Journal of Econometrics 189 (2015) 492–506

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Generalized ARMA models with martingale difference errors[☆]

Tingguo Zheng[a], Han Xiao[b], Rong Chen[b,*]

[a] Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China

[b] Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA

ARTICLE INFO

Article history:
Available online 20 March 2015

Keywords:
Beta time series
Ergodicity
Gamma time series
Gaussian pseudo-likelihood
Realized volatility

ABSTRACT

The analysis of non-Gaussian time series has been studied extensively and has many applications. Many successful models can be viewed as special cases or variations of the generalized autoregressive moving average (GARMA) models of Benjamin et al. (2003), where a link function similar to that used in generalized linear models is introduced and the conditional mean, under the link function, assumes an ARMA structure. Under such a model, the 'transformed' time series, under the same link function, assumes an ARMA form as well. Unfortunately, unless the link function is an identity function, the error sequence defined in the transformed ARMA model is usually not a martingale difference sequence. In this paper we extend the GARMA model in such a way that the resulting ARMA model in the transformed space has a martingale difference sequence as its error sequence. The benefit of such an extension are four-folds. It has easily verifiable conditions for stationarity and ergodicity; its Gaussian pseudo-likelihood estimator is consistent; standard time series model building tools are ready to use; and its MLE's asymptotic distribution can be established. We also proposes two new classes of non-Gaussian time series models under the new framework. The performance of the proposed models is demonstrated with simulated and real examples.

1. Introduction

Researchers have shown an increasing interest in non-Gaussian time series models, with problems coming from various applications. These studies can be divided into two categories: innovations-based and data-based. The innovations-based models make distributional assumptions on the noise or innovation process. Heavy tailed and asymmetric distributions are often employed for modeling time series exhibiting heavy tails and skewness, including Student-t, Gamma, generalized error distributions (GED), generalized logistic distributions and their skewed versions, see Bollerslev (1987), Nelson (1991), Hansen (1994), Li and McLeod (1988), Tiku et al. (2000), Wong and Bian (2005), Bondon (2009) and others. The data-based approach is based on the distributional assumptions on the observed time series data. Several models under this approach have been developed recently, including the autoregressive conditional duration (ACD) models by Engle and Russell (1998), multiplicative error models (MEM) by Engle (2002) and Engle and Gallo (2006), Poisson and negative binomial models for discrete-valued or count time series models by Davis et al. (2003), Davis and Wu (2009), Fokianos et al. (2009) and Fokianos and Fried (2010), Beta autoregressive moving average (ARMA) models for rates or proportional time series by Rocha and Cribari-Neto (2009), binomial ARMA models for binary data by Startz (2008), and others. These models are either special cases or variations of the generalized autoregressive moving average (GARMA) models under the general framework of Benjamin et al. (2003). It is also a time series extension of the generalized linear models (GLM) of McCullagh and Nelder (1989). Similar to GLM, the conditional mean (given past information in a time series setting) is modeled directly, through a link function, with an ARMA type of structure that is most commonly used for modeling time series.

The GARMA models have an intriguing feature—the time series, transformed using the link function, assumes an ARMA structure. Such a feature can potentially make model building, estimation and prediction very easy. It would also make investigation of the probabilistic properties of the series and asymptotic behavior of the estimators easier, if the error sequence in the ARMA formation is a martingale difference sequence (MDS). Unfortunately, unless

☆ Zheng's research was supported in part by the National Natural Science Foundation of China (No. 71371160, 11101341) and Program for New Century Excellent Talents in University (NCET-13-0509). Chen's research was supported in part by National Science Foundation grants DMS-0905763 and DMS-0915139. Xiao's research was supported in part by National Science Foundation grant DMS-1209091.

* Corresponding author.
E-mail address: rongchen@stat.rutgers.edu (R. Chen).

http://dx.doi.org/10.1016/j.jeconom.2015.03.040
0304-4076/

an identity link function is used, the GARMA model of Benjamin et al. (2003) does not have a MDS as its error sequence in the ARMA formation. On the other hand, the use of identity link in GARMA models is restrictive and often encounters various difficulties in modeling the underlying time series. For example, when modeling rates or proportional time series data, it is difficult to provide feasible parameter conditions to ensure all values of the conditional expectation be bounded between 0 and 1. For nonnegative time series, GARMA models with identity link function do not allow negative autocorrelations (Fokianos, 2012). Multiplicative error models and Poisson autoregressive models, where the ARMA coefficients must be constrained to be non-negative, have similar problems.

In this paper we extend the GARMA models so that the error sequence in the ARMA formulation is a MDS. We refer it as the *Martingalized GARMA* (M-GARMA) model. It continues to enjoy all the interpretations, the link to GLM models, and many of the properties of GAMRA model, but more importantly, with MDS as its error sequence, verifiable stationarity and ergodicity conditions are easy to obtain, its Gaussian pseudo-likelihood estimator is consistent, standard model building tools are ready to use, MLE's asymptotic behavior can be established, and predictions are easier to obtain.

Under the proposed setting, the model can be easily generalized to integrated M-GARMA and fractional integrated M-GARMA models, as martingale processes (i.e. integrated MDS) are well understood and well behaved. It can also be easily extended to have a joint dynamics of conditional mean and conditional variance structure.

The rest of this paper is organized as follows. Section 2 briefly reviews the existing GARMA model introduced by Benjamin et al. (2003) and introduces the M-GARMA model, including two special models, the log-Gamma-M-GARMA model and the logit-Beta-M-GARMA model. Section 3 provides a detailed analysis of the probabilistic properties of the M-GARMA model, including its stationarity and ergodicity conditions. In Section 4, we propose two estimators for M-GARMA models and investigate their theoretical properties. Model building and prediction issues are discussed as well. Section 5 carries out a simulation study of the two M-GARMA models introduced earlier. The finite sample properties of the estimators are studied and a fast model building approach is compared with full model diagnostic. Finally, in Section 6 we use the log-Gamma-M-GARMA model and logit-Beta-M-GARMA model to study realized volatilities and US personal saving rates, respectively.

2. GARMA models and martingalized GARMA models

2.1. GARMA models

Let $\{y_t\}$ be a (non-Gaussian) time series and $\mathcal{F}_t = \{y_t, y_{t-1}, \ldots\}$ be the σ-field generated by the information up to time t. We also denote μ_t as the conditional expectation of y_t given $\mathcal{F}_{t-1}$.

Benjamin et al. (2003) formulated the framework of GARMA (p, q) model under an exponential family distribution. Specifically, it assumes that the conditional distribution of y_t given its past follows

$$f(y_t|\mathcal{F}_{t-1}) = \exp\left\{\frac{y_t\vartheta_t - b(\vartheta_t)}{\varphi} + a(y_t, \varphi)\right\}, \tag{1}$$

where ϑ_t and φ are the canonical and scale parameters, and the conditional expectation and variance of y_t given $\mathcal{F}_{t-1}$ is given by $\mu_t = b'(\vartheta_t) = E(y_t|\mathcal{F}_{t-1})$ and $\mathrm{Var}(y_t|\mathcal{F}_{t-1}) = \varphi b''(\vartheta_t)$, respectively. Benjamin et al. (2003) further assumed that the conditional mean μ_t has the following ARMA structure

$$\eta_t \equiv g(\mu_t) = \nu + \sum_{j=1}^{p}\phi_j g(y_{t-j}) + \sum_{j=1}^{q}\delta_j[g(y_{t-j}) - \eta_{t-j}], \tag{2}$$

where $\phi = (\phi_1, \ldots, \phi_p)'$ and $\delta = (\delta_1, \ldots, \delta_q)'$ are the autoregressive and moving average parameters. The function $g(\cdot)$ is called a link function. It is assumed that the transformed mean follows a seemingly ARMA process. The quantity η_t is called the linear predictor. The link function $g(\cdot)$ is restricted to a one-to-one function and hence it can be inverted to obtain $\mu_t = g^{-1}(\eta_t)$. Benjamin et al. (2003) also included deterministic covariates in (2). For cleaner notation, we will exclude it in our model development but it can be easily included in all our models.

By adding $g(y_t) - \eta_t$ to both sides of (2), we have

$$g(y_t) = \nu + \sum_{j=1}^{p}\phi_j g(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q}\delta_j\varepsilon_{t-j}, \tag{3}$$

where $\varepsilon_t = g(y_t) - \eta_t = g(y_t) - g(\mu_t)$. Obviously (3) shows that under GARMA model, $g(y_t)$ assumes exactly a standard ARMA model formulation (Box and Jenkins, 1976). The only difference is that the error sequence is not necessarily a white noise sequence. Note that $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = E[g(y_t) \mid \mathcal{F}_{t-1}] - g(\mu_t)$ is not necessarily zero, unless $g(\cdot)$ is an identity function. Thus under this formulation, the noise sequence ε_t is not a MDS in most of the cases.

2.2. M-GARMA models

In the following, we propose to extend the GARMA model in such a way so to ensure the error sequence in (3) is a MDS. We call it a *martingalized* GARMA model (M-GARMA). Specifically, we assume the conditional distribution $p(y_t \mid \mathcal{F}_{t-1})$ can be parametrized as

$$p(y_t \mid \mathcal{F}_{t-1}) = f(y_t \mid \mu_t, \varphi), \tag{4}$$

where φ is a collection of time invariant parameters hence all past information is summarized in μ_t. In addition, let

$$g_\varphi(\mu_t) = \nu + \sum_{j=1}^{p}\phi_j h(y_{t-j}) + \sum_{j=1}^{q}\delta_j[h(y_{t-j}) - g_\varphi(\mu_{t-j})], \tag{5}$$

where $g_\varphi(\mu_t) = E[h(y_t) \mid \mathcal{F}_{t-1}]$ serves as the link function in the terminology of GLM.

By adding $h(y_t) - g_\varphi(\mu_t)$ to both sides of (5), we have

$$h(y_t) = \nu + \sum_{j=1}^{p}\phi_j h(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q}\delta_j\varepsilon_{t-j}, \tag{6}$$

where $\varepsilon_t = h(y_t) - g_\varphi(\mu_t)$. It is clear by this construction of the pair of link functions $(h(\cdot), g_\varphi(\cdot))$ that ε_t is now a MDS. In the following we refer to $g_\varphi(\cdot)$ as the *link function* and $h(\cdot)$ the *y-link function*.

Some remarks on different issues of the model are in order:

Remark 2.1 (*The Link Functions*)**.** In practice, it is often more convenient to start with a parameter free y-link function $h(\cdot)$ and work backwards to obtain the link function $g_\varphi(\cdot)$. As we demonstrate later that modeling is easier if the y-link function $h(\cdot)$ does not involve any unknown parameters. In this case, the ARMA orders and ARMA coefficients can be obtained directly using $h(y_t)$ series. On the other hand, there is no special difficulty when the link function $g_\varphi(\cdot)$ involves some unknown parameters.

Remark 2.2 (*From a Transformation Point of View*)**.** The impact of y-link function $h(\cdot)$ in M-GARMA model can be viewed as

Table 1
Some commonly used conditional distributions, their recommended y-link functions, the corresponding link functions, and the conditional variances of the resulting MDS.

	Density	$E[y_t \mid \mathcal{F}_{t-1}]$	$\mathrm{Var}[y_t \mid \mathcal{F}_{t-1}]$
Lognormal	$\log N(\log(\mu_t) - \sigma^2/2, \sigma^2)$	μ_t	$(e^{\sigma^2} - 1)\mu_t^2$
Gamma	$\mathrm{Gam}(c\mu_t^d, c\mu_t^{d-1})$	μ_t	μ_t^{2-d}/c
Inverse-Gamma	$\text{Inv-Gam}(c\mu_t^{d-1} + 1, c\mu_t^d)$	μ_t	$c\mu_t^{1+d}/(c\mu_t^{d-1} - 1)$
Weibull	$\mathrm{Weibull}(k, \mu_t/\Gamma(1 + k^{-1}))$	μ_t	$\mu_t^2\left[\frac{\Gamma(1+2k^{-1})}{\Gamma^2(1+k^{-1})} - 1\right]$
Beta	$\mathrm{Beta}(\tau\mu_t, \tau(1 - \mu_t))'$	μ_t	$\frac{\mu_t(1-\mu_t)}{1+\tau}$
Poisson	$\mathrm{Poisson}(\mu_t)$	μ_t	μ_t
	$h(y_t)$	$g_\varphi(\mu_t)$	$\mathrm{Var}[h(y_t) \mid \mathcal{F}_{t-1}]$
Lognormal	$\log(y_t)$	$\log\mu_t - \frac{1}{2}\sigma^2$	σ^2
Gamma	$\log(y_t)$	$\psi(c\mu_t^d) - (d - 1)\log(\mu_t) - \log(c)$	$\psi_1(c\mu_t^d)$
Inverse-Gamma	$\log(y_t)$	$d\log(\mu_t) + \log(c) - \psi(c\mu_t^{d-1} + 1)$	$\psi_1(c\mu_t^{d-1} + 1)$
Weibull	$\log(y_t)$	$\approx \log\mu_t - \frac{1}{2}\left[\frac{\Gamma(1+2k^{-1})}{\Gamma^2(1+k^{-1})} - 1\right]$	$\approx \frac{\Gamma(1+2k^{-1})}{\Gamma^2(1+k^{-1})} - 1$
Beta	$\log(y_t/1 - y_t)$	$\psi(\tau\mu_t) - \psi(\tau(1 - \mu_t))$	$\psi_1(\tau\mu_t) + \psi_1(\tau(1 - \mu_t))$
Poisson	$\sqrt{y_t}$	$\approx\sqrt{\mu_t}$	$\approx\frac{1}{4}$

Note: The functions $\psi(\cdot)$ and $\psi_1(\cdot)$ are the digamma and trigamma functions, respectively.

transforming the time series to a linear process. Transformation is an important tool for statistical analysis for improvements in model simplicity, variance stabilization, and precision of estimation. In time series analysis, it has been widely employed, often as the first step of analysis. However, it is simply and often irrationally assumed that the transformed series follows an ARMA model with a (Gaussian) white noise error sequence and all subsequent inferences are done under such an assumption. Also there have been a few general guidance and theoretical foundation on how to properly transform a time series. Although (6) is exactly under the form of an ARMA model with transformation, the proposed M-GARMA model starts with the conditional distribution, which serves as a foundation for parametrization of the model, the determination of the y-link function and parameter estimation. Eqs. (4) and (5) can be viewed as a special case of model (6), but it allows for a more solid foundation for theoretical development.

