AN INNOVATION APPROACH TO
NON-GAUSSIAN TIME SERIES ANALYSIS

TOHRU OZAKI\textsuperscript{1} AND
MITSUNORI INO,\textsuperscript{2} Institute of Statistical Mathematics, Tokyo

Abstract

The paper shows that the use of both types of random noise, white noise and Poisson noise, can be justified when using an innovations approach. The historical background for this is sketched, and then several methods of whitening dependent time series are outlined, including a mixture of Gaussian white noise and a compound Poisson process: this appears as a natural extension of the Gaussian white noise model for the prediction errors of a non-Gaussian time series. A statistical method for the identification of non-linear time series models with noise made up of a mixture of Gaussian white noise and a compound Poisson noise is presented. The method is applied to financial time series data (dollar–yen exchange rate data), and illustrated via six models.

Keywords: Innovations; non-linear modeling; non-Gaussian time series analysis; diffusion process; point process; white noise; Poisson noise; Lévy process; exchange rate data

AMS 2000 Subject Classification: Primary 62M10
Secondary 60G35

1. Introduction

Statistical analysis of dependent time series sequences has been one of the most important issues in many scientific fields such as geophysics, biosciences, and economics. A significant achievement in the analysis of dependent time series sequences has been attained by taking advantage of mathematical time series models in the last few decades. Natural mathematical models for such time series sequences are stochastic process models which are classified into two types; one is for continuous type processes such as diffusion processes and others are for jump type processes such as point processes, marked point processes etc. It was the first author’s fortune to meet David Vere-Jones in the mid 1970s before starting work on the estimation problem of continuous time non-Gaussian diffusion process models. At that time, the attention of time series analysts had been directed from the Gaussian setting to non-linear and non-Gaussian time series analysis. A point process is clearly a typical non-Gaussian process; another example of a non-Gaussian process is any diffusion process defined by a non-linear stochastic differential equation model. These two types of non-Gaussian processes seem to have nothing in common: amongst point processes, the Poisson process plays an essential role, while for non-Gaussian diffusion processes, Gaussian white noise, an increment

Received August 2000.
\textsuperscript{1} Postal address: Institute of Statistical Mathematics, 4-6-7 Minami-Azabu Minato-Ku, Tokyo 106-8569, Japan. Email: ozaki@ism.ac.jp
\textsuperscript{2} Email: iino@ypri.co.jp
of Brownian motion, plays an essential role. However, after decades of studying non-Gaussian time series analysis, it turns out that both Gaussian white noise and Poisson noise are very useful in the statistical analysis of financial time series data which, as is widely known, are not adequately described by Gaussian distributions alone.

We show below that using the two types of random noise can be justified from the standpoint of the innovation approach, whose historical background is briefly explained in Section 2. We consider several methods of whitening dependent time series in Sections 3 and 4: a mixture of Gaussian white noise and a compound Poisson process appears as a natural extension of the Gaussian white noise model for the prediction errors of a non-Gaussian time series. A statistical method for the identification of non-linear time series models with noise made up of a mixture of Gaussian white noise and a compound Poisson noise is presented. The application of the model to financial time series data is presented in Section 5, together with numerical results, followed by a brief concluding remark in Section 5.

2. Innovation approach

Theories of the innovation approach to time series analysis have been developed by both mathematicians and statisticians. Historically they go back to Wold (1938), Kolmogorov (1941) and Wiener (1942) who developed the theories in close relation with the idea of predicting dynamic phenomena from time series observations. A typical example is Wiener's theory of predicting the trajectory of flying objects or missiles for the control of anti-aircraft guns. Wiener considered the prediction error as a source of information for improving the prediction of future phenomena and gave 'prediction error' the more positive name innovation.

Wiener noticed the limit of the mathematical approach and, though a mathematician, realized the need for statistics to obtain solutions in real applications. In his celebrated book (Wiener, 1948) he correctly pointed out that the contribution of statisticians is needed to obtain a practical and useful solution for a real prediction problem. His first idea for finding a solution to the prediction problem was to write down the prediction scheme using the spectral function of the time series and to estimate the spectrum from the observed time series so that it led to the best predictor of the future series. In line with his frequency domain approach, spectral estimation methods were developed by many statisticians from the 1950s onward (Tukey, Bartlett, Parzen, Hannan, Priestley, Akaike).

It was Akaike (1968) who changed the trend of Wiener's frequency domain approach into the time domain approach for predictions in statistical time series analysis, although Whittle (1963) developed a fair body of work throughout the 1950s on autoregressive (AR) predictors after Wold (1938). Akaike pointed out that the spectrum estimation problem is solved by finding a best AR predictor using a statistical criterion FPE (final prediction error), which idea later culminated in Akaike's information criterion (AIC) as a more general statistical model selection criterion, including ARMA models and state space representation models, using Boltzmann's probabilistic interpretation of entropy (Akaike, 1977).

