# A bound on the standard error of the price of risk

John Dodson[1]

The Options Clearing Corporation

University of Minnesota Center for Financial and Actuarial Mathematics

October 18, 2013

[1]The views expressed here are the author's alone.

# Problem statement



Put yourself in the shoes of a statistical analyst considering the marginal potential of an investment asset.

- Assume that its total return process is continuous.
- Further assume that there is a risk-neutral measure.

You are interested in how much this asset can be expected to outperform a risk-free asset. Therefore, you are interested in fitting a Radon-Nikodým derivative,

$$\left.\frac{d\mathbb{P}}{d\mathbb{Q}}\right|\mathcal{F}_0 \triangleq e^{Z_t - \frac{1}{2}\langle Z\rangle_t} \quad \forall t > 0 \qquad (\dagger)$$

where

$$Z_t \triangleq \int_0^t \lambda_s \, dW_s$$

with $\mathcal{F}_t$-measurable values $\lambda_t$ the "price of risk" and $W_t$ the $\mathbb{Q}$-Wiener process driving the asset and $t = 0$ is the present.

# Problem statement



In these terms, this might seem like a difficult econometric challenge; but professional investment analysts find ostensibly satisfactory solutions every day in the course of forming their recommendations.

## Main result

I consider the class of unbiased estimators for a constant value of the price of risk using data sampled from the total return of the asset and the risk-free investment and prove there is a lower bound on the standard error.

$$\text{bias}\,\hat{\lambda} = 0 \implies \text{SE}\,\hat{\lambda} \geq \frac{1}{\sqrt{T}} \qquad (*)$$

where $T$ is the duration of the historical period used in the estimate.

# Problem statement



## Experiment design

Andrew Lo, in his study on the statistics of Sharpe ratios, says that values of the order $\lambda \approx 1.0\ \text{yr}^{-1/2}$ are typical. Analysts seem to find it useful to discriminate between $\lambda = +0.5\ \text{yr}^{-1/2}$ and zero. This requires a standard error of about half this difference. But

$$\text{bias}\,\hat{\lambda} = 0 \ \ \& \ \ \text{SE}\,\hat{\lambda} \leq 0.25\ \text{yr}^{-1/2} \implies T > 16\ \text{yr}$$

I am doubtful that investment analysts typically include data from sixteen years ago (expect possibly for indexes) to make their forecasts.

*I take this to mean that the use of biased estimators must be very common!* ☺

# Static volatility

The result is (I hope) relatively well-known in the setting of static volatility. My present contribution is to extend the result to a class of parametric models for dynamic volatility. But let us review the static volatility case first.

## Geometric Brownian motion

Consider a sample of $N+1$ joint observations over $T$ years of risky $S_t$ and risk-free $B_t$ with stochastic processes

$$\begin{aligned} dS_t &= (r_t + \lambda\sigma)\, S_t\, dt + \sigma S_t\, d\tilde{W}_t \\ dB_t &= r_t B_t\, dt \end{aligned}$$

where $\tilde{W}_t \triangleq W_t - \lambda t$ is a $\mathbb{P}$-martingale. Quantities

$$\log \frac{S_{t_i}}{B_{t_i}} - \log \frac{S_{t_{i-1}}}{B_{t_{i-1}}}$$

are independent Gaussian random variables.

# Static volatility



It is straight-forward to form the joint likelihood function for the parameters $\sigma^2$ and $\lambda$, and it is merely tedious to evaluate the inverse Fisher information,

$$\mathcal{I}^{-1}\left(\sigma^2,\lambda\right)=\begin{pmatrix}\frac{2\sigma^4}{N} & \frac{\sigma-\lambda}{2\sigma^2}\frac{2\sigma^4}{N}\\ \frac{\sigma-\lambda}{2\sigma^2}\frac{2\sigma^4}{N} & \frac{1}{T}+\left(\frac{\sigma-\lambda}{2\sigma^2}\right)^2\frac{2\sigma^4}{N}\end{pmatrix}$$

hence, by Cramér-Rao, we have

$$\operatorname{var}\hat{\lambda}\geq\frac{1}{T}+\left(\frac{\sigma-\lambda}{2\sigma^2}\right)^2\frac{2\sigma^4}{N}\geq\frac{1}{T}$$

for any unbiased estimator $\hat{\lambda}$ of $\lambda$.

Of course volatility is not static. Robert Merton wrote about this result in 1980 and took it as a sign to turn towards dynamic volatility models.

