Tim Cohn

Low Outliers

Tim Cohn
703-395-0204
tim@timcohn.com

Tim Cohn
5 Feb '10

Low Outlier Detection

Assume

$\underline{X} = \{X_1, \ldots X_N\}$ iid Gaussian deviates

Define

$$M = \frac{1}{(N-r)} \sum_{i=r+1}^{N} (X_{[i]})$$

$$S^2 = \frac{1}{(N-r-1)} \sum (X_{[i]} - M)^2$$

where

$X_{[i]} \equiv i^{th}$ order statistic of $\underline{X}$

Question:

What is the distribution of

$$(X_{[r]} - M)/\sqrt{S^2}$$

Specifically, what is

$$p(\eta) \equiv P[\,X_{[R]} \le M + \eta S\,] \qquad (1)$$

Exact Solution:

$$p(\eta) = \int_{X_N}\int_{X_{N-1}} \cdots \int_{X_1} I\left(X_{[R]} < \frac{\Sigma X_{[\cdot]}}{(N-R)} - \eta \cdot \sqrt{\frac{\Sigma \left(X_{[\cdot]} - \frac{\Sigma X_i}{\cdot}\right)^2}{N-R-1}} \right) f(X_1) \cdots f(X_N)\, dX_1\, dX_2 \cdots dX_N$$

This integral can be evaluated exactly for small values of N.

Monte Carlo simulation works well for large N.

Approximate Solution

Note that M, S^2, and $X_{[R]}$ are the only random variables in equation 1. Then:

~~If we assume that M is approximately Gaussian~~

$$\cancel{M \sim N(\mu_M, \sigma_M^2)}$$

$$p(\eta) = \int_{X_{[R]}} \int_{S^2} \int_{M} I\left(X_{[R]} \le m + \eta\sqrt{S^2}\right) f_M(m \mid s^2, x_R) \cdot f_{S^2}(s^2 \mid x_R) f_{X_R}(x) \, dm \, ds^2 \, dx$$

This assumes that $M \perp S^2$, which is true for the case $r = 0$ but only approximate otherwise; $\operatorname{cov}(M, S^2)$

Q: What are the Distribution of M, S^2?

Approach: Compute Moments.

- Condition on $X_{[R]}$ fixed. Then integrate as function of $X_{[R]}$

Lemma 1

$$E[X^k \mid X_0 < X < \infty, X \sim \text{Truncated Standard Normal}]$$

Integrate by Parts.

Define $H(\zeta) = \phi(\zeta)/(1-\Phi(\zeta))$

k	$E[X^k \mid X \geq \zeta]$
0	1
1	$H(\zeta)$
2	$1+\zeta H(\zeta)$
3	$2H(\zeta)+\zeta^2 H(\zeta)$
⋮	⋮
k	$(k-1)E[X^{k-2} \mid X > \zeta] + \zeta^{k-1}H(\zeta)$

Expected Values of M, S^2 Given $X_{[R]} = \xi$

$$E[M] = E\left[\frac{1}{N-R}\sum_{i=R+1}^{N} X_i\right]$$

$$= E[X]$$

$$= H(\xi)$$

$$E[S^2] = E\left[\frac{1}{(N-1)}\sum_{R+1}^{N}(X_i - M)^2\right]$$

$$= E\left[\left(\frac{1}{N}\right)\sum\left(X_i^2 - E^2[X]\right)\right]$$

$$= E[X^2] - E^2[X]$$

$$= 1 + \xi H(\xi) - H^2(\xi)$$

Variance - Covariance of M, S^2

$$Var[M] = Var\left[\frac{1}{N-R}\sum X_i\right]$$

$$= \left(\frac{1}{N-R}\right) \cdot Var[X]$$

$$= \left(\frac{1}{N-R}\right)\left(1 + \xi H(\xi) - H^2(\xi)\right)$$

2/11/16

$$Cov(M, S^2) = E\left[\left(\frac{1}{N'}\right)\sum(X_i - M)\left(\frac{1}{N'-1}\right)\sum(X_i - M)^2\right]$$

$$= \frac{E\left[\sum(X_i - M)^3\right]}{N'(N'-1)} + \underbrace{\sum_{i \neq j}\sum \frac{E\left[(X_i - M)(X_j - M)^2\right]}{N'(N'-1)}}_{0 \text{ because } X_i \perp X_j}$$

