Testing Nonlinear Hypotheses.

$$C_t = \alpha + \beta Y_t + \gamma C_{t-1} + u_t$$

Long Run MPC

$$\delta = \frac{\beta}{1-\gamma}$$

$H_0: \delta = 1$

$H_A: \delta \neq 1$

Use what's called the Delta Method to get the variance of a nonlinear function of LS estimator.

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + u_i$$

Let $g(\underset{\sim}{\beta})$ be a continuous function of $\underset{\sim}{\beta}$.

$H_0: g(\beta) = 0$

$H_A \; g(\beta) \neq 0$

First order Taylor series expansion

$$g(\hat{\beta}) \approx g(\beta_0) + \frac{\partial g(\beta)}{\partial \underset{\sim}{\beta}^T} (\hat{\beta} - \beta_0)$$

$\text{plim}(\hat{\beta}) = \beta$ and By Slutsky Theorem

$g(\hat{\beta}) \xrightarrow{p} g(\beta)$ and $\left.\frac{\partial g(\beta)}{\partial \underset{\sim}{\beta}^T}\right|_{\hat{\beta}} \xrightarrow{p} \frac{\partial g(\beta)}{\partial \beta^T}$

$$\text{Var}[g(\hat{\beta})] = \frac{\partial g(\beta)}{\partial \beta^T} \text{Var}(\hat{\beta}) \frac{\partial g(\beta)}{\partial \underset{\sim}{\beta}}$$

By "Δ" Theorem.

Estimate $\text{Var}(\hat{\beta})$ in usual way and subst $\hat{\beta}$ into $\frac{\partial g}{\partial \beta^T}$ And you are <u>set</u>.

Testing nonlinear Hypotheses

$$C_t = \alpha + \beta Y_t + \gamma C_{t-1} + u_t$$

SR MPC $= \beta$

LR " $= \beta / 1-\gamma = \delta$

$H_0: \delta = 1$

$H_A \ \delta \neq 1$

$$\hat{C}_t = -7.69 + .40014\, Y_t + .3800\, C_{t-1}$$

$$\hat{\delta} = \frac{.400147}{1-.3800} = .6461$$

$$\hat{G} = \begin{array}{lll} \frac{\partial \delta}{\partial \alpha} = & 0 & = 0 \\ \frac{\partial \delta}{\partial \beta} = & \frac{1}{1-\gamma} & = 1.6149 \\ \frac{\partial \delta}{\partial \gamma} = & \frac{\hat{\beta}}{(1-\hat{\gamma})^2} & = 1.0434 \end{array}$$

$$Var(\hat{\delta}) = \hat{G}^T \hat{\sigma}^2 (X^T X)^{-1} \hat{G}$$

G - Theorem

Suppose $\sqrt{n}\,(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 S_{xx}^{-1})$

$(0, \cancel{V\hat{a}r(\hat{\beta})})$

$$S_{xx}^{-1} = \underset{n \to \infty}{\text{plim}} \frac{X^T X}{n}$$

and let $g(\beta)$ be a $\overset{l \times 1}{}$ continuously differential function of β. ~~and that~~ ~~β is in~~ and

$$\underset{l \times k}{G} = \frac{\partial g(\beta)}{\partial \beta^T} \quad \begin{matrix} - l \times 1 \\ - 1 \times k \end{matrix}$$

$$\sqrt{n}\,(g(\hat{\beta}) - g(\beta)) \xrightarrow{d} N(0, \sigma^2 G S_{xx}^{-1} G^T)$$

AND this can be used as a basis for hypothesis tests.

Example

CES Production

$$\ln(y) = \ln\gamma - \frac{v}{\rho}\ln\left[\delta K^{-\rho} + (1-\delta)L^{-\rho}\right] + \varepsilon.$$

A Taylor's series expansion around $\rho = 0$ yields

$$\ln(y) = \ln\gamma + v\delta\ln(K) + v(1-\delta)\ln L$$
$$+ \rho v\delta(1-\delta)\left\{-\tfrac{1}{2}(\ln K - \ln L)^2\right\} + \varepsilon'$$

$$\ln(y) = \beta_1 + \beta_2\ln(K) + \beta_3\ln(L) + \beta_4\left(-\tfrac{1}{2}\ln^2(K/L)\right) + u.$$

which can be estimated by LS.

$$\theta = \begin{pmatrix}\gamma\\ \delta\\ v\\ \rho\end{pmatrix} = \cancel{g(\beta)}$$ AND your interest lies in estimating these easily.

