THE EFFICIENT MARKET HYPOTHESIS AND THE DYNAMIC BEHAVIOR OF SUGAR FUTURE PRICES

Ricardo Chaves Lima[1]
Alberto Ohashi[2]

RESUMO: o presente trabalho examina a hipótese de eficiência de mercado estimando um modelo de series temporais com coeficientes que variam no tempo. A metodologia utilizada foi a do Filtro de Kalman para um modelo autoregressivo e de média móvel (ARMA) com erros modelados com heteroscedasticidade condicional autoregressiva generalizada (GARCH). O modelo foi estimado pelo método da máxima verossimilhança para os retornos dos preços futuros do açúcar. As variáveis de maior ordem de defasagem foram as que mostraram a maior queda em valor absoluto no tempo; o que pode sugerir que variáveis com maior ordem de defasagem perdem peso no tempo a medida que os mercados tornam-se mais eficientes.

Palavras-chave: Séries Temporais, Eficiência de Mercado e Filtro de Kalman

SUMMARY: this paper examines the market efficiency hypothesis by estimating time-varying coefficients using Kalman Filter techniques for an Autoregressive Moving Average model (ARMA) with Generalized Autoregressive Conditional Heteroskedasticity (GARCH) errors. The estimation technique utilized was maximum likelihood for sugar future returns. The higher lag order variable coefficients were the ones which showed greater fall in absolute value over time, which may suggest that variables with higher lags may lose weight as markets get more efficient.

Keywords: Time Series Analysis, Market Efficiency and Kalman Filter

[1] Professor Adjunto do Departamento de Economia/PIMES – UFPE (rlima@npd.ufpe.br)
[2] Departamento de Economia - UFPE, PET/CAPES (amfo@decon.ufpe.br)

1. Introduction

The behavior of financial markets has been a theme of ample discussion amongst economists and financial market experts. Discussions about the Efficient Market Hypothesis (EMH) and the random walk hypothesis versus dependence are some topics of great weightiness in recent studies. Themes of great interest and equally controversial, they are very important for the study of market microstructure.

The origins of the (EMH) can be traced back at least as far as the pioneering theoretical contribution of Bachelier (1900) and empirical research of Cowles (1933). The modern literature begins with Samuelson (1965), whose contribution is neatly summarized in his article "Proof that Properly Anticipated Prices Fluctuate Randomly". Price changes must be unforecastable if they are properly antecipated, i.e., if they fully incorporate the expectations and information of all market participants. Fama (1970) summarizes this idea by writing: "A market in which prices always 'fully reflect' available information is called 'efficient' ". Fama's use of quotation marks around the words "fully reflect" indicates that these words are a form of shorthand and need to be explained more fully. More recently, Malkiel (1992) has offered a more explicit definition which can be summarized as follows: "A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set ... "

Fundamentally, it is said that the capital market is efficient if: a) all security prices fully reflect all known market information, and b) no traders in the market have monopoly control of information. There are three possibilities of efficient market: 1) a strong form, which encompasses all information, including that possessed by insiders; 2) a semi-strong form, which includes all public information; and 3) a weak form, which includes only that information which can be gleaned from an examination of an historical series of security prices. Specifically, the future prices reflect the action of producers, consumers and speculators about the price of a commodity at a later date. To be of value to hedgers, the futures prices must respond quickly and accurately to relevant new information. The concept of efficiency referred in this paper concerns to the weak form of efficiency.

An autoregressive memory of a time series model indicates how fast information is processed by economic agents in the market. That is, a long memory model shows a slower assimilation of information, whereas a short memory model shows a faster assimilation of market information. This paper intends to examine the behavior of the autoregressive memory of sugar future returns over time. The methodological procedure used is the estimation of an autoregressive model with Kalman filter time-varying coefficients to show the dynamic behavior of the return's autoregressive memory. Thus, this paper has the aim of rising some questions about the efficiency in commodity futures markets, especifically for sugar future prices, suggesting that the increasing market efficiency is related to decreasing autoregressive memory.

This paper is organized as follows. In Section 2, it is presented a discussion about the random walk hypothesis and some considerations about commodity futures markets. In sections 3 it is presented the methodology and in section 4 it is presented the empirical results. Section 5 presents the conclusions of the study.

