A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

November 2009

EPL, **88** (2009) 30007
www.epljournal.org
doi: 10.1209/0295-5075/88/30007

Non-extensitivity vs. informative moments for financial models —A unifying framework and empirical results

K. HERRMANN[(a)]

School of Business and Economics, University of Erlangen-Nürnberg
Lange Gasse 20, 90403 Nürnberg, Germany, EU

received 5 June 2009; accepted in final form 20 October 2009
published online 16 November 2009

PACS 05.45.Tp – Time series analysis
PACS 05.90.+m – Other topics in statistical physics, thermodynamics, and nonlinear dynamical systems
PACS 89.65.Gh – Economics; econophysics, financial markets, business and management

Abstract – Information-theoretic approaches still play a minor role in financial market analysis. Nonetheless, there have been two very similar approaches evolving during the last years, one in the so-called econophysics and the other in econometrics. Both generalize the notion of GARCH processes in an information-theoretic sense and are able to capture kurtosis better than traditional models. In this article we present both approaches in a more general framework. The latter allows the derivation of a wide range of new models. We choose a third model using an entropy measure suggested by Kapur. In an application to financial market data, we find that all considered models – with similar flexibility in terms of skewness and kurtosis – lead to very similar results.

Introduction. – Although information-theoretic approaches still play only a minor role in financial market analysis, there have been two recent approaches developing independently, one in econophysics and the other in econometrics. Both approaches generalize the notion of ARCH or GARCH models.

Furthermore, both approaches have successfully been applied to financial market data – possibly due to the fact that both generalizations allow to capture kurtosis in given variance models. Following Kesavan and Kapur [1] we give a more general framework for information-theoretic models for time-varying moments and derive GARCH models as well as the above-mentioned approaches as special cases. We briefly present the recent information-theoretic approaches and – using the general framework – derive a third approach based on a entropy suggested by Kapur. In an application to financial market data we compare these models' capability to capture the behavior of financial returns. We find very similar results for all models.

Models for time-varying moments based on the generalized principle of maximum entropy. – We consider time-series models that assume the conditional distribution of some random variable X at time t given $\mathcal{F}_t$, the information set available at time $t-1$, to be the distribution maximizing some generalized entropy measure

$$H(f) = -\int_{D(X)} \phi(f(x))\,\mathrm{d}x, \tag{1}$$

for some convex function[1] ϕ and $D(X)$ the random variables support, subject to constraints of $k+1$ conditional expectation values of suitable functions $g_i(\cdot)$ as

$$E(g_0(X_t)\,|\mathcal{F}_t) = 1, \quad E\,(g_i(X_t)\,|\mathcal{F}_t) = a_i(\mathcal{F}_t), \tag{2}$$

$\forall\, i = 2,\ldots,k$, with $g_0(x) = x$, where a_i may be some deterministic functions depending only on the information set available.

The information-theoretic interpretation of such models is that we model only some expectation values' motion in time and for all information missing to completely determine the corresponding density functions we maximize entropy. In the case of the Shannon entropy measure such a way of modeling is nothing else but the consequent application of Jaynes' Principle of Maximum Entropy [2]. Using generalized entropy measures is justified by Kesavan and Kapur's generalized maximum entropy principle [1].

Following their suggestion, we restrict ϕ to the set of differentiable convex functions. A variational approach

(a) E-mail: klaus.herrmann@wiso.uni-erlangen.de

[1] Here we follow the suggestion in [1].

shows that, under some weak conditions for ϕ, it holds for the maximum entropy density (if it exists) that

$$\phi'(f(x)) = \sum_{i=0}^{k} \lambda_i g_i(x), \tag{3}$$

where the λ_i have to be chosen such that the constraints are fullfilled[2].

