Time Series Models With Asymmetric Laplace Innovations

A. Alexandre Trindade *　　Yun Zhu†　　Beth Andrews‡

May 18, 2009

Abstract

We propose autoregressive moving average (ARMA) and generalized autoregressive conditional heteroscedastic (GARCH) models driven by Asymmetric Laplace (AL) noise. The AL distribution plays, in the geometric-stable class, the analogous role played by the normal in the alpha-stable class, and has shown promise in the modeling of certain types of financial and engineering data. In the case of an ARMA model we derive the marginal distribution of the process, as well as its bivariate distribution when separated by a finite number of lags. The calculation of exact confidence bands for minimum mean-squared error linear predictors is shown to be straightforward. Conditional maximum likelihood-based inference is advocated, and corresponding asymptotic results are discussed. The models are particularly suited for processes that are skewed, peaked, and leptokurtic, but which appear to have some higher order moments. A case study of a fund of real estate returns reveals that AL noise models tend to deliver a superior fit with substantially less parameters than normal noise counterparts, and provide both a competitive fit and a greater degree of numerical stability with respect to other skewed distributions.

Keywords: ARMA, GARCH, conditional maximum likelihood, joint distribution, prediction, financial returns

1 Introduction

The classical linear autoregressive moving average (ARMA) model has enjoyed widespread popularity in the time series literature. A theoretical justification for its applicability is in part provided by The Wold Decomposition, which states that a stationary process can essentially be expressed as a linear combination of current and past values of a serially uncorrelated (white noise) sequence.

Distributional assumptions on the noise or innovations process, although not needed for optimal point forecasts, are a prerequisite for accurate confidence bands (see section 2). For estimation one can by default maximize a Gaussian likelihood, resulting in consistent

*To whom correspondence should be addressed at: Texas Tech University, Department of Mathematics & Statistics, Lubbock, TX 79409-1042, U.S.A. Phone: (806)742-2580 ext 233, Fax: (806)742-1112, E-mail: alex.trindade@ttu.edu.

†CitiGroup, Consumer Analytics & Modeling Unit, O'Fallon, Missouri.

‡Department of Statistics, Northwestern University, Evanston, Illinois. Supported in part by NSF Grant DMS0806104.

and asymptotically normal parameter estimates by assuming only that the the noise is independent and identically distributed (IID), and not necessarily normal (e.g. Brockwell and Davis, 1991). However, this does not in general result in (asymptotically) efficient estimates, unless the noise is itself Gaussian. There can therefore be some gains by more carefully specifying the marginal and joint distributions of the process, or equivalently of the noise process. This is particularly true of financial data, which tend to exhibit heavy tails and sometimes skewness, neither of which is consistent with a Gaussian distribution.

There has been some interest in the literature, mostly in engineering but more recently also in finance, in using the Laplace and related distributions in data modeling contexts that involve time. Davenport (1952) modeled speech waves using the Laplace distribution. McGill (1962) showed that the Laplace distribution provides a characterization of the error in a timing device that is under periodic excitation. Hsu (1979) finds that navigation errors for aircraft position are best fitted by a mixture of two Laplace distributions. Damsleth and El-Shaarawi (1989) employ an ARMA model driven by Laplace noise to fit weekly data on sulphate concentration in a Canadian watershed. Anderson and Arnold (1993) observed that IBM daily stock price returns are adequately modeled by Linnik processes. The Linnik (also known as α-Laplace) is a symmetric distribution supported on the real line and parameterized by a tail index $\alpha \in (0, 2]$. For $\alpha = 2$ the Linnik coincides with the Laplace. Thomas and Jayakumar (2003) discuss a generalized Linnik distribution and process.

The efforts in using the Laplace distribution in ARMA models have focused on two primary directions. The NLAR(1) and NLAR(2) models of Dewald and Lewis (1985) and the NAREX(1) model of Novković (1998), assume a marginal Laplace distribution and find the noise process to be a mixture of Laplace densities. Similar results are obtained by Jayakumar and Kuttykrishnan (2007), who assume an asymmetric Laplace marginal distribution. On the other hand, Damsleth and El-Shaarawi (1989) start with Laplace noise and find the marginal distribution of the process to be a linear combination of Laplace distributions. Damsleth and El-Shaarawi (1989) also show that the requirements of having Laplace distributions for both the marginal and noise processes, cannot be simultaneously achieved within the class of linear time series models. Other notable ARMA modeling attempts at incorporating Laplace-like distributions include the Linnik processes of Anderson and Arnold (1993) which can be viewed as random coefficient autoregressive (AR) schemes, and the AR processes of Lekshmi and Jose (2004) which have geometric α-Laplace marginals.

Recently, Kotz, Kozubowski, and Podgórski (2001) introduced a generalization of the symmetric Laplace location-scale family of distributions that allows for skewness. They find that this *asymmetric Laplace* (AL) distribution is leptokurtic (its kurtosis exceeds 3) and shows promise in the modeling of data from fields as diverse as engineering, finance, astronomy, and the biological and environmental sciences. We sometimes use the term "heavy-tailed" when referring to this and similar distributions whose kurtosis exceeds 3 but, since all its moments are finite, the AL is not heavy-tailed in the sense that $P(|X| > x) \sim x^{-\alpha}$, for some $\alpha > 0$ as $x \to \infty$, whence $\mathbb{E}|X|^{\delta} = \infty$ for $\delta > \alpha$. Its rich structure also allows natural extensions to stable laws, commonly used in financial applications.

Stable distributions enjoy the property of stability with respect to scaled sums. In the case of the *stable* (Paretian stable or α-stable) distribution Y, this defining property is that for a sequence $Y_1, Y_2, \ldots$ of IID copies of Y, there exist sequences $a_n > 0$ and b_n such that

$Y \stackrel{d}{=} a_n(Y_1 + \cdots + Y_n) + b_n$ for all n. These distributions form a location-scale family, with a parameter controlling skewness and an exponent $\alpha \in (0, 2]$ governing the heaviness of the tail. The normal is the only non-degenerate stable distribution with a finite variance ($\alpha = 2$). See e.g. Samorodnitsky and Taqqu (1994) for more on the stable distribution.

Geometric stable distributions are similar to Paretian stable, but the defining stability property is now with respect to the limit of scaled random sums: For a sequence $Y_1, Y_2, \ldots$ of IID copies of Y, there exist sequences $a_p > 0$ and b_p such that $a_p(Y_1 + \cdots + Y_{G_p} + G_p b_p) \xrightarrow{d} Y$ as $p \to 0$, where G_p is geometric with mean $1/p$ and is independent of the Y_i's (Kozubowski, 1994). These distributions also form a location-scale family, with a parameter controlling skewness and a tail index parameter $\alpha \in (0, 2]$ governing the tail behavior. In fact, the Geometric stable can be viewed as a Paretian stable with random location and scale parameters (Kozubowski and Rachev, 1994). Analogous to the normal, the AL is the only non-degenerate Geometric stable distribution with a finite variance ($\alpha = 2$). See Kotz *et al.* (2001) for more on Geometric stable laws.

It is commonly accepted that stable distributions provide useful models for certain types of financial data like asset returns. Recent work by Andrews, Calder, and Davis (2009) suggests it may also be a plausible model for traded stock volume. They propose a method of maximum likelihood estimation for autoregressive processes and fit a stable AR(2) model to Wal-Mart stock volume. To accommodate the possibility of market crashes, Kozubowski and Rachev (1994) preserve the stability of stock price changes up to a geometrically-distributed random time, leading to the class of Geometric stable laws. They go on to empirically compare the fit of several distributions to various financial returns data, and find that the Geometric stable provides a better fit than several others, including the Paretian stable. Hence, although both Paretian and Geometric stable distributions are handicapped by a lack of explicit expressions for densities and distribution functions, the latter can provide a better fit if the data are both heavy-tailed and peaked.

In Trindade and Zhu (2007) we approximated the sampling distributions of estimators of financial risk under IID sampling from the AL distribution, and found the AL to be a good (marginal) model for currency exchange rate returns. Our goal in the present paper is to extend that work to a dependent data setting, by investigating time series models where the AL distribution is the driving white noise (or innovations) process. To this end, in section 2 we consider ARMA models driven by AL noise, and derive both the marginal and bivariate distributions of the process. Parameter estimation via conditional maximum likelihood, and prediction of the process are also discussed. Section 3 extends this to ARMA models driven by GARCH AL noise, where parameters are again estimated by maximizing the conditional likelihood. Simulations in section 4 shed light on the quality of the finite-sample estimates. We end in section 5 with an illustration and comparison of the models applied to a data set of real estate returns.

