Statistics Surveys
Vol. 4 (2010) 232–274
ISSN: 1935-7516
DOI: 10.1214/09-SS060

The ARMA alphabet soup: A tour of ARMA model variants*

Scott H. Holan

Department of Statistics
University of Missouri
Columbia, MO 65211-6100
e-mail: holans@missouri.edu

Robert Lund

Department of Mathematical Sciences
Clemson University
Clemson, SC 29634-0975
e-mail: lund@clemson.edu

and

Ginger Davis

Department of Systems and Information Engineering
University of Virginia
Charlottesville, VA 22904-4747
e-mail: gingerdavis@Virginia.edu

Abstract: Autoregressive moving-average (ARMA) difference equations are ubiquitous models for short memory time series and have parsimoniously described many stationary series. Variants of ARMA models have been proposed to describe more exotic series features such as long memory autocovariances, periodic autocovariances, and count support set structures. This review paper enumerates, compares, and contrasts the common variants of ARMA models in today's literature. After the basic properties of ARMA models are reviewed, we tour ARMA variants that describe seasonal features, long memory behavior, multivariate series, changing variances (stochastic volatility) and integer counts. A list of ARMA variant acronyms is provided.

AMS 2000 subject classifications: Primary 62M10; secondary 91B84.
Keywords and phrases: Autocovariance function, counts, long memory, short memory, stochastic volatility, time series.

Received November 2009.

Contents

*This paper was accepted by Mark Griffin, Associate Editor for the ASA.

1. Introduction

Autoregressive moving-average (ARMA) models are fundamental stationary time series models (stationary here refers to covariance or second order stationary and not strict stationarity). The ARMA class is dense in all short memory stationary series; moreover, the class is parsimonious in that it flexibly generates a variety of different stationary autocovariance shapes from a few parameters. Many extensions and variants of ARMA models have been developed to describe departures from short memory stationary characteristics, such as long memory autocovariances, periodicities, stochastic volatility (changing variances), multivariate series and discrete counts. In this paper, we enumerate many of the ARMA variants, discussing what they intend to achieve, and compare and contrast their probabilistic and statistical structures.

The time series literature is by now extensive and many general and specialized texts exist. For example, Brockwell and Davis (2002), Chatfield (2003), Shumway and Stoffer (2006), Box, Jenkins and Reinsel (2008) and Cryer and Chan (2008) are comprehensive course texts for introductory material, emphasizing mainly the univariate case. Brockwell and Davis (1991) and Fuller (1996) are more advanced treatments, the former walking the reader through many theoretical and conceptual details and the latter focusing on statistical issues. Texts considering the multivariate case are more limited but include Hannan (1970), Reinsel (1997) and Lütkepohl (2005). Hamilton (1994) casts the material from an econometric standpoint.

As we proceed, citations to detailed references are presented and avenues for future research are discussed. The last section provides a list of ARMA variant acronyms — the proverbial alphabet soup. Since the area is so voluminous, this paper is necessarily incomplete; however, it does study a large portion of the subject. Some elements of this paper, such as Sections 2 and 3, are well known but are included for completeness; other aspects, such as Sections 4.2 and 6.2, are relatively new.

2. ARMA background

We begin with a univariate covariance stationary time series $\{X_t\}$. Stationarity requires that $E[X_t]$ does not depend on t and that $\text{Cov}(X_t, X_{t+h})$ is finite and only depends on the "lag" h. For simplicity, we consider the case where $E[X_t] \equiv 0$; if the series does not have a zero mean, one simply examines $\{X_t - E[X_t]\}$. Estimation and removal of a general non-zero first moment via regression methods is typically straightforward; Fuller (1996) is an excellent reference for this endeavor in time series settings. The autocovariance and autocorrelation of $\{X_t\}$ at lag h are denoted by $\gamma(h) = \text{Cov}(X_t, X_{t+h})$ and $\rho(h) = \gamma(h)/\gamma(0)$ respectively.

The series $\{X_t\}$ is said to be an ARMA series with autoregressive order $p \geq 0$ and moving-average order $q \geq 0$ if it is stationary and a solution to the difference equation

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}. \tag{2.1}$$

Here, $\{Z_t\}$ is zero mean white noise; specifically, Z_t and Z_s are uncorrelated whenever $t \neq s$ and $\text{Var}(Z_t) \equiv \sigma^2$. Equation (2.1) is a stochastic linear recursion driven by white noise. In most cases, the solution to (2.1) is unique (in mean square) and can be expressed as

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}, \tag{2.2}$$

where the ψ_ks are functions of $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$. The Wold Decomposition (Wold, 1954); states that all infinite stationary series can be written as (2.2) plus a so-called deterministic component (the deterministic component, as it is perfectly predictable from past observations, can be estimated and removed).

When $q = 0$ in (2.1), $\{X_t\}$ is called an autoregression of order p (AR(p)); when $p = 0$, $\{X_t\}$ is referred to as a moving-average of order q (MA(q)). Note that one does not need a ϕ_0 coefficient multiplying X_t or a θ_0 coefficient multiplying Z_t in (2.1). Introduction of such coefficients would make parameters non-identifiable as constant multiples of stationary series/white noise are again stationary series/white noise. Authors are inconsistent about the plus/minus signs on the ARMA coefficients. For example, Brockwell and Davis (1991) use (2.1) verbatim, while Box, Jenkins and Reinsel (2008) place a minus sign on their moving-average coefficients.

Two ubiquitous ARMA quantities are the autoregressive and moving-average polynomials ϕ and θ defined respectively by

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \cdots + \phi_q z^q$$

for a complex-valued z (as roots of the polynomials will arise later, it is necessary to view polynomials as functions of a complex argument). The ARMA equation (2.1) is frequently written in the compact form

$$\phi(B)X_t = \theta(B)Z_t, \tag{2.3}$$

where the backshift operator B applied to X_t is defined as X_{t-1}. Extending this logic to higher powers, we set $B^k X_t = X_{t-k}$ for $k \geq 0$.

The autoregressive and moving-average polynomials play critical roles in the structure of ARMA solutions. To proceed, two assumptions are necessary. First, we assume that ϕ and θ have no common roots. Solutions to (2.1) may not be unique if ϕ and θ have a common root. Second, we assume that $\phi(z) \neq 0$ when $|z| = 1$. When $\phi(z)$ is non-zero on the complex unit circle, the ARMA equation has a unique stationary solution. The ψ_ks in (2.2) can be determined by expanding $\theta(z)/\phi(z)$ into a power series over some annulus containing the unit circle and equating coefficients — say

$$\psi(z) := \sum_{k=-\infty}^{\infty} \psi_k z^k = \frac{\theta(z)}{\phi(z)}. \tag{2.4}$$

When $\phi(z) \neq 0$ for all z with $|z| = 1$, one has absolute summability of the ψ_ks; i.e.,

$$\sum_{k=-\infty}^{\infty} |\psi_k| < \infty \tag{2.5}$$

(e.g., Shumway and Stoffer, 2006, Chapter 3).

An ARMA model is said to be causal if X_t can be written explicitly as a function of $Z_t, Z_{t-1}, \ldots$; i.e., X_t cannot depend on Z_{t+k} for any $k \geq 1$ and $\psi_k = 0$ for all $k < 0$. A fundamental result is that ARMA models are causal if and only if $\phi(z) \neq 0$ for all z with $|z| \leq 1$. For example, causality of the first-order autoregression (AR(1)) model

$$X_t = \phi X_{t-1} + Z_t \tag{2.6}$$

takes place when $|\phi| < 1$ (parameter subscripts are omitted for first order models). Stationary solutions to the AR(1) equation do exist when $|\phi| > 1$; for example

$$X_t = -\sum_{k=1}^{\infty} \frac{Z_{t+k}}{\phi^k}$$

satisfies (2.6). As X_t depends on future Z_js here, this solution is non-causal and is frequently discarded (see Breidt, Davis and Trindade (2001) for an exception). Stationarity and causality are frequently confused (Hamilton (1994) is a prominent example). When $\phi = \pm 1$, no stationary solution to the AR(1) equation exists.

It is possible for two distinct sets of ARMA parameters to give the same autocovariance function (at all lags). A necessary and sufficient condition for an ARMA model to be identifiable through its autocovariance function (that is, to a second order) is the so-called invertibility condition that $\theta(z) \neq 0$ for all z with $|z| \leq 1$. ARMA model causality and invertibility is typically assumed in statistical practice. If the model is not causal and/or invertible, one can change the ARMA parameters to those of a causal and invertible model without altering any autocovariances (Chapter 4 of Brockwell and Davis (1991) provides the construction).

Zero mean time series models are frequently compared via their second moment structures. An expression for the ARMA autocovariances can be obtained from (2.2):

$$\gamma(h) = \sigma^2 \sum_{k=-\infty}^{\infty} \psi_k \psi_{k+h}. \tag{2.7}$$

Further, (2.5) and (2.7) give

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| \leq \sigma^2 \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \left| \psi_k \psi_{k+h} \right| = \sigma^2 \left(\sum_{k=-\infty}^{\infty} |\psi_k| \right)^2 < \infty.$$

Such summability is typically referred to as short memory in the literature. Observe that any ARMA series where $\phi(z)$ is non-zero on the complex unit circle has short memory. It is known that when ϕ has a root on the complex unit circle, no stationary solution to the ARMA equation exists (Problem 4.28 in Brockwell and Davis (1991) states this — the accompanying solution is non-trivial). In short, ARMA models are good stationary short memory time series models. Some of the variants below are designed to induce long memory (also known as long-range dependence) into the model. It is important to note that, in practice, checking assumptions is an essential aspect of model development.

For a causal ARMA model, a difference equation for $\gamma(h)$ can be obtained by multiplying both sides of (2.1) by X_{t-h} for $h \geq 0$ and taking expectations. Invoking causality, one gets

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p), \tag{2.8}$$

for $h > q$. Hence, $\gamma(\cdot)$ obeys a pth order linear difference equation. Solutions to (2.8), in the case where $\phi(z)$ has no repeated roots, are linear combinations of geometric sequences. In fact, autocovariances of causal ARMA models decay geometrically in that $|\gamma(h)| \leq \kappa r^{-h}$ for some $\kappa < \infty$ and $r > 1$ (one might say that a causal ARMA model has "very short memory"). The largest possible r can be taken as the magnitude of the largest root of ϕ. The initial conditions in (2.8) needed to identify all autocovariances are

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p) + \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k}, \quad h \leq q. \tag{2.9}$$

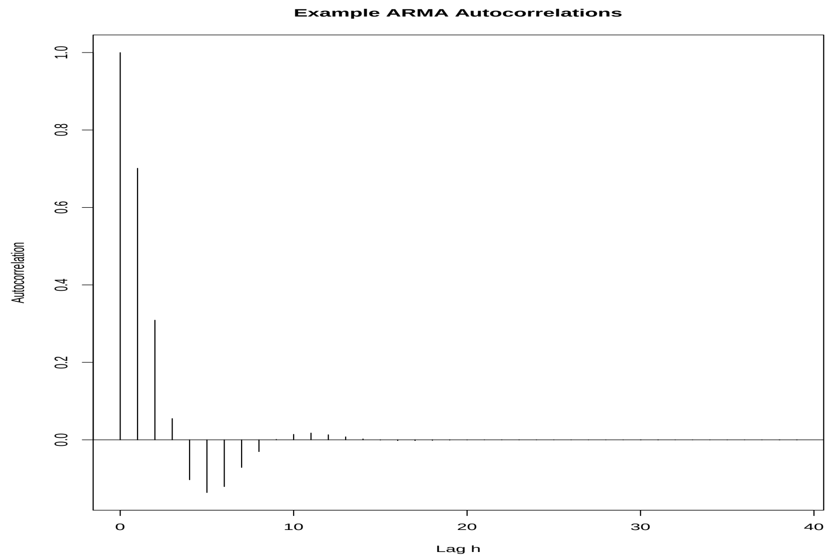

Fig 1. *Autocorrelations of the causal and invertible ARMA(3,2) series when* $\phi_1 = 1/2$, $\phi_2 = 1/3$, $\phi_3 = -1/3$, $\theta_1 = 1/2$, $\theta_2 = -1/3$, *and* $\sigma^2 = 1$.

Here, the convention $\theta_0 = 1$ is made. Simple explicit expressions for ARMA autocovariances are available for all moving-averages, autoregressions of first and second order, and ARMA(1,1) models. Explicit expressions for ARMA autocovariances for higher order models are not available in general; for these reasons, ARMA autocovariance evaluation is typically a numerical task (see Tunnicliffe-Wilson (1979) for an efficient algorithm). Figure 1 plots the autocorrelations of the causal and invertible ARMA(3,2) series when $\phi_1 = 1/2$, $\phi_2 = 1/3$, $\phi_3 = -1/3$, $\theta_1 = 1/2$, $\theta_2 = -1/3$, and $\sigma^2 = 1$. Notice how quickly the autocorrelations decay to zero with increasing lag.

Stationary series can be equivalently described in the spectral (Fourier) domain. This is because every stationary autocovariance $\gamma(\cdot)$ admits the spectral representation

$$\gamma(h) = \int_{[-\pi,\pi)} e^{ih\lambda} dF(\lambda), \tag{2.10}$$

where F is a symmetric distribution function on $[-\pi, \pi)$: $F(\lambda)$ is nondecreasing in λ, $F(-\pi) = 0$, $F(\pi) = \gamma(0)$, and $dF(\lambda) = dF(-\lambda)$ for $\lambda \in [0, \pi]$. Here, $i = \sqrt{-1}$.

In the case where $\gamma(\cdot)$ has short memory, a spectral density exists in the sense that (2.10) becomes

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda,$$

and the spectral density takes the form

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda}\gamma(h).$$

Note that $f(\lambda)$ is real, symmetric in λ, and non-negative.

The ARMA spectral density is known to be

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}, \qquad -\pi \le \lambda < \pi,$$

which is a rational function of λ (the ratio of two polynomials). Since AR and/or MA spectral densities can each approximate any spectral density uniformly over $\lambda \in [-\pi, \pi)$ and the spectral density determines the autocovariances, AR and MA series are dense in the class of stationary series.

