

Testing intra-daily seasonality using Maximum Entropy Density

F. Chan ^a and R. Singh ^a

^a*School of Economics and Finance, Curtin University, GPO BOX U1987, Perth, Western Australia, 6845
Email: ranjodh.singh@curtin.edu.au*

Abstract: The objective of this paper is to test for intra-daily seasonality using Maximum Entropy Density (MED). Specifically, this paper attempts to investigate seasonal patterns over weekdays and through the hours of a given trading day. The MED estimation is essentially a data driven technique which produces a density function. A comparison of MEDs across different time segments of the data allows one to test for the existence of intra-daily seasonality. More importantly with regard to comparisons, an MED presents a richer source of information compared to a singular metric such as the mean or variance. In other words, such a comparison enables one to measure different facets of a distribution. In particular, different moments of the distribution. Repeated patterns in one or more moments between time segments over a period implies the presence of intra-daily seasonality.

MED estimation techniques assume that a random variable is independent and identically distributed (iid). This condition ensures that consistent estimators are produced. However, return data has a correlation structure and as such the observations are not iid. It is proposed that this correlation structure be filtered out prior to the construction of the MED using an ARMA(1,1) - GARCH(1,1) model.

The overall methodology of this paper is as follows. A data series consisting of returns is segmented into weekdays. For example, all the Mondays are extracted from the data. Each Monday segment is referred to as a block. Note that there is a time discontinuity between two consecutive Monday blocks. This has important implications with regard to filtering the correlation structure. An ARMA(1,1) - GARCH(1,1) model is only implemented at the block level since time is continuous within the block. The residuals are standardised and checked for autocorrelation. A MED is computed on the residuals of each block. The first four moment constraints are used for the MED construction. As such, there are four MED parameters i.e. λ_i where $i \in 1, 2, 3, 4$. Subsequently, the mean values for each λ_i are computed over all blocks corresponding to a given weekday. These values represent the final MED parameters for a given weekday. For example, the λ_i values are averaged over all Monday blocks to get the overall MED parameters for the Monday segment. Testing for intra-daily seasonality is done in two parts. Firstly in order to verify the structure of the resulting MED, tests are conducted to assess if the resulting mean values are significantly different from zero. Secondly in order to check for intra-daily seasonality, tests are conducted to assess if the resulting mean values are significantly different across the weekdays. Significant differences in the mean values of λ_i across weekdays indicates that the distribution of returns changes during the week. This pattern over a period of time corresponds to intra-daily seasonality. This process repeated to check for intra-daily seasonality across different time intervals over a trading day.

The results indicate that one of the mean values of the MED parameters for Wednesday is significantly different from the rest of the weekdays. Similarly, one of the mean values of the MED parameters is significant for the 12p.m.- 2p.m. interval.

Keywords: *Maximum Entropy, Moments, GARCH, Time of day effect, Weekday effect*

1 INTRODUCTION

Investors take advantage of seasonal patterns in return data. Traditionally, analysts have searched for these patterns over longer time horizons such as quarterly, bi-annually or yearly time periods. However, there are cases where seasonality exists over a much smaller time horizon such as daily or even hourly time periods. This has been made possible due to the availability of ultra-high frequency data. This type of seasonality is more commonly referred to as intra-daily seasonality. Examples of intra-daily seasonality include the weekday effect, time of the day effect and weekend effect.

French (1980) and Gibbons and Hess (1981) showed that the average daily return for Monday was negative compared to the positive returns for the rest of the weekdays. Rogalski (1984) investigated the behaviour of returns over trading and non-trading periods. Smirlock and Starks (1986) conducted a similar study replacing daily data with hourly data. More recently, Doyle and Chen (2009) introduced the wandering weekday effect which states that the weekday effect is not fixed, but changes over time.

The objective of this paper is to test for intra-daily seasonality using Maximum Entropy Density (MED). Specifically, this paper attempts to investigate seasonal patterns over weekdays and through the hours of a given trading day. The MED derivation produces a density function. A comparison of MEDs across different time segments of the data allows one to test for the existence of intra-daily seasonality.