Remark 2.3 (*The Selection of the y-Link Function*)**.** In practice, the y-link function $h(\cdot)$ can be selected as a common strictly monotone continuous function such as logarithm, power, logit, reciprocal, probit and others. For example, the logarithm and square root functions are often used for positive data, the logit transformation is used for the data in the unit interval such as rates and proportions, and the reciprocal transformation can be used for non-zero data. Considerations of the range of μ_t and its corresponding space of the ARMA parameters are often of practical importance. For example, for proportional observations, an identical link function would result in difficult constraints on the ARMA parameter space to ensure the conditional mean stays within the meaningful boundary. For conditional Gamma distributions, an identical link function would often restrict the ARMA parameters to be positive.

A useful tool for constructing the y-link function is the Bartlett (1947) variance-stabilizing transformation. It aims to remove a mean/variance relationship, so that the variance becomes constant relative to the mean. Often the Delta method is used. For example, logarithm transformation for Gamma distribution, square root transformation for Poisson distribution, arcsine square root transformation for proportions (binomial data), are often used as the variance-stabilizing transformation. The Box–Cox power transformation of Box and Cox (1964), denoted as y_t^λ when $\lambda \neq 0$ and $\log y_t$ when $\lambda = 0$, is a family of transformations parametrized by λ that includes the logarithm, square root, and multiplicative inverse as special cases. It is often used as Gaussian transformation as well as for variance stabilization, hence can be considered as a y-link function.

When the conditional distribution belongs to an exponential family, and if $h(y_t)$ is a canonical sufficient statistic, then its conditional mean and variance depends on the log-partition function of the exponential distribution, with certain properties that are useful for estimation.

As the y-link function is the core component of the model and often a suitable choice may not be obvious, one may consider a nonparametric approach, estimating it from the data. For example, $h(\cdot)$ can be approximated by an expansion $\sum_{j=1}^K \gamma_j z_j(\cdot)$, where $\{z_j(\cdot)\}$ is a family of basis functions. In practice, it is often desirable to have a monotone y-link function. So existing methods on smoothing monotone functions may be applied here. See for example, Ramsay (1988, 1998) and He and Shi (1998) among others. With the nonparametrically estimated y-link function, we can further develop model specification tests to validate simple and "natural" choices of $h(\cdot)$ such as logarithm, logit and power transformations. Such an adaptive approach may make M-GARMA model more flexible in more applications, as well as more confident in its formation. We hope to study this semiparametric approach in the future.

Remark 2.4 (*Canonical Link Functions*)**.** In some cases, $g_\varphi(\mu_t) - h(\mu_t)$ is a constant with respect to μ_t. We call such a pair of link functions $(g_\varphi(\cdot), h(\cdot))$ *canonical* link functions. A pair of identity functions are canonical link functions. When the link functions are canonical, the proposed M-GARMA model can be reformulated to the GARMA model of Benjamin et al. (2003). For example, if the conditional distribution is a Gamma distribution in the form of $\mathrm{Gamma}(\alpha, \alpha/\mu_t)$, then $\log(\cdot)$ is a canonical link function since $E[\log(y_t) \mid \mathcal{F}_{t-1}] = \log(\mu_t) + \psi(\alpha) - \log(\alpha)$, where $\psi(\cdot)$ is the digamma function. In this case, let $\varepsilon_t = h(y_t) - h(\mu_t) - \psi(\alpha) + \log(\alpha)$, which is a MDS. Then (6) becomes $h(y_t) = \nu^* + \sum_{j=1}^p \phi_j h(y_{t-j}) + \varepsilon_t + \sum_{j=1}^q \delta_j \varepsilon_{t-j}$. When the conditional distribution is log-normal, the log function is canonical, see Table 1.

Remark 2.5 (*Inverting $g_\varphi(\cdot)$*)**.** To calculate the likelihood function, we will need the inverse of the link function $g_\varphi(\cdot)$. When the y-link function $h(\cdot)$ is monotone and the conditional cumulative probability function of the conditional distribution $f(\mu_t, \varphi)$ is monotone in μ_t (such as when the conditional distribution belongs to a location family), the link function $g_\varphi(\mu_t)$ is also monotone with respect to μ_t, hence invertible. It is often possible to restrict the parameter space of φ to make $g_\varphi(\cdot)$ invertible, when $g_\varphi(\cdot)$ is not invertible but continuous, there are only a finite number of distinct solutions of μ_t for $g_\varphi(\mu_t) = \eta_t$ in the practical range of usual problems. In such cases, we evaluate of the likelihood function at each of the solutions and use the one that maximizes $f(y_t \mid \mu_t, \varphi)$ as its 'generalized inverse'.

Remark 2.6 (*Approximation of the Link Function*)**.** The link function $g_\varphi(\mu_t)$ is determined by the conditional distribution and the y-link function $h(\cdot)$. In certain situations, the exact link function may be too complex and an approximation may be sufficient. Specifically, one may use a second-order Taylor approximation. Since $E[y_t - \mu_t|\mathcal{F}_{t-1}] = 0$ and $E[(y_t - \mu_t)^2|\mathcal{F}_{t-1}] = \mathrm{Var}(y_t|\mathcal{F}_{t-1})$, we have

$$g_\varphi(\mu_t) = E[h(y_t)|\mathcal{F}_{t-1}] \approx h(\mu_t) + \frac{1}{2!}h''(\mu_t)\mathrm{Var}(y_t|\mathcal{F}_{t-1}). \quad (7)$$

Although higher-order Taylor expansion may be more accurate, the approximation may be complex and not invertible. Experience shows in many problems the second-order Taylor expansion is sufficient. Linear approximation results in $g_\varphi(\mu_t) \approx h(\mu_t)$, the link function suggested by Benjamin et al. (2003).

Although the inverse of g_φ may not always exist analytically, the solution can always be found with numerical procedures as this is a one dimensional function. For our simulation and data analysis, we use the bisection method.

Remark 2.7 (*Linear Predictor*)**.** Note that, under the M-GARMA model, $\eta_t = g_\varphi(\mu_t)$ is a linear predictor of $h(y_t)$ and the above formulations can also be rewritten as

$$\eta_t \equiv g_\varphi(\mu_t) = \nu + \sum_{j=1}^{m} \tilde{\phi}_j h(y_{t-j}) + \sum_{j=1}^{q} \tilde{\delta}_j \eta_{t-j}, \quad (8)$$

where $m = \max\{p, q\}$, $\tilde{\phi}_j = \phi_j + \delta_j$ for $j = 1, \ldots, m$ and $\tilde{\delta}_j = -\delta_j$ for $j = 1, \ldots, q$, and $\phi_j = 0$ if $m > p$ and $\delta_j = 0$ if $m > q$. That is, the linear predictor η_t is a moving average of past transformed responses $h(y_{t-1}), \ldots, h(y_{t-p})$ and the past predictors $\eta_{t-1}, \ldots, \eta_{t-q}$. GARCH model assumes a very similar formulation.

Remark 2.8 (*Martingale Difference Tests*)**.** The fact that $\{\varepsilon_t\}$ is a MDS is critical for the validity of the pseudo Gaussian estimation discussed in Section 4.1. After a model has been fitted to the data, we can apply the martingale difference tests on the residuals as a model diagnostic tool. We refer the reader to Escanciano and Lobata (2009) for a thorough review of such tests. In our numerical analysis, we will apply the tests in Domínguez and Lobata (2003), Escanciano and Velasco (2006) and Kuan and Lee (2004).

Table 1 shows some conditional distributions that may be used in practice, along with recommended y-link functions. For lognormal, Gamma, inverse-Gamma, Weibull, Beta and Poisson distributions, we have analytic forms of the link functions under mean-parametrization with the logarithm y-link function. For the conditional Beta distribution with logit y-link function, we also have an analytic form for the corresponding link function. For Weibull with the logarithm y-link function and Poisson with the square-root y-link function, their approximated link functions, obtained with Taylor expansion, are shown.

2.3. Two specific M-GARMA models

Here we propose two specific M-GARMA models. They are designed to model two types of commonly encountered non-Gaussian time series. Their probabilistic properties will be established later in Section 3 and we will use them for the empirical study of various estimators.

Log Gamma-M-GARMA model: Consider the following Gamma-M-GARMA(p, q) model with the y-link function $h(y_t) = \log y_t$,

$$y_t|\mathcal{F}_{t-1} \sim \mathrm{Gam}(c\mu_t^d, c\mu_t^{d-1}),$$
$$\log y_t = \nu + \sum_{j=1}^{p} \phi_j \log y_{t-j} + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}, \quad (9)$$

with $\varepsilon_t = \log y_t - g_{c,d}(\mu_t) = \log y_t - \psi(c\mu_t^d) + (d-1)\log\mu_t + \log c$, where $c\mu_t^d$ and $c\mu_t^{d-1}$ are the shape and rate parameters of Gamma distribution, and μ_t is the conditional expectation of y_t based on the past information up to time $t-1$. The link function is given by $g_{c,d}(\mu_t) = \psi(c\mu_t^d) - (d-1)\log\mu_t - \log c$, and the conditional variance is $\psi_1(c\mu_t^d)$, i.e., $\mathrm{Var}[\varepsilon_t|\mathcal{F}_{t-1}] = \psi_1(c\mu_t^d)$, where $\psi(\cdot)$ and $\psi_1(\cdot)$ are digamma and trigamma functions, respectively.

This is a special parametrization for the Gamma distribution. It maintains the conditional mean μ_t as one of the parameters while allows flexibility to control the shape and scale. Although a Gamma distribution is determined by two parameters, our three-parameter parametrization is identifiable as long as μ_t changes over time, similar to settings such as random effect models.

When $d = 0$, the link function $g_{c,d}(\mu_t) = \log\mu_t + \psi(c) - \log c$ differs from the y-link function by a constant, and the conditional variance is also a constant, i.e. $\mathrm{Var}[\varepsilon_t|\mathcal{F}_{t-1}] = \psi_1(c)$. In this case, the model reduces to a canonical M-GARMA model in which the error process remains to be a MDS with the same function link and y-link functions.

Fig. 1 shows simulated processes under the model using different d and Fig. 2 shows the histogram of the corresponding marginal distributions. It can be seen that the parameter d controls the shape of the marginal distribution of the process significantly and the smaller d makes the process less 'normal'. Fig. 3 shows the link functions for different d (solid line). The dash line is the corresponding linear approximation $g_\varphi(\cdot) = h(\cdot)$ (see Remark 2.6). It can be seen that the linear approximation is more accurate with large d.

This specification of the ARMA process with logarithm y-link function is different from that used for the MEM models given by Engle (2002), Engle and Gallo (2006) and Brownlees et al. (2012) because they suggested using an identity transformation, i.e., $h(y_t) = y_t$.

Logit-Beta-M-GARMA model: Beta-M-GARMA model can be used for proportion time series where the observations take value in (0, 1). We consider the following Beta-M-GARMA(p, q) model with the logit y-link $h(y_t) = \mathrm{logit}(y_t) = \log[y_t/(1-y_t)]$,

$$y_t|\mathcal{F}_{t-1} \sim \mathrm{Beta}(\tau\mu_t, \tau(1-\mu_t)),$$
$$\mathrm{logit}(y_t) = \nu + \sum_{j=1}^{p} \phi_j \mathrm{logit}(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}, \quad (10)$$

with $\varepsilon_t = \mathrm{logit}(y_t) - g_\tau(\mu_t)$, where $\tau\mu_t$ and $\tau(1-\mu_t)$ are two positive shape parameters of Beta distribution, and μ_t is the conditional mean based on the past information up to time $t-1$. The link function and conditional variance are given by $g_\tau(\mu_t) = \psi(\tau\mu_t) - \psi(\tau(1-\mu_t))$ and $\mathrm{Var}[h(y_t)|\mathcal{F}_{t-1}] = \mathrm{Var}(\varepsilon_t|\mathcal{F}_{t-1}) = \psi_1(\tau\mu_t) + \psi_1(\tau(1-\mu_t))$ respectively. Rocha and Cribari-Neto (2009) proposed another form of Beta-GARMA model based on the class of Beta regression models of Ferrari and Cribari-Neto (2004). Their model is similar to those of Benjamin et al. (2003) and the error terms are not a MDS.

2.4. Some extensions

In this section we discuss several possible extensions of the M-GARMA model.

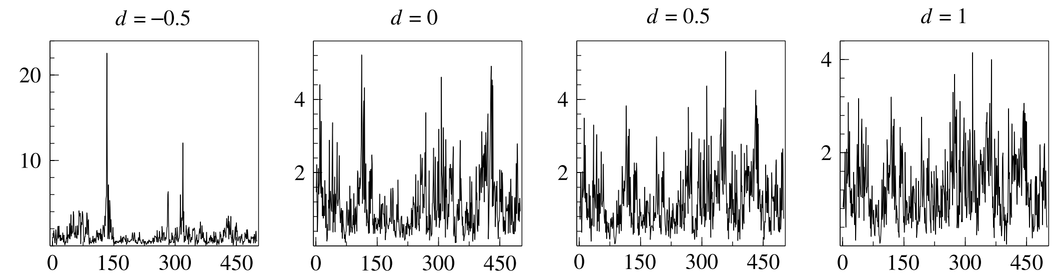

Fig. 1. Simulated log-Gamma-M-GARMA series with $c = 3$.