** Test that term 2 = 0 by MC experiment

$$= \frac{1}{(N'-1)}\left\{E[X^3] - 3E[X^2]E[X] + 2E^3[X]\right\}$$

$$Var[S^2] = E\left[(S^2 - E[S^2])^2\right]$$

$$= \left(\frac{1}{N'}\right)\left\{E[(X - E[X])^4] - E^2[(X - E[X])^2]\right\}$$

$$= \left(\frac{1}{N'}\right)\left\{E[X^4] - 4E[X^3]E[X] + 6E[X^2]E^2[X] - 3E^4[X] - (E[X^2] - E^2[X])^2\right\}$$

Crazy Analysis

$$\mathrm{Var}[\sqrt{X}] = E[X] - E[\sqrt{X}]^2$$

$$\mathrm{Var}\left[\frac{(X-X_0)}{2\sqrt{X_0}} - \frac{(X-X_0)^2}{8\sqrt{X_0}^3}\right] = \frac{1}{64X_0^3}\,\mathrm{Var}\left[4X_0(X-X_0) - (X-X_0)^2\right]$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{\mathrm{Var}\left[4X_0X - 4X_0^2 - X^2 + 2X_0X - X_0^2\right]\right\}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{\mathrm{Var}\left[-X^2 + 6X_0X\right]\right\}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{E\left[\left(-X^2 + 6X_0X\right)^2\right] - E^2\left[-X^2 + 6X_0X\right]\right\}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{E\left[X^4 - 12X_0X^3 + 36X_0^2X^2\right] - E^2\left[-X^2 + 6X_0X\right]\right\}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{E[X^4] - 12X_0E[X^3] + 36X_0^2E[X^2] - \left(-E[X^2] + 6X_0E[X]\right)^2\right\}$$

** Check this with MC simulation

Note: HOT may be important.

Taylor series for $\sqrt{X}$ not fast-convergent.

$(x-x_0)^2]$

Covariance of M, S

$$\mathrm{Cov}[M, S] = \mathrm{Cov}[M, \sqrt{S^2}]$$

$$= \left(\frac{1}{2\cdot\sqrt{E[S^2]}}\right)\cdot \mathrm{Cov}[M, S^2]$$

$$\mathrm{Var}[S] = \mathrm{Var}[\sqrt{S^2}]$$

$$\approx \mathrm{Var}[\text{Taylor Series Summation}]$$

$\}$

$f(x) = \sqrt{x} \cong$

$x])^2\}$

$$= \sqrt{x_0 + (x - x_0)}$$

$$= (x_0)^{1/2} + \left(\tfrac{1}{2}\right)(x_0)^{-1/2}(x-x_0) - \left(\tfrac{1}{4}\right) x_0^{-3/2}\frac{(x-x_0)^2}{2} + \left(\tfrac{3}{8}\right) x_0^{-5/2}\frac{(x-x_0)^3}{6}$$

$$= x_0^{1/2} + \frac{x_0^{-1/2}(x-x_0)}{2} - \frac{x_0^{-3/2}(x-x_0)^2}{8} + \frac{x_0^{-5/2}(x-x_0)^3}{16}$$

y cav.

$$\mathrm{Var}\left[x_0^{1/2} + \frac{x_0^{-1/2}(x-x_0)}{2} - \frac{x_0^{-3/2}(x-x_0)^2}{8} + \frac{x_0^{-5/2}(x-x_0)^3}{16}\right]$$

0

$=$

$$= \mathrm{Var}\left[\frac{(x-x_0)}{2\cdot\sqrt{x_0}} - \frac{(x-x_0)^2}{8\sqrt{x_0}^3} + HOT\right]$$

?

