Journal of Econometrics 150 (2009) 219–230

Contents lists available at ScienceDirect

Journal of Econometrics

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

Maximum entropy autoregressive conditional heteroskedasticity model☆

Sung Y. Park[a], Anil K. Bera[b,*]

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

[b] Department of Economics, University of Illinois, 1206 S.6th Street, Champaign, IL 61820, USA

ARTICLE INFO

Article history:
Available online 8 January 2009

JEL classification:
C40
C61
G10

Keywords:
Maximum entropy density
ARCH models
Excess kurtosis
Asymmetry
Peakedness of distribution
Stock returns data

ABSTRACT

In many applications, it has been found that the autoregressive conditional heteroskedasticity (ARCH) model under the conditional normal or Student's *t* distributions are not general enough to account for the excess kurtosis in the data. Moreover, asymmetry in the financial data is rarely modeled in a systematic way. In this paper, we suggest a general density function based on the maximum entropy (ME) approach that takes account of asymmetry, excess kurtosis and also of high peakedness. The ME principle is based on the efficient use of available information, and as is well known, many of the standard family of distributions can be derived from the ME approach. We demonstrate how we can extract information functional from the data in the form of moment functions. We also propose a test procedure for selecting appropriate moment functions. Our procedure is illustrated with an application to the NYSE stock returns. The empirical results reveal that the ME approach with a fewer moment functions leads to a model that captures the stylized facts quite effectively.

1. Introduction

There have been a number of theoretical and empirical studies in the area of density estimation. Since complete information about the density function is not available, a parametric form is generally assumed before performing estimation. In non-parametric approach, estimated tail-behavior of the density, which is of substantial concern in most financial applications, is not satisfactory due to the scarcity of data in the tail part of the distribution. If the density function is correctly specified, then classical maximum likelihood estimation preserves efficiency and consistency. The true density, however, is not known in almost all cases; therefore, an assumed density function could be misspecified. The main contribution of this paper is to show that how can one extract useful information about the unknown density from a given data by imposing some well-defined moment functions in analyzing financial time-series data. By so doing one can reduce the degree of model misspecification considerably. We use the maximum entropy density (MED) as conditional density function in the autoregressive conditional heteroskedasticity (ARCH) framework. Since Engle's (1982) pioneering work and its generalization by Bollerslev (1986), ARCH-type models have been widely used, and various extensions have been suggested, primarily in two directions. *First* extension has concentrated on generalizing the conditional variance function. *Second* extension deals with the form of the conditional density function. Various non-normal conditional density functions have been proposed to explain high leptokurtic behavior. Although these two extension are inter-related, in this paper we concentrate on the second extension, namely, finding a suitable general form of the conditional density. If we impose certain moment conditions, we can obtain normal, Student's *t*, generalized error distribution (GED) and Pearson type-IV distribution through MED formulation. In this sense, our proposed maximum entropy ARCH (MEARCH) model is a very general one.

MEARCH model is quite related to other moment-based estimation, such as generalized method of moments (GMM) and maximum empirical likelihood (MEL) estimation. All these estimations could also be considered within estimating function (EF) approach, for example, see Bera et al. (2006). The purpose of this paper is twofold. First, we present the characterization of MED, and show how, within an ARCH framework, our selected moment conditions capture asymmetry and excess kurtosis of

☆ We are grateful to the editors Chung-Min Kuan and Yongmiao Hong, and two anonymous referees for many pertinent comments and suggestions. We would also like to thank the participants of the First Symposium on Econometric Theory and Application (SETA) at the Institute of Economics, Academia Sinica, Taipei, Taiwan, May 18–20, 2005, and at some other conferences for helpful comments and discussions. In particular, we are thankful to Jin-Chuan Duan, Alastair R. Hall, George Judge, Nour Meddahi, Eric Renault, and Vicky Zinde-Walsh. However, we retain the responsibility for any remaining errors. Financial support from the Research Board, University of Illinois at Urbana-Champaign is gratefully acknowledged.

* Corresponding author. Tel.: +1 217 333 4596; fax: +1 217 244 6678.
E-mail addresses: sungpark@sungpark.net (S.Y. Park), abera@uiuc.edu (A.K. Bera).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.12.014

financial data. Second, we introduce estimation procedure of the MEARCH model, and suggest moment selection criteria based on Rao's score test.

The rest of the paper is organized as follows. In the next section we present some basic characteristics of MED and discuss estimation of a basic model. In Section 3, we propose our MEARCH model along with its estimation and the moment selection test. Section 4 provides an empirical application to the daily return of NYSE with specific moment functions that generate a skewed and heavy tail distribution. The paper is concluded in Section 5.

2. Maximum entropy density

The maximum entropy density is obtained by maximizing Shannon's (1948) entropy measure

$$H(f) = -\int f(x)\ln f(x)\mathrm{d}x, \tag{1}$$

satisfying

$$\mathbb{E}[\phi_j(x)] = \int \phi_j(x)f(x)\mathrm{d}x = \mu_j, \quad j = 0, 1, 2, \ldots, q, \tag{2}$$

where the μ_j's are known values. The normalization constraint corresponds to $j = 0$ by setting $\phi_0(x)$ and μ_0 to 1. The Lagrangian for the above optimization problem is given by

$$\mathcal{L} = -\int f(x)\ln f(x)\mathrm{d}x + \sum_{j=0}^{q}\lambda_j\left[\int \phi_j(x)f(x)\mathrm{d}x - \mu_j\right], \tag{3}$$

where λ_j is the Lagrange multiplier corresponding the j-th constraint in (2), $j = 0, 1, 2, \ldots, q$. The solution to the above optimization problem, obtained by simple calculus of variation, is given by [Zellner and Highfield (1988) and Golan et al. (1996), p. 36)]

$$f(x) = \frac{1}{\Omega(\boldsymbol{\lambda})}\exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(x)\right], \tag{4}$$

where $\Omega(\lambda)$ is calculated by $\int f(x)\mathrm{d}x = 1$, and can be expressed in terms of the Lagrangian multipliers as $\Omega(\lambda) = \int \exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(x)\right]\mathrm{d}x$. $\Omega(\lambda)$ is known as the "partition function" that converts the relative probabilities to absolute probabilities. In the maximization problem (1) and (2), $\phi_j(x)$ in the moment constraint equation (hereafter, we call ϕ as moment function) is only function of data x. Due to this characteristics of ϕ_j, we obtain simple exponential forms as a solution to the maximization problem. Since (4) belongs to exponential family, λ_j's and $\phi_j(x)$'s are the natural parameters and the corresponding sufficient statistics, respectively.

We extend simple exponential solution forms to more general exponential forms by introducing additional parameters, γ, in ϕ_j. Consider the following optimization problem,

$$\max_f H(f) = -\int f(x)\ln f(x)\mathrm{d}x, \tag{5}$$

satisfying

$$\int \phi_j(x, \boldsymbol{\gamma})f(x)\mathrm{d}x = C_j(\boldsymbol{\gamma}), \quad j = 0, 1, 2, \ldots, q. \tag{6}$$

The solution to (5) and (6), obtained by applying the same Lagrangian's procedure, is the general exponential density

$$f(x) = \frac{1}{\Omega(\lambda, \gamma)}\exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(x, \gamma)\right], \tag{7}$$

where $\Omega(\lambda, \gamma) = \int \exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(x, \gamma)\right]\mathrm{d}x$. Thus, by adding additional parameter γ to moment functions, the ME formulation provides a more general family of distributions.

The moment conditions can be interpreted as known prior information, and using these we can achieve a least biased distribution by the ME principle. Suppose we have no prior information except for the normalization constraint, then the solution is the uniform distribution which is a "perfect smooth" density. If we have an additional information, say $\int xf(x)\mathrm{d}x = \mu_1 > 0$, then the solution takes the form, $f(x|\lambda_1) = \lambda_1\exp[-\lambda_1 x]$ where $x \in [0, \infty)$. With a further constraint, say $\int x^2 f(x)\mathrm{d}x = \mu_2$, the associated solution is the normal distribution. In the next subsection, we discuss in detail the role of different moment conditions in some of the commonly used distributions, and also suggest new densities by constructing and selecting certain moment functions judiciously.

All solutions are functions of Lagrangian multipliers which represent *marginal contribution* (shadow price) of constraints to the objective value. For example, suppose $\hat{\lambda}_2$ corresponding to $\int x^2 f(x)\mathrm{d}x = \mu_2$ is estimated to be close to 0, then, there is little contribution of this moment constraint to the objective function. Consequently, the Lagrangian multiplier reflect the information content of each constraint.