Thus linear time series models, including AR models and state space models, have been the major models used by applied scientists in whitening or predicting time series. We will briefly see, in the next section, how these models have been developed and supported by both mathematicians and statisticians.
3. Whitening filters

3.1. Mathematicians’ solutions

Since the prediction error variance of stationary time series is given by the spectral density function of the process and the spectral density function of stationary processes is best approximated by a rational function, the natural solution for the mathematicians’ prediction problem is an ARMA model (Doob, 1953). Once a most appropriate ARMA model is obtained, the spectral density function of the process, prediction and whitening of the process are given by the model. Here, a practical problem is that slightly different orders and coefficients of ARMA models yield different predictors. Finding out how to estimate the model coefficients from finite set of observed time series data or finding out how to choose optimal model orders are not mathematicians’ major concern.

Wiener (1942) considered the mathematics of the prediction problem from time series observations contaminated by observation errors. By taking full advantage of the idea of ‘whitening’ in his innovation approach, he introduced the Wiener–Hopf equation; this provides us with a method, called the Wiener filter, of obtaining the best linear predictor from past and present observations contaminated by observation errors. Here the impulse response function and auto-covariance function play a key role. Wiener tried to estimate an impulse response function from observed time series that led to the optimal predictor. However a weak point was soon recognized by many engineers at the application stage of the Wiener filter. It turned out that using impulse response functions was not an efficient way of obtaining prediction errors. An ingenious solution was given by Kalman (1960), using a Markovian representation to describe the dynamics. Since the Kalman filter implements filtering and prediction in a recursive scheme, the methods became an indispensable technique in some applications such as satellite orbit prediction or space rocket control, where the dynamics of the object are fairly simple and known from physical theories. The help of statisticians was required later when people started applying the Kalman filter to more complex dynamic phenomena.

Soon after Kalman’s introduction of the recursive filter with a linear Markov state space representation several interesting works followed. First a recursive smoothing algorithm based on the Kalman filter was introduced by Bryson and Frazier (1963). Later a Bayesian interpretation of the Kalman filter was pointed out by Ho and Lee (1964) and also a Bayesian interpretation of the recursive smoother was given by Rauch (1963) and Meditch (1967). Use of the recursive Kalman filter for Bayesian modeling of time series was proposed much later (Harrison and Stevens, 1976).

3.2. Solutions found by statisticians

Whether we like it or not, statisticians should acknowledge that many important concepts, models and algorithms in the prediction problem are owed to mathematicians. On the other hand, statisticians could make claims of their own regarding solutions to the prediction problem that are practical and useful in real applications.

For example, different ARMA models yield different predictions, and we need statistical guidelines and criteria to lead us to the optimal order and efficient estimation of the coefficients of such models (Hannan, 1969; Akaike, 1973; Box and Jenkins, 1970).

Kalman’s state space models are not free from statistical problems when such models are applied to more complex phenomena, where the dynamics behind the time series
are not so obvious. In some cases, state dynamics prediction is inaccurate when the variance-covariance matrix of the state prediction errors cannot be given a priori; the matrix needs to be estimated from the observed time series by a statistical method (Astrom and Kallstrom, 1973; Mehra, 1971).

In both ARMA model identification and state space model identification, Akaike’s idea of entropy and AIC can provide us with a useful tool. The work of these statisticians has helped Wiener’s idea of innovation approach to find useful applications through ARMA models and linear state space models for prediction and control of time series in many applied sciences. Of all the statisticians working in this area, Box and Jenkins (1970) were foremost in promoting the use of whitening ideas, suggesting the use of linear ARMA models in many fields of application involving time series analysis. Here models in the time domain play an essential role in whitening the time series into prediction errors by subtracting from each observation, its past-dependent components. However, after decades of experience of applications of ARMA models and state space representation models in applied fields, it has been noticed that linear models are not sufficient for some prediction problems.

One driving force in the search of more general prediction models comes from mathematical considerations. Mathematics tell us that the linear predictor can be optimal only when the process is Gaussian. When the process is non-Gaussian, a better predictor may be given by a non-linear dynamic model (Masani and Wiener, 1959). Another driving force came from statistics. In some case studies, the prediction errors (residuals) of the fitted model do not behave like Gaussian white noise. Common patterns of discrepancy of the residuals from Gaussian white noise are classified into the following two types:

1. time-inhomogeneous prediction errors; and
2. time-homogeneous Gaussian errors mixed with a few outlying large errors.

