Normal inverse Gaussian autoregressive model using EM algorithm

Monika S. Dhull1 · Arun Kumar1

Abstract In this article, normal inverse Gaussian (NIG) autoregressive model is introduced. The parameters of the model are estimated using expectation maximization (EM) algorithm. The efficacy of the EM algorithm is shown using simulated and real-world financial data. It is shown that NIG autoregressive model fit very well on the considered financial data and hence could be useful in modelling of various real-life time-series data.

Keywords Normal inverse Gaussian distribution · Autoregressive model · EM algorithm · Monte Carlo simulations

Mathematics Subject Classification 91B02 · 62P05 · 62F10

1 Introduction

The most simple and intuitive time-series model encountered in the time-series modelling is the standard first-order autoregressive process which is also denoted by AR(1) [1]. In the AR(1) model, each new entry is the sum of two terms; one is proportional to the previous entry and the another is a white-noise process also called the error term or the innovation term. In a white-noise process, variables are independent and identically distributed (iid) with zero mean. If the variables are Gaussian, this is called Gaussian white noise. However, in financial markets, the observed time-series (log-returns) distributions are non-Gaussian and have tails heavier than Gaussian and lighter than power law. These kind of distributions are also called semi-heavy-tailed distributions (see e.g. [2–4]). For a literature survey on different innovation distributions or marginal distributions in the AR(1) model [5], the AR(1) models with non-Gaussian innovation terms are very well-considered in the literature. Sim [6] considered AR(1) model with Gamma process as the innovation term. For AR models with innovations following a Student’s t distribution, see e.g., [7–11] and references therein. Note that t distribution is used in modelling of asset returns [12].

Normal inverse Gaussian (NIG) distribution introduced by Barndorff-Nielsen [13] is a semi-heavy-tailed distribution with tails heavier than the Gaussian but lighter than the power law tails. NIG distribution is defined as the normal variance–mean mixture where the mixing distribution is the inverse Gaussian distribution. NIG distributions and processes are used to model the returns from the financial time series, see e.g. [14, 15]. A more general distribution which is obtained by taking normal variance–mean mixture with mixing distribution as generalised inverse Gaussian distribution is called generalised hyperbolic distribution [16].

In this article, we introduce an AR(1) model with NIG innovation terms. Due to heavy-tailedness of the innovation term, it can model large jumps in the observed data. The introduced process very well model the Google stock price time series. The non-Gaussian behaviour of the innovation term of the Google stock price is shown using QQ plot and Kolmogorov–Smirnov test. From a market risk management perspective, obtaining reasonable extreme observation levels is a crucial objective in modelling. Since, it captures the market extreme movements which a Gaussian-based model does not capture. The
parameters of the model are estimated using EM algorithm. The efficacy of the estimation procedure is shown on the simulated data. The rest of the paper is organised as follows. In Sect. 2, the NIG autoregressive model is defined along with important properties of the NIG distribution. Section 3 discusses the estimation procedure of the parameters of the introduced model using EM algorithm. The efficacy of the estimation procedure on simulated data and the real-world financial data application is discussed in Sect. 4. Last section concludes.

2 NIG autoregressive model

We introduce the AR(1) model with iid normal inverse Gaussian (NIG) innovations \( e_t \). Consider a univariate time series \( Y_1, Y_2, \ldots, Y_t \) following

\[
Y_t = \rho Y_{t-1} + e_t, \tag{2.1}
\]

where \( e_t \sim \text{NIG}(\alpha, \beta, \mu, \delta) \). NIG is a semi-heavy-tailed distribution which can be obtained as normal mean–variance mixture with inverse Gaussian as mixing distribution. The conditional distribution of \( Y_t \) given all the parameters \( \rho, \alpha, \beta, \mu, \) and \( \delta \) conditioning on all the preceding data \( \mathcal{F}_{t-1} \) (i.e. \( Y_1, Y_2, \ldots, Y_{t-1} \)) only depend on previous sample \( Y_{t-1} \) and has the form

\[
p(Y_t | \rho, \alpha, \beta, \mu, \delta, \mathcal{F}_{t-1}) = p(Y_t | \rho, \alpha, \beta, \mu, \delta, Y_{t-1}) = f_t(y_t; \rho Y_{t-1}, \alpha, \beta, \mu, \delta),
\]

where \( f_t(\cdot) \) denotes the probability density function (pdf) of NIG(\( \alpha, \beta, \mu, \delta \)), which is given by

\[
f(x; \alpha, \beta, \mu, \delta) = \frac{\alpha}{\pi} \exp \left( \frac{\delta}{\sqrt{x^2 - \beta^2 - \beta \mu}} \right) \phi(x)^{-1/2}
\]

\[
K_1(\delta \sqrt{x^2}) \exp(\beta x), \quad x \in \mathbb{R},
\]

where \( \phi(x) = 1 + [(x - \mu) / \delta]^2 \), \( \alpha = \sqrt{\delta^2 + \beta^2} \) and \( K_1(x) \) denotes the modified Bessel function of the third kind of order \( v \) evaluated at \( x \) and is defined by

\[
K_1(x) = \frac{1}{2} \int_0^\infty y^{x-1} e^{-y-y^{-1}} dy.
\]

