Prices of financial instruments with stochastic volatility

Isaac Meilijson
School of Mathematical Sciences
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel-Aviv University, 69978 Tel-Aviv, Israel
E-mail: MEILIJSON@MATH.TAU.AC.IL

August 9, 2006

Abstract

This interim draft represents work in progress. It is submitted for discussion purposes at the Aggregation of Opinions conference sponsored by the Cowles Foundation, Yale University, September 15–17, 2006.

1 Introduction

Let $\cdots, S_0, S_1, \cdots, S_n, \cdots$ be the prices along time of some financial instrument and let

$$X_n = log\frac{S_n}{S_{n-1}} \tag{1}$$

be its logarithmic increments.

The $GARCH(1,1)$ estimator $\hat{\sigma}_n^2$ of the variance σ_n^2 toward the n-th entry of the time series X is defined as

$$\hat{\sigma}_n^2 = (1-\alpha-\beta)V_L + \alpha X_{n-1}^2 + \beta\hat{\sigma}_{n-1}^2 \tag{2}$$

where V_L is a typical historic variance around which the stochastic variance σ^2 should be centered. This $GARCH(1,1)$ variance estimator can also be expressed as the weighted average

$$\hat{\sigma}_n^2 = (1-\frac{\alpha}{1-\beta})V_L + \frac{\alpha}{1-\beta}s_n^2 \tag{3}$$

of the historic variance V_L and the $EWMA$ variance estimator s_n^2, the (Exponentially Weighted Moving Average) empirical smoothed variance estimator

$$\begin{aligned} s_n^2 &= (1-\beta)x_{n-1}^2 + \beta s_{n-1}^2 \\ &= (1-\beta)[x_{n-1}^2 + \beta x_{n-2}^2 + \beta^2 x_{n-3}^2 + \beta^3 x_{n-4}^2 + \cdots] \end{aligned} \tag{4}$$

We propose to re-derive $\hat{\sigma}_n^2$ in terms of Bayesian expert opinion pooling, giving a meaning to the parameter α, as V_L and β look more self justified (think of β's order of magnitude as its J. P. Morgan value 0.94). As will be seen, the $GARCH(1,1)$ variance estimator $\hat{\sigma}_n^2$ is reasonable in a heavy-tail setup that may be appropriate for Value at Risk (VaR) assessments but inappropriate for option pricing, as it makes S have infinite expected value. However, $\hat{\sigma}_n^2$ will arise as a sufficient statistic for a broader class of models, including light-tailed distributions. This means that - in this class - there is no need to search for stochastic variance estimators more general than those obtained as functions of $\hat{\sigma}_n^2$. Special attention will be given to the transformation

$$\breve{\sigma}_n^2 = V_U[\sqrt{1 + 2\frac{\hat{\sigma}_n^2}{V_U}} - 1] \tag{5}$$

where V_U is positive. This modified variance estimator $\breve{\sigma}_n^2$, essentially the same as $\hat{\sigma}_n^2$ if V_U is large, ranges from $\hat{\sigma}_n^2$ for small $\hat{\sigma}_n^2$ to twice the geometric average of $\hat{\sigma}_n$ and V_U for large $\hat{\sigma}_n^2$. The parameter V_U should be thought of as an "upper" variance that shouldn't be allowed to be exceeded too often.

The most common Gaussian conjugate Bayesian model assumes that

$$X_n = \frac{Z_n}{\sqrt{Y}} \tag{6}$$

where the Z_n are independent and identically normally distributed random variables, independent of the variance $\frac{1}{Y}$, where the *precision* Y has a prior Γ distribution. The Γ prior on Y yields (heavy-tailed) t-distributed X. The generalized inverse Gaussian prior GIG (see e.g. Jørgensen [5], Barndorff–Nielsen [2]) yields the lighter tails announced above. The GIG prior admits a further expansion of the model, under which $\frac{1}{Y}$ is not only the variance of X but also a linear component of its mean. This additional freedom, that introduces dependence between drift and volatility as in Heston [4], is further called for under portfolio design: the CAPM formulas derived via Itô calculus show that continuous trading portfolios produce financial indices whose drifts are linear combinations of the drifts, variances and covariances of the individual instruments.

