 | 1.0008 | 1.0004 |
| $V^3/V^1$ | 1.5472 | 1.0779 | 1.0143 | 1.0016 | 1.0000 | 1.0000 | 1.0016 | 1.0143 | 1.0779 | 1.5472 |
| $V^4/V^1$ | 1.0008 | 1.0126 | 1.0124 | 1.0073 | 1.0012 | 1.0020 | 1.0237 | 1.0950 | 1.2682 | 1.6263 |

Exponential marginal distribution and AR model $y_t + Ay_t - 1 = \varepsilon_t$.

results, as it is easily shown that the Hermite coefficient of the implied non-linear transformation are identical except for a factor $\lambda^{-1}$, so that the forecast MSEs are the same up to a factor of $\lambda^{-2}$.

We can see that the MSE of forecast 2 (the naive forecast) is less than that of forecast 3 except when the AR parameter is near to zero. In fact, for highly correlated series (absolute value of AR parameter near unity) forecast 2 shows a marked improvement on forecast 3 in MSE terms. Forecast 2 always fares slightly better than the linear forecast (forecast 4) but this is accentuated for positive values of the AR parameter away from zero. The 'naive' forecast works well because the MSE of forecast 2 is close to that of the optimal forecast (forecast 1) for the whole range of AR parameters,

whereas forecast 3's MSE is more than 50% greater than the optimal for AR parameters close to unity. For AR parameters 0.7 and 0.9 the MSE of forecast 4 (obtained by fitting a linear model to the autocovariance structure implied by the model) is very much greater that the optimal MSE.

We conclude therefore that, even for this trivial example, the naive forecast (forecast 2) is the 'best' forecast (in the MSE sense) based on our model and gives a near-optimal MSE throughout the parameter space. The linear-model-based forecast 4 is consistently inferior to the naive forecast and is in fact only satisfactory for a limited range of AR parameter values.

## 5. CONCLUSIONS AND FUTURE DEVELOPMENTS

We have introduced a general model which reflects both the marginal probability distribution and the autocorrelation structure of a set of sequential data. This model is intuitively attractive and can be estimated by a slight adaptation of existing estimation techniques. The residuals from the model are asymptotically normally distributed, facilitating diagnostic checking. Forecasts from the model compare favourably in theoretical MSE terms with linear or normal linear models, and confidence limits can easily be calculated.

As the reader will have noticed, up to now we have tacitly assumed that the class of marginal distribution is known, except perhaps for some parameters. While this is often a reasonable assumption it would be much nicer to determine the marginal from the data. To this end we are currently considering developing the model and procedure for a general class of marginal distributions so that very little a prior knowledge of the distribution is required. One such possibility is the GB2 distribution discussed by Bookstaber and MacDonald (1987). This is a generalization beta distribution which contains a larger number of well-known distributions, such as the gamma, log normal and exponential, as special or limiting cases. Results have been promising although we have encountered some computational problems.

We could gauge the forecasting ability of our model by considering the theoretical MSE (forecast 2 or 3) relative to the lower bound on the MSE of a linear model fitted to the same autocorrelation structure (forecast 4). For a particular type of marginal distribution, our model is defined by the parameter values of this and the autocorrelations of the underlying Gaussian model (or equivalently the autocorrelations of the non-normal process). We intend to enlarge our study of the effect of these on the relative MSEs.

When using the Sondhi algorithm to generate data with a specified autocorrelation function and marginal probability density function we have occasionally had difficulties in fitting a linear model to the autocorrelations implied for the underlying series. As these autocorrelations are obtained by assuming that the underlying series is Gaussian we must deduce that certain combinations of marginal distribution and autocorrelation structure cannot be modelled as the transformation of a Gaussian process. There is scope for

further work to identify such combinations.

Our model regards the data as a particular instantaneous transformation of a linear Gaussian model. This implies the third and higher cross-moments of the non-normal series. The nature of these restrictions has not yet been investigated.

## ACKNOWLEDGEMENTS

We wish to acknowledge the support of the Science and Engineering Research Council under contract GRD/27271, which enabled the bulk of the software used to be written. It is also a pleasure to acknowledge the help and assistance of K. Lever, previously at the University of East Anglia and now at Brunel University. We owe a debt to an anonymous referee whose comments materially improved this paper.

## REFERENCES

AKAIKE, H. (1974) A new look at statistical model identification. *IEEE Trans. Autom. Control* 19, 716–22.

BARRETT, J. F. and LAMPARD, D. G. (1955) An expansion for some second order probability distributions and its application to noise problems. *IRE Trans. PGIT* 1, 10–15.

BHANSALI, R. J. (1974) Asymptotic properties of the Wiener–Kolmogorov predictor. I. *J. R. Statist. Soc. Ser. B* 36, 61–73.

BOOKSTABER, R. M. and MACDONALD, J. B. (1987) A general distribution for describing security price returns. *J. Business*, 60, 401–24.

BOX, G. E. P. and JENKINS, G. M. (1976) *Time Series Analysis, Forecasting and Control*. San Francisco; CA: Holden Day.

EPPS, R. W. (1987) Testing that a stationary time series is Gaussian. *Ann. Statist.* 15, 1683–98.

ERDERLYI, A. *et al.* (1953) *Higher Transcendental Functions*. New York: McGraw-Hill.

GARDNER, G., HARVEY, A. C. and PHILLIPS, G. D. A. (1980) An algorithum for exact maximum likelihood estimation of autoregressive moving average models by means of Kalman filtering. *Appl. Statist.* 29, 311–22.

GRANGER, C. W. J. and NEWBOLD, P. (1976) Forecasting transformed series. *J. R. Statist. Soc. Ser. B* 38 (2), 189–203.

HOPWOOD, W. S., MACKEOWN, J. C. and NEWBOLD, P. (1982) Time series forecasting models involving power transformations. *J. Forecast.* 3, 57–61.

JANACEK, G. J. and SWIFT, A. L. (1987) Random number generator user manual. Internal Report GR/D272211, School of Mathematics, University of East Anglia, Norwich.

KITAGAWA, G. (1987) Non-Gaussian state-space modelling of non-stationary time series. *J. Am. Statist. Ass.* 82, 1032–41.

LAWRANCE, A. J. and LEWIS, P. A. W (1985) Modelling and residual analysis of nonlinear autogressive time series in exponential variables. *J. R. Statist. Soc. Ser. B* 47, 165–202.

MARTIN, R. D. (1981) Robust methods for time series. *Applied Time Series II* (ed. D. F. Findley). New York: Academic press.

—— (1983) Robust resistant spectral analysis. *Handbook of Statistics 3* (ed. P. R. Krishnaiah). Amsterdam: North-Holland.

SONDHI, M. M. (1983) random processes with specified spectral density and first order probability density. *Bell Syst. Tech.* 62, 679–701.

WHITTLE, P. (1983) *Prediction and Regulation*. London: EUP (reprinted).

WILSON, G. T. (1969) Factorization of the generating function of a pure moving average process. *SIAM J. Numer. Anal.* 6, 1–6.