Standard Errors and Tests

Leonid Kogan

MIT, Sloan

15.450, Fall 2010

Outline

Outline

Vector Notation

- Suppose θ is a vector. We always think of θ as a column:

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_N \end{pmatrix}, \qquad \theta' = \begin{pmatrix} \theta_1 & \ldots & \theta_N \end{pmatrix}$$

- Partial derivatives of a smooth scalar-valued function $h(\theta)$:

$$\frac{\partial h(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial h(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial h(\theta)}{\partial \theta_N} \end{pmatrix}, \qquad \frac{\partial h(\theta)}{\partial \theta'} \equiv \begin{pmatrix} \frac{\partial h(\theta)}{\partial \theta_1} & \ldots & \frac{\partial h(\theta)}{\partial \theta_N} \end{pmatrix}$$

- If $h(\theta)$ is a vector of functions, $(h_1(\theta), ..., h_M(\theta))'$,

$$\frac{\partial h(\theta)}{\partial \theta'} = \begin{bmatrix} \frac{\partial h_1(\theta)}{\partial \theta_1} & \ldots & \frac{\partial h_1(\theta)}{\partial \theta_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_M(\theta)}{\partial \theta_1} & \ldots & \frac{\partial h_M(\theta)}{\partial \theta_N} \end{bmatrix}$$

Multi-Variate Normal Distribution

- Linear combinations of normal random variables are normally distributed:

$$x \sim \mathcal{N}(0, \Omega) \Rightarrow Ax \sim \mathcal{N}(0, A\Omega A')$$

- The distribution of the sum of squares of n independent $\mathcal{N}(0,1)$ variables is called χ^2 with n degrees of freedom:

$$\varepsilon \sim \mathcal{N}(0, I) \Rightarrow \varepsilon'\varepsilon \sim \chi^2(\dim(\varepsilon))$$

- Distribution of a common quadratic function of a normal vector

$$x \sim \mathcal{N}(0, \Omega) \Rightarrow x'\Omega^{-1}x \sim \chi^2(\dim(x))$$

- Density function of $x \sim \mathcal{N}(\mu, \Omega)$:

$$\phi(x) = \left((2\pi)^N|\Omega|\right)^{-1/2} e^{-\frac{1}{2}(x-\mu)'\Omega^{-1}(x-\mu)}$$

The Delta Method

- Given the estimator $\widehat{\theta}$, want to derive the asymptotic distribution of the vector of smooth functions $h(\widehat{\theta})$.
- Locally, a smooth function is approximately linear:

$$h(\widehat{\theta}) \approx h(\theta_0) + \left.\frac{\partial h(\theta)}{\partial \theta'}\right|_{\theta_0} (\widehat{\theta} - \theta_0)$$

- Let $\widehat{\theta} - \theta_0 \sim \mathcal{N}(0, \Omega)$, $\Omega = \text{Var}(\widehat{\theta})$ *is small* ($\propto 1/T$), then

$$h(\widehat{\theta}) - h(\theta_0) \sim \mathcal{N}\left(0, A\Omega A'\right)$$

$$A = \left.\frac{\partial h(\theta)}{\partial \theta'}\right|_{\theta_0}$$

- In estimation, replace A and Ω with consistent estimates $\widehat{A} = \left.\frac{\partial h(\theta)}{\partial \theta'}\right|_{\widehat{\theta}}$ and $\widehat{\Omega}$:

$$h(\widehat{\theta}) - h(\theta_0) \sim \mathcal{N}\left(0, \widehat{A}\widehat{\Omega}\widehat{A}'\right)$$

Example: Sharpe Ratio Distribution by Delta Method

- Estimate mean and standard deviation of excess returns $(\widehat{\mu}, \widehat{\sigma})$.
- Asymptotic variance-covariance matrix of parameter estimates $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma})'$ is estimated to be $\widehat{\Omega}$.
- Sharpe ratio is estimated to be $\widehat{SR} = h(\widehat{\theta}) \equiv \widehat{\mu}/\widehat{\sigma}$.
- Compute

$$\widehat{A} = \left.\frac{\partial h(\theta)}{\partial \theta'}\right|_{\widehat{\theta}} = \begin{pmatrix} \frac{1}{\widehat{\sigma}} & -\frac{\widehat{\mu}}{\widehat{\sigma}^2} \end{pmatrix}$$