2 ARMA Models With Asymmetric Laplace Noise

In this section we will consider ARMA models driven by IID AL noise. We will first derive the marginal distribution of the process, and give explicit expressions for its cumulative distribution function (CDF) and probability density function (PDF). We will then consider the joint distribution of the process and discuss the prediction problem. Finally, we will propose a parameter estimation approach.

2.1 Marginal distribution of the process

The AL location-scale family of distributions recently introduced by Kotz *et al.* (2001) is a welcome extension of the ordinary (symmetric) Laplace that already enjoys widespread popularity in applications from fields as diverse as engineering, finance, astronomy, and the biological and environmental sciences. Random variable Y is AL distributed with location parameter θ, scale parameter $\tau > 0$, and skewness parameter $\kappa > 0$, $Y \sim \mathcal{AL}(\theta, \kappa, \tau)$, if its PDF is of the form

$$f(y; \theta, \kappa, \tau) = \frac{\kappa\sqrt{2}}{\tau(1+\kappa^2)} \exp\left\{ -\operatorname{sgn}(y-\theta)\frac{\sqrt{2}}{\tau}\kappa^{\operatorname{sgn}(y-\theta)}(y-\theta) \right\}.$$

We have expressed the PDF here in a form that is more suitable for programming, rather than the original version. Its mode, mean, and variance are respectively, θ, $\mu = \theta + \tau(\kappa^{-1} - \kappa)/\sqrt{2}$, and $\sigma^2 = (\mu - \theta)^2 + \tau^2$. Values of κ in the intervals $(0, 1)$ and $(1, \infty)$, correspond to positive (right) and negative (left) skewness, respectively. The (adjusted) kurtosis of an AL distribution varies between 3 (the least value for the symmetric Laplace distribution when $\kappa = 1$) and 6 (the largest value attained for the limiting exponential distribution as $\kappa \to 0$). By contrast the normal has a kurtosis of 0, and the AL is therefore leptokurtic.

For density derivation and simulation purposes, an important property of $Y \sim \mathcal{AL}(\theta, \kappa, \tau)$, is that it admits the representation

$$Y \stackrel{d}{=} \theta + \tau(E_1/\kappa - \kappa E_2)/\sqrt{2}, \tag{1}$$

where E_1 and E_2 are IID standard exponential random variables. Thus the moment generating function of an AL exists, and consequently all its moments are finite.

Definition 1 (ARMA model with AL noise) *Let $Z_t \sim$ IID $\mathcal{AL}(\theta, \kappa, \tau)$, with $\theta = -\tau(\kappa^{-1} - \kappa)/\sqrt{2}$. Then $\{X_t\}$ is an ARMA(p, q) process driven by AL noise if it is a stationary solution of the equations*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \lambda_1 Z_{t-1} + \cdots + \lambda_q Z_{t-q} + Z_t, \tag{2}$$

where the polynomials $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\lambda(z) = 1 + \lambda_1 z + \cdots + \lambda_q z^q$ have no common factors.

Note that a stationary solution exists if and only if $\phi(z) \neq 0$ for $|z| = 1$. In this case, the Laurent series expansion of $1/\phi(z)$, $1/\phi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$, exists on some annulus $\{z : a^{-1} < |z| < a\}$, $a > 1$, and the unique stationary solution to (2) is given by $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ (Brockwell and Davis, 1991, Chap. 3). The ψ_j coefficients decay at a roughly geometric rate as $j \to \pm\infty$ and are therefore absolutely summable.

The specification

$$\theta = -\tau(\kappa^{-1} - \kappa)/\sqrt{2}, \tag{3}$$

ensures that Z_t has zero mean, as is typically required in the definition of an ARMA model. Other specifications, such as Z_t having zero median, would also make sense. For the rest of the paper, the notation $f(z_t; \kappa, \tau)$ will denote the PDF of Z_t as in Definition 1.

The model in this definition has the flexibility to describe a stationary autocorrelated process with asymmetric excursions about the mean. The allowance for asymmetry is an important requirement in many modeling scenarios. In finance for example, the cost to buy a call option is limited, but the return can be substantial. An improvement in credit quality brings limited returns to investors, but in case of defaults or downgrades the loss could be large. The model also has the flexibility to describe data with heavy tails. Figure 1 displays three simulated realizations from the MA(1) process,

$$X_t = 0.8Z_{t-1} + Z_t, \qquad \{Z_t\} \sim \text{IID } W,$$

with W a standard normal distribution in the left panels, $W \sim \mathcal{AL}(0, 1, 1)$ (a symmetric Laplace) in the middle panels, and $W \sim \mathcal{AL}(-0.73, 0.50, 0.69)$ in the right panels. In the last two cases the scale parameter τ was chosen so that the variance of Z_t is 1, thus coinciding with that from the first case. The histograms of the data displayed in the bottom panels confirm that the marginal distributions of the three processes are quite different. In the left panels, X_t is actually normal with mean 0 and variance 2.66. In the middle and right panels the distributions are clearly more peaked and (in the latter) skewed, and therefore look AL-like.

We now derive the marginal distribution of an ARMA(p, q) model driven by AL noise. In so doing, we will extend the results of Damsleth and EL-Shaarawi (1989) who derive this in the special case of symmetric Laplace noise. Their method is in turn an application of the method used by Box (1954) to derive the distribution of any linear combination of independent χ^2 random variables with even degrees of freedom. Our approach will be slightly different, and hinges upon representation (1). Although not strictly necessary for the result of this subsection, we will assume for simplicity of exposition that series (2) is *causal*, meaning that it can be expressed as a linear combination of lagged values of the noise series,

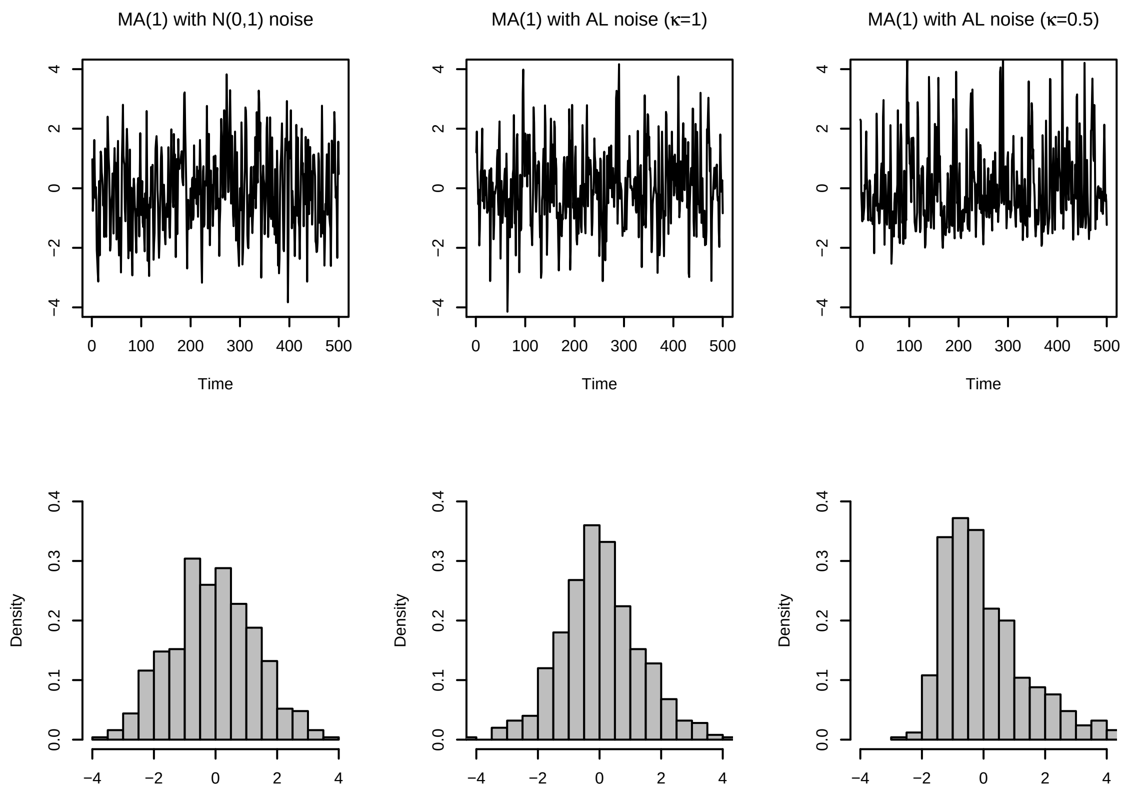
$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \tag{4}$$

for some absolutely summable sequence of constants $\{\psi_j\}$. We may sometimes also assume *invertibility*, i.e. $Z_t = \sum_{i=0}^{\infty} \pi_i X_{t-i}$ for some sequence of absolutely summable coefficients $\{\pi_i\}$. (Equivalently, the series is causal if and only if $\phi(z) \neq 0$ for $|z| \leq 1$, and invertible if and only if $\lambda(z) \neq 0$ for $|z| \leq 1$.) Causality and invertibility are standard assumptions in the time series literature, since non-causal and/or non-invertible models have second-order equivalent causal, invertible representations (Breidt and Davis, 1992).