Shannon (1948) proposed the idea of entropy as a measure of the amount of uncertainty or randomness. Following that Jaynes (1957) proposed the principle of maximum entropy. This allowed one to maximize Shannon's entropy subject to certain moment conditions. This was presented as a non-linear optimization problem. The solution to this problem became known as the maximum entropy density. Zellner and Highfield (1988) focused on the computational aspects of this problem and provided some practical examples. Rockinger and Jondeau (2002) used the entropy principle to develop a method to model time varying conditional moments. Chan (2009) proposed a similar method which was more computationally efficient and modelled the structure of the moments in terms of the MED parameters.

Section 2 provides some brief details about the MED and its advantages. This section also outlines the practical considerations of implementing this technique. Section 3.1 shows the properties of the data used in this study. Sections 3.2 and 3.3 provide evidence whether or not intra-daily seasonality exists over different time segments of the data. Finally, section 4 summarizes the major findings of this study along with its limitations.

2 METHODOLOGY

The continuous version of the entropy presented in Shannon (1948) and Jaynes (1957) is defined as

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

where $p(x)$ is a probability density function and \mathbf{A} represents the set in which the integration occurs. The principle of Maximal Entropy involves maximizing E subject to various moment conditions. The moment conditions are as follows:

$$\int_{\mathbf{A}} p(x) dx = 1 \quad (2)$$

$$\int_{\mathbf{A}} x^i p(x) dx = m_i \text{ where } i = 1, 2, \dots, k. \quad (3)$$

In the above equations, m_i represents the i^{th} raw moment of the distribution. Solving this non-linear optimization problem yields the following solution:

$$p(x) = Q^{-1} \exp \left(\sum_{i=1}^k \lambda_i x^i \right) \quad (4)$$

where $Q^{-1} = \int_{\mathbf{A}} \exp \left(\sum_{i=1}^k \lambda_i x^i \right) dx$. This quantity denotes the normalizing constant which ensures that the first condition (equation (2)) is satisfied. The λ_i values represent parameters of the MED. These are essentially functions of the moments (see Chan (2009)). Using equation (2) as the only moment condition produces the uniform distribution. This is expected since no other information is used in constructing the density function.

As such, the resulting distribution allocates equal probabilities to each outcome. As additional information is added i.e. adding more moment conditions moves the resulting distribution away from the uniform distribution. Furthermore, varying the structure of the moment conditions such as using log moment conditions or absolute value of moments etc. produces different classes of distributions. For further details on this topic refer to Park and Bera (2009).

From the above derivation, one can see that the resulting density is the generalized exponential distribution (equation (4)). The MED estimation is essentially a data driven technique which estimates the parameters (λ_i) of this distribution function. More importantly with regard to comparisons, an MED presents a richer source of information compared to a singular metric. In other words, such a comparison enables one to measure different facets of a distribution. In particular, different moments of the distribution. Repeated patterns in one or more moments between time segments over a period implies the presence of intra-daily seasonality.

Straight forward computation of the MED for return data is problematic since MED estimation techniques assume that a random variable is independent and identically distributed (iid). This condition ensures that consistent estimators are produced. However, return data has a correlation structure and as such the observations are not iid. It is proposed that this correlation structure be filtered out prior to the construction of the MED. Consider the following model:

$$\phi_r(L)r_t = \mu + \theta_s(L)\epsilon_t \tag{5}$$

$$\epsilon_t = \eta_t \sqrt{h_t} \quad \eta_t \sim \text{iid}(0, 1) \tag{6}$$

$$h_t = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^q \beta_i h_{t-i}. \tag{7}$$

