An Example on Modelling Conditional Higher Moments using Maximum Entropy Density with High Frequency Data

[1]**Felix Chan**

[1]School of Economics and Finance, Curtin University of Technology
GPO Box U1987 WA 6845, Australia.
E-Mail:F.Chan@curtin.edu.au

Keywords: *Maximum Entropy Density, Higher Conditional Moments, Conditional Skewness, Conditional Kurtosis*

EXTENDED ABSTRACT

Since the introduction of the Autoregressive Conditional Heteroscedasticity (ARCH) model of Engle (1982), the literature of modelling the conditional second moment has become increasingly popular in the last two decades. This popularity is reflected by the numerous volatility models being proposed in the literature and their multivariate counterparts (see McAleer (2005) for an excellent survey on the various volatility models and related issues on estimation and specification). Interestingly, the Quasi Maximum Likelihood Estimator (QMLE) with normal density is typically used to estimate the parameters in these models. As such, the higher moments of the underlying distribution are assumed to be the same as the normal distribution. However, various studies reveal that the higher moments, such as skewness and kurtosis of the distribution of financial returns are not likely to be the same as the normal distribution, and in some cases, they are not even constant over time. This has significant implications in risk management, especially in the calculation of Value-at-Risk (VaR), which focuses on the negative quantile of the return distribution. Failed to accurately capture the shape of the negative quantile, which is determined by the skewness and the kurtosis of the distribution, would produce inaccurate measure of risk, and subsequently lead to misleading decision in risk management.

This paper proposes a general framework to model the distribution of financial returns using Maximum Entropy Density (MED). The main advantage of MED is that it provides a general framework to estimate the distribution function directly based on a given set of data, and it provides a convenient framework to model higher order moments up to any arbitrary finite order k. However this flexibility comes with a high cost in computation time as k increases, therefore this paper proposes an alternative model that would reduce computation time substantially. Moreover, the sensitivity of the parameters in the MED with respect to the dynamic changes of moments is derived analytically. This result is important as it relates the dynamic structure of the moments to the parameters in the MED.The usefulness of this approach will be demonstrated using 5 minutes intra-daily returns of the Euro/USD exchange rate.

1 INTRODUCTION

Modelling conditional second moment has now become a standard practice in analysing financial time series following the success of the Autoregressive Conditional Heteroscedasticity (ARCH) model of Engle (1982) and the Generalized ARCH (GARCH) model of Bollerslev (1986). A natural extension to modelling time varying second moment is to model the dynamic of higher order moments such as the third and fourth moments, which relate to the skewness and the kurtosis of the underlying distribution, respectively. However, the values of the third and fourth moments are pre-determined by the first and second moments under the standard assumption of normality. Although empirical evidence, such as Mandelbrot (1963) and Mandelbrot (1967), show that the normality assumption is often unrealistic for financial time series, the parameters of most volatility models, especially those belong to the GARCH-family, are typically estimated by Quai Maximum Likelihood Estimator (QMLE) with normal density. Moreover, Value-at-Risk (VaR) forecasts based on the normality assumption often leads to excessive violation due to the restrictive assumptions on the third and fourth moments, see for example da Veiga et al. (2005). This is particular important as VaR has now become a standard tool for forecasting, evaluating and managing market risk (see Jorion (2000)), excessive violation would imply that the VaR forecasts were consistently underestimating market risk, which could lead to devastating consequence for the financial market. Therefore, it is important to improve the methodology for VaR forecasts by accommodating the higher order moment structure and consider more flexible distributions.

The standard approach to relax the normality assumption in the literature is to replace the normal distribution by more flexible distributions, see for examples, Bollerslev (1987), Nelson (1991), Hansen (1994) and Harvey and Siddique (1999). Although these studies considered more flexible distributions with the possibility of time varying higher moments, the distributional assumption may still be too restrictive for the following reasons: (i) these distributions usually only allow the first four moments to be time varying, and the values of the higher moments are subsequently pre-determined by the first four moments; and (ii) the properties of the associated (Quasi) Maximum Likelihood Estimator ((Q)MLE) is unclear, especially if the distribution assumption was violated.