The marginal CDF of X_t when only a finite number of the ψ_j's are nonzero, is now easily computed by appealing to the result concerning the distribution of a linear combination of exponential random variables derived by Ali and Obaidullah (1982).

Proposition 1 *Let* $X_t^{(m)} = \sum_{j=0}^{m} \psi_j Z_{t-j}$, *where* $Z_t, Z_{t-1}, \ldots, Z_{t-m}$ *are IID* $\mathcal{AL}(\theta, \kappa, \tau)$,

Figure 1: Simulated realizations from an MA(1) process driven by normal noise (left), AL noise with $\kappa = 1$ (middle), and AL noise with $\kappa = 0.5$ (right). Corresponding histograms of the data are shown in the bottom panels.



and the ψ_j *are distinct and nonzero for all* j*. Define* $\nu = \theta \sum_{j=0}^{m} \psi_j$*, the coefficients*

$$a_i = \begin{cases} \tau\psi_{i/2}/(\kappa\sqrt{2}), & \text{if } i \text{ is even,} \\ -\tau\psi_{(i-1)/2}\kappa/\sqrt{2}, & \text{if } i \text{ is odd,} \end{cases} \qquad A_i = \prod_{k=0,\ k\neq i}^{2m+1} (a_i - a_k), \qquad i = 0, 1, \ldots, 2m+1,$$

and the index sets $J_- = \{k = 0, 1, \ldots, 2m+1 \,|\, a_k < 0\}$*,* $J_+ = \{k = 0, 1, \ldots, 2m+1 \,|\, a_k > 0\}$*. Then the CDF of* $X_t^{(m)}$ *is given by*

$$F_{X_t^{(m)}}(x) = \begin{cases} \sum_{i \in J_-} a_i^{2m+1} A_i^{-1} \exp\{-(x-\nu)/a_i\}, & \text{if } x < \nu, \\ 1 - \sum_{i \in J_+} a_i^{2m+1} A_i^{-1} \exp\{-(x-\nu)/a_i\}, & \text{if } x \geq \nu, \end{cases} \tag{5}$$

and upon differentiation we obtain immediately its PDF

$$f_{X_t^{(m)}}(x) = \begin{cases} -\sum_{i \in J_-} a_i^{2m} A_i^{-1} \exp\{-(x-\nu)/a_i\}, & \text{if } x < \nu, \\ \sum_{i \in J_+} a_i^{2m} A_i^{-1} \exp\{-(x-\nu)/a_i\}, & \text{if } x \geq \nu. \end{cases} \tag{6}$$

The marginal distribution of X_t for an ARMA process can then be obtained by letting $m \to \infty$ in Proposition 1, whence we obtain that $X_t^{(m)}$ converges absolutely with probability one and in mean square to X_t (Brockwell and Davis, 1991, Prop. 3.1.1). However, the infinite sum in (4) can essentially be truncated after a few terms since (in general) the ψ_j's decay at a roughly geometric rate (Brockwell and Davis, 1991, Chap. 3). In the case of an MA(q) process, the summation is finite since $m = q$, and explicit closed-form expressions for the A_i's could in principle be derived. For example, if $q = 1$ routine computations give the following result.

Corollary 1 (Marginal distribution of MA(1) with AL noise) *With $\nu = (1+\lambda)\theta$, the marginal PDF of the MA(1) process $\{X_t\}$ in Definition 1, where the MA coefficient satisfies $\lambda > 0$ and $\lambda \neq 1$, is given by*

$$f_{X_t}(x) = \begin{cases} -b_2 \exp\left\{\frac{\sqrt{2}}{\tau\kappa}(x-\nu)\right\} + b_4 \exp\left\{\frac{\sqrt{2}}{\lambda\tau\kappa}(x-\nu)\right\}, & \text{if } x < \nu, \\ -b_1 \exp\left\{-\frac{\kappa\sqrt{2}}{\tau}(x-\nu)\right\} + b_3 \exp\left\{-\frac{\kappa\sqrt{2}}{\lambda\tau}(x-\nu)\right\}, & \text{if } x \geq \nu. \end{cases}$$

where

$$b_2 = \frac{\kappa^3\sqrt{2}}{\tau(1+\kappa^2)(\lambda+\kappa^2)(\lambda-1)}, \qquad b_4 = \frac{\lambda\kappa^3\sqrt{2}}{\tau(1+\kappa^2)(1+\lambda\kappa^2)(\lambda-1)},$$
$$b_1 = \frac{\kappa\sqrt{2}}{\tau(1+\kappa^2)(1+\lambda\kappa^2)(\lambda-1)}, \qquad b_3 = \frac{\lambda\kappa\sqrt{2}}{\tau(1+\kappa^2)(\lambda+\kappa^2)(\lambda-1)}.$$

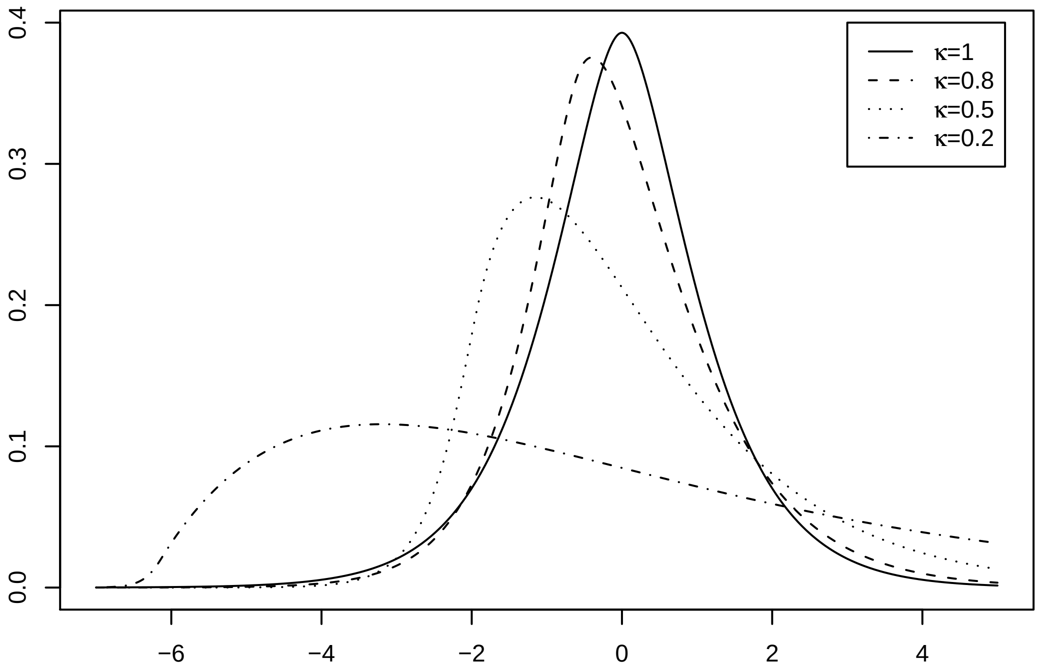
Some plots of these densities for the case $\lambda = 0.8$ and $\tau = 1$ appear in Figure 2. A similar expression for the PDF results when $\lambda < 0$. For $\lambda = 1$, $\psi_0 = 1 = \psi_1$, and a slightly different formula than the one presented in Proposition 1 can be similarly obtained following the algorithm of Ali and Obaidullah (1982). (This different form is to be used whenever there are coincident ψ_j's.) Note that the density of an MA(1) is not quite a mixture of AL densities; it would be a mixture of $\mathcal{AL}(\nu,\kappa,\tau)$ and $\mathcal{AL}(\nu,\kappa,\lambda\tau)$ if the pairs of normalizing constants (b_1, b_2) and (b_3, b_4) coincided (a fact which happens if and only if $\kappa = 1$). This result extends to a general ARMA(p,q) when $\kappa = 1$, as noted by Damsleth and EL-Shaarawi (1989): the marginal density is a mixture of AL densities, but although the mixing coefficients sum to unity, this is not necessarily a proper mixture since some of the coefficients may be negative.

