IWQW

Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 06/2008

Models for Time-varying Moments Using Maximum Entropy Applied to a Generalized Measure of Volatility

Klaus Herrmann
University of Erlangen-Nürnberg

ISSN 1867-6707

Models for Time-varying Moments Using Maximum Entropy Applied to a Generalized Measure of Volatility

Klaus Herrmann[1]

Department of Statistics and Econometrics
University of Erlangen-Nürnberg, Germany

Email:
klaus.herrmann@wiso.uni-erlangen.de

3.12.2008

SUMMARY

We use an information-theoretic approach to interpret Engle's (1982) and Bollerslev's (1986) GARCH model as a model for the motion in time of the expected conditional second power moment. This interpretation is used to show how these models may be generalized, if we use alternative measures of volatility. We choose one feasible alternative and derive a generalized volatility model. Applying this model to some exemplary market indices, we are able to give some empirical evidence for our method.

Keywords and phrases: Information Theory, Maximum Entropy, GARCH, Volatility.

JEL classification: C22.

[1] I thank Ingo Klein, Matthias Fischer, Vlad Ardelan and Stephan Schlüter for helpful discussions and ideas.

1 Introduction

Information-theoretic approaches have become more and more popular in econometrics in recent years.[2] Especially the idea of characterizing observable but unknown distributions through their moments and some information measure has found new applications from income distributions[3] to distributions of financial returns[4].

In this article, we use information theory to interpret GARCH models as models for the motion of variance in time. Applying the principle of Maximum Entropy (ME) we get the normal distribution as return distribution as it is the ME distribution for a given variance.

Similar interpretations of GARCH models have been used in Rockinger and Jondeau (2002), Fischer and Herrmann (2008), Bera and Park (2008) and Queirós and Tsallis (2005). In the first three of these articles volatility is measured by variance, additional knowledge of higher moments is included to give a better characterization of the observed data. Queirós and Tsallis (2005) use the above interpretation to model variance and derive return distributions using the ME principle and a generalized information measure.

All these approaches use variance to measure volatility. But to our view this is only one possible way of measuring what the notion of volatility should mean. We briefly discuss how volatility should by defined and derive a first suggestion of how this concept may be measured more generally as a measure of dispersion. As a first suggestion, we use Bickel and Lehmann's (1976) proposition of a generalized measure of dispersion where variance and average absolute deviation appear as special cases, to derive a new volatility model.

Applying this model to daily returns of the market indices S&P 500, FTSE 100 and Nikkei 225 from January 2001 to August 2008[5], we are able to give some empirical evidence for our method.

Our argument is structured as follows: First we give a brief introduction on some concepts of information theory and models for time-varying moments. Then we show how Engle (1982)'s and Bollerslev (1986)'s GARCH models may be interpreted more generally as information theoretic models for time-varying moments. After a brief discussion on volatility we propose a first generalization of the model and derive the corresponding ME density. In the last chapter we are able to give some empirical evidence for our method.

2 Some Concepts of Information Theory

Information theory bases on Shannon (1948)'s idea that the abstract concept of information can be quantified. From coding theory it is known that the most efficient way to store a message M out of a given set of m possible messages is to code every message with $log_k(\frac{1}{f(M)})$ digits, where k denotes the number of signs of the alphabet, log_k the logarithm to base k and $f(M)$ the messages relative frequency. Efficient coding here means that there

[2]Compare e.g. Golan and Maasoumi (2008) or Golan (2002).
[3]Compare Wu (2003).
[4]Compare e.g. Borland (2005).
[5]The data has been downloaded from http://de.finance.yahoo.com/.

is no other possibility of coding messages such that the expected code length for an unknown message is smaller than

$$\sum_{i=1}^{m} f(M_i) \cdot log_k \left(\frac{1}{f(M_i)} \right),$$

where m is the number of possible messages. So, for given relative frequencies we can derive a lower bound for the expected code length. Using this result, we can give a lower bound for the information contained in some unknown message as its minimal expected code length.