- Variance of the Sharpe ratio estimate is

$$\begin{pmatrix} \frac{1}{\widehat{\sigma}} & -\frac{\widehat{\mu}}{\widehat{\sigma}^2} \end{pmatrix} \left[\widehat{\Omega}\right] \begin{pmatrix} \frac{1}{\widehat{\sigma}} \\ -\frac{\widehat{\mu}}{\widehat{\sigma}^2} \end{pmatrix}$$

Outline

GMM Standard Errors

- Under mild regularity conditions, GMM estimates are consistent: asymptotically, as the sample size T approaches infinity, $\widehat{\theta} \to \theta_0$ (in probability).
- Define

$$\widehat{d} = \left.\frac{\partial \widehat{\mathrm{E}}(f(x_t, \theta))}{\partial \theta'}\right|_{\widehat{\theta}}, \quad \widehat{S} = \widehat{\mathrm{E}}[f(x_t, \widehat{\theta}) f(x_t, \widehat{\theta})']$$

GMM estimates are asymptotically normal:

$$\sqrt{T}(\widehat{\theta} - \theta_0) \Rightarrow \mathcal{N}\left[0, \left(\widehat{d}'\widehat{S}^{-1}\widehat{d}\right)^{-1}\right]$$

- Standard errors are based on the asymptotic var-cov matrix of the estimates,

$$T\mathrm{Var}[\widehat{\theta}] \approx \left(\widehat{d}'\widehat{S}^{-1}\widehat{d}\right)^{-1}$$

Example: Mean and Standard Deviation

- Compute standard errors for estimates of mean and standard deviation

$$f_1(x_t,\theta) = x_t - \mu, \qquad f_2(x_t,\theta) = (x_t-\mu)^2 - \sigma^2$$

$$\widehat{d} = \left.\frac{\partial \widehat{\mathrm{E}}(f(x_t,\theta))}{\partial \theta'}\right|_{\widehat{\theta}} = \begin{bmatrix} -1 & 0 \\ -2(\widehat{\mathrm{E}}(x_t)-\widehat{\mu}) & -2\widehat{\sigma} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -2\widehat{\sigma} \end{bmatrix}$$

$$\widehat{S} = \widehat{\mathrm{E}}[f(x_t,\widehat{\theta})f(x_t,\widehat{\theta})'] = \widehat{\mathrm{E}}\begin{bmatrix} f_1^2 & f_1 f_2 \\ f_1 f_2 & f_2^2 \end{bmatrix}$$

$$\widehat{\theta} - \theta_0 \sim \mathcal{N}(0, \frac{1}{T}\widehat{V}), \qquad \widehat{V} = \left(\widehat{d}'\widehat{S}^{-1}\widehat{d}\right)^{-1}$$

Mean and Standard Deviation, Gaussian Distribution

- Recall that for Gaussian distribution, $\mathrm{E}[(x-\mu_0)^3] = 0$, $\mathrm{E}[(x-\mu_0)^4] = 3\sigma_0^4$.
- Using LLN,

$$\text{plim}_{T\to\infty}\, \widehat{d} = d \equiv \begin{bmatrix} -1 & 0 \\ 0 & -2\sigma_0 \end{bmatrix}$$

$$\text{plim}_{T\to\infty}\, \widehat{S} = S \equiv \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & 2\sigma_0^4 \end{bmatrix}$$

$$\widehat{\theta} - \theta_0 \sim \mathcal{N}(0, \frac{1}{T}\widehat{V})$$

$$\text{plim}_{T\to\infty}\, \widehat{V} = \left(d' S^{-1} d\right)^{-1} = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \frac{1}{2}\sigma_0^2 \end{bmatrix}$$

