An empirical investigation of Australian Stock Exchange Data.

William Bertram
School of Mathematics and Statistics
University of Sydney

January 27, 2004

Abstract

We present an empirical study of high frequency Australian equity data examining the behaviour of distribution tails and the existence of long memory. A method is presented that allows us to deal with Australian Stock Exchange data by splitting it into two separate data series representing an intraday and overnight component. Power law exponents for the empirical density functions are estimated and compared with results from other studies. Using the autocorrelation and variance plots we find there to be a strong indication of long memory type behaviour in the absolute return, volume and transaction frequency.

1 Introduction

The past decade has seen an explosion in the popularity of financial modelling. Covering a rich variety of disciplines, mathematicians and physicists, have brought together many techniques and constructions from their fields of study to investigate and classify the behaviour of financial markets. Recent advances in the availability of high frequency data has opened the door to an increasing number of empirical studies into these complex systems, with the aim of gaining a better understanding of the true nature of the market's behaviour leading to more advanced and realistic models.

This paper contributes to this aim through an empirical examination of the behaviour of a large collection of high frequency equity data on the Australian Stock Exchange (ASX) spanning the period January 1993 through July 2002. The fundamental rules of the ASX will be shown to influence the price dynamics of the securities that trade on it. Markets such as the ASX allow for price development during non-trading periods, such as after hours trading and dual-listed stocks. A new approach for working with securities that trade on exchanges similar to the ASX will be introduced. This approach amounts to representing the stock returns as two separate stochastic processes, a 'discontinuous' overnight return process and a 'continuous' intraday return process. We examine the distributional properties of the returns, trade volume, and transaction frequency, estimating the tail indicies empirically. Using the method of *variance plots* described in [10] we detect the existence of long memory in the absolute return, trading volume and transaction frequency. Our analyses are compared to results found in studies by [2], [3] and [7] and indicate that the behaviour of Australian equities is significantly different from the reported behaviour of other financial markets.

The rest of the paper is organised as follows. Section 2 describes the data, operations of the ASX and how these operations influence the behaviour of the data. This section contains a new approach for the analysis and modelling of financial data, which provides greater insight when dealing with ASX stock data. Section 3 contains an empirical investigation of the distributional properties of the ASX data. We examine the tails of the distributions for returns, absolute returns, volume and number of transactions. Section 4 examines the correlation structure of the returns, volume and transactions. We use variance plots to detect and estimate long memory in the studied series. Section 5 reviews the results and concludes.

2 Data

The data set contains a record of every transaction that took place over the period January 1993 to July 2002, for each of the 200 most actively traded [1] stocks on the ASX. If p_t represents the price of a stock at time t then we define the return over interval size τ as

$$r_t = \log(p_t) - \log(p_{t-\tau}). \tag{1}$$

We also define V_t to be the volume traded in the interval $(t-\tau, t]$ and N_t to be the transaction frequency in $(t-\tau, t]$. The ASX operates using the Stock Exchange Automated Trading System (SEATS). SEATS is an electronic order book that trades continuously between the hours of 10am and 4pm Monday to Friday. Before the opening and after the close, the market enters a special *'pre open'* mode, where orders may be entered, adjusted and cancelled but not executed until a fixed time determined by the ASX. In the morning this starts at 07:00, giving three hours for traders to adjust trades to compensate for overnight information flow . A similar process occurs after the close at 16:00 until a random time between 16:05 and 16:06,when a single market auction takes place to clear the order book and set the official closing price. This is performed in order to reduce the end of day volatility and the possibility of market manipulation by large market participants. The ASX allows for after hours trades between brokers, requiring that they report their activities to the ASX. Further, the exchange contains several dual listed stocks trading on exchanges such as London or New York. These ASX operations clearly impact on the price forming process over non-trading hours. This is significant as it affects the way we analyse and model our data. In Fig. 1 we show a typical month of trading for Rio Tinto sampled every 10 minutes. The price shows large changes occurring during the overnight/closed interval. If we compare the time series for the 10 minute returns over the whole period 1993-2002, as seen in Fig. 2, with the returns where the overnight price jumps have been removed, we immediately see the impact the overnight returns have on the series as a whole and can speculate about the large effect they must have on a stock's volatility. This leads to the interesting notion of treating the stock price as two separate processes: an intraday process and an overnight process.

[1] Stocks were ranked by total turnover during the studied period.