The precision Y, or rather Y_n, will be allowed to be a stochastic process, modelling stochastic volatility. Dynamics will be dictated by a model of fusing the priors of two experts, one of which claims fixed variance and the other ever-changing variance. This is a technically simple version of sorts of the general Hidden Markov stochastic volatility modelling idea as in Barndorff–Nielsen [2].

If multivariate X is represented as in (6), with Z_n multivariate normally distributed and Y a scalar random variable as above, then X has the so called multivariate t distribution.

This description has been observed to be adequate for financial instrument modelling. Following Barndorff–Nielsen [2], a *GIG* prior on Y yields a tractable light-tail multivariate model. The role of *GARCH* will be played by its natural multivariate extension via exponentially smoothed empirical covariance matrices.

2 A Bayesian model based on a family of conjugate priors

Suppose $X = \frac{1}{\sqrt{Y}} Z + \text{drift}$, where Z is a standard normal variable independent of its "local variance" $\frac{1}{Y}$.

The drift $\mu + \frac{\delta}{Y}$ is allowed to depend linearly on variance. This is natural in lognormal variables, and may be as modelling correlation between price and volatility, for which there is ample evidence (see Heston [4], Roth [6]).

Given a prior density $g_Y(\cdot)$ on Y, the joint density of (X, Y) at (x, y) is

$$\begin{aligned} g_{Y|X=x}(y) f_X(x) &= g_Y(y) f_{X|Y=y}(x) = g_Y(y) \frac{1}{\sqrt{2\pi}} y^{\frac{1}{2}} \exp\{-\frac{(x-\mu-\frac{\delta}{y})^2}{2} y\} \\ &= g_Y(y) \frac{1}{\sqrt{2\pi}} \exp\{\delta(x-\mu)\} y^{\frac{1}{2}} \exp\{-\frac{(x-\mu)^2}{2} y - \frac{\delta^2}{2y}\} \end{aligned} \quad (7)$$

This suggests the application of conjugate Generalized Inverse Gaussian priors (*GIG*, see Jørgensen [5] and Barndorff–Nielsen [2]) of the form

$$g_Y(y) = g_{GIG}(y\ ;\ \gamma, \theta, \eta) = c(\gamma, \theta, \eta) y^{\gamma-1} \exp\{-\frac{1}{2}(\theta y + \frac{\eta}{y})\} \quad (8)$$

to get the *GIG* posterior density

$$g_{Y|X=x}(y) = g_{GIG}(y\ ;\ \gamma + \frac{1}{2}, \theta + (x-\mu)^2, \eta + \delta^2) \quad (9)$$

and Generalized Hyperbolic (*GH*) predictive density of X evaluated at x (substitute (8) and (9) into (7))

$$f_X(x) = f_{GH}(x\ ;\ \gamma, \theta, \eta) = \frac{c(\gamma, \theta, \eta)}{c(\gamma + \frac{1}{2}, \theta + (x-\mu)^2, \eta + \delta^2)} \frac{1}{\sqrt{2\pi}} \exp\{\delta(x-\mu)\} \quad (10)$$

This marginal density can be a main tool for pricing derivatives on X or evaluating VaR assessments, once (i) its scope is generalized from the time length between measurements to time intervals of arbitrary lengths, and (ii) a proper dynamic update of the prior is developed. This is the subject of the next sections. In principle, (i) should offer no problem in the absence of (ii): Barndorff–Nielsen [2] shows that *GH* is closed under convolutions and even identify a continuous time Lévy process representation for it.

The common Γ density is the special case of GIG corresponding to $\eta = \delta = 0$. The normalizing function c takes an especially simple form and $f_X(\cdot)$ is the Student t distribution. This distribution is heavy tailed and induces on prices infinite expected value. From this point of view, it is advisable to think in terms of $\eta > 0$ even when modelling $\delta = 0$, i.e., independent logarithmic price and volatility increments.