Numerical implementation. – If a solution exists for f, $\lambda = (\lambda_1, \ldots, \lambda_k)$ from eq. (3) can be found by minimizing a convex function Z of some dual problem with

$$\max_f(H(f)) = \min_\lambda(Z(\lambda)). \tag{4}$$

Such dual problems are given, *e.g.*, in [3]. For the cases relevant to this letter we use

$$Z_S(\lambda) = \sum_{i=1}^{k} a_i\lambda_i + \int_{-\infty}^{\infty} \exp\left(\sum_{i=1}^{k} \lambda_i g_i(x)\right) \mathrm{d}x, \tag{5}$$

$$Z_T(\lambda) = \frac{2-q}{q-1}\sum_{i=1}^{k} a_i\lambda_i + \int_{-\infty}^{\infty} \left(\sum_{i=1}^{k} \lambda_i g_i(x)\,\mathrm{d}x\right)^{\frac{2-q}{1-q}} \mathrm{d}x, \tag{6}$$

$$Z_K(\lambda) = \sum_{i=1}^{k} a_i\lambda_i - \frac{1}{c}\int_{-\infty}^{\infty} \ln\left(\exp\left(\sum_{i=1}^{k} \lambda_i g_i(x)\right) - c\right) + \sum_{i=1}^{k} \lambda_i g_i(x)\,\mathrm{d}x, \tag{7}$$

where Z_S, Z_T and Z_K denote the dual problems for the Shannon, Tsallis and Kapur entropy, respectively. The algorithm for the numerical implementation given in [4] may be extended to any information measure. It relies for the integration involved a Gauss-Legendre approach and for the minimization gradient-based optimization[3]. Using this, we may write, *e.g.*, Z_S as

$$Z_S(\lambda) = \sum_{i=1}^{k} a_i\lambda_i + \sum_{j=1}^{n} \exp\left(\sum_{i=1}^{k} \lambda_i g_i(x_j)\right) w_j, \tag{8}$$

where w_j denotes the Gauss-Legendre weights, n its number and x_j the corresponding transformed abscissa, with

$$x_j = [(u-l)z_j + (u+l)]/z_j \tag{9}$$

where u and l denote the upper and lower limit for the support to be used for the integration[4] and z_j the abscissa in $[-1, 1]$. Using this integration scheme, we may derive λ by iterating k with

$$\lambda^k = \lambda^{k-1} + \delta^k, \tag{10}$$

and δ^k as the solution of

$$\frac{\partial^2 Z(\lambda)}{\partial\lambda^2}\delta^k = -\frac{\partial Z(\lambda)}{\partial\lambda}. \tag{11}$$

This algorithm proved to be very efficient for all dual problems given in [3], where sufficient convergence was achieved after about 14 iterations.

GARCH models. – Bollerslev's original GARCH (p,q) model in [6] may be written as

$$X_t|\mathcal{F}_t = Z_t \cdot \sigma_t, \quad Z_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{12}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_{1,i} x_{t-i}^2 + \sum_{i=1}^{q} \alpha_{2,i}\sigma_{t-i}^2, \tag{13}$$

where X_t is the random variable at time t and x_t its realization.

We may present this model in the above framework, if we derive the conditional density functions of X_t by maximizing *e.g.*, Shannon's entropy measure

$$H_S(f) = -\int_{D(X)} f(x)\ln(f(x))\,\mathrm{d}x \tag{14}$$

subject to the constraints

$$E(X_t|\mathcal{F}_t) = \mu, \qquad E((X_t-\mu)^2|\mathcal{F}_t) = \sigma_t^2, \tag{15}$$

assuming again eq. (13) for σ_t. The distribution maximizing eq. (14) subject to eq. (15) is again the normal distribution.

But it is a well known fact that GARCH models assuming Gaussian innovations do not sufficiently describe financial market data. Apart from a vast literature concerned with that topic[5], this can be seen by looking at the GARCH-filtered innovations. Their distribution should be close to a Gaussian, but it is regularly found that this assumption does not hold.

The observation that financial market GARCH innovations deviate from the normal distribution mainly in terms of skewness and higher kurtosis has been frequently found in several studies and is now known to econometricians as a so-called stylized fact. This observation gave rise to a large number of approaches which try to model these phenomena by assuming z_t to follow some flexible parametric distribution, such as, *e.g.*, the SGT2, EGB2 or the family of stable distributions[6].

