The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

MONASH

WP 10/98

ISSN 1440-771X
ISBN 0 7326 1047 8

MONASH UNIVERSITY

AUSTRALIA

LIBRARY
WITHDRAWN
NOV 24 1998
GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS

A General Volatility Framework and the Generalised Historical Volatility Estimator

Bernard Bollen and Brett Inder

Working Paper 10/98
August 1998

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

A GENERAL VOLATILITY FRAMEWORK AND THE GENERALISED HISTORICAL VOLATILITY ESTIMATOR.

Bernard Bollen and Brett Inder
Department of Econometrics and Business Statistics
Monash University
Clayton, Australia, 3168.

Email: bernie.bollen@buseco.monash.edu.com
Phone: (03)99052438
Fax: (03)99055474

(Current Version : August 1998)

ABSTRACT

This study proposes a new approach to the estimation of the time series properties of daily volatility in financial markets. The estimation technique is a two stage procedure which initially estimates the volatility of any particular trading day from intraday data. This procedure is implemented over a number of trading days to produce a series of daily volatility estimates. A general volatility framework is also developed and the series of daily volatility estimates can be put into this framework to estimate the time series properties of daily volatility. Furthermore, with this new approach it is shown that the time series properties of daily volatility can be modelled in a wide range of functional forms, including those functional forms which capture asymmetric information effects.

• The authors wish to thank Paul Kofman, Keith McLaren, Maxwell King, Neil Shephard, Chris Ulph, Rodney Strachan and participants at the 1998 Econometric Society Australasian Meeting for their many invaluable comments. All errors and omissions are however the sole responsibility of the authors.

A General Volatility Framework and the Generalised Historical Volatility Estimator.

ABSTRACT

This study proposes a new approach to the estimation of the time series properties of daily volatility in financial markets. The estimation technique is a two stage procedure which initially estimates the volatility of any particular trading day from intraday data. This procedure is implemented over a number of trading days to produce a series of daily volatility estimates. A general volatility framework is also developed and the series of daily volatility estimates can be put into this framework to estimate the time series properties of daily volatility. Furthermore, with this new approach it is shown that the time series properties of daily volatility can be modelled in a wide range of functional forms, including those functional forms which capture asymmetric information effects.

In recent financial market literature, the price of a security is typically assumed to follow a geometric Brownian motion process in continuous time. A discrete time version of the geometric Brownian motion model is a random walk in the logarithm of prices,

$$Log_e(p_t) = \mu + Log_e(p_{t-1}) + e_t \ , \qquad (1)$$

where $e_t \sim iid\ N(0,\sigma^2)$, p_t is the security price at time t and μ is the expected rate of return on p_t. Defining the return at time t to be $r_t = Log_e\left(p_t \big/ p_{t-1}\right)$, the return series can thus be modelled as $r_t = \mu + e_t$. An estimate of the volatility of the return series is given by Figlewski (1997).

$$\hat{\Lambda} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(r_t - \bar{r})^2} \qquad (2)$$

where $\hat{\Lambda}$ is the volatility estimator and $\bar{r} = \frac{1}{n}\sum_{t=1}^{n} r_t$. Equation (2) is usually referred to as the historical volatility estimator and implicitly assumes that daily volatility is constant. However daily volatility modelling in recent times has been dominated by the ARCH family of volatility models and the more recent range of stochastic volatility models, both of which capture the notion that daily volatility is time varying.

This paper proposes a new approach to the estimation of the time series properties of daily volatility estimated from intraday data. This goal is achieved with the development of two theoretical models. Initially a general volatility framework is developed in section I which is general enough to encompass most existing time varying volatility models, that is the ARCH family of volatility models and the stochastic volatility models.

The second theoretical development is the generalised historical volatility estimator which can be used to estimate daily volatility from intraday data. Given that the intraday return series on any particular trading day is highly autocorrelated, this volatility estimator can estimate the volatility of an autocorrelated return series to produce a daily volatility estimate $\hat{\Lambda}_t$. Then a series of daily volatility estimates can be estimated $\left\{\hat{\Lambda}_t\right\}_{t=1}^{T}$ for T trading days from the intraday data from these trading days. It is this series of daily volatility estimates placed with the general volatility framework that allows the researcher to model the time series properties of daily volatility.

The literature concerning the estimation of daily volatility from intraday data is remarkably scarce. Beckers (1983), Anderson (1995), Parkinson (1980) and Rogers and Satchell (1991) all propose estimators of daily volatility based upon daily open, close, high and low prices. Anderson and Bollerslev (1998) propose an efficient daily volatility estimator based upon the sum of five minute absolute returns. All of these daily volatility estimators are based upon the assumption that the intraday return series can be modelled as a continuous time Brownian motion process or as strict white noise. However recent research into finance market microstructure suggests that the intraday return series is autocorrelated, a stylised fact not captured in the geometric Brownian motion model.

In section II, a model of the data generating process for the intraday return series is developed. This model explicitly assumes the intraday return series for each trading day can be characterised by a unique ARMA model. In section III two estimators of daily volatility are considered, the historical volatility estimator and the simple volatility estimator. The historical volatility estimator will be seen to be a biased estimator of daily volatility whilst the simple daily volatility estimator is an unbiased estimator of daily volatility, however it is inefficient, because it fails to use all available intraday information.

In section IV, new specifications of volatility are proposed which allow us to model the volatility of an autocorrelated return series. Given that the intraday return series is often autocorrelated, this new specification will allow the finance researcher to estimate daily volatility from the intraday return series. This new specification of volatility can be seen as a generalisation of the historical volatility estimator, and thus shall be referred to as the generalised historical volatility estimator. Anderson and Bollerslev (1998, p255) note that

> " A significant finding to emerge from our study is that high-frequency returns contain valuable information for the measurement of volatility at the daily level. .... These results encourage the development of new and improved techniques for the estimation and prediction of daily or lower frequency volatility that explicitly incorporate the information in high frequency returns."

This paper attempts to address the problem articulated by Anderson and Bollerslev.

The discussion will then explore various methods of modelling the type of autocorrelation typically found in intraday data. A key finding of this section is that if the intraday return process can be modelled as an ARMA(p,q) process, then daily volatility can be expressed in terms of the parameters of that ARMA(p,q) model.

Furthermore, it will be shown that daily volatility can be estimated from, for example transaction returns, five minute returns, ten minute returns or any other sampling interval. It will be demonstrated that by utilising higher sampling frequencies, usually an increase the efficiency of the proposed daily volatility estimator will result. Monte Carlo evidence is presented in section V which shows that the generalised historical volatility estimator is both unbiased and generally more efficient as the intraday sampling frequency of returns increases.

In section VI, the generalised historical volatility estimator is applied to S&P 500 Index futures contracts. Evidence suggests that the generalised historical volatility estimator correlates well with proxies for information arrival. In section VII, the validity of the generalised historical volatility estimator is further verified by showing that scaled daily returns, conditioned upon an unbiased daily volatility estimator, should follow a standard normal distribution. This is empirically verified with the generalised historical volatility estimator based upon transactions data. In section VIII, the estimation of general volatility framework models is considered under a number of different functional forms. Evidence is presented in section IX which brings into question the specification of the GARCH(1,1) model of daily volatility. Section X concludes the discussion.

I. A General Volatility Framework

Anderson (1994) defines a volatility process $\{\Lambda_t\}_{t=1}^{T}$ in terms of the model

$$R_t = \mu_t + \Lambda_t Z_t, \tag{3}$$

where R_t is the proportional daily return, μ_t is the mean daily return, Λ_t is the daily volatility of day t and $Z_t \sim$ iid(0,1). In the spirit of Anderson (1994), we shall furthermore assume that daily volatility in a financial market is driven by an unobserved activity variable K_t, and that daily volatility is a function of this activity variable $\Lambda_t = g(K_t)$. This activity variable we also assume drives the daily volume and the number of intraday transactions. We shall further assume that K_t has an autoregressive and moving average representation.

$$K_t = \omega + \beta\Theta_{t-1} + \phi_1 K_{t-1} + \cdots + \phi_p K_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_{t-q}\varepsilon_{t-q} \tag{4}$$

where Θ_{t-1} is a vector of explanatory variables determined at time *t-1*; ω , ϕ_i, θ_i $(i > 0)$ are fixed parameters with $\sum_i \phi_i < 1$; β is a vector of parameters; $Var(\varepsilon_t) = \sigma_\varepsilon^2$, and ε_t is strict white noise. Equations (3) and (4) constitute a well defined stochastic volatility model if Z_t and ε_t are independent. Note that Λ_t is interpreted as the volatility of R_t on day t in the sense that $Var(R_t | \ell_{t-1}, \varepsilon_t) = \Lambda_t^2$, where ℓ_{t-1} is the information set at time *t-1*. This functional form for K_t clearly differs from Anderson's (1994) specification of K_t, namely

$$K_t = \omega + \beta K_{t-1} + [\gamma + \alpha K_{t-1}]\mu_t\ , \tag{5}$$

where $\alpha, \beta, \gamma > 0$, $\alpha + \beta > 0$ and $\alpha + \gamma > 0$. Anderson specifies this functional form because by appropriately restricting the parameters α, β and γ as well as the functional form of K_t, it can be seen that stochastic volatility generalisations of the GARCH(1,1) and EGARCH(1,1) models are obtained. The lognormal stochastic volatility model is also obtained with appropriate restrictions upon (4). Certainly the ARCH family of volatility models cannot be defined within this framework. It is for this reason that we believe (4) will be general enough to encompass a wide range of stochastic volatility models for the volatility process as well as the ARCH family of volatility models. The specification of volatility in (4) shall be referred to as the general volatility framework.

The variance of ε_t is of particular importance in that if we restrict $Var(\varepsilon_t) = 0$, we can, with other suitable restrictions define the ARCH family of volatility models, with the exception of integrated and fractionally integrated ARCH volatility models. For example the GARCH(1,1) model (see Bollerslev, 1986) is defined as

$$R_t = e_t,\ h_t = Var(e_t|e_{t-i})\ i > 0,\ h_t = \omega + \alpha e_{t-1}^2 + \beta h_{t-1}. \tag{6}$$

Now letting $\Lambda_t^2 = K_t$, $Var(\varepsilon_t) = 0$, $p=1$ and $\Theta_{t-1} = R_{t-1}^2 = e_{t-1}^2$ then (6) reduces to $\Lambda_t^2 = \omega + \phi\Lambda_{t-1}^2 + \beta e_{t-1}^2$. This GARCH(1,1) model is not a genuine stochastic volatility model in the sense that the level of volatility at time t is driven by a stochastic process, but rather it is fully determined at time $t-1$ by Λ_{t-1} and R_{t-1}. It is the inclusion of the contemporaneous random variable ε_t that defines a genuine stochastic volatility process. The EGARCH(1,1) model (see Nelson, 1989), captures the effect of asymmetric information on volatility. It is defined as

$$R_t = e_t\ ,\ h_t = Var(e_t|e_{t-i})\ i > 0, \tag{7}$$
$$Log_e(h_t) = \omega + \beta Log_e(h_{t-1}) + \gamma.\frac{e_{t-1}}{\sqrt{h_{t-1}}} + \alpha\left[\frac{|e_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}}\right],$$

where ω, β, γ and α are constant parameters. The EGARCH(1,1) model can also be defined within the general volatility framework with the restrictions $K_t = Log_e(\Lambda_t^2)$, $Var(\varepsilon_t) = 0$, $p=1$ and $\Theta_{t-1} = \gamma.\frac{e_{t-1}}{\Lambda_{t-1}} + \alpha\left[\frac{|e_{t-1}|}{\Lambda_{t-1}} - \sqrt{\frac{2}{\pi}}\right]$. The GJR model (see Glosten, Jagannathan and Runkle, 1990) like the EGARCH(1,1) model, captures the effect of asymmetric information on volatility. The model is defined as

$$R_t = e_t \;,\; h_t = Var\left(e_t | e_{t-i}\right) \; i > 0,$$
$$h_t = \omega + \beta h_{t-1} + \alpha e_{t-1}^2 + \gamma S_{t-1}^- e_{t-1}^2, \tag{8}$$

where ω, β, γ and α are constant parameters, and $S_t^- = 1$ if $\varepsilon_t < 0$, $S_t^- = 0$ otherwise. Again, this model can be put into the general volatility framework with the restrictions $\Lambda_t^2 = K_t$, $Var(\varepsilon_t) = 0$, $p=1$ and $\Theta_{t-1} = S_{t-1}^- \Lambda_{t-1}^2$. The lognormal stochastic volatility model (see Anderson, 1994) is obtained with the restrictions $K_t = Log_e(\Lambda_t)$, $\theta_i = 0$ $(i>0)$, $Var(\varepsilon_t) > 0$, $p=1$, and $\Theta_{t-1}=0$,

$$Log_e\left(\Lambda_t\right) = \omega + \phi Log\left(\Lambda_{t-1}\right) + \varepsilon_t. \tag{9}$$

Model (9) is of particular interest because $Var(\varepsilon_t) > 0$ and it is this non zero variance that allows the model to be a genuine stochastic volatility model in the sense that the level of volatility at time *t* is not fully determined by any other set of variables determined at time $t-1$. A wide range of other functional forms could in principle be defined by specifying the functional form of g(K_t) and Θ_{t-1} along with p, θ_i $(i > 0)$, and $Var(\varepsilon_t)$.