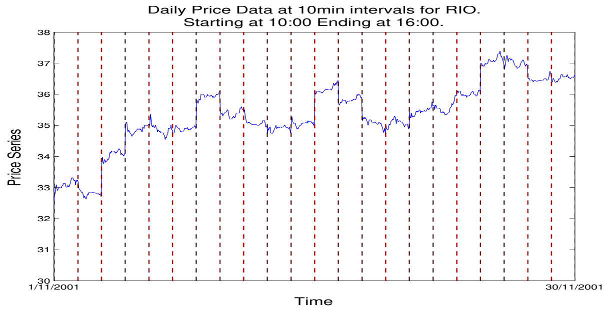

Figure 1: Month of trading of Rio Tinto (RIO). The dashed lines correspond to the opening and closing times. Apparent in this plot are the large jumps in price taking place between the close and open of consecutive days.

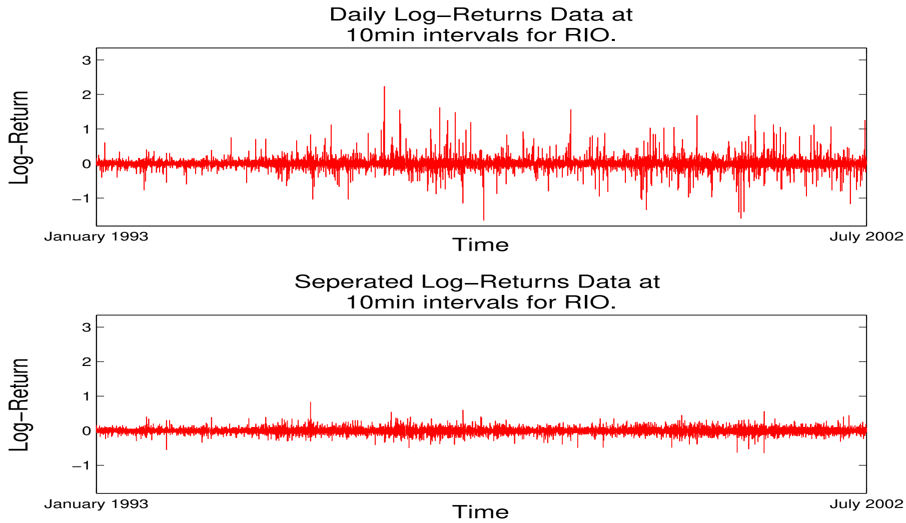

Figure 2: The return process r_t for Rio Tinto (RIO) from 1993 to 2002 calculated in for 10 minute intervals. The top picture shows the return process including the overnight return, while the bottom picture has had the overnight return removed.

We propose that an equity return on the Australian Stock Exchange consists of two processes, one that drives the stock during trading times and another that operates during non-trading times. The implication is that because Australia is a relatively small market, then the general behaviour of the ASX will be influenced mainly by the overnight values since any major market shifting information will arrive from the large markets of US and Europe while Australian markets are closed. It is apparent that the overnight process is discrete and should no longer be modelled by a continuous process. We term this process of overnight jumps the 'Jump Process'. In comparison to the Jump Process, the intraday traded process is more like a continuous time process. It has a multitude of scales and looks more like a classical random noise process. Consequently we term it the 'Noise Process'. Hence when we look at daily price series we see the total stock price consisting of mainly the overnight jump process plus a smaller contribution due to the intraday noise.

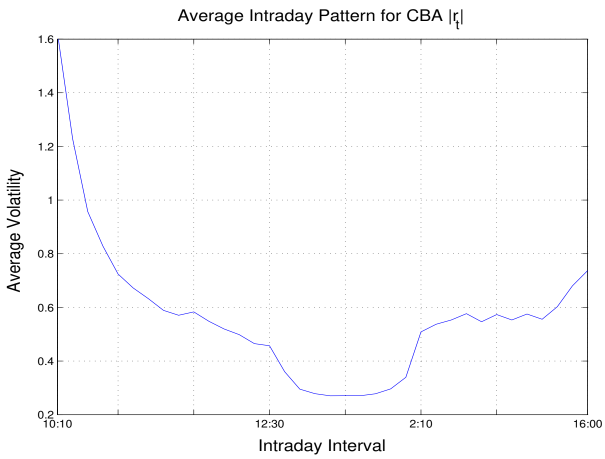

Figure 3: The average absolute value of Commonwealth Bank (CBA) returns, r_t, averaged over 10 years ($\approx$ 2500 days). The pattern shows a deterministic pattern of behaviour of ASX traders.