In a seminal paper, Rockinger and Jondeau (2002) proposed to estimate the distribution function directly using Maximum Entropy Density (MED). The main advantage of MED is that it provides a general framework to estimate the density function directly based on a given set of data, and it provides a convenient framework to model higher order moment up to any arbitrary finite order k. That is, it is possible to investigate the dynamic nature of higher order moments up to any finite order k using MED. More importantly, it is possible to determine k before specifying the dynamic structure of the moments using techniques such as those proposed in Wu (2003). However, this flexibility comes with a high cost in computation time as k increases.

The aim of this paper is to propose a flexible framework to estimate MED for financial returns accommodating the dynamic structure of higher order moments. Moreover, the new specification reduces the computation burden substantially relative to the standard approach. It is much simpler to implement in practice and it also avoids various computational issues caused by the Stieltjes and Hamburger moment problems. Moreover, the sensitivity of the parameters in the MED with respect to the changes of moments is derived analytically. This result is important as it relates the dynamic structure of the moments to the parameters in the MED. The empirical usefulness of the model will be investigated using 5 minutes intra-daily Euro/USD exchange rate data.

The paper is organised as follows: Section 2 introduces the concept of Maximum Entropy Density and its estimation methods, with a special emphasis on the computational issue as well as various specifications for the dynamic of higher order moments. A new model will be introduced in Section 3. This is followed by an empirical example in Section 4 and Section 5 contains some concluding remarks. The proofs of all the Propositions are omitted due to page constraints but they are available upon request.

2 MAXIMUM ENTROPY DENSITY

The basic idea of Maximum Entropy Density (MED) is to estimate the density function by maximising a certain entropy functional subject to a set of moment constraints. Shannon (1948) proposed the following entropy functional:

$$E = -\int_{\mathbf{A}} p(x) \log p(x) dx, \qquad (1)$$

to measure the difference in the information content between the density $p(x)$ and the density for the uniform distribution, where $\mathbf{A}$ denotes the appropriate set in which the integration takes place. The motivation of the Shannon's entropy relies on the fact that the uniform distribution is used in the absence of any information and therefore, the distance between $p(x)$ and the uniform density would provide a measure of information content in $p(x)$. Given this interpretation, it would then natural to estimate $p(x)$ by maximising its information content through a set of moment constraints. This is equivalent to maximise equation (1) subject to a set of moment constraints up to order k, namely

$$p(x) = \arg\max_{p} - \int_{\mathbf{A}} p(x) \log p(x) dx$$

subject to

$$\begin{aligned} \int_{\mathbf{A}} p(x) dx &= 1 \\ \int_{\mathbf{A}} x^i p(x) dx &= m_i \quad i = 1, .., k \end{aligned} \tag{2}$$

where m_i denotes the i^{th} raw moment of the distribution. Since the moments of a distribution can be estimated through a given set of data, therefore MED essentially provides a density that capture as much information from the data as the k moments could provide.

The conventional way to solve the optimisation problem (2) is to define the following Hamiltonian:

$$\begin{aligned} H(p) = - \int_{\mathbf{A}} p(x) \log p(x) dx + \lambda_0' \bigg(\int_{\mathbf{A}} p(x) dx - 1 \bigg) + \\ \sum_{i=1}^{k} \lambda_i \left[\int_{\mathbf{A}} x^i p(x) dx - m_i \right], \end{aligned} \tag{3}$$

where the maximisation of $H(p)$ can be easily solved using calculus of variation, which leads to the following closed form solution:

$$p(x) = \exp(\lambda_0) \exp\left(\sum_{i=1}^{k} \lambda_i x^i \right) \tag{4}$$

where $\lambda_0 = \lambda_0' - 1$. Since $\int p(x)dx = 1$, it is straightforward to show that $\exp(\lambda_0) = \left[\int_{\mathbf{A}} \exp\left(\sum_{i=1}^{k} \lambda_i x^i \right) dx \right]^{-1}$. Thus, the MED is defined to be

$$p(x) = Q^{-1} \exp\left(\sum_{i=1}^{k} \lambda_i x^i \right) \tag{5}$$

where $Q = \int_{\mathbf{A}} \exp\left(\sum_{i=1}^{k} \lambda_i x^i \right) dx$. It is straightforward to show that the density of the normal distribution is a special case of equation (5). The following propositions are useful for examining the properties of MED relative to the normal density:

Proposition 1. *For $k = 4$, the Maximum Entropy Density as defined in equation (5) is an even function and hence symmetric around 0 if and only if $\lambda_2 < 0$, $\lambda_4 \leq 0$, $\lambda_1 = 0$ and $\lambda_3 = 0$.*

Proposition 2. *The Maximum Entropy Density as defined in equation (5) is equivalent to a normal distribution, $N(\lambda_1/2\lambda_2, 1/2\lambda_2)$ if and only if $\lambda_2 < 0$ and $\lambda_{it} = 0 \; \forall i = 1, ..., k, i > 2$.*

Proposition 2 implies that if $k = 4$ and $\lambda_3 \neq 0$ or $\lambda_4 \neq 0$ then the Maximum Entropy Density is non-normal.

Obviously, the Lagrange multiplier, λ_i is a nonlinear function of the moments, m_i, for all i. Therefore if there is a dynamic structure underlying the moments, then the λ_i must also be time-varying. This can be seen formally from the following proposition:

Proposition 3. *Let μ_i denotes the i^{th} moment of a MED and λ_i be the parameters in the MED, for $i = 1, .., k$. If $\mu_{2k} < \infty$ then*

$$\frac{\partial \lambda_i}{\partial \mu_j} = (\mu_{i+j} - \mu_i \mu_j)^{-1} \qquad \forall i, j = 1, ..., k.$$

Therefore the values of the language multipliers, λ_i, $i = 1, ..., k$ will be affected by any changes in any of the moments. Thus, if the moment changes over time, then the λ_i will be affected accordingly. Hence, the MED at a given time t is

$$p(x_t) = Q_t^{-1} \exp\left(\sum_{I=1}^{k} \lambda_{it} x_t^i \right) dx_t. \tag{6}$$

where $Q_t = \int_{\mathbf{A}} \exp\left(\sum_{I=1}^{k} \lambda_{it} x_t^i \right) dx_t$. Note that if k is even, then a sufficient condition for $Q_t < \infty$ is $\lambda_{kt} < 0$ for all t. Rockinger and Jondeau (2002) proposed to a set of parametric models for the dynamic of the first four moments, $k = 4$. The parameters in the model can then be estimated by standard Maximum Likelihood approach with density equal to the corresponding MED as given in equation (6). While this approach is conceptually straightforward and easy to understand, it imposes significant computational burden on the estimation of the parameters. Under the time varying assumption of the moments, the estimation of MED would require the computation of λ_{it} for every t at every iteration in the optimisation routine. Since there is no closed-form solution for λ_{it} given a set of moments, $\{m_{it}\}_{i=1}^{k}$, for $k > 2$, the computation of λ_{it} must rely on numerical procedure. This makes the parameter estimation a time consuming exercise, and may not be feasible for large sample set such as those typically seen in financial time series. Another drawback of imposing dynamic structure on the moments directly is related to the Stieltjes and Hamburger moment problems. Essentially, the problem seeks to find the necessary and sufficient condition in which a sequence of number, $\{m_i\}_{i=1}^k$, must satisfy in order to ensure the existence of a proper density function such that its i^{th} moment is m_i for $i = 1, ..., k$. Although such conditions were derived in Mead and Papanicolaou (1984) and Frontini and Tagliani (1997) , it is virtually impossible to restrict the parameters so that the model could always produce a sequence of number that satisfies the conditions for every t.

In order to resolve these issues, this paper propose to model the dynamics of λ_{it} directly. This method only requires the computation of λ_{it} once and therefore significantly reduces the computation time. Model specification and estimation issues will be discussed in the next section.