In principle, the density in (8) (hence, also in (9)) could (but will not) be further multiplied by some arbitrary function $h(y)$. The normalizing function c in (8) has a simple closed form representation in terms of one of the standard Bessel functions, so the literature opts for this parsimonious choice $h \equiv 1$.

$$\begin{aligned}\int_0^\infty y^{\gamma-1}\exp\{-\frac{\theta y}{2}\}\exp\{-\frac{\eta}{2y}\}dy &= \int_0^\infty y^{\gamma-1}\exp\{-\frac{\sqrt{\theta\eta}}{2}(\sqrt{\frac{\theta}{\eta}}y+\sqrt{\frac{\eta}{\theta}}\frac{1}{y})\}dy \\ &= [\frac{\eta}{\theta}]^{\gamma/2}\int_0^\infty z^{\gamma-1}\exp\{-\frac{\sqrt{\theta\eta}}{2}(z+\frac{1}{z})\}dz \\ &= 2[\frac{\eta}{\theta}]^{\gamma/2}\mathrm{BesselK}[\gamma,\sqrt{\theta\eta}] \qquad (11)\end{aligned}$$

Compare (8) and (11) to obtain the GIG normalizing function

$$c(\gamma,\theta,\eta) = c_{GIG}(\gamma,\theta,\eta) = \frac{[\frac{\theta}{\eta}]^{\gamma/2}}{2\mathrm{BesselK}[\gamma,\sqrt{\theta\eta}]} \qquad (12)$$

or its Γ version

$$c_\Gamma(\gamma,\theta) = \frac{\theta^\gamma}{\Gamma(\gamma)} \qquad (13)$$

3 Pooling expert opinion

Suppose there are two (or more) experts on an unknown parameter Y, assigning to it respective priors g_1 and g_2. You know nothing about Y but pride yourself in being an expert on experts. How do you pool their priors into yours? Genest & Zidek [3] attribute to Peter Hammond the *logarithmic opinion pool*, characterized by the *commuting-with-experiment* axiom: if you pool the priors and compute a posterior given new information, you should end up with the same result as if you used the same rule to pool the posteriors evaluated by the experts from the same new information.

The only solutions to this axiom are: fix positive weights β and $1-\beta$ and let your compromise prior be

$$g(\cdot) = K(\beta)g_1(\cdot)^\beta g_2(\cdot)^{1-\beta} \qquad (14)$$

Since $g_1(\cdot)^\beta g_2(\cdot)^{1-\beta}$ is a sub-density, $K(\beta) > 1$. Or, if the two experts agree on the probability of some value of Y, you should assign it a higher probability! If a *unanimity* axiom is added postulating that whenever the experts concur on the probability of some value, you should adopt their assessment, then the only solutions are the dictatorships (well, the coordinate mappings...), in which β can only be 0 or 1. Logarithmic expert opinion pooling shares this remarkable feature with Arrow's [1] impossibility theorem.

For details, a critique and comments, consult [3].

3.1 Stochastic volatility via expert opinion pooling

You have two experts, a *Bayesian* expert that tells you to do as if the unknown variance is constant and you are progressively learning about it, and an *Empirical Bayesian* expert that tells you that variance is re-sampled independently from the prior at every time point. If the current prior density is

$$c_1(\gamma,\theta,\eta)m(y)y^{\gamma-1}\exp\{-\frac{\theta y}{2}\}\exp\{-\frac{\eta}{2y}\} \tag{15}$$

then the posterior density after observing $X = x$ is up to a multiplicative factor

$$m(y)y^{\gamma-1}\exp\{-\frac{\theta y}{2}-\frac{\eta}{2y}\}y^{\frac{1}{2}}\exp\{-\frac{(x-\mu)^2}{2}y-\frac{\delta^2}{2y}\} \tag{16}$$

If the standard empirical Bayesian prior density is

$$c_2(\gamma_0,\theta_0,\eta_0)h(y)y^{\gamma_0-1}\exp\{-\frac{\theta_0 y}{2}\}\exp\{-\frac{\eta_0}{2y}\} \tag{17}$$

then the pooled posterior density after observing $X = x$ is up to a multiplicative factor

$$\begin{aligned} & [m(y)y^{\gamma-1}\exp\{-\frac{\theta y}{2}-\frac{\eta}{2y}\}y^{\frac{1}{2}}\exp\{-\frac{(x-\mu)^2}{2}y-\frac{\delta^2}{2y}\}]^{\beta} \\ * \quad & [h(y)y^{\gamma_0-1}\exp\{-\frac{\theta_0 y}{2}-\frac{\eta_0}{2y}\}]^{1-\beta} \end{aligned} \tag{18}$$

