Friedrich-Alexander-Universität Erlangen-Nürnberg

Lehrstuhl für Statistik und Ökonometrie

Diskussionspapier

84 / 2009

An Alternative Model for Time-varying Moments using Maximum Entropy

Matthias Fischer and Klaus Herrmann

Lange Gasse 20 · D-90403 Nürnberg

An Alternative Model for Time-varying Moments using Maximum Entropy

MATTHIAS FISCHER AND KLAUS HERRMANN
Department of Statistics and Econometrics
University of Erlangen-Nürnberg, Germany
Email:
Matthias.Fischer@wiso.uni-erlangen.de
Klaus.Herrmann@wiso.uni-erlangen.de

SUMMARY

With their 2002 article on Maximum Entropy (ME) densities for time-varying moments Rockinger and Jondeau (2002) set a new milestone for the application of information theoretic principles to the analysis of financial market data. In this note we will apply their approach to financial data, point out some shortcomings that it encounters and show how these can be overcome.

Keywords and phrases: Entropy density; skewness; kurtosis, GARCH
JEL C22

1. Introduction

With their 2002 article on Maximum Entropy (ME) densities for time-varying moments Rockinger and Jondeau (2002) set a new milestone for the application of information theoretic principles to the analysis of financial market data. In their article, they implicitly propose a new framework for time-varying power moments up to order 4 and give an very efficient algorithm for implementing the corresponding ME densities. When applying their approach to financial data, we found that the approach with power moments meets some technical problems, especially in cases typical for financial data. Therefore, we propose to use robust moment functions, such as e.g. tan^{-1}, to model higher moments and show that we can overcome that problem. In addition to that we also find improved godness-of-fit using daily return's of the market indices S&P 500, FTSE 100 and Nikkei 225 from January 2001 to August 2008[1] as exemplary data.

Our paper is structured as follows. First we introduce a very general approach to models for time-varying moments which include Engle (1982)'s and Bollerslev (1986)'s GARCH model and Rockinger and Jondeau (2002)'s as special cases. Then we give a brief introduction on how the corresponding ME densities can be derived. The following three sections point out the limitation of an approach with power moments and give suggestions of moments that do not encounter that problem. In the the last chapter we apply both methods to financial data.

[1]The data has been downloaded from http://de.finance.yahoo.com/.

2. Models for Time-varying Moments using Maximum Entropy

Models for time-varying moments of some random variable X may generally be written as

$$X_t|\Im_{t-1} \sim F(m_{1,t}, ..., m_{k,t}), \tag{2.1}$$

where X_t is the random variable at time t, $\Im_t$ the information available at time t and F the random variables distribution at time t of which the only known information is that

$$m_{i,t} = E\left(g_i(X_t)|\Im_{t-1}\right), \tag{2.2}$$

where g_i is some moment function, $i = 1, ..., k$ and k the number of moments to be modelled.

Standard GARCH-models, introduced by Engle (1982) and Bollerslev (1986), can be considered to be a model for a time varying moment. These models try to capture the fact, that financial markets volatility is not stable over time. Using variance as a measure of volatility, these models explain volatility movements by past squared returns and past variances. In its simplest form, a GARCH(1,1) model is given by

$$X_t|\Im_{t-1} \sim F(\mu, \sigma_t^2), \quad \mu_t|\Im_{t-1} = \mu,$$
$$\sigma_t^2|\Im_{t-1} = \alpha_0 + \alpha_1 x_{t-1}^2 + \alpha_2 \sigma_{t-1}^2, \quad \alpha_0 > 0, \alpha_1 + \alpha_2 < 1, \tag{2.3}$$

where F is the (unknown) distribution of e.g. an asset return X_t at a given time t, μ the distribution's mean and σ_t^2 its variance.

We can express this model as a time-varying moment model defined above by choosing $m_{1,t} = \mu$, $g_1(a) = a$, $m_{2,t} = \sigma_t^2 = \alpha_0 + \alpha_1 x_t^2 + \alpha_2 \sigma_{t-1}^2$ and $g_2(a) = (a - \mu)^2$.