These two typical patterns of the prediction errors have motivated both mathematicians and statisticians to study non-linear and/or non-Gaussian prediction models.

4. Whitening non-Gaussian time series

Non-linear time series analysis has attracted the attention of scientists since the late 1970s. Especially, chaos, a feature of non-linear time series models, attracted the interest of many scientists in the 1980s and 1990s. By extending the traditional idea of the whitening filter in time series analysis and by bringing chaos, non-linear time series modeling and stochastic differential equation models together, an innovation approach has been presented and has found important applications (see e.g. Ozaki, 1992a, b; Valdes et al., 1999). On the other hand, non-Gaussian time series analysis seems not to have received much attention inside the circle of scientists, in spite of its intrinsic importance in applied science and time series analysis.

If possible, researchers would like to obtain an optimal result from the most general model with a very powerful and flexible computational method. The reality is that such a situation is not as easy to attain as might be expected. General purpose numerical methods for a non-linear non-Gaussian filter are still in the early stages of development. It is often found that a general state space model identified by the maximum likelihood method with the numerical Monte Carlo filter is inferior to the linear model identified by the maximum likelihood method with the ordinary linear Kalman filter algorithm (Ikoma, 1999).
In this section we will focus upon the non-Gaussian nature of time series, and will introduce a compatible innovation approach to non-Gaussian time series analysis, which can be shared by both non-linear and non-Gaussian time series modelers.

4.1. The mathematics behind non-Gaussian time series analysis

Theories of non-Gaussian Markov processes were developed from the 1950s to 1970s by mathematicians. It is unfortunate that by the time linear statistical methods for prediction were used in practice and their limitation was recognized by practitioners, the honeymoon of mathematics and statistics of the Wold–Kolmogorov–Wiener era was already over. Several non-linear time series models have been introduced in statistical time series analysis independently of Markov diffusion process theory. It took some time to realize that the non-linear time series modeling in statistics and non-Gaussian Markov diffusion process modeling in mathematics are essentially the same.

Markov process theory tells us that under certain regularity conditions any continuous Markov process \( x(t) \) has a stochastic differential equation representation,

\[
dx = a(x) \, dt + b(x) \, dw(t).
\]

This implies that for a short time interval \([t, t + dt]\) the prediction \( x(t + dt) \) of any continuous Markov process \( x(t) \) is given approximately by \( x(t) + a(x(t)) \, dt \) and its prediction error variance is approximately \( b(x)^2 \, dt \). This explains that the residuals of ARMA models fitted to some time series data are time-inhomogeneous. Their variance is \( x(t) \)-dependent. If we introduce a variable transformation satisfying \( dx/b(x) = dy \), we have another stochastic differential equation representation for \( x(t) \), namely

\[
x(t) = h(y(t)) \quad \text{and} \quad dy = f(y) \, dt + dw(t),
\]

for some \( h(\cdot) \) and \( f(\cdot) \). This means that any continuous Markov process can be transformed into homogeneous Gaussian white noise \( dw(t) \) by the combination of a variable transformation \( h(\cdot) \) and the removal from the present \( y(t) \) of the past-dependent non-linear component \( f(y) \, dt \) as in the non-linear dynamic model \( dy = f(y) \, dt + dw(t) \).

**Example 4.1.** (Gamma-distributed process.) The process \( x(t) \) defined by the Fokker–Planck–Kolmogorov equation

\[
dx = \{ (\alpha - \frac{1}{2}) \beta - x \} \, dt + \sqrt{2x} \, dw(t)
\]

is known to be Gamma-distributed, i.e. it has the stationary density function

\[
p(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}
\]

This example gives us useful guidelines for the prediction of non-Gaussian time series as follows:

- The prediction error variance is \( x \)-dependent: i.e. residuals derived by fitting a time series model, whether the model is linear or non-linear, are time inhomogeneous.
- Although the process is non-Gaussian distributed, it is generated from Gaussian white noise by a non-linear dynamic model, combined with a variable transformation \( x = \frac{1}{\beta} (\beta y)^2 \). This implies the possibility of transforming the non-Gaussian
distributed time series \( x_t \) back by using the inverse transformation and a non-linear dynamic model
\[
dy = \left( \alpha - \frac{1}{2} \right) y + \frac{1}{2} y \, dt + dw(t).
\]

- The stochastic differential equation is linear when \( \alpha = \frac{1}{2} \). Then the process \( y(t) \) is Gaussian distributed. This means that sometimes a non-Gaussian process can be transformed into Gaussian white noise by a variable transformation combined with a linear model, and a non-linear dynamic model is not necessary.