Further, the following properties of \( K_1(\cdot) \) are used in subsequent calculations.

\[
\begin{align*}
K_{-1}(x) &= K_1(x); \quad K_2(x) = K_0(x) + 2x K_1(x). \\
\end{align*}
\]

A pdf \( h(x) \) is called a semi-heavy-tailed pdf if

\[
h(x) \sim e^{-c x} g(x), \quad c > 0,
\]

where \( g \) is a regularly varying function [4]. Recall that a positive function \( g \) is regularly varying with index \( x \) if

\[
\lim_{x \to \infty} \frac{g(dx)}{g(x)} = d^x, \quad d > 0.
\]

Using \( K_1(\alpha) \sim \sqrt{\frac{\alpha}{2\pi}} e^{-\alpha^2/2} \), as \( \alpha \to \infty \) [17], we have

\[
f(x; \alpha, \beta, \mu, \delta) \sim \frac{\alpha}{\pi} e^{(\delta \sqrt{x^2 - \beta^2 - \beta \mu})} \frac{\sqrt{\pi}}{2} \left( \frac{\delta \sqrt{x^2}}{\sqrt{x^2 - \beta^2 - \beta \mu}} \right)^{1/2} e^{(\beta x)} + \frac{\sqrt{\pi}}{2} \left( \frac{\delta \sqrt{x^2}}{\sqrt{x^2 - \beta^2 - \beta \mu}} \right)^{-1/2} e^{-(\beta x)} e^{(-x \sqrt{\alpha^2 - \beta^2})} \quad \alpha > \beta,
\]

as \( x \to \infty \). Thus, the pdf of NIG is semi-heavy-tailed. Due to exponentially decaying tails, the moments of all order exist for NIG distribution. The NIG-distributed random variable can be generated from its normal variance–mean mixture form, such that

\[
X = \mu + \beta G + \sqrt{\beta} Z,
\]

where \( Z \) is standard normal, i.e. \( Z \sim N(0, 1) \) and \( G \) has an inverse Gaussian distribution (IG) with parameters \( \gamma \) and \( \delta \) denoted by \( G \sim IG(\gamma, \delta) \). The mixing distribution \( IG(\gamma, \delta) \) having pdf

\[
g(x; \gamma, \delta) = \frac{\delta}{\sqrt{2\pi}} \exp(\delta x) x^{-3/2} \exp \left( -\frac{1}{2} \left( \frac{\delta^2}{x} + \gamma^2 x \right) \right),
\]

\[
x > 0.
\]

This form of NIG distribution makes it suitable to use EM algorithm for ML estimation of parameters. For AR(1) model, consider that the innovations \( e_t \) follow NIG distribution and has the form defined in (2.1). The conditional distribution \( e_t | G_t = g \sim N(\mu + \beta g, g) \), here \( G_t \) are independent copies of \( G \sim IG(\gamma, \delta) \).

3 Parameter estimation using EM algorithm

In this section, we provide a step-by-step procedure to estimate the parameters of the proposed normal inverse Gaussian autoregressive model of order 1 called NIGAR(1) model using EM algorithm. The EM algorithm for NIG innovation is based on [18]. Note that EM algorithm is a general iterative algorithm for model parameter estimation by maximizing the likelihood in the presence of missing data [19]. An alternative to numerical optimization of the likelihood function, the EM algorithm was introduced in
1977 by Dempster, Laird and Rubin [20]. The EM algorithm is popularly used in estimating Gaussian mixture models (GMMs), estimating hidden Markov models (HMMs) and model-based data clustering [19, 21]. Recently, EM algorithm is also used in parameter estimation of time-series model [22, 23]. Some extensions of EM algorithm include expectation conditional maximization (ECM) algorithm proposed by Meng and Rubin [24]; expectation conditional maximization either (ECME) algorithm of Liu and Rubin [25]. A detailed discussion on the theory of EM algorithm and its extensions is given in McLachlan and Krishnan [26].

The algorithm is proven to be numerically stable. Also, as a consequence of Jensen’s inequality, log-likelihood function at the updated parameters \( \theta^{(k+1)} \) will not be less than that at the current values \( \theta^{(k)} \). Although there is always a concern that the algorithm might get stuck at local extrema, but it can be handled by starting from different initial values and comparing the solutions. In next proposition, we provide the estimates of the parameters if NIGAR(1) model uses EM algorithms.