Example: Mean and Standard Deviation, Gaussian Distribution

MATLAB ® Code

```
T = 100; mu = 0.1; sigma = 0.2;
x = mu + sigma*randn(1,T); % Simulated sample

mu_hat = (1/T) * sum(x); % GMM parameter estimates
sigma_hat = sqrt((1/T) * sum((x - mu_hat).^2));

f = [x - mu_hat; (x - mu_hat).^2 - sigma_hat^2];
d = [-1  0;  0  -2*sigma_hat];
S = (1/T) * (f * f');
V = inv(d' * inv(S) * d);

SE_mu = sqrt(1/T * V(1,1)); % Compute standard errors
SE_sigma = sqrt(1/T * V(2,2));
```

Example: Mean and Standard Deviation, Gaussian Distribution

MATLAB® Output

```
mu =      0.1000    sigma     = 0.2000
mu_hat = 0.1196    sigma_hat = 0.1936
SE_mu =  0.0194    SE_sigma  = 0.0142
```

Example: Mean and Standard Deviation, Gaussian Distribution

- 95% Confidence intervals for parameter estimates can be constructed as

$$\widehat{CI}(\theta_i) = [\widehat{\theta}_i - 1.96 \times SE(\widehat{\theta}_i), \widehat{\theta}_i + 1.96 \times SE(\widehat{\theta}_i)], \quad i = 1, 2$$

  Asymptotically, these should contain the true values with 95% probability.
- How good are the CI's in a finite sample?
- Perform a Monte Carlo experiment: simulate N independent artificial samples and compute the coverage frequency.

Example: Mean and Standard Deviation, Gaussian Distribution

MATLAB® Code

```
coverage = zeros(2,N);
for n=1:N
  x = mu + sigma*randn(1,T); % Simulated sample

  [mu_hat, sigma_hat, SE_mu, SE_sigma] = GMMGaussian(x);

  coverage(1,n) = (abs(mu_hat - mu) < 1.96*SE_mu);
  coverage(2,n) = (abs(sigma_hat - sigma) < 1.96*SE_sigma);
end
y = mean(coverage');
```

100,000 simulations: coverage frequencies are (0.945, 0.929).

Example: Sharpe Ratio Distribution by Delta Method

Gaussian distribution

- Asymptotic variance-covariance matrix of the parameter estimates $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma})'$ is

$$\widehat{\Omega} = \frac{1}{T} \begin{bmatrix} \widehat{\sigma}^2 & 0 \\ 0 & \frac{1}{2}\widehat{\sigma}^2 \end{bmatrix}$$

- Asymptotic variance of the Sharpe ratio is

$$\begin{pmatrix} \frac{1}{\widehat{\sigma}} & -\frac{\widehat{\mu}}{\widehat{\sigma}^2} \end{pmatrix} \left[\widehat{\Omega}\right] \begin{pmatrix} \frac{1}{\widehat{\sigma}} \\ -\frac{\widehat{\mu}}{\widehat{\sigma}^2} \end{pmatrix} = \frac{1}{T}\left(1 + \frac{1}{2}\widehat{SR}^2\right)$$

Outline

Ordinary Least Squares (OLS) and GMM

- Consider a linear model

$$y_t = x_t'\beta + u_t$$

- OLS is based on the assumption that the residuals have zero mean conditionally on the explanatory variables and each other:

$$\mathrm{E}[u_t | x_t, x_{t-1}, ..., u_{t-1}, u_{t-2}, ...] = 0$$

- If we define

$$f(x_t, y_t, \beta) = x_t\,(y_t - x_t'\beta)$$

  then β can be estimated using GMM:

$$\mathrm{E}[x_t\,(y_t - x_t'\beta)] = \mathrm{E}[x_t u_t] \overset{\text{Iterated Expectations}}{=} \mathrm{E}[x_t\,\mathrm{E}[u_t | x_t]] = 0$$

Ordinary Least Squares (OLS) and GMM

- GMM estimate is based on

$$\widehat{\mathrm{E}}[x_t\,(y_t - x_t'\beta)] = 0 \quad \Longrightarrow \quad \widehat{\beta} = \widehat{\mathrm{E}}(x_t x_t')^{-1}\widehat{\mathrm{E}}(x_t y_t)$$

  which is the standard OLS estimate.