Estimation of stochastic volatility models is difficult using maximum likelihood methods since K_t is unobservable. A number of other estimation techniques for stochastic volatility models have thus been proposed. In particular Harvey *et al.* (1994) use quasi-maximum likelihood techniques, Jacquier *et al.* (1994) used Bayesian Markov chain Monte Carlo methods, Danielsson (1994) suggested simulated maximum likelihood whilst Anderson (1994) employed a GMM technique. In later sections we shall implement a rather unorthodox procedure to estimate general stochastic volatility models (that is when $Var(\varepsilon_t) > 0$). We exploit the fact that in many situations high frequency data is available, and this data can be used to estimate daily volatility. It is a two stage estimation procedure in which daily volatility on any particular trading day Λ_t, is estimated from intraday data to produce a series a daily volatility estimates $\left\{\hat{\Lambda}_t\right\}_{t=1}^{T}$ for T trading days with the generalised historical volatility estimator. The second stage of the estimation procedure involves initially specifying the functional form of g(K_t) and Θ_{t-1} along with p and allowing θ_i $(i > 0)$ to be non-zero. By letting $Var(\varepsilon_t) > 0$, we have in effect an ARMA(p,q) model of daily volatility and this can be estimated using maximum likelihood methods. To estimate (4) however we require an estimate of the volatility of any particular trading day. In later sections, we consider three estimators of daily volatility from intraday data, the historical, the simple, and the generalised historical volatility estimators.

II. A Model for Intraday Returns

In recent years the finance researcher has had access to the intraday price series, whether this be a transaction price series, five minute price series or any other sampling frequency of prices. Let $p_{i,t}$ be the i^{th} price on day t, and let n_t be the number of prices on this day. A series of intraday returns are defined from this price series as $r_{i,t} = Log_e\left(\frac{p_{i,t}}{p_{i-1,t}}\right)$, where $r_{i,t}$ is the i^{th} intraday return on day t. The relationship between intraday returns and the daily return R_t is given by $R_t = \sum_{i=1}^{n_t} r_{i,t}$. Combining this definition with (3), we have daily volatility defined as

$$\Lambda_t^2 = Var\left(R_t | \ell_{t-1}, \varepsilon_t\right) = Var\left(\sum_{i=1}^{n_t} r_{i,t} | \ell_{t-1}, \varepsilon_t\right). \tag{10}$$

Now intraday data can be used to estimate daily volatility Λ_t, if one is able to specify a particular model for how the intraday return series $\{r_{i,t}\}_{i=1}^{n_t}$ evolves throughout any particular trading day. Previous empirical research into the intraday return series has found returns to be highly autocorrelated. Autocorrelation in the intraday return series could be the result of the bid-ask spread (see Roll, 1984), trading strategies employed by informed traders (see Kyle, 1985), discreteness in prices (see Harris, 1990), non-continuous trading (see Stoll and Whaley, 1990), and other sources. We consequently postulate that intraday returns can be modelled as a unique ARMA process each trading day $r_{i,t} = \frac{\mu_t}{n_t} + \phi_{1,t} r_{i-1,t} + \cdots + \phi_{p_t,t} r_{i-p_t,t} + e_t + \theta_{1,t} e_{i-1,t} + \cdots + \theta_{t-q_t} e_{i-q_t,t}$ which can be equivalently written as

$$\phi_t(L) r_{i,t} = \frac{\phi_t(1)\mu_t}{n_t} + \frac{\Lambda_t \phi_t(1)\theta_t(L)}{\sqrt{n_t}\,\theta_t(1)} e_{i,t}, \tag{11}$$

where L is the usual lag operator, $\phi_t(L) = 1 - \phi_{1,t}L - \phi_{2,t}L^2 - \cdots - \phi_{p_t,t}L^{p_t}$, $\theta_t(L) = 1 - \theta_{1,t}L - \theta_{2,t}L^2 - \cdots - \theta_{p_t,t}L^{q_t}$, p_t and q_t are the orders of the autoregressive and moving average components respectively on day t, $\phi_{i,t}$ and $\theta_{i,t}$ are the i^{th} autoregressive and moving average parameters respectively on day t, n_t is the number of intraday returns, μ_t is the mean daily return for day t and $e_{i,t} \sim iid\ N(0,1)$.

Provided the orders of the AR(.) and MA(.) process (p_t and q_t) are chosen using an appropriate data dependent criterion such as the BIC, this class of models can encompass a wide range of stationary processes (see Berk, 1974 and Ng and Perron, 1993). This assumption is very much in the spirit of Hasbrouck and Ho (1987) and Stoll and Whaley (1990) who also model the intraday return series as a stationary ARMA process.

However (11) does not capture some stylised facts regarding the data generating process for intraday returns. In particular, intraday seasonality in volatility is well established in the literature (see Bollerslev and Anderson, 1997), particularly at the open and close of trading as well as at predicable times when other markets open and close. Furthermore, strategic (informed) trading may occur at any time during the trading day giving rise to unpredictable periods when trading is more intense and more volatile than usual. Hence intraday seasonality in volatility introduces predictable periods of volatility clustering whilst strategic trading introduces unpredictable periods of volatility clustering in the intraday return series. However, to date no definitive model for the data generating process of the intraday return process which includes volatility clustering has yet been developed which is supported by rigorous empirical evidence. Consequently the covariance stationarity assumption of the intraday return series may at first sight be considered a rather bold assumption.

Clearly then (11) is a considerable simplification of the true data generating process. However our focus is on the variance of daily returns and how this relates to the variance of intraday returns. Volatility clustering is a form of heteroskedasticity and consequently the parameters of the ARMA model for a particular trading day will be less efficiently estimated but nonetheless consistently estimated if we choose to ignore the time variation in conditional variance through the trading day. Given the very large intraday datasets, the inefficiency of the parameter estimates will be of little consequence in practice. The evidence we present in later sections suggests that very little if any systematic bias in daily volatility estimates does occur by assuming a covariance stationary intraday return series.

We now show in the following theorem that the assumed data generating process for $r_{i,t}$ in (11) is consistent with how R_t is assumed to evolve in (3).

Theorem. If $r_{i,t}$ is defined by (11), and $R_t = \sum_{i=1}^{n_t} r_{i,t}$ then for large n_t we can define R_t by $R_t = \mu_t + \Lambda_t Z_t$ where $Z_t \sim N(0,1)$.

Proof. See appendix (1).

Our aim is to estimate the parameters of (11) and various restricted versions of (4) for a given $g(.)$. The general approach we advocate is as follows

(a) Ignore the interday relationships between the Λ_t's which are implied by (4), and estimate Λ_t for a given trading day using the intraday return series and the assumed data generating process in (11).

(b) Use the series of estimates of Λ_t to estimate the parameters of (11) by standard methods for estimating ARMA models.

Point (a) above is unconventional in that daily volatility is estimated from intraday data. However from the data generating process we have assumed for the intraday return process (11), $r_{i,t}$ can be considered to be a drawing from a conditionally normal distribution with variance $Var(r_{i,t}|\varepsilon_t) \propto \frac{\Lambda_t^2}{n_t}$. Our sample of n_t intraday observations can thus be used to estimate Λ_t .

Section III will focus on two estimators of Λ_t, the simple and the historical daily volatility estimators and the limitations of these two daily volatility estimators, namely inefficiency and bias respectively. In section IV the generalised historical daily volatility estimator, which is both an efficient and consistent estimator of daily volatility is developed.

III. Two Estimators of Daily Volatility

If we use no intraday data to estimate daily volatility, then daily volatility can nonetheless be estimated from daily close to close prices. Initially we define a random variable P_t^C, as the closing price in a financial market on day *t*. We have already defined the daily volatility parameter Λ_t for day t from the model $R_t = \mu_t + \Lambda_t Z_t$ as $\Lambda_t^2 = Var(R_t | \ell_{t-1}, \varepsilon_t)$, $R_t = Log_e\left(\frac{P_t^C}{P_{t-1}^C}\right)$. A possible estimator for Λ_t is $\hat{\Lambda}_t = |R_t|$ if we assume $\mu_t = 0$ (an assumption that we will return to later). If it is assumed that Z_t is normally distributed then $Z_t = \frac{R_t}{\Lambda_t} \sim \chi^2_{(1)}$ conditional on Λ_t, and it is well known that $E\left[\sqrt{Z_t}\right] = \sqrt{2/\pi}$. Thus $|R_t|$ would be a biased estimator of Λ_t. However, an unbiased estimator is easy to construct

$$\hat{\Lambda}_t = \frac{|R_t|}{\sqrt{2/\pi}}. \tag{12}$$

This estimator of daily volatility is unbiased and shall be referred to as the simple daily volatility estimator. Of course, this estimator is inconsistent, as it only uses one observation in constructing the estimate, so it is not a very viable estimator, but its unbiasedness property will be useful for later comparison. Anderson and Bollerslev (1998) also use this estimator of daily volatility.

Other estimators such as the historical and generalised historical volatility estimator use averages of $r_{i,t}^2$ to compute $\hat{\Lambda}_t^2$, and hence will have distributions which are functions of a $\chi^2_{(n_t)}$ random variable where n_t is the number of intraday returns on day *t*. For any reasonably sized n_t, it is easily seen that $E\left[\hat{\Lambda}_t\right] \approx \sqrt{E\left[\hat{\Lambda}_t^2\right]}$, and for a very large n_t no correction for bias would be necessary.

We now consider the historical volatility estimator of daily volatility calculated from all intraday returns. If in (11) we assume intraday returns are not autocorrelated then $\phi_t(L) = \theta_t(L) = 1$ and (11) reduces to

$$r_{i,t} = \frac{\mu_t}{n_t} + \frac{\Lambda_t}{\sqrt{n_t}} e_{i,t} . \quad (13)$$

Note that in (13), $Var(r_{i,t}|\Lambda_t) = \frac{\Lambda_t^2}{n_t}$ and hence Λ_t can be estimated as $\hat{\Lambda}_t = \sqrt{n_t V\hat{a}r(r_{i,t})}$ where $V\hat{a}r(r_{i,t}) = \frac{1}{n_t - 1}\sum_{i=1}^{n_t}(r_{i,t} - \bar{r})^2$.

However at the intraday level of analysis, the return series may well contain some autocorrelation, violating the assumption that r_i is *iid*. Thus equation (13) will be of little practical use in estimating daily volatility from an intraday return series since equation (13) assumes an *iid* return series.