Akin to moment generating functions in probability, spectral methods provide a convenient isomorphism in stationary time series analysis. Some results are easily proven in the spectral domain, but are considerably more difficult to establish in the time domain (the reverse is also true, so it is wise to be proficient in both domains). Consider, for example, a proof of the fact that the sum of two independent ARMA series is again ARMA. A spectral argument is merely that the sum of two rational functions is again rational; a time-domain argument is considerably more involved. Also, until the work of Ansley (1979), spectral-based approximations to the Gaussian ARMA likelihood dominated the literature. Today, Gaussian ARMA likelihoods can be rapidly evaluated with established time-domain algorithms (which also provide the exact likelihood).

Parameter estimation in ARMA models can be performed via moment, least squares, and Gaussian maximum likelihood methods. All methods work well for AR(p) models and give asymptotically equivalent estimators. Here and in what follows, we assume that the ARMA orders p and q are known. If this is not the case, optimizing a likelihood penalized by some selection criterion such as the AIC or BIC frequently gives consistent estimates of the AR and/or MA orders (Brockwell and Davis (1991) discuss this issue in detail).

We will not elaborate on least squares methods here. Moment estimation techniques for AR(p) models simply plug the sample autocovariances

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) \tag{2.11}$$

into (2.9) and solve for the resulting parameter estimators. Here, $\bar{X} = \sum_{t=1}^{n} X_t/n$ and $X_1, \ldots, X_n$ is the data sample. The denominator n is used in (2.11) instead of $n - h$ for technical reasons rooted in non-negative definiteness of the sample autocovariance function. The moment estimators for causal AR(p) models are asymptotically normal and $\sqrt{n}$-consistent; these aspects are quantified in Fuller (1996).

Likelihood methods are preferred when the ARMA model has a moving-average component as solutions to moment equations based on sample autocovariances may not exist. Optimization of the likelihood function is a numerical task, easily accomplished with gradient step and search methods. No simple explicit closed forms for the likelihood estimators exist, even in the simplest cases (for example, the AR(1) likelihood estimator of ϕ requires solving a cubic equation). Hence, the major statistical issue entails rapidly evaluating the likelihood. Let $\boldsymbol{\alpha} = (\phi_1, \ldots, \phi_p; \theta_1, \ldots, \theta_q; \sigma^2)$ be the $p + q + 1$-dimensional vector of ARMA parameters and let $L(\boldsymbol{\alpha})$ denote the Gaussian likelihood of a causal and invertible ARMA model. An orthogonal decomposition known as the Innovations Algorithm provides

$$-2\log(L(\boldsymbol{\alpha})) = n\log(2\pi) + \sum_{t=1}^{n}\log(v_t) + \sum_{t=1}^{n}\frac{(X_t - \widehat{X}_t)^2}{v_t}, \tag{2.12}$$

where $\widehat{X}_t = P[X_t | X_{t-1}, \ldots, X_1]$ is the best one-step-ahead prediction of X_t from linear combinations of $X_1, \ldots, X_{t-1}$ and $v_t = E[(X_t - \widehat{X}_t)^2]$ is the unconditional mean squared error. Because of (2.12), likelihood optimization boils down to rapid computation of $\{\widehat{X}_t\}$ and $\{v_t\}$.

For AR(p) models and $t > p$,

$$\widehat{X}_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} \quad \text{and} \quad v_t = \sigma^2.$$

This forecasting relation makes the autoregressive class extremely attractive to practitioners. Computing $\widehat{X}_t$ and v_t for $t \leq p$ is also straightforward (see Chapter 5 of Brockwell and Davis (1991)). As there is no asymptotic loss of precision in studying a conditional likelihood of $X_{p+1}, \ldots, X_n$ given $X_1, \ldots, X_p$, "edge-effects" from the first p observations can be ignored; this tactic will be used below.

Forecasting recursions are more complicated for ARMA models with a moving-average component. A classic result of Ansley (1979) provides, for $t > \max(p, q)$,

$$\widehat{X}_t = \sum_{k=1}^{p}\phi_k X_{t-k} + \sum_{j=1}^{q}\theta_{t,j}(X_{t-j} - \widehat{X}_{t-j}), \tag{2.13}$$

along with simple recursions for v_t and the coefficients $\theta_{t,j}$ for $1 \leq j \leq q$. Notice the similarities between (2.13) and (2.1). For causal and invertible ARMA models, it can be shown that $\theta_{t,j}$ converges at a geometric rate to θ_j for each j with $1 \leq j \leq q$ and that v_t converges monotonically downwards to σ^2 at a geometric rate. Because of this, there is no asymptotic loss of precision in examining a likelihood of $X_{\max(p,q)+1}, \ldots, X_n$ conditional on $X_1, \ldots, X_{\max(p,q)}$ using the asymptotic relations

$$\widehat{X}_t = \sum_{k=1}^{p}\phi_k X_{t-k} + \sum_{j=1}^{q}\theta_j(X_{t-j} - \widehat{X}_{t-j}) \quad \text{and} \quad v_t = \sigma^2 \tag{2.14}$$

for $t > \max(p, q)$. This said, the reader is warned that asymptotic proofs for ARMA models with moving-average components are usually much harder than those for autoregressive series. Shumway and Stoffer (2006, Chapter 3) and Fuller (1996) are comprehensive ARMA estimation references that elaborate on much of the above.

ARMA models can be also cast in a state-space (dynamic linear models) framework. Observe that (2.1) can be written as

$$X_t = \sum_{j=1}^{r} \phi_j X_{t-j} + Z_t + \sum_{j=1}^{r-1} \theta_j Z_{t-j} \qquad t \geq r,$$

where $r = \max\{p, q+1\}$ and some of the ϕ_js and/or θ_js are taken as zero. Let $W_t = (1, 0, 0, \ldots, 0)$ and

$$\boldsymbol{\alpha}_t = \begin{pmatrix} X_t \\ \phi_2 X_{t-1} + \cdots + \phi_r X_{t-r+1} + \theta_1 Z_t + \cdots + \theta_{r-1} Z_{t-r+2} \\ \phi_3 X_{t-1} + \cdots + \phi_r X_{t-r+2} + \theta_2 Z_t + \cdots + \theta_{r-1} Z_{t-r+3} \\ \vdots \\ \phi_r X_{t-1} + \theta_{r-1} Z_t \end{pmatrix}.$$

Then for $t \geq r$,

$$\begin{aligned} X_t &= W_t \boldsymbol{\alpha}_t, \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}\boldsymbol{\alpha}_t + \mathbf{R}\eta_t, \end{aligned}$$

with

$$\mathbf{T} = \begin{bmatrix} \phi_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{r-1} & 0 & \cdots & 1 \\ \phi_r & 0 & \cdots & 0 \end{bmatrix}; \quad \mathbf{R} = \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{r-1} \end{pmatrix}; \quad \eta_t = Z_{t+1},$$

defines one convenient ARMA state-space representation (see Durbin and Koopman (2001, Pages 38-47) and Harvey (1989)). State-space representations have aided Bayesian approaches to ARMA parameter estimation. For more here, see Marriott *et al.* (1996), West and Harrison (1997), and Prado and West (2010).

Two generalizations of ARMA models that will not be extensively dealt with later are worth mentioning here: models that accommodate covariates and nonzero means and heavy tailed models. A model that allows X_t to have the nonzero mean $E[X_t] = \mu_t$ simply examines the difference equation

$$X_t - \mu_t = \sum_{k=1}^{p} \phi_k (X_{t-k} - \mu_{t-k}) + Z_t + \sum_{k=1}^{q} \theta_k Z_{t-k}.$$

When μ_t is posited to be of the form $\mu_t = \mathbf{H}'\mathbf{u}_t$, where $\mathbf{H}$ is a known $r \times 1$ regression design vector and $\mathbf{u}_t$ is an r-dimensional vector of covariates at time

t, the model is frequently referred to as an ARMAX model, with X standing for exogenous. Hannan (1976), Chan (2002, Page 138), and Shumway and Stoffer (2006) discuss ARMAX issues further. Because the analysis of ARMAX models is similar to that for ARMA models, we work with the zero mean case in what follows.

In some fields (especially finance, internet traffic, and telecommunications), the innovations $\{Z_t\}$ may not have finite polynomial moments of all orders (or even second moments). Kokoszka and Taqqu (1995), for example, consider ARMA models when $\{Z_t\}$ is a stable infinite variance sequence. Davis and Resnick (1996) rigorously quantify notions of (2.2) when $\{Z_t\}$ has an infinite second moment and go on to analyze one such nonlinear stationary series that is termed a bilinear process and is related to first order autoregressive models. A literature on extreme values of ARMA models (Rootzén, 1986; Chernick, Hsing and McCormick, 1991; Scotto, 2007, among others) is tangential here and relates the tail distribution properties of Z_t to those of X_t.

3. ARIMA models

Some time series exhibit behavior that is not adequately described by ARMA models. Nonstationary series with random-walk type characteristics can often be transformed to stationary series by differencing. This section discusses the class of autoregressive integrated moving-average (ARIMA) models. ARIMA models have the property that the dth order difference of $\{X_t\}$ is an ARMA series. Specifically, if d is a non-negative integer, $\{X_t\}$ is said to be an ARIMA(p, d, q) series if $\{Y_t\}$, defined for a fixed $t > d$ by $Y_t = (1-B)^d X_t$, is a causal ARMA(p, q) series. Hence, $\{X_t\}$ satisfies a difference equation analogous to (2.3):

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t,$$

where $\{Z_t\}$ is zero mean white noise with $\mathrm{Var}(Z_t) \equiv \sigma^2$. As before, $\phi(z)$ and $\theta(z)$ are polynomials of degree p and q respectively, with no common roots, and $\phi(z) \neq 0$ on $|z| = 1$. Since $\phi(z)(1 - z)^d$ has a zero of order d at $z = 1$, it follows that $\{X_t\}$ is stationary if and only if $d = 0$. The I in the ARIMA acronym stands for integration (or summation in discrete time). When $d = 1$, one has the random walk type representation

$$X_t = X_0 + \sum_{k=1}^{t} Y_k,$$

where $\{Y_t\}$ is ARMA.

A common modeling strategy successively differences the series until a difference order d (assumed minimal) is found where $\{(1 - B)^d X_t\}$ has the rapidly decaying sample autocovariances indicative of an ARMA series whose AR polynomial is root free inside and near the unit circle. As was the case with ARMA series, checking model assumptions is essential. Figure 2 shows sample autocorrelations of an ARIMA(3,1,2) series of length 1000 whose ARMA coefficients are

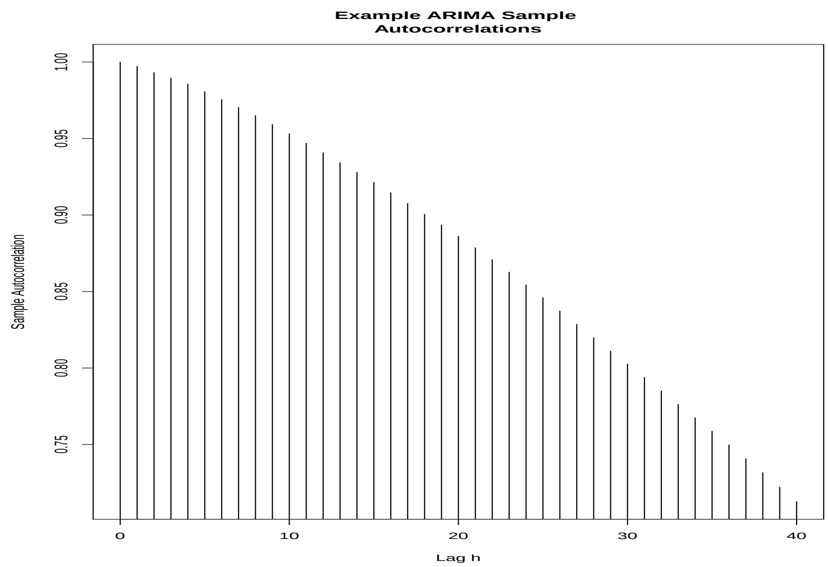

Fig 2. *Sample autocorrelations of the causal and invertible ARIMA(3,1,2) series with* $d = 1$, $\phi_1 = 1/2$, $\phi_2 = 1/3$, $\phi_3 = -1/3$, $\theta_1 = 1/2$, $\theta_2 = -1/3$, *and* $\sigma^2 = 1$.

taken as those used in Figure 1. The autocorrelations decay to zero much slower than those of the ARMA(3,2) series. Here, true stationary autocorrelations do not exist as the ARIMA model is not stationary. Rather, the theme is that if an ARIMA series were to be misspecified as stationary, sample autocorrelations with such a structure would arise.

Differencing also removes some trends in the mean of the series. Because a polynomial trend of degree $d - 1$ can be added to the series without altering dth order differences, ARIMA methods are often uses to eliminate polynomial trends in the series. This said, ARIMA models still adequately describe many trend-free series. Comprehensive discussions of ARIMA models can be found in Brockwell and Davis (1991), Shumway and Stoffer (2006) and Box, Jenkins and Reinsel (2008).

Parameters in ARIMA models are easily estimated: simply apply the ARMA estimation methods above to $\{(1-B)^d X_t\}$. While such a scheme shortens the series length by d observations (one cannot define $(1-B)^d X_t$ for $t \leq d$), no asymptotic loss of precision is incurred. For a comprehensive discussion of ARIMA estimation, see Chapter 7 of Box, Jenkins and Reinsel (2008).

4. Periodic and seasonal ARMA models

Many observed series display periodicities in their autocovariance structure. For example, day-to-day precipitations on the Pacific Coast of the United States are strongly correlated during the summer where rain is infrequent and less correlated during the winter when fronts pass through the region in rapid succession. Temperature variances in Miami, FL are some three times greater during winter months than they are during summer months. The ARMA models of Section 2 are stationary and do not describe periodic features. This section considers two ARMA variants devised to describe periodic behavior: seasonal autoregressive moving-average (SARMA) models and periodic autoregressive moving-average (PARMA) models. The PARMA class has yet to be popularized by any mainstream textbook but deserves more attention in our opinion.