**Proposition 3.1** Suppose the innovation terms \( \eta_t \) in the AR(1) time-series model \( Y_t = \rho Y_{t-1} + \eta_t \) are from NIG

\[
\hat{\theta} = \frac{\sum_{t=1}^{n} w_t y_{t-1} - \sum_{t=1}^{n} w_t y_{t-1} \sum_{t=1}^{n} y_{t-1}}{\left( \sum_{t=1}^{n} w_t y_{t-1} - \sum_{t=1}^{n} w_t \sum_{t=1}^{n} y_{t-1} \right)} + \frac{\sum_{t=1}^{n} w_t y_{t-1} - \sum_{t=1}^{n} w_t y_{t-1} \sum_{t=1}^{n} y_{t-1}}{\left( \sum_{t=1}^{n} w_t y_{t-1} - \sum_{t=1}^{n} w_t \sum_{t=1}^{n} y_{t-1} \right)}; \quad (3.1)
\]

\[
\hat{\mu} = \frac{\sum_{t=1}^{n} w_t y_{t-1} - \sum_{t=1}^{n} w_t y_{t-1} \sum_{t=1}^{n} y_{t-1}}{\left( \sum_{t=1}^{n} w_t y_{t-1} - \sum_{t=1}^{n} w_t \sum_{t=1}^{n} y_{t-1} \right)}; \quad (3.2)
\]

McLachlan and Krishnan [26]. The EM algorithm iterates between two steps:

**E-Step:** In this step, create a function \( Q(\theta|\theta^{(k)}) \) to compute the expectation of log-likelihood of complete data \( (X, G) \) with respect to the conditional distribution of \( G \) given \( X \), using the current estimate for the parameters.

\[
Q(\theta|\theta^{(k)}) = \mathbb{E}_{G|X, \theta^{(k)}}[\log f(X, G|\theta)|X, \theta^{(k)}].
\]

**M-Step:** Compute the parameters by maximizing the expected log-likelihood of complete data found on the E-step such that

\[
\theta^{(k+1)} = \arg\max_{\theta} Q(\theta|\theta^{(k)}).
\]

The algorithm is proven to be numerically stable. Also, as a \((\alpha, \beta, \mu, \delta)\) distribution. Then, the ML estimates of the model parameters using EM algorithm are as follows:

\[
\hat{\beta} = \frac{\sum_{t=1}^{n} w_t y_t - \rho \sum_{t=1}^{n} w_t y_{t-1} - \mu \sum_{t=1}^{n} w_t}{n}; \quad (3.3)
\]

\[
\hat{\delta} = \sqrt{\frac{\bar{s}}{\bar{w}}} \hat{\gamma} = \frac{\hat{\delta}}{\bar{s}}, \quad \text{and} \quad \hat{\xi} = (\gamma^2 + \beta^2)^{1/2}; \quad (3.4)
\]

where \( \bar{s} = \sum_{t=1}^{n} s_t, \bar{w} = \sum_{t=1}^{n} w_t \).

**Proof** For AR(1) model, to implement EM algorithm let \((\eta_t, G_t)\), for \( t = 1, 2, \ldots, n \) denote the complete data. The observable data \( \eta_t \) is assumed to be from NIG\((\alpha, \beta, \mu, \delta)\) and the unobserved data \( G_t \) follows IG\((\gamma, \delta)\). Also, given the time-series data \( y_1, y_2, \ldots, y_n \), the innovations term can
be rewritten as
\[ e_t = y_t - \rho y_{t-1}, \quad \text{for } t = 1, 2, \ldots, n. \]

Note that \( e_t | G = g \sim N(\mu + \beta g, g) \) and the conditional pdf is
\[
f(e | G = g_t) = \frac{1}{\sqrt{2\pi g_t}} \exp \left(-\frac{1}{2g_t} (y_t - \rho y_{t-1} - \mu - \beta g_t)^2 \right).\]

Let \( \hat{\theta} = (\alpha, \beta, \mu, \delta, \rho) \) denote the vector of parameters we need to estimate. At E-step, we need to find the conditional expectation of log-likelihood of complete data \( (e, G) \) with respect to the conditional distribution of \( G \) given \( e \). In our case, unobserved data is from \( IG(\gamma, \delta) \), or we can say prior distribution for \( G \) is \( IG(\gamma, \delta) \); therefore, the posterior distribution is generalized inverse Gaussian (GIG) distribution. The distribution of \( G | e, \alpha, \beta, \mu, \delta, \rho \) is \( GIG(-1, \delta \sqrt{\phi(e)}, \alpha) \). The conditional first moments and inverse first moment are as follows:
\[
\mathbb{E}(G | e) = \frac{\delta \phi(e)^{1/2}}{\alpha} K_0(\alpha \delta \phi(e)^{1/2}) \quad \frac{1}{K_1(\alpha \delta \phi(e)^{1/2})},
\]
\[
\mathbb{E}(G^{-1} | e) = \frac{\alpha}{\delta \phi(e)^{1/2}} K_{-2}(\alpha \delta \phi(e)^{1/2}) \quad \frac{1}{K_{-1}(\alpha \delta \phi(e)^{1/2})}.
\]