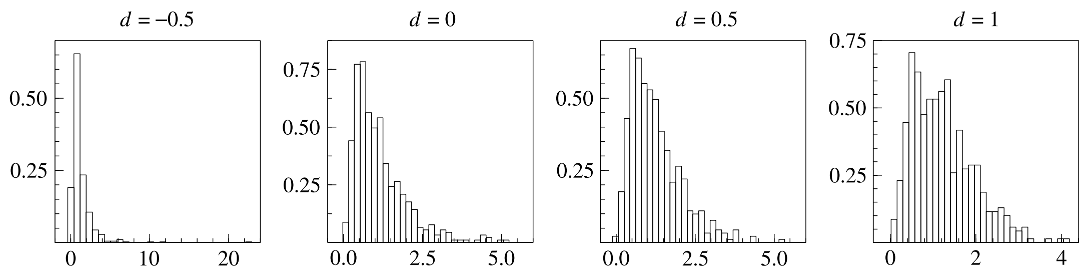

Fig. 2. Histogram of simulated log-Gamma-M-GARMA series with $c = 3$.

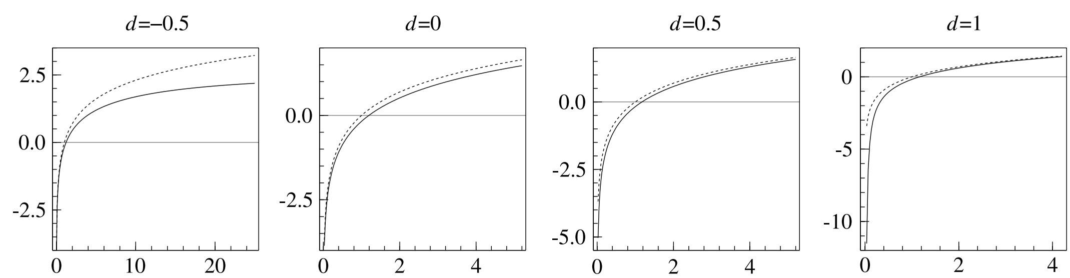

Fig. 3. The link function of the log-Gamma-M-GARMA model with $c = 3$.

M-GARIMA model: The M-GARMA model can be extended to have an integrated ARMA structure. Specifically, one may extend (6) to

$$\Delta^d h(y_t) = \sum_{j=1}^{p} \phi_j \Delta^d h(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}, \tag{11}$$

where $p(y_t \mid \mathcal{F}_{t-1}) = f(y_t \mid \mu_t, \varphi)$, $g_\varphi(\mu_t) = E[h(y_t) \mid \mathcal{F}_{t-1}]$ and $\varepsilon_t = h(y_t) - g_\varphi(\mu_t)$.

By expanding the difference operator in the left side of (11) and rearranging, and let $\eta_t = g_\varphi(\mu_t)$, we obtain the following representation similar to (5)

$$\eta_t = \sum_{k=1}^{d} \binom{d}{k} (-1)^k h(y_{t-k}) + \sum_{j=1}^{p} \phi_j \Delta^d h(y_{t-j}) + \sum_{j=1}^{q} \delta_j (h(y_{t-j}) - \eta_{t-j}). \tag{12}$$

M-GARFIMA model: Another extension is to include the fractionally integrated operator $(1 - L)^d$ $(0 < d < 1)$ into the M-GARMA model. Specifically, the new model is given by

$$\phi(L)(1 - L)^d h(y_t) = \nu + \delta(L)\varepsilon_t, \tag{13}$$

with $\varepsilon_t = h(y_t) - g_\varphi(\mu_t)$, where L is the backshift operator, $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, $\delta(L) = 1 + \delta_1 L + \cdots + \delta_q L^q$, and $g_\varphi(\mu_t) = E[h(y_t) \mid \mathcal{F}_{t-1}]$. The M-GARFIMA$(p, d, q)$ model can be seen as an extension of standard ARFIMA models (see Sowell (1992); Beran (1995)), and it can be used to model the long memory behavior of y_t.

M-GARMA–GARCH model: One can also extend the conditional distribution to have two or more time varying parameters, with joint or separate dynamics. For example, suppose the conditional distribution is

$$y_t \mid \mathcal{F}_{t-1} \sim f(y_t \mid \xi_{1t}, \xi_{2t}, \varphi). \tag{14}$$

Let $\eta_t = g_\varphi(\xi_{1t}, \xi_{2t}) = E[h(y_t) \mid \mathcal{F}_{t-1}]$ and $\sigma_t^2 = g_\varphi^*(\xi_{1t}, \xi_{2t}) = \mathrm{Var}[h(y_t) \mid \mathcal{F}_{t-1}]$. We can assume

$$\eta_t = g_\varphi(\xi_{1t}, \xi_{2t}) = \nu + \sum_{j=1}^{p} \phi_j h(y_{t-j}) + \sum_{j=1}^{q} \delta_j (h(y_{t-j}) - \eta_{t-j}), \tag{15}$$

$$\sigma_t^2 = g_\varphi^*(\xi_{1t}, \xi_{2t}) = \alpha_0 + \sum_{j=1}^{k} \alpha_j (h(y_{t-j}) - \eta_{t-j})^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2.$$

By adding $h(y_t) - g_\varphi(\xi_{1t}, \xi_{2t})$ to both sides of (15), we have

$$h(y_t) = \nu + \sum_{j=1}^{p} \phi_j h(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j},$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{k} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2,$$

with $\varepsilon_t = h(y_t) - g_\varphi(\xi_{1t}, \xi_{2t})$ and $\sigma_t^2 = \mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1})$. Hence $h(y_t)$ follows a standard ARMA–GARCH model. Combining it with (14), we have the M-GARMA–GARCH model.

3. Probabilistic properties of M-GARMA Model

It is of interest to study whether the M-GARMA models admit stationary distributions. The ergodicity of generalized ARMA processes has been discussed in Davis et al. (2003), Neumann (2011), Woodard et al. (2011) and Douc et al. (2013) among others. The M-GARMA models bear ARMA representations (6) with MDS as innovations. Let $\phi(z) = 1 - \sum_{i=1}^{p} \phi_p z^i$. Because the distribution of ε_t depends on $\mathcal{F}_{t-1}$, the standard condition $\phi(z) \neq 0$ for all $|z| \leq 1$ is not sufficient for the existence of a stationary distribution of the process $\{h(y_t)\}$. To answer this question, we use the theory of Markov chains on a general state space. For all the terminology related to Markov chains, we refer the reader to Meyn and Tweedie (2009).

Let $\boldsymbol{X} = \{X_n\}_{n\geq 0}$ be a Markov chain on the state space $\mathfrak{X}$, equipped with some σ-field $\mathcal{B}(\mathfrak{X})$. Let $\{P(x, A),\ x \in \mathfrak{X}, A \in \mathcal{B}(\mathfrak{X})\}$ be the transition probability kernel. The main tool we use is presented as Lemma 1. We omit the proof, because it can be obtained by applying the results in Chapter 15 of Meyn and Tweedie (2009). The major condition is the so called *geometric drift condition*: there exists an extended-valued function $\mathcal{V} : \mathfrak{X} \to [1, \infty]$, a measurable set C, and constants $b < \infty$, $\beta > 0$ such that

$$\Delta\mathcal{V}(x) := \int P(x, dy)\mathcal{V}(y) - \mathcal{V}(x) \leq -\beta\mathcal{V}(x) + bI_C(x),\quad x \in \mathfrak{X}. \tag{D}$$

We call $\boldsymbol{X}$ *geometrically ergodic*, if $\boldsymbol{X}$ is positive Harris recurrent, and there exists a constant $r > 1$ such that

$$\sum_{n=1}^{\infty} r^n \|P^n(x, \cdot) - \pi\| < \infty, \quad \text{for all } x \in \mathfrak{X},$$

where π is the unique invariant probability measure, and $\|\cdot\|$ denotes the total variation norm.

Lemma 1. *Suppose $\boldsymbol{X}$ is ψ-irreducible and aperiodic. If for some m, the skeleton $\boldsymbol{X}^m$ satisfies the drift condition (D) for a petite set C and a function $\mathcal{V}$ which is everywhere finite. Then $\boldsymbol{X}$ is geometrically ergodic, and $\int_{\mathfrak{X}} V(x)\pi(dx) < \infty$, where π is the unique invariant probability measure.*

Consider the model defined by (4) and (5). Recall that $g_\varphi(\mu) = \int h(y)f(y \mid \mu, \varphi)dy$ and define $V_\varphi(\mu) = \int (h(y) - g_\varphi(\mu))^2 f(y \mid \mu, \varphi)dy$. Since φ is fixed, we sometimes omit the subscript φ in g_φ and V_φ. Throughout this section we assume $g(\mu)$ and $V(\mu)$ are continuous. Our conditions depend on the growth rate of the variance function $V(\mu)$ relative to the mean function $g(\mu)$

$$\lambda := \limsup_{|g(\mu)|\to\infty} \frac{V(\mu)}{g(\mu)^2}. \tag{16}$$

To provide a definition of the stationary M-GARMA process, we need to use the Markov chain representation. Without loss of generality, assume $q = p - 1$, and at least one of ϕ_p and δ_{p-1} is nonzero. We also assume $\nu = 0$ for simplicity. As can be seen from the proof, including a constant term ν in (5) does not affect the ergodicity condition. Define the square matrices

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$
$$\Phi_1 = \begin{pmatrix} \phi_1+\delta_1 & \phi_2+\delta_2 & \cdots & \phi_{p-1}+\delta_{p-1} & \phi_p \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{17}$$

and set $\delta = (1, \delta_1, \ldots, \delta_{p-1})'$. We first define a Markov Chain on $\mathbb{R}^p$. Given X_{t-1}, we generate y_t as

$$y_t \sim f(y_t \mid \mu_t, \varphi), \quad \text{where } g(\mu_t) = \delta'\Phi X_{t-1}. \tag{18}$$

We then set $\varepsilon_t = h(y_t) - g(\mu_t)$, and define

$$X_t = \Phi X_{t-1} + (1, 0, \ldots, 0)'\varepsilon_t. \tag{19}$$

Clearly $\boldsymbol{X} = \{X_t\}_{t\geq 0}$ is a time-homogeneous Markov chain. It can be shown that $h(y_t) = \delta' X_t$, and the $\{y_t\}$ process defined in (18) satisfies the dynamic (4) and (5).

Theorem 1. *Assume $\boldsymbol{X}$ is ψ-irreducible and aperiodic, and every compact set is a petite set. If either of the following conditions hold, $\boldsymbol{X}$ is geometrically ergodic, and under the invariant probability measure π, $E_\pi[h(y_t)]^2 < \infty$.*

(i) *The quantity $\lambda = 0$, and $\phi(z) \neq 0$ for all $|z| \leq 1$.*
(ii) *When $0 < \lambda < \infty$, define Ψ_k recursively as $\Psi_0 = I$, and $\Psi_k = \Phi'\Psi_{k-1}\Phi + \lambda\Phi_1'\Psi_{k-1}\Phi_1$ for $k \geq 1$. For some h, the operator norm of Ψ_h is strictly less than one.*

Remark 3.1. Suppose conditional on X_{t-1}, ε_t has a density $p(\cdot \mid \mu_t, \varphi)$ (with respect to the Lebesgue measure), and for every μ, φ, the density $p(\cdot \mid \mu, \varphi)$ is positive everywhere, then similarly as Chan and Tong (1985), it can be shown that $\boldsymbol{X}$ is ψ-irreducible and aperiodic, where ψ can be taken as the Lebesgue measure on $\mathbb{R}^p$. This condition holds for both the log-Gamma-M-GARMA model (9) and the logit-Beta-M-GARMA model (10).

Remark 3.2. The assumption that every compact set is a petite set is technical. It is true when $\boldsymbol{X}$ is a ψ-irreducible Feller chain, and the support of ψ has non-empty interior (Meyn and Tweedie, 2009, Proposition 6.2.8). Clearly the support of the Lebesgue measure is $\mathbb{R}^p$, and thus has non-empty interior. On the other hand, if we assume the density $p(y \mid \mu, \varphi)$ is jointly continuous in y and μ, then it can be shown that $\boldsymbol{X}$ is Feller. This condition also holds for both log-Gamma-M-GARMA model (9) and logit-Beta-M-GARMA model (10).

Remark 3.3. The condition introduced through Ψ_k is stronger than $\phi(z) \neq 0$ for all $|z| \leq 1$. It is interesting that when $\lambda = 0$, the ergodicity condition reduces to $\phi(z) \neq 0$ for all $|z| \leq 1$, which extends the standard result for ARMA processes with i.i.d innovations.

There are many ways to represent ARMA models as state space models. Different representations lead to different ergodicity conditions. We take Harvey's representation (Harvey, 1993) as another example. Again without loss of generality, assume $q = p - 1$, and at least one of ϕ_p, δ_{p-1} is nonzero. Let $s_{1t} = h(y_t)$, and

$$s_{kt} = \sum_{i=k}^{p} \phi_i h(y_{t+k-i-1}) + \sum_{j=k-1}^{p-1} \delta_j \varepsilon_{t+k-j-1}$$

for $2 \leq k \leq p$. Set $S_t = (s_{1t}, \ldots, s_{pt})'$ and $\delta = (1, \delta_1, \ldots, \delta_{p-1})'$, then Harvey's representation is given by

$$S_t = \Phi' S_{t-1} + \delta\varepsilon_t, \tag{20}$$

and $h(y_t) = (1, 0, \ldots, 0)S_t$. Conversely, if given S_{t-1}, we generate y_t as

$$y_t \sim f(y_t \mid \mu_t, \varphi), \quad \text{where } g(\mu_t) = (1, 0, \ldots, 0)\Phi' S_{t-1}, \tag{21}$$

set $\varepsilon_t = h(y_t) - g(\mu_t)$, and define S_t as (20), then $\boldsymbol{S} = \{S_t\}_{t\geq 0}$ is a Markov Chain on $\mathbb{R}^p$. It holds that $h(y_t) = (1, 0, \ldots, 0)S_t$, and the $\{y_t\}$ process defined in (21) satisfy the dynamic (4) and (5).