It is for this reason that the finance researcher typically uses daily, weekly or even monthly returns. These sampling intervals are necessary to avoid the autocorrelation found in financial return time series which can bias volatility estimates. Unfortunately, by using these large sampling intervals a price is paid in terms of losing a large number of observations. It is generally found that data based on small sampling intervals are more highly autocorrelated than those based on larger sampling intervals. Thus the estimation of volatility based upon these small sampling intervals will be more efficient (in the sense of using more data), but will also be more biased if we fail to take account of the autocorrelation in returns.

IV. The Generalised Historical Volatility Estimator

It is well known that the intraday return series can be autocorrelated and consequently we will assume that conditional on Λ_t, $r_{i,t}$ is stationary and that it can be modelled as a covariance stationary ARMA process. Recall from (11) that an ARMA process in the intraday return series for day t can be written in the form $\phi_t(L)r_{i,t} = \frac{\phi_t(1)\mu_t}{n_t} + \frac{\Lambda_t\phi_t(1)\theta_t(L)}{\sqrt{n_t}\theta_t(1)}e_{i,t}$. Now defining $z_{i,t} = \frac{\Lambda_t\phi_t(1)}{\sqrt{n_t}\theta_t(1)}e_{i,t}$ and $\mu_t^* = \frac{\phi_t(1)\mu_t}{n_t}$, equation (11) can be written as $\phi_t(L)r_{i,t} = \mu_t^* + \theta_t(L)z_{i,t}$, where $Var(z_{i,t}) = \sigma_t^2 = \frac{\Lambda_t^2\phi_t^2(1)}{n_t\theta_t^2(1)}$. We thus have

$$\Lambda_t^2 = \left[\frac{\theta_t(1)}{\phi_t(1)}\right]^2 n_t \sigma_t^2 . \tag{14}$$

An expression for daily volatility has thus been developed in terms of the parameters of an ARMA(p,q) model for $r_{i,t}$. The model can be estimated by firstly estimating the parameters of the ARMA(p,q) process and then substituting these estimates into (14). This methodology will give us consistent estimates of daily volatility since as $n_t \to \infty$, $plim\left(\hat{\theta}_{i,t} - \theta_{i,t}\right) = 0$, $plim\left(\hat{\phi}_{i,t} - \phi_{i,t}\right) = 0$ and $plim\left(\hat{\sigma}_t^2 - \sigma_t^2\right) = 0$. The estimator is thus defined as

$$\hat{\Lambda}_t^2 = \left(\frac{1 - \hat{\theta}_{1,t} - \hat{\theta}_{2,t} - \cdots - \hat{\theta}_{q_t,t}}{1 - \hat{\phi}_{1,t} - \hat{\phi}_{2,t} - \cdots - \hat{\phi}_{p_t,t}}\right)^2 n_t \hat{\sigma}_t^2 . \tag{15}$$

Thus far it has been assumed that the transaction return series can be modelled as an ARMA process, and consequently the parameters of that ARMA process can be used to calculate daily volatility. However the data from the transaction return series may be unavailable, and only data for the one minute or five minute return series may be available for example. The generalised historical volatility estimator (15) can be estimated from an intraday return series of any sampling frequency. However as more intraday returns are used to calculate daily volatility, an increase in the efficiency of the estimator should result.

V. A Monte Carlo Experiment

A Monte Carlo experiment was conducted to determine the relative efficiency of the generalised historical volatility estimator. A series of 2000 intraday returns were generated from an ARMA(1,1), MA(1), AR(0), AR(1), AR(2), AR(3) and AR(4) process. The disturbance term in each model was drawn from the standard normal distribution. An ARMA(1,1), MA(1), AR(1), AR(2) to AR(10) model was then fitted to the generated return series and the Schwartz Information Criterion (BIC) statistic $\left({}^{-}Log_e(BIC) = Log_e\left(\tilde{\sigma}_t^2\right) + \frac{K_t Log_e(n_t)}{n_t}\right)$ of each of these models was recorded where $\tilde{\sigma}_t^2$ is the maximum likelihood estimate of the error variance and K_t is the number of parameters in the model on day t. The 'best' model of the generated return series was chosen according the criteria of maximising the BIC. The parameters of this 'best' model were then used to estimate the volatility of the generated return series using the generalised historical estimator.

Two subsets of the 2000 generated returns were then also considered. A subset using every 4^{th} price, and a subset using every 20^{th} price were then used to generate a series of returns with 500 and 100 observations respectively. This process is in some ways analogous to estimating daily volatility from either transaction returns, one minute returns or the five minute return series. Also the simple volatility estimator was calculated from the return series and the results reported in the table titled 'n=1'. The two new return series (of 500 and 100 observations) were then modelled as an ARMA(p,q) process and the parameters of the 'best' model were used to estimate the volatility of the return series using the generalised historical volatility estimator. This experiment was repeated 1000 times. The mean, standard deviation, bias and root mean square error (RMSE) of the estimated daily volatility were calculated for the seven ARMA models considered. Results are shown in Appendix 2.

Note that in all the models of the intraday return series, volatility estimates are relatively unbiased no matter what the sampling frequency. However the standard deviation of volatility estimates (or efficiency of volatility estimates) shows that the estimator is most efficient with the larger sample size regardless of the nature of the 'true' series. In all models tested the efficiency decreased as sampling frequencies of n=500, n=100 and n=1 were used respectively. These results lend evidence to the belief that higher sampling frequencies generally lead to more efficient estimates of generalised historical volatility when the return series are correctly modelled as an ARMA(p,q) process. The RMSE is generally at its minimum when higher sampling frequencies are used.

It is of some interest to see what biases may result in daily volatility estimates if the intraday return process is generated by some type of ARMA process, but daily volatility is estimated assuming the return series can be modelled as white noise or as an MA(1) process. The motivation for modelling the return process as an MA(1) model is that it is not unusual in the finance literature to assume that the intraday return series can be approximated by an MA(1) process (see for example Anderson and Bollerslev, 1997). The Monte Carlo experiment considered earlier in this section is repeated, only this time we do not find the 'best' ARMA model to fit the data, rather we calculate the generalised historical volatility estimator modelling the given return series as an MA(1) or white noise process. This exercise allows us to quantify the biases and inefficiencies that occur in estimating daily volatility when the time series properties of the intraday return process is mis-specified. In Appendix 3 we consider the case when the historical volatility estimator is used estimate daily volatility when the 'true' return process is either an ARMA(1,1), MA(1), AR(1), AR(2), AR(3) or AR(4) process. In Appendix 4 the Monte Carlo experiment is repeated, however this time it is assumed the return process can be modelled as an MA(1) process.

In the case where it is assumed the return process is white noise, the bias in volatility estimates is at its maximum when all transactions are used (*n=2000*). Although in all cases the efficiency of the historical volatility estimator increased as the sampling frequency increased, the RMSE was at its maximum when the highest sampling frequency was used, due to worsening bias. It is thus concluded from the Monte Carlo results in Appendix 3, that calculating historical volatility on a return series which follows an ARMA process leads to very biased estimates of daily volatility, however by using lower sampling frequencies this bias problem can be alleviated to some extent.

In the case where it is assumed the return process follows an MA(1) process, the bias in volatility estimates is again at its maximum when higher sampling frequencies are used and the efficiency is at its maximum when higher sampling frequencies are used. The RMSE was generally at its highest when all intraday returns were used with an exception being the case when the true return generating process is ARMA(1,1). What is clear from the Monte Carlo experiments reported in Appendix 3 and 4, is that mis-specifying the intraday return process leads to biased estimates of daily volatility.

VI. An Analysis of S&P 500 Index Futures Daily Volatility

In this section we apply the generalised historical volatility estimator as proposed in section III to S&P 500 Index Futures. Intraday data on S&P 500 Index Futures from the Chicago Mercantile Exchange (CME) for the period January 1993 to December 1995 were obtained. Data for the period 19/3/94 to 4/5/94 was not available. Trading times for the CME are 8.30 am to 3.30 pm. 1,900,374 trades were recorded over this period. These recorded trades are not necessarily a record of each transaction. Transactions with zero price changes are frequently not recorded.

It is generally agreed in the literature that ultra high frequency data on intraday returns are highly autocorrelated. An AR(0), AR(1), AR(2) to AR(10) model with and without a constant was estimated in the intraday transaction return series and the five minute return series for each trading day of the S&P 500 Index futures data set in the years 1993 to 1995. The reason for modelling the intraday return series as an AR(.) process rather than as a more general ARMA(p,q) process is for computational tractability. An AR(.) model with sufficient lag length can in practice capture the autocorrelation present in a ARMA(p,q) model and consequently was considered to be an adequate model for the intraday return data generating process. Treating event spaced data (for example transaction returns) as an ARMA process is rarely done in practice, but should adequately capture autocorrelated behaviour if the series is not too thinly traded. The criteria used to select the 'best' model was maximising the BIC. The frequency of the AR(.) models chosen by the BIC for both the transaction and five minute return series are shown in table I.

< INSERT TABLE I HERE >

S&P 500 Index Futures intraday transaction return data is often highly autocorrelated as shown by the number of AR(.) models chosen by the BIC in table I. In fact, since this is a intraday return series with zero returns deleted, the degree of autocorrelation present has been considerably reduced relative to the original transaction by transaction price series (which we do not have access to). It is of some interest that over the 740 trading days considered, no trading days were chosen with a constant by the BIC in the transaction return series. This result lends evidence to the hypothesis that $E[R_t]=0$. Some 92% of the 'best' models chosen in the five minute return series were white noise. What is apparent is that the five minute return series is far less autocorrelated than the transaction return series. Furthermore only seven trading days were chosen with a constant by the BIC in the five minute return series. We should also note that selecting an ARMA model with the BIC will underestimate the lag length of the AR(.) model chosen as the 'best' model of the day's trading relative to using an information criteria such as the AIC. Hence table I is based, if anything on conservative estimates of the lag length of the AR(.) model of the intraday return series.

Six different estimators of daily volatility are estimated from the data. These daily volatility estimators are

1. Simple Volatility.

2. MA volatility based upon transactions returns .... (denoted by MA(T)).
3. MA volatility based upon 5 minute returns ....(MA(5)).
4. Historical volatility based upon transactions returns .... (Hist).
5. Generalised historical volatility based upon transactions returns ...... (GH(T)).
6. Generalised historical volatility based upon 5 minute returns .... (GH(5)).

The MA volatility estimator $\left(\Lambda_t^2 = (1-\theta_{1,t})n_t\sigma_t^2\right)$ is included because in the literature it is sometimes believed that the residuals from an MA(1) estimation on a five minute return series should be uncorrelated and hence can be used for the estimation of an MA(1)-GARCH(1,1) volatility model (see Anderson and Bollerslev 1997). The historical volatility estimator is included to quantify what bias may occur in the estimation of daily volatility by assuming the intraday return series is white noise when in fact the intraday return series is highly autocorrelated. Finally both transaction and five minute returns are used to estimate daily volatility to see if any systematic differences emerge from estimating the appropriate ARMA model in chronological time (five minute returns) or event time (transaction returns). The six daily volatility estimators were then used to calculate daily volatility of S&P 500 Index futures over the period 1993 to 1995. Results are shown in figure I.

< INSERT FIGURE I HERE >

A simple visual inspection of these six graphs shows that the simple volatility estimator has a number of estimates which are substantially larger and smaller than those for the other estimators, probably due to the inefficiency of this daily volatility estimator. All daily volatility estimators appear to move in a similar way through time, and appear to cluster in distinct high and low periods.

To further compare these volatility estimators, a correlogram for these daily volatility series, as well as for some measures (proxies) for information arrival, namely volume and the transaction count are shown in Appendix 5. The reason for including proxies for information arrival is that from the Mixture of Distributions Hypothesis (see Harris, 1986 and 1987), it is a desirable property of any proposed daily volatility estimator that it should positively correlate with proxies for information arrival. Note that two measures of volume are given.