2.1. Maximum entropy characterization of thick tail, peakedness and asymmetry

Maximum entropy distribution has a very flexible functional form. By choosing a sequence of moment functions $\phi_j(x)$, $j = 1, 2, \ldots, q$, we can generate a sequence of various flexible MED functions. Many well-known families of distributions can be obtained as special cases of MED function. Kagan et al. (1973) provided characterization of many distributions, such as, the beta, gamma, exponential and Laplace distributions as ME densities. In Table 1, we present a number of well-known distributions under various moment constraints. These common distributions can be interpreted in an information theoretic way that they are least biased density functions obtained by imposing certain moment constraints which are inherent in the data. If we add more and more moment constraints, the resulting density, $f(x)$ will be more unsmooth.

Let us consider three moment functions x^2, $\ln(\gamma^2 + x^2)$ and $\ln(1 + x^2)$ that correspond to normal, Student's t and Cauchy distributions, respectively. These three and two other moment functions from the generalized error distribution (GED) are plotted in Fig. 1. From Eq. (7), we note that a moment function $\phi_j(x, \gamma)$ adds to the log-density $\ln f(x)$ an extra term $-\lambda_j\phi_j(x, \gamma)$ when the moment constraint is binding. $\phi_j(x, \gamma) = \ln(\gamma^2 + x^2)$ penalizes the tail events *less* severely than the function $\phi_j(x, \gamma) = x^2$ (that generate the normal density) to adhere the maximum value of the entropy under the constraints.

This intuitive penalization mechanism results in *heavier* tails for the Student's t density. As the value of γ^2 decreases, $\phi_j(x, \gamma)$ takes less extreme values at tails and that in turn makes the tails of the densities thicker. Therefore, in some sense, the shape of the resulting density has a close link with the "inverted" shape of $\phi_j(x, \gamma)$. This observation, in general, leads us to the choice of different moment functions. Various financial data such as, stock returns, inflation rates and exchange rates display both thick tails and high peakedness. Wang et al. (2001) proposed the exponential generalized beta distribution of the second kind to explain thick tails and high peakedness of financial time-series data. From Fig. 1 and the above discussion on the link between $\phi_j(x, \gamma)$ and $f(x)$, we can say that $\phi_j(x, \gamma) = \ln(\gamma^2 + x^2)$ type of functions cannot capture peakedness.

Table 1
Characterization of some common densities as MED.

Type	Moment constraints	Corresponding form of the density, $f(x)$	Common form of the density function, $f(x)$	
Uniform	None	$\exp[-\lambda_0]$	$\frac{1}{b-a}$,	$a \le x \le b$, $(b > a)$
Exp.	$\int xf(x)\mathrm{d}x = m\ (m > 0)$	$\exp[-\lambda_0 - \lambda_1 x]$	$\frac{1}{m}\exp\left[-\frac{x}{m}\right]$,	$0 \le x < \infty$
Normal	$\int xf(x)\mathrm{d}x = \mu$	$\exp\left[-\lambda_0 - \lambda_1 x - \lambda_2 x^2\right]$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$,	$-\infty < x < \infty$
	$\int (x-\mu)^2 f(x)\mathrm{d}x = \sigma^2$			
Log-normal	$\int \ln x f(x)\mathrm{d}x = \mu$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 (\ln x)^2\right]$	$\frac{1}{\sigma\sqrt{2\pi x}}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$,	$0 < x < \infty$
	$\int (\ln x-\mu)^2 f(x)\mathrm{d}x = \sigma^2$			
Gen. exp.	$\int x^i f(x)\mathrm{d}x = \mu_i\ i = 1, 2, \ldots N$	$\exp\left[-\sum_{n=0}^{N} \lambda_n x^n\right]$	$\exp\left[-\sum_{n=0}^{N} \lambda_n x^n\right]$	$-\infty < x < \infty$
Double exp.	$\int \lvert x-\mu\rvert f(x)\mathrm{d}x = \sigma^2$	$\exp\left[-\lambda_0 - \lambda_1 x - \lambda_2 \lvert x\rvert^2\right]$	$C(\theta)\exp\left[-\lvert x-\mu\rvert/\sigma^2\right]$	$-\infty < x < \infty$
Gamma	$\int xf(x)\mathrm{d}x = a\ (a > 0)$	$\exp[-\lambda_0 - \lambda_1 x - \lambda_2 \ln x]$	$\frac{1}{\Gamma(a)}\exp^{-x} x^{a-1}$,	$0 < x < \infty$
	$\int \ln x f(x)\mathrm{d}x = \frac{\Gamma'(a)}{\Gamma(a)}$			
Chi-squared with ν df	$\int xf(x)\mathrm{d}x = \nu$	$\exp[-\lambda_0 - \lambda_1 x - \lambda_2 \ln x]$	$\frac{1}{2^{\nu/2}\Gamma(\nu/2)}\exp^{-x/2} x^{\nu/2-1}$,	$0 < x < \infty$
	$\int \ln x f(x)\mathrm{d}x = \frac{\Gamma'(\frac{1}{2})}{\Gamma(\frac{1}{2})} + \ln 2$			
Weibull[a]	$\int x^a f(x)\mathrm{d}x = 1\ (a > 0)$	$\exp[-\lambda_0 - \lambda_1 x^a - \lambda_2 \ln x]$	$ax^{a-1}\exp[-x^a]$,	$0 < x < \infty$
	$\int \ln x f(x)\mathrm{d}x = -\gamma/a$			
GED	$\int \lvert x\rvert^{\nu} f(x)\mathrm{d}x = c(\nu)$	$\exp[-\lambda_0 - \lambda_1 \lvert x\rvert^{\nu}]$	$C(\nu)\exp[\lvert x\rvert^{\nu}]$,	$-\infty < x < \infty$
Beta	$\int \ln x f(x)\mathrm{d}x = \frac{\Gamma'(a)}{\Gamma(a)} - \frac{\Gamma'(a+b)}{\Gamma(a+b)}$	$\exp[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1-x)]$	$\frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$,	$0 < x < 1$
	$\int \ln(1-x) f(x)\mathrm{d}x = \frac{\Gamma'(b)}{\Gamma(b)} - \frac{\Gamma'(a+b)}{\Gamma(a+b)}$			
Cauchy	$\int \ln(1+x^2) f(x)\mathrm{d}x = 2\ln 2$	$\exp\left[-\lambda_0 - \lambda_1 \ln(1+x^2)\right]$	$\frac{1}{\pi(1+x^2)}$,	$-\infty < x < \infty$
Student's t	$\int \ln(r^2+x^2) f(x)\mathrm{d}x = \ln(r^2) + \frac{\Gamma'(\frac{1+r^2}{2})}{\Gamma(\frac{1+r^2}{2})} - \frac{\Gamma'(\frac{r^2}{2})}{\Gamma(\frac{r^2}{2})}$	$\exp\left[-\lambda_0 - \lambda_1 \ln(r^2+x^2)\right]$	$\frac{\Gamma[(r^2+1)/2]}{\sqrt{\pi r^2}\,\Gamma(r^2/2)}\frac{1}{(1+x^2/r^2)^{(r^2+1)/2}}$,	$-\infty < x < \infty$
Pearson type-IV	$\int \tan^{-1}\left(\frac{x}{r}\right) f(x)\mathrm{d}x = c_1(r)$	$\exp\left[-\lambda_0 - \lambda_1 \tan^{-1}\left(\frac{x}{r}\right) - \lambda_2 \ln(r^2+x^2)\right]$	$K\left(1+\frac{x^2}{r^2}\right)^{-m}\exp\left[\delta\tan^{-1}\left(\frac{x}{r}\right)\right]$,	$-\infty < x < \infty$
	$\int \ln(r^2+x^2) f(x)\mathrm{d}x = c_2(r)$			
Generalized Student's t	$\int x^{i-2} f(x)\mathrm{d}x = \mu_i\ i = 3, 4, \ldots k$	$\exp\left[-\lambda_0 - \lambda_1 \tan^{-1}(\frac{x}{r}) - \lambda_2 \ln(r^2+x^2) - \sum_{i=3}^{k} \lambda_i x^{i-2}\right]$		$-\infty < x < \infty$
	$\int \tan^{-1}(\frac{x}{r}) f(x) = c_1(r)$			
	$\int \ln(r^2+x^2) f(x) = c_2(r)$			
Generalized log-normal	$\int x^{i-2} f(x)\mathrm{d}x = \mu_i\ i = 3, 4, \ldots k$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 (\ln(x))^2 - \sum_{i=3}^{k} \lambda_i x^{i-2}\right]$		$0 < x < \infty$
	$\int \ln x f(x) = c_1(r)$			
	$\int (\ln(x))^2 f(x) = c_2(r)$			

[a] γ denotes the Euler constant.