In practical applications, e.g. in order to calculate risk measures or for the ML estimation of the above model the probability density function (pdf) f is required. Of course, the knowledge of some moments does not completely determine the whole pdf. As a solution to this problem one can use the principle of Maximum Entropy (ME) to make up for the information that the model does not capture.[2] In the above setting where the only available information are the distribution's mean and variance, the resulting maximum entropy distribution is the normal distribution.[3]

There is empirical evidence, that financial returns may be well explained by GARCH(1,1)-models.[4] But looking at the distribution of GARCH-filtered residuals (ϵ_t), high values for skew and leptokurtic are often found, which implies that the assumption of normal distributed innovations may not hold.[5]

A solution to this problem can be to assume more informative pdfs for the innovation's distribution, such as e.g. a skewed t-distribution[6]. But in consequent application of information theoretic methods in the above framework, one should rather try not only to

[2] For the justification of the Maximum Entropy approach in problems of given moments see Jaynes (1957).
[3] Compare e.g. Cover and Thomas (2006).
[4] Compare e.g. Bera and Higgins (1993).
[5] Compare e.g. Bollerslev (1987) or Hansen (1994).
[6] See Hansen et al. (2007) or Mittnik et al. (1998).

model mean's and variance's motion in time, but include additional knowledge in the form of higher moments. Rockinger and Jondeau (2002) suggest to examine time-varying models for return distribution's third and fourth moments, namely skewness and kurtosis. As a measure for skewness and kurtosis they use third and fourth standardized power moments as

$$s = E\left(g_3(X)\right) = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right) \quad \text{and} \tag{2.4}$$

$$k = E\left(g_4(X)\right) = E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right), \tag{2.5}$$

and assume these moments to be constant over time or in motion of the form

$$m_{3,t} = \beta_1 + \beta_2 x_{t-1}, \quad m_{4,t} = \gamma_1 + \gamma_2 |x_{t-1}| \quad \text{or alternatively} \tag{2.6}$$

$$m_{3,t} = \beta_1 + \beta_2 \frac{x_{t-1}}{\sigma_{t-1}}, \quad m_{4,t} = \gamma_1 + \gamma_2 |\frac{x_{t-1}}{\sigma_{t-1}}|, \quad \text{with} \quad (m_{3,t}, m_{4,t}) \in \mathcal{E} \tag{2.7}$$

where β_1, β_2, γ_1 and γ_2 are constants and $\mathcal{E}$ is the set of values of s and k for which a ME density is defined.

3. Determining Maximum Entropy Densities

Let X denote a random variable following a maximum entropy distribution. From standard literature of information theory (see e.g. Cover and Thomas (2006)), we know that a ME distribution pdf's (here f_{ME}) functional form with given four moment functions $g_i(\cdot)$ is

$$f_{ME}(x) = \exp\left(\sum_{i=0}^{4} \lambda_i g_i(x)\right), \tag{3.1}$$

where $g_0(x) = 1$ and $\lambda_0, ..., \lambda_4$ are uniquely determined through the side conditions of the given moment restrictions

$$E(g_i(X)) = \int_{\mathcal{D}} g_i(x) \cdot f_{ME}(x) dx = m_i. \tag{3.2}$$

If e.g. $g_i(a) = a^i$, for $i = 1, ..., 4$, we obtain $m_0 = 1$, $m_1 = \mu$, $m_2 = \sigma^2 + \mu^2$, $m_3 = s\sigma^3 + 3\sigma^2\mu + \mu^3$ and $m_4 = k\sigma^4 + 4s\sigma^3\mu + 6\sigma^2\mu^2 + \mu^4$.

While the functional dependence of λ_i from the moment target values m_j is known for the case of given first two power moments as

$$\lambda_0 = -\ln(\sqrt{2\pi}\sigma) - \frac{\mu^2}{2\sigma^2}, \quad \lambda_1 = \frac{\mu}{\sigma^2}, \quad \lambda_2 = \frac{-1}{2\sigma^2}, \tag{3.3}$$

which is the normal distribution, it is, to our best knowledge, still unknown for cases including higher moments.

