18.S096 Problem Set 5 Fall 2013
Volatility Modeling
Due Date: 10/29/2013

1. **Sample Estimators of Diffusion Process Volatility and Drift**

Let $\{X_t\}$ be the price of a financial security that follows a geometric Brownian motion process:

$$\frac{dX(t)}{X(t)} = \mu_* dt + \sigma dW(t),$$

where

- $\sigma > 0$, is the volatility parameter
- $\mu_* \in (-\infty, \infty)$, is the drift parameter
- $dX(t)$ is the infinitesimal increment in price.
- $dW(t)$ is the increment of a standard Wiener Process, i.e, infinitesimal increments $W(t+dt)-W(t)$ are i.i.d. Normal random variables with zero mean and variance equal to 'dt'.

Consider sampling values of the price process over a fixed time period $t \in [0, T]$, at equal time increments $h = T/n$. Define

$$t_i = i \times h,\ i = 0, 1, \ldots, n$$
$$X_i = X(t_i),\ i = 0, 1, \ldots, n$$
$$Y_i = log(X_i/X_{i-1}),\ i = 1, 2, \ldots, n$$

Accept as given that:

$$Y_i \text{ are i.i.d. } N(\mu \cdot h, \sigma^2 \cdot h) \text{ random variables,}$$

(this is proven with the theory of diffusion processes/stochastic differential equations, with $\mu = \mu_* - \frac{1}{2}\sigma^2$).

1(a) Prove that the Maximum-Likelihood Estimates: $\hat{\mu}$ and $\hat{\sigma}$ for a sample: $y_1, y_2, \ldots, y_n$, are given by
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2$$

1(b) Derive the distribution of $\hat{\mu}$; give specific formulas for the expectation and variance of $\hat{\mu}$.

1(c) Derive the distribution of $\hat{\sigma}^2$; give specific formulas for the expectation and variance of $\hat{\sigma}^2$.

1(d) Consider increasing the number of increments n on the fixed time period $[0, T]$, and let $\hat{\mu}_n$ and $\hat{\sigma}^2_n$ be the corresponding MLEs of the parameters. Determine the limiting distributions of $\hat{\mu}_n$ and $\hat{\sigma}^2_n$.

1(e) A sequence of estimators $\hat{\theta}_n$ for a parameter θ, is weakly consistent if

$$\lim_{n\to\infty} Pr(|\hat{\theta}_n - \theta|) = 0.$$

For each of $\hat{\mu}_n$ and $\hat{\sigma}^2_n$, determine whether the sequence of estimators is weakly consistent.

2. Consider the same process as in problem 1, but now, for fixed values of μ and σ, consider sampling n values of the price process over a fixed time period $t \in [0, T]$, at variable increments $h_i > 0$, $i = 1, 2, \ldots, n$, such that $\sum_{i=1}^{n} h_i = T$. Define

   $t_i = \sum_{j=1}^{i} h_j$, $i = 0, 1, \ldots, n$
   $X_i = X(t_i)$, $i = 0, 1, \ldots, n$
   $Y_i = log(X_i/X_{i-1})$, $i = 1, 2, \ldots, n$

   Accept as given that:

   Y_i are i.i.d. $N(\mu \cdot h_i, \sigma^2 \cdot h_i)$ random variables,

   (this is proven with the theory of diffusion processes/stochastic differential equations).

2(a) Derive the MLE for μ and its distribution for a fixed set of sampling increments $\{h_i\} : \sum_{i=1}^{n} h_i = T$.

2(b) Derive the MLE for σ^2 and its distribution for a fixed set of sampling increments $\{h_i\} : \sum_{i=1}^{n} h_i = T$.

2(c) If limited to sampling $n + 1$ price points of $\{X_t\}$, (including X_0 and X_T) prove that

- For estimating σ^2, sampling, the ML estimators vary with the increment spacing, but the variance of these estimators are all equal, regardless of the increment spacing.
- For estimating μ, all ML estimators are the same and have the same variance, regardless of the increment spacing.