(i) Volume(1) - the total number of contracts traded on each trading day.

(ii) Volume(2) - Volume over the period 1993 to 1995 displays both seasonal and trend components. The trend component is due to growth in the market whilst the seasonal component arises because volume increases toward contract maturity time. S&P 500 Index futures contracts expire on the third Friday of March, June, September and December of each year. Hence Volume(2) is a deseasonalised and detrended index of trading volume. Appendix 6 shows how this deseasonalised and detrended index of trading volume is developed.

Referring to Appendix 5, the simple daily volatility estimates showed the least correlation with proxies for information arrival. All other estimators performed reasonably well, with the generalised historical volatility estimates and MA volatility estimates based upon transactions data being the most highly correlated with proxies for information arrival overall.

VII. Conditional Normality in Returns

Recall that we define a process for daily returns by equation (3) as $R_t = \mu_t + \Lambda_t Z_t$, where $Z_t \sim iid(0,1)$. If we condition on Λ_t and assume Z_t is normally distributed, then it follows that

$$\left.\frac{R_t}{\Lambda_t}\right|\Lambda_t \sim N\left(\mu_t/\Lambda_t, 1\right). \tag{16}$$

If it is further assumed that $\mu_t = 0$ then (16) becomes $\frac{R_t}{\Lambda_t} \sim N(0,1)$. We can now define our standardised return Z_t as $Z_t = R_t/\Lambda_t$. If Λ_t is adequately estimated then our estimate of Z_t, $\hat{Z}_t = R_t/\hat{\Lambda}_t$ should be distributed as $N(0,1)$. Thus examining the distribution of $\hat{Z}_t$ serves as a powerful test of the validity of the proposed volatility estimators. The standard deviation of $\hat{Z}_t$ is of particular interest in that it indicates the existence of systematic bias in a proposed volatility estimator if standard deviation of $\hat{Z}_t$ is significantly different from one. Furthermore, if a volatility estimator is valid we would expect insignificant skewness, kurtosis and consequently the Jacque-Bera normality test would not reject $\hat{Z}_t$ as being normally distributed. If the assumed data generating process for intraday returns (11) is an adequate model then no endogenety problem should be present in the calculation of the $\hat{Z}_t$ statistic; in particular the independence between $e_{i,t}$ and ε_t is crucial to being able to estimate Λ_t from intraday data.

In the ARCH family of models, $\hat{\Lambda}_t$ is estimated as a 1-step ahead forecast from the given estimated model. However, empirical research has shown that $\hat{Z}_t$ is typically not normally distributed under this scenario. This may well be because of contemporaneous shocks impacting upon Λ_t which are not captured within the ARCH specification of daily volatility. In this sense the generalised historical volatility estimator has an unfair advantage over ARCH-type estimators in this evaluation.

In this analysis we have added two further assumptions to Anderson's (1994) model of daily volatility (3) in two significant ways. Firstly we have assume $\mu_t = 0$. This assumption is consistent with

(a) The empirical finding that the BIC chose models without a constant for the intraday transaction return series without exception.
(b) Finance theory which suggests that daily returns 'should' be white noise if arbitrage opportunities do not exist.

The second departure from Anderson's (1994) model of daily volatility involves restricting Z_t to be *iid N(0,1)*. This restriction is validated in Appendix 1. Descriptive statistics on $\hat{Z}_t$ for the various volatility estimators were obtained using all daily data for the years 1993 to 1995. Results are shown in table II.

< INSERT TABLE II HERE >

The mean of $\hat{Z}_t$ was close to zero but positive for all volatility estimators, indicating a positive mean return over the period of analysis. The standard deviation of $\hat{Z}_t$ was close to one (1.0094) for the Gen Hist(T) volatility estimator. The standard deviation of $\hat{Z}_t$ was further away from one for all of the other volatility estimators confirming the systematic biases inherent in these estimators as shown in Appendix 5.

Given that the Jacque-Bera normality test statistic is asymptotically distributed as a χ^2_2 random variable, the $\hat{Z}_t$ statistic based on the historical, MA(T), GH (T), and GH (5) volatility estimators could not be rejected as being normally distributed. This result is important because it indicates that the generalised historical volatility estimator based upon transactions data yields estimates of Λ_t which allow the restricted version of (5) to capture the behaviour of daily returns.

VIII. Estimation of General Volatility Framework Models

Implicit in the ARCH daily volatility modelling methodology is the belief that daily volatility not only varies on different trading days, but that daily volatility 'clusters' in high and low periods. It is these stylised facts regarding daily volatility that leads us to believe that modelling the series of daily activity variables $\{K_t\}_{t=1}^{T}$ in a general volatility framework (4), will allow us to model the time series properties of daily volatility, where daily volatility $\Lambda_t = g(K_t)$ is a function of the daily activity variable K_t. However we do not observe the daily activity variable K_t. Nonetheless K_t can be estimated for a given functional form $g(.)$ and an estimate of daily volatility $\hat{\Lambda}_t$ where $\hat{K}_t = g^{-1}\left(\hat{\Lambda}_t\right)$. To see that $g^{-1}\left(\hat{\Lambda}_t\right)$ is a consistent estimator of K_t, we note that $\operatorname{plim}\left(g^{-1}\left(\hat{\Lambda}_t\right)\right) = g^{-1}\left(\operatorname{plim}\left(\hat{\Lambda}_t\right)\right)$ by Slutsky's theorem and $g^{-1}\left(\operatorname{plim}\left(\hat{\Lambda}_t\right)\right) = g^{-1}(\Lambda_t) = K_t$ provided $\hat{\Lambda}_t$ is consistent for Λ_t, and subject to some weak regularity conditions on $g(.)$. Now given an estimate of K_t, we denote the estimation error η_t in K_t as

$$\hat{K}_t = K_t + \eta_t, \tag{17}$$

where $E[\eta_t] = 0$, $E\left[\eta_t^2\right] = \sigma_\eta^2$, $E[\eta_t, \eta_{t-i}] = 0, i \geq 1$ and $E[\eta_t, \varepsilon_{t-i}] = 0\ , i \geq 0$. Substituting (17) into (4) and rearranging we obtain

$$\begin{aligned}\hat{K}_t = \omega + \beta\Theta_{t-1} + \phi_1 \hat{K}_{t-1} + \cdots + \phi_p \hat{K}_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_t \varepsilon_{t-q} \\ + \eta_t - \phi_1 \eta_{t-1} - \cdots - \phi_p \eta_{t-p}.\end{aligned} \tag{18}$$

Equation (18) cannot be estimated with maximum likelihood estimation techniques because the error term in (18) is correlated with the regressors. Hence estimating (18) using estimates of K_t instead of K_t will produce biased estimates of the parameters in (18) and their standard errors. However we note that as the number of intraday transactions used to estimate

Λ_t increases, the variance of the estimation error σ_η^2 approaches zero provided $\hat{\Lambda}_t$ is a consistent estimator of Λ_t. Consequently replacing K_t with $\hat{K}_t$ in (18) and then implementing maximum likelihood estimation techniques will produce consistent parameter estimates of (18) as the number of intraday observations approaches infinity. Due to the high efficiency of the estimates of K_t, we thus do not believe the inherent bias would effect the results much, given the large number of intraday observations available for estimating Λ_t.

The daily generalised historical volatility estimates based upon transactions data over the period 1993 to 1995 was thus modelled as an ARMA(0,0), AR(1), AR(2), AR(3), MA(1), MA(2), ARMA(1,1), ARMA(1,2), ARMA(2,1) and ARMA(2,2) process with a constant. The functional form of $g(.)$ was initially defined as $K_t = \Lambda_t^2$. That is, models of the form

$$\Lambda_t^2 = \delta + \phi_1 \Lambda_{t-1}^2 + \phi_2 \Lambda_{t-2}^2 + \phi_3 \Lambda_{t-3}^2 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \tag{19}$$

were estimated. Clearly the models defined in (19) are all special cases of the general volatility framework defined in (4). Given that Λ_t^2 is unobserved, estimates of Λ_t^2 based upon intraday transaction data were used as proxies for Λ_t. The BIC of the above models were recorded and the model with the maximum BIC was considered to be the 'best' model of the time-series behaviour of daily volatility. The model of choice according to the criteria of maximising the BIC for (19) was

$$\underset{(2.992)}{\Lambda_t^2 = 0.000002014} + \underset{(41.16)}{0.93074\Lambda_{t-1}^2} + \varepsilon_t - \underset{(18.12)}{0.742198\varepsilon_{t-1}} \tag{20}$$

(t-ratio's are shown in parenthesis)*

That is, an ARMA(1,1) model was selected as the 'best' model of the time series properties of daily volatility. This is important because it shows that daily volatility is an autocorrelated time series and that it is mean reverting since the AR(.) component indicates that Λ_t^2 is stationary. Furthermore the coefficient of $\hat{\Lambda}_{t-1}^2$ was both highly significant and had a value of 0.93074, indicating that daily volatility is highly persistent. The coefficient of ε_{t-1} was -0.742198 and is both highly significant and had a large value, indicating that their exists a significant amount of information in the previous error term.

To further test the effect of the estimation error η_t on parameter estimates in equation (20), the equation was estimated using instrumental variable techniques where $\hat{\Lambda}^2_{t-2}$ was used as the instrument for Λ^2_{t-1}. It is easily verified that $\hat{\Lambda}^2_{t-2}$ is uncorrelated with the error term $\xi_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \eta_t - \phi_1\eta_{t-1}$ and is highly correlated with Λ^2_{t-1}. The coefficient of $\hat{\Lambda}^2_{t-1}$ was estimated as (0.9485) and further testing revealed that this was not significantly different from (0.9307), the coefficient of Λ^2_{t-1} in (20). This result lends some evidence to the belief that the effect of replacing estimates of K_t in (4) rather than K_t has little significant effect on parameter estimates in (20).

The lognormal functional form for general stochastic volatility models was also estimated within the general stochastic volatility framework. The daily generalised historical volatility estimates based upon transactions data over the period 1993 to 1995 were thus again modelled as ARMA(0,0), AR(1), AR(2), AR(3), MA(1), MA(2), ARMA(1,1), ARMA(1,2), ARMA(2,1) and ARMA(2,2) processes with a constant. The functional form of $g(.)$ was defined as $K_t = Log_e(\Lambda_t)$. The model of choice according to maximum *BIC* criteria was again an ARMA(1,1) model.

$$Log_e(\Lambda_t) = \underset{(-153.6)}{-2.300} + \underset{(45.81)}{0.9302} Log_e(\Lambda_{t-1}) + \varepsilon_t - \underset{(17.35)}{0.6928}\varepsilon_{t-1}. \tag{21}$$

(t-ratio is shown in parenthesis)*

This stochastic volatility model differs from Anderson's (1994) lognormal stochastic volatility model in that an MA(1) disturbance term is included in (21). Thus to test further whether model (21) could be rejected in favour of (9), likelihood ratio tests were conducted. We note that equation (9) is a restricted version of equation (21) where by restricting the coefficient of the moving average term to zero $(\theta_1 = 0)$, we obtain (9). However performing likelihood ratio (*LR*) tests to examine these restrictions is not easy. Evaluating the likelihood in stochastic volatility models is difficult, due to the unobservability of Λ_t. Furthermore, the *LR* statistics are unlikely to have standard asymptotic distributions. These theoretical questions are the subject of future research. However to get an indication of the likely plausibility of the different models, we compute *LR* statistics for this hypothesis where the likelihood for the stochastic volatility models is based on treating $\hat{\Lambda}_t$ as the observed Λ_t. Hence we define $L_{(21)}$ = maximum log of the likelihood function for (21), and $L_{(9)}$ is similarly defined. The likelihood ratio test statistic under the null hypothesis $H_0\colon \theta_1 = 0$ and alternative $H_1\colon \theta_1 \neq 0$ is given by $LR = -2\left(L_{(21)} - L_{(9)}\right) \sim \chi_k^2$ where k = number of restrictions.