$\beta_1 = \ln\gamma$

$\beta_2 = v\delta$

$\beta_3 = v(1-\delta)$

$\beta_4 = \rho v\delta(1-\delta)$

$$\gamma = e^{\beta_1}$$

$$\delta = \beta_2/(\beta_2+\beta_3)$$

$$\nu = \beta_2 + \beta_3$$

$$\rho = \beta_4(\beta_2+\beta_3)/\beta_2\cdot\beta_3$$

$$\mathrm{Cov}(\theta) = G \cdot \mathrm{Var}(\hat{\beta})\, G^T$$

$$\underset{4\times 4}{G} = \begin{array}{ccccc} & \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ \partial\gamma/\partial\beta_1 & e^{\beta_1} & 0 & 0 & 0 \\ \partial\delta/\partial\beta_1 & 0 & & & \\ \partial\nu/\partial\beta_1 & 0 & & & \\ \partial\rho/\partial\beta_1 & 0 & & & \end{array}$$

$$G = \begin{array}{c} \partial\gamma/\partial\beta^T \\ \partial\delta/\partial\beta^T \\ \partial\nu/\partial\beta^T \\ \partial\rho/\partial\beta^T \end{array}$$

$$G = \begin{bmatrix} e^{\beta_1} & 0 & 0 & 0 \\ 0 & \dfrac{\beta_3}{(\beta_2+\beta_3)^2} & -\dfrac{\beta_2}{(\beta_2+\beta_3)^2} & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -\beta_3\beta_4/(\beta_2^2\beta_3) & -\dfrac{\beta_2\beta_4}{(\beta_2\beta_3^2)} & (\beta_2+\beta_3)/\beta_2\beta_3 \end{bmatrix}$$

$\hat{G} = G$ replace β_i with $\hat{\beta}_i$

$$\mathrm{Cov}[g(\hat{\beta})] = \hat{G}\,\hat{\sigma}^2 (X^T X)^{-1} \hat{G}^T$$

$$\mathrm{Cov}(\hat{\theta}) = \hat{G}\,\hat{\sigma}^2 (X^T X)^{-1} \hat{G}^T$$

Nonlinear Regression

A linear regression model consists of all DGPs for which the expectation of y_t conditional on Ω_t (INFO set) can be expressed as a linear combination of $X_t'\beta$.

- X_t' can be nonlinear, but model must be linear in params.

$$y_t = \beta_1 + \beta_2 X_t + \beta_3 X_t^2 + u_t$$

$$\ln(y_t) = \beta_1 + \beta_2 X_t + u_t$$

$$y_t = \beta_1 + \beta_2 \ln(X_t) + u_t$$

$$\ln(y) = \beta_1 + \ln(X) + u.$$

$$\ln(y) = \beta_1 + \beta_2 \frac{1}{X_t} + u_t$$

etc.

All of these satisfy this. There are times when this is not enough to capture the DGP.

$$y_t = \gamma_t(\beta) + u_t \qquad u_t \text{ iid } (0, \sigma^2)$$
$$t = 1, \dots, n$$

$\underset{\sim}{\beta}$ is still $k \times 1$

γ_t is a nonlinear function that depends on $\underset{\sim}{\beta}$.

There may be more or fewer than k explanatory vars.

Example:

Linear Regression with AR(1) errors.

1st

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

$$u_t = \rho u_{t-1} + \gamma \varepsilon_t$$

$$\varepsilon_t \text{ iid } (0, \sigma_\varepsilon^2) \quad |\rho| < 1$$

ε_t are homoscedastic and not correlated with other innovations.

u_t are correlated. 1st order $\Rightarrow$ u_t is a function of $u(t-1)$.