This note, of course, is devoted to information-theoretic approaches that either explicitly model such features by including measures of skewness and kurtosis into the entropy maximization task or, as the non-extensive approaches do, implicitly by generalizing the entropy measure.

[2]Compare, *e.g.*, [3].
[3]The corresponding implementation in *R* will be delivered to the interested reader by the author on request.
[4]Problems potentially arising by this restriction are discussed in [5].
[5]See, *e.g.*, [7] for an overview.
[6]Compare, *e.g.*, [8–10]. Some of the parametric approaches may also find some representation in the above framework – but because of the vast supply of such suggestions, we restrict this letter to the most basic case.

Table 1: Suggestions for measures of asymmetry m_3 and measures of kurtosis m_4 as proposed in the recent literature.

m_3	m_4
$E(\tan^{-1}(X))$	$E(\tan^{-1}(X)^2)$
$E\left(\frac{X}{1+X^2}\right)$	$E(\ln(1+X^2))$

Higher informative moments. – To the best of our knowlegde, the first approach of including higher moments in an information-theoretic framework for financial models was given in [4]. Extensions have been given in [5,11,12]. The idea behind these approaches is to interpret stylized facts as information given to the analyst, such that it consequently has to be included in the entropy maximization task. Formally, these econometric approaches assume for the conditional distributions of X_t given the information set available at $t-1$ the distribution maximizing the Shannon entropy subject to

$$E\left(X_t|\mathcal{F}_t\right)=\mu, \quad E\left((X_t-\mu)^2|\mathcal{F}_t\right)=\sigma_t^2, \tag{16}$$

$$E\left(g_3\left(\frac{X_t-\mu}{\sigma_t}\right)|\mathcal{F}_t\right)=m_3, \tag{17}$$

$$E\left(g_4\left(\frac{X_t-\mu}{\sigma_t}\right)|\mathcal{F}_t\right)=m_4. \tag{18}$$

The approach in [4] uses third and fourth power moments to measure skewness and kurtosis. But as [5] point out, these measures do not allow the derivation of proper maximum Shannon entropy densities for kurtosis values higher than implied by the lower moments. Suggestions for more appropriate measures of the form $m_i=E(g_i(X))$ can be found in [11] or [5], as given in table 1, where m_3 denotes measures of asymmetry and m_4 denotes measures of kurtosis.

Non-extensive approach. – The information-theoretic models in econophysics have been developed mainly in [13–16]. Contrary to the econometric models, these approaches do not explicitly aim at modeling stylized facts, but they rather try to generalize the entropy measure to capture situations where the Shannon entropy does not reflect the *nature of the system*, compare [14]. Consequently these approaches generalize GARCH models in the above-discussed form by assuming the conditional density functions to maximize the generalized entropy measure suggested by Tsallis [17][7],

$$H_T=\int_{D(X)}\frac{1-f(x)^q}{q-1}\,\mathrm{d}x, \tag{19}$$

$$\text{s.t. } E\left(X_t|\mathcal{F}_t\right)=\mu, \quad E\left((X_t-\mu)^2|\mathcal{F}_t\right)=\sigma_t^2, \tag{20}$$

with different suggestions for σ_t's motion in time.

[7]Tsallis suggestion is also known as the entropy of Havrda/Charvat. It is non-additive and non-extensive for $q\neq 1$, compare, *e.g.*, [18], and Shannon's entropy appears as the limiting case for $q\to 1$.

