

A Note on the Informational Content of Option Prices

The concept of market efficiency has yield researchers to test whether options were accurately priced and provided, therefore, information about the future behavior of the volatility of the underlying. A traditional hypothesis for this kind of test is the continuous time model of BLACK and SCHOLES (1973) which has the particularity of displaying a linearity in the volatility for at-the-money options. Therefore it is possible to extract the expected volatility over the remaining lifetime of the option from at the money market prices of options. An often tested hypothesis (PAGAN and SCHWERT (1990), DAY and LEWIS (1992), LASTRAPES and LAMOUREUX (1995)) is the ability of the implied volatility to forecast the future standard deviation of return, and to contrast it with other forecast given by historical standard deviation or time series model. This procedure is a joint test of model specification and market efficiency. If the model chosen is the correct one then in an efficient market all past information should be included in the option price and there is no reason, in that case, for a time series modelisation to dominate in terms of forecasting ability. However, these studies have found that the time series

* Current address: Boston University SMG Finance and Economics 595 Commonwealth Avenue 02467 Boston MA USA email: berrada@bu.edu. I thank the referee, Thomas Stucki, for insightful comments and my colleagues Christophe Perignon and Dusan Isakov for helpful discussions. All remaining errors are my own.

specification yields better results in particular for long term forecast (20 days), suggesting then that either the market was wrong or the pricing model was inappropriate. In this note we will show that this apparent dominance may be caused by an estimation error which occurs in the procedure.

In the context of the BLACK and SCHOLES model it is important to distinguish between the standard deviation of daily return and the volatility of the geometric brownian motion with drift defined by

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \quad (1)$$

$$S_0 = s_0$$

The volatility implied by the option price is σ in equation (1) and we will show in a simulation that this measure differs from the standard deviation of daily return. In the simulation that we will perform we will assume that equation (1) is indeed the governing process for the stock price. We will further assume that the implied volatility obtained from the option prices is

$$\sigma_{iv} = \sigma + \varepsilon \quad (2)$$

where ε is a white noise. Therefore the implied volatility is an unbiased measure of the true volatility. Note that we add here a noise term to take

account of the fact that implied volatility is usually subject to measurement error due to transaction costs, non-synchronicity or other frictions. We compare its forecasting power to the historical standard deviation of daily return defined as

$$\sigma_{\text{hist}} = \frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^2 \quad (3)$$

where r_i is the return on day i defined as

$$r_i = \ln \left[\frac{S_i}{S_{i-1}} \right] \quad (4)$$

\bar{r} is the mean return, and N is the number of days used for the computation. We will use the first 20 days to compute σ_{hist} , on day 20 the option price is observed, and the observed volatility is then computed as in equation (3) but for the next T days.

$$\sigma_{\text{obs}} = \frac{1}{T} \sum_{i=1}^T (r_i - \bar{r})^2 \quad (5)$$

In the simulation we use a value of 0.00075 for μ (18.75% yearly), 0.02 for σ (31.6% yearly) and 20 for T . Figure A1 and A2 in the appendix display the empirical distribution of σ_{hist} and σ_{obs} obtained from 300 simulations. We clearly see that the two distributions are very close and have the same mean, which is superior to the true volatility σ . We perform the following regressions, using OLS, to test the forecasting power of the historical standard deviation and implied volatility

$$\sigma_{\text{obs}} = \beta_1 \sigma_{\text{hist}} + \eta \quad (6)$$

$$\sigma_{\text{obs}} = \gamma_1 \sigma_{\text{iv}} + \nu \quad (7)$$

with σ and ν white noise. The better estimator (historical or implied) will be the one with a coefficient closer to one. We obtain the following results:

Table 1: Regression Results

	Value	T-stat	R ²
β_1	0.99	70.3	0.69
γ_1	1.36	69.4	0.73

The t-statistics imply that the parameters β_1 and γ_1 are highly significant, and β_1 is almost equal to one.

Using the standard deviation as a measure of observed volatility yields a misleading result. The forecasting ability of the historical volatility, computed in the same way, seems excellent. The implied volatility extracted from option prices, which we know is the correct one in our setting seems to underestimate the future volatility. This means that according to the model, option prices are too low, since the price is an increasing function of the volatility.