$$\left(\frac{1}{64 S_0^6}\right)\left\{E[S^8] - 12X_0 E[S^6] + 36X_0^2 E[S^4] - E^2[S^4] + 12X_0^3 E[S^4] - 36X_0^4\right\}$$

$$\left(\frac{1}{64X_0^6}\right)\left\{36 \cdot X_0^2\left(E[S^4] - E^2[S^2]\right) + 12X_0\left(X_0^2 E[S^4] - E[S^6]\right) + \left(E[S^8] - E^2[S^4]\right)\right\}$$

$$= \left(\frac{1}{64X_0^6}\right)\left\{36X_0^2 \, Var[S^2] + Var(S^4)\right)$$

$[S^4]-36X_0^4\}$

$[S^6]) +$

$$Var[S] = Var[\sqrt{S^2}]$$

$$= \left(\frac{1}{64 S_0^6}\right)\left\{E[S^8] - 12X_0 E[S^6] + 36X_0^2 E[S^4] - \left(-E[X^2] + 6X_0 E[X]\right)^2\right\}$$

Note: This is __not__ the same result that I derived in my notes of 12/16/09. The 12/16/09 results contained an error. (:-( ) [Maybe not ?? **

** Check __both__ for other errors.

Complete Case

$$E[S] \approx S_0 - \frac{Var[S^2]}{8 S_0^3}$$

$$= \sigma - \left(\frac{2\sigma^4}{N-1}\right) / (8 \cdot \sigma^3)$$

$$= \sigma\left(1 - \frac{1}{4(N-1)}\right) \quad \checkmark$$

[see p. 31]

2/11/10

Expected Values of M, S^2, S

$$E[M] = E[\tfrac{1}{N}\Sigma X_i]$$

$$= E[X_i] = E[X]$$

$$E[S^2] = E[(\tfrac{1}{N-1})\Sigma(X_i - M)^2]$$

$$= E[X^2] - E^2[X]$$

$$E[S] = E\left[(S_0^2)^{1/2} + \left(\tfrac{1}{2}(S_0^2)^{-1/2}\right)(S^2 - S_0^2) - \tfrac{1}{4}\left((S_0^2)^{-3/2}\right)\frac{(S^2 - S_0^2)^2}{2} + \ldots\right]$$

$$= S_0 + \quad 0 \quad - \frac{Var[S^2]}{8 S_0^3}$$

$$= S_0 - \frac{Var[S^2]}{8 S_0^3}$$

Notes: This looks right!

However, see p. 45

2/11/10

Definition of Generalized Grubbs-Beck Test

$$\hat{\omega}_i \equiv \left(\frac{X_{[i:N]} - M}{S}\right)$$

$$p(\eta) \equiv P[\hat{\omega}_i < \eta \mid H_0]$$

$$= P[X_{[i:N]} < M + \eta S]$$

$$= P[X_{[i:N]} < (M - \lambda S) + (\eta + \lambda) S]$$

$$= P\left[\frac{X_{[i:N]} - M'}{S} < (\eta + \lambda)\right]$$

where

$$\lambda = \frac{Cov[M, S]}{Var[S]}$$

Note: λ is non-dimensional

$$M' = M - \lambda S$$

which guarantees

$$M' \perp S \qquad \text{"First Order"}$$

$$\mathrm{Cov}(M, S) = \mathrm{Cov}(M, \sqrt{S^2})$$

$$= \mathrm{Cov}(M, \sqrt{(S^2 - \sigma^2) + \sigma^2})$$

$$= \frac{\mathrm{Cov}(M, S^2)}{2\sqrt{\sigma^2}}$$

$$= \frac{\mathrm{Cov}(M, S^2)}{2 \cdot \sigma}$$

2/11/10

$$\mathrm{Cov}(M', S) = \mathrm{Cov}[M, S] + \mathrm{Cov}(\lambda S, S)$$

$$= \mathrm{Cov}[M, S] - \lambda \mathrm{Var}[S]$$

$$= \mathrm{Cov}[M, S] - \frac{\mathrm{Cov}[M, S]}{\mathrm{Var}[S]} \cdot \mathrm{Var}[S]$$

$$= 0 \qquad \blacksquare$$

Notes: For fixed $X_{[i:N]}$, $(X_{[i:N]} - M')$ is orthogonal to S

The ratio of $(X_{[i:N]} - M')$ to S is thus (approx) proportional to a non-central Student's t variate.

