Risk & Asset Allocation
Homework for Week 4

John Dodson

September 24, 2014

A solution to these problems is due at the beginning of the next session, which is 5:30 PM on Wednesday, October 1.

Timeseries

An asset price *timeseries* is a set of ordered pairs,

$$\{(\text{timestamp } t_i, \text{value } p_i) : i \in \{0, 1, \ldots, N\}\}$$

Without loss of generality, we can assume $t_0 < t_1 < \cdots < t_N$. Usually, we can also assume that the periods between timestamps are regular.

Geometric Brownian motion

A classic model for the dynamics of an asset price (paying no dividends or coupons) is *geometric Brownian motion* (GBM), which we can write as a stochastic differential equation $dP = P\mu\, dt + P\sigma\, dB_t$ in terms of a standard Brownian motion B_t and constants μ ("drift") and $\sigma > 0$ ("volatility"); hence

$$P(t) = P(0)e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma B_t}$$

Since (non-overlapping) increments of B_t (hence non-overlapping increments of $\log P$) are normal and independent,

$$B(t)|\mathcal{F}_s \sim \mathcal{N}\left(B(s), t-s\right) \quad \forall\, t > s \geq 0$$

a likelihood function for the parameters μ and σ can be formed from a time series of observations of P.

Problems

Let us work with FB (Facebook) and GOOGL (Google) common equity[1] daily closing price time series for two years through September 24.

1. Evaluate the maximum likelihood estimate and Cramér-Rao lower bound for the standard error of the estimate of the GBM parameters (μ, σ) of each. (**8 points**)

[1] Google split its public shares on March 27, 2014. The existing "class A" shares were assigned the new GOOGL ticker, while the new "class C" shares were assigned the existing GOOG ticker (the "class B" shares remain unlisted).

Kendall's rank correlation for a sample $\left(\left(x_{1,1}, x_{1,2}\right), \left(x_{2,1}, x_{2,2}\right), \ldots, \left(x_{N,1}, x_{N,2}\right)\right)$ of a bivariate random variable X is

$$\hat{\tau}_{1,2} \triangleq \binom{N}{2}^{-1} \sum_{1 \leq i < j \leq N} \operatorname{sgn}\left(\left(x_{i,1} - x_{j,1}\right)\left(x_{i,2} - x_{j,2}\right)\right)$$

where $\operatorname{sgn}(\cdot)$ is the signum function, which evaluates to $-1, 0, 1$ according to the sign of the argument.

2. Assuming again that continuous daily returns are i.i.d., evaluate Kendall's rank correlation for FB and GOOGL. (**1 point**)
3. By method of moments, what is the corresponding elliptical copula pseudo-correlation estimate? (**1 point**)

Technical Hint

In MATLAB, the principal routine in the core language for numerical optimization is `fminsearch()`. For example, if you are trying to solve the optimization problem,

$$x^{\star} = \arg \min_{x \in \mathbb{R}^2} \left(x_1 + 1\right)^2 + \left(x_2 - 2\right)^2$$

you can use calculus to conclude $x^{\star} = (-1, 2)'$. In MATLAB, you can code this as

```
>> x=fminsearch(@(x)(x(1)+1)^2+(x(2)-2)^2,[0 0])

x =

   -1.0000    2.0000
```

Obviously, if you are seeking a local maximum rather than a local minimum, you can reverse the sign of the objective function.

Solution

First we have to assemble our data.

```
>> raw=yahoo_prices({'FB','GOOGL'},'2012-09-21','2014-09-24'); % adjust start for weekend
>> tsc=resample(raw,raw.Time(~isnan(raw.FB.Data))); % remove holidays
```

Note that I am using a timeseries function to remove non-trading days here.

Next we need to extract the log-returns.

```
>> ret.FB=log(tsc.FB.Data(2:end))-log(tsc.FB.Data(1:end-1));
>> ret.GOOGL=log(tsc.GOOGL.Data(2:end))-log(tsc.GOOGL.Data(1:end-1));
```

Let's measure time in trading days, and let's define

$$x_i = \log p_i - \log p_{i-1}$$

In these terms, the log-likelihood function for the GBM parameters is

$$
\begin{aligned}
L(\mu, \sigma) &= \log \prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(x_i - \mu + \frac{1}{2}\sigma^2\right)^2 / \left(2\sigma^2\right)} \\
&= \text{const.} - \sum_i \left(\frac{\left(x_i - \mu + \frac{1}{2}\sigma^2\right)^2}{2\sigma^2} + \log \sigma \right)
\end{aligned}
$$