Remark 1 *Since stationary, non-causal ARMA processes have an infinite-order moving average representation $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ with $\psi_j \to 0$ as $j \to \pm\infty$, a result similar to Proposition 1 can be obtained for truncated non-causal series.*

2.2 Joint distribution of the process

The joint distribution of a sequence of observations $(X_1, \ldots, X_n)$ from an ARMA model driven by AL noise would be much more difficult to derive. Damsleth and EL-Shaarawi (1989) compute the bivariate distribution of (X_{t-h}, X_t) for any lag h in the simpler special case of an AR(1) driven by (symmetric) Laplace noise. Similar results can be obtained in

Figure 2: Marginal density functions from an MA(1) process with coefficient $\lambda = 0.8$ driven by AL noise with scale parameter $\tau = 1$, for various settings of the skewness parameter κ.



our setting on a case by case basis. For example consider the ARMA(1,1) with AL noise, $X_t = \phi X_{t-1} + \lambda Z_{t-1} + Z_t$. By iterating this expression $h-1$ times ($h \geq 1$) we obtain,

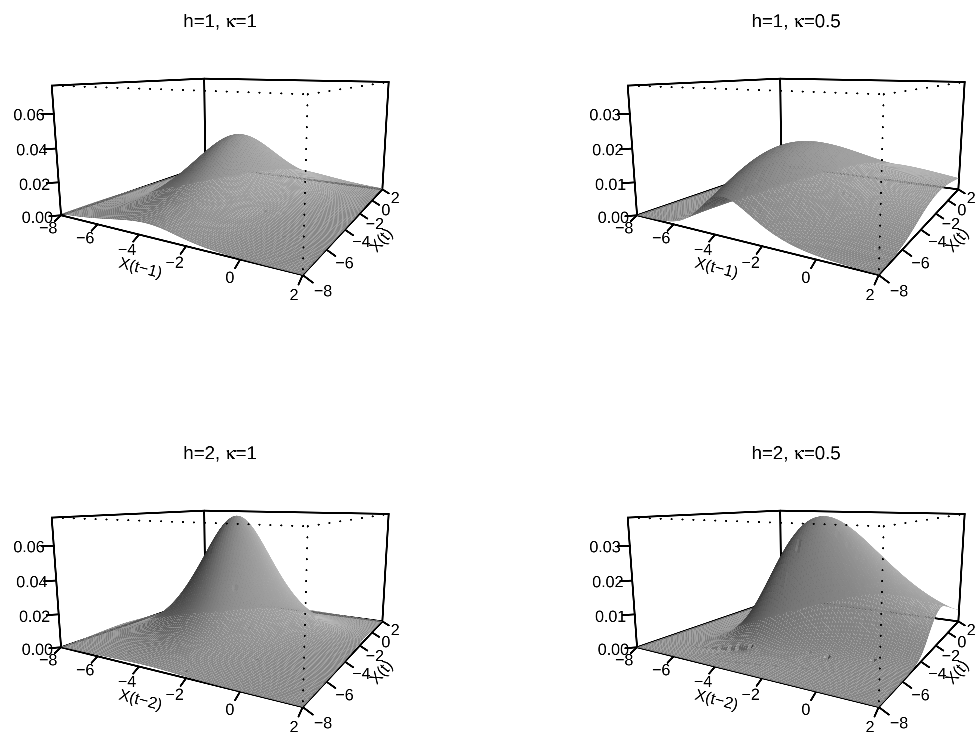
$$X_t = \phi^h X_{t-h} + \lambda\phi^{h-1} Z_{t-h} + \sum_{j=0}^{h-1} \psi_j Z_{t-j}, \qquad \text{where } \psi_j = \begin{cases} 1, & j = 0, \\ (\phi+\lambda)\phi^{j-1}, & j \geq 1. \end{cases}$$

Define $W_1 = \sum_{j=0}^{h-1} \psi_j Z_{t-j}$, $W_2 = Z_{t-h}$, and $W_3 = \sum_{j=1}^{\infty} \psi_j Z_{t-h-j}$, and note that the W_i are independent. This means the joint PDF of (W_1, W_2, W_3) is straightforward to derive from Proposition 1, since each W_i is a linear combinations of AL random variables. Now, $X_{t-h} = W_2 + W_3$ and $X_t = W_1 + aW_2 + bW_3$, where $a = \phi^h + \lambda\phi^{h-1}$ and $b = \phi^h$. This defines a one-to-one transformation: $(W_1, W_2, W_3) \mapsto (X_{t-h}, X_t, W_1)$, with Jacobian $(b-a)^{-1}$. The joint PDF of (X_{t-h}, X_t) can then be obtained by (numerically) integrating the joint PDF of (X_{t-h}, X_t, W_1) over W_1, yielding

$$f_{X_{t-h},X_t,}(x,y) \quad = \quad \frac{1}{|b-a|} \int_{-\infty}^{\infty} f_{W_1}(w) f_{W_2}\left(\frac{bx-y+w}{b-a}\right) f_{W_3}\left(\frac{y-w-ax}{b-a}\right) dw.$$

Figure 3 illustrates some of the shapes obtained for the joint PDF of (X_{t-h}, X_t) when $\phi = 0.5$ and $\lambda = 0.8$. The value of h is 1 and 2 in the top and bottom panels, respectively, while κ is 1 and 0.5 in the left and right panels, respectively. There is a moderate amount of correlation in the panels corresponding to $h = 1$, since the model autocorrelation at lag 1 is $\rho(1) = 0.75$. For lag $h = 2$, the model autocorrelation is only $\rho(2) = 0.37$,

Figure 3: Bivariate density function of (X_{t-h}, X_t) from the ARMA(1,1) model $X_t - 0.5X_{t-1} = 0.8Z_{t-1} + Z_t$, with $Z_t \sim$ IID $\mathcal{AL}(\theta, \kappa, \tau = 1)$.



2.3 Prediction of the process

We consider the classical best linear h-step ahead (finite past) prediction problem, i.e. the linear function of observations $X_1, \ldots, X_n$ from a causal invertible ARMA driven by IID AL noise that minimizes the mean squared prediction error. From (4), it is clear that the resulting predictor will itself be of the form of (4), whence its distribution can be readily derived as in the preceding sections.

To simplify the exposition, we will focus on the h-step ahead best linear predictor (BLP), $\tilde{X}_{n+h}$, based on the infinite past. Then, standard results from, say, Brockwell and Davis

(1991), gives $\tilde{X}_{n+h} = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j}$, whence

$$X_{n+h} - \tilde{X}_{n+h} = \sum_{j=0}^{h-1} \psi_j Z_{n+h-j} := \tilde{E}_h, \quad \text{and,} \quad \mathbb{E}(X_{n+h} - \tilde{X}_{n+h})^2 = \sigma_Z^2 \sum_{j=0}^{h-1} \psi_j^2 := \tilde{\sigma}_h^2,$$

where σ_Z^2 is the variance of (the noise) Z_t. As is well known, the BLP on the infinite past coincides with the *best* predictor on the infinite past, $\mathbb{E}(X_{n+h}|X_n, X_{n-1}, \ldots)$. In applications, if n is large the difference between the BLPs on the infinite vs. finite pasts is negligible.

If $\tilde{E}_{h,\alpha}$ denotes the αth quantile of $\tilde{E}_h$, then we have

$$P(\tilde{X}_{n+h} + \tilde{E}_{h,\alpha/2} < X_{n+h} < \tilde{X}_{n+h} + \tilde{E}_{h,1-\alpha/2}) = 1 - \alpha,$$

whence the end terms of the inequality can be taken to be $(1-\alpha)100\%$ prediction bounds. If $Z_t \sim \mathrm{N}(0, \sigma_Z^2)$, then $\tilde{E}_{h,\alpha} = \Phi_\alpha \tilde{\sigma}_h$, where Φ_α is the αth quantile of a standard normal. If $Z_t \sim \mathcal{AL}(\theta, \kappa, \tau)$, then $\tilde{E}_h$ is a linear combination of AL distributions, whence its quantiles can be readily obtained from Proposition 1.