2/11/10

Distribution of M', S and $X_{[i:n]}$

Assume $M' \sim N(\mu_{M'}, \sigma^2_{M'})$

$$\mu_{M'} = E[M'] = E[M] - \lambda E[S]$$

$$\sigma^2_{M'} = Var[M - \lambda S]$$

$$= Var[M] - 2\lambda Cov[M, S] + \lambda^2 Var[S]$$

$$= Var[M] - Cov^2[M, S] / Var[S]$$

Assume $S^2 \sim \Gamma(\alpha, \beta)$

$$\alpha = E^2[S^2] / Var[S^2]$$

$$\beta = \left(\frac{1}{\alpha}\right) E[S^2]$$

Derivation of P-Value Computation (Assume $X \sim N(\mu, \sigma^2)$)

$$p(\eta) \equiv P\left[\frac{X_{[i:N]} - M}{S} < \eta\right] =$$

$$= P\left[\frac{\frac{(X_{[i:N]}-\mu)}{\sigma} - \frac{(M-\mu)}{\sigma}}{S/\sigma}\right]$$

$$= P\left[Z_{[i:N]} - M_Z + \eta' S_Z\right] \quad ; \quad \eta' = \eta + \lambda$$

$$= \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{z-\eta' s}^{\infty} f_{M'_Z, S_Z}(m, s^2 | z)\, dm \cdot ds^2\, f_{Z_{[i:N]}}(z)\, dz$$

$z \quad s^2 \quad m$

$$= \int_{-\infty}^{\infty} \int_{0}^{\infty} \left(1 - F_{M'_Z}(z - \eta' s)\right) \cdot f_{S_Z^2} \cdot ds^2\, f_{Z_{[i:N]}}(z)\, dz$$

$z \quad s^2$

$$F_{M'_Z}(y) \equiv \Phi\left(\frac{y - E[M'_Z]}{\sqrt{Var[M'_Z]}}\right)$$

$$f_{s_z^2}(s^2) = \frac{(s^2/\beta)^{\alpha-1} e^{-(s^2/\beta)}}{\beta \cdot \Gamma(\alpha)}$$

$$f_z(z) \equiv N \cdot \binom{N-1}{i} \Phi(z)^{i-1} (1-\Phi(z))^{N-i} \phi(z)$$

Note that this is a classical order-statistics result.

Check on 2 Derivations of Var[S]

$$Q: \ Var[\sqrt{X}] = \frac{Var[X]}{4X_0} - \frac{Var^2[X]}{64X_0^3}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{16X_0^2 \cdot Var[X] - Var^2[X]\right\}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{16X_0^2 \cdot E[X^2] - 16X_0^2 E^2[X] - \left(E[X^2] - E^2[X]\right)^2\right\}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{-E^2[X^2] + 2E[X]E^2[X] - E^4[X] + 16X_0^2 E[X^2] - 16X_0^2 E^2[X]\right\}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{(+-E^2[X^2] + 2X_0^3 - X_0^4 + 16X_0^2 E[X^2] - 16X_0^4\right\}$$

$$= \left(\frac{1}{64X_0^3}\right)\left\{-\left(E[X^2]\right)^2 + 16X_0^2 \cdot E[X^2] - 17X_0^4 + 2X_0^3\right\}$$

$=$

See p. 176 for analysis based on p. 31 results.

cls.

Check for Complete Normals

$(s^2/\sigma^2)/\sqrt{N-1} \sim \chi^2_{N-1}$

For standard normals

$$s^2 \sim \Gamma\left(\alpha = \frac{(N-1)}{2}, \beta = \sigma^2\left(\frac{2}{N-1}\right)\right)$$

$E[s^2] = \sigma^2$

$E[s^{2k}] = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \cdot \beta^k$

$$= (\sigma^2)^k \left(\frac{2}{N-1}\right)^k \Gamma\left(\frac{N-1}{2}+k\right) / \Gamma\left(\frac{N-1}{2}\right)$$

k	$E[s^{2k}]$	
0	1	
1	σ^2	$\sigma^2\left(\frac{2}{N-1}\right)\left(\frac{N-1}{2}\right)$
2	$\sigma^4\left(\frac{N+1}{N-1}\right)$	$\sigma^4\left(\frac{2}{N-1}\right)^2\left(\frac{N-1}{2}\right)\left(\frac{N+1}{2}\right) = \sigma^4\left(\frac{N+1}{N-1}\right)$
3	$\sigma^6\left(\frac{(N+1)(N+3)}{(N-1)^2}\right)$	$\sigma^6\left(\frac{(N+1)(N+3)}{(N-1)^2}\right)$
4	$\sigma^8\left(\frac{(N+1)(N+3)(N+5)}{(N-1)^3}\right)$	$\sigma^8\left[\frac{(N+1)(N+3)(N+5)}{(N-1)^3}\right]$
⋮		
k	$\sigma^{2k}\left(\frac{2}{N-1}\right)^k\left(\frac{\Gamma\left(\frac{N-1}{2}+k\right)}{\Gamma\left(\frac{N-1}{2}\right)}\right)$	