4.1. SARMA models

For a known period T, the SARMA idea is to drive an ARMA equation at data taken at multiples of T. The difference equation governing the model is

$$X_t - \phi_1 X_{t-T} - \cdots - \phi_p X_{t-pT} = Z_t + \theta_1 Z_{t-T} + \cdots + \theta_q Z_{t-Tq}, \tag{4.15}$$

or in compact form, $\phi(B^T)X_t = \theta(B^T)Z_t$. The autocovariances of solutions to (4.15) have the property that $\gamma(h) = 0$ unless h is a multiple of T, which is not a feature commonly exhibited by series in practice. Remedies to this typically allow $\{Z_t\}$ to be another ARMA series instead of white noise, say

$$\phi^*(B)Z_t = \theta^*(B)\epsilon_t, \tag{4.16}$$

where $\{\epsilon_t\}$ is zero mean white noise with variance σ^2_ϵ and the superscript $*$ indicates that the AR and MA polynomials in (4.15) and (4.16) are in general different. Combining (4.15) and (4.16) gives

$$\phi(B^T)\phi^*(B)X_t = \theta(B^T)\theta^*(B)\epsilon_t. \tag{4.17}$$

The autocovariances of this "two-layered" SARMA model tend to be relatively larger at lags that are multiples of T, but they will not be zero at other lags in general. Figure 3 plots the autocorrelations of a SARMA model with $T = 4$. Here, the model in (4.15) is taken as that which produced Figure 1 and the component in Equation (4.16) is taken as AR(1) with $\phi = 1/2$. Note that the autocorrelations at lags 4 and 8 are larger than their preceding neighbors, i.e., autocorrelations at lags 3 and 7.

The SARMA acronym is perhaps a misnomer as solutions to (4.17) do not have periodic features in a strict sense. In fact, (4.17) is an ARMA model with autoregressive order $pT + p^*$ and moving-average order $qT + q^*$ (most of the coefficients in this representation are zero, however); here, p^* and q^* are the autoregressive and moving-average orders in (4.16). It follows that SARMA series are actually stationary.

One method of SARMA parameter estimation is done in layers. First, a good model is fitted to the autocovariances $\widehat{\gamma}(0), \widehat{\gamma}(T), \widehat{\gamma}(2T), \ldots$. A second layer is then chosen to describe $\widehat{\gamma}(1), \ldots, \widehat{\gamma}(T-1)$. A better scheme maximizes the ARMA$(pT + p^*, qT + q^*)$ Gaussian likelihood of the two-layered model (that is, jointly) imposing constraints on the parameters that entail that the AR

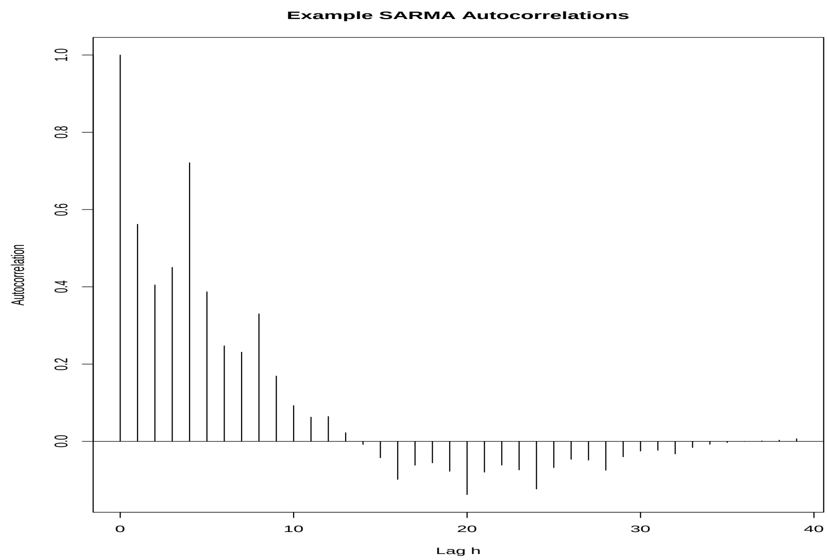

FIG 3. *Autocorrelations of the SARMA series with* $T = 4$*, Layer One Parameters* $\phi_1 = 1/2$*,* $\phi_2 = 1/3$*,* $\phi_3 = -1/3$*,* $\theta_1 = 1/2$*,* $\theta_2 = -1/3$*, and Layer Two Parameters* $\phi_1 = 1/2$ *and* $\sigma_\epsilon^2 = 1$*.*

and MA polynomials factor as in (4.17). See Brockwell and Davis (1991) and Box, Jenkins and Reinsel (2008) for more on estimation of SARMA parameters.

SARMA methods have proven useful in modeling economic processes. This is because the dependence structure of such series is most prominent at lags that are multiples of T. January housing starts, for example, are more heavily correlated with housing starts of the previous January than with housing starts of last month (December). SARMA methods are not very useful in describing periodic series whose autocovariances decay monotonically in lag. For example, in describing daily temperature anomalies about a periodic mean (take $T = 365$), the SAR(1) difference equation $X_t = \phi X_{t-T} + Z_t$ is not useful as it attempts to explain today's temperature fluctuations from anomalies 365 days ago, a dubious task considering the poor quality of weather forecasts even a week in advance. In many settings, one needs a model that has periodic features and where the most recent observations are included in the autoregressive component of the difference equation. This brings us to PARMA series.

4.2. PARMA models

PARMA methods allow the ARMA coefficients to vary periodically with time. The difference equation governing the PARMA model is

$$X_t - \phi_1(t)X_{t-1} - \cdots - \phi_p(t)X_{t-p} = Z_t + \theta_1(t)Z_{t-1} + \cdots + \theta_q(t)Z_{t-q} \quad (4.18)$$

and the coefficients satisfy $\phi_i(t+T) = \phi_i(t)$ for $1 \leq i \leq p$ and $\theta_j(t+T) = \theta_j(t)$ for $1 \leq j \leq q$. In this setting, $\{Z_t\}$ is zero mean periodic white noise: Z_t and Z_s are uncorrelated when $t \neq s$ and $\text{Var}(Z_t) = \sigma^2(t) > 0$ with $\sigma^2(t+T) = \sigma^2(t)$. One can allow p and q to vary periodically if needed (Lund and Basawa, 2000; Vecchia, 1985b); we will not do this here.

A notation that emphasizes seasonality uses $X_{nT+\nu}$ as the data point during the νth season of the nth cycle of data. Here, ν is a seasonal suffix that satisfies $1 \leq \nu \leq T$; we allow a 0th cycle of data so that X_1 denotes the first observation (season 1 of cycle 0). The PARMA model in (4.18) is equivalently written as

$$X_{nT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} = Z_{nT+\nu} + \sum_{j=1}^{q} \theta_j(\nu) Z_{nT+\nu-j}, \tag{4.19}$$

where $\{Z_t\}$ is zero mean periodic white noise with $\text{Var}(Z_{nT+\nu}) = \sigma^2(\nu)$.

PARMA models were first used in Hannan (1955); Jones and Brelsford (1967), Pagano (1978) and Troutman (1979) popularized the idea further. Most prominent introductory time series texts discuss SARMA models but omit PARMA models (Brockwell and Davis, 1991; Box, Jenkins and Reinsel, 2008; Shumway and Stoffer, 2006; Cryer and Chan, 2008). This omission is unfortunate as PARMA models, unlike SARMA models, are actually periodic. Specifically, PARMA autocovariances satisfy

$$\text{Cov}(X_{t+T}, X_{s+T}) = \text{Cov}(X_t, X_s) \tag{4.20}$$

for each integer t and s. In other words, each season ν has its own stationary covariance function. Let $\gamma_h(\nu) = \text{Cov}(X_{nT+\nu}, X_{nT+\nu-h})$ for $h \geq 0$ and $\nu \in \{1, \ldots, T\}$. Shao and Lund (2004) present an algorithm that rapidly computes the $\gamma_h(\nu)$s. Figure 4 shows the seasonal autocovariance function of a PARMA(2,1) model with $T = 4$. Here, the model coefficients used are listed in Table 1. This model is causal and invertible (as defined below). Observe that the autocovariances decay to zero quickly in each season and that there are differences in the seasonal autocovariance functions.

PARMA series can be viewed as T-variate stationary series. To see this, block $\{X_{nT+\nu}\}$ into $T \times 1$ vectors via

$$\boldsymbol{X}_n = (X_{nT+1}, \ldots, X_{nT+T})'.$$

Gladyšhev (1961) proved the intuitive result that $\{\boldsymbol{X}_n\}$ is T-variate stationary if and only if (4.20) holds. While this result is convenient in proofs, periodic series have additional structure beyond general multivariate stationarity: for a fixed n, the components of $\boldsymbol{X}_n$ are time-ordered. Hence, PARMA series are T-variate stationary series where the T components within each $\boldsymbol{X}_n$ are also time ordered. This aspect has practical ramifications. For example, consider forecasting a July temperature from past monthly temperatures ($T = 12$). A suboptimal forecast (the naive multivariate forecast) is obtained by extracting the 7th component of the 12-dimensional forecast from data that contains all monthly temperatures occurring in the *previous years*. This forecast does not

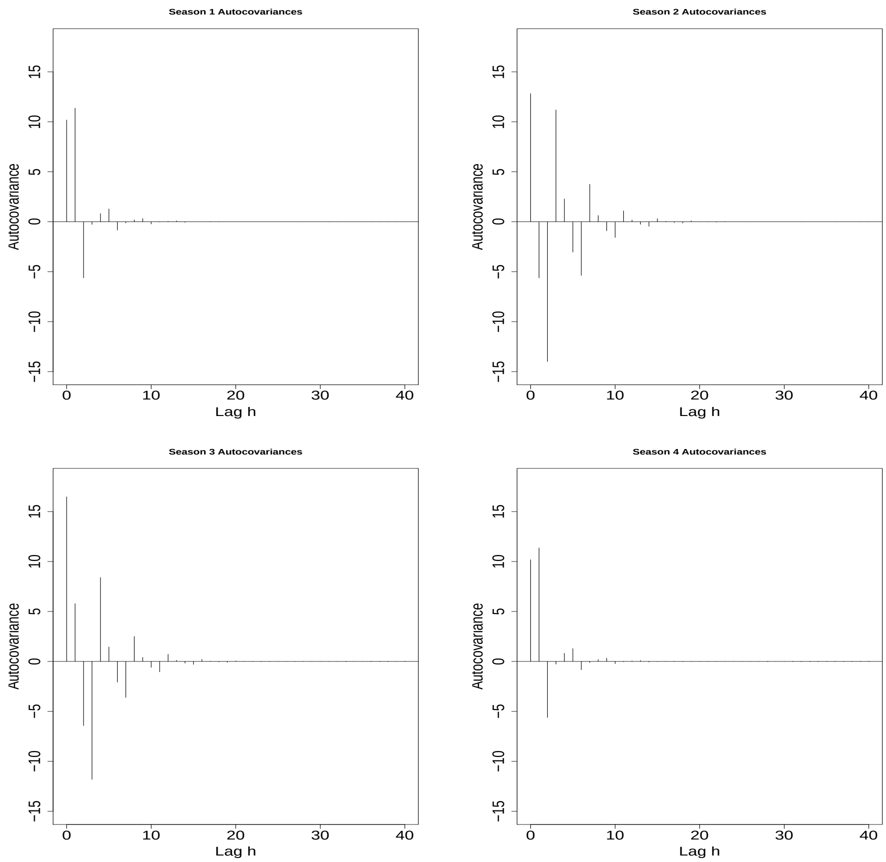

FIG 4. *Autocovariances of the PARMA model whose coefficients are listed in Table 1.*

TABLE 1
PARMA(2,1) coefficients

Season ν	$\phi_1(\nu)$	$\phi_2(\nu)$	$\theta_1(\nu)$	$\sigma^2(\nu)$
1	0.8	0.1	0.5	1.0
2	0.2	0.7	0.3	9.0
3	-0.2	0.7	-0.3	9.0
4	-0.8	0.1	-0.5	1.0

contain January-June temperatures of the current year, arguably some of the most important regressands. A better forecast is obtained by using the PARMA model setup above.

The PARMA model in (4.19) can be written in the multivariate ARMA form

$$\boldsymbol{\Phi}_0 \boldsymbol{X}_n - \sum_{k=1}^{p^*} \boldsymbol{\Phi}_k \boldsymbol{X}_{n-k} = \boldsymbol{\Theta}_0 \boldsymbol{Z}_n + \sum_{j=1}^{q^*} \boldsymbol{\Theta}_j \boldsymbol{Z}_{n-j},$$

where $\boldsymbol{Z}_n = (Z_{nT+1}, \ldots, Z_{nT+T})'$ is multivariate white noise with a diagonal covariance matrix. Construction of the coefficients $\boldsymbol{\Phi}_\ell$, $0 \le \ell \le p^*$, $\boldsymbol{\Theta}_\ell$, $0 \le \ell \le q^*$, the model orders p^* and q^* in terms of the model coefficients in (4.19) and p and q is detailed in Vecchia (1985b). From this, one can extract conditions for when a causal and invertible solution to the PARMA difference equation exists. For example, unless $\det(\boldsymbol{\Phi}(z)) = 0$ for some z on the complex unit circle, where

$$\boldsymbol{\Phi}(z) = \boldsymbol{\Phi}_0 - \boldsymbol{\Phi}_1 z - \cdots - \boldsymbol{\Phi}_{p^*} z^{p^*},$$

the PARMA equation will have a unique (in mean square) solution that may be written as

$$X_{nT+\nu} = \sum_{k=-\infty}^{\infty} \psi_k(\nu) Z_{nT+\nu-k},$$

with $\sum_{k=-\infty}^{\infty} |\psi_k(\nu)| < \infty$ for each season ν. Causality conditions for PARMA series are simply that $\det(\Phi(z)) \neq 0$ for all z with $|z| \le 1$. Expressing this condition in terms of the PARMA model coefficients is a difficult unresolved problem, though it is known that a PAR(1) model is causal if and only if $|\prod_{\nu=1}^{T} \phi_1(\nu)| < 1$. Explicit conditions for causality of a PAR(2) model are unknown (explicit conditions for AR(2) causality are $|\phi_2| < 1$, $\phi_1 + \phi_2 < 1$, and $\phi_2 - \phi_1 < 1$).

Estimation of PARMA model coefficients parallels that for ARMA models. Specifically, moment methods work well in PAR settings (see Pagano (1978)). Maximum likelihood methods are considered in Vecchia (1985a) and Basawa and Lund (2001) and are better for models with a moving-average component. The Innovations likelihood equation (2.12) still holds but (2.14) is modified to

$$\widehat{X}_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} + \sum_{j=1}^{q} \theta_j(\nu)(X_{nT+\nu-j} - \widehat{X}_{nT+\nu-j})$$

and $v_{nT+\nu} = \sigma^2(\nu)$ for $nT + \nu > \max(p, q)$.