2.4 Parameter estimation

The fact that computation of just the joint distribution of (X_s, X_t) requires numerical integration means that, with the goal of maximum likelihood estimation in mind, it is generally infeasible to pursue a full exact likelihood approach. *Conditional* maximum likelihood estimation on the other hand is quite straightforward to implement, and we pursue this approach instead. The resulting conditional maximum likelihood estimates, sometimes called quasi-maximum likelihood estimates (QMLEs), are essentially indistinguishable from (unconditional) maximum likelihood estimates (MLEs) for large sample sizes. The reason for this is clear from (7) below; for MLEs the summation starts at $t = 1$.

Assuming the ARMA(p, q) model of Definition 1 is both causal and invertible, the log-likelihood of $\{X_1, \ldots, X_n\}$ conditional on $\{X_1, \ldots, X_p\}$ is given by (Li and McLeod, 1988):

$$l(\phi_1, \ldots, \phi_p, \lambda_1, \ldots, \lambda_q, \kappa, \tau) = \sum_{t=p+1}^{n} \log f(z_t; \kappa, \tau), \tag{7}$$

where the z_t's represent model residuals which can be computed from (2) using the observations $X_1, \ldots, X_n$, with $z_t = \mathbb{E}Z_t = 0$, for $t = \min(p - q + 1, p), \ldots, p$. QMLEs are then obtained by maximizing (7).

Results concerning consistency and asymptotic normality of ARMA model QMLEs for a general noise density $f(\cdot)$ is derived under fairly mild assumptions by Li and McLeod (1988). One of these assumptions requires the (almost everywhere) existence and continuity of first and second order derivatives of $f(z; \kappa, \tau)$ with respect to both z and the parameters. This does not however present a problem since the AL density is non-differentiable only at $z = \theta$. The zero-mean restriction (3) on estimation necessitates some care when computing the last 2×2 block of the Fisher Information matrix corresponding to the AL noise contribution. Details are provided in the appendix.

3 ARMA Models With GARCH Asymmetric Laplace Noise

When linear models are not appropriate, nonlinear time series models such as bilinear models, random coefficient autoregressive models, and threshold models, are possible alternatives. The nonlinear class of AutoRegressive Conditionally Heteroscedastic (ARCH) models was introduced by Engle (1982) upon observing that the volatility of certain series depends on the past; a common occurrence with financial data. This class was generalized by Bollerslev (1986) to the GARCH process. In this section we will consider ARMA models driven by AL GARCH noise, the definition of which is as follows.

Definition 2 (ARMA model with AL GARCH noise) *Let $e_t \sim$ IID $\mathcal{AL}(\theta, \kappa, \tau)$, with $\theta = -\tau(\kappa^{-1} - \kappa)/\sqrt{2}$ and $\tau^2 = 2[2 + (\kappa^{-1} - \kappa)^2]^{-1}$. Then $\{X_t\}$ is an ARMA(p, q) process driven by GARCH(u, v) AL noise if it is a stationary solution of the equations*

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \lambda_j Z_{t-j} + Z_t,$$

and

$$Z_t = \sigma_t e_t, \tag{8}$$

where σ_t is a positive function of $Z_s, s < t$, defined by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2, \tag{9}$$

with $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, u$, and $\beta_j \geq 0$, $j = 1, \ldots, v$.

The specified values of θ and τ ensure that $\{e_t\}$ is a zero-mean and unit variance series, as is commonly stipulated in the formulation of GARCH models.

The ARMA-GARCH model has the property that the conditional mean and variance are given by,

$$\mathbb{E}[X_t | Z_s, s < t] = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \lambda_j Z_{t-j} := \mu_t,$$

and $\mathrm{Var}[X_t | Z_s, s < t] = \sigma_t^2$, respectively. Derivation of the marginal distribution of just the GARCH process Z_t (let alone X_t) is intractable already in the case of $e_t \sim$ IID Gaussian, and thus will not be attempted here under the more complicated AL model. Similar statements hold for the distributional properties of predictors.

Model fitting can be accomplished by maximizing the joint conditional likelihood as described in section 2. This is more commonly known as quasi-maximum likelihood estimation (QMLE) in the GARCH literature. The joint conditional likelihood of an ARMA-GARCH model is

$$L(\phi_1, \ldots, \phi_p, \lambda_1, \ldots, \lambda_q, \alpha_0, \ldots, \alpha_u, \beta_1, \ldots, \beta_v, \kappa) = \prod_{t=p+1}^{n} \frac{1}{\sigma_t} f\left(\frac{x_t - \mu_t}{\sigma_t}; \kappa\right),$$

where $f(\cdot; \kappa)$ is the PDF of e_t. Standard deviations σ_t, $t \geq 0$, can be computed recursively from (8) and (9), with $Z_t = 0$ and $\sigma_t^2 = \hat{\sigma}^2$ for all $t \leq 0$, where $\hat{\sigma}^2$ is the sample variance of the GARCH residuals $\{\hat{Z}_1, \ldots, \hat{Z}_n\}$. As before, this conditional likelihood function can be maximized via numerical optimization methods.

Asymptotic results for Gaussian QMLEs of ARMA-GARCH model parameters are nicely summarized by Francq and Zakoian (2004), who establish these under weaker conditions than previous authors. To the best of our knowledge, no extensions have yet been worked out for non-Gaussian QMLEs. Berkes and Horvath (2004) do however consider ML and QML estimation for the parameters of pure GARCH processes under a variety of distributional assumptions, and one possibility would be to extend that work to ARMA-GARCH processes with AL innovations. Another interesting direction would be to consider the least absolute deviations estimators for GARCH model parameters introduced in Peng and Yao (2003), and use those for ARMA-GARCH parameter estimation. The least absolute deviations estimators are robust in the sense that they are asymptotically normal and $n^{1/2}$-consistent under more general conditions than Gaussian MLEs, and they are more efficient for processes with a heavy-tailed noise distribution.

Notwithstanding the lack of more explicit asymptotic results for ARMA-GARCH models, appropriate regularity conditions guarantee the usual consistency and asymptotic normality of MLEs with covariance matrix equal to the inverse of the Fisher Information. If analytical expressions for the latter prove to be intractable, standard errors for all parameter estimates can nevertheless be obtained by inverting the (numerically evaluated) Hessian matrix, H, and taking the square root of the appropriate diagonal entries. This is in fact an option available in many software packages, e.g. "S+FinMetrics" of Splus and "fGarch" of R. If the model is correctly specified this provides a reasonable alternative to explicit evaluation of the Information matrix. An alternative that is robust to model misspecification is to calculate standard errors based on the *sandwich estimator* of the covariance matrix, $H^{-1}\mathbf{g}\mathbf{g}^T H^{-1}$, where $\mathbf{g}$ denotes the gradient vector, but the increase in robustness is gained at the expense of efficiency. For details see Davidson and MacKinnon (2004, Chapt.10).

4 Simulations

In this section we present a small simulation study to assess the quality of the QMLEs from ARMA and ARMA-GARCH models with AL innovations. The PDF, CDF, quantiles, and random numbers from an AL distribution are available from the R package "VGAM". At the time of writing there does not seem to be any package for fitting ARMA models with AL noise. The estimates in Tables 1 and 2 below were thus obtained with our own program written in Matlab.

The ARMA-GARCH estimates in Tables 3 and 4 were obtained with the R package "fGarch" developed by Rmetrics (www.rmetrics.org). Skewness is introduced into the normal, generalized error (also known as exponential power), and t distributions, via the method of Fernandez and Steel (1998), and thus the generalized error distribution with shape parameter $\nu = 1$ (and skewness κ) coincides with the AL. Standard errors are based on the numerically evaluated Hessian. The package is nicely documented in a forthcoming paper by Würtz, Chalabi, and Luksan.

Tables 1 and 2 list the fitted parameters of the first five simulations from an ARMA(1,1) and an AR(2) model under IID AL noise. The calculation of means and mean squared errors is based on 100 simulations. The sample size in each simulation is $n = 100$.

Table 1: Estimated parameter values for the first 5 of 100 simulations from an ARMA(1,1) model under AL noise with a sample size of $n = 100$. Means and mean squared errors (MSE) are for all 100 simulations.