3 A MODEL

Let $\lambda_t = (\lambda_{1t}, ..., \lambda_{kt})'$ and $m_t = (m_{1t}, ..., m_{kt})'$. Consider the following specification for the dynamic of λ_t:

$$\lambda_t = \Omega + \sum_{j=1}^{p} \Theta_j m_{t-j} + \sum_{l=1}^{q} \Gamma_l \lambda_{t-l} \qquad (7)$$

where Ω is a $k \times 1$ vector, Θ_j and Γ_l are $k \times k$ matrices for $j = 1, .., p$ and $l = 1, ..., q$, respectively. Let $vec()$ denotes the vec operator of a matrix, the parameter vector, $\Theta = (\Omega', vec(\Theta_1)', ..., vec(\Theta_p)', vec(\Gamma_1)', ..., vec(\Gamma_q)')'$, can then be estimated by minimising the Hellinger distance, which leads to the Minimum Hellinger Distance Estimator (MHDE) as follows:

$$\hat{\Theta} = \arg\min_{\Theta} \sum_{t=1}^{T} \int_{\mathbf{A}} (\hat{p}(x_t)^{1/2} - p(x_t)^{1/2})^2 dx_t, \quad (8)$$

where

$$\hat{p}(x_t) = \hat{Q}_t^{-1} \exp\left(\sum_{i=1}^{k} \hat{\lambda}_{it} x_t^i\right) \qquad (9)$$

such that

$$\hat{\lambda}_t = \hat{\Omega} + \sum_{j=1}^{p} \hat{\Theta}_j m_{t-j} + \sum_{l=1}^{q} \hat{\Gamma}_l \lambda_{t-l}$$

$$\hat{Q}_t = \int_{\mathbf{A}} \exp\big(\sum_{i=1}^{k} \hat{\lambda}_i x^i\big) dx.$$

Moreover, $p(x)$ can be constructed using the estimated moments from the sample. Consider the standard raw moment estimator

$$\hat{m}_{it} = t^{-1} \sum_{\tau=1}^{t} x_\tau^i, \qquad \forall i = 1, .., k, \qquad (10)$$

λ_{it} can then be calculated numerically based on $\hat{m}_{it}$. Rockinger and Jondeau (2002) provides an excellent account on the different efficient methods to compute λ_{it} given a set of $\hat{m}_{it}$. More importantly, following the result from Mead and Papanicolaou (1984), Rockinger and Jondeau (2002) proved that the corresponding λ_{it} exists and is unique given a sequence of moments, for all $i = 1, ..., k$.

Given the recent availability of intra-daily data, the MED can be constructed based on sample moments from intra-daily data for every day, then the dynamic of the MED can be modelled directly using equation (7) and the parameter estimates can be obtained through MHDE as defined in equation (8). This is the methodology used for the empirical example in this paper.

Divide the total trading time in a day into h equally spaced intervals and let $p_{t,j}$ and $p_{t,j+\triangle}$ denote the price of a particular asset at the beginning and the end of the j^{th} interval in day t, respectively for all $t = 1, ..., T$ and $j = 1, ..., h$, that is, the price of the asset is being recorded on equally spaced intervals, $h + 1$, times a day, for T days. Note that $p_{t,j+\triangle} = p_{t,j+1}$. Then the return within each interval can be calculated as

$$r_{t,j} = 100 \log(p_{t,j+\triangle}/p_{t,j}), \quad \forall j \geq 2 \qquad (11)$$

which produces h intra-daily returns for all $t = 1, .., T$. The sample moments, and subsequently, the associated λ_{it} and the entropy density of the daily return can then be constructed based on the h intra-daily returns for all $t = 1, ..., T$. Given the set of λ_{it}, the parameters in model (7) can then be estimated using the MHDE as stated in equation (8). The procedure can be summarised into the following steps:

Step. 1 For every $t = 1, ..., T$, calculate the intradaily return using equation (11).