Equation (5) represents the Autoregressive Moving Average (ARMA) model where r and s denote the order of autoregressive and moving average parts of the model respectively. It attempts to capture the dynamics of the mean of the return data r_t . In equation (5), μ represents the drift term and $\phi_r(L) = 1 - \phi_1(L) - \phi_2(L) - \dots - \phi_r(L)$ is the autoregressive operator where L is the lag operator such that $Ly_t = y_{t-1}$. Similarly, $\theta_s(L) = 1 + \theta_1(L) + \theta_2(L) + \dots + \theta_s(L)$ represents the moving average operator. Both these polynomials have their unit roots outside the unit circle and share no common roots. This model assumes that the conditional variance of residuals is constant over time. However, estimating the ARMA model for the return data shows that the residuals are not constant over time. Furthermore, the ARCH test results are significant for the first and second lags. This suggests evidence of GARCH effects. Therefore, the variance of the process is modelled using a GARCH(p, q) model (equation (7)) where p and q represent the order of the autoregressive and moving average parts of the model respectively. Bollerslev (1986) introduced the GARCH model whilst extending the work of Engle (1982) on ARCH models. This allows one to model a time varying conditional variance. An ARMA(1,1) - GARCH(1,1) model is used to filter out the time dynamics of the return data. Higher orders did not result in an improved fit. The residuals of this model are standardised and checked to ensure that no autocorrelation is present in the first and second moment. Subsequently, the MED is estimated from these residuals.

MED estimation assumes the existence of higher moments in $p(x)$. In practise, sample moments are estimated and used in place of population moments. These estimates are derived from finite data and as a consequence always provide finite moments. Whilst a large sample maybe helpful in alleviating some of this concern, this does not guarantee the existence of higher moments. Berkes(2003) suggests a method of determining the highest finite moment for given set of data. This method basically relies on estimating the tail index and thus enables one to justify the use of higher moment conditions in the MED estimation.

The overall methodology of this paper is as follows. A data series consisting of returns is segmented into weekdays. For example, all the Mondays are extracted from the data. Each Monday segment is referred to as a block. Note that there is a time discontinuity between two consecutive Monday blocks. This has important implications with regard to filtering the correlation structure. An ARMA(1,1) - GARCH(1,1) model is only implemented at the block level since time is continuous within the block. The residuals are standardised and checked to ensure that no autocorrelation exists. A MED is computed using the standardised residuals of each block. The first four moment constraints are used for the MED construction. As such, there are four MED parameters i.e. λ_i where $i \in 1, 2, 3, 4$. Subsequently, the mean values for each λ_i are computed over all blocks corresponding to a given weekday. These values represent the final MED parameters for a given weekday. For example, the λ_i values are averaged over all Monday blocks to get the overall MED parameters for the Monday segment. Testing for intra-daily seasonality is done in two parts. Firstly in order to verify the structure

of the resulting MED, tests are conducted to assess if the resulting mean values are significantly different from zero. Secondly in order to check for intra-daily seasonality, tests are conducted to assess if the resulting mean values are significantly different across the weekdays. Significant differences in the mean values of λ_i across weekdays indicates that the distribution of returns changes during the week. This pattern over a period of time corresponds to intra-daily seasonality. This process repeated to check for intra-daily seasonality across different time intervals over a trading day.

3 EMPIRICAL RESULTS

3.1 DATA

The data used in this study consists the exchange rate between the US and Australian dollar. This data is sourced from the SIRCA database. The data series is captured at a one minute frequency from the 30th of May 2008 to the 1st of February 2012. The returns are calculated using the bid price. Figure 1 illustrates a sample of the return data.