Turning our attention to the Noise Process, we find that this process contains some remarkably consistent behaviour. Averaging $|r_t|$ at each intraday interval gives us a measure of the intraday volatility over a day. We find that a well defined intraday volatility pattern emerges, shown in Fig. 3. The volatility starts high at the opening and drops off rapidly over the morning as fund managers move quickly to correct their positions due to the overnight jump in information. Conversely, in the afternoon the volatility rapidly increases in anticipation of the close. The above pattern will also be reinforced by the presence of the many so called *'day traders'* on the ASX, whose practice is to close out all their positions at the end of each trading day and reopen their positions the following morning. The rationale of *day traders* is to avoid overnight exposure to risk. Interestingly this plot also provides us with a picture of the social behaviour of ASX equity traders. The volatility can be seen to drop off suddenly after the interval 12.20-12.30pm and pick up again on the interval 14.00-14.10pm. These times correspond to the close of options trading on the ASX and is typically the preferred lunchtime of most traders. Examination of the average volume traded or the average transaction frequency during each intraday interval yields similar measures for the intraday trading activity, shown in Fig. 4. The data must be corrected for these intraday trends as any failure to do so will result in the analysis of correlation and dependence structure being dominated by these strong periodic trends. We have found that the intraday trend can be successfully removed by using the methods described in [1]. Choosing a suitable measure for the intraday trading activity, a new time scale $T(t)$ is constructed such that the activity level is on average a constant for all intervals in $T(t)$. Data is then sampled at the new constantly spaced intervals of $T(t)$, to provide a time series with the intraday activity spread evenly through each day.

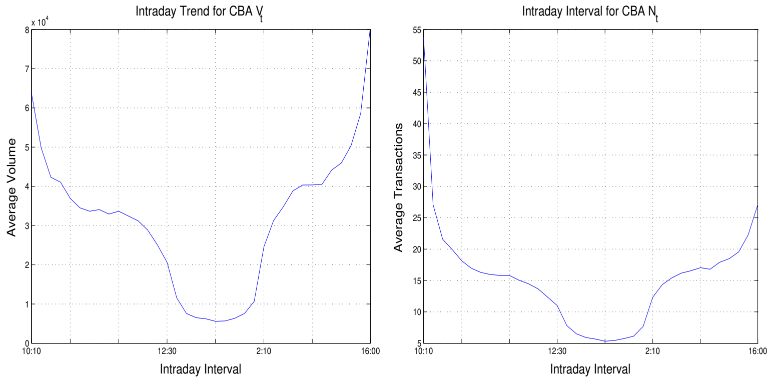

Figure 4: The average value of both volume, V_t, and transactions, N_t, averaged over 10 years ($\approx$ 2500 days) of Commonwealth Bank (CBA) data. The pattern shows a deterministic pattern of behaviour of ASX traders.

3 Distributional properties

Distributional properties of financial data have been studied in many forms over the past 50 years. Mandelbrot [9] and Fama [11] were early challengers to the assumption of the normality of returns, introducing Lévy stable distributions. The drift away from Gaussian behaviour was continued by authors such as Clark [6], Praetz [8], and more recently authors such as Gopikrishnan et.al. [4] and Gorski et.al. [7] have found tails of financial data corresponding to power-law behaviour with exponent exceeding that of the Lévy regime. Studying the empirical distributions of ASX data has revealed that the returns series contains an unusually high proportion of zero values. This behaviour was also reported by [7] who termed the effect 'zero return enhancement'. We have found that this effect is produced in three ways: no trade taking place in the sampling interval, trading at a constant price across the interval, and starting and ending the interval at the same price (with a deviation in price in between). Currently this effect is the subject of ongoing research and is not examined in this study where we focus on the distribution tails. In this section we examine the tails of the cumulative distribution function for the returns $r_t \sim x^{-\alpha_{r_t}}$, absolute returns $|r_t| \sim x^{-\alpha_{|r_t|}}$, volume $V_t \sim x^{-\alpha_{V_t}}$ and transaction frequency $N_t \sim x^{-\alpha_{N_t}}$.

In Fig. 5 we show the left and right tails of the intraday returns r_t for two typical stocks in the data set, Commonwealth Bank (CBA) and Rio Tinto (RIO), sampled with $\tau = 10$ minutes. Estimates for the power-law index for our data set yields values in the range $\alpha_{r_t} \approx 3.6$. This value is well outside that of the Lévy Stable Distributions. Table 1 shows the estimated power-law values for 10 stocks in the dataset. Least squares fitting was performed over different tail ranges (measured in standard deviations). The best power-law behaviour, measured using Pearsons correlation coefficient, was found by fitting the distribution tails across the range of $3 < \sigma < 15$ standard deviations as shown in Table 1. Increasing the sampling time, τ, we find that the power-law index remains unchanged as shown in Fig. 6. This result shows the scale invariance of the ASX data and conflicts with the findings of studies such as [7] that find tail exponents do change with τ in DAX returns. For the overnight