**Example 4.2.** (Cauchy-distributed process.) Ozaki (1985) pointed out that the combination of the instantaneous variable transformation \( x(t) = \sinh \sqrt{2} y(t) \) and the non-linear dynamics,
\[
dy = -\sqrt{2} \alpha \tanh(\sqrt{2} y) \, dt + dw(t),
\]
leads to a non-Gaussian distributed diffusion process with a heavy-tailed distribution \( W(x) \), where
\[
W(x) = \frac{\Gamma(\alpha + 1)}{\Gamma(\frac{1}{2}) \Gamma(\alpha)} (1 + x^2)^{-\alpha + 1/2}
\]
and \( W(x) \) is Cauchy distribution when \( \alpha = \frac{1}{2} \). In other words, by applying the inverse transformation \( \sinh^{-1} \) to \( x_t \) and by whitening the transformed series \( y_t = \sinh^{-1} x_t \), the series \( x_t \) can be transformed into Gaussian white noise \( w_t \). Here the whitening of \( y_t \) is realized by removing the \( y_t\)-dependent part from \( y_{t+dt} \) to give
\[
w_{t+dt} = y_{t+dt} - y_t + \sqrt{2} \alpha \tanh(\sqrt{2} y_t) \, dt.
\]

In Section 5, as in Iino and Ozaki (2000), there are examples of the non-linear modeling of a financial time series, where each model shows a different way of whitening the time series of a heavy-tailed distribution.

**4.2. Outlying residuals and Lévy’s theorem**

We mentioned in Section 1 that one of the typical non-Gaussian patterns in the residuals derived from fitting linear time series models such as ARMA models or linear state space models consists of a few large outlying residuals mixed with other homogeneous more or less Gaussian looking residuals. This pattern of residuals, which is quite common in financial time series modeling, leads to a linear or non-linear dynamic model with heavy-tailed distributed driving noise. Dynamic models with general non-Gaussian distributed noise have been introduced by Bayesian statisticians mainly to cope with this kind of heavy-tailed distributed time series. From a mathematical point of view, however, the raison d'être for general non-Gaussian noise is not so obvious. Prediction theory tells us that the distribution of prediction errors should be in the family of infinitely divisible distributions (Feller, 1966). Paul Lévy, a French mathematician, was one of those who carried on investigating the mathematics of non-Gaussian prediction errors after Wiener. His theorem on prediction errors is intriguing. It says that under regularity conditions, prediction errors with bounded jumps decompose into the sum of Gaussian white noise and compensated Poisson processes (Lévy, 1956; Protter. 1990). These theorems seem to provide us with a theoretical explanation for the large outlying residuals, often seen in fitting linear prediction models, as well as hinting at a more computationally efficient estimation method for prediction.
4.3. Statistical solutions

As we saw in the previous section, many statisticians have presented solutions to the non-Gaussian prediction error case. For state space modeling, both system and observation noises are replaced by general non-Gaussian distributed white noise. For ARMA modeling, white noise having a general non-Gaussian distribution was considered and the maximum likelihood method was extended to these non-Gaussian cases. Several schemes for the calculation and maximization of the likelihood function of these non-Gaussian noise models have been presented by Bayesians.

Ozaki (1992a) pointed out the relevance of the work of Lévy (1956) on independent increment processes with jumps to non-Gaussian predictors in time series analysis. Later (Iino and Ozaki, 1999a), the idea was used in real applications, where a computationally simple and mathematically consistent method was presented and applied to financial time series analysis.

The basic idea of Iino and Ozaki (1999a) is to assume that the system noise and/or observation noise is a mixture of Gaussian white noise and compound Poisson noise. Here the size of the compound Poisson noise could be noticeably large, i.e. much larger than the twice the standard deviation of the associated Gaussian white noise. The state space representation and its filtering and prediction can be calculated in the same way as in the linear case in the following example.

**Example 4.3.** (State space model representation.) We describe various aspects based on the model for the underlying and observed processes $x_t$ and $z_t$,

$$x_t = F_t x_{t-1} + n_t \quad \text{and} \quad z_t = H x_t + m_t,$$

where

$$m_t \sim \begin{cases} N(0, r), & \text{when no pulse noise is present at } t, \\ N(0, R), & \text{when pulse noise is present at } t. \\ \end{cases} \quad n_t \sim \begin{cases} N(0, q), & \text{when no pulse noise is present at } t, \\ N(0, Q), & \text{when pulse noise is present at } t. \\ \end{cases}$$

Let $Q^{(0)}_t = q$ and $Q^{(1)}_t = Q$, and $R^{(0)}_t = r$ and $R^{(1)}_t = R$. On the basis of this model we give formulae for various purposes:

**Prediction formulae :**

$$x^{(i,j)}_{t/t-1} = F_t x^{(i,j)}_{t-1/t-1}$$ and

$$P^{(i)}_{t/t-1} = F_t P^{(i,j)}_{t-1/t-1} F_t^T + Q^{(i)}_t.$$

**Filtering :**

$$x^{(i,i)}_{t/t} = x^{(i,j)}_{t/t-1} + K^{(i,i)}_{t} (z_t - H x^{(i,i)}_{t/t-1}),$$

$$P^{(i,j)}_{t/t} = (I - K^{(i,j)}_{t} H) P^{(i,j)}_{t/t-1}$$ and

$$K^{(i,i)}_{t} = P^{(i,j)}_{t/t-1} H^T (H P^{(i,j)}_{t/t-1} H^T + R^{(j)}_t)^{-1}.$$

**Prediction error :**

$$v^{(i,j)}_{t/t-1} = z_t - H x^{(i,i)}_{t/t-1}.$$

**Prediction error variance :**

$$V^{(i,j)}_{t/t-1} = H P^{(i,j)}_{t/t-1} H^T + R^{(j)}_t.$$

The filtering and prediction of the model are specified in terms of the state variable $M_t$ defined by

$$M_t = \begin{cases} M^{(0,0)}_t, & \text{when no pulse in either } n_t \text{ or } m_t, \\ M^{(0,1)}_t, & \text{when no pulse in } n_t \text{ and a pulse in } m_t, \\ M^{(1,0)}_t, & \text{when a pulse in } n_t \text{ and no pulse in } m_t, \\ M^{(1,1)}_t, & \text{when pulses in both } n_t \text{ and } m_t. \\ \end{cases}$$
log-likelihood function:

\[-2 \log p(z_1, \ldots, z_N) \]
\[= \sum_{t=1}^{N} \left\{ \sum_{i=0}^{1} \sum_{j=0}^{1} I_{\{M_t = M_t^{(i,j)}\}} \log V_t^{(i,j)} \right\} + \sum_{t=1}^{N} \left\{ \sum_{i=0}^{1} \sum_{j=0}^{1} I_{\{M_t = M_t^{(i,j)}\}} \frac{(V_t^{(i,j)})^2}{V_t^{(i,j)}} \right\},\]

where \(I_{\{M_t = M_t^{(i,j)}\}} = 1\) when \(M_t = M_t^{(i,j)}\), and 0 otherwise.

In principle, by using the maximum likelihood method with this likelihood function, we can estimate parameters of the model and at the same time detect jumps in either or both of \(n_t\) and \(m_t\). The difficulty with this estimation and detection procedure is that it requires enormous computation since there are so many possible states for \(M_t\) for \(t = 1, \ldots, N\), namely \(4^N\) possible cases. Thus it is hardly a practical method if we have to maximize the likelihood function in terms of parameters for all these possible cases. However a simple and practical recursive algorithm can be given for the calculation and maximization of the likelihood (see the appendix for details). We show in the next section how the method works in practice for the analysis of non-Gaussian financial time series.

5. Application to financial time series

5.1. Volatility models

In financial modeling, one of the typical characteristics is that the prediction error variance is not constant. Here the estimation of the prediction error variance is an important issue for the risk assessment of investment. Several mathematical models for financial time series have been considered.

(i) Subordinate process model.

\[\begin{align*}
\text{d}S(t) &= \{a + bS(t)\kappa(t)\} \text{d}t + c\sqrt{\kappa(t)} \text{d}w_1(t), \\
\text{d}\log \kappa(t) &= \{\alpha \log \kappa(t) + \beta u(t)\} \text{d}t + \gamma \text{d}w_2(t).
\end{align*}\]

This model comes from the Gaussian process \(\text{d}S(\tau) = \{a + bS(\tau)\} \text{d}\tau + c \text{d}w_1(\tau)\). Here the operational time \(\tau\) is different from the calendar time \(t\). When we specify the relation between the operational time \(\tau\) and the calendar time \(t\) by \(\text{d}\tau(t) = \kappa(t) \text{d}t\), and if we specify the dynamics of \(\kappa(t)\) by \(\text{d}\log \kappa(t) = \{\alpha \log \kappa(t) + \beta u(t)\} \text{d}t + \gamma \text{d}w_2(t)\), we have the above two dimensional diffusion process model for \(S(t)\) (see Iino and Ozaki (1999b) for numerical studies).