In statistics we often deal with the different problem of searching an useful assumption for some random variable X's density $f(X)$. The idea from information theory is, that if our assumption shall include the least knowledge possible, the information generated by some random draw should be maximal. Formally we look for some density f for which the expression

$$H(X) = \sum_{x \in \mathcal{D}} f(x) ln \left(\frac{1}{f(x)} \right) = - \sum_{x \in \mathcal{D}} f(x) ln \left(f(x) \right)$$

or its continuous analogue

$$H(X) = \int_{x \in \mathcal{D}} f(x) ln \left(\frac{1}{f(x)} \right) dx = - \int_{x \in \mathcal{D}} f(x) ln \left(f(x) \right) dx$$

called information entropy or simply entropy, is maximal, where $\mathcal{D}$ is the distribution's support.[6] In some cases we might have some prior knowledge, e.g. derived from some model, about F, e.g. in the form of expected values $E(g(X))$. This problem can be solved using an Euler-Lagrange approach for calculus of variations[7], where we can include additional knowledge as side constraints for the maximization task.

3 Models for Time-varying Moments

Models for time-varying moments have been introduced implicitly in Rockinger and Jondeau (2002). These models may generally be written as

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

where X_t is some random variable at time t, $\Im_t$ the set of information available at time t and F its conditional distribution for which the only known information is that

$$m_{i,t} | \Im_{t-1} = E(g_i(X_t) | \Im_{t-1}), \quad i \in \{1, ..., k\}$$

where g_i is the i-th moment function, with $E\left(g_i(X_t)\right) < \infty$, and $m_{i,t}$ the i-th moment's motion in time, e.g. as

$$m_{i,t} | \Im_{t-1} = \alpha_{i,0} + \sum_{j=1}^{p} \alpha_{i,j} g_i(x_{t-j}) + \sum_{j=1}^{q} \beta_{i,j} m_{i,t-j},$$

[6] As $log_k(x) = c \cdot ln(x)$, we can use any base for the maximization regardless of the numeral system used.
[7] Compare e.g. Brunt (2004).

$i = 1, ..., k$ and $k \in \mathbb{N}$ the number of moments to be modeled.

If there is no additional assumption on the functional form of the conditional distribution F, using the information-theoretic concepts described above, we should assume for F the maximum entropy distribution (MED) under constraints of the expected moment values known from our model.[8] That means that for every point in time t we find the conditional density function f for x_t as solution to the problem

$$\max_f \left(- \int_{z \in \mathcal{D}} f(z) log(f(z)) dz \right)$$

under the constraints that

$$\int_{z \in \mathcal{D}} f(z) dz = 1 \quad \text{and} \quad \int_{z \in \mathcal{D}} g_i(z) f(z) dz = m_{i,t} \quad i \in \{1, ..., k\},$$

where $\mathcal{D}$ is the support for x_t. Solutions to this problem can be found e.g. in Cover and Thomas (2006). We will denote the corresponding distributions as $MED(E(g_1(X)) = m_1, ..., E(g_k(X)) = m_k)$.

4 GARCH-Models

The idea of Engle (1982) and Bollerslev (1986)'s GARCH models is to capture the fact that the distribution of asset returns seems not to be stable over time, see figure 1. Assuming that there are clusters where the returns' volatility is higher or lower, volatility could be explained as an autoregressive process. Using variance, that is for standardized returns x_t $E(x_t^2)$, as measure for volatility, the general form, a GARCH(p,q) model, may be given as[9]

$$x_t | \Im_{t-1} = \sigma_t z_t, \quad z_t \sim P$$
$$\sigma_t^2 | \Im_{t-1} = \alpha_0 + \sum_{i=1}^{p} \alpha_{1,i} x_{t-i}^2 + \sum_{i=1}^{q} \beta_{1,i} \sigma_{t-i}^2,$$

where x_t is the return, z_t some innovation, P its distribution and σ_t^2 the variance of the returns' distribution at time t. Traditional GARCH models give explicit assumptions on the innovation's distribution P. The simplest of these may be the assumption of normal distributed innovations.

In rewriting the above equations as

$$m_{1,t} | \Im_{t-1} = 0, \quad g_1(x) = x,$$
$$m_{2,t} | \Im_{t-1} = \alpha_0 + \sum_{i=1}^{p} \alpha_{1,i} g_2(x_{t-i}) + \sum_{i=1}^{q} \alpha_{2,i} m_{2,t-i}, \quad g_2(x) = x^2,$$

we can easily interpret the above model as a model for time-varying moments in the above framework. As we only model the variance's motion in time the corresponding ME distribution again is the normal distribution.