The above test yielded a test statistic of 69.38 and consequently rejected $H_0\colon \theta_1 = 0$ where $\chi_1^2 = 3.84$ at 95% confidence. Even though this result is not based on valid asymptotic theory, the rejection is so strong that it is almost certain to stand up to a more proper analysis. The importance of the above testing lies in the result that the LR test rejected an AR(1) functional form in favour of a lognormal general stochastic volatility model with an ARMA(1,1) functional form.

A similar model to the EGARCH(1,1) model within the general stochastic volatility framework was also estimated with an MA(1) disturbance term.

$$Log_e\left(\Lambda_t^2\right) = -4.5961 + \underset{(47.59)}{0.9342} Log_e\left(\Lambda_{t-1}^2\right) - \underset{(-4.155)}{0.02652} \frac{R_{t-1}}{\Lambda_{t-1}} + \varepsilon_t - \underset{(17.96)}{0.7045} \varepsilon_{t-1}. \tag{22}$$

(t-ratio's are shown in parenthesis)*

The coefficient of $\frac{R_{t-1}}{\Lambda_{t-1}}$ is negative as expected, and it is significant. Hence 'bad' news as reflected in the previous day's standardised return significantly impacts on today's level of volatility. A daily volatility model with a similar functional form to the GJR model was also estimated with an MA(1) disturbance term.

$$\Lambda_t^2 = 0.00002784 + \underset{(47.76)}{0.9450} \Lambda_{t-1}^2 + \underset{(6.420)}{0.06588} S_{t-1}^{-} \Lambda_{t-1}^2 + \varepsilon_t - \underset{(20.94)}{0.7897} \varepsilon_{t-1}. \tag{23}$$

(t-ratio's are shown in parenthesis)*

As expected the coefficient of $S_{t-1}^{-} \hat{\Lambda}_{t-1}^2$ is positive and significant, again reflecting the notion that 'bad' news impacts positively on daily volatility levels.

IX. Testing the GARCH(1,1) Specification of Daily Volatility

Recall that a GARCH(1,1) model can be written in the form $R_t|\ell_{t-1} \sim N(0,h_t)$, where ℓ_{t-1} is the information set at time *t-1* and $h_t = \alpha_0 + \alpha_1 R_{t-1}^2 + \beta h_{t-1}$. This GARCH model can also be written in the form of an ARMA(1,1) model (see Bollerslev, 1986). If it is assumed that R_t^2 is an unbiased estimator of the true daily volatility h_t, then h_t can be written as $h_t = E[R_t^2]$ or equivalently

$$R_t^2 - h_t = \upsilon_t \quad , \tag{24}$$

where $E[\upsilon_t] = 0$ and $E[\upsilon_t, \upsilon_{t-i}] = 0$ where $i \geq 1$. Substituting (24) into (6) we obtain

$$R_t^2 = \alpha_0 + (\alpha_1 + \beta) R_{t-1}^2 + \upsilon_t - \beta \upsilon_{t-1}. \tag{25}$$

That is, if daily volatility can be modelled as the GARCH(1,1) process defined in (6), then the unbiased estimator of daily volatility R_t^2 will follow the ARMA(1,1) process in (25). This analysis can be further extended by noting that any unbiased estimator of daily volatility; and in particular $\hat{\Lambda}_t^2$ can be written as

$$\hat{\Lambda}_t^2 - h_t = \vartheta_t \tag{26}$$

where $E[\vartheta_t] = 0$ and $E[\vartheta_t, \vartheta_{t-i}] = 0$, $i \geq 1$. Substituting (26) into (9) we obtain

$$\begin{aligned} \hat{\Lambda}_t^2 - \vartheta_t &= \alpha_0 + \alpha_1 R_{t-1}^2 + \beta\left(\hat{\Lambda}_{t-1}^2 - \vartheta_{t-1}\right) \\ \Rightarrow \quad \hat{\Lambda}_t^2 &= \alpha_0 + \alpha_1 R_{t-1}^2 + \beta \hat{\Lambda}_{t-1}^2 + \vartheta_t - \beta \vartheta_{t-1}. \end{aligned} \tag{27}$$

What is significant about this result is that if the GARCH(1,1) specification of the 'true' daily volatility h_t is correct then for any unbiased estimator of daily volatility, the estimator will follow the ARMA(1,1) process defined in (27) where both the coefficient of $\hat{\Lambda}_{t-1}^2$ and ϑ_{t-1} are equal. If the coefficient of $\hat{\Lambda}_{t-1}^2$ and ϑ_{t-1} are not equal then the GARCH(1,1) model given in (9) may not be the correct specification of the time series model of daily volatility. Note that this specification test requires the assumption that $E[\vartheta_t, \vartheta_{t-i}] = 0$, an assumption that cannot be empirically verified but does however seem to be a reasonable assumption. Hence equation (27) was estimated using the generalised historical volatility estimator based on transactions data and results are reported in (28).

$$\hat{\Lambda}_t^2 = \underset{(13.85)}{0.00002765} + \underset{(44.14)}{0.9386}\hat{\Lambda}_{t-1}^2 + \underset{(4.213)}{0.0358}R_{t-1}^2 + \vartheta_t - \underset{(19.47)}{0.7748}\vartheta_{t-1}. \qquad (28)$$

BIC=-14,228 *(* t-ratio's are shown in parenthesis)*

To test if the coefficient of $\hat{\Lambda}_{t-1}^2$ and ϑ_{t-1} are significantly different we define β_{AR} as the coefficient of $\hat{\Lambda}_{t-1}^2$ in (28), β_{MA} as the coefficient of ϑ_{t-1} in (28), b_{AR} as an estimate of the coefficient of $\hat{\Lambda}_{t-1}^2$ in (28) and b_{MA} as an estimate of the coefficient of ϑ_{t-1} in (28). Under the null hypothesis $H_0: \beta_{AR} - \beta_{MA} = 0$ against the alternative $H_1: \beta_{AR} - \beta_{MA} \neq 0$ a natural test statistic is defined as $t = \dfrac{b_{AR} - b_{MA}}{\sqrt{V\hat{a}r(b_{AR} - b_{MA})}} = \dfrac{b_{AR} - b_{MA}}{\sqrt{V\hat{a}r(b_{AR}) + V\hat{a}r(b_{MA}) - 2C\hat{o}v(b_{AR}, b_{MA})}} = 6.502$. At 5% significance, H_0 is rejected and consequently evidence has been presented that the GARCH(1,1) specification of the time series behaviour of daily volatility is incorrect.

A further observation that can be made about (28) is that the coefficient of R_{t-1}^2 is 0.036 whilst the coefficient of Λ_{t-1}^2 is 0.939. Clearly the previous day's estimate of daily volatility based upon transactions data with the generalised historical volatility estimator contains far more information in regard to the current level of volatility Λ_t^2 than does R_{t-1}^2.

X. Conclusions

After specifying a general stochastic volatility framework, a new estimator of daily volatility which is a generalisation of the historical volatility estimator has been proposed. This new estimator of volatility is called the generalised historical volatility estimator, and has been developed to take into account the autocorrelation typically found in an intraday return series.

We then went on to consider the case when the intraday return series is stationary, and adequately modelled as an ARMA(p,q) process. Under this assumption it was established that generalised historical volatility is given by $\Lambda_t^2 = \left(\dfrac{1 - \theta_{1,t} - \theta_{2,t} - \cdots - \theta_{q_t,t}}{1 - \phi_{1,t} - \phi_{2,t} - \cdots - \phi_{p_t,t}}\right)^2 n_t\sigma_t^2$. Substituting parameter estimates into this expression will yield a consistent and efficient estimator of daily volatility. Evidence also suggests that unless the transaction return series is used to estimate daily volatility, daily volatility estimates may be systematically biased. We conclude from this analysis that intraday data can be used to effectively estimate daily volatility. In particular the generalised historical volatility estimator based upon transactions data outperformed all of the other daily volatility estimators considered on the basis that

(a) it correlates more highly with proxies for information arrival
(b) daily returns, when conditioned upon the generalised historical volatility estimates based upon transactions data, are normally distributed with variance close to one.

This paper has employed a two step procedure to model the time series behaviour of daily volatility, namely

(a) to estimate daily volatility with the generalised historical volatility estimator for T trading days to produce a series of daily volatility estimates $\left\{\hat{\Lambda}_t\right\}_{t=1}^{T}$,
(b) then model the series $\left\{\hat{K}_t\right\}_{t=1}^{T}$ as an ARMA process within the general stochastic volatility framework for a given functional form $g(.)$.

The generalised historical volatility estimator based upon transactions data can within the general volatility framework, successfully model the time series properties of daily volatility utilising a wide range of functional forms.

The literature concerning the time series properties of daily volatility in financial markets since Engle's (1982) seminal paper has been dominated by the ARCH family of volatility models and also the stochastic volatility models. The modelling methodology proposed in this paper may well be a viable alternative to existing volatility modelling methodologies. Furthermore the generalised historical volatility estimator could potentially be utilised in further financial market analysis particularly at the market micro-structure level where previous research has found the ARCH family of models to be mis-specified in regards to levels of volatility persistence. Research in option pricing theory, and in particular the effects of time varying volatility on the Black-Scholes formula may also benefit from the kind of highly efficient volatility analysis that the generalised historical volatility estimator can be used for. Given that the notion of risk (or volatility) is a key concept in finance theory, it is a matter of further theoretical development and empirical research to see how the generalised historical volatility estimator linked with the general stochastic volatility framework may be implemented to shed light on other problems.

Appendix 1 The Volatility of an ARMA(p,q) Return Process

Recall from (11) that an ARMA process in the intraday return series can be written in the form,

$$\text{(A1)} \quad \phi_t(L)r_{i,t} = \frac{\phi_t(1)\mu_t}{n_t} + \frac{\Lambda_t\phi_t(1)\theta_t(L)}{\sqrt{n_t}\,\theta_t(1)}e_{i,t},$$

where L = usual lag operator,

$\phi_t(L) = 1-\phi_{1,t}L-\phi_{2,t}L^2-\phi_{3,t}L^3-\cdots-\phi_{p_t,t}L^{p_t}$,

$\theta_t(L) = 1-\theta_{1,t}L-\theta_{2,t}L^2-\theta_{3,t}L^3-\cdots-\theta_{p_t,t}L^{q_t}$,

p_t and q_t are the orders of the autoregressive and moving average components respectively on day t,

$\phi_{i,t}$ and $\theta_{i,t}$ are the i^{th} autoregressive and moving average parameters respectively on day t,

n_t = number of intraday returns,

μ_t = mean daily return for day t,

$e_{i,t} \sim iid\ N(0,1)$.