A practical issue with PARMA models lies with parsimony. For example, a PAR(1) model for daily temperatures (ignoring leap years and taking $T = 365$) has 365 autoregressive parameters and 365 white noise parameters. Jones and Brelsford (1967) suggest constraining PARMA parameters to be a Fourier series with a few harmonics. This tactic works well with meteorological data, where seasonal changes in the coefficients are relatively "smooth". Wavelet expansions can be considered if season to season changes are more abrupt (e.g. Vidakovic, 1999). Lund, Shao and Basawa (2006) and Anderson, Tesfaye and Meerschaert (2007) develop an asymptotic theory of parameter estimation for PARMA models having general parameter constraints.

5. The ARCH/GARCH paradigm

The models considered so far have been built from homoscedastic (constant variance) errors or periodic homoscedastic errors. Yet, many series, particularly those in financial settings, display periods of large local variabilities followed by relative sojourns of tranquil behavior. The autoregressive conditional heteroscedastic (ARCH) model of Engle (1982) and the generalized autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986) were introduced to describe such behavior. ARCH/GARCH models permit the conditional variance of the next observation to depend on the last few observations, thus allowing the conditional variance to change over time while leaving the unconditional variance constant. These models are well suited for financial series where highly variable observations cluster in time. As risk is a key component in many financial decisions, ARCH/GARCH models have become staples for forecasting volatility changes in the financial markets.

It is common to examine the log price ratios in finance. Specifically, let P_t be the price of an asset at time t and let $X_t = \log(P_t/P_{t-1})$. An ARCH(p) model for $\{X_t\}$ obeys

$$\begin{aligned} X_t &= \sigma_t Z_t \\ \sigma_t^2 &= \phi_0 + \phi_1 X_{t-1}^2 + \cdots + \phi_p X_{t-p}^2, \end{aligned} \tag{5.21}$$

where $\phi_0, \phi_1, \ldots, \phi_p \geq 0$ and $\{Z_t\}$ is zero mean unit variance independent and identically distributed (iid) noise, independent of $\{X_t\}$. Thus, an ARCH(p) series models log price ratios as noise with a time-varying conditional variance depending on the squares of the p previous log price ratios. From the upper equation in (5.21), it is easily seen that $\{X_t\}$ is uncorrelated (assuming that moments exist). However, the squares of an ARCH process will be dependent, implying that ARCH series cannot be iid or have marginal Gaussian distributions. Combining the two equations in (5.21) gives

$$X_t^2 = Z_t^2(\phi_0 + \phi_1 X_{t-1}^2 + \cdots + \phi_p X_{t-p}^2),$$

from which one can show that $\{X_t^2\}$ is stationary when all roots of $1 - \phi_1 z - \cdots - \phi_p z^p$ lie outside the complex unit circle. This solution is causal in the sense that X_t does not depend on $Z_{t+1}, Z_{t+2}, \ldots$. In the causal case where $\{Z_t\}$ is iid, $\{X_t\}$ is strictly stationary (i.e., $(X_1, \ldots, X_k)'$ and $(X_{1+h}, \ldots, X_{k+h})'$ have the same joint distribution for all positive integers k and integers h). From the assumed nonnegativity of the ϕ_js, all roots of $1 - \phi_1 z - \cdots - \phi_p z^p$ lie outside the complex unit circle if and only if $\phi_1 + \cdots + \phi_p < 1$. When this inequality holds, $\{X_t\}$ is zero mean strictly stationary noise with $\mathrm{Var}(X_t) = (1 - \phi_1 - \cdots - \phi_p)^{-1}\phi_0$ and

$$\mathrm{Var}(X_t | X_{t-1}, \ldots, X_{t-p}) = \phi_0 + \phi_1 X_{t-1}^2 + \cdots + \phi_p X_{t-p}^2. \tag{5.22}$$

Several properties of an ARCH(1) process deserve mention. If $\phi_1 \in [0, 1)$, then $E[X_t^2] < \infty$ and $\{X_t\}$ is strictly stationary noise that can be expressed in the causal form

$$X_t = \left(\phi_0 \sum_{j=0}^{\infty} \phi_1^j Z_t^2 Z_{t-1}^2 \cdots Z_{t-j}^2 \right)^{1/2}.$$

While the marginal distribution of X_t is symmetric about zero, it is also leptokurtic in that its kurtosis is high (the fourth central moment is greater than three times the second central moment). If $\phi_1 \in (0, 3^{-1/2})$, fourth moments of X_t are finite and $\{X_t^2\}$ has the AR(1) autocorrelation form $\text{Corr}(X_t^2, X_{t+h}^2) = \phi_1^{|h|}$. When $\phi_1 \in [3^{-1/2}, 1)$, $\{X_t^2\}$ is strictly stationary but has an infinite variance.

Moving-average components were introduced in the ARCH model for parsimony (a large p would otherwise frequently be needed), giving the GARCH (generalized ARCH) model. In a GARCH(p, q) model, the second equation in (5.21) is modified to

$$\sigma_t^2 = \phi_0 + \sum_{k=1}^{p} \phi_k X_{t-k}^2 + \sum_{j=1}^{q} \theta_j \sigma_{t-j}^2. \tag{5.23}$$

A GARCH(p, q) series is strictly stationary noise whenever $\phi_1 + \cdots + \phi_p < 1$. However, as with ARCH series, squares of a GARCH series may be correlated. A GARCH series is causal and has the finite variance

$$\text{Var}(X_t) = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p - \theta_1 - \cdots - \theta_q}$$

if and only if $\phi_1 + \cdots + \phi_p + \theta_1 + \cdots + \theta_q < 1$. GARCH expressions for $\text{Var}(X_t | X_{t-1}, \ldots, X_{t-\ell})$ are not as simplistic as those in the ARCH case of (5.22), but recursions that resemble (2.13) can be derived (Aknouche and Guerbyenne, 2006). Figure 5 shows autocorrelations of the square of a GARCH(1,1)

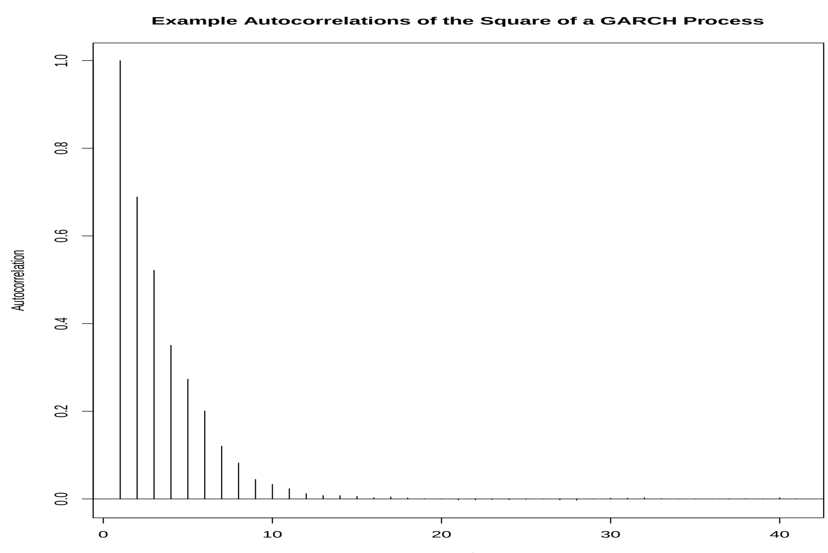

FIG 5. *Autocorrelations of the square of a GARCH(1,1) series with* $\phi_0 = 0.01, \phi_1 = 0.5$, *and* $\theta_1 = 0.2$.

series with parameters $\phi_0 = 0.01, \phi_1 = 0.5$, and $\theta_1 = 0.2$ (of course, autocorrelations of the non-squared series are zero).

Estimation of parameters in a causal and invertible GARCH model can be done via maximum likelihood. Suppose for instance that $\{Z_t\}$ is Gaussian. In this case, the marginal distribution of X_t cannot be Gaussian (otherwise $\{X_t\}$ would be iid); however, the conditional distribution of X_t given $X_{t-1}, X_{t-2}, \ldots,$ (the infinite past) is Gaussian. Hence, an approximate conditional likelihood L of all model parameters $\boldsymbol{\alpha}$, conditioned on $X_1, \ldots, X_m$, where $m = \max(p, q)$, is

$$L(\boldsymbol{\alpha}) = \prod_{t=m+1}^{n} f_{\boldsymbol{\alpha}}(X_t | X_{t-1}, \ldots, X_{t-m}). \tag{5.24}$$

In (5.24), $f_{\boldsymbol{\alpha}}(X_t | X_{t-1}, \ldots, X_{t-m})$ is the conditional density of X_t given $X_{t-1}, \ldots, X_{t-m}$. When $\{Z_t\}$ is Gaussian, $f_{\boldsymbol{\alpha}}$ can be taken as a Gaussian density with zero mean. For an ARCH model with $t > p$, the conditional variance $\mathrm{Var}(X_t | X_{t-1}, \ldots, X_{t-m})$ coincides with the right hand side of (5.22). For general GARCH processes, $\mathrm{Var}(X_t | X_{t-1}, \ldots, X_{t-m})$ can be approximated, without asymptotic loss of precision in the resulting likelihood estimators, by η_t^2, which is recursively defined by

$$\eta_t^2 = \phi_0 + \sum_{k=1}^{p} \phi_k X_{t-k}^2 + \sum_{j=1}^{q} \theta_j \eta_{t-j}^2,$$

with the understanding that $\eta_t = X_t = 0$ for $t \le 0$. In some financial applications, a non-normal and/or heavy-tailed $f_{\boldsymbol{\alpha}}$ distribution is preferred. Non-normal parametric choices for $f_{\boldsymbol{\alpha}}$ include the t, double exponential, and generalized double exponential distributions.

GARCH models have been generalized to allow positive and negative returns to impart different volatility effects. For example, the exponential GARCH (EGARCH) (Nelson, 1991) model allows for asymmetric return dynamics. This model is governed by

$$h_t = \phi_0^* + \sum_{k=1}^{p} \phi_k^* \frac{|X_{t-k}| + \gamma_k X_{t-k}}{\sigma_{t-k}} + \sum_{j=1}^{q} \theta_j^* h_{t-j},$$

where $h_t = \log(\sigma_t^2)$, the γ_i are nonnegative, and the superscript * is placed on the ARMA coefficients to signify that they possibly differ from those in (5.23) because of the logarithm. In this setup, when X_{t-k} is positive (i.e., there is "good news"), one can regard the total effect of X_{t-k} as $(1 + \gamma_k)|X_{t-k}|$; when X_{t-k} is negative (i.e., there is "bad news"), one can regard the total effect of X_{t-k} as $(1 - \gamma_k)|X_{t-k}|$. See Zivot and Wang (2006) for further discussion.

Another GARCH variant is the IGARCH model of Engle and Bollerslev (1986). Analogous to ARIMA series, a unit root is placed in the model by requiring that $\phi_1 + \cdots + \phi_p + \theta_1 + \cdots + \theta_q = 1$. IGARCH models are not stationary, but produce sample paths that have relatively longer sojourns of high variability in comparison to causal GARCH series. Mikosch and Starica (2004) is a good recent reference about this process.

5.1. Comments and open problems

Similar to models built from homoscedastic errors, checking assumptions is vital. Much current research examines multivariate GARCH (MGARCH) models (see Section 7.2). Several open problems, as detailed by Bauwens, Laurent and Rombouts (2006) and Silvennoinen and Teräsvirta (2009), exist in this area. These include: developing more flexible specifications of the MGARCH conditional correlation structure, developing two and higher step-ahead forecasting techniques, quantifying asymptotic properties of maximum likelihood estimates, and developing multivariate diagnostic tests. Also, estimation for multivariate GARCH series with a large dimension and intricate conditional correlations merits further consideration (Chib, Nardari and Shephard, 2006; Engle, 2002). Several other variants of the ARCH (GARCH) model have been proposed; Bollerslev (2008) provides a comprehensive glossary that introduces many of these models.

6. Long memory models

As Section 2 shows, ARMA models have short memory autocovariances in that $|\gamma(h)|$ decays to zero at a geometric rate in h. Situations where autocovariances decay more slowly arise in econometrics, hydrology, and other scientific disciplines. This section considers ARMA type models that allow for longer memory autocovariances, including autoregressive fractionally integrated moving-average (ARFIMA) models and ARFIMA variants that describe seasonal characteristics or changing variances. Although multivariate and continuous-time long memory ARMA models exist, for example the vector ARFIMA (VARFIMA) model of Ravishanker and Ray (2002) and the continuous-time autoregressive fractionally integrated moving-average (CARFIMA) model (Tsai and Chan, 2005; Brockwell and Marquardt, 2005; Tsai, 2009), we restrict our discussion to discrete-time univariate processes.

Several definitions of long memory (also termed long-range dependence) have been proposed and studied (see Palma, 2007, Chapter 3). For us, long memory has the Section 2 meaning that $\sum_{h=0}^{\infty} |\gamma(h)| = \infty$.

6.1. ARFIMA models

The series $\{X_t\}$ is called an ARFIMA(p, d, q) process if it satisfies

$$\phi(B)X_t = \theta(B)(1-B)^{-d}Z_t, \tag{6.25}$$

where $-1/2 < d < 1/2$. As before, we assume that ϕ and θ have no common roots and that $\{Z_t\}$ is zero mean white noise with variance σ^2. The quantity $(1-B)^{-d}$ is the fractional differencing operator defined by the binomial expansion

$$(1-B)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} B^j,$$

where $\Gamma(\cdot)$ is the usual gamma function with the convention that $\Gamma(x) = \Gamma(1+x)/x$ for $x \in (-1,0)$. The ARFIMA process was originally proposed by Granger and Joyeux (1980) and Hosking (1981).

If the zeros of $\phi(\cdot)$ lie outside the complex unit circle and $d \in (-1/2, 1/2)$, (6.25) has the unique zero mean stationary solution

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}, \tag{6.26}$$

where the ψ_ks are determined by expanding the ratio

$$\Psi(z) := \sum_{k=-\infty}^{\infty} \psi_k z^k = \frac{(1-z)^{-d}\theta(z)}{\phi(z)}$$

into a power series and equating coefficients (it is instructive to compare this to (2.4)). Further, if $\phi(\cdot)$ has no roots inside the complex unit circle, then (6.26) is causal in that $\psi_k = 0$ for $k < 0$ in (6.26). Theorem 13.2.2 of Brockwell and Davis (1991) establishes these and other ARFIMA facts. Bondon and Palma (2007) extend these results to $d \in (-1, 1/2)$.