that becomes $c_3(\gamma_N,\theta_N,\eta_N)m_N(y)y^{\gamma_N-1}\exp\{-\frac{\theta_N y}{2}-\frac{\eta_N}{2y}\}$ upon setting (N for "New")

$$\begin{aligned} m_N(\cdot)=m(\cdot)^{\beta}h(\cdot)^{1-\beta}\ ,\ \gamma_N &= \beta\gamma+(1-\beta)\gamma_0+\frac{\beta}{2} \\ \theta_N=\beta\theta+(1-\beta)\theta_0+(x-\mu)^2\beta\ ,\ \eta_N &= \beta\eta+(1-\beta)\eta_0+\delta^2\beta \end{aligned} \tag{19}$$

But along time the new m converges to h (as long as both are strictly positive) and the new γ and η converge to the respective fixed points

$$\gamma_1=\gamma_0+\frac{\beta}{2(1-\beta)}\ ,\ \eta_1=\eta_0+\frac{\delta^2\beta}{(1-\beta)} \tag{20}$$

of $\beta\gamma+(1-\beta)\gamma_0+\frac{\beta}{2}$ and $\beta\eta+(1-\beta)\eta_0+\delta^2\beta$. Hence, since the process has been going on for some long time, it can be assumed from the outset that $m(\cdot)\equiv h(\cdot)$ and γ,η

are statically fixed as γ_1, η_1 in (20). Thus, the prior density (15) and posterior density (18) belong to the same h-modified GIG family, sharing the same $h(\cdot)$ term (henceforth $h \equiv 1$, thus $c_1 = c_2 = c_3 = c_{GIG}$), inverse scale parameter η_1 and shape parameter γ_1. They differ only in the value assigned to the scale parameter θ, updated from its prior value θ_{n-1} to its posterior value θ_n after observing $X = x_{n-1}$ by (see θ_N in (19)) $\theta_n - \theta_0 = \beta(\theta_{n-1} - \theta_0) + (x_{n-1} - \mu)^2\beta$. In other words,

$$\theta_n = \theta_0 + \frac{s_n^2 \beta}{(1-\beta)} \tag{21}$$

expressed directly in terms of the smoothed variance estimator s_n^2 (see (4)) applied to the μ-centered logarithmic increments, to wit

$$\begin{aligned} s_n^2 &= (1-\beta)(x_{n-1} - \mu)^2 + \beta s_{n-1}^2 \\ &= (1-\beta)[(x_{n-1}-\mu)^2 + \beta(x_{n-2}-\mu)^2 + \beta^2(x_{n-3}-\mu)^2 + \cdots] \end{aligned} \tag{22}$$

Sufficiency of the smoothed variance estimator s_n^2 (see (22)) has thus been established.

Remark: Care must be exercised here for estimation purposes of $\theta_0, \gamma_0, \eta_0$ and δ once a likelihood model has been formulated: the parameter μ enters into the definition of the statistic s_n^2, so it should be interpreted as a known constant or smooth empirical centering.

Differentiation of the posterior density with respect to y will show that the MAP (maximal a-posteriori probability) variance estimator $\breve{\sigma}_n^2$ (see (5)) is

$$\breve{\sigma}_n^2 = \frac{\hat{1}}{Y} = V_U[\sqrt{1 + 2\frac{\hat{\sigma}_n^2}{V_U}} - 1] \ , \tag{23}$$

expressed in terms of the upper variance parameter V_U

$$V_U = \frac{\gamma_1 - 1}{2\eta_1} = \frac{2(\gamma_0 - 1)(1-\beta) + \beta}{2\eta_0(1-\beta) + \delta^2\beta} \tag{24}$$

and the $GARCH(1,1)$ variance estimator $\hat{\sigma}_n^2$

$$\hat{\sigma}_n^2 = \frac{\theta_n}{\gamma_1 - 1} = \frac{2(1-\beta)(\gamma_0 - 1)}{2(1-\beta)(\gamma_0-1) + \beta} \, \frac{\theta_0}{\gamma_0 - 1} + \frac{\beta}{2(1-\beta)(\gamma_0 - 1) + \beta} \, s_n^2 \tag{25}$$

As is common, the $GARCH(1,1)$ variance estimator has been represented as a weighted average of the exponentially smoothed empirical variance s_n^2 and a standard historical value. This value is of the same shape as $GARCH(1,1)$, a θ term divided by a $\gamma - 1$ term.