It is therefore important to define a correct measure of observed volatility when testing the informational content of option prices. One could use an indirect inference estimator, as in GOURRIEROUX and al. (1993) of the standard deviation as an alternative choice. We briefly recall the methodology, which consists of the following steps:

1. Estimation based on a given criterion (intermediary model) from the original data
2. Simulation of the trajectory using the initial continuous time model and time interval equal to the original data
3. Estimation based on the same criterion from the simulated data
4. Calibration of the model used for simulation to minimize the distance between the estimation in step 1 and 3

Let us point the key assumptions about the original model. The process must be stationary and independent of the innovations, the distribution of which is known (this is verified in our setting for

the return process dS_t/S_t). Under this hypothesis it is possible to simulate values of S_t which will be denoted $\tilde{Z}(\theta, z_0)$ conditional on some initial value $z_0 = S_0$ (and given values for the parameters $\theta = (\mu, \sigma)$). In order to perform these simulations only the innovations have to be drawn from their known distribution, in our case the wiener increments are normally distributed with mean 0 and variance equal to t . We can now define the intermediary model, the auxiliary parameter and its estimator. The intermediary model is a function of the simulated data $\tilde{Z}(\theta, z_0)$. The auxiliary parameter $\beta = (m, s)$ is estimated by maximizing a chosen criterion

$$\max_{\beta} Q_T(S_t; \beta) \tag{8}$$

In our case the criterion $Q_T(S_t; \beta)$ will be the likelihood of the intermediary model. In our estimation the intermediary model is given by a simple EULER discretization scheme

$$Z_{t+1} = Z_t + mZ_t + sZ_t \varepsilon_t \tag{9}$$

where ε_t is a white noise. The auxiliary parameter vector β has in fact the same dimension as the original parameter vector θ , but this is not required by the estimation technique.

The auxiliary parameter is estimated on the original data and the estimator $\hat{\beta}_T$ is obtained. Note that the value obtained here for the volatility s , is different from σ , and is actually equal to the standard deviation of the series. Using the simulation described above H paths are generated and for each of these paths the auxiliary parameter $\hat{\beta}_h(\theta, z_0^h)$ is estimated. The parameter vector θ is indirectly obtained by calibration in order to have $\frac{1}{H} \sum_{h=1}^H \hat{\beta}_h(\theta, z_0^h)$ close to $\hat{\beta}_T$, more formally the following minimization is performed

$$\min_{\theta} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \hat{\beta}_h(\theta, z_0^h) \right]' \Omega_T \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \hat{\beta}_h(\theta, z_0^h) \right] \tag{10}$$

Where Ω is a positive definite matrix which converges to the identity matrix (in our case we use the identity matrix directly).

Using this estimator for the observed standard deviation we perform the following regression

$$\sigma_{ind} = \beta_2 \sigma_{hist} + \eta \tag{11}$$

$$\sigma_{ind} = \gamma_2 \sigma_{iv} + v \tag{12}$$

where σ_{ind} is the indirect inference estimator of the standard deviation, the results of the regression are displayed in Table 2. The empirical distribution of the standard deviation obtained from the indirect inference estimation is given in Figure A3. We see that when the standard deviation is correctly estimated, the implied volatility, which is, as we have assumed, the best forecast, dominates the historical standard deviation in term of explanatory power. We see that the coefficient γ_2 is very close to one.

In this note we have used an extremely simple setting where the coefficients were constant, however one can think that a more complex setting will produce an even bigger bias in the volatility measure, for example in a context of stochastic volatility.

Table 2: Regression Results

	Value	T-stat	R ²
β_2	0.69	70.1	0.67
γ_2	0.95	68.8	0.72

References

- BLACK, F. and SCHOLLES, M. (1973): "The pricing of options and corporate liabilities", *Journal of Political Economy*, 81, pp. 637-54.
- DAY, T.E. and C. M. LEWIS (1992): "Stock Market Volatility and the Information Content of Stock Index Options", *Journal of Econometrics*, 52, pp. 267-287.
- GOURIEROUX, C., A. MONFORT and E. RENAULT (1993): "Indirect Inference", *Journal of Applied Econometrics*, 8, pp. 58-118.
- LAMOUREUX, C.G. and W. D. LASTRAPES (1993): "Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities", *Review of Financial Studies*, 6, pp. 293-326.
- PAGAN, A. R. and G. W. SCHWERT (1990): "Alternative Models for Conditional Stock Volatility", *Journal of Econometrics*, 45, pp. 267-290.