- To find standard errors, compute

$$\widehat{S} = \widehat{\mathrm{E}}(f_t f_t') = \widehat{\mathrm{E}}(\widehat{u}_t^2 x_t x_t'), \qquad \widehat{u}_t \equiv y_t - x_t'\widehat{\beta}$$

$$\widehat{d} = \frac{\partial \widehat{\mathrm{E}}[f]}{\partial \beta'} = -\widehat{\mathrm{E}}(x_t x_t')$$

  Then

$$\mathrm{Var}[\widehat{\theta}] = \frac{1}{T}\left(\widehat{d}'\widehat{S}^{-1}\widehat{d}\right)^{-1} = \frac{1}{T}\widehat{\mathrm{E}}(x_t x_t')^{-1}\widehat{\mathrm{E}}(\widehat{u}_t^2 x_t x_t')\widehat{\mathrm{E}}(x_t x_t')^{-1}$$

Outline

Standard Errors: Correlated Observations

- When $f(x_t, \theta)$ are correlated over time, formulas for standard errors must be adjusted to account for autocorrelation.
- Correlated observations affect the effective sample size.
- The relation

$$\mathrm{Var}[\widehat{\theta}] = \frac{1}{T}\left(\widehat{d}^{-1}\widehat{S}\left(\widehat{d}'\right)^{-1}\right) = \frac{1}{T}\left(\widehat{d}\widehat{S}^{-1}\widehat{d}'\right)^{-1}$$

  is still valid. But need to modify the estimate $\widehat{S}$.
- In an infinite sample,

$$S = \sum_{j=-\infty}^{\infty} \mathrm{E}\left[f(x_t, \theta_0) f(x_{t-j}, \theta_0)'\right]$$

Estimating S: Newey-West

- Newey-West procedure for computing standard errors prescribes

$$\widehat{S} = \sum_{j=-k}^{k} \frac{k-|j|}{k} \frac{1}{T} \sum_{t=1}^{T} f(x_t, \widehat{\theta}) f(x_{t-j}, \widehat{\theta})' \qquad \text{(Drop out-of-range terms)}$$

- k is the band width parameter. The larger the sample size, the larger the k one should use. Suggested growth rate is $k \propto T^{1/3}$.
- In a finite sample, need k to be small compared to T, but large enough to cover the intertemporal dependence range.
- Consider several values of k and compare the results.

OLS Standard Errors With Correlated Residuals

- Linear model

$$y_t = x_t' \beta + u_t$$

- Assume that

$$\mathrm{E}[u_t | x_t, x_{t-1}, ...] = 0$$

but allow u_t to be autocorrelated.

- Since $f(x_t, \theta) = x_t u_t$, Newey-West estimate of $\widehat{S}$ is

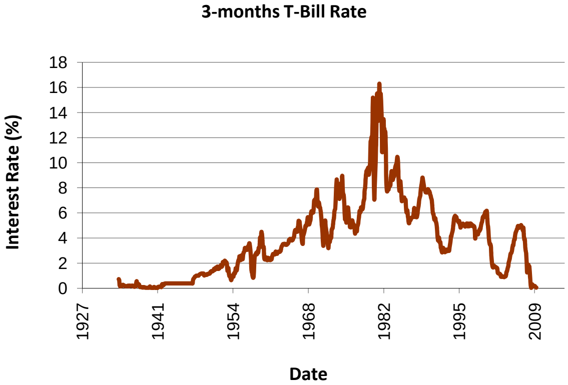
$$\widehat{S} = \sum_{j=-k}^{k} \frac{k - |j|}{k} \frac{1}{T} \sum_{t=1}^{T} \left[u_t x_t x_{t-j}' u_{t-j} \right] \qquad \text{(Drop out-of-range terms)}$$

- Asymptotic var-cov matrix of the regression coefficients:

$$\mathrm{Var}[\widehat{\theta}] = \frac{1}{T} \widehat{\mathrm{E}}(x_t x_t')^{-1} \, \widehat{S} \, \widehat{\mathrm{E}}(x_t x_t')^{-1}$$

Example: Estimating Average Interest Rate

- We want to estimate the average 3-months TBill rate using historical data.
- 3-Month Treasury Bill secondary market rate, monthly observations.



Source: Federal Reserve Bank of St. Louis.