(ii) Market microstructure model. Recently mathematical models based on the market microstructure have been proposed by several authors (O'Hara, 1995; Bouchard and Cont, 1998; Farmer, 1998). A generalized version of such models is presented by Iino and Ozaki (2000) as

\[\begin{align*}
\text{d}S(t) &= \lambda(t)\phi(t) + \sigma \lambda(t) \text{d}w_1(t), \\
\text{d}\phi(t) &= \{\alpha + \beta \phi(t)\} \text{d}t + \gamma \text{d}w_2(t), \\
\text{d}\log \lambda(t) &= \{\delta + \varepsilon \log \lambda(t)\} \text{d}t + \zeta \text{d}w_3(t).
\end{align*}\]

Here the price \(S(t)\) of financial asset is driven by excess demand \(\phi(t)\), and amplitude of price change depends on the liquidity \(1/\lambda(t)\) of the market. When the liquidity is high,
the market can absorb excess demand by small price changes. When the liquidity is low, one unit of excess demand causes large price changes. Iino and Ozaki (2000) showed that the model can be actually estimated from real financial time series data using state space representations.

(iii) Volatility model.

\[
\begin{align*}
\mathrm{d}S(t) &= \{\alpha + \beta S(t)\} \, \mathrm{d}t + \sigma(t) \, \mathrm{d}w_1(t), \\
\mathrm{d}\log \sigma^2(t) &= \{\gamma + \delta \log \sigma^2(t)\} \, \mathrm{d}t + \varepsilon \, \mathrm{d}w_2(t).
\end{align*}
\]

This model is one of the most commonly used models in mathematical finance for explaining the non-linear dynamics of financial time series. Here the prediction of \(S_t\) is locally Gaussian with mean \(\{\alpha + \beta S(t)\}\) \, \mathrm{d}t and the variance \(\sigma(t)^2 \, \mathrm{d}t\).

The processes defined by all three models above are Markov diffusion processes, and the prediction errors of these Markov diffusion processes are Gaussian white noise (Frost and Kailath, 1971). Therefore, using the Kalman filters derived from these models we can find three different ways of whitening the series \(S_1, \ldots, S_N\).

We can also allow shot noise into the dynamic model for \(S_t\) to give models with jumps. For example for the volatility model (iii) we have

\[
\begin{align*}
\mathrm{d}S(t) &= \{\alpha + \beta S(t)\} \, \mathrm{d}t + \sigma(t) \, \mathrm{d}w_1(t) + \kappa(t) \, \mathrm{d}q(t), \\
\mathrm{d}\log \sigma^2(t) &= \{\gamma + \delta \log \sigma^2(t)\} \, \mathrm{d}t + \varepsilon \, \mathrm{d}w_2(t),
\end{align*}
\]

while if we allow shot noise into the dynamic component of the model for \(\log \sigma^2(t)\) in (iii), we have

\[
\begin{align*}
\mathrm{d}S(t) &= \{\alpha + \beta S(t)\} \, \mathrm{d}t + \sigma(t) \, \mathrm{d}w_1(t), \\
\mathrm{d}\log \sigma^2(t) &= \{\gamma + \delta \log \sigma^2(t)\} \, \mathrm{d}t + \varepsilon \, \mathrm{d}w_2(t) + \kappa(t) \, \mathrm{d}q(t).
\end{align*}
\]

Here \(\mathrm{d}q(t)\) is a Poisson process with intensity \(\lambda\), and \(\kappa(t)\) is the size of the jump at time point \(t\).

5.2. State space representation

Constructing a discrete-time version of the volatility model using the Euler scheme leads to

\[
\begin{align*}
S_t &= \alpha + (1 + \beta)S_{t-1} + \sigma_{t-1} \xi_{1,t}, \\
\log \sigma_t^2 &= \gamma + (1 + \delta) \log \sigma_{t-1}^2 + \varepsilon \xi_{2,t},
\end{align*}
\]

where \(\xi_{2,t} \sim N(0,1)\). Taylor (1994) points out the difficulty of directly estimating the discretized volatility model. The ARCH model has been commonly recognized as a convenient approximate model for the volatility model and has been used in financial studies (Nelson, 1990; Nelson and Foster, 1994).

Here we take the approach of Nelson and Foster (1994) and use the approximation

\[
\varepsilon \xi_{2,t} \approx \varepsilon \frac{\xi_{2,t-1}^2 - 1}{\sqrt{2}}.
\]

Then we have

\[
\begin{align*}
S_t &= \alpha + (1 + \beta)S_{t-1} + \sigma_{t-1} \xi_{1,t}, \\
\log \sigma_t^2 &= \gamma + (1 + \delta) \log \sigma_{t-1}^2 + \varepsilon \frac{\xi_{2,t-1}^2 - 1}{\sqrt{2}},
\end{align*}
\]
for the approximation of the volatility model. By using its state space representation,

\[ x_t = [S_t \quad \log \sigma_t^2 \quad 1]^T, \]

\[ F_t = \begin{bmatrix} 1 + \beta & 0 & \alpha \\ 0 & 1 + \delta & \gamma + \varepsilon (\xi_{t-1}^2 - 1) / \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}, \]

\[ H_t = [1 \quad 0 \quad 0], \]

we can transform the time series \( S_1, \ldots, S_N \) into the innovation sequence \( \nu_1, \ldots, \nu_N \). By using Gaussian likelihood, we can estimate the parameters of the model by the maximum likelihood method.