Step. 2 Calculate the k sample moments, m_{it}, $i = 1, ..., k$ using the intradaily returns as calculated in Step (1) for every $t = 1, ..., T$.

Step. 3 Compute λ_{it} for $i = 1, ..., k$ given the sample moment $\hat{m}_{it}$, for every $t = 1, ..., T$.

Step. 4 Given the set of λ_{it}, construct $p(x_t)$ for every $t = 1, ..., T$.

Step. 5 Estimate the parameter vector Θ by minimising the Hellinger distance as defined in equation (8).

Notice this approach allows the specification of λ_{it} to be flexible with minimum computational cost. Using the theorems derived in Beran (1977), (see also Chandra and Taniguchi (2006)), the MHDE for Θ, $\hat{\Theta}$, as defined in equation (8) is shown to be asymptotically normal, that is

$$\sqrt{T}(\hat{\Theta} - \Theta) \stackrel{A}{\rightarrow} N\left(0, 4^{-1}\left[\int_{\mathbf{A}} \left(\frac{\partial \hat{p}(x_t)}{\partial \Theta}\right)\left(\frac{\partial \hat{p}(x_t)}{\partial \Theta}\right)'\Bigg|_{\Theta=\hat{\Theta}} dx\right]^{-1}\right). \tag{12}$$

Given, $\hat{\Theta}$, $\hat{\lambda}_{it}$ can then be calculated using equation (7), and hence the estimated MED can then be constructed using equation (9).

4 EMPIRICAL RESULTS

This section provides an empirical example of estimating MED incorporating higher conditional moment using intradaily data of exchange rate between Euro and US dollar. The data used in this paper was collected through the Philadelphia exchange with the exchange rate being recorded every 5 minutes from 3/1/2005-3/6/2006. As a result, there are 84 intra-daily returns for each day over 330 days, which makes 27,720 observations in total. For this empirical example, $k = 4$, $p = q = 1$ and the coefficient matrices, Θ_1 and Γ_1, are restricted to be diagonal matrices, that is

$$\lambda_{it} = \omega_i + \theta_i m_{it-1} + \gamma_i \lambda_{it-1} \qquad \forall i = 1, .., 4. \tag{13}$$

All the estimation in this paper was conducted using Ox version 4.10 and the computing codes are available upon request. Since most of the integrals required for the estimation do not have closed form solutions, integration must be computed numerically. The implication is that, even though $\mathbf{A} = (-\infty, \infty)$ in the present context, it is necessary to restrict $\mathbf{A}$ into a bounded interval. In this paper, integration were computed using the Newton-Cotes approximation with $\mathbf{A} = [-10, 10]$. The range was chosen so that an unit expansion in the range resulted in less than 10^{-6} changes in the final values in integration.

Figure 1 contains the plots of 330 days sample moments calculated using 84 intra-daily returns for each day, where Figure 2 contains the sample estimates of mean, variance, skewness and kurtosis of the return distribution for each day. As shown in both figures, the mean, the third moment and subsequently the skewness are all centered around 0 indicating the distribution of daily returns are symmetric on average. However, the plots also reveal two negative outliers in the third moment, and subsequently the skewness. Interestingly, these two outliers in the third moment (skewness) match by the two positive outliers in the fourth moment (kurtosis) as shown in Figures 1 and 2. Moreover, the dynamic of the second moment and the variance follow quite closely to the typical financial time series for returns, especially the clustering of high and low volatility.

Figure 3 contains the plots of the corresponding λ_t for $t = 1, ..., 330$. Notice all $\lambda_{4t} < 0$ for all t which satisfies $Q_t < \infty$ for all t and therefore MED exists for every t. Interestingly, while λ_{1t} and λ_{3t} fluctuate around 0 and do not exhibit any definitive pattern, both λ_{2t} and λ_{4t} exhibit pattern very similar to the pattern found in the variance and kurtosis. However, it is important to note that each λ_{it} is a function of all four moments, so it would be misleading to associate λ_{it} with the moment of a particular order.