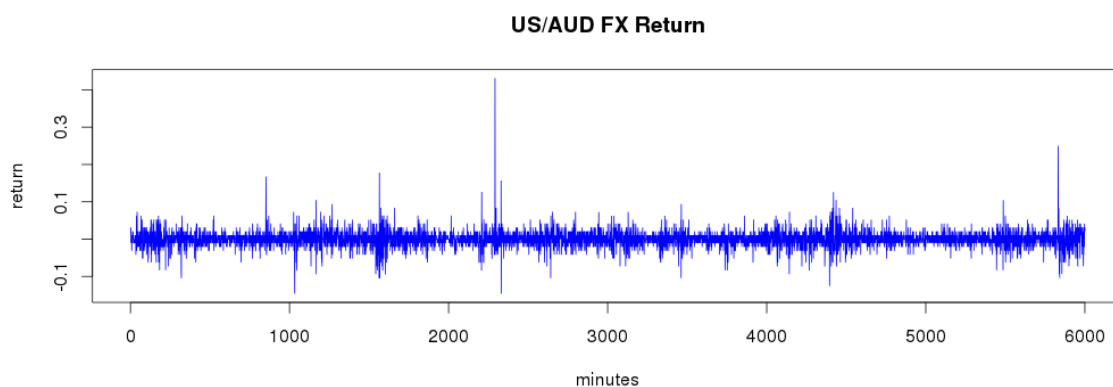


Figure 1. Sample: US/AUD FX Return Data

Observations that occur in non-trading hours¹ are excluded from the data set. The data is segmented into weekdays. Each weekday consists of 192 blocks, each containing 360 observations. Table 1 shows the summary statistics of the weekday returns over all the blocks.

Table 1. Weekday summary statistics

Weekday	Minimum	Q1	Median	Mean	Q3	Maximum
Monday	-0.9664	-0.0196	0.0000	0.0002	0.0196	0.6122
Tuesday	-0.6609	-0.0204	0.0000	0.0002	0.0209	1.1370
Wednesday	-0.8187	-0.0209	0.0000	0.0000	0.0208	2.5496
Thursday	-13.9569	-0.2109	0.0000	-0.0004	0.0207	0.7478
Friday	-0.6654	-0.0204	0.0000	0.0002	0.0206	0.8563

In order to test for intra-daily seasonality within a trading day, three time intervals are introduced. These are 10a.m. to 12p.m., 12p.m. to 2p.m. and 2p.m. to 4p.m. Similar to the weekday segmentation, each time intervals consists of 959 blocks, each containing 120 observations.

As mentioned in Section 2, an ARMA(1,1) - GARCH(1,1) model is implemented on each block (weekday and time interval). The residuals of each block are standardised and checked for autocorrelation. The moments

¹Trading hours are assumed to be 10a.m. to 4p.m. weekdays

of these residuals are computed as part of MED estimation. Table 2 provides the summary statistics for the skewness and kurtosis of the standardised residuals for the weekdays over all blocks. In the subsequent sections, the MED parameters derived from these statistics are shown.

Table 2. Weekday moment data

Weekday	Moment	Min	Q1	Median	Mean	Q3	Max
Monday	Skewness	-3.1519	-0.1187	0.0307	0.0123	0.1939	2.1901
	Kurtosis	2.8961	3.6611	4.2172	5.1861	5.1082	49.945
Tuesday	Skewness	-1.9224	-0.1648	0.0116	0.0073	0.1668	3.3627
	Kurtosis	3.0591	3.6942	4.2753	5.7201	5.3201	42.9433
Wednesday	Skewness	-1.5613	-0.1234	0.0242	0.0933	0.2298	4.2119
	Kurtosis	3.0731	3.8502	4.2851	5.9034	5.2064	59.4311
Thursday	Skewness	-23.1089	-0.2194	0.0099	-0.1222	0.1968	2.5607
	Kurtosis	2.8301	3.9681	4.6131	8.6342	6.1283	536.011
Friday	Skewness	-5.3876	-0.2045	0.0221	0.0306	0.2062	5.9749
	Kurtosis	2.7491	4.1522	5.0972	8.3374	7.6032	84.3871

3.2 WEEKDAY EFFECT

As mentioned in the section 2, the first test is to assess whether or not the final MED parameters for a given weekday are significantly different from zero. Equation (8) states the hypothesis of the test where $\bar{\lambda}_{i,w}$ represents the mean value of the λ_i over all blocks corresponding to a given weekday w .