3. **ARCH(1) Model Properties**

Let $y_t = log(S_t/S_{t-1})$ be the discrete returns of the price of a security/portfolio $\{S_t, t = 1, 2, \ldots\}$, and supppose that $y_t \sim ARCH(1)$, i.e.

$$y_t = \mu_t + \epsilon_t,$$

where μ_t is the mean return, conditional on $\mathcal{F}_{t-1}$, the information available up to time $(t-1)$ and

$$\epsilon_t = Z_t \sigma_t,$$

where Z_t iid with $E[Z_t] = 0$, and $var[Z_t] = 1$, and

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2.$$

Additionally, suppose that $E[Z_t^3] = 0$, and $E[Z_t^4] = \kappa$. (The parameter κ is the Kurtosis of the Z_t distribution with unit variance; if Z_t is Gaussian/normal, then $\kappa = 3$.

Prove that:

3(a) $E[\epsilon_t^2] = \alpha_0/(1-\alpha_1)$

3(b) $E[\epsilon_t^3] = 0$

3(c) $E[\epsilon_t^4] = \frac{\kappa\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-\kappa\alpha_1^2)}$

3(d) What constraints on α_0, α_1 must be made in (c), to maintain 4-th order stationarity (bounded).

3(e) The kurtosis of ϵ_t is

$$\kappa_\epsilon = E[\epsilon_t^4]/(E[\epsilon_t^2])^2.$$

(The fourth moment is normalized to be scale-free). If the distribution Z_t is Gaussian/normal (i.e., the scaled, conditional error distribution of ϵ_t), does the unconditional distribution of ϵ_t, have a higher than that of the Gaussian distribution, (i.e., heavier tails)?

4. Using Daily Open/High/Low/Close Data on the S&P500 Index from 2006-20012 , annual sample variances were computed of changes in the log index value of the daily Close.

The following table gives the annual sample variances, day counts, and annualized volatilities

Annual Sample Variances of Logarithmic Returns:

	daily.variance	days	volatility
2006	3.981351e-05	251	0.09996595
2007	1.018599e-04	251	0.15989632
2008	6.677100e-04	253	0.41101171
2009	2.950132e-04	252	0.27265972
2010	1.294545e-04	252	0.18061706
2011	2.164385e-04	252	0.23354335
2012	6.459046e-05	250	0.12707327

The differences in the sample variances and volatilities appears quite large for some years. Are the year-by-year differences significant?

To address this question, consider modeling the returns for any given year as a simple random sample from a Gaussian distribution:

$$\{y_1, y_2, \ldots, y_n\} : y_i \text{ } i.i.d. \text{ } N(\mu, \sigma^2).$$

The table gives values of $\hat{\sigma}^2$, n, and $\sqrt{n}\hat{\sigma}$, where

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \hat{\mu})^2,$$

with $$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n y_i.$$

4(a) Under the Gaussian model, given n, μ, σ^2, prove that the distribution of $\hat{\sigma}^2$ is

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n-1} \times \chi^2_{n-1}.$$

that is, a scaled Chi-square distribution with degrees of freedom equal to $(n-1)$, and scale factor equal to $\frac{\sigma^2}{n-1}$

4(b) Statistical methodology defines confidence intervals for unknown parameters by computing a likely interval for the parameter estimate given the unknown parameter, and then inverting the interval to correspond to the parameter instead of the estimate.

For a 95% confidence interval (two-sided), the development is as follows:

- With 95% probability, the χ^2_{n-1} random variable will fall within the interval from the 0.025 percentile, to the 0.975 percentile, i.e.,

  $$Pr(q_{0.025} < \chi^2_{n-1} \leq q_{0.975}) = 0.975 - 0.025 = 0.95$$

  where

  $$Pr(\chi^2_{n-1} \leq q_{0.025}) = 0.025$$
  $$Pr(\chi^2_{n-1} \leq q_{0.975}) = 0.975$$

- Replacing the random variable χ^2_{n-1} with $(\frac{n-1}{\sigma^2})\hat{\sigma}^2$ gives:

  $$Pr(q_{0.025} < (\frac{n-1}{\sigma^2})\hat{\sigma}^2 \leq q_{0.975}) = 0.95$$

  which can be inverted to

  $$Pr(\hat{\sigma}^2 \frac{(n-1)}{q_{0.975}} \leq \sigma^2 \leq \hat{\sigma}^2 \frac{(n-1)}{q_{0.025}}) = 0.95$$