When shot noise is present in the dynamic equation for the state variable \( S_t \), the discrete time approximate volatility model becomes

\[ S_t = \alpha + (1 + \beta) S_{t-1} + \sigma_{t-1} \xi_{1,t} + \kappa_{t-1} \nu_{k,t}, \]

\[ \log \sigma_t^2 = \gamma + (1 + \delta) \log \sigma_{t-1}^2 + \varepsilon (\xi_{t-1}^2 - 1) / \sqrt{2}, \]

where \( \kappa_t \sim N(0, \theta_{\kappa}^2) \) and \( \nu_{k,t} \) is a Bernoulli number with \( E(\nu_{k,t}) = \lambda \).

### 5.3. Numerical results

In this section we aim to see how these considerations affect data analysis. Here the three important concepts in modeling are:

1. variable transformation,
2. non-linear dynamic model, and
3. a mixture of Gaussian white and Poisson noise.

We analyzed the daily dollar–yen exchange rate data \( P_1, \ldots, P_N \) from January 1994 to December 1997 (see Figure 1), using the following six models. In them, \( a, b, c, d \) and \( e \) denote parameters to be fitted.
\[
\begin{align*}
\text{Model 1:} & \quad \begin{cases} 
S_t = \log P_t, \\
S_t = a + bS_{t-1} + \sigma_t \varepsilon_t, \\
\sigma_t^2 = c + d\sigma_{t-1}^2 + e\sigma_{t-1}^2 \varepsilon_{t-1}^2.
\end{cases} \\
\text{Model 2:} & \quad \begin{cases} 
S_t = \log(P_t/P_{t-1}), \\
S_t = a + bS_{t-1} + \sigma_t \varepsilon_t, \\
\sigma_t^2 = c + d\sigma_{t-1}^2 + e\sigma_{t-1}^2 \varepsilon_{t-1}^2.
\end{cases} \\
\text{Model 3:} & \quad \begin{cases} 
S_t = [\sinh^{-1}(\phi \log(P_t/P_{t-1}))]/\phi, \\
S_t = a + bS_{t-1} + c\varepsilon_t.
\end{cases} \\
\text{Model 4:} & \quad \begin{cases} 
S_t = [\sinh^{-1}(\phi \log(P_t/P_{t-1}))]/\phi, \\
S_t = a + bS_{t-1} + \sigma_t \varepsilon_t, \\
\sigma_t^2 = c + d\sigma_{t-1}^2 + e\sigma_{t-1}^2 \varepsilon_{t-1}^2.
\end{cases} \\
\text{Model 5:} & \quad \begin{cases} 
S_t = \log P_t, \\
S_t = a + bS_{t-1} + \sigma_t \varepsilon_t + \kappa_{t-1} \varepsilon_{t-1}, \\
\sigma_t^2 = c + d\sigma_{t-1}^2 + e\sigma_{t-1}^2 \varepsilon_{t-1}^2.
\end{cases} \\
\text{Model 6:} & \quad \begin{cases} 
S_t = \log P_t, \\
S_t = a + bS_{t-1} + \sigma_t \varepsilon_t, \\
\sigma_t^2 = c + d\sigma_{t-1}^2 + e\sigma_{t-1}^2 \varepsilon_{t-1}^2 + \kappa_{t-1} \varepsilon_{t-1}.
\end{cases}
\end{align*}
\]

Residuals (i.e., prediction errors), conditional variances and histograms of the normalized prediction errors of each fitted model are given in Figures 2–4. Here the prediction errors of the Model 5 and the Model 6 are a mixture of the Gaussian white noise and Poisson shot noise. Mean squared errors (MSE), AIC and log-likelihood of each fitted model are given in Table 1.

Models 5 and 6 are generalized version of Model 1. When we add shot noise to \( S_t \) in Model 1 we obtain Model 5, while adding shot noise to \( \log \sigma_t^2 \) of Model 1 yields Model 6. The AIC values in the table show that a big improvement comes from generalizing the Gaussian white noise for \( S_t \) to a mixture of Gaussian white noise and shot noise (compare Models 1 and 5), i.e, 148.9. By generalizing the noise model for \( \sigma_t^2 \) (compare Models 1 and 6) the AIC value improves by 112.1.

| \hline
| Model 1 | -993.4 | 1996.9 | 0.4553 |
| Model 2 | -994.0 | 1998.0 | 0.4555 |
| Model 3 | -957.0 | 1920.0 | 0.4549 |
| Model 4 | -939.0 | 1890.1 | 0.4550 |
| Model 5 | -918.0 | 1848.0 | 0.4559 |
| Model 6 | -936.4 | 1884.8 | 0.4554 |
| \hline

\textbf{Table 1: Estimation results.}
Figure 2: Innovations of Models 1–6 (top to bottom, in order).