You can code this as

```
>> f=@(mu,sigma,x)(x-mu+sigma^2/2).^2/(2*sigma^2)+log(sigma);

>> theta.FB=fminsearch(@(theta)sum(f(theta(1),theta(2),ret.FB)),[0 0]);
>> theta.GOOGL=fminsearch(@(theta)sum(f(theta(1),theta(2),ret.GOOGL)),[0 0]);
```

For the next part, I am asking for the standard errors of the estimates. There are now two different kinds of standard deviations here: the standard deviation of the random variable, σ, and the standard deviation of the estimators, $\operatorname{se}\hat{\mu}$ and $\operatorname{se}\hat{\sigma}$. Please do not get confused! The latter is a property of the sample. The former is a property of the model. For example, as the sample size goes to infinity, the standard errors go to zero, while this has no bearing on the standard deviation of the random variable.

There are several ways to approach this. You could note the underlying normality of the model and make use of the results for the normal. It is relatively uncommon that there are analytical result available. Instead, we will note that the Fisher information is the covariance of the gradient of the log of the joint density and that

$$
\begin{aligned}
\frac{\partial}{\partial \mu} \log f_{X|\mu,\sigma}(X) &= \frac{X - \mu}{\sigma^2} + \frac{1}{2} \\
\frac{\partial}{\partial \sigma} \log f_{X|\mu,\sigma}(X) &= \frac{(X - \mu)^2}{\sigma^3} - \frac{1}{\sigma} - \frac{\sigma}{4}
\end{aligned}
$$

We can use the sample of log-returns and the MLE parameter estimates to estimate this covariance.

```
>> Info.FB=cov((ret.FB-theta.FB(1))/theta.FB(2)^2,...
    (ret.FB-theta.FB(1)).^2/theta.FB(2)^3);
>> Info.GOOGL=cov((ret.GOOGL-theta.GOOGL(1))/theta.GOOGL(2)^2,...
    (ret.GOOGL-theta.GOOGL(1)).^2/theta.GOOGL(2)^3);
```

While not strictly true, we can assume here that the estimator is unbiased, so the (co)variance of the estimator is bounded below by the inverse of the Fisher information of the sample.

```
>> N=length(ret.FB);
>> SE.FB=sqrt(diag(inv(N*Info.FB)));
>> SE.GOOGL=sqrt(diag(inv(N*Info.GOOGL)));
```

The results are summarized in Table 1.

Notice that the significance of our volatility estimates is much greater than the significance of our drift estimates. In fact, our drift estimates are so weak that we cannot even be confident that the signs are correct! This is typical, and it signifies a deep problem in the objective analysis of asset returns.

ticker	daily drift (μ)	daily volatility (σ)
FB	0.003 ± 0.001	0.0282 ± 0.0003
GOOGL	0.0011 ± 0.0006	0.0137 ± 0.0001

Table 1: Geometric Brownian motion parameters to one significant digit of the standard error estimate.

For the last part, you are to consider a measure of dependence. Let me discuss the difference between correlation and pseudo-correlation. If the joint model of the data were elliptical (e.g. bivariate normal), we could estimate the copula parameter with an MLE $\hat{\rho}$ in conjunction with the marginal parameters.

If we are not prepared to assume that the join model is elliptical, but we still want to characterize the dependence structure with pair-wise parameters, we can use something like Kendall's rank correlation metric, which is based on matching moments of the distribution of concordances.

Elliptical copulas could be expressed directly in terms of pair-wise Kendall's taus, but it is conventional to express them in terms of pseudo-correlations,

$$\rho = \sin \tfrac{\pi}{2} \tau$$

This is because psedo-correlation coincides with (Pearson's) correlation for jointly elliptical models, and elliptical copulas are a relatively new idea.

For the sample concordance measurement, we have to loop over pairs of dates.

```
pairs=0;conc=0;
for i=1:N-1
    for j=i+1:N
        pairs=pairs+1;
        conc=conc+sign(ret.FB(i)-ret.FB(j))...
            *sign(ret.GOOGL(i)-ret.GOOGL(j));
    end
end
>> tau=conc/pairs

tau =

   0.2767

>> rho=sin(tau*pi/2)

rho =

   0.4211
```

That is, the Kendall's rank-order correlation is about 0.28 and the corresponding elliptical pseudo-correlation is about 0.42.

By the way, the Pearson's correlation (based on an assumption of bivariate normality) is only about 0.28, which would seem to be a significant underestimate compared to the pseudo-correlation.