The MAP variance estimator $\breve{\sigma}_n^2$ is a slowly-growing version of $\hat{\sigma}_n^2$ and the two are practically equal if $\frac{\hat{\sigma}_n^2}{V_U}$ is small, that is the case if V_U is indeed an “upper” value of the unit-time variance.

The Bayes variance estimator is

$$E[\frac{1}{Y}] = \frac{c(\gamma_1,\theta_n,\eta_1)}{c(\gamma_1-1,\theta_n,\eta_1)} = \sqrt{\frac{\theta_n}{\eta}}\frac{\mathrm{BesselK}(\gamma_1-1,\sqrt{\theta_n\eta_1})}{\mathrm{BesselK}(\gamma_1,\sqrt{\theta_n\eta_1})} \tag{26}$$

It should be noticed that the MAP estimate of variance has been defined as the reciprocal of the mode of Y rather than as the mode of $\frac{1}{Y}$. This is so because, as will be seen in the next section, for the regular Γ case (corresponding to $\eta = 0$), the reciprocal of the mode of Y, the mean of $\frac{1}{Y}$ and the $GARCH(1,1)$ estimate of variance are identical. Since we see the GIG model as a light-tail perturbation of the Γ model, it is reasonable to expect our choice to be closer. We don't dare to try to analyze this issue except by numerical comparisons of the four estimates in question. Indeed, for positive values of the shape parameter γ, MAP and Bayes are very close to each other and reasonably close to $GARCH(1,1)$), but the mode of $\frac{1}{Y}$ is inadequately far off. Some further intuition for the effect of the sign of γ may be gained by the observation that if $Y \sim GIG(\gamma,\theta,\eta)$, then $\frac{1}{Y} \sim GIG(-\gamma,\eta,\theta)$.

The next section analyzes the Γ case where $\eta = \delta = 0$, for which solutions are easier to represent, all three variance estimators above (Bayes, MAP and $GARCH(1,1)$) coincide and X is distributed according to the t-distribution. This last property may be useful and practical for the evaluation of VaR, but is inappropriate as a theoretical model as well as for pricing options because it yields infinite expected prices. Further sections will consider general $\eta > 0$ and δ, to model correlated price and volatility as well as lighter tails.

4 The Γ case: $\eta = \delta = 0$

The inverse variance Y follows the Γ distribution with shape parameter $\gamma_1 = \gamma_0 + \frac{1}{2(1-\beta)}$ (see (20)) and scale parameter $\theta_n = \theta_0 + \frac{s_n^2}{2(1-\beta)}$ (see (21)),

$$\frac{\theta_n^{\gamma_1}}{\Gamma(\gamma_1)} y^{\gamma_1-1}\exp\{-\theta_n y\} \tag{27}$$

The Bayes variance estimator, well defined if $\gamma_0 \geq 1$ or if $\beta \geq \frac{1}{2}$ (i.e., if the Bayesian expert is given more weight than the Empirical Bayesian expert, such as in the J. P. Morgan choice $\beta = 0.94$), is the expectation of inverse Y when Y is distributed (27),

$$E[\frac{1}{Y}] = \frac{\Gamma(\gamma_1-1)}{\theta_n^{\gamma_1-1}}\frac{\theta_n^{\gamma_1}}{\Gamma(\gamma_1)} = \frac{\theta_n}{\gamma_1-1} = \frac{2(1-\beta)\theta_0 + s_n^2}{2(1-\beta)(\gamma_0-1)+1} \tag{28}$$

This is also the point at which the density in (27) is maximal. As announced in the last subsection, the $GARCH(1,1)$ variance estimator on the RHS of (28) is thus justified as the

Bayes and MAP estimator. The Bayesian model is thus seen to lend itself as a natural tool for the interpretation of the $GARCH(1,1)$ weights.

4.1 VaR for one time unit and approximate VaR for longer intervals

> These approximations, for Γ as well as for GIG, are being studied by Eliahu Arviv as his M.Sc. thesis research at Tel Aviv University, under my guidance.

If the current (time n) distribution of Y is given by (27), how is the new X_n distributed? How is S_{n+t} distributed?