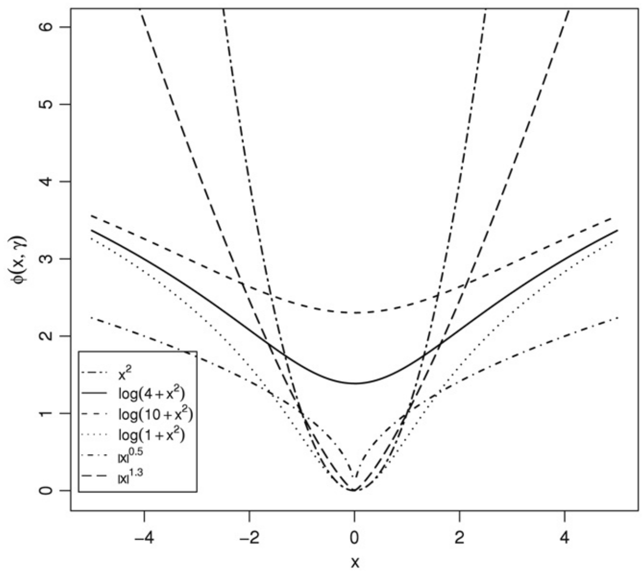

Fig. 1. Moment functions $\phi_j(x, \gamma)$ representing thick tail.

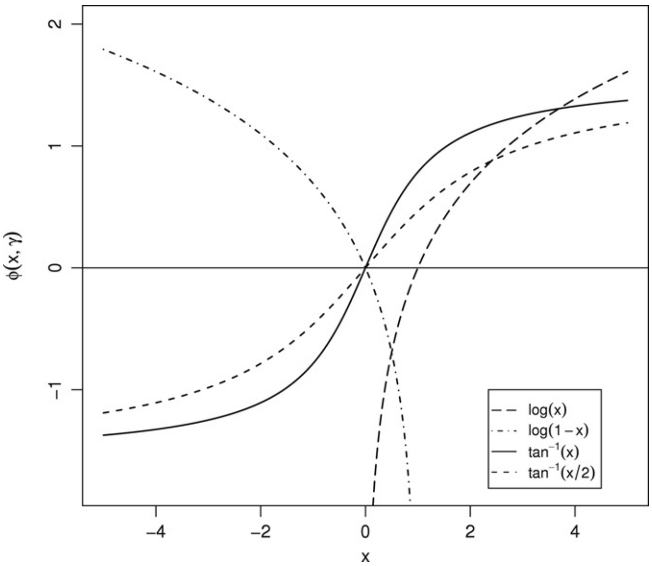

Fig. 3. Moment functions $\phi_j(x, \gamma)$ representing skewness.

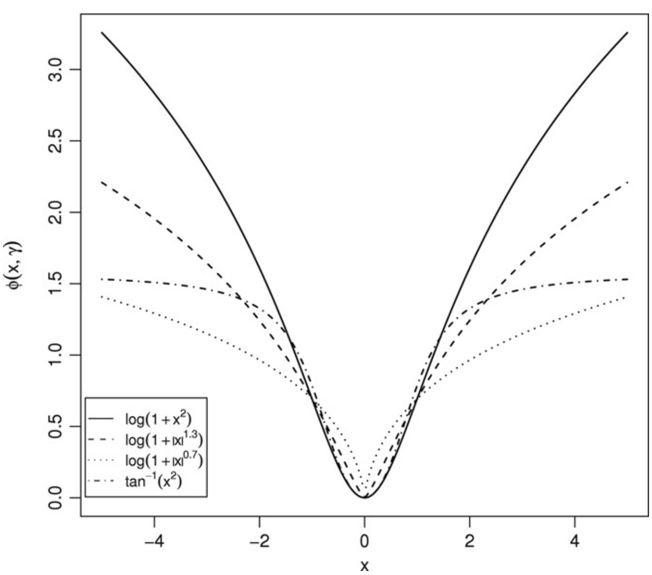

Fig. 2. Moment functions $\phi_j(x, \gamma)$ representing high peakedness.

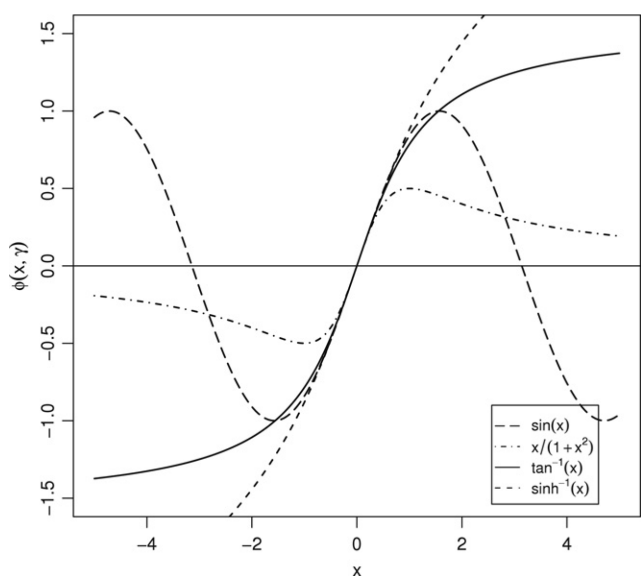

Fig. 4. Moment functions $\phi_j(x, \gamma)$ representing general skewness.

To take account of high peakedness, we suggest functions $\ln(1+|x/r|^p)$ and $\tan^{-1}(x^2/r^2)$. In Fig. 2, these functions along with $\ln(1+x^2)$ are plotted. We note that $\ln(1+|x|^{1.3})$ and $\ln(1+|x|^{0.7})$ have cusp at $x=0$, while the Cauchy moment function $\ln(1+x^2)$ does not. In $\ln(1+|x/r|^p)$ the parameter p (<2) appears to capture peakedness, while r takes account of the tail thickness. For $p \geq 2$, the cusp behavior disappears, and for this case, both p and r together capture the tail behavior of the underlying distribution. The moment function $\tan^{-1}(x^2)$ also captures high peakedness and penalizes the tails less than that of $\ln(1+x^2)$.

As is well known, financial data also displays asymmetry (skewness), see for instance, Premaratne and Bera (2000). Moment functions, $\ln x$, $0 < x < \infty$, $\ln(1-x)$, $0 < x < 1$ and $\tan^{-1}(x/\gamma)$, $-\infty < x < \infty$ can capture asymmetry, and these are plotted in Fig. 3. $\ln x$ and $\ln(1-x)$ generates the beta distribution over the range $0 \leq x \leq 1$; $\tan^{-1}(x/\gamma)$ is part of the moment functions of the Pearson type-IV density. $\mathbb{E}[\ln x] = \frac{\Gamma'(a)}{\Gamma(a)}$ is used as a moment condition in gamma density and chi-squared is a special case of gamma distribution when $\mathbb{E}[\ln x] = \frac{\Gamma'(1/2)}{\Gamma(1/2)} + \ln 2$ for $0 < x < \infty$. Premaratne and Bera (2005) used $\tan^{-1}(x/\gamma)$ to test asymmetry in leptokurtic financial data.

In general, any odd function can serve as a moment function to capture asymmetry. However, the benefits of a function like $\tan^{-1}(x/\gamma)$ are that it is bounded over the whole range and "robust" to outliers. Chen et al. (2000) used $\sin(x)$ and $\beta x/(1+\beta^2 x^2)$ with a specific value of β to test asymmetry. Tests based on such bounded functions will be more robust compared to those based on the third moment, i.e., moment function like x^3 [for more on this, see, Premaratne and Bera (2005)]. Some robust-type functions that capture general skewness are plotted in Fig. 4.

As we assign more and more moment constraints in maximization problem (1) and (2) or (5) and (6), the solution is likely to become more unsmooth (rough) functional if given moment constraints are informative. There is a close relationships between MED and the penalization method. The ME method starts with a very smooth density and, adding more moment constraints, MED is likely to have more "roughness" but with improved goodness-of-fit at the same time. Here we do not face the problem of selecting smoothing parameter or the bandwidth. Instead, we need choose moment functions priori. On the other hand, a non-parametric approach begins with a rough density (histogram), and then uses a smoothing procedure (such as selecting a proper bandwidth) to control the balance between smoothness and goodness-of-fit. Gallant (1981), Gallant and Nychka (1987) and Ryu (1993) considered (semi-) non-parametric density estimators using flexible polynomial series approaches such as Fourier series, Hermite polynomial and any orthonormal basis. These approaches are useful to fit the underlying density or functional form and to analyze asymptotic properties of estimators since very high orders of polynomial series can be easily considered. However, if one can select only a few informative functions that explain underlying density enough instead of using high orders of polynomials, the complexity and computational burden can be significantly reduced, and, moreover, some valuable interpretation can be made using the selected informative functions.