The spectral density of a causal ARFIMA series is

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|1-e^{-i\lambda}|^{-2d} |\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}, \quad \lambda \in [-\pi, \pi),$$

and the lag h autocovariance decays to zero asymptotically at the power law rate with exponent $2d-1$. Thus, the key ARFIMA feature is that ARFIMA series have long memory when $d \in (0, 1/2)$. When $d = 0$, the process reduces to an ARMA(p,q) series. The long memory case where $d \in (0, 1/2)$ is our focus here; however, ARFIMA series with $d \in (-1/2, 0)$ are sometimes referred to as having intermediate memory. Notice that $\lim_{\lambda \downarrow 0} f(\lambda) = \infty$ when $d \in (0, 1/2)$, implying that frequency zero aspects (i.e., long period structures) are prominent in ARFIMA series.

The autocovariance and autocorrelation functions of the ARFIMA$(0,d,0)$ model are known, for $h \geq 0$, as

$$\begin{aligned} \gamma(0) &= \sigma^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}, \\ \text{Corr}(X_t, X_{t+h}) &= \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)} = \prod_{k=1}^{h} \frac{k-1+d}{k-d}. \end{aligned}$$

The autocovariances for general ARFIMA(p,d,q) process are more tedious to quantify. Exact expressions are given in Palma (2007) and were first established by Sowell (1992). Figure 6 shows the autocorrelations of an ARFIMA(3,d,2) model with $d = 0.3$ and ARMA coefficients taken as those used in the ARMA(3,2) model of Section 2. Notice how slowly $\rho(h)$ decays to zero.

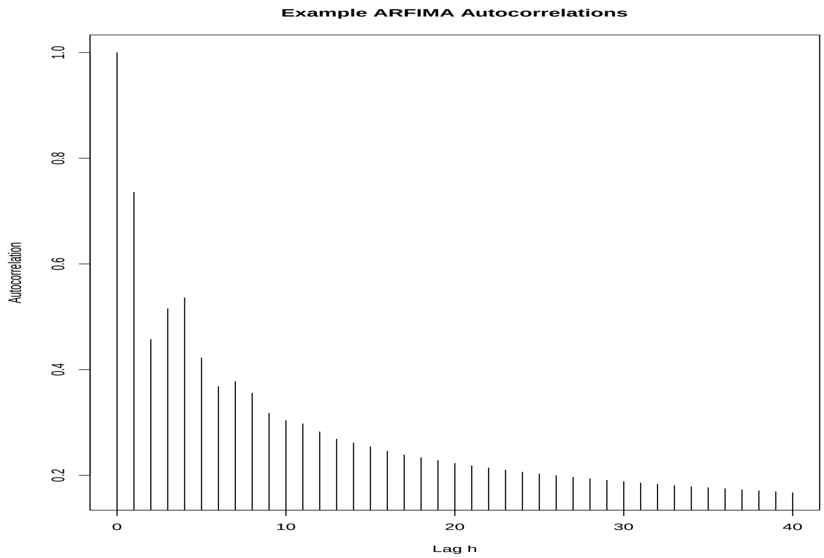

FIG 6. *Autocorrelation function of the ARFIMA(3,d,2) model with* $d = 0.3$ *and* $\phi_1 = 1/2$, $\phi_2 = 1/3$, $\phi_3 = -1/3$, $\theta_1 = 1/2$, $\theta_2 = -1/3$, *and* $\sigma^2 = 1$ *(identical ARMA parameters as in Figures 1 and 2).*

ARFIMA parameters can be estimated with Gaussian maximum likelihood methods. As no versions of (2.13) or (2.14) exist, one reverts back to the multivariate normal density function to evaluate the likelihood. Specifically, with $\boldsymbol{\alpha}$ denoting all ARFIMA parameters, the Gaussian likelihood function $L(\boldsymbol{\alpha})$ (unconditional) is

$$-2\log L(\boldsymbol{\alpha}) = n\log(2\pi) + \log(\det \Gamma_{\boldsymbol{\alpha}}) + \boldsymbol{X}'\Gamma_{\boldsymbol{\alpha}}^{-1}\boldsymbol{X}, \tag{6.27}$$

where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ and $\Gamma_{\boldsymbol{\alpha}}$ is the covariance matrix of $\boldsymbol{X}$. Exact evaluation of (6.27) for ARFIMA models was first considered by Sowell (1992), who first calculated the exact autocovariances of the model with parameter $\boldsymbol{\alpha}$ and then computed the quadratic form $\boldsymbol{X}'\Gamma_{\boldsymbol{\alpha}}^{-1}\boldsymbol{X}$ via the Durbin-Levinson algorithm (see Brockwell and Davis, 1991, Chapter 5). For large sample sizes, evaluating the exact likelihood function becomes a serious computational issue. A different approach for calculating the likelihood proceeds using the so-called splitting method proposed by Bertelli and Caporin (2002) to calculate the ARFIMA autocovariances. This method was implemented by Chen, Hurvich and Lu (2006) in conjunction with a preconditioned conjugate gradient method that facilitates fast accurate evaluation of the likelihood function (6.27). For a complete description of this approach, see Chen, Hurvich and Lu (2006) and the references therein.

Alternative methods for ARFIMA parameter estimation include state space techniques (Chan and Palma, 1998; Palma, 2007), Bayesian approaches

(Pai and Ravishanker, 1998; Ko and Vannucci, 2006), the Haslett-Raftery method (Haslett and Raftery, 1989), and approximate maximum likelihood based on the so-called Whittle approximation (e.g., Shumway and Stoffer, 2006; Palma, 2007). These methods are described in Chapters 4 and 8 of Palma (2007). A good review of ARFIMA likelihood estimation methods is found in Chan and Palma (2006).

For ARFIMA(p, d, q) models, it is difficult to identify p and q from plots of the sample autocorrelation and partial autocorrelation functions (as is often done in ARMA(p, q) model fitting). One method for estimating p and q proceeds by first estimating d using the approach of Geweke and Porter-Hudak (1983) (which does not depend on p and q). With this estimated d, one then estimates the remaining ARMA parameters by exploiting the fact that $\{(1-B)^d X_t\}$ is an ARMA(p, q) series. For more detail, see Brockwell and Davis (1991).

6.2. Periodic and seasonal long memory models

Long memory models with periodic and seasonal features are possible to construct. For example, a PARFIMA model would combine PARMA and long-memory dynamics by examining solutions to a periodic difference equation of the form

$$\phi_\nu(B) X_{nT+\nu} = \theta_\nu(B)(1-B)^{-d_\nu} Z_{nT+\nu}. \tag{6.28}$$

In (6.28), $\{Z_{nT+\nu}\}$ is zero mean periodic white noise with variance $\sigma^2(\nu)$ during season ν, $1 \le \nu \le T$; $\phi_\nu(\cdot)$ and $\theta_\nu(\cdot)$ are the season ν AR and MA polynomials, respectively, which are assumed to have no common roots or individual roots inside or on the unit circle; and $d_\nu \in [0, 1/2)$ for each season ν. Similar formulations to (6.28) have been proposed, see Ooms and Franses (2001) and Hui and Li (1995). It is expected that solutions to (6.28) would be periodic in the sense of (4.20) and that the autocovariance function of the seasonal series $\{X_{nT+\nu}\}_{n=-\infty}^{\infty}$ (which is stationary) would exhibit long memory for each fixed season ν. However, the properties of solutions to PARFIMA difference equations need to be formalized. In the above formulation, each season is allowed a distinct long memory parameter (d_ν for season ν).

Long memory variants of SARMA series have been more extensively explored. In particular, Gegenbauer autoregressive moving-average (GARMA) processes (Gray, Zhang and Woodward, 1989), k-factor GARMA processes (Woodward, Cheng and Gray, 1998), and seasonal autoregressive fractionally integrated moving-average (SARFIMA) processes have been proposed. Below, we briefly tour these processes.

The SARFIMA model permits long memory autocovariances and persistence at period T (loosely meaning that autocovariances at lags which are multiples of the period T are relatively larger than other autocovariances). In particular, the SARFIMA model satisfies the difference equation

$$\phi(B^T)\phi^*(B) X_t = \theta(B^T)\theta^*(B)(1-B^T)^{-d_T}(1-B)^{-d} Z_t, \tag{6.29}$$

where $\phi(\cdot)$, $\phi^*(\cdot)$, $\theta(\cdot)$, and $\theta^*(\cdot)$ are polynomials assumed to have no common roots or roots inside or on the unit circle, $d, d_T \in [0, 1/2)$ and $0 < d + d_T < 1/2$. In (6.29), $\{Z_t\}$ is zero mean time-homogeneous white noise with variance σ^2. Under these conditions, the SARFIMA model is stationary and has the spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta^*(e^{-i\lambda})|^2}{|\phi^*(e^{-i\lambda})|^2} \frac{|\theta(e^{-i\lambda T})|^2}{|\phi(e^{-i\lambda T})|^2} [2\{1 - \cos(\lambda)\}]^{-d} [2\{1 - \cos(T\lambda)\}]^{-d_T},$$

for $\lambda \in [-\pi, \pi]$. The distinction for the practitioner is that an ARFIMA model is appropriate for general long memory processes (the pole in the spectral density is at frequency zero), whereas SARFIMA models have additional poles in the spectrum at frequencies $\lambda = \pm 2\pi j/T$ $(j = 1, 2, \ldots, \lfloor T/2 \rfloor)$, meaning that persistence of features with period T are also present in SARFIMA series.

Gray, Zhang and Woodward (1989) introduced a different way of modeling long memory accompanied with a persistent periodic component. This approach is known as the Gegenbauer ARMA model (GARMA) and is governed by

$$\phi(B)(1 - 2uB + B^2)^d X_t = \theta(B) Z_t, \tag{6.30}$$

where $u \in [-1, 1]$ is a parameter that controls the frequency at which the long memory occurs and d governs the rate of decay of the autocovariance. Such a model is called a Gegenbauer ARMA (GARMA) series because of the Gegenbauer expansion

$$(1 - 2uB + B^2)^{-d} = \sum_{n=0}^{\infty} C_n B^n,$$

where

$$C_n = \sum_{k=0}^{[n/2]} \frac{(-1)^k (2u)^{n-2k} \Gamma(d - k + n)}{k!(n - 2k)! \Gamma(d)}.$$

Note that when $u = 1$, the GARMA model reduces to an ARFIMA model.

Gray, Zhang and Woodward (1989) show that when $|u| < 1$ and $0 < d < 1/2$, solutions to the GARMA difference equation are zero mean and stationary; moreover, solutions are causal when all roots of $\phi(\cdot)$ lie outside the unit circle. In the case where $|u| = 1$, a stationary causal solution exists when $0 < d < 1/4$. When there is no AR or MA component in the model, autocovariances take the explicit form

$$\gamma(h) = \sigma^2 \sum_{m=0}^{\infty} C_m C_{m+h}; \tag{6.31}$$

however, this sum is slow to converge and much care is needed when calculating $\gamma(h)$ via (6.31) (see Woodward, Cheng and Gray (1998) and the references therein).

The spectral density of a GARMA series is

$$f(\lambda) = c \frac{|\theta(e^{i\lambda})|^2}{|\phi(e^{i\lambda})|^2} (\cos(\lambda) - u)^{-2d}, \quad \lambda \in [-\pi, \pi),$$

where $c = \sigma^2/(\pi 2^{2d+1})$ is a constant whose value is not overly important in what follows. The spectral density at frequency $\cos^{-1}(u)$ is unbounded and is called the Gegenbauer frequency.

The model in (6.30) extends to describe series with k persistent periodic components simultaneously. Such a model is called the k-factor GARMA model (Woodward, Cheng and Gray, 1998) and is governed by

$$\phi(B)\prod_{j=1}^{k}(1-2u_jB+B^2)^{d_j}X_t = \theta(B)Z_t. \tag{6.32}$$

The spectral density of a k-factor GARMA process is given by

$$f(\lambda) = c\frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}\prod_{j=1}^{k}|\cos(\lambda)-u_j|^{-d_j}, \tag{6.33}$$

where $c > 0$ is a constant and the u_js are in [-1,1] and are assumed unrepeated.

The spectral density in (6.33) has k poles, one at each frequency $\cos^{-1}(u_j)$, $j = 1, \ldots, k$. Suppose all roots of $\phi(\cdot)$ lie outside the unit circle and that the u_js are distinct. If $d_j \in (0, 1/2)$ whenever $|u_j| < 1$ and $d_j < 1/4$ whenever $|u_j| = 1$, then $\{X_t\}$ in (6.32) is stationary and has long memory. The proof is given in Theorem 2.3 of Woodward, Cheng and Gray (1998).

When the u_js are known and the model is causal and invertible, Gaussian maximum likelihood estimators can be devised and are consistent, asymptotically normal, and efficient (see Palma and Chan (2005) and Theorem 12.1 in Palma (2007) for details).

6.3. Heteroscedastic long memory models

Models with conditionally heteroscedastic variances and long memory autocovariances are worth briefly mentioning. One such model is the ARFIMA(p, d, q)-GARCH(r, s), which satisfies the set of equations

$$\begin{aligned}\phi(B)X_t &= \theta(B)(1-B)^{-d}Z_t,\\ Z_t &= \epsilon_t\sigma_t,\\ \sigma_t^2 &= \phi_0+\sum_{k=1}^{r}\phi_kZ_{t-k}^2+\sum_{j=1}^{s}\theta_j\sigma_{t-j}^2.\end{aligned} \tag{6.34}$$

It is instructive to compare to (5.21) and (5.23) to (6.34). In (6.34), $\sigma_t^2 = E(X_t^2|\mathcal{F}_{t-1})$, $\mathcal{F}_{t-1}$ is the sigma algebra generated by the infinite past X_{t-1}, $X_{t-2}, \ldots$, and $\{Z_t\}$ is zero mean unit variance noise assumed to be independent of $\{X_t\}$. The GARCH coefficients $\phi_1, \ldots, \phi_r$ and $\theta_1, \ldots, \theta_s$ are assumed to be positive (so that variances cannot become negative). When $\sum_{k=1}^{r}\phi_k+\sum_{j=1}^{s}\theta_j < 1$ and $d \in [0, 1/2)$, a zero mean strictly stationary solution $\{X_t\}$ exists that is white noise; however, $\{X_t^2\}$ will have a stationary long memory autocovariance structure. It is possible to relax the positivity constraints on the GARCH coefficients somewhat; Nelson and Cao (1992) and Tsai and Chan (2008) provide the necessary conditions.