Parameters	κ	τ	ϕ	λ
True values	0.8	1	0.7	0.5
simulation 1	0.740	0.921	0.760	0.505
simulation 2	0.815	0.984	0.711	0.481
simulation 3	0.828	1.029	0.712	0.477
simulation 4	0.836	0.998	0.642	0.606
simulation 5	0.742	1.019	0.724	0.435
Mean	0.807	0.985	0.689	0.500
MSE	0.008	0.011	0.006	0.009

Table 2: Estimated parameter values for the first 5 of 100 simulations from an AR(2) model under AL noise with a sample size of $n = 100$. Means and mean squared errors (MSE) are for all 100 simulations.

Parameters	κ	τ	ϕ_1	ϕ_2
True values	0.8	1	0.7	-0.1
simulation 1	0.671	0.930	0.801	-0.176
simulation 2	0.674	0.749	0.749	-0.166
simulation 3	0.906	1.042	0.855	-0.200
simulation 4	0.860	1.065	0.710	-0.083
simulation 5	0.692	0.890	0.825	-0.105
Mean	0.789	0.975	0.697	-0.105
MSE	0.008	0.014	0.009	0.006

These simulations were repeated for the same ARMA models, but now driven by AL GARCH(1,1) noise (Tables 3 and 4). The ARMA(1,1) had GARCH parameter values of $\alpha_0 = 0.1$, $\alpha_1 = 0.3$, $\beta_1 = 0.2$, whereas the AR(2) had $\alpha_0 = 0.2$, $\alpha_1 = 0.4$, $\beta_1 = 0.5$.

Table 3: Estimated parameter values for the first 5 of 100 simulations from an ARMA(1,1) model under AL GARCH(1,1) noise with a sample size of $n = 100$. Means and mean squared errors (MSE) are for all 100 simulations.

Parameters	κ	ϕ	λ	α_0	α_1	β_1
True values	0.8	0.7	0.5	0.1	0.3	0.2
simulation 1	0.954	0.535	0.682	0.412	0.020	0.005
simulation 2	0.930	0.740	0.547	0.425	0.189	0.000
simulation 3	0.954	0.612	0.602	0.294	0.308	0.456
simulation 4	0.964	0.652	0.486	0.354	0.027	0.000
simulation 5	0.942	0.756	0.402	0.408	0.000	0.000
Mean	0.926	0.681	0.504	0.339	0.223	0.102
MSE	0.032	0.016	0.021	0.076	0.047	0.056

Table 4: Estimated parameter values for the first 5 of 100 simulations from an AR(2) model under AL GARCH(1,1) noise with a sample size of $n = 100$. Means and mean squared errors (MSE) are for all 100 simulations.

Parameters	κ	ϕ_1	ϕ_2	α_0	α_1	β_1
True values	0.8	0.7	-0.1	0.2	0.4	0.5
simulation 1	0.899	0.477	-0.088	0.159	0.529	0.342
simulation 2	0.827	0.831	-0.247	1.146	0.000	0.000
simulation 3	0.648	0.749	-0.260	0.028	0.740	0.323
simulation 4	0.904	0.667	-0.137	0.817	0.000	0.007
simulation 5	0.494	0.610	-0.020	0.073	0.685	0.288
Mean	0.818	0.685	-0.112	0.417	0.376	0.256
MSE	0.025	0.013	0.018	0.193	0.099	0.113

We can see that reasonable estimates are obtained when a numerical optimization method is used to maximize the conditional likelihood. There seems to be more bias in the GARCH parameter estimates.

5 Application: Modeling TIAA-CREF Real Estate Returns

In this section we present a case study to illustrate the proposed methodology. We analyze the returns of the real estate variable annuity account managed by the Teachers Insurance and Annuity Association - College Retirement Equities Fund (TIAA-CREF), based in New

York City. This account seeks favorable long-term returns primarily through rental income and appreciation of real estate investments. The data we consider range from June 15, 2004 to December 31, 2006; a total of 650 daily values. Returns were computed by taking differences of successive log values i.e. by differencing the natural logarithm values at lag 1.

The plots in Figure 4 indicate the data is peaked, heavy-tailed, and right-skewed. There is also a persistent autocorrelation (ACF) that exhibits a somewhat periodic behavior, possibly due to a monthly seasonal effect. This is confirmed by the periodogram which has a moderate peak at a period of about 21 (trading) days. A suspected cause may be the accummulation of monthly rental income arriving at the end of each month. However, some large returns are also observed throughout the month. The fact that the calendar end-of-month may not occur on a trading day (and the number of days in a month varies), may be part of the reason that standard attempts at de-seasonalizing, such as differencing at lags around 21, do not work here. With enough expert knowledge about the account one could conceivably construct a suitable set of regressors that may successfully extract such deterministic features (and possibly any remaining autocorrelation along with it). In the spirit of illustrating our methodology, we will proceed by attempting to fit stationary ARMA models directly to the returns without any further pre-processing.

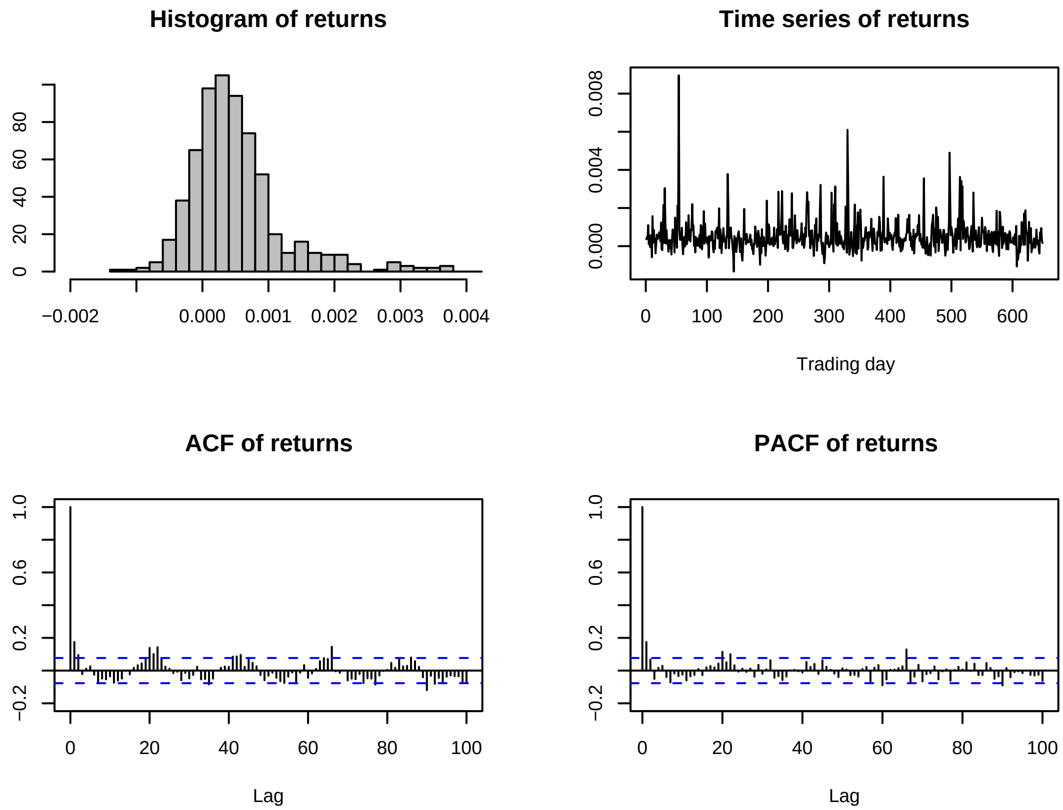
We first searched for the best-fitting ARMA models driven by IID AL noise, and by IID normal noise. Our criterion for *best* is the Akaike Information Criterion (AIC), coupled with a check for lack of serial correlation in the resulting residuals. Our version of AIC is the classical one, -2(log likelihood)+2(number of parameters), see e.g. Brockwell and Davis (1991). We searched the space of all ARMA(p, q) models, up to maximum values of $p = 20 = q$. The AL and normal noise models were obtained by maximizing respectively, the conditional and full likelihoods. Parameter estimates for the resulting best model with IID AL noise, an MA(1) with AIC=-7704, appear in Table 5. The best model with IID normal noise was an ARMA(5,4) with AIC=-7391 (estimates not shown).

Table 5: Estimated parameters (and standard errors) for 4 MA(1) models fitted to the TIAA-CREF returns with the following noise structures: IID Asymmetric Laplace (IID AL), GARCH(1,1) Asymmetric Laplace (GCH AL), GARCH(1,1) Asymmetric Normal (GCH AN), GARCH(1,1) Asymmetric t (GCH AT). (NaN=undefined.)