For this representation, in order to get the ψ-irreducibility of $\boldsymbol{S}$, we introduce the *controllability* condition

$$\text{the } p \text{ vectors } \delta, \Phi'\delta, (\Phi')^2\delta, \ldots, (\Phi')^{p-1}\delta \text{ are linearly independent.} \tag{22}$$

Theorem 2. *Assume (22) holds, and $\boldsymbol{S}$ is ψ-irreducible and aperiodic, and every compact set is petite. Under either of the following, $\boldsymbol{S}$ is geometrically ergodic, and $E_\pi[h(y_t)]^2 < \infty$, where π is the invariant probability measure.*

(i) *The quantity $\lambda = 0$, and $\phi(z) \neq 0$ for all $|z| \leq 1$.*
(ii) *When $0 < \lambda < \infty$, let $\zeta = (\phi_1, 1, 0, \ldots, 0)'$, and define Υ_k recursively as $\Upsilon_0 = I$, and $\Upsilon_k = \Phi\Upsilon_{k-1}\Phi' + \lambda \cdot \delta'\Upsilon_{k-1}\delta \cdot \zeta\zeta'$ for $k \geq 1$. For some h, the operator norm of Υ_h is strictly less than one.*

Remark 3.4. Again, suppose conditional on S_{t-1}, ε_t has a density $p(\cdot \mid \mu_t, \varphi)$, which is positive everywhere, then under the controllability condition (22), it can be shown that $\boldsymbol{X}$ is ψ-irreducible and aperiodic, where ψ can be taken as the Lebesgue measure on $\mathbb{R}^p$. If we assume the density $p(y \mid \mu, \varphi)$ is jointly continuous in y and μ, then it can be shown that $\boldsymbol{X}$ is Feller. These conditions hold for both log-Gamma-M-GARMA model (9) and logit-Beta-M-GARMA model (10).

Remark 3.5. Let $\delta(z) = 1 + \sum_{j=1}^{p-1} \delta_j z^j$. It seems that if $\phi(z)$ and $\delta(z)$ have no common zeros, then the controllability condition (22) holds. But we currently do not have a proof for it.

Remark 3.6. Both Theorems 1 and 2 guarantee that the marginal variances of $h(y_t)$ and ε_t are finite. Since $\{\varepsilon_t\}$ is a MDS, and $\phi(z) \neq 0$ for all $|z| \leq 1$, the autocovariances of $\{h(y_t)\}$ can be calculated using the ARMA representation (6).

Remark 3.7. Positive Harris recurrence implies that under π, the invariant σ-field of the stationary process $\boldsymbol{X}$ is trivial, see Theorem 17.1.5 of Meyn and Tweedie (2009). So $\boldsymbol{X}$ under π is ergodic. The same is true for $\boldsymbol{S}$.

Remark 3.8. Under the conditions of either of Theorems 1 and 2, if the y-link function $h(\cdot)$ is one-to-one, then under π, the process $\{y_t\}$ is stationary and ergodic in the sense that its invariant σ-field is trivial.

Remark 3.9. For the special case with $\lambda > 0$, and $p = q = 1$, it is possible to focus on the process $\{g(\mu_t)\}$, and provide an alternative ergodicity condition: if $\phi_1^2 + (\phi_1 + \delta_1)^2 < 1$, then the process $\{g(\mu_t)\}$ is geometrically ergodic, and $E_\pi[g(\mu_t)]^2 < \infty$. The proof is given in the Appendix.

The proofs of Theorems 1 and 2 are presented in the Appendix.
In the following we consider log-Gamma-M-GARMA model and logit-Beta-M-GARMA model in detail. By Lemma 2, we have the following corollary.

Corollary 1. *For the model (9), when $d \geq 0$, there exists an initial distribution on*

$$\{\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{1-q}, y_0, y_{-1}, \ldots, y_{1-p}\}$$

such that the process $\{y_t\}_{t\geq 1}$ defined by (9) is stationary and ergodic, and $E[h(y_t)]^2 < \infty$.

The corollary can be seen by noting that when $d > 0$, we have $\lambda = 1$ and Theorem 1(ii) and Theorem 2(ii) apply. When $d = 0$ then $\lambda = 0$ and Theorem 1(i) and Theorem 2(i) apply. Similar conditions may be derived when $d < 0$.

Corollary 2. *For the model (10), there exists an initial distribution on*

$$\{\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{1-q}, y_0, y_{-1}, \ldots, y_{1-p}\}$$

such that the process $\{y_t\}_{t\geq 1}$ defined by (10) is stationary and ergodic, and $E[h(y_t)]^2 < \infty$.

This can be seen that, for logit-Beta-M-GARMA model (10), similarly as Lemma 2, we have $\lambda = 1$ and Theorem 1(ii) and Theorem 2(ii) apply.

4. Estimation, model selection and prediction

This section considers general estimation procedure of the M-GARMA models, along with general model selection guidance and prediction procedures.

4.1. Estimation

We propose two approaches for parameter estimation. The first is based on Gaussian pseudo-likelihood with additional conditional likelihood estimation, the second is based on the likelihood. It is often convenient and possible to use an approximation of $g_\varphi(\cdot)$ for evaluating the likelihood.

Based on (6), one can use Gaussian pseudo-likelihood to estimate the ARMA parameters quickly, following Yao and Brockwell (2006). This estimate will be referred to as GMLE. Recall that $\phi(z) = 1 - \sum_{i=1}^p \phi_i z^i$, and let $\delta(z) = 1 + \sum_{j=1}^q \delta_j z^j$. Assume for simplicity that $\nu = 0$. Let $\boldsymbol{\theta} = (\phi_1, \ldots, \phi_p, \delta_1, \ldots, \delta_q)'$. We assume $\boldsymbol{\theta} \in \mathcal{B}$, which is a closed set contained in the set

$$\{\boldsymbol{\theta} \in \mathbb{R}^p \times \mathbb{R}^q : \phi(z)\delta(z) \neq 0 \text{ for all } |z| \leq 1, \ \phi(\cdot) \text{ and } \delta(\cdot) \text{ have no common zeros}\}.$$

Given that the error sequence is a stationary and ergodic MDS, the following theorem ensures its asymptotic consistency.

Theorem 3 (*Hannan, 1973*)**.** *Consider the M-GARMA model (4) and (5). Assume $\{\varepsilon_t\}$ is a stationary and ergodic time series, and $E\varepsilon_t^2 < \infty$. If the true value $\boldsymbol{\theta}_0$ is an interior point of $\mathcal{B}$, then as $T \to \infty$, the GMLE $\tilde{\boldsymbol{\theta}}$ based on $h(y_1), \ldots, h(y_T)$ converges to the true $\boldsymbol{\theta}_0$ with probability one.*

Hannan (1973) also obtained the central limit theorem of the GMLE, under the additional condition that $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$, which is a constant over time. For the general M-GARMA(p, q) models, the innovations ε_t are conditional heteroscedastic, and the asymptotic distribution of the GMLE is currently unavailable. Here we consider the M-GARMA$(p, 0)$ model as a special case. Let $\boldsymbol{\phi} = (\nu, \phi_1, \ldots, \phi_p)'$. For this model, maximizing the conditional Gaussian likelihood is equivalent as minimizing the sum of squares

$$\bar{\boldsymbol{\phi}} = \arg\min_{\nu,\phi_1,\ldots,\phi_p} \sum_{t=p+1}^T [h(y_t) - \nu - \phi_1 h(y_{t-1}) - \cdots - \phi_p h(y_{t-p})]^2.$$

It turns out even if the innovations ε_t are not Gaussian, and are conditional heteroscedastic, the asymptotic distribution of $\bar{\phi}$ can be shown to be normal. Let $X_t = (1, h(y_t), h(y_{t-1}), \ldots, h(y_{t-p+1}))'$, then

$$\bar{\boldsymbol{\phi}} = \left(\sum_{t=p+1}^T X_{t-1}X_{t-1}'\right)^{-1} \sum_{t=p+1}^T X_{t-1}\varepsilon_t.$$

We use $V(\mu_t) = V_\varphi(\mu_t) := \mathrm{Var}(\varepsilon_t|\mathcal{F}_{t-1})$ to denote the variance function.

Theorem 4. *Consider the M-GARMA model (4) and (5). Assume $\{h(y_t)\}$ is a stationary and ergodic time series, and $E[h(y_t)]^4 < \infty$, $E[V(\mu_t)]^2 < \infty$, then as $T \to \infty$,*

$$\sqrt{T}(\bar{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \Rightarrow N(\mathbf{0}, \boldsymbol{V}^{-1}\boldsymbol{W}\boldsymbol{V}^{-1}),$$

where $\boldsymbol{V} := E(X_tX_t')$ and $\boldsymbol{W} := E[V(\mu_t)X_{t-1}X_{t-1}']$ are all assumed to be positive definite.

The proof of Theorem 4 is given in the Appendix. Unlike the ARMA models with i.i.d. innovations, the asymptotic distribution here depends on moments of the process with orders higher than two. Consider the Gamma-M-GARMA(p, 0) model (9), for which

$$g(\mu_t) = \psi(c\mu_t^d) - \log(c\mu_t^{d-1}) = \nu + \sum_{i=1}^{p}\phi_i h(y_{t-i}),$$

$$V(\mu_t) = \psi_1(c\mu_t^d).$$

By Lemma 2, there exists $\kappa > 0$ and $B_\kappa > 0$ such that $V(\mu_t) \le (1+\kappa)[g(\mu_t)]^2 + B_\kappa$. Therefore, the finiteness of $E[V(\mu_t)]^2$ is a consequence of $E[h(y_t)]^4 < \infty$.

Proposition 1. *Consider the model (9) with $c > 0$, $d > 0$ and $q = 0$. Let ξ be the positive root of the polynomial z^3+3z^2-1. Define $\Xi_0 = I$, and $\Xi_k = \Phi'\Xi_{k-1}\Phi + (3+2\xi)\Phi_1'\Xi_{k-1}\Phi_1$. If the operator norm of Ξ_h is strictly less than one for some h, then at the stationary distribution, $E[h(y_t)]^4 < \infty$.*

The proof is presented in the Appendix. Similarly we can show that the same result holds for the Beta-M-GARMA(p, 0) model (10).

The estimated innovation variance using GMLE seems not meaningful for the M-GARMA model. However, with the estimated ARMA parameters and residuals, one can estimate μ_t by solving

$$g_\varphi(\mu_t) = h(y_t) - \hat{\varepsilon}_t,$$

and construct an approximated likelihood for other parameters in the conditional distribution

$$L(\psi) = \prod_{t=1}^{T} f(y_t \mid \hat{\mu}_t, \varphi),$$

to obtain an estimator for φ.

In the following we consider full maximum likelihood estimation. Let the parameter vector $\boldsymbol{\theta}' = (\nu, \phi_1, \ldots, \phi_p, \delta_1, \ldots, \delta_q)$ and $\boldsymbol{\beta}' = (\boldsymbol{\theta}', \varphi, \varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{1-q})$. The log-likelihood function is

$$L_T(\boldsymbol{\beta}) = \sum_{t=1}^{T} \ell_t(\boldsymbol{\beta}) = \sum_{t=1}^{T} \log f(y_t \mid \mu_t, \varphi), \tag{23}$$

where μ_t satisfying (5) for $t = 1, \ldots, T$.

Suppose the available data set is $\{y_{1-p}, \ldots, y_0, y_1, \ldots, y_T\}$. Given a set of parameters $\boldsymbol{\theta}$, including the initial values of ε_t for $t = 0, -1, \ldots, 1-q$, $\eta_t = g_\varphi(\mu_t)$ can be recursively obtained for $t = 1, \ldots, T$. Let $\mu_t = g_\varphi^{-1}(\eta_t)$ for $t = 1, \ldots, T$, where $g_\varphi^{-1}(\eta_t)$ is a generalized inverse function of $g_\varphi(\cdot)$ discussed in Remark 2.5. By plugging μ_t into (23), we can evaluate the log-likelihood function $L_t(\theta)$. The maximum likelihood estimate is then obtained by maximizing (23), using nonlinear optimization procedures.

In practice, one can set the initial values $\varepsilon_t = h(y_t) - g_\varphi(\mu_t)$ to zero, for $t = 0, -1, \ldots, 1-q$ to reduce the complexity, a common practice in time series estimation. The resulting estimate $\hat{\boldsymbol{\theta}}$ then becomes conditional maximum likelihood estimate. One can also use Gaussian ML estimates for the ARMA parameters and its corresponding ML estimate for φ as initial values to obtain the full likelihood estimate. The theory of Hall and Heyde (1980) can be applied to study the asymptotic distribution of the MLE. However, for concrete M-GARMA models, sufficient conditions that ensure the asymptotic normality are under investigation. Regardless of the technical issues, we provide a reasonable formula for the asymptotic covariance matrix. Let

$$u_t(\boldsymbol{\theta}) = \frac{\partial \log f(y_t|\mu_t, \varphi)}{\partial \boldsymbol{\theta}}.$$

Let $\boldsymbol{\theta}_0$ be the true parameter. Under regularity conditions, $\{u_t(\boldsymbol{\theta}_0)\}$ is a MDS with respect to $\{\mathcal{F}_t\}$. Define

$$I_T(\boldsymbol{\theta}_0) = \sum_{t=1}^{T} E_{\boldsymbol{\theta}_0}[u_t(\boldsymbol{\theta}_0)(u(\boldsymbol{\theta}_0))'|\mathcal{F}_{t-1}].$$

Under regularity conditions, it holds that

$$[I_T(\boldsymbol{\theta}_0)]^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \Rightarrow N(0, I).$$

For many M-GARMA models, for a given $\boldsymbol{\theta}_0$, the conditional information $E_{\boldsymbol{\theta}_0}[u_t(\boldsymbol{\theta}_0)(u(\boldsymbol{\theta}_0))'|\mathcal{F}_{t-1}]$ can be calculated explicitly. By substituting the estimate $\hat{\boldsymbol{\theta}}$, we get an estimate $I_T(\hat{\boldsymbol{\theta}})$. In practice, the standard errors of the estimates can also be obtained by evaluating the Hessian matrix of the log-likelihood function (23) at the MLE. Our empirical experiences have shown that the estimator works very well.