These expectations will be useful in finding the conditional expectation of the log-likelihood function. The complete data likelihood is given by
\[
L(\theta) = \prod_{t=1}^{n} f(e_t, g_t) = \prod_{t=1}^{n} f(e_t | G = g_t) f(G = g_t)
\]
\[
= \prod_{t=1}^{n} \frac{1}{2\pi g_t} \exp \left(-\frac{1}{2g_t} (y_t - \rho y_{t-1} - \mu - \beta g_t)^2 \right) \exp(-\delta^2 \frac{g_t}{2} - \frac{\gamma^2}{2} g_t - \frac{\delta^2}{2} g_t^2 \sum_{t=1}^{n} y_t - \rho y_{t-1} - \mu)^2 - \frac{\beta^2}{2} g_t + \beta \sum_{t=1}^{n} (y_t - \rho y_{t-1} - \mu)).
\]

The log likelihood function will be
\[
l(\theta) = n \log(\delta) - n \log(2\pi) + n \delta \gamma - n \beta \mu - 2 \sum_{t=1}^{n} \log(g_t)
\]
\[
- \frac{\delta^2}{2} \sum_{t=1}^{n} g_t^{-1}
\]
\[
- \frac{\gamma^2}{2} \sum_{t=1}^{n} g_t^{-1} \sum_{t=1}^{n} (y_t - \rho y_{t-1} - \mu)^2
\]
\[
- \frac{\beta^2}{2} \sum_{t=1}^{n} g_t + \beta \sum_{t=1}^{n} (y_t - \rho y_{t-1}).
\]

Now to implement the E-step of EM algorithm, we need to compute \( Q(\theta | \theta^{(k)}) \), which is the expected value of complete data log likelihood and can be expressed as
\[
Q(\theta | \theta^{(k)}) = \mathbb{E}_{G|e, \theta^{(k)}} [\log f(e, G | \theta)] = \mathbb{E}_{G|e, \theta^{(k)}} [L(\theta | \theta^{(k)})]
\]
\[
= n \log(\delta) + n \delta \gamma - n \beta \mu - n \log(2\pi)
\]
\[
- 2 \sum_{t=1}^{n} \mathbb{E}(\log g_t | e_t, \theta^{(k)}) - \frac{\delta^2}{2} \sum_{t=1}^{n} w_t
\]
\[
- \frac{\gamma^2}{2} \sum_{t=1}^{n} s_t - \frac{\beta^2}{2} \sum_{t=1}^{n} s_t + \beta \sum_{t=1}^{n} (y_t - \rho y_{t-1})
\]
\[
- \frac{1}{2} \sum_{t=1}^{n} (y_t - \rho y_{t-1} - \mu)^2 w_t,
\]

where \( s_t = \mathbb{E}_{G|e, \theta^{(k)}} (g_t | e_t, \theta^{(k)}) \) and \( w_t = \mathbb{E}_{G|e, \theta^{(k)}} (g_t^{-1} | e_t, \theta^{(k)}) \). At M-step, to update the parameters, maximize the \( Q \) function by solving the following equations:
\[
\frac{\partial Q}{\partial \rho} = \sum_{t=1}^{n} w_t y_t y_{t-1} - \beta \sum_{t=1}^{n} y_{t-1}
\]
\[
- \mu \sum_{t=1}^{n} w_t y_{t-1}^2 - \rho \sum_{t=1}^{n} w_t y_{t-1}^2,
\]
\[
\frac{\partial Q}{\partial \mu} = -n \beta + \sum_{t=1}^{n} w_t y_t - \mu \sum_{t=1}^{n} w_t y_{t-1} - \rho \sum_{t=1}^{n} w_t y_t - \rho \sum_{t=1}^{n} w_t y_{t-1},
\]
\[
\frac{\partial Q}{\partial \beta} = -n \mu + \sum_{t=1}^{n} y_t - \beta \sum_{t=1}^{n} s_t - \rho \sum_{t=1}^{n} y_{t-1},
\]
\[
\frac{\partial Q}{\partial \gamma} = \frac{n}{\beta} - \frac{\delta}{\gamma} \sum_{t=1}^{n} w_t,
\]
\[
\frac{\partial Q}{\partial \delta} = n \delta - \gamma \sum_{t=1}^{n} s_t.
\]

Table 1: Actual and estimated parameters using different stopping criterion for case 1

| Actual | Estimated    |
|--------|-------------|
| \( \alpha = 2.24, \beta = 1, \mu = 1, \delta = 2 \) | \( \hat{\alpha} = 2.091, \hat{\beta} = 0.892, \hat{\mu} = 1.042, \hat{\delta} = 1.962 \) |
| \( \alpha = 2.24, \beta = 1, \mu = 1, \delta = 2 \) | \( \hat{\alpha} = 2.090, \hat{\beta} = 0.892, \hat{\mu} = 1.042, \hat{\delta} = 1.962 \) |
Solving the above equations, the following estimates of the parameters are obtained:

\[
\hat{\rho} = \left( \frac{n}{n} \sum_{t=1}^{n} w_t y_{t-1} - \sum_{t=1}^{n} w_t y_{t-1} \sum_{t=1}^{n} y_{t-1} \right) \left( \sum_{t=1}^{n} s_t \sum_{t=1}^{n} w_t - n^2 \right) + \left( \frac{n}{n} \sum_{t=1}^{n} w_t y_{t-1} - \sum_{t=1}^{n} w_t y_{t-1} \sum_{t=1}^{n} y_{t-1} \right) \left( \sum_{t=1}^{n} y_t - \sum_{t=1}^{n} w_t y_{t-1} \sum_{t=1}^{n} s_t \right);
\]

\[
\hat{\beta} = \frac{n}{n} \sum_{t=1}^{n} w_t y_{t-1} - \rho \sum_{t=1}^{n} w_t y_{t-1} - \mu \sum_{t=1}^{n} w_t;
\]

\[
\hat{\delta} = \frac{\overline{s}}{\sqrt{sw - 1}} \quad \hat{\gamma} = \frac{\delta}{\overline{s}} \quad \text{and} \quad \hat{\zeta} = (\gamma^2 + \beta^2)^{1/2},
\]

where \( \overline{s} = \sum_{n=1}^{n} s_t \), \( \overline{w} = \sum_{n=1}^{n} w_t \). □

**Proposition 3.2** For the time series defined in (2.1), we have

\[
\mathbb{E}(Y_t) = \frac{\mu\gamma + \delta\beta}{\gamma} \left( 1 - \rho^{t+1} \right) - \frac{1}{1 - \rho};
\]

\[
\text{Var}(Y_t) = \frac{\delta^2}{\gamma^2} \left( 1 - \rho^{2t+2} \right).
\]

**Proof** For \( \epsilon_t \sim \text{NIG}(\alpha, \beta, \mu, \delta) \), we have \( \mathbb{E}(\epsilon_t) = \frac{\mu\gamma + \delta\beta}{\gamma} \) and \( \text{Var}(\epsilon_t) = \frac{\delta^2}{\gamma^2} \). For an AR(1) time series, it follows

\[
\mathbb{E}(Y_t) = \rho \mathbb{E}(Y_{t-1}) + \mathbb{E}\epsilon_t = \mathbb{E}\epsilon_t + \rho \mathbb{E}\epsilon_{t-1} + \rho^2 \mathbb{E}\epsilon_{t-2} + \cdots + \rho^t \mathbb{E}\epsilon_1
\]

\[
= \mathbb{E}(\epsilon_1) \left( 1 - \rho^{t+1} \right) = \frac{\mu\gamma + \delta\beta}{\gamma} \left( 1 - \rho^{t+1} \right).
\]

Using law of total variance to calculate \( \text{Var}(Y_t) \) yields to

\[
\text{Var}(Y_t) = \mathbb{E}[\text{Var}(Y_t | Y_{t-1} = y_{t-1})] + \text{Var}[\mathbb{E}(Y_t | Y_{t-1} = y_{t-1})]
\]

\[
= \mathbb{E}[\text{Var}(\rho Y_{t-1} + \epsilon_t | Y_{t-1} = y_{t-1})] + \text{Var}[\mathbb{E}(\rho Y_{t-1} + \epsilon_t | Y_{t-1} = y_{t-1})]
\]

\[
= \mathbb{E}[\text{Var}(\rho Y_{t-1} + \epsilon_t | Y_{t-1} = y_{t-1})] + \text{Var}[\mathbb{E}(\epsilon_t) + \rho Y_{t-1}]
\]

\[
= \text{Var}(\epsilon_t) + \text{Var}(\rho Y_{t-1}) = \text{Var}(\epsilon_t) + \rho^2 \text{Var}(Y_{t-1}).
\]

Recursively using the above relation and since \( \epsilon_t \) are iid, we can write

\[
\text{Var}(Y_t) = \left( 1 + \rho^2 + \rho^4 + \cdots + \rho^{2t} \right) \text{Var}(\epsilon_t)
\]

\[
= \frac{1 - \rho^{2t+2}}{1 - \rho^2} \text{Var}(\epsilon_t)
\]

\[
= \frac{\delta^2}{\gamma^2} \left( 1 - \rho^{2t+2} \right) \text{Var}(\epsilon_t)
\]

\[
= \frac{\delta^2}{\gamma^2} \left( 1 - \rho^{2t+2} \right).
\]

In [18], an EM algorithm was developed to estimate the parameters of NIG distribution. The estimate of the parameter \( \rho \) can be evaluated directly by applying the conditional least square method on the time-series data \( y_t \) which is obtained by minimizing

\[
\sum_{i=1}^{n} (Y_{t+1} - \mathbb{E}(Y_{t+1} | Y_t, Y_{t-1}, \ldots, Y_1))^2,
\]

with respect to \( \hat{\theta} \) and leads to

\[
\hat{\rho} = \frac{\sum_{i=0}^{n} (y_i - \overline{y})(y_{i+1} - \overline{y})}{\sum_{i=0}^{n} (y_i - \overline{y})^2}.
\]

For remaining parameters \( \theta = (\alpha, \beta, \mu, \delta) \), we will use the EM algorithm proposed by Karlis [18] for ML estimation of NIG distribution. Using the steps given in [18], the following estimates of the parameters are obtained:
In this section, we discuss the parameter estimation of the proposed time-series model on the simulated data. Further, the application of the NIGAR(1) model is shown in Fig 1. It is intuitive from the plot that the series has roughly constant mean and constant variance over time with occasional large jumps. This kind of behaviour cannot be explained by using standard Gaussian innovation term autoregressive model.