2. The Random Walk Hypothesis, Efficient Markets and Brownian Motion

With respect to the commodity markets, we can find some results rejecting the random walk hypothesis. Taylor (1985), for several series of futures prices, strongly supports the conclusion that a small amount of relevant information is reflected slowly by prices, causing price trends. Another interesting result presented were the substantial changes in standard deviations from contract to contract. Cargill and Rausser (1975) show the strong evidence that the random walk must be rejected as a realistic description of commodity markets. That is, the random walk model does not represent a reasonably accurate explanation of commodity market behavior. Leuthold (1972), using spectral analysis and filter rules, examines the live cattle futures markets indicating that a simple stochastic process (random walk) appears consistent with the price behavior of some of the contracts but not with others.

Peterson, Ma and Ritchey (1992) presented evidences of dependence in commodity price using variance ratio test proposed by Lo and MacKinlay (1988). They investigated 17 commodity spot-prices and identified three theoretical components in commodity price: 1) a systematic component reflecting price drift or the expected arrival of information; 2) a negatively autocorrelated component that is attributed to the bid-ask spread of market makers; and 3) a noise term that represents the pricing of unexpected information. The variance ratio test rejected the random walk hypothesis since many short-term realized returns exhibit either positive or negative persistence over different time horizons. Another aspect that they concluded is that the positive serial correlation between successive price changes goes beyond the structure of the underlying fundamentals. For many commodities (especially grains and other crops), there is some evidence that positive serial correlation exists in price changes over short and intermediate time horizons.

The investors are constantly subjected to a vast quantity of diverse information and, as we saw in the first section, the concept of (EMH) is very close to the idea of a market processing quickly and efficiently the information as they arrive to the market. In the early commentaries of (EMH), the statement that the current price of a security "fully reflects" available information was assumed to imply that successive price changes are independent. That is, price changes can be determined only by new information. Thus, today's market returns are unrelated to yesterday's returns, as that information has already been processed. In addition, it was usually assumed that successive changes (or returns) are identically distributed.

The random walk hypothesis states that present and past prices cannot be used to find a more accurate forecast of the next price than today's price. There is then no correlation between the price changes on different days and no information in past prices useful for forecasting future prices. Moreover, the concept of market rationality, which is consistent with the random walk theory, asserts that assets are priced by traders who use all available information to make unbiased predictions of future prices.

Perhaps the simplest version of the random walk hypothesis is the case of an independently and identically distributed (i.i.d) disturbance in which the dynamics of $[P_t]$ are given as follows:

$$P_t = P_{t-1} + e_t, \quad e_t \sim \text{i.i.d } N\,(0,\, s^2) \quad (1)$$

where *(0, s²)* denotes that ε_t is distributed with mean 0 and variance σ^2. The independence of the disturbance [ε_t] implies that the random walk is also a "fair game". The "fair game" model just says that the conditions of market equilibrium can be stated in terms of expected returns, and thus it says little about the details of the stochastic process generating returns. Independence implies not only that disturbances are uncorrelated, but that any nonlinear functions of the disturbances are also uncorrelated.

The derivation and definitions given below, taken from Hamilton (1994) and Campbell, Lo and MacKinlay (1997) will show the relation between the random walk model and the brownian motion. Consider again the equation (1),

$$P_t = P_{t-1} + e_t, \quad e_t \sim \text{i.i.d } N(0, 1)$$

If the process is started with $P_0 = 0$, then it follows that

$$P_t = e_1 + e_2 + \ldots + e_t$$

$$P \sim N(0, t)$$

Moreover, the change in the value of P between dates t and s,

$$P_s - P_t = e_{t+1} + e_{t+2} + \ldots + e_s,$$

is itself $N(0, (s-t))$ and is independent of the change between dates r and q for any dates $t < s < r < q$.