[8] Compare Jaynes (1957).
[9] Compare Bollerslev (1986).

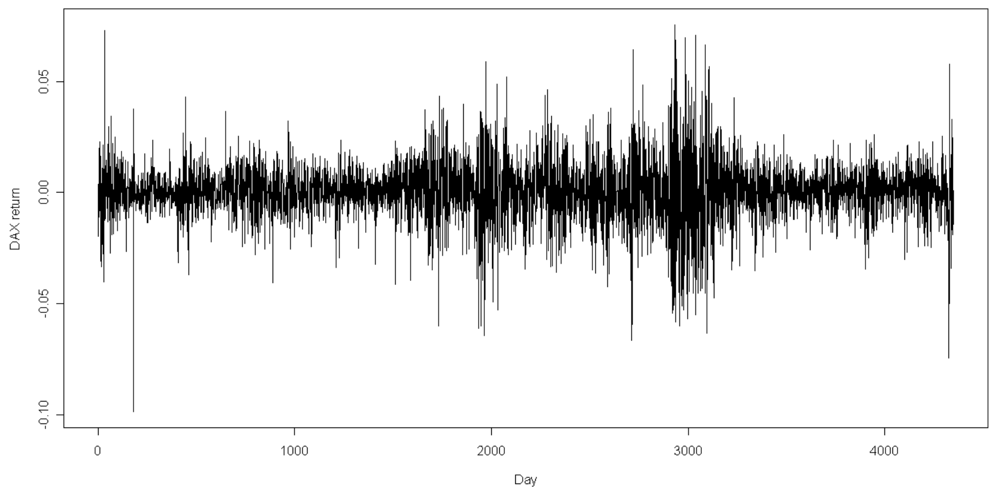

Figure 1: Daily DAX returns from 1990-11-26 to 2008-02-22.

5 Measuring Volatility

The term volatility is used in many fields of science, mainly describing the notion of instability or variability. In financial literature it is often directly referred to the variance rate of some generalized Wiener process.[10] For time series analysis, including GARCH models, volatility is usually measured by variance.

But to our view, volatility in financial analysis should also be defined more general, e.g. as a distribution's time-varying dispersion. Using this definition we could use any measure of dispersion as a measure of volatility.

Measures of dispersion and required properties are discussed e.g. in Bickel and Lehmann (1976) or Oja (1981). For the above proposed maximum entropy framework for time-varying moments, we are only able to consider measures of dispersion that can be expressed in form of an expected value of some function of the random variable $g(X)$. This condition cancels out frequently used measures of dispersion depending on quantiles, such as interquantile range or mean absolute deviation from median.

But we can use the common measures such as variance or mean absolute deviation from mean. Both of these are special cases of the more general class of measures of dispersion proposed by Bickel and Lehmann (1976), called the pth power deviation, formally written as

$$\tau_p = \left(E\left(|X-\mu|^p\right)\right)^{\frac{1}{p}},$$

[10] Compare Hull (1993).

for $1 \leq p \leq 2$. In our setting it is more convenient to change the notation and use the monotone transformation of the pth power deviation

$$\tau^* = E\left(|X-\mu|^\gamma\right).$$

6 A Feasible Model with a Generalized Variance

Using τ^* as a measure of dispersion instead of variance, which means to replace g_2, we derive a more general model for a time-varying moment of dispersion as

$$m_{1,t}|\Im_{t-1} = 0, \quad g_1(x) = x$$

$$m_{2,t}|\Im_{t-1} = \alpha_0 + \sum_{i=1}^{p} \alpha_{1,i} g_2(x_{t-i}) + \sum_{i=1}^{q} \alpha_{2,i} m_{2,t-i}, \quad g_2(x) = |x|^\gamma,$$

where the traditional GARCH(p,q) model assuming Gaussian innovations appears as a special case, if we set $\gamma = 2$.