4 Simulation study and real-world data application

In this section, we discuss the parameter estimation of the proposed model on the simulated data. Further, the application of the proposed time-series model is shown on the Google equity price.

Simulation study: For simulation study, it is required to generate the iid inverse Gaussian distributed random numbers \( G_i \sim IG(\mu_i, \lambda_1) \), \( i = 1, 2, \ldots, N \). We use the algorithm mentioned in [27], which involves following steps:

**Step 1**: Generate standard normal variate \( N \) and set \( Y = N^2 \).

**Step 2**: Set \( X_1 = \mu_1 + \frac{1}{2} \frac{\mu_1^2}{\lambda_1} - \frac{\mu_1}{\lambda_1} \sqrt{4\mu_1\lambda_1 Y + \mu_1^2 Y^2} \).

**Step 3**: Generate uniform random variate \( U[0, 1] \).

**Step 4**: If \( U < \frac{\hat{\mu}_t}{\hat{\mu}_t + X_t} \), then \( G = X_t \); else \( G = \frac{\hat{\mu}_t^2}{X_t} \).

Note that the form of the pdf taken in Devroye [27] is different than the form given in (2.3), so it is required to make the following substitutions for the parameters \( \mu_1 = \frac{\delta}{\gamma} \) and \( \hat{\lambda}_1 = \hat{\delta}^2 \). These substitutions using above steps will generate random numbers from \( IG(\gamma, \delta) \) with pdf given in (2.3). In order to generate the iid NIG innovation terms, we use the normal variance–mean mixture form of the NIG with inverse Gaussian as mixing distribution and standard normal distribution representation given in (2.2), which will give \( \epsilon_t \sim NIG(\alpha, \beta, \mu, \delta) \). A simulated series of NIGAR(1) model is shown in Fig 1. It is intuitive from the plot that the series has roughly constant mean and constant variance over time with occasional large jumps. This kind of behaviour cannot be explained by using standard Gaussian innovation term autoregressive model.

Next, we simulate an NIG random sample of size 10,000 with true parameter values as \( \alpha = 2.24, \beta = 1, \mu = 1 \) and \( \delta = 2 \). Using the relationship \( y_t = \rho y_{t-1} + \epsilon_t \), the simulated data is used to generate the NIGAR(1) time series with \( \rho = 0.5 \) and \( y_0 = \epsilon_0 \). To analyse the behaviour of the EM algorithm, we use two different stopping criteria, first is based on the relative change of the log-likelihood (criterion 1) and the other is based on the relative change of the parameter values (criterion 2). Let \( L(k) \) denote the log-likelihood value after \( k \) iterations, then we stop the iterations when \( \frac{L(k) - L(k+1)}{L(k)} < 10^{-5} \). In other criterion, we stop the iterations when

\[
0.001 \leq P(Y 
\leq 0.01)
\]

Fig. 1 Simulated NIGAR(1) series (size = 1000) for parameters \( \delta = 2, \gamma = 2, \mu = 1, \beta = 1 \) with (a) \( \rho = 0.5 \) and (b) \( \rho = 0.9 \)
Table 2  Actual and estimated parameters using different stopping criterion for Case 2

| Criterion  | Actual | Estimated |
|------------|--------|-----------|
| Criterion 1| $\alpha = 2.24, \beta = 1, \mu = 1, \delta = 2$ | $\hat{\alpha} = 2.073, \hat{\beta} = 0.880, \hat{\mu} = 1.077, \hat{\delta} = 1.953$ |
| Criterion 2| $\alpha = 2.24, \beta = 1, \mu = 1, \delta = 2$ | $\hat{\alpha} = 1.980, \hat{\beta} = 0.824, \hat{\mu} = 1.125, \hat{\delta} = 1.897$ |

\[
\max \left\{ \left| \frac{\beta(k+1) - \beta(k)}{\beta(k)} \right|, \left| \frac{\mu(k+1) - \mu(k)}{\mu(k)} \right|, \left| \frac{\gamma(k+1) - \gamma(k)}{\gamma(k)} \right|, \left| \frac{\delta(k+1) - \delta(k)}{\delta(k)} \right| \right\} < 10^{-5}.
\]

**Case 1:** We use the parameter estimates derived from Proposition 3.1 on simulated data.

Using the first stopping criterion, the estimated values of parameters are $\hat{\alpha} = 2.073, \hat{\beta} = 0.880, \hat{\mu} = 1.077, \hat{\delta} = 1.953$, and $\hat{\gamma} = 1.877$.