Table 1 contains the parameter estimates for equation (13) by minimising the Hellinger distance as defined in equation (8) with the corresponding t-ratios in the parenthesis. Interestingly only three parameters are significant, namely, ω_2, ω_4 and θ_2. This has the following implications about the distribution of the daily return for Euro. Firstly, only λ_{2t} evolves over time, as only θ_2 is significant. Secondly, since ω_3, θ_3 and γ_3 are not statistically significant, implying that λ_{3t} is 0 on average. Moreover, both λ_{2t} and λ_{4t} are negative $\forall t$. Therefore, the distribution of the daily return for Euro is symmetric on average. Thirdly, Proposition 2 implies that the distribution of the daily return for Euro is non-normal since ω_4 is statistically difference from 0 indicating λ_{4t} is not 0 on average.

For the purposes of demonstration, Figures 4 and 5 contain the plots of the estimated MED, MED and the distribution of the Euro/USD exchange rate under the assumption of normality for 3/6/2006. As shown in Figure 4, the shape of the estimated MED resembled closely to the MED, which exhibited a small negative skewness which could not be captured by the normal

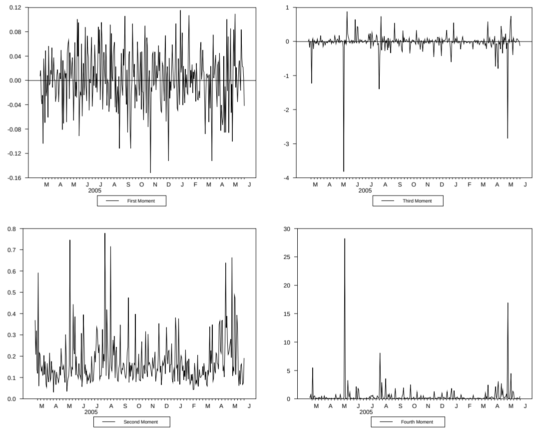

Figure 1. Daily Sample Moments

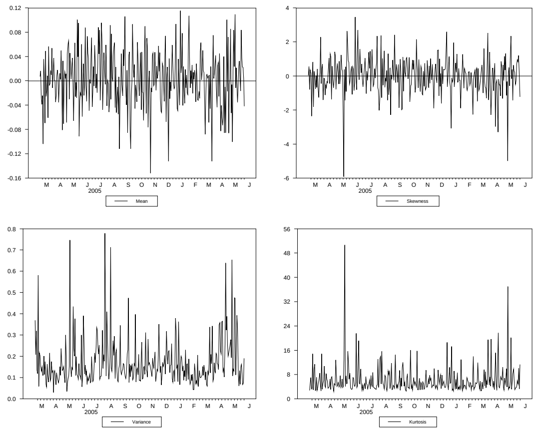

Figure 2. Description Statistics

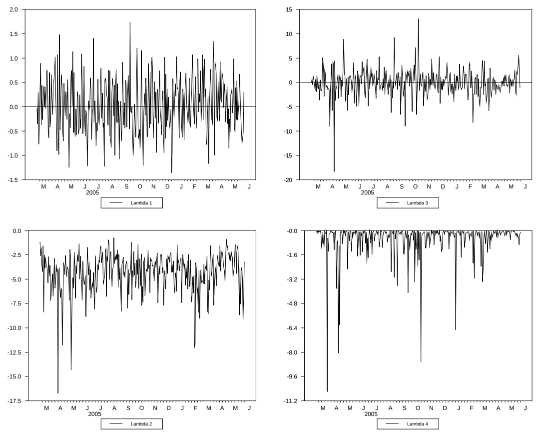

Figure 3. Daily λ's

Parameters	Estimates
$\hat{\omega}_1$	0.0380
	(1.151)
$\hat{\omega}_2$	-4.446**
	(-16.566)
$\hat{\omega}_3$	-0.039
	(-0.258)
$\hat{\omega}_4$	-0.549**
	(-8.356)
$\hat{\theta}_1$	0.052
	(0.877)
$\hat{\theta}_2$	0.192**
	(2.715)
$\hat{\theta}_3$	-0.024
	(-0.427)
$\hat{\theta}_4$	-0.012
	(-0.219)
$\hat{\gamma}_1$	1.251
	(1.717)
$\hat{\gamma}_2$	0.559
	(0.423)
$\hat{\gamma}_3$	0.047
	(0.099)
$\hat{\gamma}_4$	0.023
	(0.708)