Consider the change between P_{t-1} and P_t. The disturbance e_t was taken to be $N(0,1)$. Suppose we view e_t as the sum of two independent Gaussian variables:

$$e_t = e_{1t} + e_{2t}, \quad e_{it} \sim \text{i.i.d } N\left(0, \tfrac{1}{2}\right)$$

We can then associate e_{1t} with the change between P_{t-1} and the value P at some interim point ($P_{t-(1/2)}$) and e_{2t} with the change between $P_t - P_{t-(1/2)}$ as follows

$$P_{t-(1/2)} - P_{t-1} = e_{1t} \qquad (2)$$

$$P_t - P_{t-(1/2)} = e_{2t} \qquad (3)$$

Sampled at integer dates t = 1,2, ... , the process of (2) and (3) will have the same properties as (1), since

$$P_t - P_{t-1} = e_{1t} + e_{2t} \sim \text{i.i.d. } N(0,1)$$

These process (2) and (3) also relate at noninteger dates $\left\{t + \frac{1}{2}\right\}_{t=0}^{\infty}$, and retains the property for both integer and noninteger dates that $P_s - P_t \sim N(0, s-t)$ with $P_s - P_t$ independent of the change over any other nonoverlapping interval.

By the same reasoning, we could imagine partitioning the change between $t-1$ and t into N separate subperiods:

$$P_t - P_{t-1} = e_{1t} + e_{2t} + \ldots + e_{Nt} \, ,$$

with e_{it} ~ i.i.d. $N(0{,}1/N)$. The result would be a process with all the same properties as (1), defined at a finer and finer grid of dates as N increases. The limit as $N \to \infty$ is a continuous-time process known as continuous-time random walk or standard brownian motion, that has a central role in modern derivative pricing models[3] and in the context of (EMH). The value of this process at date t is denoted $W(t)$ that sometimes is called as a Wiener process[4]. A realization of a continuous-time process can be viewed as a stochastic function, denoted $W(\cdot)$, where $W : t \in [0, \infty) \to \Re^1$.

Say that $W(\cdot)$ is a continuous-time stochastic process, associating each date $t \in [0, T]$ with the scalar $W(t)$ such that:

a) For any t_1 and t_2 such that $0 \le t_1 \le t_2 \le T$:

$$W(t_2) - W(t_1) \sim N\left(m\,(t_2 - t_1), s^{\,2}\,(t_2 - t_1)\right)$$

b) For any t_1, t_2, t_3 and t_4 such that $0 \le t_1 \le t_2 \le t_3 \le t_4 \le T$, the increment $W(t_2) - W(t_1)$ is statistically independent of the increment $W(t_4) - W(t_3)$.

c) The sample paths of $W(t)$ are continuous.

If we set $m = 0$ and $s = 1$, we obtain standard Brownian motion which we shall denote by $Z(t)$. Accordingly, we may re-express $W(t)$ as

$$W\;(t) \;=\; m\,t \,+\, s\,Z\,(t)\,, \qquad t \in [0, T]$$

This continuous-time process is closely related to the discrete-time versions of the random walk described and, as we can note, the discrete-time random walk can be defined as a sequence of continuous-time process which converges to a continuous-time analog of the random walk in the limit.

Thus, the change of the financial returns is called "Brownian" if it vary randomly so that: a) The motion at any one time is independent of the motion at any other time. That is, it has no "memory" of which way it was going a little while ago; b) The expected change over time is zero. It doesn't have "preferred" direction in which to drift; c) The expected distance of the change is greater than zero. In other words, it doesn't just sit still !

---

[3] See, Merton (1990)
[4] See, Neftci (1996)

3. Methodological Procedure

The data set used in the present work correspond to a series of future prices returns for sugar during the period of January 17, 1985 to June 15, 1998, summing up to a total of 3318 observations. The origin of the data is the Coffee, Sugar and Cocoa Exchange (New York).

The returns of an asset price (P_t) can be calculated following Cooper (1982). Consider the following identity:

$$P_{t+1} / P_t = exp\ [log_e\ (P_{t+1} /P_t)]$$

Which is similar to

$$P_{t+1t} = P_t\ exp[log_e\ (P_{t+1}/P_t)]$$

Moreover, considering A the sum to which P dollars will amount after t periods at a continuous rate of return r is, A is given by the following expression:

$$A = P\ exp\ (rt)$$

After 1 time period (t = 1), the expression above could be compared in a way that the returns r can be represented as follows:

$$r = log_e\ (P_{t+1}/P_t)$$

To examine if returns are uncorrelated, the autocorrelation function was calculated, and the statistical test was performed according to Box, Jenkins and Reisel (1994). The autocorrelation function is

$$r(k) = \frac{Cov[r_t, r_{t+k}]}{\sqrt{Var[r_t]}\sqrt{Var[r_{t+k}]}} = \frac{Cov[r_t, r_{t+k}]}{Var[r_t]} = \frac{g(k)}{g(0)}$$

The statistical test for $r(k)$ is given by the Ljung-Box-Pierce Q-statistics, which can be represented as follows:

$$Q = T(T+2)\sum_{k=1}^{K}(T-k)^{-1}\,\hat{r}_k^{\,2}$$

For T observations and k lags. The Q statistic is asymptotically c^2 distributed with s degrees of freedom (Enders, 1995). The null hypothesis is of no autocorrelation. The Q statistic was used as a random walk test for the daily returns.