The corresponding conditional maximum entropy density f_{ME} at a given time t with information set $\Im_{t-1}$ can be derived as the solution to the problem

$$f_{ME} = \underset{f}{argmax} \left(\int_{\mathbb{R}} f(v) ln(f(v)) dv \right)$$

where f is the density function, under the side conditions

$$\int_{\mathbb{R}} f(v) dv = 1$$

and

$$\int_{\mathbb{R}} |v|^\gamma f(v) dv = m_{2,t}|\Im_{t-1}.$$

The condition

$$\int_{\mathbb{R}} v f(v) dv = m_{1,t}|\Im_{t-1} = 0.$$

can be neglected, as it is not binding.[11] The solution's functional form is known from standard literature on information theory as[12]

$$f(x_t) = \exp(\lambda_0 + \lambda_1 |x_t|^\gamma).$$

This form coincides with the Box/Tiao error distribution (BT)[13], which we will note as

$$f_{BT}(x_t) = \frac{e^{-\frac{1}{\gamma}\left(\frac{|x_t|}{\sigma_t}\right)^\gamma}}{c(\gamma)\sigma_t}$$

[11] Compare Cover and Thomas (2006).
[12] Compare e.g. Cover and Thomas (2006) or Kapur (1989).
[13] Compare Box and Tiao (1962).

where

$$c(\gamma) = \int_{\mathbb{R}} e^{-\frac{1}{\gamma}|v|^{\gamma}} dv.$$

For $\gamma = 2$ we find the normal distribution, for $\gamma = 1$ the Laplace distribution. As we find for a given moment value[14]

$$E(|X|^{\gamma}) = \int_{\mathbb{R}} |v|^{\gamma} \frac{e^{-\frac{1}{\gamma}(\frac{|v|}{\sigma})^{\gamma}}}{c(\gamma)\sigma} dv = \sigma^{\gamma},$$

we can derive the dependence between λ_0 and λ_1 and $m_{2,t}$ as

$$\lambda_0 = ln\left(-c(\gamma)\sigma_t\right) = ln\left(-c(\gamma)m_{2,t}^{-\gamma}\right) \quad \text{and} \quad \lambda_1 = \frac{1}{\gamma\sigma_t^{\gamma}} = \frac{1}{\gamma m_{2,t}}.$$

Using the scalability[15] of the Box/Tiao error distribution, we can rewrite this model in a notation similar to the original notation of GARCH(p,q) models as

$$x_t|\Im_{t-1} = \sigma_t z_t \quad \text{with } z_t \sim MED\left(E\left(|z_t|^{\gamma}\right) = 1\right)$$

and

$$\sigma_t^{\gamma}|\Im_{t-1} = E\left(|X_t|^{\gamma}|\Im_{t-1}\right) = \alpha_0 + \sum_{i=1}^{p} \alpha_i |x_{t-i}|^{\gamma} + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^{\gamma}.$$

This model resembles to a model proposed by Higgins and Bera (1992) or the general model proposed by Hentschel (1995), who applied similar specifications to model variance under parametric assumptions for the distribution of z_t.[16]

7 Application to Financial Market Time Series

For the application to data we will use a reduced version of the above model where we set $p = 1$ and $q = 1$, as there is some evidence that for traditional GARCH models this parametrization is sufficient.[17] For the estimation of the models parameters we use the maximum likelihood method. As exemplary market indices we will use daily returns of the S&P 500, FTSE 100 and Nikkei 225 from January 2001 to August 2008. Numerical evidence for our method is given by likelihood, see figure 2.

Here we consider three different cases, where in the first and last case γ is fixed to a value of 1 or 2 and estimated as additional parameter by maximum likelihood in the second case. Using asymptotic normality and the variance as estimated from the Hesse matrix (given in small font below the estimates), we can reject the null hypothesis of γ being equal to 2 or

[14]For the proof see Appendix.

[15]See Appendix.

[16]Assuming Box/Tiao error distribution for z_t, as suggested by Hentschel (1995), would result in a very similar but slightly more general representation, but could no more be interpreted as an information-theoretic model for time-varying moments, as defined above.

[17]Compare Bera and Higgins (1993).