However, it is possible to determine the values for λ_0, ..., λ_4 for given $m_1, ..., m_4$ through minimizing a function $Q(\lambda_0, ..., \lambda_4)$, that is convex in $\lambda_0, ..., \lambda_4$ and for which holds that

$$\frac{\partial Q}{\partial \lambda_i} = \int_{\mathcal{D}} (g_i(x) - m_i) \cdot f_{ME}(x) dx, \tag{3.4}$$

where g_i is i-th moment function, m_i the i-th moment target value and $\mathcal{D}$ the distribution's support. This minimum in λ_i is then at the same time solution to the maximum entropy density problem. In this setting such a dual problem is given by[7]

$$Q(\lambda_0, ..., \lambda_4) = \int_{\mathcal{D}} \exp\left(\sum_{i=0}^{4} x^i - m_i\right) dx. \tag{3.5}$$

There exist different suggestions on how to implement ME densities most efficiently, see e.g. Ormoneit and White (1999) or Agmon et al. (1979). In this article we follow the approach proposed by Rockinger and Jondeau (2002).

4. Limitations of the Maximum Entropy Solution

A simple consideration shows that the maximum entropy density in equation (3.1) is not able to build useful models for all combinations of $m_0, ..., m_N$ if X is defined on $\mathbb{R}$. To see this, let us have a look at how the moment target values m_i in equation (3.2) depend on λ_i:

$$\frac{\partial m_i}{\partial \lambda_i} = \frac{\partial \int_{\mathbb{R}} g_i(x) \cdot f_{ME}(x) dx}{\partial \lambda_i} = \int_{\mathbb{R}} g_i(x)^2 \cdot f_{ME}(x) dx > 0. \tag{4.1}$$

So, a slightly higher (lower) target value m_i will always result in a higher (lower) value for the corresponding λ_i. But taking the limit for the density in equation (3.1) for $|x| \to \infty$ in the case of $g(x_i) = x^i$ and even N[8]

$$\lim_{|x|\to\infty} \exp(\sum_{i=0}^{N} \lambda_i g_i(x)) = \lim_{|x|\to\infty} \exp(\sum_{i=0}^{N} \lambda_i x^i) = \lim_{|x|\to\infty} \exp(\lambda_N x^N) = \begin{cases} 0 & \text{for } \lambda_N < 0, \\ 0 \text{ or } \infty & \text{for } \lambda_N = 0, \\ \infty & \text{for } \lambda_N > 0. \end{cases} \tag{4.2}$$

We can see, that λ_N is restricted to negative values or zero to obtain a proper density. Consequently, there is an upper bound to λ_N and also to m_N. In the case of $N = 4$ there is an upper bound to m_4[9]. This bound is reached when λ_N is equal to zero, which is, when the corresponding moment restriction is not binding. So m_4 cannot be modeled to be bigger than implied by mean, variance and skewness, which in the case of no skewness is for standardized data a value of 3.

[7] This can be seen by looking at $\frac{\partial Q}{\partial \lambda_i}$. For the convexity of Q in λ compare Rockinger and Jondeau (2002).

[8] A similar problem arises for odd N, see e.g. Cover and Thomas (2006).

[9] This has already been observed by Einbu (1977) for ME distributions defined on $\mathbb{R}^+$.

Of course, a maximum entropy solution for cases of greater m_4 still exists. But as Cover and Thomas (2006) point out for the case of given first three power moments, this value is achieved through a density of equation (3.1), but with some small 'wiggles' somewhere in the infinity to make up for the higher moments.

5. Solutions on a Truncated Support

Analyzing log-return densities r_t of the form

$$r_t = log(p_t) - log(p_{t-1}), \tag{5.1}$$

where p_t is the asset price at time t, as we do here, means using densities with infinite support, so that $\mathcal{D} = \mathbb{R}$ in the above setting.

As we will show in the sequel, it is sometimes convenient to run numerical procedures on finite supports. Standardizing the returns to a mean of zero and a variance of one, for convenience, it is advisable to restrict the support for the above integration to some finite interval $[-z; z]$ with $z \in \mathbb{R}^+$.