The use of Tsallis entropy may be justified from a theoretic perspective by the generalized Khinchin conditions, compare [19], or, for the application to financial returns, by its ability to allow for long-range interactions, compare [14]. From a mere statistical point of view, the use of this entropy measure would rather be justified by its ability to model kurtosis, not as information explicitly included to the constraints, but by flexibly varying q. The conditional distribution of the model is given by

$$f_{ME,T}=\left(\sum_{i=0}^{k}\lambda_i g_i(x)\right)^{\frac{1}{1-q}}=\left(\lambda_0+\lambda_1 x+\lambda_2(x-\mu)^2\right)^{\frac{1}{1-q}}, \tag{21}$$

and can, for $q>1$, be interpreted as some generalized t-distribution, where q drives the degrees of freedom[8]. For $q>1$, a higher deviation from extensitivity ($q=1$) implies higher kurtosis.

A Kapur entropy-based approach. – From the presented framework we may easily derive a vast range of equally flexible models by choosing appropriate measures of entropy. As an example, we suggest a third approach, where we proceed as above, but maximize the non-additive entropy measure suggested by Kapur [3],

$$H_K(f) = \int_{D(X)}\left(-f(x)\ln(f(x))+\frac{1}{c}\,\mathrm{d}(x)\ln(\mathrm{d}(x))\right)\mathrm{d}x, \quad \text{with } \mathrm{d}(x)=1+c\cdot f(x), \tag{22}$$

where Shannon's entropy appears for $c=0$. This entropy measure can be related to the Fermi-Dirac statistics for $c=-1$ or to the Bose-Einstein statistics for $c=1$, compare [21]. The corresponding density function has the form [3]

$$f_{ME,K}(x)=\left(\exp\left(\sum_{i=0}^{k}\lambda_i g_i(x)\right)-c\right)^{-1}. \tag{23}$$

Its application here shall be motivated only from a statistician's perspective by its ability to model higher kurtosis by increasing $c>0$.

Application to financial market data. – We compare the above-presented approaches, using specifications with equivalent flexibility, in an application to three different time-series typical for financial markets. We will compare results by likelihood and likelihood-based goodness-of-fit measures, by the distance of their empirical innovations' distribution to the theoretical model and by their capability to explain their empirical quantiles in the tails.

Models. For all models we use a constant mean and time-varying variance as

$$E\left(X_t|\mathcal{F}_t\right)=\mu, \qquad E\left((X_t-\mu)^2|\mathcal{F}_t\right)=\sigma_{t^2}, \tag{24}$$

[8]For $q<1$ the distribution may be related to the r-distribution, see [20].

Table 2: Descriptive statistics for the illustrative data sets.

	Gold	DJIA	EurUS
Mean	−1.363e-03	2.316e-05	−1.657e-04
Stan. Dev.	3.479e-02	1.286e-02	6.499e-03
$\hat{m}_3$	−1.596e-01	−5.802e-02	7.545e-02
$\hat{m}_4$	6.961	14.301	7.254
Observations	1606	1603	1632

where we assume for σ_t's motion in time

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-i}^2 + \alpha_2 \sigma_{t-i}^2, \qquad (25)$$

as there is empirical evidence that $p=1$ and $q=1$ sufficiently describe financial returns behavior, compare, *e.g.*, [22].

For the inclusion of stylized facts we model skewness as

$$E\left(\tan^{-1}\left(\frac{X_t-\mu}{\sigma_t}\right)\bigg|\mathcal{F}_t\right) = \beta_0, \qquad (26)$$

and for the Shannon model (E) kurtosis as

$$E\left(\ln\left(1+\left(\frac{X_t-\mu}{\sigma_t}\right)^2\right)\bigg|\mathcal{F}_t\right) = \gamma_0, \qquad (27)$$

since in some previous study these moments performed best. For the Tsallis model (N) and the Kapur model (K) kurtosis is modeled by the flexibility of the parameters of the corresponding entropy measures q and c. For all models we assume skewness and kurtosis to be constant over time.

Data. In order to sample different kinds of financial market indices, we chose the daily returns between January 1st, 2003 and March 20th, 2009 for the gold price, the Dow Jones Industrial Average, both from `yahoo.finance.com`, and for the Euro-US-Dollar exchange rate from `www.ecb.int`. Some descriptive statistics for the data are given in table 2 (see footnote [9]).