Table 1. Parameter Estimates

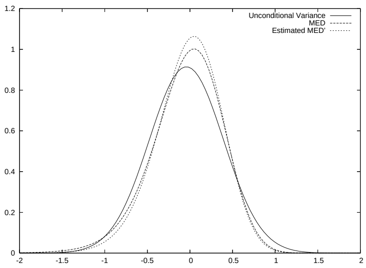

Figure 4. Estimated Maximum Entropy Density

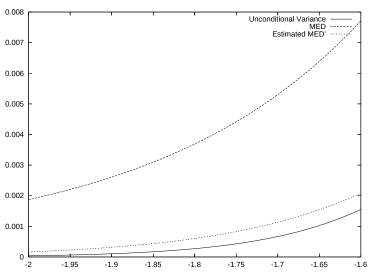

Figure 5. Negative Quantile of the Estimated Maximum Entropy Density

distribution. Moreover, the MED also has a much thicker negative tail than the normal distribution. This implies that the empirical distribution suggested a much higher probability for a negative return than it is implied by the normality assumption. This has also been captured by the estimated MED which has a slightly thicker tail at the negative quantile than the normal distribution. This suggests that the estimated MED as proposed in this paper can capture the probability of negative return more accurately than the standard assumption of normality.

5 CONCLUSIONS

This paper proposed a new method to analyse the distribution of financial time series using Maximum Entropy Density by accommodating potential dynamic structure of higher order moment. The new method is more computational efficient than the conventional MED methods. The dynamic structure of the moments are modelled through their corresponding parameters in the MED. The parameters in the model are then estimated through minimising the Hellinger Distrance (MHDE). The usefulness of this approach was demonstrated by using 5 minutes intra-daily data of the Euro/US exchange rate. The results provide useful insight into the dynamic of the distribution for Euro/US exchange and showed that the method proposed in this paper can improve the accuracy in modelling the probability of negative returns.

6 ACKNOWLEDGEMENT

The author would like to thank Hiroaki Suenaga and Zdravetz Lazarov for insightful discussions. The financial support from the School of Economics and Finance, Curtin University of Technology is gracefully acknowledged.

7 References

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, 5(3):445–463.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69:542–547.

Chandra, S. A. and Taniguchi, M. (2006). Minimum α-divergence estimation for ARCH models. *Journal of Time Series Analysis*, 27(1):19–39.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007.

Frontini, M. and Tagliani, A. (1997). Entropy-convergence in Stieltjes and Hamburger moment problem. *Applied Mathematics and Computation*, 88:39–51.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3):705–730.

Harvey, C. and Siddique, A. (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, 34(4):465–486.

Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36:394–419.

Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, 40:393–413.

McAleer, M. (2005). Automated inference and learning in modeling financial volatility. *Econometric Theory*, 21:232–261.

Mead, L. and Papanicolaou, N. (1984). Maximum entropy in the problem of moment. *Journal of Mathematical Physics*, 25(8):2404–2417.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns - a new approach. *Econometrica*, 59(2):347–370.

Rockinger, M. and Jondeau, E. (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics*, 106:119–142.

Shannon, C. (1948). The mathematical theory of communication. *Bell Systems Technical Journal*, 27:349–423.

da Veiga, B., McAleer, M., and Chan, F. (2005). It pays to violate: model choice and critical value assumption for forecasting value-at-risk thresholds. In Zerger, A. and Argent, R., editors, *MODSIM 2005 International Congress on Modelling and Simulation*, pages 2304–2311.

Wu, X. (2003). Calculation of maximum entropy density with application to income distribution. *Journal of Econometrics*, 115:347–354.