We wish to prove that if the intraday return series can be correctly modelled as it is specified in (A1), then daily returns can be written as $R_t = \mu_t + \Lambda_t Z_t$. Now (A1) can also be rewritten as

$$\text{(A2)} \quad r_{i,t} = \frac{\phi_t(1)\mu_t}{n_t\phi_t(L)} + \frac{\Lambda_t\phi_t(1)\theta_t(L)}{\sqrt{n_t}\,\theta_t(1)\phi_t(L)}e_{i,t}.$$

However daily returns R_t are simply the sum of intraday returns and consequently daily returns can be written as

$$\text{(A3)} \quad R_t = \sum_{i=1}^{n_t} r_{i,t} = \frac{\phi_t(1)\mu_t}{n_t}\sum_{i=1}^{n_t}\frac{1}{\phi_t(L)} + \frac{\Lambda_t\phi_t(1)}{\sqrt{n_t}\,\theta_t(1)}\sum_{i=1}^{n_t}\frac{\theta_t(L)}{\phi_t(L)}e_{i,t}.$$

Our strategy in developing this proof is to show that

$$\text{(A4)} \quad \mu_t = \frac{\phi_t(1)\mu_t}{n_t}\sum_{i=1}^{n_t}\frac{1}{\phi_t(L)}$$

and that

$$\text{(A5)} \quad \Lambda_t Z_t = \frac{\Lambda_t\phi_t(1)}{\sqrt{n_t}\,\theta_t(1)}\sum_{i=1}^{n_t}\frac{\theta_t(L)}{\phi_t(L)}e_{i,t} \quad \text{where } Z_t = \frac{1}{\sqrt{n_t}}\sum_{i=1}^{n_t}e_{i,t}.$$

Substituting (A4) and (A5) into (A3) will establish that daily returns can be correctly modelled as $R_t = \mu_t + \Lambda_t Z_t$ if the intraday return data generating process is given by (A1). We initially show that $\sum_{i=1}^{n_t}\frac{1}{\phi_t(L)} = \frac{n_t}{\phi_t(1)}$. Now

$$\text{(A6)} \quad \sum_{i=1}^{n_t} \frac{1}{\phi_t(L)} = \frac{1}{1+\phi_{1,t}+\cdots+\phi_{1,p_t}} + \frac{1}{1+\phi_{1,t}+\cdots+\phi_{1,p_t}} + \cdots + \frac{1}{1+\phi_{1,t}+\cdots+\phi_{1,p_t}}$$
$$= \frac{n_t}{\phi_t(1)}.$$

Substituting (A6) into (A4) the desired result is obtained. We now go on to prove (A5) by initially showing that

$$\text{(A7)} \quad \sum_{i=1}^{n_t} \frac{\theta_t(L)}{\phi_t(L)} e_{i,t} = \frac{\sqrt{n_t}\,\theta_t(L)}{\phi_t(L)} Z_t.$$

Showing (A7) and substituting into (A5) completes the proof. Now the Wold's decomposition (1938) allows the expression $\frac{\theta_t(L)}{\phi_t(L)} e_{i,t}$ in equation (A7) to be written in the form of an infinite order MA process.

$$\text{(A8)} \quad \frac{\theta_t(L)}{\phi_t(L)} e_{i,t} = \Psi_t(L) e_{i,t} \quad \text{where } \Psi_t(L) = 1 + \Psi_{1,t} L + \Psi_{2,t} L^2 + \Psi_{3,t} L^3 + \cdots$$
$$= \sum_{j=0}^{\infty} \Psi_{j,t} L^j \quad \textit{and } \Psi_{0,t} = 1.$$

If we assume that the intraday return series is a stationary process we will have to assume that

$$\text{(A9)} \quad \sum_{j=0}^{\infty} \Psi_{j,t}^2 < \infty \quad \textit{and} \quad \sum_{j=0}^{\infty} j\left|\Psi_{j,t}\right| < \infty .$$

Now we define $\Psi_{k,t}^*$ as

$$\text{(A10)} \quad \Psi_{k,t}^* = -\left(\Psi_{k+1,t} + \Psi_{k+2,t} + \Psi_{k+3,t} + \cdots\cdots\right) = -\sum_{j=k+1}^{\infty} \Psi_{j,t}$$

We now employ the Beveridge-Nelson (1981) decomposition to simplify the analysis.

$$\text{(A11)} \quad \Psi_t(L) = \Psi_t(1) + (1-L)\Psi_t^*(L)$$
$$\text{where } \Psi_t(1) = \Psi_t = 1 + \Psi_{1,t} + \Psi_{2,t} + \Psi_{3,t} + \cdots = \sum_{j=0}^{\infty} \Psi_{j,t},$$
$$\text{and } \Psi_t^*(L) = \Psi_{0,t}^* + \Psi_{1,t}^* L + \Psi_{2,t}^* L^2 + \cdots + \Psi_{i,t}^* L^i = \sum_{j=0}^{i} \psi_{j,t}^* L^j .$$

Therefore

$$\text{(A12)} \quad \begin{aligned} \Psi_t(L) e_{i,t} &= \left(\Psi_t(1) + (1-L)\Psi_t^*(L)\right) e_{i,t} \\ &= \Psi_t(1) e_{i,t} + (1-L)\Psi_t^*(L) e_{i,t} \\ &= \Psi_t(1) e_{i,t} + \Psi_t^*(L) e_{i,t} - \Psi_t^*(L) e_{i-1,t} \\ &= \Psi_t(1) e_{i,t} + \Psi_t^*(L) \Delta e_{i,t} \end{aligned}$$

Substituting (A12) into (A7), we have

$$\text{(A13)} \quad \sum_{i=1}^{n_t} \frac{\theta_t(L)}{\phi_t(L)} e_{i,t} = \Psi_t(1) \sum_{i=1}^{n_t} e_{i,t} + \Psi_t^*(L) \sum_{i=1}^{n_t} \Delta e_{i,t}$$

But

$$\text{(A14)} \quad \begin{aligned} \sum_{i=1}^{n_t} \Delta e_{i,t} &= \Delta e_{1,t} + \Delta e_{2,t} + \cdots + \Delta e_{n_t,t} \\ &= (e_{1,t} - e_{0,t}) + (e_{2,t} - e_{1,t}) + (e_{3,t} - e_{2,t}) + \cdots + (e_{n_t,t} - e_{n_t-1,t}) \\ &= e_{n_t,t} - e_{0,t} \end{aligned}$$

Hence combining (A13) and (A14) we have

$$\text{(A15)} \quad \sum_{i=1}^{n_t} \frac{\theta_t(L)}{\phi_t(L)} e_{i,t} = \Psi_t(1) \sum_{i=1}^{n_t} e_{i,t} + \Psi_t^*(L)\left(e_{n_t,t} - e_{0,t}\right)$$

Now as $n_t \to \infty$ the term $\Psi_t(1) \sum_{i=1}^{n_t} e_{i,t}$ in (A15) becomes larger whilst the term $\Psi_t^*(L)\left(e_{n_t,t} - e_{0,t}\right)$ remains constant and consequently

$$\text{(A16)} \quad \sum_{i=1}^{n_t} \frac{\theta_t(L)}{\phi_t(L)} e_{i,t} \approx \frac{\sqrt{n_t}\,\theta_t(1)}{\phi_t(1)} Z_t \quad \text{when } n_t \text{ is large,}$$

and the proof is complete. Furthermore, since we have defined Z_t as $Z_t = \frac{1}{\sqrt{n_t}} \sum_{i=1}^{n_t} e_{i,t}$, Z_t can be rewritten as $Z_t = \frac{1}{\sqrt{n_t}} \sum_{i=1}^{n_t} e_{i,t} = \sqrt{n_t}\,\overline{e}_t$ where $\overline{e}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} e_{i,t}$ and so by the Central Limit theorem it is readily seen that $Z_t \sim N(0,1)$.

Appendix 2 - The Efficiency and Bias of the Generalised Historical Volatility Estimator when Returns Follow an ARMA(p,q) Process.

Model	ARMA(1,1)	MA(1)	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
True Volatility	33.5410	26.8328	44.7214	63.8877	68.8021	68.8021	74.5356

n=2000

Mean	35.6576	26.8325	44.6810	63.9175	67.1858	68.9967	75.5096
Std Dev	1.9671	1.0252	0.9900	2.3371	5.0253	5.7410	5.7624
Bias	2.1165	-0.0003	-0.0404	0.0298	-1.6163	0.1947	0.9740
RMSE	2.8895	1.0252	0.9908	2.3373	5.2788	5.7443	5.8442

n=500

Mean	33.5018	26.9155	44.6365	64.3026	69.0422	68.9487	74.1762
Std Dev	1.8688	1.6321	2.5082	3.4920	3.6630	3.8821	4.8655
Bias	-0.0392	0.0827	-0.0849	0.4150	0.2401	0.1466	-0.3594
RMSE	1.8692	1.6341	2.5097	3.5165	3.6708	3.8848	4.8788

n=100

Mean	32.9212	26.3021	44.3393	63.7164	67.6980	68.6357	74.5455
Std Dev	3.9193	3.1542	5.5740	8.2713	7.9566	8.7389	9.6819
Bias	-0.6198	-0.5307	-0.3820	-0.1713	-1.1041	-0.1664	0.0099
RMSE	3.9680	3.1986	5.5871	8.2731	8.0329	8.7405	9.6819

n=1

Mean	32.7597	26.1810	45.5190	63.0706	67.1952	67.9819	73.5764
Std Dev	24.7212	19.7564	32.8473	47.8533	50.4678	51.4427	55.7266
Bias	-0.7813	-0.6518	0.7977	-0.8171	-1.6069	-0.8202	-0.9592
RMSE	24.7335	19.7672	32.8570	47.8603	50.4934	51.4493	55.7349

<u>True model used to generate the intraday return series.</u>

ARMA(1,1) $\left(r_i = 0.2r_{i-1} + e_i - 0.4e_{i-1}\right)$

MA(1) $\left(r_i = e_i - 0.4e_{i-1}\right)$

AR(0) $\left(r_i = e_i\right)$

AR(1) $\left(r_i = 0.3r_{i-1} + e_i\right)$

AR(2) $\left(r_i = 0.25r_{i-1} + 0.1r_{i-2} + e_i\right)$

AR(3) $\left(r_i = 0.2r_{i-1} + 0.1r_{i-2} + 0.05r_{i-3} + e_i\right)$

AR(4) $\left(r_i = 0.2r_{i-1} + 0.1r_{i-2} + 0.05r_{i-3} + 0.05r_{i-4} + e_i\right)$

Appendix 3 - The Efficiency and Bias of the Generalised Historical Volatility Estimator when Returns are assumed to follow a White Noise Process.

ARMA(1,1) $\left(r_i = 0.2r_{i-1} + e_i - 0.4e_{i-1}\right)$

True Volatility = 33.5410

	n=2000	n=500	n=100
Mean	45.6706	37.7527	34.3921
Std Dev	0.7441	1.2090	2.3356
Bias	12.1296	4.2117	0.8510
RMSE	12.1524	4.3818	2.4858

AR(2) $\left(r_i = 0.25r_{i-1} + 0.1r_{i-2} + e_i\right)$

True Volatility = 68.8021

	n=2000	n=500	n=100
Mean	46.8136	59.6112	66.9109
Std Dev	0.8280	1.9318	4.6270
Bias	-21.9885	-9.1909	-1.8912
RMSE	22.0041	9.3917	4.9985

MA(1) $\left(r_i = e_i - 0.4e_{i-1}\right)$

True Volatility = 26.8328

	n=2000	n=500	n=100
Mean	48.1913	33.4679	28.2504
Std Dev	0.8383	1.1084	1.9092
Bias	21.3584	6.6351	1.4176
RMSE	21.3749	6.7270	2.3779

AR(3) $\left(r_i = 0.2r_{i-1} + 0.1r_{i-2} + 0.05r_{i-3} + e_i\right)$

True Volatility = 68.8021

	n=2000	n=500	n=100
Mean	46.3260	57.7926	66.4508
Std Dev	0.8036	1.8873	4.5899
Bias	-22.4760	-11.0095	-2.3513
RMSE	22.4904	11.1701	5.1571

AR(1) $\left(r_i = 0.3r_{i-1} + e_i\right)$

True Volatility = 63.8877

	n=2000	n=500	n=100
Mean	46.9105	58.4868	62.7863
Std Dev	0.8208	1.8646	4.3386
Bias	-16.9771	-5.4008	-1.1013
RMSE	16.9969	5.7136	4.4762

AR(4) $\left(r_i = 0.2r_{i-1} + 0.1r_{i-2} + 0.05r_{i-3} + 0.05r_{i-4} + e_i\right)$

True Volatility = 74.5356

	n=2000	n=500	n=100
Mean	46.5281	58.5527	70.8396
Std Dev	0.8161	1.9575	4.8956
Bias	-28.0075	-15.9829	-3.6960
RMSE	28.0194	16.1023	6.1341

(Note - the 'true' model is shown in the top row of each table.)