Empirical results. We use numerical optimization routines to implement maximum likelihood estimation of the model parameters. Estimates and standard errors (in brackets) are given in table 3.

We find for all models similar estimates for variance motion in time as well as similar values for skewness. The kurtosis parameters differ of course, as all models have their own way to capture kurtosis. For all models and all data sets we find significant non-normal kurtosis[10].

Table 4 gives an overview over the model fit, where LogL denotes log-likelihood, AIC the Akaike information criterion, BIC the Bayesian information criterion, KS the Kolgomorov-Smirnov distance and χ^2 the χ^2 test statistic assuming 10 classes.

Table 3: Estimates for the applied models.

	$\widehat{\alpha_0}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{\beta_0}$	$\widehat{\gamma_0}$ / $\widehat{q}$ / $\widehat{c}$
E	7.63e-06	0.0410	0.952	0.00409	0.508
	(4.16e-06)	(0.0087)	(0.010)	(0.00393)	(0.005)
N	7.03e-06	0.0394	0.955	0.00537	1.240
	(3.96e-06)	(0.0082)	(0.009)	(0.00385)	(0.058)
K	7.68e-06	0.0416	0.952	0.00520	3.883
	(4.16e-06)	(0.0089)	(0.010)	(0.00375)	(1.011)
Gold price					
E	7.08e-07	0.0707	0.923	−4.0e-05	0.509
	(3.14e-07)	(0.0115)	(0.012)	(0.00399)	(0.005)
N	7.38e-07	0.0702	0.923	−0.00114	1.221
	(3.16e-07)	(0.0114)	(0.011)	(0.00393)	(0.061)
K	7.31e-07	0.0702	0.923	−0.00284	4.071
	(3.17e-07)	(0.0115)	(0.012)	(0.00384)	(1.083)
Dow Jones industrial average					
E	1.47e-07	0.0337	0.963	−0.00144	0.509
	(9.24e-08)	(0.0070)	(0.007)	(0.00386)	(0.005)
N	1.61e-07	0.0313	0.964	−0.00272	1.220
	(9.02e-08)	(0.0064)	(0.007)	(0.00379)	(0.056)
K	1.45e-07	0.0343	0.962	−0.00236	3.640
	(9.27e-08)	(0.0068)	(0.007)	(0.00369)	(0.991)
Euro-USD exchange rate					

Table 4: Goodness-of-fit measures for the applied models.

	LogL	AIC	BIC	KS	χ^2
E	3283.85	−6557.69	−6530.80	0.5238	4.666
N	3284.06	−6558.13	−6531.23	0.5626	4.396
K	3283.85	−6557.70	−6530.81	0.4707	4.037
Gold price					
E	5255.65	−10521.29	−10474.39	1.6563	23.776
N	5252.58	−10515.16	−10468.26	1.7200	29.956
K	5255.70	−10521.40	−10474.50	1.4537	21.874
Dow Jones industrial average					
E	6076.47	−12162.94	−12115.95	0.6817	5.747
N	6076.81	−12163.62	−12116.63	0.7017	9.754
K	6075.93	−12161.86	−12116.87	0.6125	5.217
Euro-USD exchange rate					

All models behave well – except for the Dow Jones Industrial Average, where KS and χ^2 test would reject all suggested models. In the sense of likelihood-based measures we cannot derive a uniformly "best" model. The Kapur model outperforms all other models if we use only KS and χ^2 as criteria.

In a last step, we compare the model fit in terms of tails. Therefore we compare some extreme theoretical quantiles with their sample anologa. Table 5 gives the differences for the standardized estimates. We find that all models similarly under- or over-estimate the most extreme quantiles.

[9] $\hat{m}_3$ ($\hat{m}_4$) denotes the empirical third (fourth) standardized power moment.
[10] The value for m_4 (used in the "E" model) implied by the normal distribution is about 0.53345.

Table 5: Absolute difference between theoretical quantiles and sample estimates.