Example: Long-Horizon Return Predictability

- Predict S&P 500 returns using the log of the dividend-price ratio (1934/01 – 2008/12)

$$r_{t\to t+h} = \alpha + \beta \ln\left(\frac{D}{P}\right)_{t-1} + u_{t+h}$$

- Returns are cumulative over 6 or 12 months. Sum of monthly returns.

h	β	Standard Error				
		$k=0$	$k=5$	$k=12$	$k=24$	$k=36$
6	0.0530	0.0089	0.0185	0.0215	0.0233	0.0232
12	0.1067	0.0129	0.0297	0.0378	0.0428	0.0431

Discussion

- Classical OLS is based on very restrictive assumptions.
- In practice, the RHS variables are stochastic, and not uncorrelated with lagged residuals.
- GMM provides a powerful framework for dealing with regressions: OLS is valid as long as the moment conditions are valid.
- Important to treat standard errors correctly. GMM offers a general recipe.

Outline

MLE and GMM

- MLE or QMLE can be related to GMM.
- Optimality conditions for maximizing $\mathcal{L}(\theta) = \sum_{t=1}^{T} \ln p(x_t|\text{past } x; \theta)$ are

$$\sum_{t=1}^{T} \frac{\partial \ln p(x_t|\text{past } x; \theta)}{\partial \theta} = 0$$

- If we set $f(x_t, \theta) = \partial \ln p(x_t|\text{past } x; \theta)/\partial \theta$ (the *score vector*), then MLE is "GMM" with the moment vector f.
- Scores are uncorrelated over time because $\mathrm{E}_t[f(x_{t+1}, \theta_0)] = 0$ (Campbell, Lo, MacKinlay, 1997, Appendix A.4). Standard errors using GMM formulas:

$$\widehat{d} = \widehat{\mathrm{E}}\left[\frac{\partial^2 \ln p(x_t|\text{past } x; \theta)}{\partial \widehat{\theta} \partial \widehat{\theta}'}\right], \quad \widehat{S} = \widehat{\mathrm{E}}\left[\frac{\partial \ln p(x_t|\text{past } x; \theta)}{\partial \widehat{\theta}} \frac{\partial \ln p(x_t|\text{past } x; \theta)}{\partial \widehat{\theta}'}\right]$$

$$T\mathrm{Var}[\widehat{\theta}] = \left(\widehat{d}'\widehat{S}^{-1}\widehat{d}\right)^{-1}$$

Nonlinear Least Squares (NLS)

- Consider a nonlinear model

$$y_t = h(x_t, \beta) + u_t, \qquad \mathrm{E}[u_t|x_t] = 0$$

- We use QMLE to estimate this model. Pretend that errors u_t are IID $\mathcal{N}(0, \sigma^2)$.
- Minimize log-likelihood

$$\mathcal{L}(\beta) = \sum_{t=1}^{T} -\ln\sqrt{2\pi\sigma^2} - \frac{(y_t - h(x_t, \beta))^2}{2\sigma^2}$$

- First-order conditions can be viewed as moment conditions in GMM:

$$\widehat{\beta} = \arg\min_{\beta} \mathrm{E}\left[(y_t - h(x_t, \beta))^2\right] \Rightarrow \mathrm{E}\left[\frac{\partial h(x_t, \beta)}{\partial \beta}(y_t - h(x_t, \beta))\right] = 0$$

- Nonlinear Least Squares. Can use GMM formulas for standard errors.
- Why not choose other moments, e.g., $f = g(x_t)(y_t - h(x_t, \beta))$ with pretty much arbitrary $g(x_t)$, e.g., $g(x_t) = x_t$?
- We could. But this may result in less precise estimates of $\widehat{\beta}$ or invalid moment conditions. In fact, if u_t are Gaussian, NLS is optimal (see MLE).

Outline

Hypothesis Tests

- Sample of independent observations $x_1, ..., x_T$ with distribution $p(x, \theta_0)$.
- Want to test the null hypothesis H_0, which is a set of restrictions on the parameter vector θ_0, e.g., $b'\theta_0 = 0$.
- Statistical test is a decision rule, rejecting the null if some conditions are satisfied by the sample, i.e.,

$$\text{Reject if } \ (x_1, ..., x_T) \in \mathbb{A}$$

- *Test size* is the upper bound on the probability of rejecting the null hypothesis over all cases in which the null hypothesis is correct.
- Type I error is false rejection of the null H_0. Test size is the maximum probability of false rejection.