$$\begin{aligned}
 H_0 & : \bar{\lambda}_{i,w} = 0 \\
 H_1 & : \bar{\lambda}_{i,w} \neq 0
 \end{aligned}
 \tag{8}$$

The test statistic for this hypothesis test is

$$t = \frac{\bar{\ell}_{i,w} - 0}{s_{\ell_{i,w}} / \sqrt{n_{i,w}}}
 \tag{9}$$

where $\bar{\ell}_{i,w}$ is the sample estimate of $\bar{\lambda}_{i,w}$. The standard deviation of ℓ_i over all blocks corresponding to a given weekday w is given by $s_{\ell_{i,w}}$ and $n_{i,w}$ denotes the number of blocks corresponding to ℓ_i for a given weekday w . This test statistic follows a Student's t -distribution if the null hypothesis is supported. The results of this test are shown in table 3 below. This table shows the actual values of the final MED parameters along with their corresponding t test statistic (in brackets).

Table 3. MED parameters for weekdays

	Monday	Tuesday	Wednesday	Thursday	Friday
ℓ_1	0.0058 (0.7712)	0.0112 (1.5200)	-0.0081 (-1.1837)	0.0205 (2.4501)	-0.0027 (-0.3595)
ℓ_2	-0.7927 (-69.2149)	-0.7895 (-65.1616)	-0.8117 (-69.5279)	-0.8162 (-26.8868)	-0.7997 (-59.6185)
ℓ_3	0.0011 (0.2502)	-0.0046 (-1.1495)	0.0053 (1.5008)	-0.0046 (-1.0745)	0.0026 (0.6091)
ℓ_4	0.0000 (0.2237)	0.0001 (0.6389)	0.0007 (4.8581)	-0.0001 (-0.1813)	0.0000 (0.3128)

The results indicate that the $\bar{\lambda}_2$ values are significant for all weekdays. Additionally, the value of $\bar{\lambda}_4$ is significant for Wednesday. These results show the structure of the resulting MED. As a consequence, differences in these structure translate to differences in the cyclical behaviour of returns across the weekdays. A second test is conducted to assess whether or not the $\bar{\lambda}_i$ values differ across the weekdays. Equation (10) states the hypothesis of this test.

$$\begin{aligned}
 H_0 & : \bar{\lambda}_{i,w} = \bar{\lambda}_{i,w^*} \text{ where } w \neq w^* \\
 H_1 & : \bar{\lambda}_{i,w} \neq \bar{\lambda}_{i,w^*}
 \end{aligned}
 \tag{10}$$

To test whether the means are different, the t-test statistic is

$$t = \frac{\bar{\ell}_{i,w} - \bar{\ell}_{i,w^*}}{\sqrt{\frac{s_{\ell_{i,w}}}{n_{i,w}} + \frac{s_{\ell_{i,w^*}}}{n_{i,w^*}}}}. \quad (11)$$

This test assumes that the population variances are not equal. The results of this test indicate that the values of $\bar{\lambda}_2$ values are not significantly different across all the weekdays. Secondly, $\bar{\lambda}_4$ value for the Wednesday is significantly different from the other weekdays. This result is as expected, since the MED pertaining to Wednesday has a different structure compared to the rest of the weekdays.

3.3 TIME OF THE DAY EFFECT

Similar tests are carried out for the three time intervals over a trading day. The first test is to assess whether or not the final MED parameters for the three time intervals are significantly different from zero. Equation (12) states the hypothesis of the test where $\bar{\lambda}_{i,t}$ represents the mean value of the λ_i over all blocks corresponding to a given time interval t .

$$\begin{aligned} H_0 &: \bar{\lambda}_{i,t} = 0 \\ H_1 &: \bar{\lambda}_{i,t} \neq 0 \end{aligned} \quad (12)$$

The corresponding test statistic for this hypothesis test is of the same form as equation (9). The results of this test are shown in table 4 below. This table shows the actual values of the final MED parameters along with their corresponding t test statistic (in brackets).