The empirical time series was assumed to follow an ARMA process with heteroskedastic errors. Statistical tests were performed to examine stationarity, and the Box-Jenkins procedure was used to specify the ARMA model. The error volatility of the model was examined a Lagrange multiplier (LM) test which was compared to a chi-squared statistic. A proper GARCH error structure was adjusted, and the ARMA model with GARCH errors were specified as follows:

$$y_t = f_1 y_{t-1} + ¼ + f_p y_{t-p} + d + e_t - q_1 e_{t-1} - ¼ - q_q e_{t-q}$$

where f and q represent the coefficients for the autoregressive and the moving average part of the model, respectively, and d stands for the mean of the process. The error term of the ARMA process is assumed to follow the specification represented next (Enders, 1995) :

$$e_t = n_t \sqrt{h_t} \text{ where } s_n^2 = 1$$

where and ν is the multiplicative disturbance and h is represented as follows:

$$h_t = a_0 + \sum_{i=1}^{q} a_i e_{t-i}^2 + \sum_{i=1}^{p} b_i h_{t-i}$$

That is, the ARMA error squared is assumed to follow a new ARMA process

The coefficients of the ARMA process were allowed to vary using the Kalman filter algorithm. The time-varying Kalman filter algorithm can be parsimoniously represented in the state-space representation as follows (Hamilton, 1994):

$$\mathrm{x}_{t+1} = \mathbf{F}\,\mathrm{x}_t + \mathbf{v}_{t+1}$$

$$\mathbf{y}_t = \mathbf{A'x}_t + \mathbf{H'}\,\mathrm{x}_t + \mathbf{w}_t$$

where the first expression is the state equation and the second is the observational equation. **F, A'** and **H'** are matrices of parameters of dimension $(r \times r)$, $(n \times k)$ and $(n \times r)$ respectively, $\mathbf{x}_t$ is a $(k \times 1)$ of exogenous or pre-determined variables and x_t is a vector of nonobserved variables. The $(r \times 1)$ vector $\mathbf{v}_t$ and the $(n \times 1)$ vector $\mathbf{w}_t$ are white noise vectors such that $E(\mathbf{v}_t\mathbf{v}_\tau) = \mathbf{Q}$ for ($t = t$) and 0 otherwise, and $E(\mathbf{w}_t\mathbf{w}_t) = \boldsymbol{R}$ for ($t = t$) and 0 otherwise. The errors $\mathbf{v}_t$ and $\mathbf{w}_t$ are assumed to be uncorrelated at all lags, that is, $E(\mathbf{v}_t\mathbf{w}'_\tau) = 0$. For the time-varying coefficient model $\mathbf{F}(x)$, $\mathbf{Q}(x)$, $\mathbf{H}(x)$ and $\mathbf{R}(x)$ are matrix-valued functions of $\mathbf{x}_t$. That is, the coefficient vectors and the error variance of the state and observational equations are allowed to vary over time as a function of **x.**

All computational work was performed using the software RATS for Windows version 4.31.

4. Discussion of the Results

Figures 1 and 2 show the returns of sugar future prices and the autocorrelation function for this series, respectively. An augmented dickey-fuller test showed that the series for sugar is stationary. An exam of the autocorrelation function of the series for the returns of sugar future prices, using the Ljung-Box-Pierce statistics, showed that the series is not random walk (table 1). It is clear from visual inspection of figure 1 that the returns are not i.i.d. For example, volatility was clearly higher in the begining of the series, during the 1980's than during the next years. This result was confirmed by a Lagrange multiplier (LM) test (161.314) for 2 degrees of freedom, according to Enders (1995).