Sometimes the exact link function $g_\varphi(\mu_t)$ may be too complicated to invert. In such a situation we can use an approximation $\tilde{g}_\varphi(\mu_t)$ to replace it for evaluating the likelihood. Specifically, given $\mathcal{F}_{t-1}$, we replace $g_\varphi(\mu_t)$ by $\tilde{g}_\varphi(\mu_t)$ in (5), and then invert $\tilde{g}_\varphi(\mu_t)$ to get μ_t, which allows us to calculate $\ell_t(\boldsymbol{\beta}) = \log f(y_t \mid \mu_t, \varphi)$. See Remark 2.6 for some examples of the approximation.

4.2. Model building and selection procedures

The ARMA formation in (6) provides a convenient approach for order determination. One can simply use the standard methods for building Gaussian time series models, based on information such as autocorrelation function (ACF), partial autocorrelation function (PACF) and extended autocorrelation function (EACF) of Tsay and Tiao (1984), though caution needs to be excised since it has been shown that for ARMA model with varying conditional variances, estimators of ACF, PACF and EACF may have more complex standard errors, hence model determination based on such measures may be slightly biased (Chen et al., 2013). With several tentative models, BIC and out-of-sample prediction performance can be obtained to select the more appropriate models.

4.3. Prediction

A benefit of having a M-GARMA model is the easiness of performing out-of-sample forecast. For a sequence $\{y_t\}$ following the M-GARMA model defined by (4) and (6), if the process is causal and invertible, then the linear least square predictor of $h(y_{t+h})$ based on the information set $\mathcal{F}_t$ is

$$E(h(y_{t+h}) \mid \mathcal{F}_t) = \eta_{t+h|t} = \nu + \sum_{i=1}^{p}\phi_i\eta_{t+h-i|t} + \sum_{j=1}^{q}\delta_j\varepsilon_{t+h-j|t}$$

where $\eta_{t+h-i|t}$ is the $h-i$ step ahead prediction for $i < h$ and $\eta_{t+h-i|t} = h(y_{t+h-i})$ for $i \ge h$; and $\varepsilon_{t+h-j|t}$ is the linear prediction of ε_{t+h-j} based on $\{h(y_1), \ldots, h(y_t)\}$ for $j \ge h$, and $\varepsilon_{t+h-j|t} = 0$ for $j < h$. This is easily seen as ε_t is a MDS.

It is more complicated to predict y_{t+h} since $E(y_{t+h} \mid \mathcal{F}_t)$ does not equal to the inversion $g_\varphi^{-1}(\eta_{t+h|t})$. A feasible way is through forward simulation. Specifically, at time t, let $\tilde{\eta}_{t-j} = h(y_{t-j}) - \varepsilon_{t-j|t}$ for $j \ge 0$, we first obtain the predictor $\tilde{\eta}_{t+1}$ using (8), where it is understood that every η is replaced by $\tilde{\eta}$. Then with

Table 2
Simulation results of the Gamma-M-GARMA model.

Parameter	True	GMLE	MLE	AMLE0	AMLE1
ν	−0.01	−0.1327 (0.0510)	−0.0191 (0.0408)	0.2345 (0.0291)	−0.0126 (0.0350)
ϕ_1	0.90	0.7944 (0.0637)	0.8873 (0.0358)	0.9427 (0.0306)	0.8962 (0.0342)
δ_1	−0.60	−0.5705 (0.0833)	−0.5902 (0.0378)	−0.5897 (0.0385)	−0.5899 (0.0380)
c	1.00	0.9153 (0.0648)	1.0023 (0.0625)	0.9995 (0.0589)	0.9968 (0.0619)
d	−0.50	−0.6399 (0.1094)	−0.5141 (0.0801)	−0.4865 (0.0737)	−0.5142 (0.0808)
ν	−0.01	−0.0379 (0.0272)	−0.0127 (0.0276)	0.2038 (0.0290)	0.0110 (0.0246)
ϕ_1	0.90	0.8615 (0.0438)	0.8884 (0.0355)	0.8674 (0.0359)	0.8815 (0.0353)
δ_1	−0.60	−0.5816 (0.0647)	−0.5888 (0.0505)	−0.5858 (0.0530)	−0.5881 (0.0511)
c	1.00	1.0050 (0.0687)	1.0127 (0.0663)	1.0099 (0.0661)	1.0109 (0.0666)
d	0.00	−0.0394 (0.0895)	−0.0154 (0.0814)	−0.0180 (0.0818)	−0.0210 (0.0815)
ν	−0.01	−0.0095 (0.0239)	−0.0086 (0.0220)	0.1831 (0.0296)	−0.0148 (0.0205)
ϕ_1	0.90	0.8794 (0.0414)	0.8887 (0.0364)	0.8230 (0.0374)	0.8748 (0.0356)
δ_1	−0.60	−0.5845 (0.0756)	−0.5928 (0.0553)	−0.5931 (0.0549)	−0.5931 (0.0553)
c	1.00	1.0173 (0.0804)	1.0153 (0.0806)	1.0228 (0.0780)	1.0154 (0.0803)
d	0.50	0.4853 (0.1084)	0.4960 (0.1118)	0.4745 (0.1000)	0.4959 (0.1106)

$\tilde{\mu}_{t+1} = g_{\varphi}^{-1}(\tilde{\eta}_{t+1})$, $\tilde{y}_{t+1}$ can be simulated from the conditional density $f(\tilde{y}_{t+1} \mid \tilde{\mu}_{t+1}, \varphi)$. With the simulated $\tilde{y}_{t+1}$, one can repeat the procedure to obtain a simulated $\tilde{y}_{t+2}$, and so on. The h-step prediction can then be obtained using the average of many simulated $\tilde{y}_{t+h}$'s.

5. Simulation examples

In this section, we investigate finite sample performances of proposed estimators under the Gamma-M-GARMA and Beta-M-GARMA models in Section 2.3, with log and logit transformations respectively. Since the exact link function and approximate link functions can be obtained, we compare their performances for estimating the parameters. In particular, linear approximation leads to $g_{\varphi}(\mu_t) = h(\mu_t)$, the link function suggested by Benjamin et al. (2003), see Remark 2.6. The estimator obtained using this approximation is referred to as AMLE0. With the second order approximation (see Remark 2.6), the estimator is referred to as AMLE1. The MLE obtained using the exact link function $g_{\varphi}(\mu_t)$ is still referred to as MLE. We also demonstrate the finite sample performance of the pseudo Gaussian likelihood estimator GMLE. All estimates are obtained with a constraint optimization technique that uses the MaxSQPF algorithm, implementing a sequential quadratic programming technique, see Nocedal and Wright (1999).

5.1. Log-Gamma-M-GARMA model

We generate a time series of length $T = 500$ from the log-Gamma-M-GARMA(1, 1) model:

$$y_t \mid \mu_t \sim \mathrm{Gam}(c\mu_t^d, c\mu_t^{d-1}),$$
$$\log y_t = \nu + \phi_1 \log y_{t-1} + \varepsilon_t + \delta_1\varepsilon_{t-1}, \quad \varepsilon_t = \log y_t - g_{c,d}(\mu_t).$$

The estimator AMLE0 is obtained by maximizing the likelihood with the approximated link function $g_{c,d}(\mu_t) \approx \log(\mu_t)$, and AMLE1 is obtained using the second order approximation $g_{c,d}(\mu_t) \approx \log\mu_t - (2c\mu_t^d)^{-1}$. For three sets of the parameters $\{\nu, \phi_1, \delta_1, c, d\}$, the simulation is repeated for 500 times. The means and standard errors of the estimates are presented in Table 2.

It can be seen that the simple GMLE performs reasonably well and can serve as good starting values. AMLE0 performs reasonably well in this case, except for $d = 0.5$ case where the AR coefficients are off significantly. Fig. 3 has shown that the linear approximation is less accurate as d increases. The constant term given by AMLE0 is always biased due to the linear approximation. The second order approximation works well. When $d = 0$, the log-function is canonical and AMLE0 should be the same as MLE except for the constant term. In Table 2, we see the estimates given by these two methods are very close when $d = 0$. The difference is due to the fact that d also needs to be estimated.

5.2. Logit-Beta-M-GARMA model

We simulate a time series of length $T = 500$ from the Logit-Beta-M-GARMA(1, 1) model:

$$y_t \mid \mu_t \sim \mathrm{Beta}(\tau\mu_t, \tau(1-\mu_t)),$$
$$\mathrm{logit}(y_t) = \nu + \phi_1 \mathrm{logit}(y_{t-1}) + \varepsilon_t + \delta_1\varepsilon_{t-1},$$
$$\varepsilon_t = \mathrm{logit}(y_t) - g_\tau(\mu_t).$$

The approximated link functions used for AMLE0 and AMLE1 are $g_\tau(\mu_t) \approx \mathrm{logit}(\mu_t)$ and $g_\tau(\mu_t) \approx \mathrm{logit}(\mu_t) - [2(1+\tau)\mu_t(1-\mu_t)]^{-1}$, respectively. For each of the four sets of parameters, the means and standard errors of the estimates based on 500 repetitions are reported in Table 3.

It can be seen that the simple GMLE performs very well, though our experiments (not shown here) show that its performance may deteriorate when the sample size is smaller. The AMLE0 performs poorly when τ is small, but its performance improves as τ increases. This is due to the fact that the approximation is closer to the true link function $g_\tau(\cdot)$ as τ increases. The estimator AMLE1, using a second order approximation, performs better than AMLE0.

Next we study the model selection procedure using BIC for both MLE and GMLE. We set the true orders as $p = q = 1$, and fix the parameters $\nu = -.1, \phi_1 = 0.8$ and $\delta_1 = -.5$. For each combination of $\tau = 1, 5$ and $T = 200, 500$, the procedure is repeated 500 times. Table 4 gives the number of times that each (p, q) is selected, with $1 \le p, q \le 3$. It is shown that the MLE, which is computationally more intensive, performs slightly better than GMLE in terms of model identification. For moderate sample size, BIC performs quite well, though it has a tendency to select large models.

6. Applications

6.1. High-frequency realized volatility

Realized volatility has been extensively modeled and studied in financial econometrics, see for example, Barndorff-Nielsen and Shephard (2002), Hansen and Lunde (2005) and Takahashi et al. (2009). Daily realized volatility is often calculated based on high frequency intra-day asset returns to measure integrated variability (Andersen and Bollerslev, 1998; Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2001). Here we study

Table 3
Simulation results of the Beta-M-GARMA model.

Parameter	True	GMLE	MLE	AMLE0	AMLE1
ν	−0.10	−0.1159 (0.1030)	−0.1097 (0.0907)	−0.1097 (0.0907)	−0.0653 (0.0542)
ϕ_1	0.80	0.0653 (0.0542)	0.7807 (0.0672)	0.5570 (0.1120)	0.6470 (0.0728)
δ_1	−0.50	−0.4773 (0.1043)	−0.4773 (0.1043)	−0.4607 (0.1122)	−0.4607 (0.1122)
τ	1.00	1.0047 (0.0525)	1.0071 (0.0511)	0.9789 (0.1350)	1.0057 (0.0512)
ν	−0.10	−0.1100 (0.0374)	−0.1091 (0.0372)	−0.0897 (0.0305)	−0.1040 (0.0354)
ϕ_1	0.80	0.7796 (0.0599)	0.7816 (0.0597)	0.7157 (0.0617)	0.7633 (0.0600)
δ_1	−0.50	−0.4847 (0.0812)	−0.4854 (0.0807)	−0.4839 (0.0804)	−0.4850 (0.0806)
τ	5.00	5.0530 (0.2957)	5.0543 (0.2956)	5.0534 (0.2959)	5.0544 (0.2956)
ν	−0.10	−0.1071 (0.0451)	−0.1053 (0.0320)	−0.0959 (0.0291)	−0.0959 (0.0291)
ϕ_1	0.80	0.7817 (0.0995)	0.7861 (0.0588)	0.7524 (0.0596)	0.7807 (0.0588)
δ_1	−0.50	−0.4872 (0.1036)	−0.4906 (0.0783)	−0.4897 (0.0783)	−0.4904 (0.0782)
τ	10.0	10.086 (0.6275)	10.105 (0.6236)	10.094 (0.6439)	10.107 (0.6236)
ν	−0.10	−0.1107 (0.0293)	−0.1107 (0.0293)	−0.1106 (0.0284)	−0.1106 (0.0284)
ϕ_1	0.90	0.8889 (0.0291)	0.8893 (0.0288)	0.8768 (0.0288)	0.8888 (0.0288)
δ_1	−0.50	−0.4913 (0.0509)	−0.4916 (0.0503)	−0.4916 (0.0503)	−0.4916 (0.0503)
τ	50.0	50.557 (3.2916)	50.559 (3.2906)	50.565 (3.2911)	50.563 (3.2914)

Table 4
Model selection using BIC. Logit-Beta-M-GARMA(1, 1) model with $\nu = -0.10$, $\phi = 0.8$ and $\delta = -0.5$.

