MONASH WP 10/98

ISSN 1440-771X
ISBN 0 7326 1047 8

MONASH UNIVERSITY

ANCORA · IMPARO

AUSTRALIA

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 $r_t = Log_e\left(p_t / p_{t-1}\right)$, 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

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

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 $\Theta_{t-1}=0$,

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

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