Figure 1 - Daily Returns of the Sugar Future

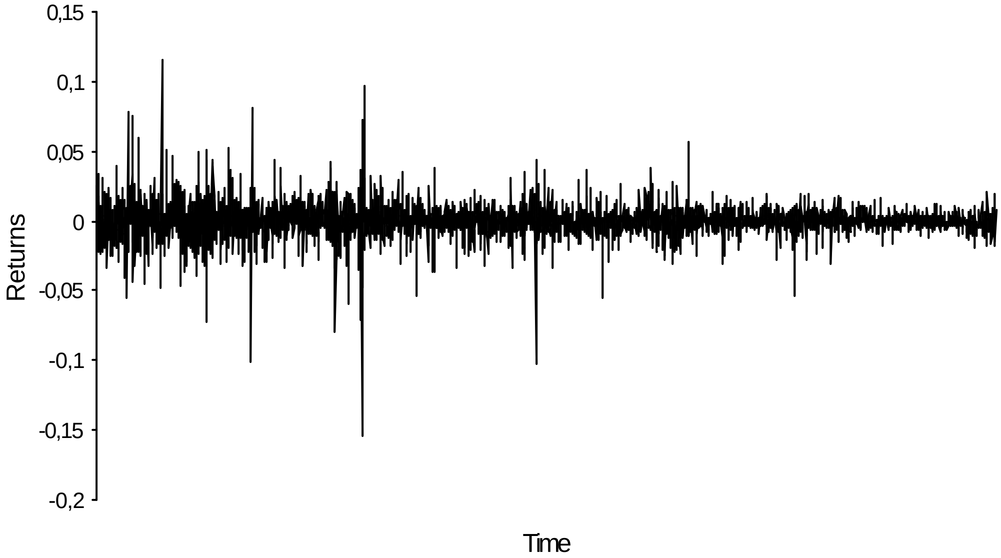


Figure 2 – Sample Autocorrelation for the Returns

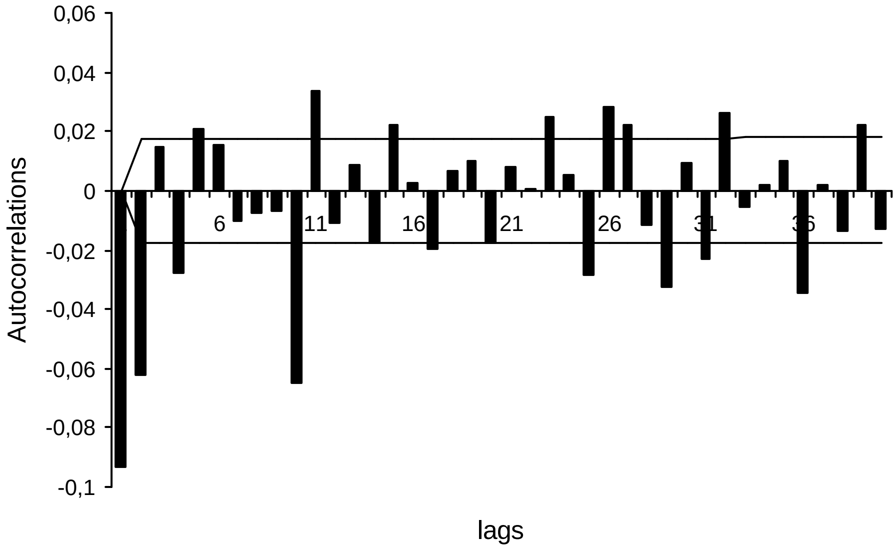


Table 1 – Ljung-Box-Pierce Statistic for the Autocorrelations of the Returns

Q (K)	Q-value	P-value
Q(2)	41.515	0.000
Q(4)	44.750	0.000
Q(6)	46.976	2×10^{-8}
Q(8)	47.528	12×10^{-8}
Q(10)	61.559	0.000
Q(12)	65.777	0.000
Q(14)	67.065	1×10^{-8}
Q(16)	68.745	2×10^{-8}
Q(18)	70.143	4×10^{-8}
Q(20)	71.495	1×10^{-7}

The series of the returns was adjusted as an ARMA process. The best fitted models was an ARMA (3,1) which is shown in table 2. The ARMA model estimated shows coefficients statistically significant for all variables but AR{3}.