St.Dev:	$1 < \sigma < 15$	$2 < \sigma < 15$	$3 < \sigma < 15$	$4 < \sigma < 15$	$5 < \sigma < 15$
Security					
ANZ	-3.0 (0.989)	-3.4 (0.994)	-3.6 (0.995)	-3.8 (0.995)	-3.9 (0.994)
AGL	-2.9 (0.987)	-3.2 (0.990)	-3.4 (0.990)	-3.6 (0.987)	-3.2 (0.983)
CBA	-2.9 (0.992)	-3.2 (0.997)	-3.4 (0.997)	-3.4 (0.994)	-3.3 (0.990)
RIO	-3.1 (0.983)	-3.6 (0.989)	-3.9 (0.990)	-4.2 (0.989)	-4.4 (0.985)
NCP	-2.7 (0.984)	-3.1 (0.989)	-3.3 (0.987)	-3.6 (0.987)	-3.9 (0.982)
WBC	-3.2 (0.986)	-3.6 (0.992)	-3.9 (0.993)	-4.3 (0.995)	-4.5 (0.993)
FGL	-3.5 (0.977)	-4.0 (0.982)	-4.4 (0.982)	-4.7 (0.980)	-5.1 (0.976)
CSR	-2.9 (0.988)	-3.3 (0.992)	-3.5 (0.991)	-3.6 (0.987)	-3.8 (0.986)
BIL	-2.7 (0.989)	-3.0 (0.994)	-3.2 (0.994)	-3.4 (0.995)	-3.4 (0.993)
SGB	-3.0 (0.990)	-3.3 (0.994)	-3.5 (0.994)	-3.5 (0.987)	-3.3 (0.984)
Av.Value:	-3.0 ± 0.2	-3.4 ± 0.2	-3.6 ± 0.3	-3.8 ± 0.4	-3.9 ± 0.5

Table 1: Estimates of the tail index, α_{r_t}, for r_t, taken over different ranges of standard deviation with corresponding values for Pearsons r in brackets. Average values for α_{r_t} were calculated over the total data set and are shown with 99% error bars.

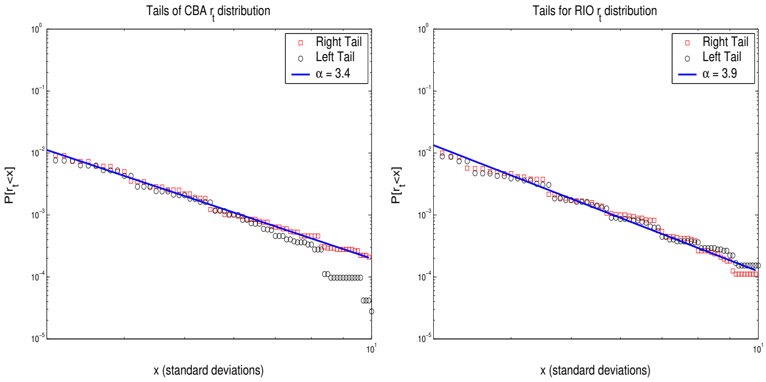

Figure 5: The above plots show the left and right tails for CBA and RIO, along with estimates for the power law index. The plot for CBA, on the left, shows the left tail decreasing much more rapidly than the right tail.

jump process we cannot estimate the distribution tails as the number of data points in this series is of the order ~ 2000 samples and is too low.

Of more interest to financial researchers is the behaviour of $|r_t|$, V_t and N_t as these series are proxies for the *volatility*, and hence measures of risk for the market. For $|r_t|$, V_t and N_t we find power-law index values of $\alpha_{|r_t|} \approx 3.6$, $\alpha_{V_t} \approx 3.4$ and $\alpha_{N_t} \approx 3.0$, as shown in Fig. 7 and Table 2. Interestingly the observed values for the ASX data disagree with values observed in other data sets [3], [4]. As a result, a recent explanation of financial power-law behaviour proposed in [2], based on the behaviour of heterogeneous agents, is not validated by our analysis. In particular, that study hypothesises that for large trading volumes it holds that $r_t \simeq k\sqrt{V_t}$, with k a constant. This is clearly not in agreement with the behaviour detected on the ASX. We feel this finding to be