When second criterion is used the estimated values are $\hat{\alpha} = 1.980, \hat{\beta} = 0.824, \hat{\mu} = 1.125, \hat{\delta} = 1.897$, and $\hat{\gamma} = 1.800$. In order to check the convergence of algorithm, we used different set of initial values which resulted in approximately same estimates. The results are summarized in following table:

**Case 2:** We use the parameter estimates of NIG proposed by Karlis [18] and estimate the $\rho$ using conditional least square estimate given in (3.6). The estimated values are given in Table 2.

In Fig. 2, we display the box plots of the parameters estimated using EM algorithm on 100 simulated data sets with the help of Proposition 3.1. Observe that the parameters are symmetric and centred on $\hat{\alpha} = 2.22, \hat{\beta} = 1.01, \hat{\mu} = 0.99, \hat{\delta} = 1.99$ and $\hat{\rho} = 0.501$.

From the whiskers size, we can observe that the distribution of the parameters may be heavy-tailed.

**Real-world data application:** The historical financial data of Google equity is collected from Yahoo finance. The whole data set covers the period from 31 December 2014 to 30 April 2021. Initially, the data contained 1594 data points with six features having Google stock’s open price, closing price, highest value, lowest value, adjusted close price and volume of each working day end-of-the-day values. In order to apply the proposed NIGAR(1) model, we take the univariate time series $y_t$ to be the end-of-the-day closing prices. The Google equity closing price is demonstrated as time-series data in Fig. 3.

We assume that the innovation terms $\varepsilon_t$ of time-series data $y_t$ follows NIG. We use ACF and PACF plot to determine the appropriate time-series model components for closing price. Figure 4 shows the ACF and PACF plot of time-series data $y_t$.

We observe that in PACF plot, there is a significant spike at lag 1, also we ignore the spike at lag 0 as it represents the correlation between the term itself which will always be 0. PACF plot indicates that the closing prices follow AR model with lag 1. In ACF plot, all the spikes are significant for lags upto 30 which implies that the closing price is highly correlated. Therefore, from ACF and PACF plot, the assumed NIGAR(1) model is expected to be good fit for data. First, we estimate the $\rho$ parameter using the conditional least square method as mentioned in (3.6)

\[
\hat{\rho} = \frac{\sum_{t=0}^{n} (y_t - \bar{y}) (y_{t+1} - \bar{y})}{\sum_{t=0}^{n} (y_t - \bar{y})^2}.
\]
The estimated value is $\hat{\rho} = 0.9941$. Using the estimated value of $\rho$ and relation $e_t = y_t - \rho y_{t-1}$, we will get the innovation terms $e_t$. The distribution plot of innovation terms is shown in Fig. 5a.

The Kolmogorov–Smirnov (KS) normality test and Jarque–Bera (JB) test are performed on $e_t$ to test if the innovation terms are Gaussian. The $p$-value in both the tests was 0, which indicates that the $e_t$ may not be from Gaussian distribution. Therefore, we fit the proposed NIGAR(1) model on Google closing price dataset. The ML estimates of parameters using EM algorithm are $\hat{\alpha} = 0.0202, \hat{\beta} = 0.0013, \hat{\mu} = 0.226, \hat{\delta} = 9.365$, and $\hat{\gamma} = 0.0201$ with initial guesses as $\hat{\alpha}^{(0)} = 0.0141, \hat{\beta}^{(0)} = 0.01, \hat{\mu}^{(0)} = 0.01, \hat{\delta}^{(0)} = 0.01$, and $\hat{\gamma}^{(0)} = 0.01$. The relative change in the log-likelihood value with tolerance value 0.0001 is used as stopping criterion. It is worthwhile to mention that the estimated $\beta$ is close to 0 and the estimated $\mu$ can be interpreted as intercept term in the AR(1) model.

Based on the Google equity prices data and estimated parameters, an equivalent model to (2.1) can be described as

$$Y_t = \mu + \rho Y_{t-1} + \epsilon_t,$$

where $\epsilon_t = \sqrt{GZ}$ with $G \sim IG(\gamma, \delta)$, with $\hat{\rho} = 0.9941, \hat{\mu} = 0.226, \hat{\delta} = 9.365$, and $\hat{\gamma} = 0.0201$.

To test the reliability of the estimated results, we used the simulated data with true parameter values as $\alpha = 0.02, \beta = 0, \mu = 0.23, \delta = 9.5$, and $\gamma = 0.02$. The QQ-plot (Fig. 5b) between the error terms $e$ of data and the NIG simulated data indicates that the innovation terms are
significantly NIG with few outliers. Also, we performed the two-sample KS test which is used to check whether the two samples are from same distribution. The two-sample KS test on the simulated data and the innovation terms resulted in the $p$-value = 0.3413. Therefore, we failed to reject the null hypothesis that the samples are from same distribution and conclude that both samples follow the same distribution, i.e. \( \text{NIG}(x = 0.02, \beta = 0, \mu = 0.23, \delta = 9.5) \).