Table 2 – Estimated Coefficients for the ARMA (3,1) Model of the Returns

Variables	Coefficients	Standard Error	t-Statistic
AR{1}	-0.83	0.13	-6.33
AR{2}	-0.14	0.03	-5.32
AR{3}	-0.03	0.02	-1.56
MA{1}	0.73	0.13	5.64

Table 3 shows the maximum likelihood estimation for the ARMA(3,1) model with GARCH(2,1) errors. For the ARMA(3,1) model, all coefficients but the constant were statistically significant. The GARCH error model showed all coefficients but MA(1) statistically significant at 5% level.

Figures 3, 4 and 5 showed the values of the coefficients of AR{1}, AR{2} and AR{3} over time. All figures show a pattern of high variance of the coefficients up to end of the first half of the sample. During the second half coefficients are more stable. The coefficients do not show a clear pattern of decreasing values for higher lags and stable values for the lower lags as was hypothesized. However, the coefficient of the highest lag variable showed the greatest fall in absolute terms when compared to other coefficient variables. This may suggest that higher lag coefficients lost weight over time when compared to lower lags. Also, the pattern of high variance in coefficients in the first half of the

Table 3 – Results of the Maximum Likelihood Estimation

Variables	Coefficients	Standard Error	t-Statistic
ARMA(3,1)			
Constant	3.3×10^{-4}	3.1×10^{-4}	1.06
AR {1}	-0.64	0.24	-2.73
AR {2}	-0.13	0.03	-4.09
AR {3}	-0.07	0.02	-2.91
MA {1}	0.57	0.23	2.42
GARCH (2,1)			
Constant	7.3×10^{-5}	3.3×10^{-6}	22.24
AR {1}	0.28	0.03	8.45
AR {2}	0.36	0.03	9.98
Ma {1}	0.06	0.00	0.00

sample and the relatively lower variance of the coefficients in the second half may suggest that market is becoming more stable over time. If one assume that market stability is related to efficiency, the pattern of variance of the coefficients can be regarded as supporting the hypothesis behind this work.

5. Final Coments

The underlying hypothesis behind this paper is that financial markets are becoming more efficient due to the increasing availability of information. Market efficiency means that economic agents have most of existing information readily available to guide decision. Accordingly, if markets are becoming more efficient a short autoregressive memory models are becoming more representative of financial time series data. This is to say that higher lags in time series models are weighting less over time. This pattern can be understood as an indication that agents are incorporating more available information over time before taking decision. The findings of this paper showed evidences that higher lag variables weighed heavily in the beginning of the series as compared to more recent years. This evidence indicates a change regarding the behavior of sugar future prices considering that a strong tendency for a short autoregressive model over time can be observed. Also, the decreasing volatility of the coefficient may be seen as an indication of increasing stability.

Some considerations about the reasoning for a significant loss weight of higher lag variables over time can be drawn as follows. This market may not be rational, a profitable trading strategies may exist, or psychological factors would be important for pricing securities. For example, a time series patterns of returns would occur because investors either overreact or only partially adjust to information arriving to the market. In many cases, investors may react late to trends, thereby incorporating past information into present purchasing strategies.

Thus, people do not always behave in a linear fashion to new information, processing it immediately, as the (EMH) requires. Indeed, people, and nature in general, are often nonlinear. Thus, for "astute" investors, excess profits can exist even if financial markets are well functioning.

Even though the discussion on market efficiency is not new, empirical tests on financial time series regarding market efficiency has been growing during the past years. With advances in the field of computer science, and the recent developments of time series techniques, it is likely that this subject will be in discussion during the next years. It would be important to examining time-varying coefficients for the returns of other commodities, as well as testing a higher variety of model specifications. Futures works should include more details about the possible random walk behavior in the commodity futures markets. Other techniques such as variance ratio test, spectral analysis, nonsynchronous trading model, an investigation about the bid-ask-spread effect in the data, plausible sources for the inefficiency in these markets, etc could be used to examine the Efficient Market Hypothesis. Simulations with other futures contracts will be very elucidative to future considerations.