Table 4. MED parameters for time intervals

	10p.m. to 12p.m.	12p.m. to 2p.m.	2p.m. to 4p.m.
ℓ_1	0.0051 (1.2105)	0.0047 (0.9949)	-0.0049 (-1.0951)
ℓ_2	-0.8131 (-176.2982)	-0.8067 (-155.4148)	-0.8163 (-172.8275)
ℓ_3	-0.0010 (-0.4546)	-0.0008 (-0.3169)	0.0027 (1.1892)
ℓ_4	-0.0001 (-1.1783)	-0.0004 (-2.6468)	-0.0001 (-0.9468)

The results indicate that the $\bar{\lambda}_2$ values are significant for all time intervals. Additionally, $\bar{\lambda}_4$ value is significant for 12p.m. to 2p.m. time interval. These results show the structure of the resulting MEDs. As a consequence, differences in these structure translate to a differences in the behaviour of returns across the time intervals. A second test is conducted to assess whether or not the $\bar{\lambda}_i$ values differ across the time intervals. Equation (13) states the hypothesis of this test.

$$\begin{aligned} H_0 &: \bar{\lambda}_{i,t} = \bar{\lambda}_{i,t^*} \text{ where } t \neq t^*. \\ H_1 &: \bar{\lambda}_{i,t} \neq \bar{\lambda}_{i,t^*} \end{aligned} \quad (13)$$

The corresponding test statistic for this hypothesis test is of the same form as equation (11). The results of this test indicate that the values of $\bar{\lambda}_2$ values are not significantly different across all the time intervals. Although $\bar{\lambda}_4$ is significant for the $t = 12\text{p.m. to } 2\text{ p.m.}$ time slot, it is not significantly different when compared to other intervals.

4 CONCLUSION

This paper has introduced a method for testing intra-daily seasonality using a MED. This methodology is purely data driven and does not rely on any distributional assumptions. The resulting density allows for a more richer comparison across different time segments (weekdays or time intervals). Specifically, one is able to check for differences in higher moments. This is especially important when differences in lower moments are not significant. This is precisely the result found in this paper. It shows that the $\bar{\lambda}_4$ value for Wednesday is significantly different from the rest of the weekdays. But the remaining $\bar{\lambda}$ values are not significantly different across the weekdays. Similarly, the $\bar{\lambda}_4$ value is significant for the 12p.m. to 2p.m. interval. Lastly, the results are limited to properties of the data used in this study.

ACKNOWLEDGMENTS

The authors would like to thank the two anonymous referees for their helpful suggestions. The authors are grateful for the financial assistance provided by the Australian Research Council.

REFERENCES

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307 – 327.
- Chan, F. (2009). Modelling time-varying higher moments with maximum entropy density. *Mathematics and Computers in Simulation* 79(9), 2767–2778.
- Doyle, J. R. and C. H. Chen (2009, August). The wandering weekday effect in major stock markets. *Journal of Banking & Finance* 33(8), 1388–1399.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), pp. 987–1007.
- French, K. R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics* 8(1), 55 – 69.
- Gibbons, M. R. and P. Hess (1981). Day of the week effects and asset returns. *The Journal of Business* 54(4), pp. 579–596.
- Jaynes, E. T. (1957, May). Information theory and statistical mechanics. *Phys. Rev.* 106, 620–630.
- Park, S. Y. and A. K. Bera (2009). Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics* 150(2), 219–230.
- Rockinger, M. and E. Jondeau (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics* 106(1), 119–142.
- Rogalski, R. J. (1984). New findings regarding day-of-the-week returns over trading and non-trading periods: A note. *The Journal of Finance* 39(5), pp. 1603–1614.
- Shannon (1948). The mathematical theory of communication. *Bell Systems Technical Journal* 27, 349–423.
- Smirlock, M. and L. Starks (1986). Day-of-the-week and intraday effects in stock returns. *Journal of Financial Economics* 17(1), 197 – 210.
- Zellner, A. and R. A. Highfield (1988, February). Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *Journal of Econometrics* 37(2), 195–209.