The conditional distribution of X_n given that $Y = y$ is Gaussian with mean zero and variance $\frac{1}{y}$. The unconditional density of X_n at a point x is obtained by integrating over y the joint density

$$\frac{1}{\sqrt{2\pi}} \frac{\theta_n^{\gamma_1}}{\Gamma(\gamma_1)} y^{\gamma_1 - \frac{1}{2}} \exp\{-(\theta_n + \frac{x^2}{2})y\} \tag{29}$$

Since (29) is a Γ density up to a multiplicative factor, the unconditional density of X_n at x is

$$\frac{1}{\sqrt{\theta_n}} \frac{\Gamma(\gamma_1 + .5)}{\sqrt{2\pi}\Gamma(\gamma_1)} [1 + \frac{x^2}{2\theta_n}]^{-(\gamma_1 + .5)} \tag{30}$$

This shows that $\sqrt{\frac{\gamma_1}{\theta_n}} X$ has a (Student) t distribution with $2\gamma_1$ degrees of freedom. The larger γ_1 is, the closer does the Gaussian model fit this distribution. The Empirical Bayes value γ_0 is relatively small, and the size of the steady-state Bayes value $\gamma_1 = \gamma_0 + \frac{1}{2(1-\beta)}$ depends on the decay factor β, that measures for how long can volatility be considered unchanged, i. e., how "seriously" should the Bayesian expert be taken.

This point will be further stressed upon the analysis of the marginal distribution of the logarithmic increment X over the time interval $[n-1+t, n-1+t+m]$ for $t \geq 0$ and $m > 0$. So far only the case $t = 0$ and $m = 1$ has been modelled, corresponding to the next observation X_n.

We could but don't extend this analysis to the case where the observations themselves are lacunar or unequally spaced, in which case the shape parameter will change from observation to observation, without necessarily reaching a steady value such as in (20).

If β is the factor by which confidence in the Bayesian expert decays per unit time, then it should be replaced by β^t over an interval of length t. Hence the prior density of the inverse unit-time variance t units of time past the current time $n-1$ should be Γ with shape parameter $\gamma_1\beta^{t-1} + \gamma_0(1-\beta^{t-1}) = \gamma_0 + \frac{\beta^{t-1}}{2(1-\beta)}$ and scale parameter $\theta_n\beta^{t-1} + \theta_0(1-\beta^{t-1}) =$

$\theta_0 + \frac{\beta^{t-1} s_{n-1}^2}{2(1-\beta)}$, each of the two converging to the corresponding Empirical Bayes parameter as $t \to \infty$.

We now face (*i*) a time interval of length m and (*ii*) a serious problem. Logarithmic expert pooling provides each individual expert with veto power, in the sense that the pooler must assign zero probability to any event assigned zero probability by any of the individual experts. Hence, this type of expert pooling is mostly addressed to applications in which the experts' priors are absolutely continuous with respect to each other. From this point of view, as long as the Bayes and Empirical Bayes densities are strictly positive, one-time pooling is adequate. However, pooling is dynamically inconsistent because on two consecutive time units the Empirical Bayes prior for the corresponding *two* variances is supported by the entire plane whereas the Bayes prior is supported by the diagonal, since it postulates that variance never changes.

Expert opinion pooling will be restricted to one-time assessments, but keeping more generally the paradigm that inverse variance is Γ or GIG distributed. The parameters of such a Γ or GIG distribution will be derived from those pertaining to single time units, by the requirement that logarithmic increments X have the correct first few moments.

References

[1] Arrow, K. (1950). A Difficulty in the Concept of Social Welfare. *The Journal of Political Economy*, **58:4**, 328–346.

[2] Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand J. Statist.* **24**, 1–13.

[3] Genest, Ch. and Zidek, J. V. (1986). Combining probability distributions: a critique and an annotated bibliography. *Statistical Science*, **1**, 114–148.

[4] Heston, S. (1993). A closed form solution for options pricing with Stochastic Volatility, with applications to bond and currency options. *The Review of Financial Studies*, 327–343.

[5] Jørgensen, S. (1982). *Statistical properties of the generalized inverse Gaussian distribution.* New York: Springer–Verlag.

[6] Roth, R. (2004). M.A. thesis at Tel Aviv University.