Figure 3 - Values for the Time Varying Coefficients for AR{1}

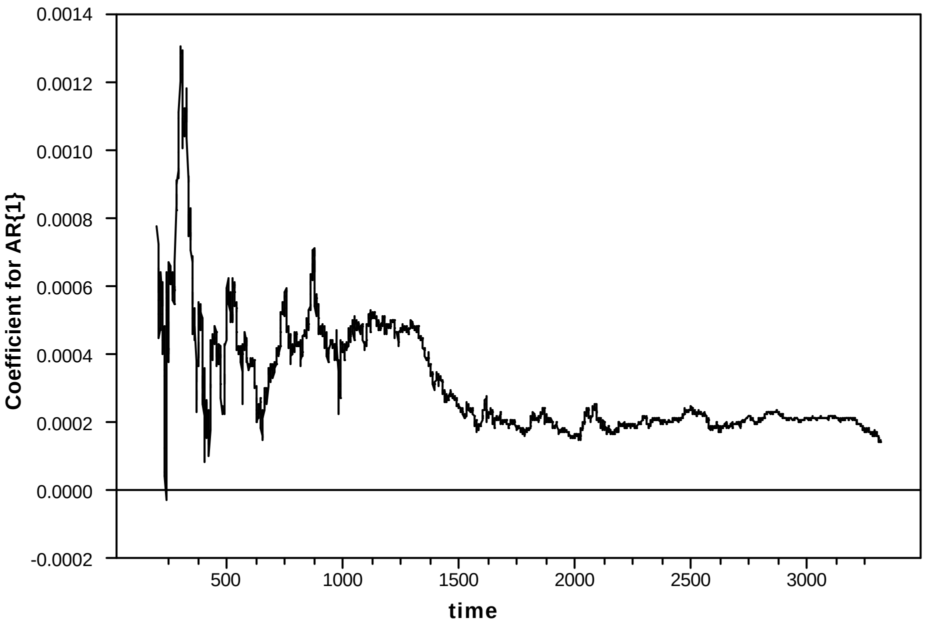


Figure 4 - Values for the Time Varying Coefficients for AR{2}

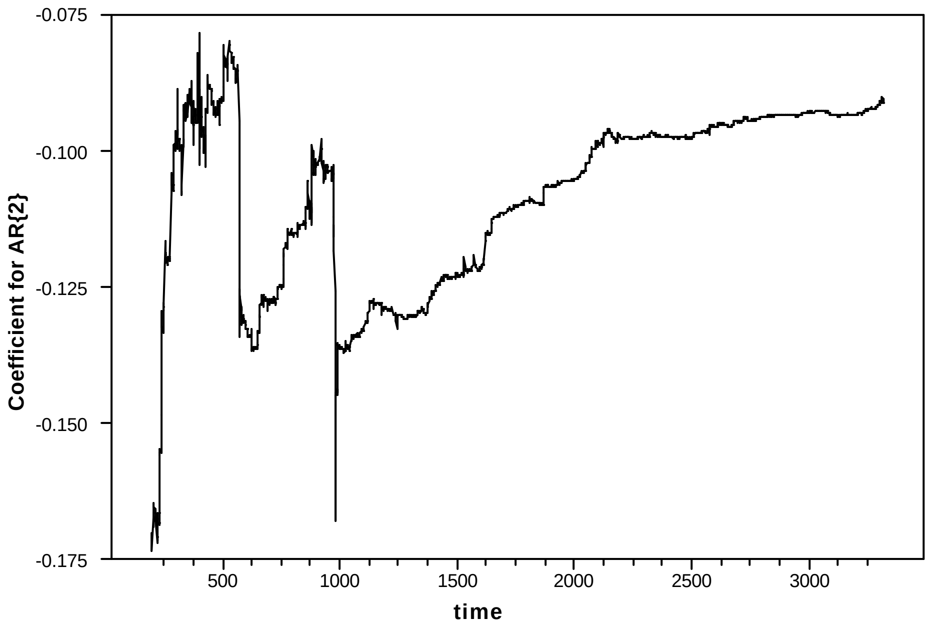
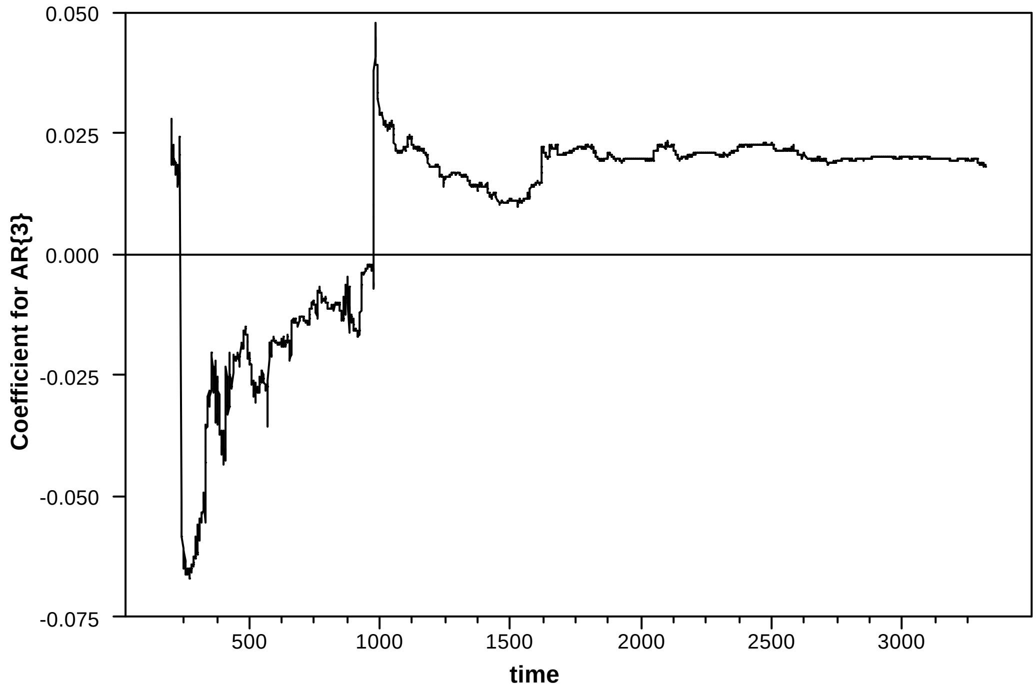