Noise Structure	Parameter Estimates (Standard Errors) λ	α_0	α_1	β_1	κ	τ or ν
IID AL	6.21E-02				6.64E-01	τ=6.31E-04
	(2.38E-02)				(3.68E-02)	(4.28E-05)
GCH AL	7.32E-02	4.97E-07	8.53E-02	1.00E-08	1.50E+00	
	(1.51E-02)	(4.96E-08)	(4.22E-02)	(9.19E-02)	(2.01E-02)	
GCH AN	9.32E-02	2.97E-07	2.63E-01	2.24E-01	1.80E+00	
	(4.65E-02)	(4.62E-08)	(8.71E-02)	(1.02E-01)	(1.03E-01)	
GCH AT	8.88E-02	6.59E-07	1.04E-01	1.00E-08	1.52E+00	ν=3.06E+00
	(2.97E-02)	(1.19E-07)	(6.21E-02)	(NaN)	(8.62E-02)	(4.19E-01)

The first panel of Figure 5 shows a relative frequency histogram of the mean-corrected

Figure 4: TIAA-CREF real estate returns data. The bottom panels show the sample autocorrelation (ACF) and partial autocorrelation (PACF) functions.



returns, with the marginal PDFs of the best-fitting ARMA with AL (solid) and normal (dashed) noises superimposed. The ARMA with normal noise naturally fails to capture the skewness and peakedness, while the fit of the AL noise model is remarkable.

As is typical of financial data, although the residuals from the ARMA model with normal noise are uncorrelated, a look at the ACFs of the squares and (especially) the absolute values of the residuals suggests dependence. The situation is somewhat different for the residuals from the MA(1) model with AL noise. In fact the returns have been (more) successfully whitened by the MA(1) filter, and there is little evidence of dependence present in the ACFs of the squares and absolute values.

To better capture the dependence we entertained the same ARMA(5,4) and MA(1) models, but now driven by GARCH(1,1) noise. The GARCH was fit using several distributions for the noise e_t: normal for the ARMA(5,4), and asymmetric versions of the normal (AN), Laplace (AL), and Student-t (AT), for the MA(1). Parameter estimates for the four MA(1) models thus fitted appear in Table 5. The ARMA model estimates were obtained with our own Matlab program; standard errors being based on the observed Fisher Information derived in the Appendix. The ARMA-GARCH model estimates were obtained with the R package "fGarch". Here standard errors are based on the numerically evaluated Hessian, and we note that there are problems with the estimate of β_1 for the GARCH(1,1) driven by asymmetric t noise (its standard error is either negative or beyond the limits of numerical precision carried by the software).

Table 6 summarizes the results in terms of the AIC values obtained with each model considered (parameter estimates for most of these are in Table 5). The lowest AIC model was the MA(1) with GARCH(1,1) Asymmetric Student-t noise (AIC=-7724), but there were numerical problems with the estimation as mentioned above. The next lowest were the MA(1)-GARCH(1,1) AL and MA(1)-IID AL, both with an AIC of essentially -7704. The remaining models are not competitive in terms of AIC.

Table 6: AIC values obtained from different models for the conditional mean and variance structures fitted to the TIAA-CREF returns.

Conditional Mean	Conditional Variance	AIC
ARMA(5,4)	IID Normal	-7391.3
ARMA(5,4)	GARCH(1,1) Normal	-7465.1
MA(1)	GARCH(1,1) Asymmetric Normal	-7578.4
MA(1)	IID Asymmetric Laplace	-7703.7
MA(1)	GARCH(1,1) Asymmetric Laplace	-7704.2
MA(1)	GARCH(1,1) Asymmetric Student-t	-7724.2

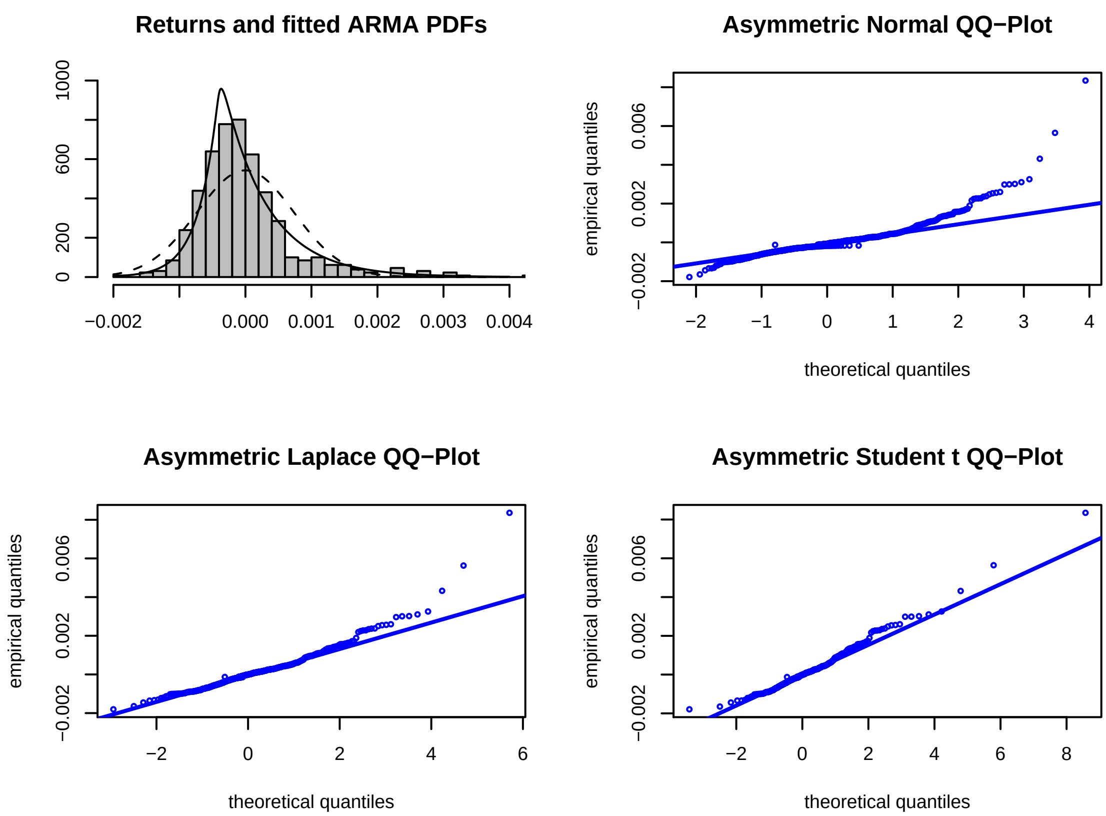
QQ-plots of the residuals from the three MA(1)-GARCH(1,1) models appearing in Table 5 are presented in Figure 5. We note that the AL and (particularly) the asymmetric t (with approximately 3 degrees of freedom) provide the best fits, however there are numerical issues with the latter in this particular dataset. The AL and t are also qualitatively very different distributions; the former has finite moments of all orders, whereas a t with ν' degrees of freedom has finite moments only up to a maximum order of ν, $\nu < \nu'$. Note that the difference between the two fits rests mainly on about 3 or 4 residuals. The AL only substantially deviates from the line in 3 upper tail values, but the t deviates both in 3 upper tail and one lower tail value.

An interesting feature of the normal vs. AL noise models is that apart from giving models with lower AIC, the latter tend to also have substantially less parameters. Note that although we are lumping the ARMA and ARMA-GARCH models in the same table, the former are intended to capture only the conditional mean, while the latter provide a model for both the conditional mean and conditional variance of the process. They should not therefore be viewed as directly competing models.

Table 7 presents predictions for the first return of 2007 using the fitted ARMA models with IID noise, along with 95% confidence bands. Note how the Gaussian prediction band is naturally symmetric, while that from the AL noise model reflects the asymmetry present in the data. Both bands do in fact capture the observed return value of 0.00036.

In conclusion, we note that ARMA models with AL noise can be useful for modeling time-varying conditional expectations of processes that tend to be peaked, skewed, and

Figure 5: Panel 1: Histogram of TIAA-CREF real estate returns and the marginal PDFs of the best-fitting ARMA models: MA(1) with IID AL noise (solid), ARMA(5,4) with IID normal noise (dashed). Remaining panels: QQ-plots of the residuals from three MA(1)-GARCH(1,1) models fitted to the TIAA-CREF returns, corresponding to different asymmetric distributions for the GARCH innovations.



leptokurtic, but which appear to have some higher order moments. In similar situations, GARCH models with AL innovations may provide a competitive fit and a greater degree of numerical stability with respect to other skewed distributions. AL noise models may also have substantially less parameters than their normal noise counterparts. These models are therefore worthy of inclusion in the applied statistician's "toolbox".