		GMLE				MLE			
		$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 0$	$q = 1$	$q = 2$	$q = 3$
$\tau = 1$	$p = 0$	0	2	5	3	0	0	2	0
$T = 200$	$p = 1$	153	199	2	4	122	227	3	3
	$p = 2$	77	0	1	6	72	1	9	5
	$p = 3$	10	2	4	32	2	1	14	38
$\tau = 5$	$p = 0$	0	1	5	3	0	25	8	4
$T = 200$	$p = 1$	140	204	1	1	0	289	1	4
	$p = 2$	65	1	14	7	81	2	12	4
	$p = 3$	4	0	5	47	5	0	17	48
$\tau = 1$	$p = 0$	0	0	0	0	0	0	0	0
$T = 500$	$p = 1$	21	372	5	0	0	387	4	2
	$p = 2$	68	2	1	2	56	2	3	4
	$p = 3$	14	1	2	0	5	1	4	21
$\tau = 5$	$p = 0$	0	0	0	0	0	0	0	0
$T = 500$	$p = 1$	8	389	1	1	0	399	3	0
	$p = 2$	55	1	10	2	58	1	2	1
	$p = 3$	6	0	5	21	5	0	2	29

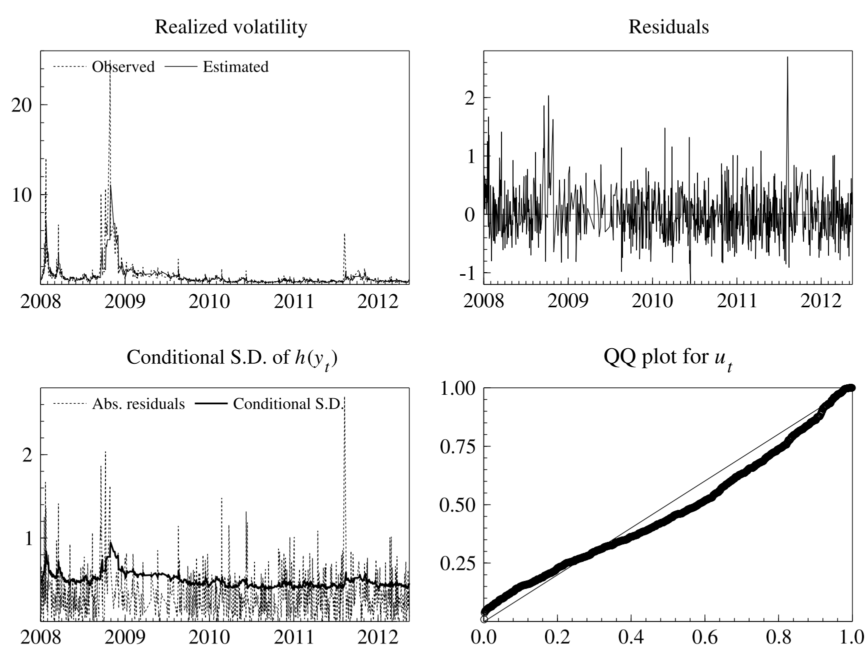

Fig. 4. Fitted results for the HSI realized volatility using the Log-Gamma-M-GARMA model.

the daily realized volatility MTRV, the *median truncated realized variance* of Hang Seng Index (HSI), taken from the "Oxford-Man Institute's realized library" (version 0.2, available at the website: http://realized.oxford-man.ox.ac.uk). The data is from January 2, 2008 to May 15, 2012 with 823 observations, see Fig. 4. As volatility is positive and often modeled by Gamma distribution, we use the

Table 5
Estimation results of the Log-Gamma-M-GARMA model for the HSI realized volatility.

Parameter	Two-parameter parametrization				Three-parameter parametrization			
	Log-Gamma-M-GARMA			MEM	Log-Gamma-M-GARMA			MEM
	GMLE	MLE	AMLE0	MLE	GMLE	MLE	AMLE0	MLE
ν	−0.0195	−0.0104	0.0329	0.0268	−0.0195	−0.0250	0.0232	0.0307
	(0.0088)	(0.0097)	(0.0095)	(0.0085)	(0.0088)	(0.0117)	(0.0097)	(0.0096)
ϕ_1	1.0414	1.0944	1.0911	0.9689	1.0414	1.1274	1.1381	0.9602
	(0.0667)	(0.0741)	(0.0738)	(0.0814)	(0.0667)	(0.0747)	(0.0756)	(0.0901)
ϕ_2	−0.0694	−0.1093	−0.1061	0.0000	−0.0694	−0.1627	−0.1589	0.0000
	(0.0612)	(0.0698)	(0.0695)	(0.0764)	(0.0612)	(0.0674)	(0.0704)	(0.0833)
δ_1	−0.6383	−0.6084	−0.6042	−0.5234	−0.6383	−0.6618	−0.6508	−0.5113
	(0.0549)	(0.0623)	(0.0623)	(0.0738)	(0.0549)	(0.0633)	(0.0633)	(0.0879)
c	4.6745	4.7277	4.7228	4.6824	4.0116	3.9394	3.9404	3.8863
	(0.2110)	(0.2253)	(0.2251)	(0.2242)	(0.2208)	(0.2586)	(0.2530)	(0.2441)
d					−0.3487	−0.3878	−0.3798	−0.4049
					(0.0616)	(0.0773)	(0.0745)	(0.0711)
Loglik	1.224	5.492	5.236	1.465	20.02	22.26	21.06	20.99
BIC	31.12	22.58	23.09	30.62	0.231	−4.243	−1.845	−1.716
RSS	1065	1030	1038	1097	1043	1035	1040	1091
$Q(1)$	0.0022	5.13*	5.11*	3.6420	0.0022	3.3889	6.062*	2.9087
$Q(5)$	4.7523	11.1*	11.2*	71.52**	4.7523	10.338	11.54*	72.51**
$Q(22)$	17.134	26.662	26.601	157.9**	17.134	23.902	26.023	157.8**
$Q^2(1)$	7.082**	9.26**	11.39**	0.0142	3.3934	4.604*	6.86**	0.0049
$Q^2(5)$	31.91**	28.07**	29.35**	0.1715	6.7814	8.1967	13.29*	0.0104
$Q^2(22)$	43.07**	39.65**	40.15**	0.3316	16.752	19.184	22.665	0.1476
C_1	0.884	0.057	0.032*	0.923	0.884	0.096	0.016*	0.924
C_2	0.646	0.076	0.076	0.886	0.646	0.262	0.112	0.948
C_3	0.320	0.012*	0.016*	0.848	0.320	0.168	0.056	0.800
K_1	0.790	0.082	0.048*	0.400	0.790	0.107	0.040*	0.496
K_2	0.808	0.096	0.124	0.990	0.808	0.224	0.136	0.992
K_3	0.454	0.042*	0.026*	0.350	0.454	0.280	0.124	0.890
GS	0.880	0.066	0.066	0.254	0.880	0.128	0.046*	0.286
KL_1	0.925	0.105	0.000**	0.489	0.925	0.172	0.000**	0.411
KL_2	0.255	0.027*	0.000**	0.694	0.255	0.071	0.000**	0.639
KL_3	0.037*	0.006**	0.000**	0.624	0.037*	0.011*	0.000**	0.644

Note: The values in parentheses are the standard errors.
* indicate that the test statistic is significant at 5% levels.
** indicate that the test statistic is significant at 1% levels.

Gamma-M-GARMA model with log y-link function (9). We also consider a simper model where the parameter d in (9) is fixed at $d = 0$, and c is to be estimated.

Based on standard order determination procedures for the Gaussian time series (Tsay and Tiao, 1984; Chen et al., 2013) and using BIC, the order of the M-GARMA process is selected as $p = 2$ and $q = 1$. That is, the following model is used,

$$y_t \mid \mu_t \sim \text{Gam}(c\mu_t^d, c\mu_t^{d-1}),$$
$$\log y_t = \nu + \phi_1 \log y_{t-1} + \phi_2 \log y_{t-2} + \varepsilon_t + \delta_1 \varepsilon_{t-1},$$

where $\varepsilon_t = \log y_t - g_{c,d}(\mu_t)$.

Table 5 shows the estimation results of three methods MLE, GMLE and AMLE0, see Section 5 for details about AMLE0. In the top panel we report the estimates and their standard errors. In the bottom panel we report a few statistics to compare different methods. The first two rows give the maximum log likelihood and the BIC. In the third row, RSS stands for residual sum of squares defined by RSS $= \sum_{t=1}^{T}(y_t - \hat{\mu}_t)^2$. The quantity $Q(m)$ denotes the Box–Ljung test statistic with m lags (Ljung and Box, 1978). The statistic $Q^2(m)$ is the portmanteau test statistic based on squared standardized residuals $\hat{e}_t^2$, which are defined as $\hat{e}_t^2 = \hat{\varepsilon}_t^2/\hat{\sigma}_t^2$, where $\hat{\sigma}_t^2$ is the estimated conditional variance $\widehat{\text{Var}}(\log y_t \mid \mathcal{F}_{t-1}) = \psi_1(\hat{c}\hat{\mu}_t^{\hat{d}})$. This statistic is used to test whether the conditional heteroscedasticity is modeled well, see McLeod and Li (1983). We also employ the martingale difference tests check whether $\{\varepsilon_t\}$ is a MDS. The statistics C_P and K_P stand for Cramer–von Mises test and Kolmogorov–Smirnov test respectively (Domínguez and Lobata, 2003), where P is the number of lagged values used in the tests. The statistic GS is the generalized spectral test proposed by Escanciano and Velasco (2006). The test KL_P were proposed by Kuan and Lee (2004), where we use the multivariate exponential density as the weight function, and the parameter β is taken as the reciprocal of the sample standard deviation. Since most martingale tests are not distribution free, and the p-values are often obtained through bootstrapping, we only report p-values of the martingale tests. GMLE results are very similar with those of MLE, suggesting the advantage of M-GARMA models. AMLE0 also provides very similar results, because in the range of μ_t of this data set, the linear approximation of the link function is very accurate.

Comparing the left and right panels of Table 5, we see the benefit of adding d as a parameter and estimating it. It allows us to model the conditional heteroscedasticity more adequately, as the statistics $Q^2(m)$ are significantly reduced in the right panel. We choose $m = 1, 5, 22$, corresponding to autocorrelations within a day, a week and a month of the squared standardized residuals. The models in the right panel also has significantly smaller BIC values. Furthermore, the martingale difference tests for the three-parameter case provide larger p-values.

In the right panel, the GMLE, MLE and AMLE0 all produce similar results, as the estimated d is small which makes the approximation used by AMLE0 quite accurate. However, the residuals from MLE estimation have slightly better properties, in terms of BIC and Q statistics.

We also compare the log-Gamma-M-GARMA model with the following MEM model

$$y_t|\mathcal{F}_{t-1} \sim \text{Gam}(c\mu_t^d, c\mu_t^{d-1}),$$
$$\mu_t = \nu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \delta_1(y_{t-1} - \mu_{t-1}).$$

For details on general MEM models, see for example Engle (2002), Engle and Gallo (2006) and Brownlees et al. (2012). Based on the

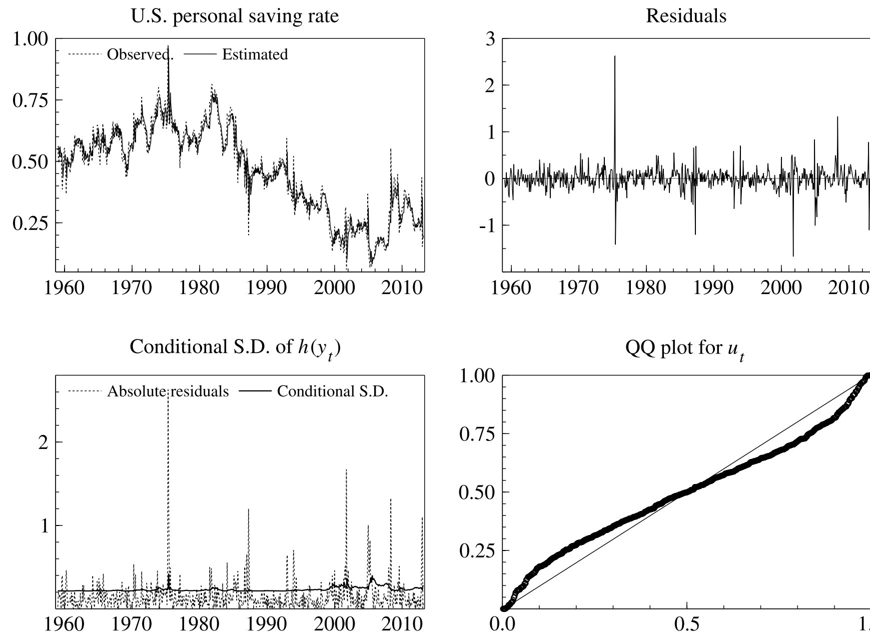

Fig. 5. Fitted results for US personal saving rate with the Logit-Beta-M-GARMA model.

likelihood, BIC, RSS and portmanteau tests, we find that the log-Gamma-M-GARMA model has a better performance.

Finally, we use a few plots in Fig. 4 to illustrate the performance of the fitted log-Gamma-M-GARMA(2, 1) model using MLE. The upper-left panel is a plot of the original time series y_t and the fitted values $\hat{\mu}_t$, showing a good fit to the data. The upper-right panel gives the residuals $\hat{\varepsilon}_t$. In the lower-left panel we show the absolute residual $|\hat{\varepsilon}_t|$, and the estimated conditional standard deviation $\hat{\sigma}_t$. We also construct a QQ plot (lower-right panel) to check the conditional distribution assumption more carefully. Let $F(\cdot \mid \mu_t, c, d)$ be the cumulative distribution function of $\mathrm{Gam}(c\mu_t^d, c\mu_t^{d-1})$, then $u_t = F(y_t \mid \mu_t, c, d)$ follows the uniform distribution on [0, 1]. Let $\hat{u}_t = F(y_t \mid \hat{\mu}_t, \hat{c}, \hat{d})$ and we show, in the lower-right panel, the QQ plot of $\hat{u}_t$ over Uniform [0, 1]. Note that such constructed QQ plot can also be viewed as residual QQ plot. Let $F_1(\cdot \mid \mu_t, c, d)$ be the conditional cumulative distribution function of ε_t given $\mathcal{F}_{t-1}$. It can be easily shown that $F_1(\varepsilon_t \mid \mu_t, c, d) = F(y_t \mid \mu_t, c, d)$. Hence a straight line in the QQ plot would indicate reasonable model (and the residual) assumption.