Figure 3: Conditional variances of Models 1–6 (top to bottom, in order).
Model 4 is a generalization of both Model 2 and Model 3. If we replace the variable transformation $S_t = \log(P_t/P_{t-1})$ of Model 2 by $S_t = \sinh^{-1}(\phi \log(P_t/P_{t-1}))/\phi$ we obtain Model 4. Replacing the constant variance of Model 3 by a first order autoregressive model driven by $\sigma_{t-1}^2 \varepsilon_{t-1}^2$ leads to the GARCH type Model 4. From Model 2 to Model 4, AIC reduces by 108.1 while from Model 3 to Model 4, AIC reduces by 29.9. This means that the proper choice of transformation is as important as non-Gaussian generalization of white noise in modeling this financial time series.

6. Conclusions

In the paper we present a compatible innovation approach to the non-Gaussian time series analysis, which can be shared by both non-linear and non-Gaussian time series modelers. We illustrated by example that all three of the factors variable transformation, non-linear dynamic model, and using a mixture of Gaussian white and Poisson noise, can play an important role in modeling non-Gaussian time series.

We also provided a computationally simple method for the filtering of non-linear state space model with a mixture of Gaussian white noise and Poisson type shot noise.

Appendix A. Online detection of jumps

To get the maximum of the likelihood of the model with shot noise, we need to maximize the likelihood function over $2^N$ or $4^N$ possible cases. This is because the observations $z_s \ (s > t)$ depend on the past states $M_t$. This formidable computational task can be drastically simplified if we use an approximate likelihood function of a Markov state space model. Here we use the property, i.e. the dependency of the Markov state $M_t$ dies out exponentially. In other words, we assume that the influence of $M_t$ persists for only a short time. Then we have

$$\Pr\{M_t = M_t^{(i)} \mid Z^{t+1}\} \approx \Pr\{M_t = M_t^{(i)} \mid Z^N\},$$
An innovation approach to non-Gaussian time series analysis

where $Z^t = (z_1, \ldots, z_t)$. Then the probability of getting a jump, i.e. $M_t = M_t^{(1)}$ or not, i.e. $M_t = M_t^{(0)}$ for $t$ is calculated by

$$
\Pr\{M_t = M_t^{(i)} \mid Z^{t+1}\} = \frac{\sum_{k=0}^{1} \Pr\{z_{t+1} \mid Z^t, M_t^{(i)}, M_t^{(k)}\} \Pr\{z_t \mid Z^{t-1}, M_t^{(i)}\} \Pr\{M_t^{(i)}\} \Pr\{M_{t+1}^{(k)}\}}{\sum_{k=0}^{1} \sum_{j=0}^{1} \Pr\{z_{t+1} \mid Z^t, M_t^{(j)}, M_t^{(k)}\} \Pr\{z_t \mid Z^{t-1}, M_t^{(j)}\} \Pr\{M_t^{(j)}\} \Pr\{M_{t+1}^{(k)}\}}.
$$

The estimated state $\hat{M}_t$ is given by

$$
\hat{M}_t = \begin{cases} 
M_t^{(0)} & \text{if } \Pr\{M_t = M_t^{(0)} \mid Z^{t+1}\} > \Pr\{M_t = M_t^{(1)} \mid Z^{t+1}\}, \\
M_t^{(1)} & \text{if } \Pr\{M_t = M_t^{(0)} \mid Z^{t+1}\} < \Pr\{M_t = M_t^{(1)} \mid Z^{t+1}\}, 
\end{cases}
$$

where the estimated state is either $M_t^{(0)}$ or $M_t^{(1)}$ according to the above posterior probability. The prior probability of having a jump at time $t$ is $\lambda$, and we have

$$
\Pr(M_t^{(i)}) = \Pr(M_t = M_t^{(i)}) = \begin{cases} 
1 - \lambda & i = 0, \\
\lambda & i = 1.
\end{cases}
$$

With this approximate method the likelihood function of the model with shot noise can be calculated and maximized easily, much as for ARMA models. In calculating the likelihood function, jumps are detected on-line and recursively.

**Acknowledgement**

T. Ozaki would like to thank the Ministry of Education, Science and Culture, Japan, for supporting the present research with the COE Research Grant in the Institute of Statistical Mathematics, Tokyo, Japan.

**References**