Appendix

Figure A1: Empirical Distribution of the Historical Standard Deviation

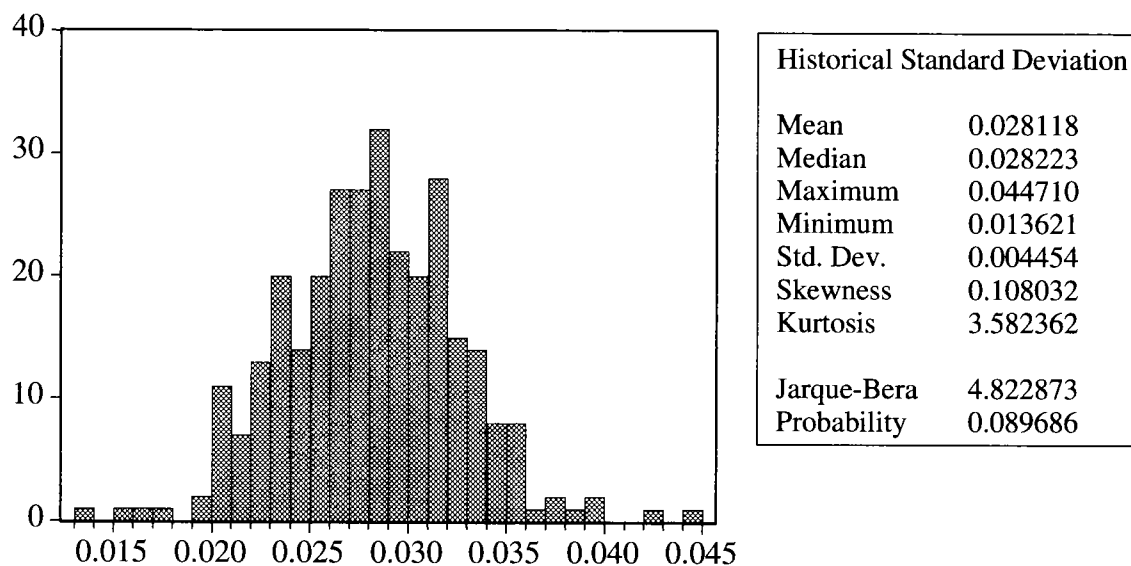


Figure A2: Empirical Distribution of the Observed Standard Deviation

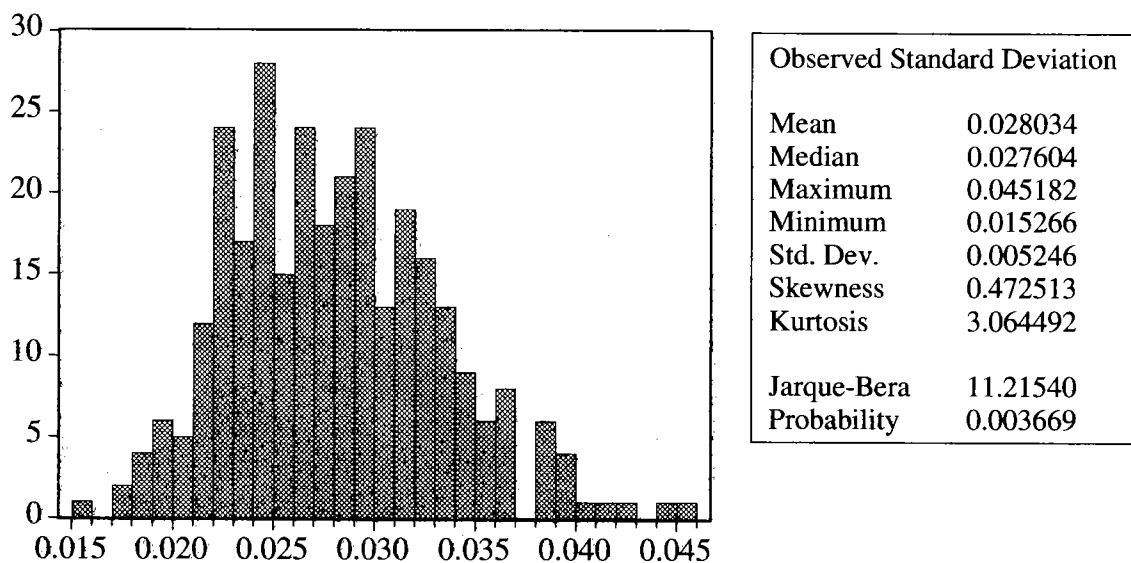


Figure A3: Empirical Distribution of the Standard Deviation estimated using Indirect Inference