6.2. US personal saving rate

In this section we study the US personal saving rate, an important economic indicator. The seasonal adjusted monthly series is available at the US Bureau of Labor Statistics (http://www.bls.gov). The series, from January 1959 to March 2013 with 651 observations, is shown in the upper left panel of Fig. 5. It is reasonable to assume that saving rate would not exceed 15% hence the series is bounded in [0, 0.15]. For simplicity, we multiply the series by 20/3 and model it with a conditional Beta distribution on [0, 1]. Specifically, we employ the Logit-Beta-M-GARMA model (10) with a time trend

$$y_t \mid \mathcal{F}_{t-1} \sim \mathrm{Beta}(\tau\mu_t, \tau(1-\mu_t)),$$

$$\mathrm{logit}(y_t) = \nu + \gamma t + \sum_{j=1}^{p} \phi_j \mathrm{logit}(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j},$$

where $\varepsilon_t = \mathrm{logit}(y_t) - g_\tau(\mu_t)$. Our analysis is parallel to that in Section 6.1. First, the order is selected as $p = 4$ and $q = 1$. We then apply the methods of MLE, GMLE, AMLE0, and AMLE1 (see Section 5 for details about AMLE0 and AMLE1), and the results are reported in Table 6. Here the residual sum of squares RSS1 is defined as $\mathrm{RSS1} = \sum_{t=1}^{T}(y_t - \hat{\mu}_t)^2$, and RSS2 is defined as $\mathrm{RSS2} = \sum_{t=1}^{T}(\mathrm{logit}(y_t) - g_{\hat{\tau}}(\hat{\mu}_t))^2$.

We see the results of MLE and AMLE1 are very close, again because the second order approximation of the link function is accurate in the range of this data set. GMLE and AMLE0 also provide very similar results. From the maximum likelihood, BIC and RSS, we see that MLE and AMLE1 provide slightly better fit than GMLE and AMLE0, but the improvements are marginal. Portmanteau tests suggest that both the dependence in the series logit(y_t), and the conditional heteroscedasticity are modeled well. This finding is also confirmed by the top panel of Fig. 5, which gives the time series plot of y_t with fitted values $\hat{\mu}_t$, as well as the residual plot of $\hat{\varepsilon}_t$. The bottom left panel shows the absolute residuals, and the conditional standard deviations $\hat{\sigma}_t$, defined by $\hat{\sigma}_t^2 = \widehat{\mathrm{Var}}[\mathrm{logit}(y_t) \mid \mathcal{F}_{t-1}] = \psi_1(\hat{\tau}\hat{\mu}_t) + \psi_1(\hat{\tau}(1-\hat{\mu}_t))$. We also construct a QQ plot using a similar procedure as described in Section 6.1. The bottom panel suggests that the conditional Beta assumption is reasonable.

Acknowledgments

We thank the editor and the referees for their helpful comments and suggestions. One anonymous reviewer has suggested to include the martingale difference tests. The other anonymous reviewer has suggested the semiparametric approach discussed in Remark 2.3.

Appendix

A.1. Proofs for Section 3

Proof of Theorem 1. For simplicity, we assume $\nu = 0$. The proof for the case $\nu \neq 0$ is very similar, but slightly more complicated. We will prove the Markov chain $\boldsymbol{X}$ in (19) is geometrically ergodic. We first consider the condition (ii).

Proof of (ii). Since $X_t = \Phi X_{t-1} + (1, 0, \dots, 0)'\varepsilon_t$, we have

$$E(\|X_t\|^2 \mid X_{t-1}) = X_{t-1}'\Phi'\Phi X_{t-1} + \mathrm{Var}(\varepsilon_t \mid \mu_t).$$

Table 6
Estimation results of the Logit-Beta-M-GARMA model for the US saving rate.

Parameter	GMLE	MLE	AMLE0	AMLE1
ν	0.0354 (0.0186)	0.0317 (0.0167)	0.0323 (0.0165)	0.0317 (0.0167)
γ	−1.43E-4 (6.1E-5)	−1.27E-4 (4.6E-5)	−1.28E-4 (5.4E-5)	−1.27E-4 (5.5E-5)
ϕ_1	0.9145 (0.1142)	0.9486 (0.1175)	0.9380 (0.1168)	0.9483 (0.1175)
ϕ_2	−0.1155 (0.0862)	−0.1413 (0.0923)	−0.1386 (0.0904)	−0.1412 (0.0922)
ϕ_3	−0.0283 (0.0532)	−0.0378 (0.0504)	−0.0369 (0.0496)	−0.0378 (0.0504)
ϕ_4	0.1895 (0.0471)	0.1978 (0.0459)	0.1946 (0.0452)	0.1977 (0.0459)
δ_1	−0.3197 (0.1128)	−0.3039 (0.1159)	−0.3050 (0.1153)	−0.3039 (0.1159)
τ	87.698 (4.7385)	87.928 (4.8120)	87.930 (4.8761)	87.934 (4.8898)
Loglik	1037.44	1038.29	1037.8	1038.27
BIC	−2023.05	−2024.75	−2023.77	−2024.71
RRS1	1.4166	1.4130	1.4142	1.4130
RRS2	30455	30459	30483	30459
$Q(12)$	10.113	11.448	10.196	11.417
$Q(24)$	16.180	18.057	15.840	18.007
$Q^2(12)$	9.5981	11.488	10.020	11.439
$Q^2(24)$	11.091	13.013	11.514	12.964
C_1	0.000**	0.046*	0.020*	0.032*
C_2	0.006**	0.068	0.036*	0.064
C_3	0.136	0.390	0.300	0.424
K_1	0.000**	0.034*	0.020*	0.020*
K_2	0.048*	0.174	0.178	0.188
K_3	0.206	0.650	0.444	0.644
GS	0.058	0.274	0.222	0.270
KL_1	0.000**	0.008**	0.004**	0.008**
KL_2	0.000**	0.000**	0.000**	0.000**
KL_3	0.000**	0.001**	0.000**	0.001**

Note: The values in parentheses are the standard errors.
* denote that the test is significant at 5% level of significance.
** denote that the test is significant at 1% level of significance.

For every $\kappa > 0$, there exists a $B_\kappa > 0$ such that

$$V(\mu) \le (1+\kappa)\lambda[g(\mu)]^2 + B_\kappa, \tag{24}$$

and it follows that

$$\begin{aligned} E(\|X_t\|^2 \mid X_{t-1}) &\le X'_{t-1}\Phi'\Phi X_{t-1} + (1+\kappa)\lambda[g(\mu_t)]^2 + B_\kappa \\ &= X'_{t-1}\Phi'\Phi X_{t-1} + (1+\kappa)\lambda X'_{t-1}\Phi'_1\Phi_1 X_{t-1} + B_\kappa \\ &\le (1+\kappa)X'_{t-1}\Psi_1 X_{t-1} + B_\kappa. \end{aligned}$$

Next, by taking a double expectation

$$E(\|X_t\|^2 \mid X_{t-2}) \le (1+\kappa)E[X'_{t-1}\Psi_1 X_{t-1} \mid X_{t-2}] + B_\kappa.$$

Applying (24) again,

$$\begin{aligned} &E[X'_{t-1}\Psi_1 X_{t-1} \mid X_{t-2}] \\ &\quad = X'_{t-2}\Phi'\Psi_1\Phi X_{t-2} + w_1 \mathrm{Var}(\varepsilon_{t-1} \mid \mu_{t-1}) \\ &\quad \le X'_{t-2}\Phi'\Psi_1\Phi X_{t-2} + w_1(1+\kappa)\lambda[g(\mu_{t-1})]^2 + w_1 B_\kappa \\ &\quad = X'_{t-2}\Phi'\Psi_1\Phi X_{t-2} + (1+\kappa)\lambda X'_{t-2}\Phi'_1\Psi_1\Phi_1 X_{t-2} + w_1 B_\kappa \\ &\quad \le (1+\kappa)X'_{t-2}\Psi_2 X_{t-2} + w_1 B_\kappa. \end{aligned}$$

It follows that

$$E(\|X_t\|^2 \mid X_{t-2}) \le (1+\kappa)^2 X'_{t-2}\Psi_2 X_{t-2} + (1+\kappa)w_1 B_\kappa + B_\kappa,$$

where w_1 is the $(1, 1)$th entry of Ψ_1.

Following the same argument, we have for any positive integer h,

$$E(\|X_t\|^2 \mid X_{t-h}) \le (1+\kappa)^h X'_{t-h}\Psi_h X_{t-h} + \sum_{j=0}^{h-1}(1+\kappa)^j w_j B_\kappa$$

where w_j is the $(1, 1)$th entry of Ψ_j for $j \ge 1$, and $w_0 = 1$.

Choose h, such that the operator norm of Ψ_h is less than $1 - 2\omega$ for some $\omega > 0$. Choosing κ such that $(1+\kappa)^h(1-2\omega) < 1-\omega$, we have

$$E(\|X_t\|^2 \mid X_{t-h}) \le (1-\omega)\|X_{t-h}\|^2 + \sum_{j=0}^{h-1}(1+\kappa)^j w_j B_\kappa.$$

Since every compact set is petite, from here it is easy to verify that for the skeleton $\boldsymbol{X}^h$, the drift condition (D) is met with $\mathcal{V}(x) = \|x\|^2 + 1$ for $x \in \mathbb{R}^p$. By Lemma 1, the chain $\boldsymbol{X}$ is geometrically ergodic, and $E_\pi\|X_t\|^2 < \infty$, where π is the unique invariant probability measure. Because $h(y_t) = \delta' X_t$, it follows that $E_\pi[h(y_t)]^2 < \infty$.

Proof of (i). Since $\lambda = 0$, for every $\kappa > 0$, there exists a $B_\kappa > 0$ such that $V(\mu) \le \kappa[g(\mu)]^2 + B_\kappa$. Define $\Psi_0 = I$, and $\Psi_k = \Phi'\Psi_{k-1}\Phi + \kappa\Phi'_1\Psi_{k-1}\Phi_1$ for $k \ge 1$. Similarly as the proof of (ii), we have

$$E(\|X_t\|^2 \mid X_{t-h}) \le X'_{t-h}\Psi_h X_{t-h} + \sum_{j=0}^{h-1} w_j B_\kappa.$$

Under the condition $\phi(z) \ne 0$ for all $|z| \le 1$, the spectrum radius of Φ is strictly less than one, so we can choose h, such that the operator norm of Φ^h is less than $1 - 2\omega$ for some $\omega > 0$, and we can then choose a κ small enough such that the operator norm of $\|\Psi_h\|$ is less than $1 - \omega$. Thus,

$$E(\|X_t\|^2 \mid X_{t-h}) \le (1-\omega)\|X_{t-h}\|^2 + \sum_{j=0}^{h-1} w_j B_\kappa.$$

The proof is completed by applying Lemma 1. □

Proof of Theorem 2. The proof of Theorem 2 is very similar with that of Theorem 1. We will outline the proof of (ii) and omit the proof of (i).

Recall $\zeta = (\phi_1, 1, 0, \ldots, 0)'$. It holds that $g(\mu_t) = \phi_1 s_{1,t-1} + s_{2,t-1} = \zeta' S_{t-1}$. For every $\kappa > 0$, there exists a $B_\kappa > 0$ such that

$$\begin{aligned} E(\|S_t\|^2 \mid S_{t-1}) &\le \|\Phi' S_{t-1}\|^2 + \delta'\delta V(\mu_t) \\ &\le S'_{t-1}\Phi\Phi' S_{t-1} + \delta'\delta(1+\kappa)\lambda S'_{t-1}\zeta\zeta' S_{t-1} + \delta'\delta B_\kappa \\ &\le (1+\kappa)S'_{t-1}\Upsilon_1 S_{t-1} + \delta'\delta B_\kappa. \end{aligned}$$

Using the same argument as the proof of Theorem 1, we can show that

$$E(\|S_t\|^2 \mid S_{t-h}) \leq (1+\kappa)^h S_{t-h}'\Upsilon_h S_{t-h} + \sum_{j=0}^{h-1}(1+\kappa)^j \delta'\Upsilon_j\delta B_\kappa.$$

The rest of the proof is the same as that of Theorem 1, and we will omit details. □

Proof of the Statement in Remark 3.9. We first observe that

$$\begin{aligned} g(\mu_t) &= \nu + \phi_1 h(y_{t-1}) + \delta_1\varepsilon_{t-1} \\ &= \nu + \phi_1 g(\mu_{t-1}) + (\phi_1+\delta_1)\varepsilon_{t-1}, \end{aligned}$$

so $\{g(\mu_t)\}$ is a Markov Chain. For every $\kappa > 0$, there exists a $B_\kappa > 0$ such that

$$\begin{aligned} &E[(g(\mu_t)^2) \mid g(\mu_{t-1})] \\ &\quad = [\nu + \phi_1 g(\mu_{t-1})]^2 + (\phi_1+\delta_1)^2 V(\mu_{t-1}) \\ &\quad \leq [\nu + \phi_1 g(\mu_{t-1})]^2 + (\phi_1+\delta_1)^2\{(1+\kappa)\lambda[g(\mu_{t-1})]^2 + B_\kappa\}. \end{aligned}$$

If $\phi_1^2 + \lambda(\phi_1+\delta_1)^2 < 1$, it is easy to check that the drift condition (D) is met, thus the process $\{g(\mu_t)\}$ is geometrically ergodic by Lemma 1, and $E_\pi[g(\mu_t)]^2 < \infty$, where π is the unique invariant probability measure. □

In order to apply Theorems 1 and 2 for the log-Gamma-M-GARMA model (9), we need the following Lemma.