Appendix 4 - The Efficiency and Bias of the Generalised Historical Volatility Estimator when Returns are assumed to follow an MA(1) process.

ARMA(1,1) $\left(r_i = 0.2r_{i-1} + e_i - 0.4e_{i-1}\right)$

True Volatility = 33.5410

	n=2000	n=500	n=100
Mean	35.4818	33.5421	33.2616
Std Dev	1.1773	1.9472	4.2239
Bias	1.9408	0.0011	-0.2794
RMSE	2.2699	1.9472	4.2331

AR(3) $\left(r_i = 0.2r_{i-1} + 0.1r_{i-2} + 0.05r_{i-3} + e_i\right)$

True Volatility = 68.8021

	n=2000	n=500	n=100
Mean	53.9072	67.4764	68.2064
Std Dev	1.3005	3.2565	8.3124
Bias	-14.8949	-1.3257	-0.5957
RMSE	14.9515	3.5160	8.3337

AR(1) $\left(r_i = 0.3r_{i-1} + e_i\right)$

True Volatility = 63.8877

	n=2000	n=500	n=100
Mean	57.4434	63.8603	63.3362
Std Dev	1.2938	3.2637	7.8097
Bias	-6.4442	-0.0273	-0.5514
RMSE	6.5728	3.2638	7.8291

AR(4) $\left(r_i = 0.2r_{i-1} + 0.1r_{i-2} + 0.05r_{i-3} + 0.05r_{i-4} + e_i\right)$

True Volatility = 74.5356

	n=2000	n=500	n=100
Mean	54.2278	70.6155	73.8869
Std Dev	1.3166	3.2668	8.8998
Bias	-20.3078	-3.9201	-0.6487
RMSE	20.3504	5.1029	8.9234

AR(2) $\left(r_i = 0.25r_{i-1} + 0.1r_{i-2} + e_i\right)$

True Volatility = 68.8021

	n=2000	n=500	n=100
Mean	55.5410	68.2719	68.2057
Std Dev	1.2944	3.3622	8.3532
Bias	-13.2611	-0.5302	-0.5964
RMSE	13.3241	3.4038	8.3745

(Note - the 'true' model is shown in the top row of each table.)

Appendix 5 Correlogram of the Volatility Estimators and Proxies for Information Arrival.

Correlogram.

	Simple	Hist (T)	MA (T)	MA (5)	Gen Hist (T)	Gen Hist (5)	TRANSCNT	Volume(1)	Volume(2)
Simple	1.0000								
Hist (T)	0.3973	1.0000							
MA (T)	0.4142	0.9352	1.0000						
MA (5)	0.4198	0.8378	0.8698	1.0000					
Gen Hist (T)	0.4152	0.9354	0.9772	0.8805	1.0000				
Gen Hist (5)	0.4102	0.8720	0.8984	0.9328	0.9087	1.0000			
TRANSCNT	0.3072	0.6704	0.7831	0.6751	0.7220	0.6913	1.0000		
Volume(1)	0.2735	0.4436	0.5174	0.4478	0.4496	0.4255	0.5405	1.0000	
Volume(2)	0.3113	0.5669	0.5844	0.5192	0.5465	0.5030	0.4358	0.8622	1.0000

Appendix 6 Detrending and Deseasonalising Volume

To deseasonalize and detrend the volume series, equation (A6.1) was estimated with OLS.

(A6.1) $$V_t = a + bT_t + ct + \varepsilon_t$$

where *a, b, c are constant parameters,*

t = *time trend (=1 on 4/1/93),*

T_t = *Total number of trading days since the last contract maturity time.*

V_t = *Total volume on day t.*

It is implicit in model (A6.1) that volume increases linearly toward contract maturity time and that their is a linear increase in volume over the trading period January 1993 to December 1995. Regression results are shown in (A6.2).

(A6.2) $$\hat{V}_t = \underset{(14.18)}{32{,}865} + \underset{(12.76)}{632.004T_t} + \underset{(9.40)}{39.007t}$$

where t-ratios are in parenthesis.

Residuals from regression (A6.2) are in effect an index of 'unexplained' volume. It is the relationship between this index of unexplained volume and volatility that we are primarily concerned with. This index of 'unexplained' volume shall be referred to as Volume(2). Figure (2) shows total contract volume (Volume(1)) as well as a deseasonalised and detrended index of trading volume (Volume(2)) over the period January 1993 to December 1995.

< INSERT FIGURE 2 HERE>

Bibliography

Anderson, Goran, 1995, Variance estimation in the presence of drift and non-trading, Working paper No 1994:15, School of Economics and Commercial Law, University of Goteborg.

Anderson, Torben G, 1994, Stochastic autoregressive volatility: A framework for volatility modelling, *Mathematical Finance*, 4, 75-102.

Anderson, Torben G and Tim Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance,* 4, 115-158.

Anderson, Torben G and Tim Bollerslev, 1998, Deutsche mark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies, *Journal of Finance*, 53, 219-265.

Beckers, S, 1983, Variances of security prices based on high, low and closing prices, *Journal of Business*, 56, 97-112.

Bollerslev, Tim, 1986, Generalised autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

Danielsson, Jon, 1994, Stochastic volatility in asset prices, estimation with simulated maximum likelihood, *Journal of Econometrics*, 64, 375-400.

Engle, Robert and Victor Ng, 1993, Measuring and testing the impact of news on volatility, *Journal of Finance*, 48, 1749-1778.

Figlewski, Stephen, 1997 , Forecasting Volatility , *New York University Salomon Center, Monograph series in Finance and Economics* (Financial Markets, Institutions and Instruments), 6.

Glosten, L, R Jagannathan and D Runkle, 1990, Relationship between the expected value and the volatility of nominal excess return on stocks, Working paper, Department of Finance, Columbia University.

Hamilton, James, 1994 , Time Series Analysis, (Princeton University Press).

Hasbrouck, Joel and Thomas Ho, 1987, Order arrival, quote behaviour, and the return-generating process, *Journal of Finance*, 42, 1035-1048.

Harris, Lawerence, 1986, Cross-security tests of the mixture of distributions hypothesis, *Journal of Financial and Quantitative Analysis,* 21, 39-46.

Harris, Lawerence, 1987, Transaction data test of the mixture of distributions hypothesis, *Journal of Financial and Quantitative Analysis,* 22, 127-141.

Harvey, Andrew, Esther Ruiz and Neil Shephard, 1994, Multivariate stochastic variance models, *Review of Economic Studies*, 61, 247-264.

Jacquier, E , N G Polson and P Rossi, 1994, Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics.*

Kyle, Albert S, 1985 ,Continuous Auctions and Insider Trading, *Econometrica*, 53, 1315-1335.

Nelson, Daniel B, 1990, Conditional hetroskedasticity in asset returns : A new approach, *Econometrica*, 59, 347-370.

Parkinson, M, 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business*, 53, 61-65.

Rogers, G and Satchell, 1991, Estimating variance from high, low and closing prices, *The Annals of Applied Probability*, 1, 504-512.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance*, 39,1127-1139.

Stoll, Hans R and Robert E Whaley, 1990, The dynamics of stock index and stock index futures returns, *Journal of Financial and Quantitative Analysis*, 25, 441-468.

Taylor, S, 1994, Modeling stochastic volatility, *Mathematical Finance*, 4, 183-204.

Table I 'Frequency of AR(.) models chosen in the transaction return series'

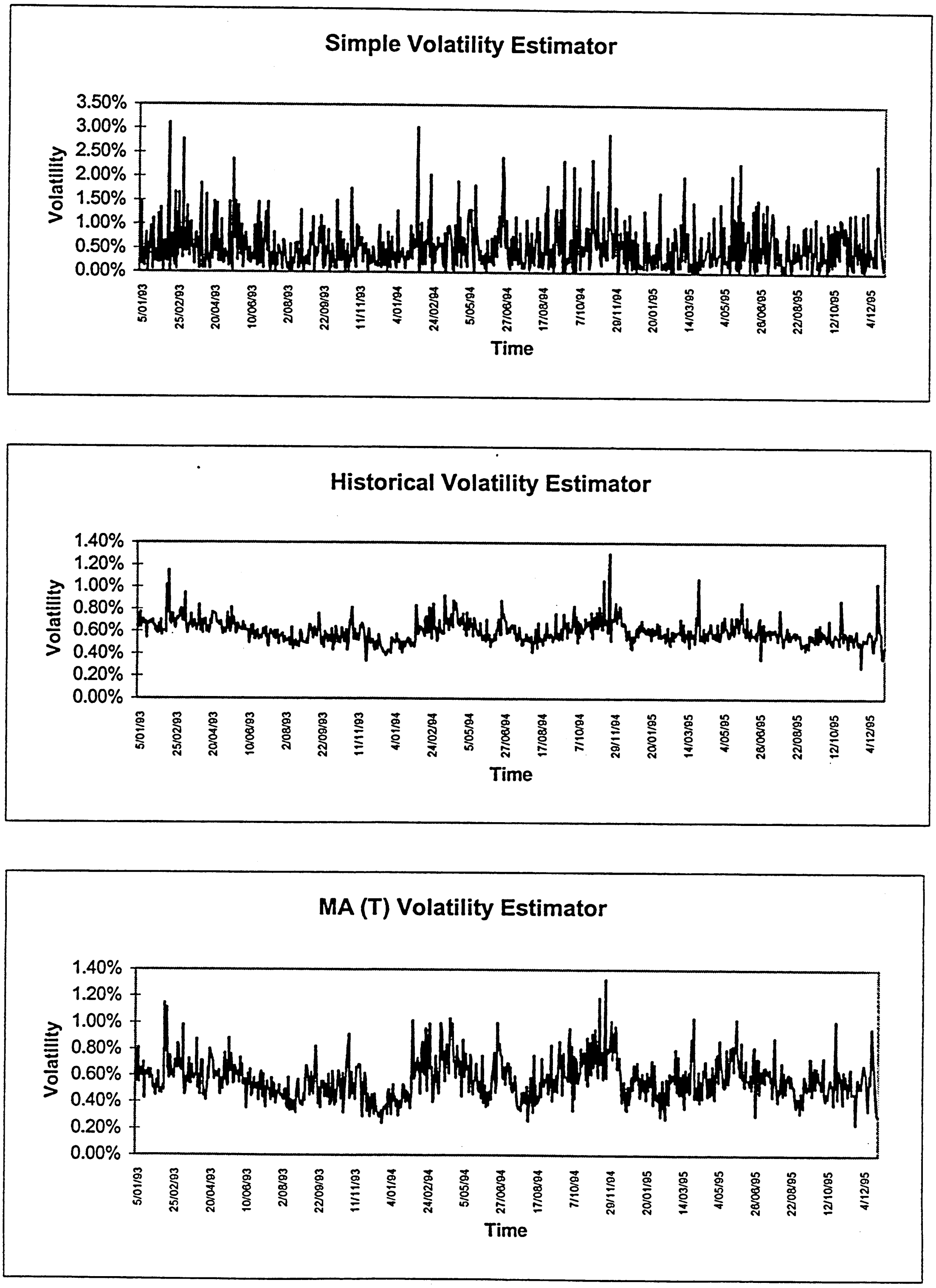
Model	Frequency (Transaction returns)	Frequency (Five minute returns)
AR(0)	32	679
AR(1)	15	41
AR(2)	588	16
AR(3)	102	8
AR(4)	4	0
AR(5)	0	0
AR(6)	0	0
AR(7)	1	0
AR(8)	1	0
AR(9)	1	0
AR(10)	0	0

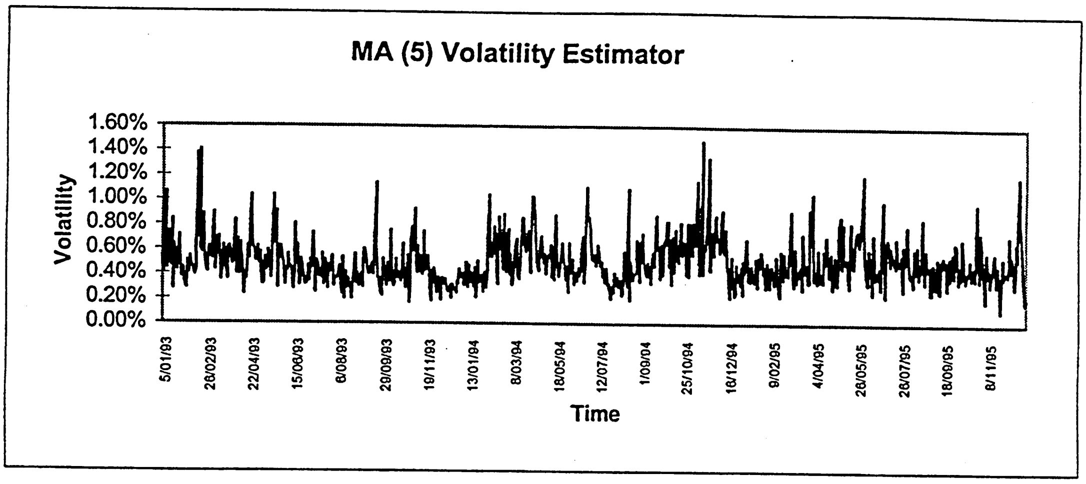
Table II Summary statistics on the $\hat{Z}_t$ statistic.