2.2. Methods of estimation

When μ_j's are unknown in (2), the maximum likelihood (ML) estimates are the same as ME estimates when μ_j's are replaced by their consistent estimates $\frac{1}{T}\sum_{t=1}^{T}\phi_j(x_t), j = 1, 2, \ldots, q$. Since exponential family distributions have a unique ML solution, the ME solution is also unique, if it exists. However, when we have general moment conditions [as in (6)], then we have to consider estimation of unknown parameter γ. Usually, this estimation problem can be solved by estimating saddle point of the objective function proposed by Kitamura and Stutzer (1997) and Smith (1997).

Let us rewrite (6) as

$$\int [\phi_j(x,\gamma) - C_j(\gamma)]f(x)\mathrm{d}x = \int \psi_j(x,\gamma)f(x)\mathrm{d}x = 0, \quad j = 1, 2, \ldots, q,$$

where $\psi_j(x,\gamma) = \phi_j(x,\gamma) - C_j(\gamma)$. The profiled objective function is obtained by substituting (7) to the Lagrangian (3)

$$\ln \int \exp\left[-\sum_{j=1}^{q}\lambda_j\psi_j(x,\gamma)\right]\mathrm{d}x. \quad (8)$$

ME estimators of the parameters γ and λ are the solution of following saddle point problem

$$\hat{\gamma}_{ME} = \arg\max_{\gamma} \ln \int \exp\left[-\sum_{j=1}^{q}\hat{\lambda}_j\psi_j(x,\gamma)\right]\mathrm{d}x,$$

where $\hat{\lambda}(\gamma)$ is given by

$$\hat{\lambda}(\gamma)_{ME} = \arg\min_{\lambda} \ln \int \exp\left[-\sum_{j=1}^{q}\lambda_j\psi_j(x,\gamma)\right]\mathrm{d}x.$$

Since the profiled objective function (8) has the exponential form it is relatively easy to calculate first order derivatives. However, $C_j(\gamma)$ could be complicated in some case or even, may not have analytic form. In such a case, $C_j(\gamma)$ can be substituted by the sample moment of $\phi_j(x,\gamma)$. Thus, we consider the following non-linear equations:

$$\int \phi_j(x,\gamma)f(x|\lambda,\gamma)\mathrm{d}x = \frac{1}{T}\sum_{t=1}^{T}\phi_j(x_t,\gamma), \quad j = 1, 2, \ldots q.$$

We can express (8) as

$$\ln \int \exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(x,\gamma) + \sum_{j=1}^{q}\lambda_jC_j(\gamma)\right]\mathrm{d}x$$
$$= \ln\left[\exp\left[\sum_{j=1}^{q}\lambda_jC_j(\gamma)\right]\cdot\int \exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(x,\gamma)\right]\mathrm{d}x\right]$$
$$= \sum_{j=1}^{q}\lambda_jC_j(\gamma) + \ln \int \exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(x,\gamma)\right]\mathrm{d}x.$$

Since from (4) $\ln\int \exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(x,\gamma)\right] = \ln\Omega(\lambda,\gamma)$ the above expression can be simplified as

$$\sum_{j=1}^{q}\lambda_jC_j(\gamma) + \ln\Omega(\lambda,\gamma). \quad (9)$$

From (7) the log-likelihood function is given by

$$l(\lambda,\gamma) = -T\ln\Omega(\lambda,\gamma) - \sum_{j=1}^{q}\lambda_j\sum_{t=1}^{T}\phi_j(x_t,\gamma). \quad (10)$$

From (9) and (10) it is clear that profiled objective function is the same as $-(1/T)l(\lambda,\gamma)$. Thus, the first order condition for the ME principle and the ML principle are the same under the general moment problem. However, the second order condition may differ between those principles because there exist restrictions that the Lagrange multipliers are the functions of γ in the ME problem. However, in the ML problem, λ_j is not a function of γ because $(1/T)\sum_{t=1}^{T}\phi_j(x_t,\gamma)$ does not affect the form of the solution (7).

3. Maximum entropy GARCH model

Various ARCH-type models under the assumption of non-normal conditional density have been proposed to explain leptokurtic and asymmetric behavior of financial data. We propose to use flexible ME density to capture such stylized facts, and consider the following model:

$$y_t = m_t(x_t;\zeta) + \epsilon_t, \quad t = 1, 2, \ldots, T,$$

where $m_t(\cdot)$ is the conditional mean function, x_t is a $K \times 1$ vector of exogenous variables and ζ is a vector of parameters. We assume that $\epsilon_t|\mathcal{F}_{t-1} \sim g(0,h_t)$, where $g(\cdot)$ is the unknown density function of ϵ_t conditional on the set of past information $\mathcal{F}_{t-1}$, and $h_t = \alpha_0 + \sum_{j=1}^{p}\alpha_j\epsilon_{t-j}^2 + \sum_{j=1}^{s}\beta_jh_{t-j}$.

Following (7), we can write the density function of the standardized error term $\eta_t\,(=\epsilon_t/\sqrt{h_t})$ in a general MED form as

$$f(\eta_t) = \frac{1}{\Omega(\lambda,\gamma)}\exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(\eta_t,\gamma)\right], \quad (11)$$

where $\phi_j(\eta_t,\gamma), j = 1, 2, \ldots, q$, denote the moment functions. We will term ARCH model with conditional density $f(\epsilon_t|\mathcal{F}_{t-1})$ implied by the above MED $f(\eta_t)$ as the maximum entropy ARCH (MEARCH) model. The (conditional) quasi-log-density function of ϵ_t is given by

$$l_t^{QME}(\theta) = -\ln\Omega(\lambda,\gamma) - \sum_{j=1}^{q}\lambda_j\phi_j\left(\frac{y_t - x_t'\zeta}{\sqrt{h_t}},\gamma\right) - \frac{1}{2}\ln h_t, \quad t = 1, \ldots, T,$$

where $\theta = (\alpha',\beta',\zeta',\lambda',\gamma')' \in \Theta$, and hence the quasi-log-likelihood function is

$$l^{QME}(\theta) = \sum_{t=1}^{T}l_t^{QME}(\theta) = -T\ln\Omega(\lambda,\gamma) - \sum_{t=1}^{T}\sum_{j=1}^{q}\lambda_j\phi_j\left(\frac{y_t - x_t'\zeta}{\sqrt{h_t}},\gamma\right) - \frac{1}{2}\sum_{t=1}^{T}\ln h_t. \quad (12)$$

The scores corresponding to the quasi-log-likelihood for ARCH regression model are

$$\frac{\partial l^{QME}(\theta)}{\partial \alpha} = \sum_{t=1}^{T}\left[\frac{1}{2h_t}\frac{\partial h_t}{\partial \alpha}\left[\sum_{j=1}^{q}\lambda_j\phi_j'(\cdot)\frac{\epsilon_t}{h_t^{1/2}} - 1\right]\right], \quad (13)$$

$$\frac{\partial l^{QME}(\theta)}{\partial \zeta} = \sum_{t=1}^{T}\left[\sum_{j=1}^{q}\lambda_j\phi_j'(\cdot)\frac{x_t'}{h_t^{1/2}} + \frac{1}{2h_t}\frac{\partial h_t}{\partial \zeta}\left[\sum_{j=1}^{q}\lambda_j\phi_j'(\cdot)\frac{\epsilon_t}{h_t^{1/2}} - 1\right]\right], \quad (14)$$

$$\frac{\partial l^{QME}(\theta)}{\partial \lambda_j} = -T\frac{\partial \ln \Omega(\lambda,\gamma)}{\partial \lambda_j} - \sum_{t=1}^{T}\phi_j\left(\frac{y_t - x_t'\zeta}{\sqrt{h_t}},\gamma\right), \quad (15)$$

$$\frac{\partial l^{QME}(\theta)}{\partial \gamma} = -T\frac{\partial \ln \Omega(\lambda,\gamma)}{\partial \gamma} - \sum_{t=1}^{T}\sum_{j=1}^{q}\lambda_j\frac{\partial \phi_j\left(\frac{y_t - x_t'\zeta}{\sqrt{h_t}},\gamma\right)}{\partial \gamma}, \quad (16)$$