5 Conclusions

The introduced AR(1) model with NIG innovations very well model the prices of the Google equity for the period from 31 December 2014 to 30 April 2021. Note that NIG distribution is semi-heavy-tailed and used to model equity returns in the literature. So, the proposed model could be useful in modelling time series which manifests large observations where the classical AR(1) model with Gaussian-based innovations cannot be used. We estimate the parameters using EM algorithm since direct maximization of likelihood function is difficult due to the modified Bessel function of the third kind term. The EM algorithm estimates the parameter efficiently, and the results are verified on the simulated data.

Acknowledgements Monika S. Dhull would like to thank the Ministry of Education (MoE), Government of India, for supporting her PhD research work.

References

1. Tsay, R.S.: Analysis of Financial Time Series, 2nd edn. Wiley, Hoboken (2005)
2. Rachev, S.T.: Handbook of Heavy Tailed Distributions in Finance: Hand-books in Finance. Elsevier, Amsterdam (2003)
3. Cont, R., Tankov, P.: Financial Modeling with Jump Processes. Chapman & Hall, London (2004)
4. Omeye, E., Gulcka, S.V., Vesilo, R.: Semi-heavy tails. Lith. Math. J. 58, 480–499 (2018)
5. Grunwald, G., Hyndman, R., Tedesco, L.: A Unified View of Linear AR(1) Models. D.P. Monash University, Clayton (1996)
6. Sim, C.H.: First-order autoregressive models for gamma and exponential processes. J. Appl. Prob. 27, 325–332 (1990)
7. Tiku, M.L., Wong, W.-K., Vaughan, D.C., Bian, G.: Time series models in non-normal situations: symmetric innovations. J. Time Ser. Anal. 21, 571–596 (2000)
8. Tarami, B., Pourahmadi, M.: Multi-variate t autoregressions: Innovations, prediction variances and exact likelihood equations. J. Time Ser. Anal. 24, 739–754 (2003)
9. Christmas, J., Everson, R.: Robust autoregression: student-t innovations using variational Bayes. IEEE Trans. Signal Process. 59, 48–57 (2011)
10. Nduka, U.C.: EM-based algorithms for autoregressive models with $t$-distributed innovations. Commun. Stat. Simul. Comput. 47, 206–228 (2018)
11. Meitz, M., Preve, D., Saikkonen, P.: A mixture autoregressive model based on Student’s $t$-distribution. Commun. Stat. Theory Methods (2021)
12. Heyde, C.C., Leonenko, N.N.: Student processes. Adv. Appl. Probab. 37, 342–365 (2005)
13. Barndorff-Nielsen, O.E.: Exponentially decreasing distributions for the logarithm of particle size. Proc. R. Soc. Lond. Ser. A. 353, 401–409 (1977)
14. Kallemanno, A., Schmid, B., Werner, R.: The Normal inverse Gaussian distribution for synthetic CDO pricing. J. Deriv. 14, 80–93 (2007)
15. Barndorff-Nielsen, O.E.: Normal inverse Gaussian distributions and stochastic volatility modelling. Scand. J. Stat. 24, 1–13 (1997)
16. Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.I.: Lévy Processes: Theory and Applications. Birkhauser, Basel (2013)
17. Jørgensen, B.: Statistical Properties of the Generalized Inverse Gaussian Distribution. Lecture Notes in Statistics, vol. 9. Springer, New York (1982)
18. Karlis, D.: An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. Stat. Prob. Lett. 57, 43–52 (2002)
19. Do, C.B., Batzoglou, S.: What is the expectation maximization algorithm? Nat. Biotechnol. 26(8), 897–899 (2008)
20. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via EM Algorithm. J. R. Stat. Soc. Ser. B 39, 1–38 (1977)
21. O’Hagan, A., Murphy, T.B., Gormley, E.C., et al.: Clustering with the multivariate normal inverse Gaussian distribution: Comput. Stat. Data Anal. 93, 18–30 (2016)
22. Metaxoglou, K., Smith, A.: Maximum likelihood estimation of VARMA models using a state-space EM algorithm. J. Time Series Anal. 28, 666–685 (2007)
23. Kim, J., Stoffer, D.S.: Fitting stochastic volatility models in the presence of irregular sampling via particle methods and the EM algorithm. J. Time Series Anal. 29, 811–833 (2008)
24. Meng, X.L., Rubin, D.B.: Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80, 267–278 (1993)
25. Liu, C.H., Rubin, D.B.: ML estimation of $t$-distribution using EM and its extensions. ECM and ECME. Stat. Sin. 5, 19–39 (1995)
26. McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions, 2nd edn. Wiley, New Jersey (2007)
27. Devroye, L.: Non-Uniform Random Variate Generation, 1st edn. Springer, New York (1986)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.