	0.5%	1%	5%	95%	99%	99.5%
E	−0.074	0.060	0.012	−0.037	0.002	−0.052
N	−0.026	0.042	−0.003	0.005	0.023	−0.060
K	−0.079	0.067	0.024	−0.036	−0.003	−0.057
			Gold price			
E	0.078	−0.046	−0.118	−0.067	−0.214	−0.115
N	0.082	−0.079	−0.146	−0.049	−0.203	−0.146
K	0.061	−0.058	−0.125	−0.091	−0.255	−0.158
			Dow Jones industrial average			
E	−0.151	0.087	0.016	−0.022	−0.074	−0.102
N	−0.134	0.061	−0.024	0.001	−0.076	−0.119
K	−0.159	0.083	0.019	−0.034	−0.083	−0.119
			Euro-USD exchange rate			

We conclude that none of the models clearly outperforms the other ones. This is not surprising, as all models have been chosen such that they exhibit the same flexibility needed to model known stylized facts. This result may thus be explained by the fact, that the more information we explicitly introduce into the model, the less information is introduced by the choice of entropy.

Summary. – The information-theoretic approaches to time-series models considered in econometrics and econophysics may both be interpreted as special cases of models for time-varying moments using the generalized maximum entropy principle. Using this interpretation we may derive further models. For illustration we introduce a new model based on a suggestion by Kapur. In application to three illustrative data sets typical for financial markets, we find that all suggested models from this class exhibit similar flexibility in capturing the behavior of financial returns where neither the econometric's nor the econophysician's approach gives superior results in a statistical sense. The explanation for this observation may be that the more information we include into the model, the less the choice of entropy matters.

REFERENCES

[1] Kesavan H. K. and Kapur J. N., *IEEE Trans. Syst. Man Cybern.*, **19** (1989) 1042.
[2] Jaynes E. T., *Phy. Rev.*, **106** (1957) 620.
[3] Kapur J. N., *Measures of Information and their Applications* (Wiley Eastern Limited, New Delhi) 1994.
[4] Rockinger M. and Jondeau E., *J. Econom.*, **106** (2002) 119.
[5] Fischer M. and Herrmann K., *Discussion Papers of the Chair of Statistics and Econometrics Erlangen-Nürnberg*, **84** (2009).
[6] Bollerslev T., *J. Econom.*, **31** (1986) 307.
[7] Hansen B. E., *Int. Econ. Rev.*, **35** (1994) 705.
[8] Theodossiou P., *Manag. Sci.*, **44** (1998) 1650.
[9] McDonald J. B., *Econ. Lett.*, **37** (1991) 273.
[10] Rachev S. T. and Mittnik S., *Stable Paretian Models in Finance* (Wiley, Chichester) 2000.
[11] Bera A. K. and Park S. Y., *J. Econom.*, **150** (2009) 219.
[12] Chan F., *Math. Comput. Simul.*, **79** (2009) 2767.
[13] Borland L., *Europhys. News*, **36** (2005) 228.
[14] Queirós S. M. D., *Quant. Finance*, **5** (2005) 475.
[15] Queirós S. M. D., *EPL*, **80** (2007) 30005.
[16] Queirós S. M. D. and Tsallis C., *Europhys. Lett.*, **69** (2005) 893.
[17] Tsallis C., *J. Stat. Phys.*, **52** (1988) 479.
[18] Tsallis C. and Queirós S. M. D., *AIP Conference Proceedings*, Vol. **965**, edited by Abe S., Herrmann H. J., Quarati P., Rapisarda A. and Tsallis C. (AIP, New York) 2007, p. 8.
[19] Abe S., *Phys. Lett. A*, **271** (2000) 74.
[20] Souza A. M. C. and Tsallis C., *Physica A*, **236** (1997) 52.
[21] Kapur J. N., *Indian J. Pure Appl. Math.*, **14** (1983) 1372.
[22] Bera A. K. and Higgins M. L., *J. Econ. Surv.*, **7** (1993) 305.