Lemma 2. *Suppose $Y \sim \mathrm{Gam}(c\mu^d, c\mu^{d-1})$, where $c, d > 0$ are fixed constants. Let $g(\mu) = E(\log Y)$ and $V(\mu) = \mathrm{Var}(\log Y)$. Then*

$$\limsup_{|g(\mu)|\to\infty} V(\mu)/[g(\mu)]^2 = 1.$$

Proof. We have

$$g(\mu) = \psi(c\mu^d) - \log(c\mu^{d-1}), \qquad V(\mu) = \psi_1(c\mu^d),$$

where ψ and ψ_1 are the digamma and trigamma functions respectively. The following recurrence relationship hold for $x > 0$

$$\psi(x+1) = \psi(x) + \frac{1}{x}, \qquad \psi_1(x+1) = \psi_1(x) - \frac{1}{x^2},$$

which, together with the fact $\psi_1(x) = \sum_{n=0}^\infty (x+n)^{-2}$ imply that

$$\begin{aligned} &\lim_{x\to 0} x\psi(x) = -1, && \lim_{x\to\infty} \psi(x)/\log(x) = 1, \\ &\lim_{x\to 0} x^2\psi_1(x) = 1, && \lim_{x\to\infty} \psi_1(x) = 0. \end{aligned}$$

Therefore, the conclusion follows. □

A.2. Proofs for Section 4.1

Proof of Theorem 4. Since $\{h(y_t)\}$ is stationary and ergodic, by the ergodic theorem,

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=p+1}^{T} X_t X_t' = E(X_t X_t') = \boldsymbol{V} \quad \text{a.s.}, \tag{25}$$

where $\boldsymbol{V}$ is assumed to be finite and positive definite. Since $\boldsymbol{W} := E[V(\mu_t)X_{t-1}X_{t-1}']$ is also finite and positive definite, again by the ergodic theorem

$$\begin{aligned} &\lim_{T\to\infty} \frac{1}{T}\sum_{t=p+1}^{T} E\left[X_{t-1}X_{t-1}'\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right] \\ &\quad = \lim_{T\to\infty} \frac{1}{T}\sum_{t=p+1}^{T} V(\mu_t)X_{t-1}X_{t-1}' = \boldsymbol{W} \quad \text{a.s.} \end{aligned} \tag{26}$$

Furthermore, because $\boldsymbol{W}$ is finite and the process $\{X_{t-1}X_{t-1}'\varepsilon_t^2\}$ is stationary, for any $\epsilon > 0$,

$$\begin{aligned} &E\left\{\frac{1}{T}\sum_{t=p+1}^{T} E\left[X_{t-1}X_{t-1}'\varepsilon_t^2 I\{X_{t-1}'X_{t-1}\varepsilon_t^2 \geq \epsilon T\} \mid \mathcal{F}_{t-1}\right]\right\} \\ &\quad = \frac{1}{T}\sum_{t=p+1}^{T} E\left[X_{t-1}X_{t-1}'\varepsilon_t^2 I\{X_{t-1}'X_{t-1}\varepsilon_t^2 \geq \epsilon T\}\right] \to 0, \end{aligned}$$

and therefore

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=p+1}^{T} E\left[X_{t-1}X_{t-1}'\varepsilon_t^2 I\{X_{t-1}'X_{t-1}\varepsilon_t^2 \geq \epsilon T\} \mid \mathcal{F}_{t-1}\right] = 0$$

in probability. (27)

By Corollary 3.1 of Hall and Heyde (1980) and a Cramer–Wold device, (26) and (27) imply that

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} X_{t-1}\varepsilon_t \Rightarrow N(\boldsymbol{0}, \boldsymbol{W}),$$

which, together with (25), completes the proof. □

We now verify the moment condition for the log-Gamma-M-GARMA model (9).

Proof of Proposition 1. Suppose $Y \sim \mathrm{Gam}(c\mu^d, c\mu^{d-1})$, where $c, d > 0$ are fixed constants. Recall $V(\mu) = \mathrm{Var}(\log Y)$, and define $V_3(\mu) = E[\log Y - g(\mu)]^3$ and $V_4(\mu) = E[\log Y - g(\mu)]^4$. Using the properties of polygamma functions, we have similarly as Lemma 2,

$$\limsup_{|g(\mu)|\to\infty} |V_3(\mu)|/|g(\mu)|^3 = 2, \qquad \limsup_{|g(\mu)|\to\infty} V_4(\mu)/[g(\mu)]^4 = 9.$$

Consequently, for every $\kappa > 0$, there exists a $B_\kappa > 0$ such that

$$\begin{aligned} &V(\mu) \leq (1+\kappa)[g(\mu)]^2 + B_\kappa, \\ &\quad |V_3(\mu)| \leq 2(1+\kappa)|g(\mu)|^3 + B_\kappa, \\ &\quad V_4(\mu) \leq 9(1+\kappa)[g(\mu)]^4 + B_\kappa. \end{aligned} \tag{28}$$

To simplify the notation in the calculation, we introduce the collection of matrices

$$\begin{aligned} \mathcal{A} = \{A :\ & A \text{ is a } p\times p \text{ non-negative definite matrix,} \\ & \text{and each entry of } A \text{ is a finite order multivariate} \\ & \text{polynomial of } \phi_1, \ldots, \phi_p.\}. \end{aligned}$$

Let $\varsigma = (1, 0, \ldots, 0)'$. For any matrix $A \in \mathcal{A}$, we calculate

$$\begin{aligned} &E[(X_t'AX_t)^2 \mid X_{t-1}] \\ &\quad = (X_{t-1}'\Phi'A\Phi X_{t-1})^2 + E\left[4(\varsigma'A\Phi X_{t-1})^2\cdot\varepsilon_t^2 \mid X_{t-1}\right] \\ &\qquad + E\left[2X_{t-1}\Phi'A\Phi X_{t-1}\cdot\varsigma'A\varsigma\cdot\varepsilon_t^2 \mid X_{t-1}\right] \\ &\qquad + E\left[4\varsigma'A\Phi X_{t-1}\cdot\varsigma'A\varsigma\cdot\varepsilon_t^3 \mid X_{t-1}\right] + E\left[(\varsigma'A\varsigma)^2\cdot\varepsilon_t^4 \mid X_{t-1}\right] \\ &\quad \leq (X_{t-1}'\Phi'A\Phi X_{t-1})^2 + 6X_{t-1}\Phi'A\Phi X_{t-1}\cdot\varsigma'A\varsigma\cdot V(\mu_t) \\ &\qquad + 4\varsigma'A\Phi X_{t-1}\cdot\varsigma'A\varsigma\cdot|V_3(\mu_t)| + (\varsigma'A\varsigma)^2\cdot V_4(\mu_t). \end{aligned}$$

Set $\alpha = X_{t-1}'\Phi'A\Phi X_{t-1}$ and $\beta = X_{t-1}'\Phi_1'A\Phi_1 X_{t-1}$. Using (28), and the fact $g(\mu_t)\varsigma = \Phi_1 X_{t-1}$, after collecting terms, we have

$$\begin{aligned} E[(X_t'AX_t)^2 \mid X_{t-1}] \leq\ & \alpha^2 + (1+\kappa)[6\alpha\beta + 4(c\alpha+\beta/c)\beta + 9\beta^4] \\ & + 10B_\kappa\cdot\varsigma'A\varsigma\cdot\alpha + 13B_\kappa(\varsigma'A\varsigma)^2 \end{aligned}$$

for any $c > 0$. Let ξ be the positive root of the polynomial $z^3 + 3z^2 - 1$. Set $c = \xi$, we have

$$\begin{aligned} E[(X_t'AX_t)^2 \mid X_{t-1}] \leq\ & (1+\kappa)[\alpha + (3+2\xi)\beta]^2 \\ & + 10B_\kappa\cdot\varsigma'A\varsigma\cdot\alpha + 13B_\kappa(\varsigma'A\varsigma)^2. \end{aligned} \tag{29}$$

For each integer $h > 0$, by repeating (29) h times, we know there exists a constant C_κ, and a matrix $A_\kappa \in \mathcal{A}$ (both depending on κ) such that

$$E(\|X_t\|^4 \mid X_{t-h}) \leq (1+\kappa)^h (X'_{t-h}\Xi_h X_{t-h})^2 + X'_{t-h}A_\kappa X_{t-h} + C_\kappa.$$

Choose h, such that the operator norm of Ξ_h is less than $1-2\omega$ for some $\omega > 0$. Choosing κ such that $(1+\kappa)^h(1-2\omega)^2 < 1-\omega$, we have

$$E(\|X_t\|^4 \mid X_{t-h}) \leq (1-\omega)\|X_{t-h}\|^4 + X'_{t-h}A_\kappa X_{t-h} + C_\kappa.$$

Now we can apply Theorem 14.3.7 of Meyn and Tweedie (2009) to obtain that under the invariant probability π, $E_\pi \|X_t\|^4 < \infty$, and the proof is complete. □

References

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Internat. Econom. Rev. 39, 885–905.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Amer. Statist. Assoc. 96, 42–55.
Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. J. R. Stat. Soc. Ser. B 63, 167–241.
Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Ser. B 64, 253–280.
Bartlett, M.S., 1947. The use of transformations. Biometrics 3, 39–52.
Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M., 2003. Generalized autoregressive moving average models. J. Amer. Statist. Assoc. 98, 214–223.
Beran, J., 1995. Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. J. R. Stat. Soc. Ser. B 57, 659–673.
Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. Rev. Econ. Stat. 69, 542–547.
Bondon, P., 2009. Estimation of autoregressive models with epsilon-skew-normal innovations. J. Multivariate Anal. 100, 1761–1776.
Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26, 211–252.
Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, CA.
Brownlees, C.T., Cipollini, F., Gallo, G.M., 2012. Multiplicative error models. In Bauwens, L., Hafner, C., and Laurent, S. (Eds.), *Handbook of Volatility Models*, pp. 225–248.
Chan, K.S., Tong, H., 1985. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. Adv. Appl. Probab. 17, 666–678.
Chen, S., Min, W., Chen, R., 2013. Model identification for time series with dependent innovations. Statist. Sinica 23, 873–899.
Davis, R.A., Dunsmuir, W.T.M., Streett, S.B., 2003. Observation-driven models for Poisson counts. Biometrika 90, 777–790.
Davis, R., Wu, R., 2009. A negative binomial model for time series of counts. Biometrika 96, 735–749.
Domínguez, M.A., Lobata, I.N., 2003. Testing the martingale difference hypothesis. Econometric Rev. 22, 351–377.
Douc, R., Doukhan, P., Moulines, E., 2013. Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. Stochastic Process. Appl. 123, 2620–2647.
Engle, R.F., 2002. New frontiers for ARCH models. J. Appl. Econometrics 17, 425–446.
Engle, R.F., Gallo, G.M., 2006. A multiple indicators model for volatility using intra-daily data. J. Econometrics 131, 3–27.
Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.
Escanciano, J.C., Lobata, I.N., 2009. Testing the martingale hypothesis. In: Patterson, K., Mills, T.C. (Eds.), Palgrave Handbook of Econometrics. Palgrave MacMillan, New York, pp. 972–1003.
Escanciano, J.C., Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. J. Econometrics 134, 151–185.
Ferrari, S.L.P., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31, 799–815.
Fokianos, K., 2012. Count Time Series Models. In: Rao, T.S., Rao, S.S., Rao, C.R. (Eds.), Handbook of Statistics Vol 30: Time Series Analysis—Methods and Applications. Elsevier B.V, Amsterdam, pp. 315–347.
Fokianos, K., Fried, R., 2010. Intereventions in INGARCH processess. J. Time Ser. Anal. 31, 210–225.
Fokianos, K., Rahbek, A., Tjøstheim, D., 2009. Poisson autoregression. J. Amer. Statist. Assoc. 104, 1430–1439.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press Inc., New York.
Hannan, E.J., 1973. The asymptotic theory of linear time-series models. J. Appl. Probab. 10, 130–145.
Hansen, B., 1994. Autoregressive conditional density estimation. Internat. Econom. Rev. 35, 705–730.
Hansen, P.R., Lunde, A., 2005. A realized variance for the whole day based on intermittent high-frequency data. J. Financ. Econom. 3, 525–554.
Harvey, A.C., 1993. Time Series Models, second ed. MIT Press, Cambridge, MA.
He, X., Shi, P., 1998. Monotone B-spline smoothing. J. Amer. Statist. Assoc. 93, 643–650.
Kuan, C.M., Lee, W.M., 2004. A new test of the martingale difference hypothesis. Stud. Nonlinear Dyn. Econom. 8, 1–26.
Li, W.K., McLeod, A.I., 1988. ARMA modelling with non-Gaussian innovations. J. Time Ser. Anal. 9, 155–168.
Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, second ed. Chapman & Hall/CRC, Boca Raton, Florida.
McLeod, A.I., Li, W.K., 1983. Diagnostic checking ARMA time series models using squared residual autocorrelations. J. Time Ser. Anal. 4, 269–273.
Meyn, S., Tweedie, R.L., 2009. Markov Chains and Stochastic Stability, second ed. Cambridge University Press, Cambridge.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.
Neumann, M.H., 2011. Absolute regularity and ergodicity of Poisson count processes. Bernoulli 17, 1268–1284.
Nocedal, J., Wright, S., 1999. Numerical Optimization. Springer-Verlag, New York.
Ramsay, J.O., 1988. Monotone regression splines in action. Statist. Sci. 3, 425–461.
Ramsay, J.O., 1998. Estimating smooth monotone functions. J. R. Stat. Soc. Ser. B 60, 365–375.
Rocha, A.V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. Test 18, 529–545.
Startz, R., 2008. Binomial autoregressive moving average models with an application to US recessions. J. Bus. Econom. Statist. 26, 1–8.
Sowell, F.B., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. J. Econometrics 53, 165–188.
Takahashi, M., Omori, Y., Watanabe, T., 2009. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. Comput. Statist. Data Anal. 53, 2404–2426.
Tiku, M.L., Wong, W.K., Vaughan, D.C., Bian, G., 2000. Time series models in non-normal situations: Symmetric innovations. J. Time Ser. Anal. 21, 571–596.
Tsay, R.S., Tiao, G.C., 1984. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. J. Amer. Statist. Assoc. 79, 84–96.
Woodard, D.B., Matteson, D.S., Henderson, S.G., 2011. Stationarity of generalized autoregressive moving average models. Electron. J. Stat. 5, 800–828.
Wong, W.K., Bian, G., 2005. Estimating parameters in autoregressive models with asymmetric innovations. Statist. Probab. Lett. 71, 61–70.
Yao, Q., Brockwell, P.J., 2006. Gaussian maximum likelihood estimation for ARMA models I: Time series. J. Time Ser. Anal. 27, 857–875.