# Dynamic volatility



To introduce dynamic volatility, start by defining our data generating process as $X_t \triangleq \log \frac{S_t}{B_t}$. In general, we have

$$X_t = X_{t_0} + \int_{t_0}^{t} \sigma_s \, dW_s - \int_{t_0}^{t} \tfrac{1}{2}\sigma_s^2 \, ds \quad \forall t \geq t_0$$

Discretize time according to $t_0 < t_1 < \cdots < t_N \leq 0$ and let

$$\int_{t_0}^{t_n} \sigma_s \, dW_s = \sum_{i=1}^{n} \sqrt{\frac{h_i}{t_i - t_{i-1}}} \left(W_{t_i} - W_{t_{i-1}}\right)$$

where each $h_i$ is $\mathcal{F}_{t_{i-1}}$-measurable. Therefore

$$X_{t_i} | \mathcal{F}_{t_{i-1}} \sim \mathcal{N}\left(X_{t_{i-1}} + m_i, h_i\right)$$

under $\mathbb{P}$ where $m_i = \lambda\sqrt{h_i\left(t_i - t_{i-1}\right)} - \frac{1}{2}h_i \approx 0$.

# Dynamic volatility



Consider some finite-dimensional specification for these conditional variances, for example Engle's GARCH

$$h_i = \omega + \alpha \epsilon_{i-1}^2 + \beta h_{i-1}$$

where $\epsilon_i \triangleq X_i - \mathrm{E}^{\mathbb{P}}\, X_i | \mathcal{F}_{t_{i-1}}$

- Denote these parameters collectively by the vector $\theta$.

## Likelihood function

We can define the likelihood function for a timeseries sample $x = (x_{t_0}, x_{t_1}, \ldots, x_{t_N})^\top$ as

$$f_X^{\mathbb{P}}(x; \theta \mathbin{+\!\!\!+} \lambda) = f_{X_{t_1}|\mathcal{F}_{t_0}}^{\mathbb{P}}(x_{t_1}) \cdots f_{X_{t_N}|\mathcal{F}_{t_{N-1}}}^{\mathbb{P}}(x_{t_N})$$

where the price of risk is now appended to the parameter vector.

# Dynamic volatility



The key observation is that we can apply (†) to separate the price of risk from the volatility parameters.

$$\log f_X^{\mathbb{P}}(X;\theta \mathbin{+\!\!\!+} \lambda) = \log f_X^{\mathbb{Q}}(X;\theta) + \lambda\left(W_{t_N} - W_{t_0}\right) - \tfrac{1}{2}\lambda^2\left(t_N - t_0\right)$$

### Fisher information

The Fisher information, $\mathcal{I} \triangleq \operatorname{cov}^{\mathbb{P}} \nabla \log f_X^{\mathbb{P}}$, can be written as

$$\mathcal{I}(\theta \mathbin{+\!\!\!+} \lambda) = \begin{pmatrix} \mathcal{I}(\theta) - \lambda\, \mathrm{E}^{\mathbb{P}} \frac{\partial^2 W}{\partial\theta^\top \partial\theta} & -\mathrm{E}^{\mathbb{P}} \frac{\partial W}{\partial\theta^\top} \\ -\mathrm{E}^{\mathbb{P}} \frac{\partial W}{\partial\theta} & T \end{pmatrix}$$

where $T \triangleq t_N - t_0$ is the duration of the historical sample period and $W \triangleq W_{t_N} - W_{t_0}$ is the cumulative increment of the latent driving process for the risky asset.

### Schur complement

To conclude, we need a simple result from linear algebra. It is a straight-forward exercise to prove that if matrix $M > 0$ (i.e. positive-definite) has the form

$$M = \begin{pmatrix} A & a \\ a^\top & \alpha \end{pmatrix} \quad \& \quad M^{-1} = \begin{pmatrix} B & b \\ b^\top & \beta \end{pmatrix}$$

for scalars $\alpha$ and $\beta$, then $\beta \geq 1/\alpha$. (N.B.: $\beta^{-1}$ is the Schur complement of $A$ in $M$.). Hence,

$$\left[\mathcal{I}^{-1}(\theta \mathbin{+\!\!\!+} \lambda)\right]_{\lambda,\lambda} \geq \frac{1}{T}$$

which leads to the main result $(*)$.

# Extensions



The Cramér-Rao lower bound is a special case of the Kullback inequality about the relative entropy of one measure with respect to another

$$D_{\mathrm{KL}}\left(\mathbb{P} \parallel \mathbb{Q}\right) \geq \Psi^{\star}_{\mathbb{Q}}\left(\mu'_1\left(\mathbb{P}\right)\right)$$

which may be useful in extending the result of this paper to a wider class of processes, such as those with non-zero Lévy measure.

The challenge this seems to present is that there may no longer be a plausible low-dimensional parameterization of the Radon-Nikodým derivative(s) that describe risk-neutrality in this setting.