χ^2 Test

- Want to test the Null Hypothesis regarding model parameters:

$$h(\theta) = 0$$

- Construct a χ^2 test:
  - Estimate the var-cov of $h(\widehat{\theta})$, $\widehat{V}$.
  - Construct the test statistic

$$\xi = h(\widehat{\theta})' \widehat{V}^{-1} h(\widehat{\theta}) \sim \chi^2(\dim h(\widehat{\theta}))$$

- Reject the Null if the test statistic ξ is sufficiently large. Rejection threshold is determined by the desired test size and the distribution of ξ under the Null.

Example: OLS

- Suppose we run a predictive regression of y_t on a vector of predictors x_t:

$$y_t = \beta_0 + x_t' \beta + u_t$$

- Compute parameter estimates $\widehat{\beta}$ by OLS. Use Newey-West to obtain var-cov matrix for $\widehat{\beta}$, $\widehat{\text{Var}}(\widehat{\beta})$.
- Test the Null of no predictability: $\beta = 0$.
- Test statistic is

$$\xi = \widehat{\beta}' \left[\text{Var}(\widehat{\beta}) \right]^{-1} \widehat{\beta} \sim \chi^2(\dim(\beta))$$

- Test of size α: reject the Null if $\xi \geqslant \overline{\xi}$, where

$$\text{CDF}_{\chi^2(\dim(\beta))}(\overline{\xi}) = 1 - \alpha$$

Testing the Sharpe Ratio

- Suppose we are given a time series of excess returns.
- We want to test whether the Sharpe ratio of returns is equal to SR_0.
- Two steps:
  - Using the *delta method*, derive the asymptotic variance of the Sharpe ratio estimate, $\widehat{SR} = \widehat{\mu}/\widehat{\sigma}$.
  - Test statistic

$$\frac{(\widehat{SR} - SR_0)^2}{\text{Var}(\widehat{SR})} \sim \chi^2(1)$$

Example: Sharpe Ratio Comparison

- Suppose we observe two series of excess returns, generated over the same period of time by two trading strategies:

$$(x_1^1, x_2^1, \ldots, x_T^1) \text{ and } (x_1^2, x_2^2, \ldots, x_T^2)$$

- We do not know the exact distribution behind each strategy, but we do know that these returns are IID over time.
- Contemporaneously, x_t^1 and x_t^2 may be correlated.
- We want to test the *null hypothesis* that these two strategies have the same Sharpe ratio.

Example: Sharpe Ratio Comparison

- Stack together the two return series to create a new observation vector

$$x_t = (x_t^1, x_t^2)'$$

- The parameter vector is

$$\theta_0 = (\mu_1^0, \sigma_1^0, \mu_2^0, \sigma_2^0)$$

- The null hypothesis is

$$H_0 : \left\{ \frac{\mu_1^0}{\sigma_1^0} - \frac{\mu_2^0}{\sigma_2^0} = 0 \right\}$$

- To construct the rejection region for H_0, estimate the asymptotic distribution of $\frac{\widehat{\mu}_1}{\widehat{\sigma}_1} - \frac{\widehat{\mu}_2}{\widehat{\sigma}_2}$.

Example: Sharpe Ratio Comparison

- Using standard GMM formulas, estimate the asymptotic variance-covariance matrix of the parameter estimates $\widehat{\theta}$, $\widehat{\Omega}$.
- Define

$$h(\theta) = \frac{\mu_1}{\sigma_1} - \frac{\mu_2}{\sigma_2}$$

- Compute

$$\widehat{A} = \left. \frac{\partial h(\theta)}{\partial \theta'} \right|_{\widehat{\theta}} = \begin{pmatrix} \frac{1}{\widehat{\sigma}_1} & -\frac{\widehat{\mu}_1}{(\widehat{\sigma}_1)^2} & -\frac{1}{\widehat{\sigma}_2} & \frac{\widehat{\mu}_2}{(\widehat{\sigma}_2)^2} \end{pmatrix}$$