Approximate equality for the solutions found on truncated and not truncated supports only holds if the ME solution $f_{ME}(x)$ is close to zero for any $x \in \mathbb{R} \setminus [-z; z]$, so that

$$E(g_i(X)) = \int_{\mathbb{R}} g_i(x) \cdot f_{ME}(x) dx \approx \int_{-z}^{z} g_i(x) \cdot f_{ME}(x) dx = m_i, \quad i = 0, 1, .., 4. \tag{5.2}$$

It is obvious that this cannot hold for cases of higher kurtosis as discussed above.

Using truncated supports will lead to numerical solutions for any possible value for kurtosis.[10] So, by choosing $z = 10$ and using numerical minimization of Q over $\lambda_0, ..., \lambda_4$, we find similar densities as those given by Rockinger and Jondeau (2002), see figure 1, where the graphs on the left panel have been produced for $m_4 = 2$ and $m_3 \in \{0.1, 0.4, 0.7\}$ and those on the right panel for $m_3 = 2$ and $m_4 \in \{6, 8, 10\}$. As all values found for λ_4 are negative, the results could be applied to infinite intervals as well.

But if we look at cases where the kurtosis is higher than implied be skewness, for example $m_3 = -0.1$ and $m_4 = 7$ in figure 2, we can see that the density is restricted to the support that it was found on. Here we increased the support only slightly to $[-10.7, 10.7]$ and obviously f_{ME} is no longer a proper density.[11] But as financial data exhibits higher kurtosis even after filtering for GARCH effects, as mentioned in the beginning, the densities proposed by Rockinger and Joundeau (2002) are restricted to the closed interval $[-z, z]$.

Of course, one could accept the ME pdfs only to be defined on a finite support. But then, as the density is increasing faster then exponentially even in the tails of the limited support, deriving risk measures like the Value at Risk (VaR) can lead to unreasonably high values.

[10] A restriction to the possible values of skewness can be found in Rockinger and Jondeau (2002).
[11] A similar problem appears for λ_5 for the mixed moment in the bivariate ME densities proposed in Miller and Liu (2002).

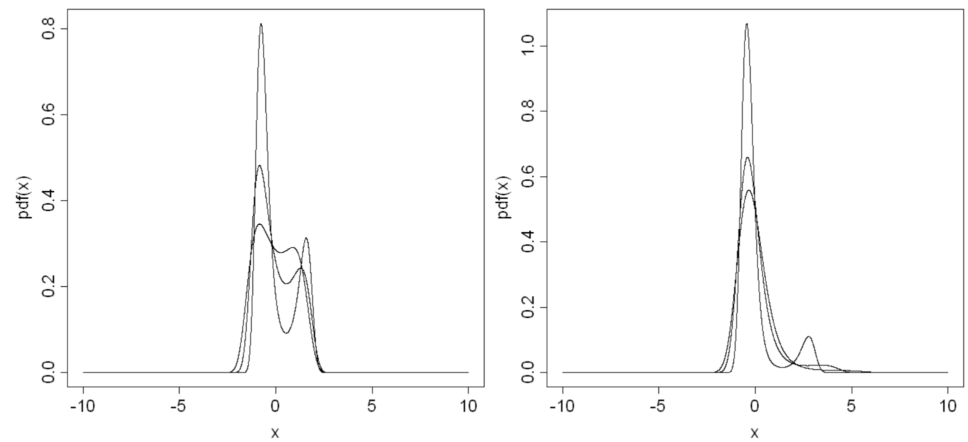

Figure 1: Pdfs similar to those given by Rockinger and Jondeau (2002).