Indice	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	$\gamma/\widehat{\gamma}$	$\log L$	AIC
S&P 500	0.00454 (0.00294)	0.05247 (0.01136)	0.94131 (0.01352)	1	-2492.809	4991.618
	0.00493 (0.00254)	0.05985 (0.01087)	0.93451 (0.01219)	1.54308 (0.07286)	-2451.000	4910.000
	0.00538 (0.00244)	0.06573 (0.01100)	0.93030 (0.01153)	2	-2466.152	4938.304
FTSE 100	0.01019 (0.00445)	0.09064 (0.01483)	0.89469 (0.01815)	1	-2419.808	4845.616
	0.01051 (0.00349)	0.10949 (0.01466)	0.87890 (0.01577)	1.7483 (0.08260)	-2352.101	4712.202
	0.01071 (0.00334)	0.11529 (0.01477)	0.87545 (0.01514)	2	-2356.182	4718.364
Nikkei 225	0.00786 (0.00424)	0.05639 (0.01119)	0.93325 (0.01398)	1	-2565.313	5136.626
	0.00912 (0.00385)	0.06904 (0.01095)	0.92100 (0.01239)	1.58373 (0.07801)	-2522.219	5052.438
	0.01076 (0.00401)	0.07915 (0.01120)	0.91231 (0.01171)	2	-2533.541	5073.082

Figure 2: Parameters, estimates, estimated standard errors (in brackets), log-likelihood and AIC for our exemplary data sets.

1 for every data set. Our model receives not only the highest Likelihood value, but also the highest value for the Akaike Information Criterion (AIC), which penalizes the inclusion of the additional parameter γ. So, the generalized model performs best.

Figure 3 shows plots of kernel density of the estimated innovations together with the theoretical distribution of the corresponding model for the S&P 500 data set, where $\gamma = 1$, $\gamma = 1.54308$ and $\gamma = 2$ from the right to left.

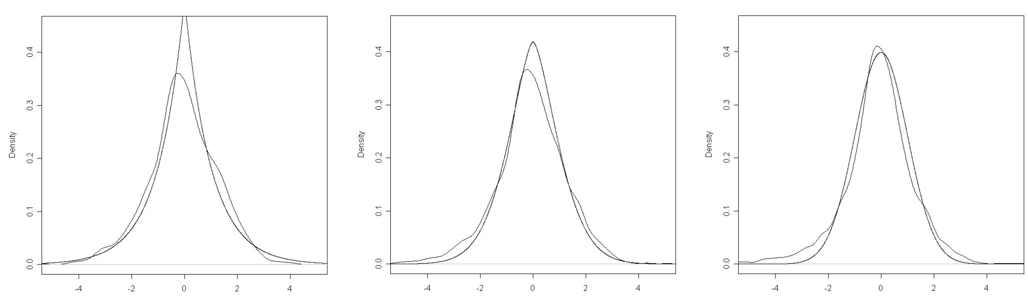

Figure 3: Kernel density from estimated innovations and theoretical density.

We can see that the variance model does better for the innovations in the center, while the generalized model slightly better captures the distribution's tails. Still, even the innovations from the best performing model exhibit skewness as well as kurtosis. Using the approach for time-varying moments, these phenomena could be included through additionally modeling suitable skewness and kurtosis moments. Propositions for such models can be found in Bera and Park (2008) and Fischer and Herrmann (2008).

8 Summary

In this article we present general models for time-varying moments. Using information theory to make up for missing parametric assumptions we are able to show that GARCH models assuming gaussian innovations appear as special case. Applying a model for a time-varying moment to Bickel and Lehmann (1963)'s pth power deviation as a generalized measure of volatility, we give a more general model for time-varying dispersion. Using exemplary data sets and their sample likelihood as criteria we are able to give some empirical evidence for this method.

Appendix

For the Box/Tiao-Error distributions holds that

$$E(|X|^\gamma) = \int_{\mathbb{R}} |v|^\gamma f_{BT}(v)dv = \int_{\mathbb{R}} |v|^\gamma \frac{e^{-\frac{1}{\gamma}(\frac{|v|}{\sigma})^\gamma}}{c(\gamma)\sigma} dv = \frac{1}{c(\gamma)\sigma} \int_{\mathbb{R}} |v|^\gamma e^{-\frac{1}{\gamma}(\frac{|v|}{\sigma})^\gamma} dv =$$

$$= \frac{1}{c(\gamma)\sigma} \int_{\mathbb{R}} \sigma^\gamma \frac{|v|^\gamma}{\sigma^\gamma} e^{-\frac{1}{\gamma}(\frac{|v|}{\sigma})^\gamma} = \frac{\sigma^{\gamma-1}}{c(\gamma)} \int_{\mathbb{R}} \frac{|v|^\gamma}{\sigma^\gamma} e^{-\frac{1}{\gamma}(\frac{|v|}{\sigma})^\gamma} dv,$$