The following table gives the percentiles of the Chi-square distributions for degrees of freedom ranging from 249 to 252 (one less than the annual day counts).

df	q0.025	q0.975	ll.factor	ul.factor
249	207.1856	294.6008	0.8452115	1.201821
250	208.0978	295.6886	0.8454840	1.201358
251	209.0102	296.7763	0.8457550	1.200898
252	209.9227	297.8637	0.8460245	1.200442

The last two columns are

$$ll.factor = \frac{n-1}{q_{0.975}} \text{ and } ul.factor = \frac{n-1}{q_{0.025}}$$

which when multiplied by the unbiased sample estimate $\hat{\sigma}^2$, define the confidence interval for σ^2.

- Using data for 2008, compute the two-sided 95% confidence interval for σ^2, based on daily log returns.
- Express the interval in terms of the annualized volatility $(\sqrt{253}\sigma)$. Does the sample annual volatility for any other year fall in the confidence interval for 2008?

4(c) The return variance / volatility varies considerably from year to year. To evaluate the statistical significance of the difference in values for any two years, we can use the F-Distribution. Consider 2007 and 2008.

Under the assumption (i.e., a null hypothesis H_0) of Gaussian/normal daily returns and that the variances of the returns are constant/ the same for all days in the two years it follows from 4(a) that:

$$X = (\frac{n_{2007}-1)}{\sigma^2})\hat{\sigma}^2_{2007} \sim \chi^2_{df_1}, \text{ where } df_X = (n_{2007} - 1)$$
$$Y = (\frac{n_{2008}-1)}{\sigma^2})\hat{\sigma}^2_{2008} \sim \chi^2_{df_2}, \text{ where } df_Y = (n_{2008} - 1)$$

and X and Y are independent random variables. The statistic

$$S = \frac{Y/df_Y}{X/df_X} = \left(\frac{\hat{\sigma}^2_{2008}}{\hat{\sigma}^2_{2007}}\right)$$

has the F-Distribution with degrees of freedom df_Y for the numerator and df_X for the denominator. (Verify by looking up the definition of the F-distribution.)

Under the null hypothesis, the numerator and denominator of S are estimates of the same return variance. Their ratio varies about 1 due to the independent variation in the numerator and denominator of scaled Chi-squared random variables.

The methodology of hypothesis testing in statistics uses the fact that the test statistic has a known distribution under the null hypothesis. The null hypothesis is accepted / rejected so long as the test statistic is not extreme. We choose a test α-*level*, the probability of (falsely) rejecting the null hypothesis if true, say $\alpha = 0.05$. From this, extreme ranges of the test statistic are defined that occur with probability α when the null hypothesis is true. For $\alpha = 0.05$, a $two - sided$ alternative is considered using $q_{0.025}$ and $q_{0.975}$, the percentiles of the F distribution given by:

$$Pr(F_{df_Y,df_X} < q_{0.025}) = 0.025$$
$$Pr(F_{df_Y,df_X} < q_{0.975}) = 0.975$$

The null hypothesis is accepted if

$$q_{0.025} < S < q_{0.975}$$

From the package R, we provide the percentiles of the F-distribution when $df_1 = n_{2008} - 1$, and $df_2 = n_{2007} - 1$:

```
> qf(0.025, df1=252, df2=250)
                [1] 0.7804173
> qf(0.975, df1=252, df2=250)
                [1] 1.281525
```

so the null hypothesis is accepted if

$$q_{0.025} = 0.7804 < S < 1.2815 = q_{0.975}$$

- Compute the test statistic $S = S_0$ for testing the daily return variance for 2008 is equal to the daily return variance for 2007.
- Given the value of the test statistic S_0, determine the $\alpha-$level at which the null hypothesis is on the boundary of being just accepted/rejected.

(This level is called the P-value of the test statistic. Reporting a test statistic's P-value provides evidence concerning for/against the test null hypothesis which can be provided without having to specify an α-level.)

- Repeat the previous two questions, for testing the equality of the return variance for 2008 to that for 2006. (Note: the degrees of freedom for 2006 are the same as those for 2007 so the same F distribution is applicable)