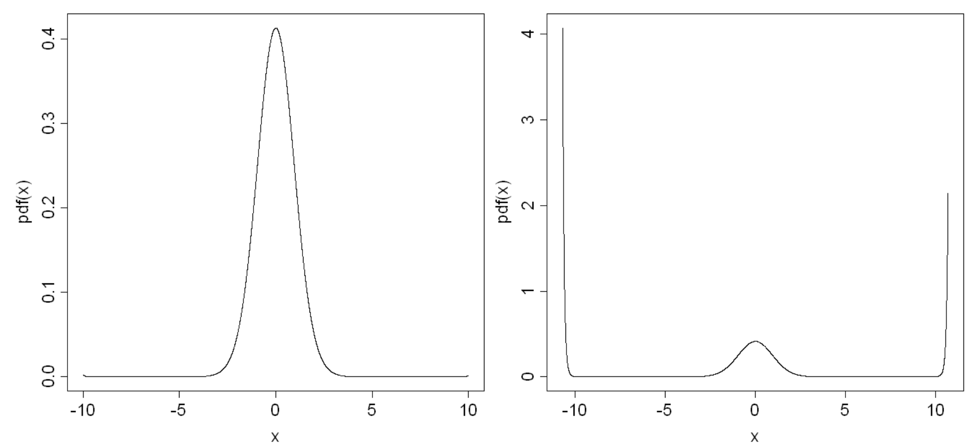

Figure 2: ME density for $m_3 = -0.1$ and $m_4 = 7$ plotted on $[-10; 10]$ and $[-10.7; 10.7]$.

6. An Alternative ME Density for higher Moments

A solution to this problem may be to alter the way to measure skewness and kurtosis. Using measures that are functions of X which grow slower in X than the moment function in $E(X^2)$ would enable us to create kurtosis higher than implied by skewness and variance. So, applying e.g. bounded functions - as it is done in robust statistics - will give better solutions. Here we propose

$$E(g_3(X)) = E\left(tan^{-1}\left(\frac{X}{\sigma}\right)\right) \quad \text{and} \quad E(g_4(X)) = E\left(tan^{-1}\left(\frac{X}{\sigma}\right)^2\right) \tag{6.1}$$

as measures of skewness[12] and kurtosis, which leads to a maximum entropy density of the form:[13]

$$f_{ME}(x) = \exp\left(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 tan^{-1}\left(\frac{x}{\sigma}\right) + \lambda_4 tan^{-1}\left(\frac{x}{\sigma}\right)^2\right). \tag{6.2}$$

The density function for m_3 and m_4 chosen that the value for the standardized third and fourth power moments equals -0.1 and 7 respectively is given in figure 3:

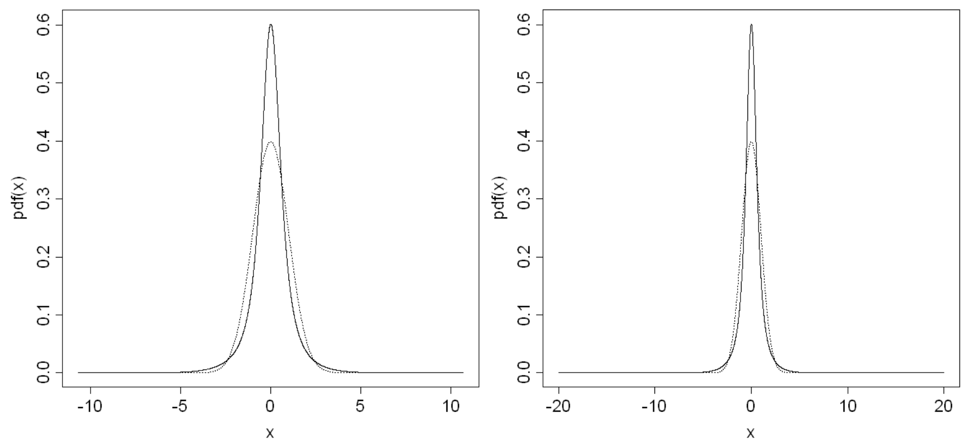

Figure 3: Normal (dotted line) and our ME density on $[-10; 10]$ and $[-20; 20]$.

Of course, the same problem as described in section 2 might also arise. Here the fastest growing moment function in X is X^2, so that we will find an upper bound to the variance. But the variance implied by tan^{-1}-moments is infinite, as

$$\lim_{|x|\to\infty} x^2 \cdot f_{ME}(x) = \lim_{|x|\to\infty} x^2 \cdot \exp(\lambda_0 + \lambda_3 tan^{-1}(x) + \lambda_4 tan^{-1}(x)^2) = \infty. \tag{6.3}$$

So, the upper bound to variance cannot be reached.