where $\phi_j'(\cdot) = \partial\phi_j(\eta,\gamma)/\partial\eta$. The quasi-log-likelihood specification (12) is related to other semi-nonparametric ARCH approaches. In parametric model, the score function is the optimal estimating function (EF) (Godambe, 1960). If underlined conditional density is correctly specified, then Eqs. (13)–(16) are the optimal estimating functions (EFs). Li and Turtle (2000) derived the optimal EFs for ARCH model as

$$\ell_1^* = -\sum_{t=1}^{T}\frac{\frac{\partial h_t}{\partial \alpha}}{h_t^2(\gamma_{2t} + 2 - \gamma_{1t}^2)}g_{2t}, \quad (17)$$

$$\ell_2^* = -\sum_{t=1}^{T}\frac{\frac{\partial x_t'\zeta}{\partial \zeta}}{h_t}g_{1t} + \sum_{t=1}^{T}\frac{h_t^{1/2}\gamma_{1t}\frac{\partial x_t'\zeta}{\partial \zeta} - \frac{\partial h_t}{\partial \zeta}}{h_t^2(\gamma_{2t} + 2 - \gamma_{1t}^2)}g_{2t}, \quad (18)$$

where $g_{1t} = y_t - x_t'\zeta$, $g_{2t} = (y_t - x_t'\zeta)^2 - h_t - \gamma_{1t}h_t^{1/2}(y_t - x_t'\zeta)$, $\gamma_{1t} = \frac{E[(y_t - x_t'\zeta)^3|\mathcal{F}_{t-1}]}{h_t^{3/2}}$, and $\gamma_{2t} = \frac{E[(y_t - x_t'\zeta)^4|\mathcal{F}_{t-1}]}{h_t^2} - 3$. (17) and (18) are actually the same as GMM moment conditions attainable by optimal instrumental variables. There is no *priori* distributional assumption for y_t conditional on $\mathcal{F}_{t-1}$ in the EF approach. Under the conditional normality assumption, $\gamma_{1t} = 0$, and $\gamma_{2t} = 0$, Eqs. (17) and (18) are identical to the first order condition of Engle (1982) [Equation (7), p. 990] up to a sign change. We can relate our approach to robust estimation through the influence function. Let us consider M-estimation that minimizes $\sum_{t=1}^{T}\rho(\eta_t,\theta)$, where $\eta_t = (y_t - x_t'\zeta)/\sqrt{h_t}$, and define the influence function as $-\nu(\eta,\theta)/E[\partial\nu(\eta,\theta)/\partial\eta]$, where $\nu(\eta,\theta) = \partial\rho(\eta,\theta)/\partial\eta$ [see McDonald and Newey (1988)]. If $\rho(\cdot)$ is negative of the natural logarithm of the true density, then we have the ML estimator of θ. Function $\nu(\cdot)$ measures the influence that η will have on the resulting estimators. For the ME density in (11), the function $\nu(\cdot)$ is

$$\nu(\eta,\theta) = \sum_{i=1}^{q}\lambda_i\frac{\partial\phi_i(\eta,\theta)}{\partial\eta}. \quad (19)$$

For $N(0,1)$ density, $\lambda\phi(\eta,\theta) = \frac{1}{2}\eta^2$ and hence $\nu(\eta,\theta) = \eta$ which is unbounded. Li and Turtle (2000) did not assume any particular conditional density and followed a semi-parametric method. In their EFs (17) and (18), the γ_{1t} and γ_{2t} should be specified in some arbitrary way. They noted that since the orthogonality of g_{1t} and g_{2t} holds for any γ_{2t}, an approximate value for γ_{2t} might be used to give near optimal estimating functions l_1^* and l_2^*. If the underlying density is Cauchy, the parameters cannot be estimated consistently by EF approach due to the non-existence of moments. Our MEARCH approach, however, can be used since the moment condition, $E\left[\ln(1+x^2)\right] = 2\ln 2$, generates Cauchy distribution and the associated influence function $\nu(\eta,\theta) = \lambda(2\eta)/(1+\eta^2)$ is bounded. Therefore, careful choice of moment function $\phi(\cdot)$ can lead to robust estimation.

3.1. Estimation

For ARCH-type models, the standardized error term, $\eta_t = \epsilon_t/\sqrt{h_t}$, should have mean 0 and variance 1. However, the MED of η_t given in (11) may not have this property. For convenience, we rewrite (11) as

$$f(\eta_t) = C^{-1}\exp\left[-\sum_{j=1}^{q}\lambda_j\phi_j(\eta_t,\gamma)\right], \quad (20)$$

where C denotes normalizing constant and the parameters vector $\gamma \equiv [\gamma_p' : \gamma_s]'$, where γ_s denotes a scale parameter. Suppose the density (20) is such that $\mathbb{E}(\eta_t) = \mu$ and $\mathbb{V}(\eta_t) = \sigma^2$. If we define $u_t = (\eta_t - \mu)/\sigma$, then $u_t \sim (0,1)$ and $\eta_t = \sigma u_t + \mu$. Due to the transformation $u_t = (\eta_t - \mu)/\sigma$, the density (20) in terms of ϵ_t changes to

$$f_\epsilon(\epsilon_t) = C^{-1}\sigma\frac{1}{\sqrt{h_t}}\exp\left[-\sum_{i=1}^{q}\lambda_i\phi_i\left(\frac{\sigma\epsilon_t}{\sqrt{h_t}} + \mu,\gamma\right)\right]. \quad (21)$$

In (21), the scale parameter γ_s, however, will not separately be identified within an ARCH framework. To make the density free of γ_s, let us define $\tilde{\eta}_t = \eta_t/\gamma_s$, so that $\mathbb{E}(\tilde{\eta}_t) = \mu/\gamma_s = \tilde{\mu}$ and $\mathbb{V}(\tilde{\eta}_t) = \sigma^2/{\gamma_s}^2 = \tilde{\sigma}^2$. The density function of $\tilde{\eta}_t$, $\tilde{f}(\tilde{\eta}_t)$ can be written as

$$\tilde{f}(\tilde{\eta}_t) = \tilde{C}^{-1}\exp\left[-\sum_{i=1}^{q}\lambda_i\phi_i\left(\tilde{\eta}_t,\gamma_p\right)\right]. \quad (22)$$

An "equivalent" density is obtained by substituting $\mu = \gamma_s\tilde{\mu}$ and $\sigma = \gamma_s\tilde{\sigma}$ in the (21) as

$$f_{\epsilon_t}(\epsilon) = C^{-1}\gamma_s\frac{\tilde{\sigma}}{\sqrt{h_t}}\exp\left[-\sum_{i=1}^{q}\lambda_i\phi_i\left(\frac{\gamma_s\tilde{\sigma}\epsilon_t}{\sqrt{h_t}} + \gamma_s\tilde{\mu},\gamma\right)\right] = \tilde{C}^{-1}\frac{\tilde{\sigma}}{\sqrt{h_t}}\exp\left[-\sum_{i=1}^{q}\lambda_i\phi_i\left(\frac{\tilde{\sigma}\epsilon_t}{\sqrt{h_t}} + \tilde{\mu},\gamma_p\right)\right]. \quad (23)$$

The quasi-log-likelihood function corresponding to the density (23) can be written as

$$l(\theta) = \sum_{t=1}^{T}\ln g\left(\frac{\epsilon_t}{h_t^{1/2}}\right) - \frac{1}{2}\sum_{t=1}^{T}\ln h_t \quad (24)$$

$$= -T\ln\tilde{C} + T\ln\tilde{\sigma} - \sum_{t=1}^{T}\sum_{i=1}^{q}\lambda_i\phi_i\left(\frac{\tilde{\sigma}\epsilon_t}{\sqrt{h_t}} + \tilde{\mu},\gamma_p\right) - \frac{1}{2}\sum_{t=1}^{T}\ln h_t, \quad (25)$$

where $\theta = (\alpha',\beta',\zeta',\lambda',\gamma_p')'$ and $g(\cdot)$ is the quasi-density function of u_t given by

$$g(u_t) = \tilde{C}^{-1}\tilde{\sigma}\exp\left[-\sum_{i=1}^{q}\lambda_i\phi_i\left(\tilde{\sigma}u_t + \tilde{\mu},\gamma_p\right)\right]. \quad (26)$$

Since $\phi_i(\cdot)$'s in (25) are not predetermined but are selected from a variety of moment functions to the underlying density, one cannot guarantee $\tilde{C}^{-1}$, $\tilde{\mu}$, and $\tilde{\sigma}$ to have analytic forms. Therefore, in practical applications these are computed using numerical integration.