Questions:

1) Is $S_0^2 = E[S^2 | S^2 = \frac{1}{N-R-1} \sum_{i=R+1}^{N} X_{[i:N]}]$?

2) If estimator for S^2 is OK, why is estimator for S apparently biased?

Check: Var[S] computation appears to be OK

3) Need to check if integrate is using valid functions. Is the function vector-capable?

4) Does "qmin" refer to non-exceedance probability or quantile? (X or P[X]?)

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{1} f(g(p))\left(\frac{dg}{dp}\right)dp$$

$$g(p) = \ln\left(p/(1-p)\right)$$

$$\frac{dg(p)}{dp} = \frac{\frac{\partial(p/(1-p))}{\partial p}}{p/(1-p)} = \frac{\frac{1\cdot(1-p)-p(-1)}{(1-p)^2}}{p/(1-p)}$$

$$= \left(p/(1-p)\right)$$

Observations 2/11/10

1) Expected values for M, S^2, S OK
2) Variances for M, S, S^2 OK.
3) Computed p-values $p(\eta, n, r)$ are not accurate. Why?
4) Need updated "integrate" function in R that can utilize non-vector functions

1)

Applying Non-Central T Dist^{n}

$$\hat{\omega} \equiv \left(\frac{X_{[k]} - m}{s}\right)$$

$$p(\eta) = P[\hat{\omega} < \eta]$$

$$= P\left[\frac{X - m + \lambda s}{s} < \eta + \lambda\right] = P\left[\frac{m - \lambda s - X}{s} > -(\eta + \lambda)\right]$$

$$= 1 - P\left[\frac{m - \lambda s - X}{s} < -(\eta + \lambda)\right]$$

$$= 1 - P\left[\frac{(m' - X)/\sigma_{m'}}{(s/\sigma_s)} < -\left(\frac{\sigma_s}{\sigma_{m'}}\right)\eta'\right] \qquad m' \equiv m - \lambda \cdot s \qquad \eta' = \eta + \lambda$$

$$= 1 - P\left[\frac{X}{\sqrt{Y^2/df}} < -\left(\frac{\sigma_s}{\sigma_{m'}}\right)\eta'\right]$$

where

$$X \sim N\left((\mu_{m'} - X)/\sigma_{m'}, 1\right)$$

$$Y^2 \sim \chi^2_{df = 2\cdot\alpha}$$

$$= 1 - P\left[T_{df = 2\alpha,\ ncp = (\mu_{m'} - X)/s_{m'}} < -\left(\frac{\sigma_s}{\sigma_{m'}}\right)\eta'\right]$$

where

$$T \sim \text{Non-Central Student's } T \ \left(X/\sqrt{Y/df}\right)$$

BPN B'day!

39
2/18/10

Monte Carlo Experiment Design

N_{Rep} ≡ # Replicate samples
$\hat{p}$ ≡ estimated probability that $\hat{\omega} < \eta$

If $\hat{p}(\eta) \approx 10\%$, then

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{N}\right)$$

So CI for $p|\hat{p}$ $= \left\{\hat{p} - 2\sqrt{\hat{p}(1-\hat{p})/N_{Rep}}, \hat{p} + 2\sqrt{\hat{p}(1-\hat{p})/N_{Rep}}\right\}$
$= \left\{\hat{p} - 0.6/\sqrt{N}, \hat{p} + 0.6/\sqrt{N}\right\}$

For

N_{REP}	CI ($\hat{p} = p$)
10^2	0.04 - 0.16
10^4	0.094 - 0.106
10^5	0.098 - 0.102
10^6	0.0994 - 0.1006

It appears that $N_{REP} = 10^6 \Rightarrow$ [One million replicate samples] should provide more than adequate precision