In a time-varying moment framework, as defined above, we could use following model for the third and fourth moment's motion in time as:

$$m_3 = \beta_0 + \beta_1 tan^{-1}\left(\frac{x_{t-1}}{\sigma_{t-1}}\right) \quad \text{and} \quad m_4 = \gamma_0 + \gamma_1 tan^{-1}\left(\frac{x_{t-1}}{\sigma_{t-1}}\right)^2, \tag{6.4}$$

with $m_3, m_4 \in \mathcal{E}$, where $\mathcal{E}$ is the area of combinations of m_3 and m_4 for which corresponding ME densities can be derived. In the sequel we will use $|m_3| < .11$ and $0.25 < m_4 < 0.5$. This area could be largely extended if we examined the relation between m_4 and m_3 more closely. But as following applications showed, this area is large enough for the cases relevant to our exemplary data set.

[12]E.g. Bera and Premaratne (2005) use tan^{-1} as a robust measure of skewness.
[13]Similar suggestions have been made in Bera and Park (2009).

7. Application to the Financial Market Data

Finally, the goodness-of-fit between Rockinger and Jondeau (2002)'s and our approach is compared. For this purpose, we chose three daily market indices, namely S&P 500, FTSE 100 and Nikkei225 from January 2001 to August 2008 such that roughly 1900 observations are available for each data set. Table 1 summarizes selected descriptive statistics for the corresponding log-return series. All of the underlying data set exhibit remarkable kurtosis and are skewed - at least to some extend and in terms of the fourth and third standardized moments. In addition, the corresponding robust skewness and kurtosis coefficients are given.

Estimates	S&P 500	FTSE 100	Nikkei 225
$\overline{x}$	1.78147e-07	4.71454e-05	2.45672e-05
$s^2 \equiv \overline{x^2} - \overline{x}^2$	0.00012	0.00013	0.00020
$\overline{(\frac{x-\overline{x}}{s})^3}$	-0.08414	0.16220	0.15840
$\overline{(\frac{x-\overline{x}}{s})^4}$	5.35543	6.09638	4.45947
$\overline{\tan^{-1}(\frac{x-\overline{x}}{s})}$	-0.00571	-0.00749	-0.00788
$\overline{\tan^{-1}(\frac{x-\overline{x}}{s})^2}$	0.38790	0.37366	0.41101
Number of observations	1925	1934	1882

Table 1: Some descriptive statistics for the data.

In a first step and for reasons of benchmarking, we estimated a Standard-Gaussian GARCH(1,1)-model where only m_2 is time-varying of the form

$$m_{2,t} = \alpha_0 + \alpha_1 x_t^2 + \alpha_2 m_{2,t-1}, \quad \alpha_0 > 0, \alpha_1 + \alpha_2 < 1, \tag{7.1}$$

while the others moments $m_3 = 0$ and $m_4 = 3$ are constant over time. The summary results are displayed in table 2, below.

Indice	α_0	α_1	α_2	LogL	$\overline{\epsilon_t^3}$	$\overline{\epsilon_t^4}$
S&P 500	0.00538	0.06573	0.93030	-2468.048	-0.28253	3.62652
	(0.00244)	(0.01100)	(0.01153)		[0.05583]	[0.11166]
FTSE 100	0.01070	0.11530	0.87547	-2357.594	-0.31952	3.39483
	(0.00334)	(0.01478)	(0.01515)		[0.05570]	[0.11140]
Nikkei 225	0.01077	0.07917	0.91228	-2535.948	-0.05553	3.45916
	(0.00401)	(0.01120)	(0.01172)		[0.05646]	[0.11293]

Table 2: GARCH(1,1) estimates and conditional moments in our exemplary data sets.