Acknowledgements

We are indebted to two anonymous referees for detailed reviews which led to a vastly improved paper. We thank T.J. Kozubowski for pointing us to the result of equation (1).

Table 7: Predictions of the first return of 2007 from two models fitted to the TIAA-CREF returns.

Model	Lower Prediction Bound	Upper Prediction Bound
ARMA(5,4) with IID Normal noise	0.00032-0.00157	0.00032+0.00157
MA(1) with IID AL noise	0.00051-0.00105	0.00051+0.00179

A QMLE Asymptotics for an ARMA Model With AL Noise

Let $\eta = (\phi_1, \ldots, \phi_p, \lambda_1, \ldots, \lambda_q, \kappa, \tau)$ denote the vector of parameters for the zero-mean ARMA(p, q) model driven by AL noise of Definition 1, and let $\hat{\eta}$ be its QMLE based on n observations. An application of the Theorem in Li and McLeod (1988), allows us to conclude that $\hat{\eta}$ is asymptotically normal with mean η and covariance matrix $n^{-1}\Omega^{-1}$, where Ω has the following block structure:

$$\Omega = \begin{bmatrix} J_{p+q} & 0 \\ 0 & \tilde{I}_2 \end{bmatrix}.$$

J_{p+q} has dimension $p+q$, its elements being functions of the AL PDF and the autocovariance of the process (see the Theorem in Li and McLeod, 1988). $\tilde{I}_2$ is the 2×2 Fisher Information matrix (per observation) corresponding to n IID observations from the AL PDF $\tilde{f}(y; \kappa, \tau)$, which represents $f(y; \theta, \kappa, \tau)$ with the constraint $\theta = -\tau(\kappa^{-1} - \kappa)/\sqrt{2}$ as in (3).

For an MA(1) model, routine calculations give $J_1 = (1 + \kappa^4)/[(1 - \lambda^2)\kappa^2]$. For a general ARMA, the elements of $\tilde{I}_2$ are given by:

$$\tilde{I}_2(1,1) = \mathbb{E}\left(\frac{\partial \log \tilde{f}}{\partial \kappa}\right)^2, \quad \tilde{I}_2(1,2) = \mathbb{E}\left(\frac{\partial \log \tilde{f}}{\partial \kappa}\frac{\partial \log \tilde{f}}{\partial \tau}\right), \quad \tilde{I}_2(2,2) = \mathbb{E}\left(\frac{\partial \log \tilde{f}}{\partial \tau}\right)^2.$$

These elements can be obtained from those for the 3×3 Fisher Information, I_3, corresponding to the (unconstrained) $f(y; \theta, \kappa, \tau)$ case, which is given in Kotz *et al.* (2001, Sec.3.5). For example, total differentiation with respect to κ gives,

$$\frac{\partial \log \tilde{f}}{\partial \kappa} = \frac{\partial \log f}{\partial \theta}\frac{\partial \theta}{\partial \kappa} + \frac{\partial \log f}{\partial \kappa},$$

so that after squaring and taking expectations,

$$\tilde{I}_2(1,1) = \left(\frac{\partial \theta}{\partial \kappa}\right)^2 I_3(1,1) + I_3(2,2) + 2\frac{\partial \theta}{\partial \kappa} I_3(1,2) = \frac{1 + \kappa^2 + 4\kappa^4 + \kappa^6 + \kappa^8}{(1+\kappa^2)^2\kappa^4}.$$

Proceeding similarly, we obtain

$$\tilde{I}_2(1,2) = \frac{\partial \theta}{\partial \kappa}\frac{\partial \theta}{\partial \tau} I_3(1,1) + \frac{\partial \theta}{\partial \kappa} I_3(1,3) + \frac{\partial \theta}{\partial \tau} I_3(1,2) + I_3(2,3) = \frac{(1-\kappa^2)(\kappa^4 - 3\kappa^2 - 1)}{(1+\kappa^2)\tau\kappa^3},$$

and

$$\tilde{I}_2(2,2) = \left(\frac{\partial \theta}{\partial \tau}\right)^2 I_3(1,1) + I_3(3,3) + 2\frac{\partial \theta}{\partial \tau} I_3(1,3) = \frac{1 - \kappa^2 + \kappa^4}{\tau^2\kappa^2}.$$

References

[1] M.M. Ali and M. Obaidullah, Distribution of linear combination of exponential variates, Comm. Statist. A—Theory & Methods 11 (1982), 1453-1463.

[2] D.N. Anderson and B.C. Arnold, Linnik distributions and processes, J. Appl. Prob. 30 (1993), 330-340.

[3] B. Andrews, M. Calder and R.A. Davis, Maximum likelihood estimation for α-stable autoregressive processes, Annals of Statistics (2009), to appear.

[4] I. Berkes and L. Horvath, The efficiency of the estimators of the parameters in GARCH processes, The Annals of Statistics, 32 (2004), 633-655.

[5] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (1986), 307-327.

[6] G.E.P. Box, Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification, Annals of Mathematical Statistics 25 (1954), 290-302.

[7] F.J. Breidt and R.A. Davis, Time-reversibility identifiability and independence of innovations for stationary time series, J. Time Series Analysis 13 (1992), 377390.

[8] P.J. Brockwell and R.A. Davis, Time Series: Theory and Methods, 2nd ed., Springer, New York, 1991.

[9] E. Damsleth and A.H. El-Shaarawi, ARMA models with double-exponentially distributed noise, J. Royal Statist. Soc. B 51 (1989), 61-69.

[10] W.B. Davenport, An experimental study of speech wave probability distributions, J. Acoust. Soc. Amer. 24 (1952), 390-399.

[11] R. Davidson and J.G. MacKinnon, Econometric Theory and Methods, Oxford University Press, New York, 2004.

[12] L.S. Dewald and P.A.W. Lewis, A new Laplace second-order autoregressive time-series model: NLAR(2), IEEE Transactions on Information Theory 31 (1985), 645-651.

[13] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50 (1982), 987-1007.

[14] C. Fernandez and M.F.J. Steel, On Bayesian modeling of fat tails and skewness, J. Amer. Statist. Assoc., 93 (1998), 359-371.

[15] C. Francq and J.M. Zakoian, Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, Bernoulli, 10 (2004), 605-637.

[16] D.A. Hsu, Long-tailed distributions for position errors in navigation, Appl. Statistics, 28 (1979), 62-72.

[17] K. Jayakumar and A.P. Kuttykrishnan, A time-series model using asymmetric Laplace distribution. Statist. Probab. Lett., 77 (2007), 1636-1640.

[18] S. Kotz, T.J. Kozubowski and K. Podgórski, The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance, Birkhäuser, Boston, 2001.

[19] T.J. Kozubowski, The inner characterization of geometric stable laws, Statistics & Decisions 12 (1994), 307-321.

[20] T.J. Kozubowski and S.T. Rachev, The theory of geometric stable distributions and its use in modeling financial data, European Journal of Operational Research 74 (1994), 310-324.

[21] V.S. Lekshmi and K.K. Jose, An autoregressive process with geometric α-Laplace marginals, Statistical Papers 45 (2004), 337-350.

[22] W.K. Li and A.I. McLeod, ARMA modelling with non-Gaussian innovations, J. Time Ser. Anal. 9 (1988), 155-168.

[23] T. Mathew and K. Jayakumar, Generalised Linnik distribution and process, Stochastic Modelling and Applications 6 (2003), 27-37.

[24] W.J. McGill, Random fluctuations of response rate, Psychometrika 27 (1962), 3-17.

[25] M. Novković, On Laplace autoregressive time series models, Matematichki Vesnik (Belgrade) 50 (1998), 53-56.

[26] L. Peng and Q. Yao, Least absolute deviations estimation for ARCH and GARCH models, Biometrika 90 (2003), 967-975.

[27] G. Samorodnitsky and M.S. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance, Chapman & Hall Ltd, London/New York, 1994.

[28] A.A. Trindade and Y. Zhu, Approximating the distributions of estimators of financial risk under an asymmetric Laplace law, Computational Statistics and Data Analysis 51 (2007), 3433-3447.

[29] D. Würtz, Y. Chalabi and L. Luksan, Parameter Estimation of ARMA Models with GARCH/APARCH Errors An R and SPlus Software Implementation , Journal of Statistical Software (forthcoming).