Although it is frequently assumed that $\{Z_t\}$ is Gaussian, t and other heavy-tailed marginal distributions can better accommodate the marginal features often encountered with financial data. Chapter 6 in Palma (2007) and Ling and Li (1997) discuss this and other estimation aspects.

7. Multivariate models

Researchers are frequently interested in modeling interrelationships among multiple variables. Multivariate time series models should be considered when the components are correlated. Examples include economic indicators across multiple countries and stock prices of competing companies. Consider K time series, $\{X_{k,t}\}$ for $k = 1, \ldots, K$. This collection of series is called K-variate stationary if for each $i \neq j$, $\{X_{i,t}\}$ and $\{X_{j,t}\}$ are stationary in a univariate sense and the cross-covariance $\text{Cov}(X_{i,t}, X_{j,t+h})$ only depends on h. When stationary, the cross-covariance function is denoted by $\gamma_{ij}(h) = \text{Cov}(X_{it}, X_{j,t+h})$.

Let $\boldsymbol{X}_t = (X_{1,t}, \ldots, X_{K,t})'$. To begin, we consider the case where $E[X_{k,t}] \equiv 0$ for all $k \in \{1, \ldots, K\}$. The covariance matrix at lag h is simply

$$\mathbf{\Gamma}(h) = E[(\boldsymbol{X}_t - \boldsymbol{\mu})(\boldsymbol{X}_{t+h} - \boldsymbol{\mu})'] = \begin{bmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \cdots & \gamma_{1k}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) & \cdots & \gamma_{2k}(h) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k1}(h) & \gamma_{k2}(h) & \cdots & \gamma_{kk}(h) \end{bmatrix}.$$

Multivariate stationary time series can be equivalently described in the spectral domain. This is because every stationary autocovariance $\mathbf{\Gamma}(\cdot)$ admits the spectral representation

$$\mathbf{\Gamma}(h) = \int_{-\pi}^{\pi} e^{ih\lambda} \boldsymbol{f}(\lambda) d\lambda,$$

where element j, k of $\boldsymbol{f}(\lambda)$, denoted as $f_{j,k}(\lambda)$, can be expressed as

$$f_{j,k}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{jk}(h) \exp(-ih\lambda)$$

assuming that $\sum_{h=-\infty}^{\infty} |\gamma_{jk}(h)| < \infty$.

An important issue in multivariate time series is that of cointegration. In particular, a K-variate series $\{\mathbf{X}_t\}$ is *cointegrated* if each component series is integrated of order one, (i.e., nonstationary with unit root behavior) while some linear combination of the series, $\{\mathbf{a}'\mathbf{X}_t\}$, is stationary for some K-dimensional nonzero vector $\mathbf{a}$. Hamilton (1994, Chapter 19) provides a comprehensive discussion.

While long-term changes can arise in the components of $\{\mathbf{X}_t\}$, cointegration still ties these components together. In this context, the nonzero vector $\mathbf{a}$ is known as the *cointegrating vector*.

Suppose that $\{\mathbf{a}'\mathbf{X}_t\}$ is stationary, then $\{b\mathbf{a}'\mathbf{X}_t\}$ is stationary for any nonzero scalar b. This implies that the cointegrating vector $\mathbf{a}$ is not unique. In practice, one typically takes the first element of $\mathbf{a}$ to be unity. A complete description of the properties of cointegrating vectors and their relation to ARMA processes is provided in Hamilton (1994).

The remainder of this Section focuses on several aspects of multivariate ARMA models. Specifically, Section 7.1 presents vector autoregressive moving-average models. Here, we describe conditions for causality and invertibility. Section 7.2 then considers multivariate GARCH sequences. Section 7.3 closes with some brief comments and open problems.

7.1. VARMA models

The multivariate generalization of an ARMA series is the vector autoregressive moving-average (VARMA or MARMA) series. A K-dimensional series $\{\boldsymbol{X}_t\}$ is said to be a VARMA series of autoregressive order p and moving-average order q if it is a solution to

$$\boldsymbol{X}_t - \boldsymbol{\Phi}_1\boldsymbol{X}_{t-1} - \cdots - \boldsymbol{\Phi}_p\boldsymbol{X}_{t-p} = \boldsymbol{Z}_t + \boldsymbol{\Theta}_1\boldsymbol{Z}_{t-1} + \cdots + \boldsymbol{\Theta}_q\boldsymbol{Z}_{t-q}, \tag{7.35}$$

with $\boldsymbol{\Theta}_p \neq \mathbf{0}, \boldsymbol{\Theta}_q \neq \mathbf{0}$, and $\{\boldsymbol{Z}_t\}$, defined for fixed t by $\boldsymbol{Z}_t = (Z_{1,t}, \ldots, Z_{K,t})'$, is zero mean multivariate white noise with covariance matrix $E[\boldsymbol{Z}_t\boldsymbol{Z}_t'] = \boldsymbol{\Sigma}$. The coefficient matrices $\boldsymbol{\Phi}_i, i = 1, \ldots, p$, and $\boldsymbol{\Theta}_j, j = 1, \ldots, q$, are $K \times K$ matrices.

The autoregressive and moving-average polynomials $\boldsymbol{\Phi}(\cdot)$ and $\boldsymbol{\Theta}(\cdot)$ are

$$\boldsymbol{\Phi}(z) = \boldsymbol{I} - \boldsymbol{\Phi}_1 z - \cdots - \boldsymbol{\Phi}_p z^p \quad \text{and} \quad \boldsymbol{\Theta}(z) = \boldsymbol{I} + \boldsymbol{\Theta}_1 z + \cdots + \boldsymbol{\Theta}_q z^q$$

for a complex-valued z; here, $\boldsymbol{I}$ is the $K \times K$ identity matrix. Using the backshift operator, (7.35) becomes

$$\boldsymbol{\Phi}(B)\boldsymbol{X}_t = \boldsymbol{\Theta}(B)\boldsymbol{Z}_t.$$

Causality and invertibility of the VARMA model imposes that solutions to $\det(\boldsymbol{\Phi}(z)) = 0$ and $\det(\boldsymbol{\Theta}(z)) = 0$ lie outside the complex unit circle, respectively. A causal VARMA model has the linear representation

$$\boldsymbol{X}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{Z}_{t-j}, \tag{7.36}$$

where the $\boldsymbol{\Psi}_j$s are found by equating coefficients in the relationship

$$\boldsymbol{\Psi}(z) := \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j z^j = \boldsymbol{\Phi}(z)^{-1}\boldsymbol{\Theta}(z).$$

Causality implies that $\sum_{j=0}^{\infty} |\boldsymbol{\Psi}_j| < \infty$ in a component-by-component sense. Evaluating covariances via (7.36) provides

$$\boldsymbol{\Gamma}(h) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{j+h}\boldsymbol{\Sigma}\boldsymbol{\Psi}_j'.$$

The spectral density matrix of the causal VARMA(p, q) process is

$$\boldsymbol{f}(\lambda) = \frac{1}{2\pi}\boldsymbol{\Psi}(e^{-i\lambda})\boldsymbol{\Sigma}\boldsymbol{\Psi}(e^{-i\lambda})'.$$

Assumptions beyond causality and invertibility are needed to ensure that VARMA parameters can be uniquely identified in terms of model autocovariances. It is possible for "seemingly unrelated" VARMA(p, q) models to generate the exact same sample path of data. Such non-identifiability occurs when two VARMA models — say $\boldsymbol{\Phi}(B)\boldsymbol{X}_t = \boldsymbol{\Theta}(B)\boldsymbol{Z}_t$ and $\boldsymbol{\Phi}^*(B)\boldsymbol{X}_t = \boldsymbol{\Theta}^*(B)\boldsymbol{Z}_t$ — are related through an invertible "polynomial operator" $\boldsymbol{U}(B)$ via

$$\boldsymbol{\Phi}^*(B) = \boldsymbol{U}(B)\boldsymbol{\Phi}(B) \quad \text{and} \quad \boldsymbol{\Theta}^*(B) = \boldsymbol{U}(B)\boldsymbol{\Theta}(B),$$

with the stipulation that the orders of $\boldsymbol{\Phi}^*$ and $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}^*$ and $\boldsymbol{\Theta}$ are the same. Then $(\boldsymbol{\Phi}^*(B)/\boldsymbol{\Theta}^*(B))X_t = (\boldsymbol{\Phi}(B)/\boldsymbol{\Theta}(B))X_t$, the two representations have the same $\boldsymbol{\Psi}_j$s and hence generate the same sample path from a fixed realization of $\{\boldsymbol{Z}_t\}$ (such VARMA models are called observationally equivalent). An example of such a phenomenon occurs in a two dimensional VARMA(1,1) process. Specifically, take

$$\begin{aligned} \boldsymbol{U}(B) &= \begin{bmatrix} 1 & \gamma B \\ 0 & 1 \end{bmatrix}; \ \boldsymbol{\Phi}(B) = \begin{bmatrix} 1-\phi_{11}B & -(\gamma+\phi_{12})B \\ 0 & 1 \end{bmatrix}; \\ \boldsymbol{\Theta}(B) &= \begin{bmatrix} 1+\theta_{11}B & (\theta_{12}-\gamma)B \\ 0 & 1 \end{bmatrix} \end{aligned}$$

and observe that the products

$$\boldsymbol{U(B)}\boldsymbol{\Phi}(B) = \begin{bmatrix} 1-\phi_{11}B & -\phi_{12}B \\ 0 & 1 \end{bmatrix} \ \text{and} \ \boldsymbol{U(B)}\boldsymbol{\Theta}(B) = \begin{bmatrix} 1+\theta_{11}B & \theta_{12}B \\ 0 & 1 \end{bmatrix}$$

do not depend on the choice of γ (one can verify that the parameter choices are causal and invertible when $|\phi_{11}| < 1$ and $|\theta_{11}| < 1$). The VARMA non-identifiability issue arises only when both p and q are both positive.

Several authors have investigated constraints beyond causality and invertibility that ensure that VARMA parameters are identifiable (Hannan, 1969, 1970, 1976, 1979; Deistler and Hannan, 1981; Hannan and Deistler, 1987; Lütkepohl, 1991, 2005; Reinsel, 1997; Dufour and Jouini, 2005). Our development here parallels Chapter 7.1 of Lütkepohl (2005).

One condition guaranteeing VARMA parameter identifiability is the so-called echelon form, which restricts where the non-zero coefficients in the VARMA matrix representation lie. To avoid trite work, we assume that the white noise covariance matrix $\boldsymbol{\Sigma}$ is nonsingular. Let $\Phi_{i,j}(z)$ and $\Theta_{i,j}(z)$ be the (i, j)th elements of the matrix AR and MA polynomials for $1 \leq i, j \leq K$; these quantities are all finite order polynomials. An echelon form requirement is that the degree of $\Phi_{i,j}(z)$ and $\Theta_{i,j}(z)$ is the same — call it p_i — and does not depend on the column index j. Clarifying, the order of $\Phi_{i,j}(z)$ and $\Theta_{i,j}(z)$ are the same for all

i, j — call this p_i — and the order of $\Phi_{i,j}(z)$ and $\Theta_{i,j}(z)$ can depend on i but not on j. The echelon form further imposes that

$$\begin{aligned}
\boldsymbol{\Phi}_{jj}(B) &= 1 - \sum_{\ell=1}^{p_j} (\boldsymbol{\Phi}_\ell)_{jj} B^\ell, \quad 1 \le j \le K; \\
\boldsymbol{\Phi}_{ij}(B) &= - \sum_{\ell=p_i-p_{ij}+1}^{p_i} (\boldsymbol{\Phi}_\ell)_{ij} B^\ell, \quad 1 \le i \neq j \le K; \\
\boldsymbol{\Theta}_{ij}(B) &= 1_{[i=j]} + \sum_{\ell=1}^{p_i} (\boldsymbol{\Theta}_\ell)_{ij} B^\ell, \quad 1 \le i, j \le K,
\end{aligned}$$

where the orders

$$p_{ij} = \begin{cases} \min(p_i + 1, p_j), & i \ge j \\ \min(p_i, p_j), & i < j \end{cases}.$$

In the above equations, $(\boldsymbol{\Phi}_\ell)_{i,j}$ denotes the i, jth element of $\boldsymbol{\Phi}_\ell$, for example. While some of the coefficients in the above representation can be zero (i.e., it is not necessary that $p = q$ in (7.35)), the echelon form restricts where the zero coefficients can appear. The mathematics behind the equations above is formidable and is rooted in McMillan degrees of polynomials and Kronecker indices; see Section 7.1 of Lütkepohl (2005) for this and other identifiable VARMA forms.

The parameters in a causal, invertible, and uniquely identified VARMA(p, q) model can be estimated using Gaussian maximum likelihood in a similar manner to ARMA(p, q) models. In particular, Innovations forms of (2.13) and (2.14) carry over with matrix quantities replacing their univariate counterparts. The Gaussian likelihood $L(\boldsymbol{\alpha})$ of all model parameters $\boldsymbol{\alpha}$, conditional on the first $\max(p, q)$ observations, is

$$\begin{aligned}
-2\log(L(\boldsymbol{\alpha})) &= d(n - \max(p,q))\log(2\pi) + \sum_{t=\max(p,q)+1}^{n} \det(\boldsymbol{V}_j) \\
&\quad + \sum_{t=\max(p,q)+1}^{n} (\boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t)' \boldsymbol{V}_j^{-1} (\boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t).
\end{aligned}$$

The likelihood estimators are asymptotically normal and $\sqrt{n}$-consistent. Hannan (1970), Dunsmuir and Hannan (1976), Reinsel (1997) and Lütkepohl (2005) are good references discussing VARMA estimation issues. When estimating VAR parameters via maximum likelihood, state-space methods can also be efficiently used; see Durbin and Koopman (2001) for a comprehensive overview. Bayesian approaches to VARMA parameter estimation have also been developed; see Ravishanker and Ray (1997) for a detailed discussion.

7.2. *Multivariate GARCH*

Multivariate GARCH (MGARCH) models have received much recent attention. The primary issue involves providing dynamics rich enough to adequately reflect the conditional variance/covariances while keeping the model parsimonious enough to permit estimation and interpretation. It is essential that any model specification have a positive definite covariance structure.