- Asymptotically, variance of $h(\widehat{\theta})$ is

$$\widehat{\mathrm{Var}}\left[h(\widehat{\theta})\right] = \begin{pmatrix} \frac{1}{\widehat{\sigma}_1} & -\frac{\widehat{\mu}_1}{(\widehat{\sigma}_1)^2} & -\frac{1}{\widehat{\sigma}_2} & \frac{\widehat{\mu}_2}{(\widehat{\sigma}_2)^2} \end{pmatrix} \left[\widehat{\Omega}\right] \begin{pmatrix} \frac{1}{\widehat{\sigma}_1} \\ -\frac{\widehat{\mu}_1}{(\widehat{\sigma}_1)^2} \\ -\frac{1}{\widehat{\sigma}_2} \\ \frac{\widehat{\mu}_2}{(\widehat{\sigma}_2)^2} \end{pmatrix}$$

Example: Sharpe Ratio Comparison

- Under the null hypothesis, $h(\theta_0) = 0$, and therefore

$$\frac{h(\widehat{\theta})}{\sqrt{\widehat{\text{Var}}\left[h(\widehat{\theta})\right]}} = \frac{h(\widehat{\theta}) - h(\theta_0)}{\sqrt{\widehat{\text{Var}}\left[h(\widehat{\theta})\right]}} \sim \mathcal{N}(0, 1)$$

- Define the rejection region for the test of the null $h(\theta_0) = 0$ as

$$\mathbb{A} = \left\{ \left| \frac{h(\widehat{\theta})}{\sqrt{\widehat{\text{Var}}\left[h(\widehat{\theta})\right]}} \right| \geqslant z \right\}$$

- A 5% test is obtained by setting $z = 1.96 = \Phi^{-1}(0.975)$, where Φ is the Standard Normal CDF.

Key Points

- Delta method.
- GMM standard errors, MLE and QMLE standard errors.
- OLS standard errors with correlated observations.
- χ^2 test.
- Testing restrictions on OLS coefficients, nonlinear restrictions.

Readings

- Cochrane, 2005, Sections 11.1, 11.3-4, 11.7, 20.1.
- Campbell, Lo, MacKinlay, 1997, Sections A.2-4.
- Cochrane, "New facts in finance.":

  http://faculty.chicagobooth.edu/john.cochrane/research/Papers/ep3Q99_3.pdf

Appendix: Intuition for GMM Standard Errors

- Consider IID observations $x_1, ..., x_T$.
- Delta method computes the var-cov of $\widehat{\mathrm{E}}[f(x_t, \widehat{\theta})]$, given the variance of $\widehat{\theta}$. By going in reverse direction, we compute the var-cov of $\widehat{\theta}$ starting from the var-cov of $\widehat{\mathrm{E}}[f(x_t, \widehat{\theta})]$.
- The latter is estimated as

$$\widehat{\mathrm{Var}}[\widehat{\mathrm{E}}(f(x_t, \widehat{\theta}))] \overset{(1)}{=} \frac{1}{T}\widehat{\mathrm{Var}}[f(x_t, \widehat{\theta})] \overset{(2)}{=} \frac{1}{T}\widehat{\mathrm{E}}[f(x_t, \widehat{\theta})\, f(x_t, \widehat{\theta})'] \equiv \frac{1}{T}\widehat{S}$$

(1) IID observations, so $\mathrm{Var}(\sum \cdot) = \sum \mathrm{Var}(\cdot)$; (2) Use $\widehat{\mathrm{E}}[f(x_t, \widehat{\theta})] = 0$

- Using the delta method on the LHS, with $\widehat{A} = \widehat{d} = \partial \widehat{\mathrm{E}}[f(x_t, \widehat{\theta})]/\partial\widehat{\theta}'$,

$$\frac{1}{T}\widehat{S} \approx \widehat{d}\, \mathrm{Var}[\widehat{\theta}]\, \widehat{d}'$$

and therefore

$$\mathrm{Var}[\widehat{\theta}] \approx \frac{1}{T}\left(\widehat{d}^{-1}\widehat{S}\left(\widehat{d}'\right)^{-1}\right) = \frac{1}{T}\left(\widehat{d}\widehat{S}^{-1}\widehat{d}'\right)^{-1}$$

15.450 Analytics of Finance

Fall 2010