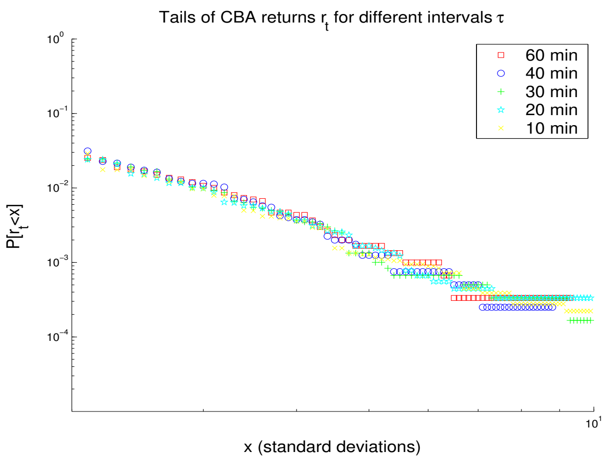

Figure 6: This plot shows the distribution tails for CBA for different time interval lengths τ. We can see that there is little change in the distributional properties over sampling interval.

Tail Index:	$\alpha_{\lvert r_t\rvert}$	α_{V_t}	α_{N_t}
Security			
ANZ	-3.6	-3.9	-3.3
	(0.995)	(0.999)	(0.999)
AGL	-3.4	-2.6	-2.7
	(0.990)	(0.997)	(0.999)
CBA	-3.4	-3.8	-3.2
	(0.996)	(0.999)	(0.999)
RIO	-3.9	-3.7	-4.4
	(0.990)	(0.997)	(0.998)
NCP	-3.3	-3.6	-2.9
	(0.987)	(0.999)	(0.998)
WBC	-3.9	-2.9	-4.0
	(0.993)	(0.996)	(0.999)
FGL	-4.4	-2.8	-2.7
	(0.982)	(0.999)	(0.999)
CSR	-3.5	-2.8	-4.0
	(0.991)	(0.999)	(0.999)
BIL	-3.2	-2.1	-2.9
	(0.994)	(0.995)	(0.998)
SGB	-3.5	-2.5	-3.8
	(0.994)	(0.998)	(0.999)
Av.Value:	-3.6 ± 0.3	-3.1 ± 0.5	-3.4 ± 0.5

Table 2: Estimates of the tail indexes for $|r_t|$,V_t and N_t Calculated over the range of standard deviation $3 < \sigma < 15$ with corresponding values for Pearsons r in brackets. Average values were calculated over the total data set and are shown with 99% error bars.