Figure 5 - Values for the Time Varying Coefficients for AR{3}



REFERENCES

BACHELIER, L.(1900): " Theory of Speculation," in Cootner, P. (ed), *The Random Character of Stock Market Prices*, Massachusetts Institute of Technology Press, Cambridge, MA, 1964; Reprint.

.

BOX, G.E.P., JENKINS, G.M AND REINSEL, G.C, *Time Series Analysis. Forecasting and Control*. Prentice-Hall, Inc. USA, 1994.

CAMPBELL, J., LO, A., and MACKINLAY, A. *The Econometrics of Financial Markets.*Princeton University Press, New Jersey, 1997.

CARGILL, T.F. and RAUSSER, G.C. (1975): " Temporal Price Behaviour in Commodity Futures Markets," *Journal of Finance*. 30, 1043-1053.

COOPER, J. C. B (1982): "World stock markets: some random walk tests. *Applied Economics*. 14, 515-531.

COWLES, A. (1933): " Can Stock Market Forecasters Forecast ? ," *Econometrica*. 1, 309-324.

ENDERS, W. *Applied Econometric Time Series*. John Wiley and Sons. 1995.

FAMA, E. (1970): " Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*. 25, 383-417.

HAMILTON, J.D. *Time Series Analysis*. Princeton University Press, New Jersey, 1994.

LEUTHOLD, R.M. (1972): " Random Walks and Price Trends: The Live Cattle Futures Market," *Journal of Finance*. 27, 879-889.

LO, A., and A.C. MACKINLAY. (1988): " Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*. 1, 41-66.

MALKIEL, B. (1992): " Efficient Market Hypothesis," in Newman, P., M. Milgate, and J. Eatwell (eds), *New Palgrave Dictionary of Money and Finance*, Macmillan, London.

MERTON, R. *Continuous Time Finance*. Blackwell, Cambridge, 1990.

NEFTCI, SALIH, N. *An Introduction to the Mathematics of Financial Derivatives*. Academic Press, USA, 1996.

PETERSON, R.L., MA, C.K., RITCHEY, R.J. (1992): " Dependence in Commodity Prices," *Journal of Futures Markets.* 12, 429-446.

SAMUELSON, P. (1965): "Proof that Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review*. 6, 41-49.

TAYLOR, S.J. (1985): " The Behavior of Futures Prices Over Time," *Applied Economics*. 17, 713-734.