Let $\{\mathbf{r}_t\}$ be a zero mean stochastic process of dimension K. Further, let $\mathcal{F}_{t-1}$ denote the sigma algebra generated by the infinite past up to and including time $t-1$. Assume that $\mathbf{r}_t$ given $\mathcal{F}_{t-1}$, is conditionally heteroscedastic in that $\mathbf{r}_t = \mathbf{H}_t^{1/2}\boldsymbol{\eta}_t$, where $\mathbf{H}_t = \{h_{ijt}\}$ is the conditional covariance matrix of $\mathbf{r}_t$ and $\{\boldsymbol{\eta}_t\}$ is iid zero mean noise with $\mathrm{E}[\boldsymbol{\eta}_t\boldsymbol{\eta}_t'] = \mathbf{I}$. This defines the general MGARCH framework (Silvennoinen and Teräsvirta, 2009), where $\{\mathbf{r}_t\}$ exhibits no linear dependence structure. There are several ways to specify $\mathbf{H}_t$, including various parametric specifications.

One such specification models $\mathbf{H}_t$ directly and is known as the VEC-GARCH model of Bollerslev, Engle and Wooldridge (1988). This model generalizes univariate GARCH models and can be written as

$$\mathrm{vech}(\mathbf{H}_t) = \mathbf{c} + \sum_{j=1}^{q} \mathbf{A}_j \mathrm{vech}(\mathbf{r}_{t-j}\mathbf{r}_{t-j}') + \sum_{j=1}^{r} \mathbf{B}_j \mathrm{vech}(\mathbf{H}_{t-j}),$$

where $\mathrm{vech}(\mathbf{c})$ is the usual operator that stacks columns of the lower triangular part of the matrix $\mathbf{c}$ into a $K(K+1)/2$ vector, and $\mathbf{A}_j$ and $\mathbf{B}_j$ are parameter matrices of dimension $K(K+1)/2 \times K(K+1)/2$. Although this model is rather flexible, conditions need to be imposed to ensure a positive definite conditional covariance $\mathbf{H}_t$. See Silvennoinen and Teräsvirta (2009) for a comprehensive discussion.

Several other MGARCH formulations and extensions have been proposed. Bauwens, Laurent and Rombouts (2006) and Silvennoinen and Teräsvirta (2009) comprehensively review multivariate GARCH models. These reviews discuss nonparametric/ semiparametric methods in addition to model-based approaches and goodness-of-fit testing.

7.3. Comments and open problems

A practical issue in VARMA modeling lies with model parsimony. In particular, the number of VARMA parameters increases rapidly with increasing K (the curse of dimensionality). Parsimony constraints for VARMA models have been considered in Tsay (1989) and Lütkepohl (2005). Much current research involves extending the ARMA variants in this article to multivariate settings. For example, Ursu and Duchesne (2009) recently studied vector PARMA models; multivariate versions of some of the count models in the next section also present an open area of research. Finally, the development of parsimonious multivariate stationary series models with long memory and intricate cross-correlation structures is an active area of current research.

8. Count models

Although the models described thus far form a rich class capable of describing many types of series, their utility is limited in modeling non-negative integer-valued time series — the so-called count series. In fact, ARMA models, while adept at describing second moment features of many series, are clumsy in accounting for marginal distribution structures. Distributional aspects are especially important when the counts are small, such as the number of claims incurred by an insurance company, yearly major hurricanes counts, or incidence rates of a rare disease.

Significant progress has been made in modeling stationary series of counts over the last 20 years. MacDonald and Zucchini (1997), Davis, Dunsmuir and Wang (1999), Davis, Dunsmuir and Streett (2003) and Kedem and Fokianos (2002) are useful references. In this section, we discuss a popular ARMA variant used to model count series, the so-called INARMA model.

8.1. INARMA models

INARMA models use a thinning operator $\circ$ in ARMA-type equations. To define the thinning operator, suppose that X is a non-negative integer random variable. Let $\rho \in [0,1]$ and set

$$\rho \circ X = \sum_{i=1}^{X} Y_i,$$

where $\{Y_i\}$ are iid Bernoulli trials, independent of X, each with success probability ρ.

INARMA methods use thinning operators in ARMA-like equations to produce integer-valued series. For example, an INAR(1) series obeys the branching process equation (with immigration)

$$X_t = \rho \circ X_{t-1} + Z_t, \tag{8.37}$$

where $\{Z_t\}$ is an iid non-negative integer-valued sequence with mean μ and variance σ^2. Here, Z_t is interpreted as the number of new immigrants joining the population from time $t-1$ to time t and $\rho \circ X_{t-1}$ is the number of inhabitants at time $t-1$ that are still alive at time t. When $\rho \in [0,1)$ and $\{X_t\}$ is stationary, one can recurse (8.37) to obtain its unique solution:

$$X_t = \sum_{j=0}^{\infty} \rho^j \circ Z_{t-j}. \tag{8.38}$$

Evaluating moments with (8.38) gives $E[X_t] = \mu/(1-\rho)$ and

$$\text{Cov}(X_t, X_{t+h}) = \frac{\rho^{|h|}(\rho\mu + \sigma^2)}{1-\rho^2}.$$

An important class of INAR(1) processes arises when $\{Z_t\}$ has a Poisson distribution (Poisson INAR(1)). Vector INAR(1) models based on multinomial thinning exist, but we do not discuss them here. Comprehensive discussions of these aspects can be found in MacDonald and Zucchini (1997), Kedem and Fokianos (2002) and the references therein.

An extension of the INAR(1) model is the pth order autoregressive INAR(p) model, which is governed by the difference equation

$$X_t = \sum_{i=1}^{p} \rho_i \circ X_{t-i} + Z_t. \tag{8.39}$$

Here, $\rho_j \in [0, 1)$. As in the ARCH section, the model in (8.39) has a stationary and causal solution when $\sum_{i=1}^{p} \rho_i < 1$ (see also Alzaid and Al-Osh (1990)).

Integer moving-average process of order q (INMA(q)) obey

$$X_t = \beta_0 \circ Z_t + \beta_1 \circ Z_{t-1} + \cdots + \beta_q \circ Z_{t-q}, \tag{8.40}$$

Here, $\beta_0 = 1$, $\beta_i \in [0, 1)$ for $1 \leq i \leq q$, and all thinnings are performed independently. Expressions for the mean, variance, and autocovariance of X_t in (8.40) can be found in Section 5.1.6 of Kedem and Fokianos (2002). INARMA(p, q) models are defined by combining (8.39) and (8.40) in the obvious manner.

Some comments about INARMA(p, q) series are worth making. First, because the ρ_js are all nonnegative, series with negative correlations cannot be made from this model class. Second, all causal INARMA(p, q) series necessarily have short memory, though longer memory models could be devised through fractional differencing schemes. It is also not evident how to make a particular marginal distribution with the INARMA class. For example, because the support set of an INARMA(p, q) series is unbounded, one cannot construct an INARMA(p, q) series with a binomial distribution for each fixed t.

An alternative to the INARMA class are the discrete autoregressive moving-average (DARMA) models (Jacobs and Lewis, 1978a,b). This class is capable of producing any discrete distribution as a marginal distribution for X_t. This class of models takes on a mixture model flavor and thus its' exposition differs slightly from the other ARMA model variants. As such, we defer discussion of this class to the comprehensive treatments in MacDonald and Zucchini (1997) and Kedem and Fokianos (2002).

8.2. Comments and open problems

Several open problems with count series are apparent. First, models for count series that have negative correlations need to be devised. Second, count models with long memory autocovariances need to be developed and studied; here, Quoreshi (2008) provides a good start. Third, count models with periodic properties need to be developed. This is especially important in that monthly rare disease counts, hurricane counts, etc., frequently show periodic behavior with a definitive season of occurrence.

We mention that a new method of generating integer count series was recently proposed in Cui and Lund (2009). Their methods simply use a renewal process to generate a correlated sequence of Bernoulli trials. Independent versions of these processes can then be mixed and/or superimposed to generate count series with *any* discrete marginal distribution structure. While these methods sometimes have ARMA-type representations and sometimes do not, they easily generate count series with negative correlations and/or long memory properties.

9. Concluding remarks

ARMA models and their variants provide a rich class of models capable of describing and forecasting a broad array of observed time series. The exposition provided here describes many of the most popular ARMA variants. Since the area is extremely voluminous, our description is necessarily limited and there remain other ARMA model variants worthy of exposition that were not described here.

Although substantial research has been devoted to ARMA models and their variants, there are still many avenues for future research. For example, parsimonious specifications and goodness-of-fit diagnostics for multivariate GARCH models provide two such opportunities. Additional open directions lie in the areas of long memory modeling for periodic processes and count data. The ever-expanding scope of ARMA models and their variants will continue to provide future research opportunities beyond what is described here. Also, as computational abilities advance, many new variants will likely be introduced.

Practical implementation of ARMA models and their variants has seen unprecedented growth since the conception of the R programming language (R Development Core Team, 2010). In fact, the *CRAN Task Views* website (`http://cran.r-project.org/web/views/`) provides up to date documentation under the "TimeSeries" link. This link provides comprehensive details to every aspect of the R infrastructure, contributed packages and (internal and contributed) datasets relating to time series. The information is comprehensive while being organized and succinct. Given this suite of computing resources, and others, opportunities for time series researchers and practitioners have become limitless.

10. An ARMA acronym list

This section serves as a simple dictionary of selected ARMA variants. Below we list the various ARMA acronyms in alphabetical order and provide references where the processes are discussed further. Some of the variants listed below are not discussed in the paper but are provided for completeness. The ARMA literature is vast and thus some acronyms are necessarily redundant (e.g., BARMA stands for both Bilinear Autoregressive Moving-Average and Binomial Autoregressive Moving-Average). Essentially, this section is a modernized version of Granger (1982).

ARCH Autoregressive Conditionally Heteroscedastic
Engle (1982).

ARFIMA Autoregressive Fractionally Integrated Moving-Average
Granger and Joyeux (1980); Hosking (1981); Geweke and Porter-Hudak (1983); Haslett and Raftery (1989); Sowell (1992); Kokoszka and Taqqu (1995, 1996); Pai and Ravishanker (1998); Chan and Palma (1998, 2006); Bertelli and Caporin (2002); Chen, Hurvich and Lu (2006); Ko and Vannucci (2006); Bondon and Palma (2007); Palma (2007).

ARFIMA-GARCH Autoregressive Fractionally Integrated Moving-Average - Generalized Autoregressive Conditionally Heteroscedastic
Ling and Li (1997); Palma (2007).

ARIMA Autoregressive Integrated Moving-Average
Brockwell and Davis (1991, 2002); Shumway and Stoffer (2006); Box, Jenkins and Reinsel (2008).

ARMA Autoregressive Moving-Average
Ansley (1979); Brockwell and Davis (1991, 2002); Box, Jenkins and Reinsel (2008); Fuller (1996); Shumway and Stoffer (2006).

ARMAX Autoregressive Moving-Average X
Hannan (1976); Shumway and Stoffer (2006).

BARMA Bilinear Autoregressive Moving-Average
Granger and Andersen (1978); Subba Rao (1981); Liu and Brockwell (1988).

BARMA Binomial Autoregressive Moving-Average
Startz (2008).

CARFIMA Continuous-time Autoregressive Fractionally Integrated Moving-Average
Brockwell and Marquardt (2005); Tsai and Chan (2005); Tsai (2009).

CARMA Continuous Autoregressive Moving-Average
Jones (1980); Brockwell (1994); Stramer, Tweedie and Brockwell (1996); Tsai and Chan (2000); Brockwell (2001); Brockwell and Davis (2002).

DAR Discrete Autoregressive
MacDonald and Zucchini (1997); Kedem and Fokianos (2002).

DARMA Discrete Autoregressive Moving-Average
Jacobs and Lewis (1978a,b, 1983); MacDonald and Zucchini (1997); Kedem and Fokianos (2002).

EARMA Exponential Autoregressive Moving-Average
Lawrance and Lewis (1980).

EGARCH Exponential Generalized Autoregressive Conditionally Heteroscedastic
Nelson (1991).

FIGARCH Fractionally Integrated Generalized Autoregressive Conditionally Heteroscedastic
Palma (2007).

FIEGARCH Fractionally Integrated Exponential Generalized Autoregressive Conditionally Heteroscedastic
Palma (2007).

GARMA Gegenbauer Autoregressive Moving-Average
Gray, Zhang and Woodward (1989); Woodward, Cheng and Gray (1998); Palma and Chan (2005); Palma (2007).

GARCH Generalized Autoregressive Conditional Heteroscedastic
Bollerslev (1986).

IGARCH Generalized Integrated Autoregressive Conditional Heteroscedastic
Engle and Bollerslev (1986).

INAR Integer Autoregressive
Alzaid and Al-Osh (1990); MacDonald and Zucchini (1997); McKenzie (1988); Kedem and Fokianos (2002).

INARMA Integer Autoregressive Moving-Average
Kedem and Fokianos (2002).

MA Moving-Average
Brockwell and Davis (1991).

PAR Periodic Autoregression
Troutman (1979).

PARMA Periodic Autoregressive Moving-Average
Hannan (1955); Gladyšhev (1961); Jones and Brelsford (1967); Pagano (1978); Vecchia (1985a,b); Lund and Basawa (2000); Basawa and Lund (2001); Shao and Lund (2004); Lund, Shao and Basawa (2006); Anderson, Tesfaye and Meerschaert (2007).

RCAR Random Coefficient Autoregression
Resnick and Willekens (1991); Bougerol and Picard (1992).

SARFIMA Seasonal Autoregressive Fractionally Integrated Moving-Average
Palma (2007).

SARMA Seasonal Autoregressive Moving-Average
Brockwell and Davis (2002); Shumway and Stoffer (2006); Box, Jenkins and Reinsel (2008).

STARMA Space-Time Autoregressive Moving-Average
Pfeifer and Deutsch (1980).

TAR Threshold Autoregression
Tong and Lim (1980); Chan (1990).

VARFIMA Vector Autoregressive Fractionally Integrated Moving-Average
Ravishanker and Ray (2002).

VARMA Vector Autoregressive Integrated Moving-Average
Dunsmuir and Hannan (1976); Hannan (1979); Kohn (1979); Hannan and Deistler (1987); Lütkepohl (2005); Shumway and Stoffer (2006).