Following our discussion in Section 2 that the first order conditions for maximizing entropy and the likelihood function are the same under general moment problem, we obtain our parameter estimator by maximizing (25). A range of numerical optimization techniques can be used to maximize (25). We adapted the Broyden, Fletcher, Goldfarb and Shannon (BFGS) algorithm. For computational convenience, the derivatives are computed numerically. We will denote our estimator as $\hat{\theta}_{QMLE}$. Lee and Hansen (1994) and Lumsdaine (1996) showed consistency and asymptotic normality for the QMLE under "conditional normal" GARCH model. Lee and Hansen (1994) established these results under the assumption that u_t is a stationary martingale difference sequence with $\mathbb{E}|u_t|^{\kappa} < \infty$ with some $\kappa \leq 4$. Ling and McAleer (2003) proved consistency and asymptotic normality of QMLE under the second-order moments of the conditional distribution and the finite fourth-order moments of the unconditional distribution of u_t. We assume that our model satisfies these conditions. The limiting distribution of $\hat{\theta}_{QMLE}$ is given by

$$\sqrt{T}\left(\hat{\theta}_{QMLE}-\theta_0\right)\longrightarrow^{d} N\left(0, A_T^{0^{-1}}B_T^0A_T^{0^{-1}}\right),$$

where θ_0 is the quasi-true parameter, $A_T^0 = -T^{-1}E\left(\frac{\partial^2 l(\theta_0)}{\partial\theta\partial\theta'}\right)$, and $B_T^0 = T^{-1}E\left(\frac{\partial l(\theta_0)}{\partial\theta}\frac{\partial l(\theta_0)}{\partial\theta'}\right)$. When our ME density function coincides with the true density, then $\hat{\theta}_{QMLE} \equiv \hat{\theta}_{MLE}$ and we have

$$\sqrt{T}\left(\hat{\theta}_{QMLE}-\theta_0\right)\longrightarrow^{d} N\left(0, B_T^{0^{-1}}\right).$$

Instead of maximizing (25) with respect to *all* the parameters, a computationally less burdensome procedure would be a two-step approach of estimation. In the first-step, using some initial consistent estimates (such as, obtained by maximizing a likelihood function assuming normality) of the conditioned mean and variance parameters, we can obtain $\hat{\eta}_t = \hat{\epsilon}_t/\sqrt{\hat{h}_t}$ and then fit a ME density to $\hat{u}_t = (\hat{\eta}_t - \hat{\mu})/\hat{\sigma}$, using proposed methods in Section 2. In the second-step, fixing the estimated density, $g(\cdot)$ in (24), we can maximize quasi-log-likelihood function with respect to the set of parameter of interest. Engle and González-Rivera (1991) suggested such an approach, where in the first-step, they used a non-parametric method of density estimation. However, based on their simulation results, they noted (p. 350): "When conditional distribution is Student's t, we cannot find any gain. We suspect that this poor performance come from the poor non-parametric estimation of the tails of the density." We can take care of the tail part of distribution by choosing moment functions targeting the tail area of the density. Another problem with the two-step procedure is that for a GARCH model with a general distribution that takes care of asymmetry and excess kurtosis, the underlying information matrix may not be block-diagonal between the conditional mean and variance parameters and the distributional parameters. Therefore, for a such a model, complete adaptive estimation is not possible. Also for this case, the usual standard errors of the parameters estimated by two-step method will not be consistent, as noted by Engle and González-Rivera (1991, p. 352). Therefore, for efficient estimation and valid inference, it is necessary to consider the *joint* estimation of *all* the parameters.

3.2. Moment selection test

As we discussed earlier, Lagrange multipliers provide marginal information of the constraints, and therefore, λ_j should be very close to zero if its associated moment function does not convey any valuable information. Now we derive a statistic for testing $H_{0j}:\lambda_j = 0$ using Rao's score (RS) principle. Detailed derivation is given in the Appendix.

Note that $\tilde{C}$, $\tilde{\sigma}$ and $\tilde{\mu}$ in the log-likelihood function $l(\theta)$ given in (25) are functions of the parameter vector $\theta = (\alpha', \beta', \zeta', \lambda', \gamma_p')'$. The first derivatives of $l(\theta)$ with respect to λ_j is given by

$$d_{\lambda_j} = -T\frac{\partial \ln\tilde{C}(\theta)}{\partial\lambda_j} + T\frac{\partial\ln\tilde{\sigma}(\theta)}{\partial\lambda_j} - \sum_{t=1}^{T}\phi_j\left(\frac{\tilde{\sigma}(\theta)\epsilon_t}{\sqrt{h_t}}+\tilde{\mu}(\theta)\right) - \sum_{t=1}^{T}\Delta_j, \tag{27}$$

where

$$\Delta_j = \sum_{i=1}^{q}\lambda_i\left[\phi_j'\left(\frac{\tilde{\sigma}(\theta)\epsilon_t}{\sqrt{h_t}}+\tilde{\mu}(\theta)\right)\times\left(\frac{\partial\tilde{\sigma}(\theta)}{\partial\lambda_j}\frac{\epsilon_t}{\sqrt{h_t}}+\frac{\partial\tilde{\mu}(\theta)}{\partial\lambda_j}\right)\right],$$

and for notational convenience, from now on the parameter vector γ_p is subassumed within $\phi(\cdot)$.

The score function (27), under $\lambda_j = 0$, reduces to [the details are in Appendix]

$$d^0_{\lambda_j} = T(\varphi_j+\xi_j) - \sum_{t=1}^{T}\phi_j\left(\frac{\omega_v^{1/2}\epsilon_t}{\sqrt{h_t}}+\omega_m\right) - \sum_{t=1}^{T}\tilde{\Delta}_j,$$

where

$$\tilde{\Delta}_j = \sum_{i=\{1,2,\ldots,q\}\backslash\{j\}}\lambda_i\phi_i'\left(\frac{\omega_v^{1/2}\epsilon_t}{\sqrt{h_t}}+\omega_m\right)\times\left(\frac{(\omega_v\omega_j-\omega_{(v,j)})\epsilon_t}{2\sqrt{\omega_v h_t}}+(\omega_m\omega_j-\omega_{(m,j)})\right),$$

and $\omega_m = \mathbb{E}_{\tilde{f}_0}[\tilde{\eta}]$, $\omega_j = \mathbb{E}_{\tilde{f}_0}[\phi_j(\tilde{\eta})]$, $\omega_{(m,j)} = \mathbb{E}_{\tilde{f}_0}[\tilde{\eta}\phi_j(\tilde{\eta})]$, $\omega_v = \mathbb{E}_{\tilde{f}_0}[(\tilde{\eta}-\omega_m)^2]$, $\omega_{(v,j)} = \mathbb{E}_{\tilde{f}_0}[(\tilde{\eta}-\omega_m)^2\phi_j(\tilde{\eta})]$, $\varphi_j = \mathbb{E}_{g_0}[\phi_j(u)]$, and $\xi_j = \mathbb{E}_{g_0}[\tilde{\Delta}_j]$ denoting each subscript represents a particular distribution with which the expectation is taken. Distributions, $\tilde{f}_0(\tilde{\eta})$ and $g_0(u)$ are given by

$$\tilde{f}_0(\tilde{\eta}) = \frac{\exp\left[-\sum_{i=\{1,2,\ldots,q\}\backslash\{j\}}\lambda_i\phi_i(\tilde{\eta})\right]}{\int\exp\left[-\sum_{i=\{1,2,\ldots,q\}\backslash\{j\}}\lambda_i\phi_i(\tilde{\eta})\right]\mathrm{d}\tilde{\eta}}, \tag{28}$$

$$g_0(u) = \frac{\exp\left[-\sum_{i=\{1,2,\ldots,q\}\backslash\{j\}}\lambda_i\phi_i\left(\omega_v^{1/2}u+\omega_m\right)\right]}{\int\exp\left[-\sum_{i=\{1,2,\ldots,q\}\backslash\{j\}}\lambda_i\phi_i\left(\omega_v^{1/2}u+\omega_m\right)\right]\mathrm{d}u}, \tag{29}$$

where $\sum_{i=\{1,2,\ldots,q\}\backslash\{j\}}$ mean summation over $i = 1, 2, \ldots, j-1, j+1, \ldots, q$. We write the score statistic for testing $\lambda_j = 0$ as $R_j(\theta) = d^0_{\lambda_j}/T$, which is given by

$$R_j(\theta) = (\varphi_j+\xi_j) - \frac{1}{T}\sum_{t=1}^{T}\phi_j\left(\frac{\omega_v^{1/2}\epsilon_t}{\sqrt{h_t}}+\omega_m\right) - \frac{1}{T}\sum_{t=1}^{T}\tilde{\Delta}_j.$$

$$= \mathbb{E}_{g_0}\left[\phi_j(u)+\tilde{\Delta}_j\right] - \frac{1}{T}\sum_{t=1}^{T}\left[\phi_j\left(\frac{\omega_v^{1/2}\epsilon_t}{\sqrt{h_t}}+\omega_m\right)+\tilde{\Delta}_j\right].$$