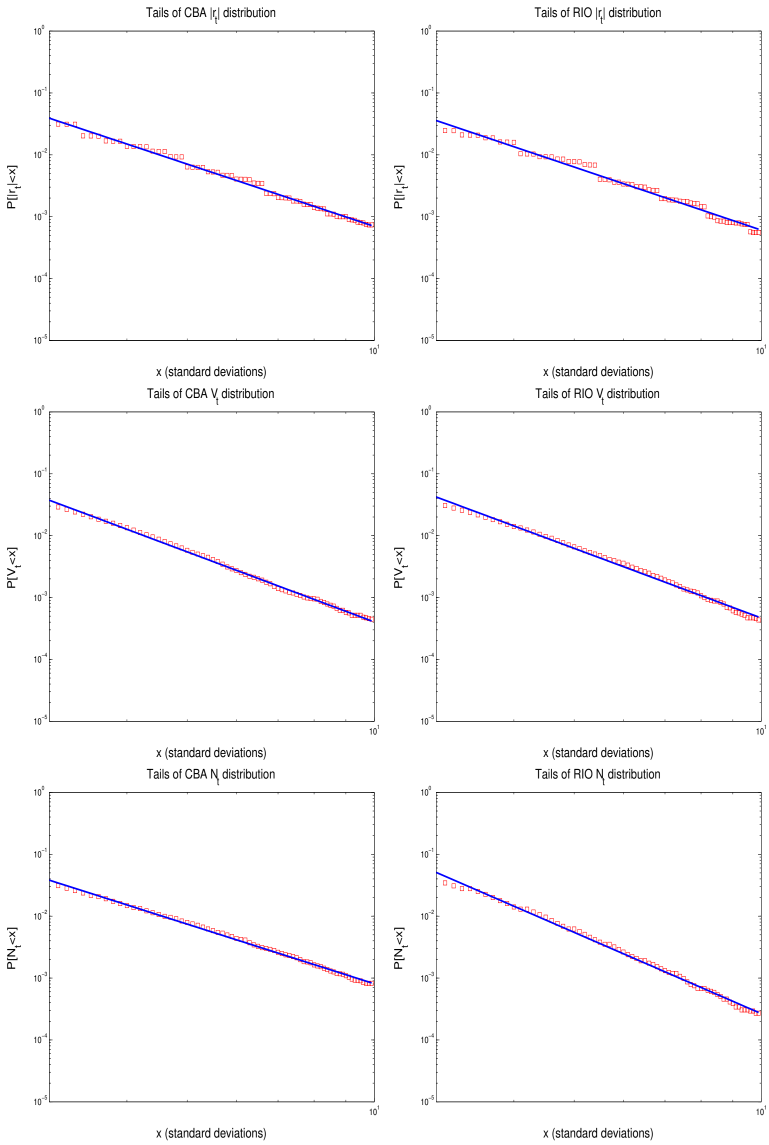

Figure 7: The above plots show the tails for $|r_t|$, V_t, N_t for both CBA and RIO. These series are proxies for the market volatility and all show strong power law type behaviour.

significant and worthy of further investigation.

4 Time Correlation

The analysis of a process' time correlations is necessary in order to properly classify its behaviour. Fig. 8 shows the autocorrelation of overnight and intraday returns for a typical stock in the data set. From these results we may be drawn to conclude stock returns are independent as proposed in [11]. However as

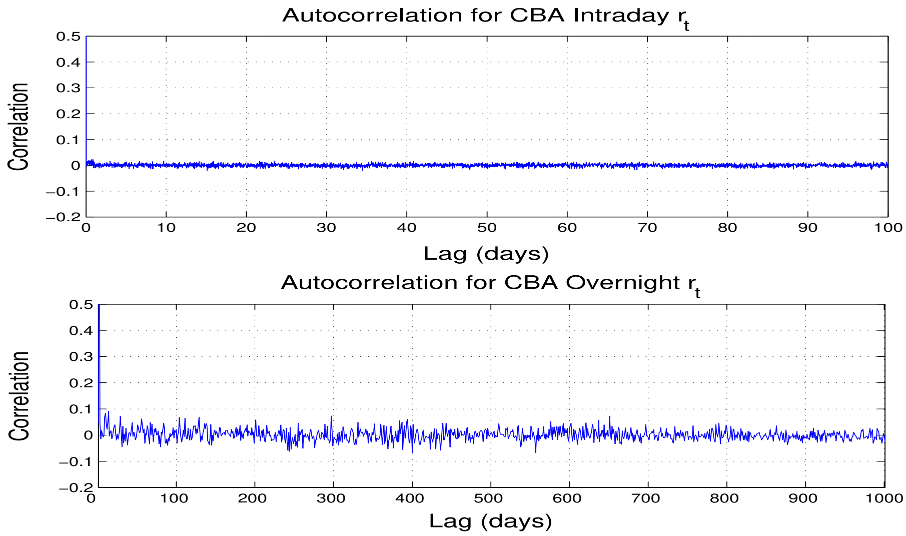

Figure 8: Autocorrelations for the Intraday (top) and Overnight (bottom) processes r_t. These both indicate the absence of any first order correlation structure.

mentioned in [4], the observed distributions appear to be scale invariant while lying outside the regime of Stable processes. This implies that the assumptions of the Central Limit Theorem are being violated in some way. Looking at autocorrelations of $|r_t|$ for the overnight and intraday data (Fig. 9) we find a strong positive autocorrelation in both series. This indicates that the stock returns though uncorrelated are not independent. For intraday values of V_t and N_t the autocorrelation shows similar positive values to that of $|r_t|$.

A process with long or infinite memory is defined as having autocorrelation function,

$$\rho(k) \sim c_\rho |k|^{-\beta} \tag{2}$$

with c_ρ a constant and k is the lag. A process with this correlation structure indicates the dependence between far apart events diminishes very slowly with increasing lag. A process can be tested for such a correlation structure by examining the variance of the process' sample mean [10].

Recall that the variance of the sample mean of a time series X can be represented in terms of its auto-correlation,

$$\mathbb{V}(\bar{X}) = n^{-2}\sigma^2 \sum_{i,j=1}^{n} \rho(i,j) \tag{3}$$

where n is the length of X, $\sigma = \mathbb{V}(X)$ and $\rho(i,j)$ is the auto-correlation matrix of X. If $\rho(i,j)$ only depends on $k = |i-j|$ then the process is said to be *stationary* and we may write

$$\mathbb{V}(\bar{X}) = n^{-1}\sigma^2 \left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho(k)\right] \tag{4}$$