EQUIVALENCE OF GB AND MGB TESTS

$$S_x = \sum_{i=1}^{N} X_{[i]} \; ; \quad SS_x = \sum_{i=2}^{N} X_{[i]}^2$$

$$S_{x_1} = \sum_{i=2}^{N} X_{[i]} = S_x - X_{[1]} \; ; \quad SS_{x_1} = \sum_{i=2}^{N} X_{[i]}^2 = SS_x - X_{[1]}^2$$

$$S^2 \equiv \frac{SS_x - S_x^2/N}{N-1}$$

$$S_1^2 = \frac{SS_{x_1} - S_{x_1}^2/(N-1)}{N-2}$$

$$\omega_1 \equiv \frac{(X_{[1]} - M_1)}{S_1} \; ; \quad M_1 = \frac{S_{x_1}}{N-1}$$

$$\omega = (X_{[1]} - M)/S \; ; \quad M = \frac{S_{x_1}}{N}$$

$$\omega_1 \equiv \frac{X_{[1]} - M_1}{S_1}$$

$$= \frac{X_{[1]} - \left(\frac{NM - X_1}{N-1}\right)}{\sqrt{S_1^2}}$$

$$= \frac{n \omega S}{(n-1)\sqrt{S_1^2}}$$

$$= \frac{nws}{(N-1)\sqrt{s_1^2}}$$

$$= \frac{nws}{(n-1)}\left\{\frac{SS_x - X_1^2 - \left(S_x^2/N - \{S_x^2/N - 2X_1S_x + NX_1^2\}/(N-1)\right)}{N-2}\right\}^{-1/2}$$

$$= \frac{nws}{(n-1)}\left\{\frac{(n-1)s^2 - (X_1 - S_x/N)^2\left(\frac{N}{N-1}\right)}{N-2}\right\}^{-1/2}$$

$$= \frac{nws}{(N-1)\sqrt{\left(\frac{s^2}{N-2}\right)\left(n-1-\left(\frac{n}{n-1}\right)w^2\right)}}$$

$$= \frac{nw}{(n-1)\sqrt{\left(\frac{1}{N-2}\right)\left(n-1-\left(\frac{n}{n-1}\right)w^2\right)}}$$

Conversely

$$w = w_1\sqrt{\frac{(n-1)^3}{n^2(n-2) + n\cdot(n-1)w_1^2}}$$

Derivation of Moments of S - Exact

9/22/10 Assume

$$S^2 \sim \Gamma(\alpha, \beta) \qquad [\text{see p. } 23]$$

$$\alpha \equiv E^2[S^2] / Var[S^2]$$
$$\beta \equiv Var[S^2] / E[S^2]$$

Then

$$E[S] = \frac{\beta^{1/2}\,\Gamma(\alpha+1/2)}{\Gamma(\alpha)}$$

$$Var[S] \equiv E[S^2] - E^2[S]$$

$$= \alpha\beta - \beta \frac{\Gamma^2(\alpha+1/2)}{\Gamma^2(\alpha)}$$

$$= \beta \left\{ \alpha - \left(\Gamma(\alpha+1/2)/\Gamma(\alpha)\right)^2 \right\}$$

This result seems to work better than approximation on p. 23

Wrong

See next page

10/26/16

$$\sigma^2_{\hat{\sigma}^2_{Z,K}} = Var[\hat{\sigma}^2_{Z,K}]$$

$$= E[(\hat{\sigma}^2_{Z,K})^2] - \mu^2_{\hat{\sigma}^2_{Z,K}}$$

$$= \left(\frac{1}{N-K-1}\right)\left(E[(Z-\hat{\mu}_{Z,K})^4 \mid Z>z] - \underbrace{(1+Hz-H^2)^2}_{=(A)}\right)$$

$$= \left(\frac{1}{N-K-1}\right)\Big\{E[Z^4] - 4E[Z^3]\mu_{\hat{\mu}_{Z,K}} + 6E[Z^2]\cdot E[(\hat{\mu}_{Z,K})^2]$$
$$- 4E[Z]\cdot E[(\hat{\mu}_{Z,K})^3] + E[(\hat{\mu}_{Z,K})^4] - (A)\Big\}$$