Therefore, the test can be viewed as the difference between population mean relating to the j-th moment function and its sample counterpart. Since φ_j, ξ_j, ω_m, ω_v, and $\tilde{\Delta}_j$ in $R_j(\theta)$ include expectation operator, those will depend on the distributions under the null hypothesis as given in (28) and (29). When $\tilde{f}_0(\tilde{\eta})$ is

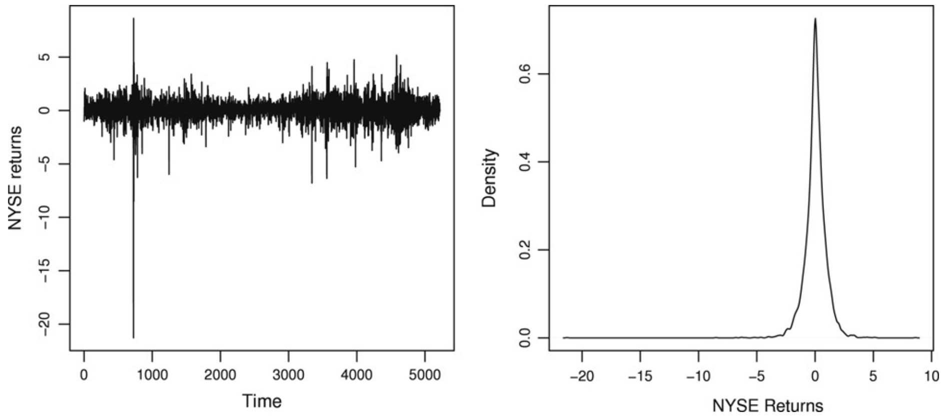

Fig. 5. NYSE return data and non-parametric kernel density. Notes: Usual Gaussian kernel is used for estimating non-parametric density in which rule-of-thumb bandwidth (Silverman, 1984) is 0.1035.

symmetric around 0, then $\mathbb{E}_{\tilde{f}_0}[\phi_i(\tilde{\eta})] = 0$ if $\phi_i(\tilde{\eta})$ is an odd function and has point symmetric property at vertex 0. For an even function the expectation is not zero. Examples of odd and point symmetric functions are $\tan^{-1}(\tilde{\eta})$, $\tilde{\eta}/(1+\tilde{\eta}^2)$, $\sinh^{-1}(\tilde{\eta})$, and $\sin(\tilde{\eta})$, while $\ln(1+\tilde{\eta}^2)$, $\ln(1+|\tilde{\eta}|^p)$, $\tan^{-1}(\tilde{\eta}^2)$, and $\cos(\tilde{\eta})$ are even functions. Premaratne and Bera (2005) developed a test of the form $R_j(\theta)$ for testing asymmetry under heavy tails distribution. They used Pearson type-IV density function under the alternative hypothesis. Thus, under the null, their distribution becomes Pearson type-VII which is symmetric around 0 and also includes Student's t as a special case. It can be easily checked that their $R_j(\theta) = -\frac{1}{T}\sum_{t=1}^{T}\tan^{-1}(\eta_t/r)$ and $\mathbb{E}[\tan^{-1}(\eta_t/r)] = 0$ under symmetry.

An operational form of Rao's score (RS) statistic would be

$$RS_j = T \cdot \frac{R_j^2(\hat{\theta})}{\hat{V}},$$

where $\hat{\theta}$ is the MLE of θ and $\hat{V}$ is an consistent estimator of asymptotic variance of $\sqrt{T}\cdot R_j(\hat{\theta})$. Under the null hypothesis, $H_0 : \lambda_j = 0$, RS_j is asymptotically distributed as χ_1^2. We obtain $\hat{V}$ by bootstrap approach, and the bootstrap score test statistic is given by

$$RS_{jB} = T \cdot \frac{R_j^2(\hat{\theta})}{\hat{V}_B}, \tag{30}$$

where $\hat{V}_B$ denotes the variance of $R_j(\hat{\theta})$ calculated by bootstrap method. Under the null hypothesis, as $B \to \infty$, RS_{jB} is asymptotically distributed as χ_1^2. For finite B, RS_{jB} is asymptotically distributed as Hotelling's T^2 with $(1, B-1)$ degrees of freedom, in short $T^2_{1,B-1}$ [see Dhaene and Hoorelbeke (2004)].

4. Empirical application

To illustrate the suitability of our methodology to financial data, we consider an empirical application of MEARCH model using the daily prices of NYSE, from Jan.1, 1985 to Dec. 30, 2004, a total of 5218 observations obtained from the Datastream. To achieve stationarity, we transform the indices prices into returns, $r_t = [\ln S_t/S_{t-1}] \times 100$, where S_t is the price index at time t. The returns data and a corresponding non-parametric density are plotted in Fig. 5. The data plot clearly shows that there is a high degree of clustering (conditional heteroskedasticity) and the estimated non-parametric density indicates high degree of non-normality with thick tails and high peakedness. The sample kurtosis, skewness and Jarque and Bera (JB) normality test statistics are 55.89, −2.43 and 613,269, respectively, and indicate not only high excess kurtosis but also distinct negative skewness. Ljung-Box (LB) test statistics for residuals from an AR(1) model at lags 12 days using the series (Q) and its squares (Q^2), cubes (Q^3) and fourth-power (Q^4) are 22.87, 424.93, 42.31 and 7.20, respectively. It appears that AR(1) model can take account of part of autocorrelation in the data. Very high values of Q^2 indicate non-linear dependence and strong presence of conditioned heteroskedasticity. The Q^3 and Q^4 statistics measure higher order dependence and some changes in the third and fourth moments over time but these changes are not as strong as for the time varying second moment as evident from the high values of Q^2. To explain such behaviors of stock return data, we need to consider a model which captures distributional characteristics and dynamic moment structure simultaneously.

For the testing and selecting different moment functions, we start with two separate ME densities as distributions under the respective null hypothesis. The first density corresponds to the moment function $\ln(1+\tilde{\eta}^2) = \ln(1+(\eta/r)^2)$ and as noted earlier, this is the Pearson type VII distribution which includes Student's t as a special case. The second density is implied by the moment function $\ln(1+|\tilde{\eta}|^p)$ and reduces to the first one when $p = 2$. Since the results of our tests based on statistic (30) with different bootstrap sample sizes $B = 50$, 100, 150 and 200 are similar, we report the results only for $B = 100$ (Table 2). When the null density comes from the moment function $\ln(1+\tilde{\eta}^2)$, the Lagrange multipliers corresponding to $\tan^{-1}(\tilde{\eta}^2)$, $\sinh^{-1}(\tilde{\eta})$ and $\tan^{-1}(\tilde{\eta})$ are all highly significant. As noted earlier, $\tan^{-1}(\tilde{\eta})$ represents high peakedness, and $\sinh^{-1}(\tilde{\eta})$ and $\tan^{-1}(\tilde{\eta})$ capture asymmetry. Our test results indicate that these three moment functions take account of certain data characteristics and convey some additional information after we have already started with the moment function $\ln(1+\tilde{\eta}^2)$, i.e., the Pearson type VII density. On the other hand, none of the additional Lagrange multipliers are significant when we test the null density implied by the moment function $\ln(1+|\tilde{\eta}|^p)$ (with maximum likelihood estimate, $\hat{p} = 1.3978$). This function behaves like a "sufficient" moment function in the sense that once we start with $\ln(1+|\tilde{\eta}|^p)$, the additional moment functions do not throw any further light on the underlying density. It is as if $\ln(1+|\tilde{\eta}|^p)$ exhausts all the available information (in the sample) that is relevant for estimating the density function.

We now use these test results and estimate the AR(1)-GARCH(1, 1) model with various ME densities. We consider the three moment functions one-by-one (in addition to $\ln(1+\tilde{\eta}^2)$) for which the associated Lagrange multipliers were significant (see Table 2). The estimation results are reported in Table 3. It is clear that additional moment functions increase the log-likelihood value substantially and make the model selection criteria AIC and SIC values more attractive. We should add that the moment function $\ln(1+|\tilde{\eta}|^p)$ where p appears as an additional parameter, by itself performs

Table 2
Moment function selection test results with bootstrap sample size $B = 100$.