VAR Vector Autoregressive
Mann and Wald (1943); Lütkepohl (2005).

Acknowledgements

The authors would like to thank three anonymous referees whose suggestions helped improve this paper. Lund's research was supported by National Science Foundation (NSF) grant DMS-0905570.

References

Aknouche, A. and Guerbyenne, H. (2006). Recursive estimation of GARCH models. *Communications in Statistics-Simulation and Computation* **35** 925–938. MR2291371

Alzaid, A. A. and Al-Osh, M. (1990). An integer-valued *p*th-order autoregressive structure (INAR (*p*)) process. *Journal of Applied Probability* **27** 314–324. MR1052303

Anderson, P. L., Tesfaye, Y. G. and Meerschaert, M. M. (2007). Fourier-PARMA models and their application to river flows. *Journal of Hydrologic Engineering* **12** 462–472.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika* **66** 59–65. MR0529148

Basawa, I. V. and Lund, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis* **22** 651–663. MR1867391

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics* **21** 79–109. MR2225523

Bertelli, S. and Caporin, M. (2002). A note on calculating autocovariances of long-memory processes. *Journal of Time Series Analysis* **23** 503–508. MR1925261

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307–327. MR0853051

Bollerslev, T. (2008). Glossary to ARCH (GARCH). *CREATES Research Paper 2008-49*.

Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *The Journal of Political Economy* **96** 116–131.

Bondon, P. and Palma, W. (2007). A class of antipersistent processes. *Journal of Time Series Analysis* **28** 261–273. MR2345660

Bougerol, P. and Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *The Annals of Probability* **20** 1714–1730. MR1188039

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*, 4th ed. Wiley, New Jersey. MR2419724

Breidt, F. J., Davis, R. A. and Trindade, A. A. (2001). Least absolute deviation estimation for all-pass time series models. *Annals of Statistics* **29** 919–946. MR1869234

Brockwell, P. J. (1994). On continuous-time threshold ARMA processes. *Journal of Statistical Planning and Inference* **39** 291–303. MR1271562

Brockwell, P. J. (2001). Continuous-time ARMA processes. In *Stochastic Processes: Theory and Methods*, (D. N. Shanbhag and C. R. Rao, eds.). *Handbook of Statistics* **19** 249–276. Elsevier. MR1861726

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York. MR1093459

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*, 2nd ed. Springer, New York. MR1894099

Brockwell, P. J. and Marquardt, T. (2005). Lèvy-driven and fractionally integrated ARMA processes with continuous-time paramaters. *Statistica Sinica* **15** 477–494. MR2190215

Chan, K. S. (1990). Testing for threshold autoregression. *Annals of Statistics* **18** 1886–1894. MR1074443

Chan, N. H. (2002). *Time Series: Applications to Finance*. John Wiley & Sons, New York. MR1896231

Chan, N. H. and Palma, W. (1998). State space modeling of long-memory processes. *Annals of Statistics* **26** 719–740. MR1626083

Chan, N. H. and Palma, W. (2006). Estimation of long-memory time series models: A survey of different likelihood-based methods. *Advances in Econometrics* **20** 89–121. MR2498114

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*, 6th ed. Chapman & Hall/CRC, Boca Raton.

Chen, W., Hurvich, C. M. and Lu, Y. (2006). On the correlation matrix of the discrete Fourier transform and the fast solution of large Toeplitz systems for long-memory time series. *Journal of the American Statistical Association* **101** 812–822. MR2281249

Chernick, M. R., Hsing, T. and McCormick, W. P. (1991). Calculating the extremal index for a class of stationary sequences. *Advances in Applied Probability* **23** 835–850. MR1133731

Chib, S., Nardari, F. and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics* **134** 341–371. MR2328413

Cryer, J. D. and Chan, K. S. (2008). *Time Series Analysis: With Applications in R*. Springer, New York.

Cui, Y. and Lund, R. (2009). A new look at time series of counts. *Biometrika* **96** 781–792. MR2564490

Davis, R. A., Dunsmuir, W. T. M. and Wang, Y. (1999). Modeling time series of count data. In *Asymptotics, Nonparametrics and Time Series*, (S. Ghosh, ed.). *Statistics Textbooks Monograph* 63–113. Marcel Dekker, New York. MR1724696

Davis, R. A., Dunsmuir, W. and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika* **90** 777–790. MR2024757

Davis, R. A. and Resnick, S. I. (1996). Limit theory for bilinear processes with heavy-tailed noise. *The Annals of Applied Probability* **6** 1191–1210. MR1422982

Deistler, M. and Hannan, E. J. (1981). Some properties of the parameterization of ARMA systems with unknown order. *Journal of Multivariate Analysis* **11** 474–484. MR0637928

Dufour, J. M. and Jouini, T. (2005). Asymptotic distribution of a simple linear estimator for VARMA models in echelon form. *Statistical Modeling and Analysis for Complex Data Problems* 209–240. MR2189538

Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. *Advances in Applied Probability* **8** 339–364. MR0451578

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford. MR1856951

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121

Engle, R. F. (2002). Dynamic conditional correlation. *Journal of Business and Economic Statistics* **20** 339–350. MR1939905

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews* **5** 1–50. MR0876792

Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed. John Wiley & Sons, New York. MR1365746

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4** 221–238. MR0738585

Gladyšhev, E. G. (1961). Periodically correlated random sequences. *Soviet Math* **2** 385–388.

Granger, C. W. J. (1982). Acronyms in time series analysis (ATSA). *Journal of Time Series Analysis* **3** 103–107.

Granger, C. W. J. and Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht Göttingen. MR0483231

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* **1** 15–29. MR0605572

Gray, H. L., Zhang, N. F. and Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis* **10** 233–257. MR1028940

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey. MR1278033

Hannan, E. J. (1955). A test for singularities in Sydney rainfall. *Australian Journal of Physics* **8** 289–297.

Hannan, E. J. (1969). The identification of vector mixed autoregressive-moving average system. *Biometrika* **56** 223–225. MR0254998

Hannan, E. J. (1970). *Multiple Time Series*. John Wiley & Sons, New York. MR0279952

Hannan, E. J. (1976). The identification and parameterization of ARMAX and state space forms. *Econometrica* **44** 713–723. MR0479509

Hannan, E. J. (1979). The Statistical Theory of Linear Systems. In *Developments in Statistics* (P. R. Krishnaiah, ed.) 83–121. Academic Press, New York. MR0554178

Hannan, E. J. and Deistler, M. (1987). *The Statistical Theory of Linear Systems*. John Wiley & Sons, New York. MR0940698

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge. MR1059059

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Applied Statistics* **38** 1–50.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176. MR0614953

Hui, Y. V. and Li, W. K. (1995). On fractionally differenced periodic processes. *Sankhyā: The Indian Journal of Statistics, Series B* **57** 19–31. MR1394866

Jacobs, P. A. and Lewis, P. A. W. (1978a). Discrete time series generated by mixtures. I: Correlational and runs properties. *Journal of the Royal Statistical Society. Series B (Methodological)* **40** 94–105. MR0512147

Jacobs, P. A. and Lewis, P. A. W. (1978b). Discrete time series generated by mixtures II: Asymptotic properties. *Journal of the Royal Statistical Society. Series B (Methodological)* **40** 222–228. MR0517443

Jacobs, P. A. and Lewis, P. A. W. (1983). Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis* **4** 19–36. MR0711293

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22** 389–395. MR0585635

Jones, R. H. and Brelsford, W. M. (1967). Time series with periodic structure. *Biometrika* **54** 403–408. MR0223041

Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*. John Wiley & Sons, New Jersey. MR1933755

Ko, K. and Vannucci, M. (2006). Bayesian wavelet-based methods for the detection of multiple changes of the long memory parameter. *IEEE Transactions on Signal Processing* **54** 4461–4470.

Kohn, R. (1979). Asymptotic estimation and hypothesis testing results for vector linear time series models. *Econometrica* **47** 1005–1030. MR0537640

Kokoszka, P. S. and Taqqu, M. S. (1995). Fractional ARIMA with stable innovations. *Stochastic Processes and their Applications* **60** 19–47. MR1362317

Kokoszka, P. S. and Taqqu, M. S. (1996). Parameter estimation for infinite variance fractional ARIMA. *Annals of Statistics* **24** 1880–1913. MR1421153

Lawrance, A. J. and Lewis, P. A. W. (1980). The exponential autoregressive-moving average EARMA(p, q) process. *Journal of the Royal Statistical Society. Series B (Methodological)* **42** 150–161. MR0583349

Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* **92** 1184–1194. MR1482150

Liu, J. and Brockwell, P. J. (1988). On the general bilinear time series model. *Journal of Applied Probability* **25** 553–564. MR0954503

Lund, R. and Basawa, I. V. (2000). Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis* **21** 75–93. MR1766175

Lund, R., Shao, Q. and Basawa, I. (2006). Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics* **48** 33–47. MR2234777

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York. MR1125072

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, New York. MR2172368

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman & Hall/CRC, Boca Raton. MR1692202

Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11** 173–220. MR0009291

Marriott, J., Ravishanker, N., Gelfand, A. and Pai, J. (1996). Bayesian analysis of ARMA processes: Complete sampling-based inference under exact likelihoods. In *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner* (D. Berry, K. Challoner and J. Geweke, eds.) 243–256. Wiley, New York. MR1367323

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability* **20** 822–835. MR0968000

Mikosch, T. and Starica, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* **86** 378–390.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59** 347–370. MR1097532

Nelson, D. B. and Cao, C. Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics* **10** 229–235.

Ooms, M. and Franses, P. H. (2001). A seasonal periodic long memory model for monthly river flows. *Environmental Modelling & Software* **16** 559–569.

Pagano, M. (1978). On periodic and multiple autoregressions. *Annals of Statistics* **6** 1310–1317. MR0523765

Pai, J. S. and Ravishanker, N. (1998). Bayesian analysis of autoregressive fractionally integrated moving-average processes. *Journal of Time Series Analysis* **19** 99–112. MR1624112

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. John Wiley & Sons, New Jersey. MR2297359

Palma, W. and Chan, N. H. (2005). Efficient estimation of seasonal long-range-dependent processes. *Journal of Time Series Analysis* **26** 863–892. MR2203515

Pfeifer, P. E. and Deutsch, S. J. (1980). A three-stage iterative procedure for space-time modeling. *Technometrics* **22** 35–47.

Prado, R. and West, M. (2010). *Time Series Modeling, Computation and Inference*. Chapman & Hall/CRC, Boca Raton. MR2655202

Quoreshi, A. M. M. S. (2008). A long memory count data time series model for financial application. *Preprint*.

R Development Core Team, (2010). R: A Language and Environment for Statistical Computing. `http://www.R-project.org`.

Ravishanker, N. and Ray, B. K. (1997). Bayesian analysis of vector ARMA models using Gibbs sampling. *Journal of Forecasting* **16** 177–194.

Ravishanker, N. and Ray, B. K. (2002). Bayesian prediction for vector ARFIMA processes. *International Journal of Forecasting* **18** 207–214.

Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*. Springer, New York. MR1451875

Resnick, S. I. and Willekens, E. (1991). Moving averages with random coefficients and random coefficient autoregressive models. *Communications in Statistics. Stochastic Models* **7** 511–525. MR1139067

Rootzén, H. (1986). Extreme value theory for moving average processes. *The Annals of Probability* **14** 612–652. MR0832027

Scotto, M. G. (2007). Extremes for solutions to stochastic difference equations with regularly varying tails. *REVSTAT–Statistical Journal* **5** 229–247. MR2365926

Shao, Q. and Lund, R. (2004). Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis* **25** 359–372. MR2062679

Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis and its Applications: With R Examples*, 2nd ed. Springer, New York. MR2228626

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. In *Handbook of Financial Time Series* (T. Andersen, R. Davis, J. Kreib, and T. Mikosch, eds.) Springer, New York.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* **53** 165–188. MR1185210

Startz, R. (2008). Binomial autoregressive moving average models with an application to U.S. recessions. *Journal of Business and Economic Statistics* **26** 1–8. MR2422056

Stramer, O., Tweedie, R. L. and Brockwell, P. J. (1996). Existence and stability of continuous time threshold ARMA processes. *Statistica Sinica* **6** 715–732. MR1410743

Subba Rao, T. (1981). On the theory of bilinear time series models. *Journal of the Royal Statistical Society. Series B (Methodological)* **43** 244–255. MR0626772

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)* **42** 245–292.

Troutman, B. M. (1979). Some results in periodic autoregression. *Biometrika* **66** 219–228. MR0548187

Tsai, H. (2009). On continuous-time autoregressive fractionally integrated moving average processes. *Bernoulli* **15** 178–194. MR2546803

Tsai, H. and Chan, K. S. (2000). A note on the covariance structure of a continuous-time ARMA process. *Statistica Sinica* **10** 989–998. MR1789804

Tsai, H. and Chan, K. S. (2005). Maximum likelihood estimation of linear continuous time long memory processes with discrete time data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **67** 703–716. MR2210688

Tsai, H. and Chan, K. S. (2008). A note on inequality constraints in the GARCH model. *Econometric Theory* **24** 823–828. MR2420388

Tsay, R. S. (1989). Parsimonious parameterization of vector autoregressive moving average models. *Journal of Business and Economic Statistics* **7** 327–341.

Tunnicliffe-Wilson, G. (1979). Some efficient computational procedures for high order ARMA models. *Journal of Statistical Computation and Simulation* **8** 301–309. MR0527226

Ursu, E. and Duchesne, P. (2009). On modelling and diagnostic checking of vector periodic autoregressive time series models. *Journal of Time Series Analysis* **30** 70–96. MR2488636

Vecchia, A. V. (1985a). Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics* **27** 375–384.

Vecchia, A. V. (1985b). Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Journal of the American Water Resources Association* **21** 721–730.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. John Wiley & Sons, New York. MR1681904

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed. Springer, New York. MR1482232

WOLD, H. (1954). *A Study in the Analysis of Stationary Time Series*. Almquist & Wiksell, Stockholm. MR0061344

WOODWARD, W. A., CHENG, Q. C. and GRAY, H. L. (1998). A k-factor GARMA long-memory model. *Journal of Time Series Analysis* **19** 485–504. MR1650550

ZIVOT, E. and WANG, J. (2006). *Modeling Financial Time Series with S-PLUS*, 2nd ed. Springer, New York. MR2000944