$$= \left(\frac{1}{N-K-1}\right)\Big\{(3+3Hz+Hz^3) - 4(2H+Hz^2) + 6(1+Hz)(Var[\hat{\mu}] + \mu^2_{\hat{\mu}})$$
$$- 4\mu_{\hat{\mu}_{Z,K}}\cdot(\mu^3_{Z,K} + 3\mu_{Z,K}\cdot Var[\hat{\mu}_{Z,K}]) + E[\hat{\mu}^4_{Z,K}] - (A)\Big\}$$

$$= \left(\frac{1}{N-K-1}\right)\Big\{(3+3Hz+Hz^3) - 4(2H+Hz^2)\mu_z + 6(1+Hz)(\sigma^2_{\hat{\mu}_{Z,K}} + \mu^2_{\hat{\mu}})$$
$$- 4\mu_{\hat{\mu}_{Z,K}}(\mu^3_{Z,K} + 3\mu_{Z,K}\cdot\sigma^2_{\hat{\mu}_{Z,K}}) +$$
$$(E[(\hat{\mu}_{Z,K} - \mu_{\hat{\mu}_{Z,K}})^4] + 6\mu^2_{\hat{\mu}_{Z,K}}\sigma^2_{\hat{\mu}_{Z,K}} + \mu^4_{\hat{\mu}_{Z,K}}) - (A)\Big\}$$

$$= \left(\frac{1}{N-K-1}\right)\Big\{3+3Hz+Hz^2) - 4(2H+Hz^2) + 6(1+Hz)(\sigma^2_{\hat{\mu}_{Z,K}} + \mu^2_{\hat{\mu}_{Z,K}})$$
$$- 4\mu_{\hat{\mu}_{Z,K}}(\mu^3_{Z,K} + 3\mu_{Z,K}\sigma^2_{\hat{\mu}_{Z,K}}) +$$
$$(3\sigma^4_{\hat{\mu}_{Z,K}} + 6\mu^2_{\hat{\mu}_{Z,K}}\sigma^2_{\hat{\mu}_{Z,K}} + \mu^4_{\hat{\mu}_{Z,K}}) - (A)\Big\}$$

10/29/10 Kendall & Stuart Vol. 1 ex. 10.13
p. 341

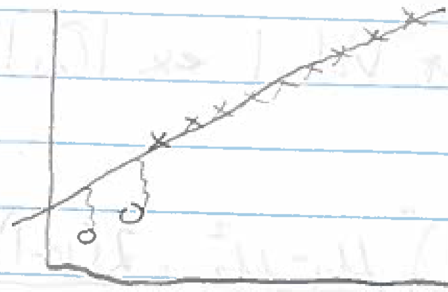
$$\operatorname{var}(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 \frac{\mu_4 - \mu_2^2}{n} + \frac{2(n-1)}{n^3} \mu_2^2$$

$$\mu_k = E[(X - E[X])^k]$$

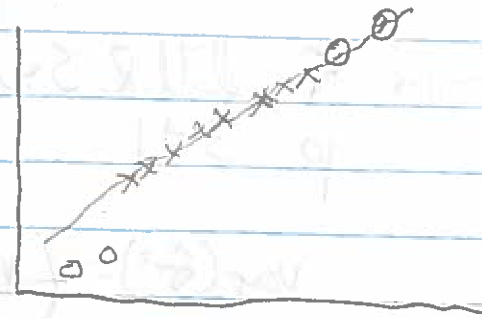
and therefore

$$\operatorname{var}(s^2) = \frac{\mu_4 - \mu_2^2}{n} + \frac{2\mu_2^2}{n(n-1)}$$

MGB



RST



$$M = \left(\frac{1}{N-k}\right)\sum_{i=k+1}^{N} X_{[i]}$$

$$S^2 = \left(\frac{1}{N-k-1}\right)\sum_{i=k+1}^{N} \left(X_{[i]} - M\right)^2$$

$$M = \frac{1}{N-2k}\sum_{i=k+1}^{N-k} X_{[i]}$$

$$S^2 = \frac{1}{N-2k-1}\sum_{i=k+1}^{N-k} \left(X_{[i]} - M\right)^2$$

Q: Does Rosner's [1975] RST test provide a solution comparable to MGB

1) Can only be applied up to 25%-ish low outliers
2) We assume high values are OK. Then they should be included.

Conclusion: MGB is smart modification to RST and GB