	$\cos(\tilde\eta)$	$\tilde\eta/(1+\tilde\eta^2)$	$\sin(\tilde\eta)$	$\tan^{-1}(\tilde\eta^2)$	$\sinh^{-1}(\tilde\eta)$	$\tan^{-1}(\tilde\eta)$
(i) $\ln(1+\tilde\eta^2)$	1.519	2.591	0.475	18.883*	8.773*	7.823*
	(0.218)	(0.108)	(0.491)	(0.000)	(0.003)	(0.005)
(ii) $\ln(1+\lvert\tilde\eta\rvert^p)$	1.708	0.596	0.762	0.151	0.831	0.773
	(0.191)	(0.440)	(0.383)	(0.698)	(0.362)	(0.379)

Notes: (i) and (ii) denote null density corresponds to the moment function $\ln(1+\tilde\eta^2)$ and $\ln(1+|\tilde\eta|^p)$, respectively. P-values are given in the parentheses and calculated using asymptotic χ^2_1 distribution. 1% critical values of Hotelling's T^2 are $T^2_{1,49}$: 7.181, $T^2_{1,99}$: 6.898, $T^2_{1,149}$: 6.808, $T^2_{1,199}$: 6.764., respectively.

* Indicates statistical significance at the 1% level.

Table 3
Estimation with different moment functions.

	(i) $\ln(1+\tilde\eta^2)$	(i) & $\tan^{-1}(\tilde\eta^2)$	(i) & $\sinh^{-1}(\tilde\eta)$	(i) & $\tan^{-1}(\tilde\eta)$	Model 1	Model 2	Model 3	Model 4
	AR(1)							
ζ_0	0.0608	0.0591	0.0484	0.0498	0.0528	0.0472	0.0474	0.0458
	(0.0089)	(0.0088)	(0.0098)	(0.0098)	(0.0091)	(0.0096)	(0.0095)	(0.0097)
ζ_1	0.0428	0.0422	0.0395	0.0396	0.0355	0.0392	0.0396	0.0388
	(0.0129)	(0.0126)	(0.0130)	(0.0130)	(0.0120)	(0.0127)	(0.0127)	(0.0126)
	GARCH(1, 1)							
α_0	0.0064	0.0066	0.0065	0.0065	0.0068	0.0066	0.0066	0.0073
	(0.0021)	(0.0021)	(0.0020)	(0.0020)	(0.0022)	(0.0020)	(0.0020)	(0.0024)
α_1	0.0577	0.0571	0.0571	0.0572	0.0588	0.0580	0.0580	0.0651
	(0.0087)	(0.0085)	(0.0083)	(0.0084)	(0.0091)	(0.0081)	(0.0080)	(0.0092)
β_1	0.9366	0.9359	0.9365	0.9364	0.9340	0.9339	0.9340	0.9314
	(0.0094)	(0.0094)	(0.0092)	(0.0092)	(0.0100)	(0.0091)	(0.0091)	(0.0097)
	Lagrange multipliers (λ_j's)							
$\ln(1+\lvert\tilde\eta\rvert^p)$					9.1912			
					(4.0198)			
$\ln(1+\tilde\eta^2)$	3.0040	2.9474	3.1284	3.1012		3.3108	3.3193	2.4649
	(0.2076)	(0.1992)	(0.2213)	(0.2200)		(0.3975)	(0.3955)	(0.1877)
$\tan^{-1}(\tilde\eta^2)$		−1.0630				−1.2067	−1.2172	−0.9024
		(0.1249)				(0.2064)	(0.2085)	(0.1586)
$\cos(\tilde\eta)$						0.6327	0.6248	0.8239
						(0.2735)	(0.2712)	(0.0978)
$\tan^{-1}(\tilde\eta)$				0.3587			0.3019	−3.3456
				(0.1265)			(0.1870)	(0.7318)
$\sin(\tilde\eta)$						−0.4075	−0.4640	
						(0.1310)	(0.1600)	
$\sinh^{-1}(\tilde\eta)$			0.2783			0.1785		2.1510
			(0.0889)			(0.1155)		(0.4830)
$\tilde\eta/(1+\tilde\eta^2)$								1.7340
								(0.3539)
p					1.3978			
					(0.1001)			
log-likelihood	−6084.15	−6073.51	−6079.49	−6080.37	−6067.36	−6059.33	−6059.32	−6041.94
AIC	2.3343	2.3306	2.3329	2.3332	2.3282	2.3263	2.3263	2.3200
SIC	2.3418	2.3394	2.3417	2.3420	2.3370	2.3389	2.3389	2.3339

Note: Standard errors are given in the parentheses.

extremely well. Also, as we discussed in Section 2.1, $\hat p = 1.3978 < 2$ captures the peakedness of the distribution. This encourages us to test various combinations of moment functions and estimate models with different sets of moment functions. To conserve space we do not report all the test and estimation results but these can be obtained from us on request. In the right panel of Table 3, we present results from four models under several combinations of moment functions for which the Lagrange multipliers were significant. The moment functions used in these models are quite clear from the lower part of the Table 3; for example, the Model 1 corresponds to moment function $\ln(1+|\tilde\eta|^p)$.

Model 4, apparently the "best" model, includes six moment functions for which all the Lagrange multipliers are highly significant. Performance of Models 2 and 3 are almost identical as the moment functions $\tan^{-1}(\cdot)$ and $\sinh^{-1}(\cdot)$ have similar shape (see Fig. 4), and as we shall also see in Figs. 6 and 7. Using our earlier discussion it is tempting to say that $\ln(1+\tilde\eta^2)$ exclusively explains excess kurtosis, $\tan^{-1}(\tilde\eta^2)$ captures high peakedness, and $\tan^{-1}(\tilde\eta)$ and other functions take care of asymmetry, etc. However, these moment functions are not orthogonal and therefore, when many are present in a single ME density, we need to consider their combined effect.

It is interesting to compare above estimation results to those of GARCH models based on some other general parametric density functions used in the current literature: standard normal; Student's t (Bollerslev, 1987); skewed-t (Fernández and Steel, 1998; Lambert and Laurent, 2001); and inverse hyperbolic sine (IHS) [Hansen et al. (2000)]. The values of log-likelihood functions and model selection criteria (AIC and SIC) for those models are reported in Table 4. We note that, in terms of goodness-of-fit, Model 4 achieves the levels of some of the very general standard distributions quite easily.

In Figs. 6 and 7 we plot, respectively, the conditional densities, and the influence functions $\nu(\cdot)$, computed using the formula (19) for our four models (presented in Table 3). Density corresponding to Model 4 is very close to the non-parametric density based on

Fig. 6. Density estimates for the standarized residuals of the final models. Notes: QMLE denotes usual Gaussian kernel density using Scott's (1992) optimal bandwidth (0.1534) for standardized residual from the estimated GARCH model under conditional normality.

Fig. 7. Influence functions for the final models.

Table 4
Goodness-of-fit for four densities.

	Normal	Student's *t*	Skewed-*t*	IHS
Log-Likelihood	−6334.76	−6085.46	−6082.82	−6079.05
AIC	2.4300	2.3348	2.3342	2.3317
SIC	2.4362	2.3423	2.3430	2.3415

the standardized residuals of the estimated GARCH model under conditional normality (QMLE). All the four influence functions are bounded, and as expected it is hard to distinguish the lines for Models 2 and 3. The influence function corresponding to the Model 4 has the least variation and comes out to be the best. Thus, after a series of estimations and tests, our maximum entropy approach leads to a model that captures stylized facts quite effectively.

5. Concluding remarks

In this paper, we provide a generalization of GARCH model by incorporating MED as the underlying probability distribution. We characterize MED and discuss various moment functions that are suitable to capture excess kurtosis, asymmetry and high peakedness generally observed in financial data. We devise a test to select appropriate moment functions. Our empirical results demonstrate that the suggested MEARCH model is quite useful in capturing the behavior of the data. Many other moment functions and their mixtures could be chosen to generate even more flexible density. Our procedure is quite different from those that use certain non-normal densities. Those densities have fixed forms and not amenable to easy modification. Ours is a completely flexible procedure where various moment functions are selected based on the information available from the data. The approach is also constructive than a (semi-) non-parametric one using orthonormal series in the sense that the ME model provides a highly parsimonious model. The extension to the multivariate MEARCH model is of particular interest since many empirical works deal with the multivariate GARCH models such as Bollerslev's (1990) constant conditional correlation and Engle's (2002) dynamic conditional correlation models [for a review