Obviously, there is still skewness and kurtosis in the residuals of the model (see $\overline{\epsilon_t^3}$ and $\overline{\epsilon_t^4}$) for all data set, such that the standard GARCH-model is not adequate (which is in line with other results from the relevant literature). We find strong non-normal kurtosis for all data sets and strong non-normal skewness for the S&P 500 and the FTSE 100.

Using the test on normality based on skewness and kurtosis proposed by Bera and Jarque (1980), we reject the null hypothesis of normality in all three cases at high significance levels. So, the assumption of normally distributed innovations does not hold. Here, LogL denotes the logarithm of the observations likelihood, in round brackets we give the standard deviation estimated by the inverse of the numerically estimated Hesse matrix and in box brackets we give the asymptotic standard deviation for the moment estimates under the assumption of normal distributed innovations.

Consider, in a second step, the results of a GARCH(1,1)-estimation where the residual distribution is one of the MED distributions considered above. For reasons of brevity, *PO* refers to the approach using power moments and *AT* to moments using tan^{-1} in table 3, below.

	S&P 500		FTSE 100		Nikkei 225	
	PO	*AT*	*PO*	*AT*	*PO*	*AT*
α_0	0.00534	0.00501	0.01095	0.01075	0.01022	0.01044
	(0.00248)	(0.00265)	(0.00353)	(0.00361)	(0.00404)	(0.00438)
α_1	0.06755	0.06717	0.12453	0.11959	0.07799	0.07952
	(0.01108)	(0.01215)	(0.01648)	(0.01647)	(0.01152)	(0.01253)
α_2	0.92810	0.92936	0.86623	0.87145	0.91385	0.91222
	(0.01150)	(0.01248)	(0.01658)	(0.01666)	(0.01198)	(0.01301)
m_3	-0.13612	0.00354	-0.15140	0.00517	0.04561	7.060e-05
	(0.11077)	(0.00355)	(0.08994)	(0.00346)	(0.16065)	(0.00355)
m_4	3.41645	0.42832	3.27046	0.43674	3.50760	0.43171
	(0.89920)	(0.00442)	(0.63467)	(0.00421)	(1.42400)	(0.00439)
LogL	-2461.798	-2453.475	-2352.495	-2350.926	-2532.598	-2526.314

Table 3: Estimates for models with constant higher moments.

Clear, our *AT*-approach outperforms that of Rockinger and Jondeau (2002) if the log-likelihood values are compared. This is also confirmed if we consider time-varying models, where also the higher moments are allowed to vary in time as autoregressive of order one in a third step (see table 4). In particular, the focus is on both specifications given by Rockinger and Jondeau (2002), denoted as *PO*va and *POvb*, and on our specification using tan^{-1}-moments denoted as *AT*v.

8. Summary

We showed that the ME approach for including knowledge of higher moments proposed by Rockinger and Jondeau (2002) produces improper densities in cases for data with high kurtosis, in particular for financial data. This motivates to use bounded functions for the incorporation of knowledge of higher moments. In addition,using daily returns of some major market indices, S&P 500, FTSE 100 and Nikkei 225, we were able to provide empirical evidence that our approach outperforms that of Rockinger and Jondeau (2002).

References

[1] Agmon, N., Y. Alhassid and R.D. Levine (1979): "An Algorithm for Finding the Distribution of Maximal Entropy", *Journal of Computational Physics*, 30, 250-258.

[2] Bera, A. K., and M. L. Higgins (1993): "ARCH Models: Properties, Estimation and Testing", *Journal of Economic Surveys*, 7(4), 305-66.

[3] Bera, A. K., and C. M. Jarque (1980): "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economics Letters*, 6(3), 255-259.

[4] Bera, A. K., and S. Y. Park (2009): "Maximum Entropy Autoregressive Conditional Heteroscedasticity Model", Working Paper.

[5] Bera, A. K., and G. Premaratne (2005): "A Test for Symmetry with Leptokurtic Financial Data", *Journal of Financial Econometrics*, 3(2), 169-187.

[6] Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 31, 307-327.

[7] Bollerslev, T. (1987): "A Conditional Heteroscedastic Time Series Model for Speculative Price and Rate of Return", *Review of Economics and Statistics*, 69(3), 542-547.