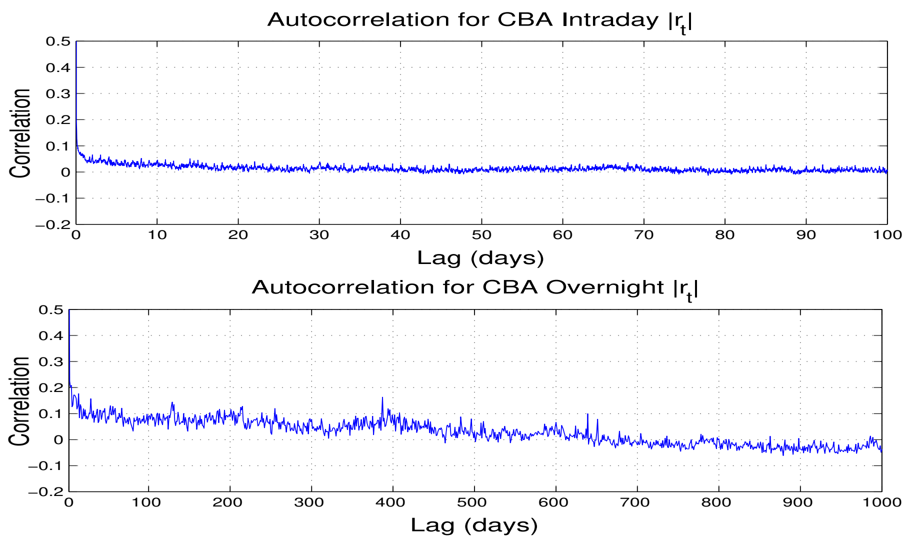

Figure 9: Autocorrelations for the absolute Intraday (top) and Overnight (bottom) processes $|r_t|$. In contrast to the returns process, these series show a strong correlation structure.

$$= \frac{\sigma^2}{n} + \frac{\sigma^2}{n^\beta} n^{\beta-1} 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \rho(k) \tag{5}$$

For large n with $\rho(k) \sim c_\rho |k|^{-\beta}$ this becomes

$$\mathbb{V}(\bar{X}) \sim \sigma^2 c(\rho) n^{-\beta} \quad \text{with} \quad c > 0 \tag{6}$$

As a test for long memory we have plotted the variance of the sample mean for different lengths n in Fig. 10 for ANZ Bank (ANZ). From these plots we are able to estimate values for the long memory parameter β for r_t, $|r_t|$, V_t and N_t using least-squares. In Table 3 we present the estimates of β for 10 stocks in the data set. For the returns process, r_t, we find a the value $\beta \sim 1$, indicating a lack of long memory in this process. However, the β values for $|r_t|$, V_t and N_t provide a strong indication of the presence of long memory in these processes.

5 Summary

This study has carried out an empirical investigation of high frequency equity data for the Australian Stock Exchange over the period January 1993 to July 2002. We examined the return series, its absolute value, the volume traded and the transaction frequency for both previously reported and unreported behaviour. This behaviour was compared to that found to exist in other studies.

It was demonstrated how, due to the time-zone of the Australian market, the ASX returns can be represented as two separate processes for the overnight and intraday periods. Further, it was found that while the intraday returns themselves contained no seasonal trend, the absolute return, volume and number of transactions all display strong periodic behaviour. This periodic behaviour is