setting $z = \frac{v}{\sigma}$,

$$= \frac{\sigma^{\gamma-1}}{c(\gamma)} \int_{\mathbb{R}} |z|^\gamma e^{-\frac{1}{\gamma}|z|^\gamma} \sigma dz = \frac{\sigma^\gamma}{c(\gamma)} \int_{\mathbb{R}} |z|^\gamma e^{-\frac{1}{\gamma}|z|^\gamma} dz = \sigma^\gamma \frac{c(\gamma)}{c(\gamma)} = \sigma^\gamma,$$

furthermore the distribution is scalable, as

$$Z = g(Z) = \frac{X}{\sigma} \quad with \quad f_X(x;\sigma) = \frac{e^{-\frac{1}{\gamma}(\frac{|x|}{\sigma})^\gamma}}{c(\gamma)\sigma}$$

$$\Rightarrow f_Z(z) = f_{g(X)}(z) = f_X(g^{-1}(z))|\frac{dg^{-1}(z)}{dz}| = \frac{e^{-\frac{1}{\gamma}(z)^\gamma}}{c(\gamma)} = f_X(x;1).$$

References

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

[2] Bera, A. K. and S. Y. Park (2008): "Maximum Entopy Autoregressive Conditional Heteroskedasticity Model", Working Paper.

[3] Bickel, P. J. and E. L. Lehmann (1976): "Descriptive Statistics for Nonparametric Models. III. Dispersion", *Annals of Statistics*, **4**(6): 1139-1158.

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

[5] Borland, L. (2005): "Long-range Memory and Nonexensitivity in Financial Markets", *Europhysics News*, November/December:228-231.

[6] Box, G. E. P. and Tiao, G. C. (1962): "A Further Look at Robustness Via Bayes Theorem", *Biometrika*, **49**:419-432.

[7] Brunt, B. v. (2004): The Calculus of Variations, Springer.

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

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

[10] Fischer, M. and K. Herrmann (2008): "An Alternative Model for Time-varying Moments using Maximum Entropy", Working Paper.

[11] Golan, A. (2002): "Information and Entropy Econometrics Editors View", *Journal of Econometrics*, **107**:1-15.

[12] Golan, A. and E. Maasoumi (2008): "Information Theoretic and Entropy Methods: An Overview", *Econometric Reviews*, **27**(4-6):317-328.

[13] Hentschel, Ludger (1995): "All in the Family Nesting Symmetric and Asymmetric GARCH Models", *Journal of Financial Econometrics*, **39**:71-104.

[14] Higgins, M. L. and A. K. Bera (1992): "A Class of Nonlinear ARCH models", *International Economic Review*, **33**:137-158.

[15] Hull, J. C. (1993): Options, Futures, and other Derivative Securities, Prentice-Hall.

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

[17] Kapur, J.N. (1989): Maximum Entropy Models in Science and Engineering, Wiley and Sons.

[18] Oja, H. (1981): "On Location, Scale, Skewness and Kurtosis of Univariate Distributions", *Scandinavian Journal of Statistics*, **8**:154-168.

[19] Queirós, S. M. D. and C. Tsallis (2005): "Bridging a Paradigmatic Financial Model and Nonextensive Entropy", *Europhysics Letters*, **69**(6):893-899.

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

[21] Shannon, E. T. (1948): "A Mathematical Theory of Communication", *Bell System Technical Journal*, **27**:379-423, continued 623-656.

[22] Wu, X. (2003): "Calculation of Maximum Entropy Densities with Application to Income Distribution", *Journal of Econometrics*, **115**:347-354.

Diskussionspapiere 2008
Discussion Papers 2008

01/2008 **Grimm, Veronika and Gregor Zoettl**: Strategic Capacity Choice under Uncertainty: The Impact of Market Structure on Investment and Welfare

02/2008 **Grimm, Veronika and Gregor Zoettl**: Production under Uncertainty: A Characterization of Welfare Enhancing and Optimal Price Caps

03/2008 **Engelmann, Dirk and Veronika Grimm**: Mechanisms for Efficient Voting with Private Information about Preferences

04/2008 **Schnabel, Claus and Joachim Wagner**: The Aging of the Unions in West Germany, 1980-2006

05/2008 **Wenzel, Tobias**: On the Incentives to Form Strategic Coalitions in ATM Markets