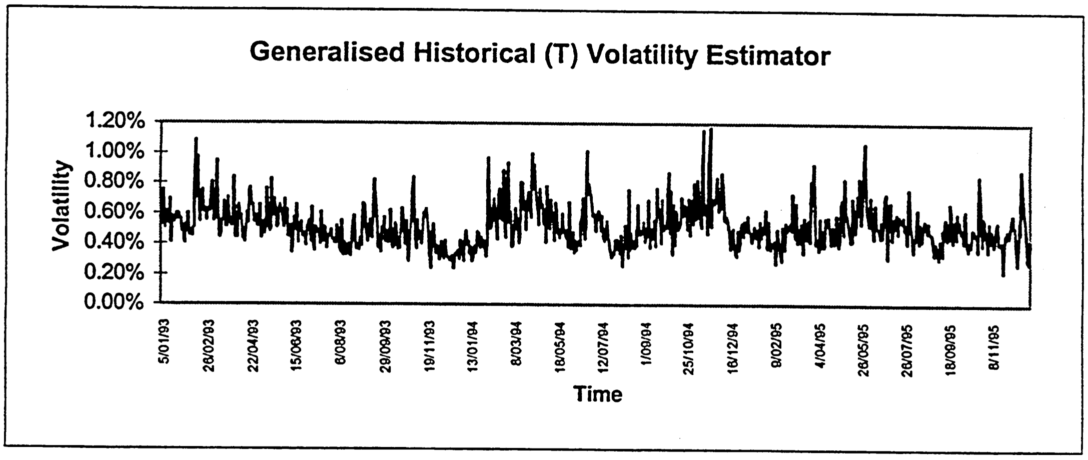
$\hat{Z}_t$	Hist (T)	MA(T)	MA (5)	Gen Hist (T)	Gen Hist (5)
Mean	0.1165	0.1364	0.1812	0.1544	0.1752
Std Dev	0.8807	0.9131	1.0543	1.0094	1.0564
Coef of Skewness	-0.0034	0.0674	0.2132	0.0912	0.1766
Coeff of Excess Kurtosis	0.2430	-0.1181	-0.2814	-0.0834	-0.1965
Jacque-Bera statistic	1.8196	0.9898	8.0341	1.2388	5.0312

Std Dev of Coef of Skewness = 0.08986
Std Dev of Coeff of Kurtosis = 0.1794
and $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n}\hat{Z}_i$, $m_k = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Z}_i - \overline{Z}\right)^k$, Coefficient of Skewness $g_1 = m_3 / m_2^{3/2}$,
Coefficient of Excess Kurtosis $g_2 = m_4 / m_2^2 - 3$, Jacque-Bera Statistic $= n\left(\frac{g_1}{6} + \frac{g_2^2}{24}\right) \sim \chi_2^2$.

Figure 1 Estimates of Daily Volatility for S&P 500 Index futures.




MA (5) Volatility Estimator
Volatility
1.60%
1.40%
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%
5/01/93
28/02/93
22/04/93
15/06/93
6/08/93
29/09/93
19/11/93
13/01/94
8/03/94
18/05/94
12/07/94
1/09/94
25/10/94
16/12/94
9/02/95
4/04/95
26/05/95
28/07/95
18/09/95
8/11/95
Time


Generalised Historical (T) Volatility Estimator
Volatility
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%
5/01/93
28/02/93
22/04/93
15/06/93
6/08/93
29/09/93
19/11/93
13/01/94
8/03/94
18/05/94
12/07/94
1/09/94
25/10/94
16/12/94
9/02/95
4/04/95
26/05/95
28/07/95
18/09/95
8/11/95
Time

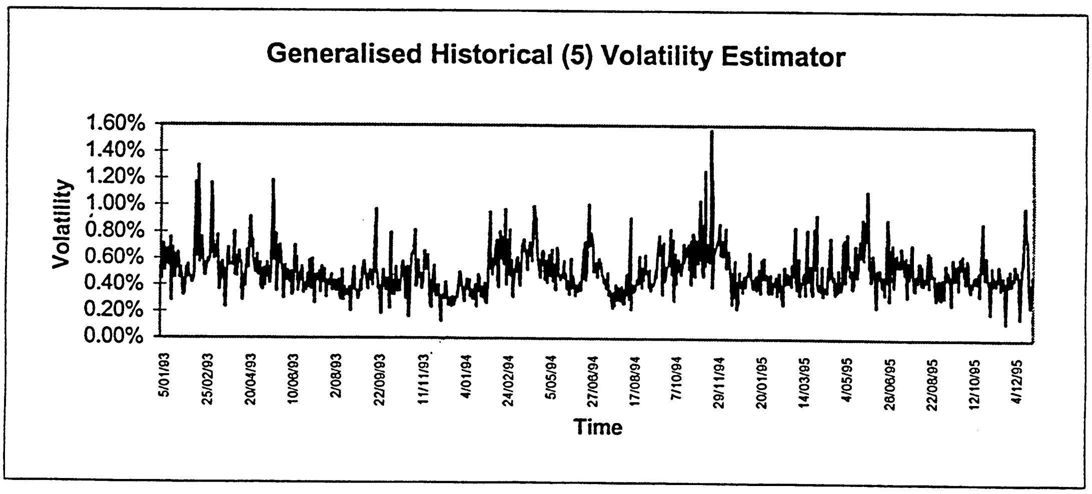
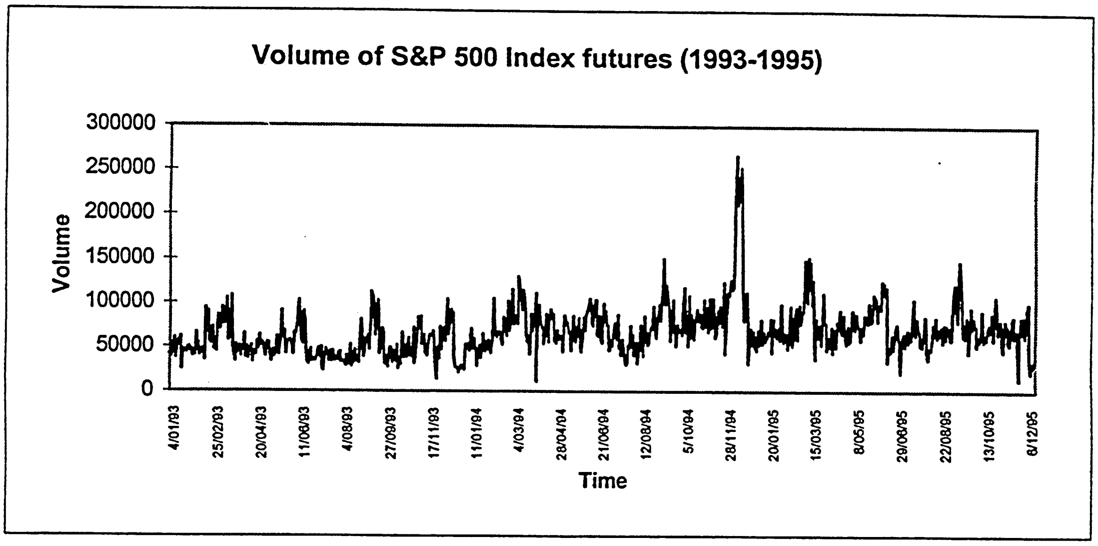
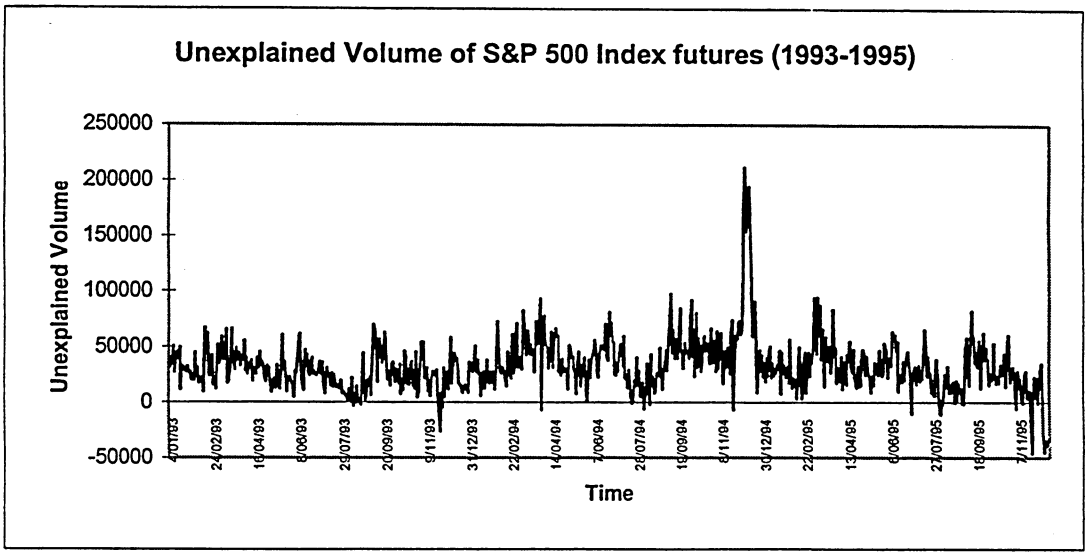

Generalised Historical (5) Volatility Estimator
Volatility
1.60%
1.40%
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%
5/01/93
25/02/93
20/04/93
10/06/93
2/08/93
22/09/93
11/11/93
4/01/94
24/02/94
5/05/94
27/06/94
17/08/94
7/10/94
29/11/94
20/01/95
14/03/95
4/05/95
28/06/95
22/08/95
12/10/95
4/12/95
Time

Figure 2


Volume of S&P 500 Index futures (1993-1995)
Volume
300000
250000
200000
150000
100000
50000
0
4/01/93
25/02/93
20/04/93
11/06/93
4/08/93
27/09/93
17/11/93
11/01/94
4/03/94
28/04/94
21/06/94
12/08/94
5/10/94
28/11/94
20/01/95
15/03/95
8/05/95
29/06/95
22/08/95
13/10/95
6/12/95
Time


Unexplained Volume of S&P 500 Index futures (1993-1995)
Unexplained Volume
250000
200000
150000
100000
50000
0
-50000
Time