[8] Cover, T. M., and J.A. Thomas (2006): *Elements of Information Theory*, New York: John Wiley & Sons.

[9] Einbu, J.M. (1977): "On the Existence of a Class of Maximum-Entropy Probability Density Functions", *IEEE Transactions on Information Theory*, 23(6), 772-775.

[10] Engle, R.F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50(4), 987-1006.

[11] Hansen, B.E. (1994): "Autoregressive Conditional Density Estimation", *International Economic Review*, 35(3), 705-730.

[12] Hansen, C.B., J. McDonald and P. Theodossiou (2007): "Some Flexible Parametric Models for Partially Adaptive Estimators of Econometric Models", *Economics: The Open-Access, Open-Assessment E-Journal*, 1(7), electronic resource.

[13] Jaynes, E.T. (1957): "Information Theory and Statistical Mechanics", *The Physical Review*, 106(4), 620-630.

[14] Miller, D. J. and W. Liu (2002): "On the Recovery of Joint Distributions from Limited Data", *Journal of Econometrics*, 107, 259-274.

[15] Mittnik, S., M. S. Paolella and S. T. Rachev (1998): "Unconditional and Conditional Distributional Models for the Nikkei Index", *Asia-Pacific Financial Markets*, 5(2), 99-128.

[16] Ornomeit, D. and H. White (1999): "An Efficient Algorithm to Compute Maximum Entropy Densities", *Econometric Reviews*, 18, 127-140.

[17] Rockinger, M. and E. Jondeau (2002): "Entropy Densities with an Application to Autoregressive Conditional Skewness and Kurtosis", *Journal of Econometrics*, 106, 119-142.

	α_0	α_1	α_2	β_0	β_1	γ_0	γ_1	LogL
S&P 500								
POva	0.00542	0.06519	0.93046	-0.20664	0.00701	4.47258	-0.58301	-2460.707
	(0.00246)	(0.01088)	(0.01129)	(0.09045)	(0.05622)	(0.77397)	(0.26331)	
POvb	0.00621	0.07869	0.91980	-0.12826	0.38303	4.25616	1.82755	-2457.461
	(0.00316)	(0.01498)	(0.01420)	(0.06268)	(0.36924)	(1.01815)	(3.58495)	
ATv	0.00515	0.07703	0.92297	0.00302	-0.01866	0.43765	-0.03351	-2447.483
	(0.00289)	(0.01489)	(0.01382)	(0.00353)	(0.00656)	(0.00620)	(0.01578)	
FTSE 100								
POva	0.01042	0.12418	0.86754	-0.23860	0.02164	3.05670	1.43632	-2350.662
	(0.00343)	(0.01637)	(0.01618)	(0.08681)	(0.07551)	(0.17491)	(1.04020)	
POvb	0.01093	0.12380	0.86762	-0.23877	0.01715	3.07572	0.85802	-2350.331
	(0.00345)	(0.01637)	(0.01630)	(0.08270)	(0.02879)	(0.17907)	(0.43905)	
ATv	0.01054	0.12239	0.87006	0.00612	-0.00243	0.44509	-0.02055	-2349.066
	(0.00357)	(0.01760)	(0.01695)	(0.00352)	(0.00562)	(0.00599)	(0.01142)	
Nikkei 225								
POva	0.01007	0.08256	0.91203	-0.13152	-0.50597	4.08602	2.31427	-2526.561
	(0.00416)	(0.01587)	(0.01306)	(0.08749)	(0.60908)	(1.12884)	(6.64341)	
POvb	0.01128	0.07921	0.91233	0.11143	-0.06931	3.15519	1.78052	-2530.722
	(0.00440)	(0.01194)	(0.01246)	(0.13987)	(0.11708)	(0.86144)	(2.06080)	
ATv	0.01020	0.08073	0.91201	0.00059	0.00487	0.43571	-0.01137	-2525.670
	(0.00436)	(0.01337)	(0.01328)	(0.00362)	(0.00615)	(0.00617)	(0.01316)	

Table 4: Estimates for models with time-varying higher moments.