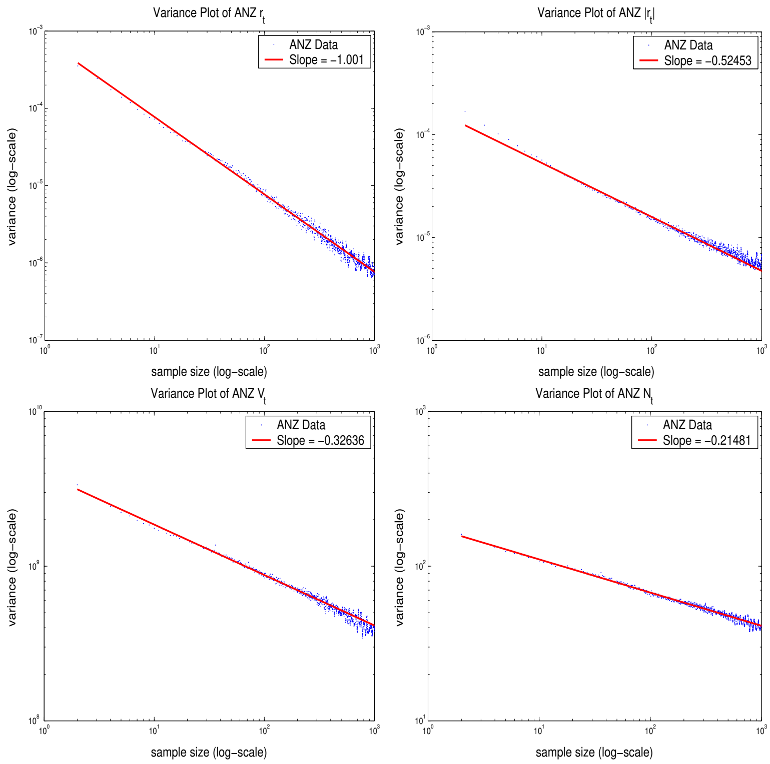

Figure 10: The variance plots shown above give estimates for the long memory parameter, β for r_t, $|r_t|$, V_t and N_t with ANZ Bank (ANZ) data. The top left plot shows $\beta \approx 1$ for r_t, which corresponds to an uncorrelated/short memory process. The other plots are indicative of the presence of long memory type behaviour.

due in part to microstructure effects unique to the ASX. Thus, we would expect studies on different exchanges to yield different results as no two exchanges operate under the same conditions. The extent to which market regulations affect the results of the commonly performed analysis can only be determined by a wider study across several markets.

We examined the tail behaviour of the empirical distributions of the data and found that ASX equities appear to possess power law type behaviour, consistent with that found in other studies. The estimated power law exponents were found to be significantly different from those presented in the literature [3], [4] on other markets. Of particular note, the traded volume was found to possess a tail index of more than twice the value reported for the S&P500. Also, the relative values of the estimated tail indicies for r_t, $|r_t|$, V_t and N_t were found to be different than those found in previous studies making the ASX incompatible with the explanation of the so-called cubic and half-cubic laws as proposed in [2]. It remains to be seen if the different power law behaviour found on the ASX is market/regional specific.

The correlation structure of the ASX data was investigated for long-memory behaviour using the property that the variance of the sample mean for such a

	β_{r_t}	$\beta_{\|r_t\|}$	β_{V_t}	β_{N_t}
Security				
ANZ	1.001	0.525	0.326	0.215
AGL	0.842	0.480	0.301	0.215
CBA	0.900	0.533	0.392	0.280
RIO	0.926	0.497	0.506	0.302
NCP	0.967	0.395	0.340	0.308
WBC	0.950	0.515	0.454	0.246
FGL	0.931	0.501	0.424	0.330
CSR	1.031	0.506	0.502	0.301
BIL	1.030	0.430	0.484	0.255
SGB	0.923	0.521	0.419	0.371
Av.Value:	0.950 ± 0.048	0.490 ± 0.037	0.415 ± 0.060	0.282 ± 0.040

Table 3: Values for long memory parameter β for r_t, $|r_t|$, V_t and N_t.

process decays faster than n^{-1}, n being the length of the sample. Results showed that while the return process appeared to be uncorrelated with exponent $\beta_{r_t} \approx 1$ the absolute returns, volume and number of transaction all displayed behaviour consistent with that of a long memory process.

References

[1] M.M. Dacorogna, R. Gencay, U.A. Muller, R.B. Olsen, O.V. Pictet, *An Introduction to High Frequency Finance*. Academic Press, London, 2001.

[2] X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, A theory of power-law distributions in financial market fluctuations, *Nature*,Vol 423 (2003), pp267-270.

[3] X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, Understanding the cubic and half-cubic laws of financial fluctuations, *Physica A*,324 (2003), pp1-5.

[4] P. Gopikrishnan, V. Plerou, Y. Liu, L.A.N. Amaral, X. Gabaix, H.E. Stanley, Scaling and correlation in financial time series, *Physica A*,287 (2000), pp362-373.

[5] R.N. Mantegna, H.E. Stanley, *Introduction to Econophysics: Correlation and Complexity in Finance*, Cambridge University Press, Cambridge, 1999.

[6] P.K. Clark, A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*, 41, pp 135-159, 1973.

[7] A.Z. Gorski, S. Drozdz and J. Speth, Financial multifractality and its subtleties: an example of DAX,*Physica A*, 316 (2002), pp496-510.

[8] P.D. Praetz, The Distribution of Share Price Changes. Journal of Business, 45, pp 49-55,1972.

[9] B. Mandelbrot, H.M. Taylor, On the Distribution of Stock Price Differences,*Operations Research*, 15, pp 1057-1062.

[10] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall/CRC, London, 1994.

[11] E.F. Fama, The Behaviour of Stock-Market Prices, *Journal of Business*, 38,pp 34-105.