$|\rho| < 1$ is a stability condition.

SUBST.

$$y_t = x_t'\beta + (\rho u_{t-1} + \varepsilon_t)$$

$$u_{t-1} = y_{t-1} - x_{t-1}'\beta$$

$$y_t = x_t'\beta + \rho y_{t-1} - \rho x_{t-1}'\beta + \varepsilon_t$$

$$= \rho y_{t-1} + (x_t' - \rho x_{t-1}')\beta + \varepsilon_t$$

This is nonlinear and dynamic.

MOM

For linear Model:

$$E(x_t u_t) = E(x_t(y_t - x_t'\beta)) = 0$$

$$\Longrightarrow \frac{1}{n} X^T(y - X\beta) = 0$$

moment condition for linear Model.

For nonlinear Model

let W_t denote $1 \times k$ vector

in the Ω_t (information set)

$$E(W_t u_t) = E\left(W_t(y_t - x_t(\beta))\right) = 0$$

Mom $\quad W^T(y - x(\beta)) = 0$

$W_{n \times k}$ with rows W_t $t = 1, \cdots, n$

$\underset{\sim}{x}(\beta)$ has t^{th} row $x_t^T(\beta)$

$$W^T(\underset{\sim}{y} - \underset{\sim}{x}(\beta)) = 0$$

k nonlinear eq:s in k unknowns.

How to pick W?

W can be just about anything as long as the t^{th} row belongs to Ω_t.

and ~~ve~~ Rank(W) = $k \le n$

Some choices of W are more efficient than others, though

NLS

In NLS, let $\underset{k\times 1}{W_t} \equiv X_t^T(\beta) = \frac{\partial x_t(\beta)}{\partial \beta^T}$

so that

$$\underset{k\times n}{X^T(\underset{\sim}{\beta})} \underset{n\times 1}{(y - x(\beta))} = 0$$

These moment conditions can be obtained using NLS

$$Q(\beta) \equiv SSR(\beta) = \sum_{i=1}^{n} u_i^2 = (y - x(\beta))^T (y - x(\beta))$$

And minimization w/resp. to β yields

$$X^T(\beta)(y - x(\beta)) = 0$$

1) $X^T(\beta)$ is <u>not</u> a matrix of const like W. It depends on β.

2) Eq's are nonlinear in β. No closed form soln.

Assuming a solution exists, the NLS estimator is

1) consistent
2) Asy normal.
3) Asy efficient

Consistency $E(\psi_t(\beta) u_t) = 0$

$E\left[X_t(\beta)\, u_t\right] = 0$

columns of the matrix of derivatives must be orthogonal to u_t.

All of this assumes that the parameters are identified by a given data set and estimator.

$\Rightarrow$ a unique value of β will be generated from the data set with the estimator..

Asympt identification requires this in the limit as $n \to \infty$

Computation

Since there is no closed form solution for $\hat{\beta}$ in terms of y & x we have to use numerical or iterative means of solving the nonlinear normal eqs.

Iterative procedure.

(1) start with an initial guess for β.

(2) computes a step that tries to improve β.

(3) keeps iterating until no further improvement can be found.

Algorithms differ by

(1) how the direction of search is chosen

(2) how big the steps are.

(3) stopping rule.

Newton's method

Based on a 2nd order Taylor's Expansion of $Q(\beta)$ around an initial guess β_0

$$Q^*(\beta) = Q(\beta_0) + g_0^T(\beta - \beta_0)$$
$$+ \frac{1}{2}(\beta - \beta_0)^T H_0 (\beta - \beta_0)$$

$\underset{n \times 1}{g(\underset{\sim}{\beta})}$ is gradient $\qquad = \frac{\partial^2 Q(\beta)}{\partial\beta\,\partial\beta^T} \; n \times n.$

$$g(\beta) = \frac{\partial Q(\beta)}{\partial \underset{\sim}{\beta}} \qquad H = \frac{\partial g(\beta)}{\partial \beta^T} \quad \text{Hessian.}$$

Minimize $Q^*(\beta)$ with respect to $\underset{\sim}{\beta}$.

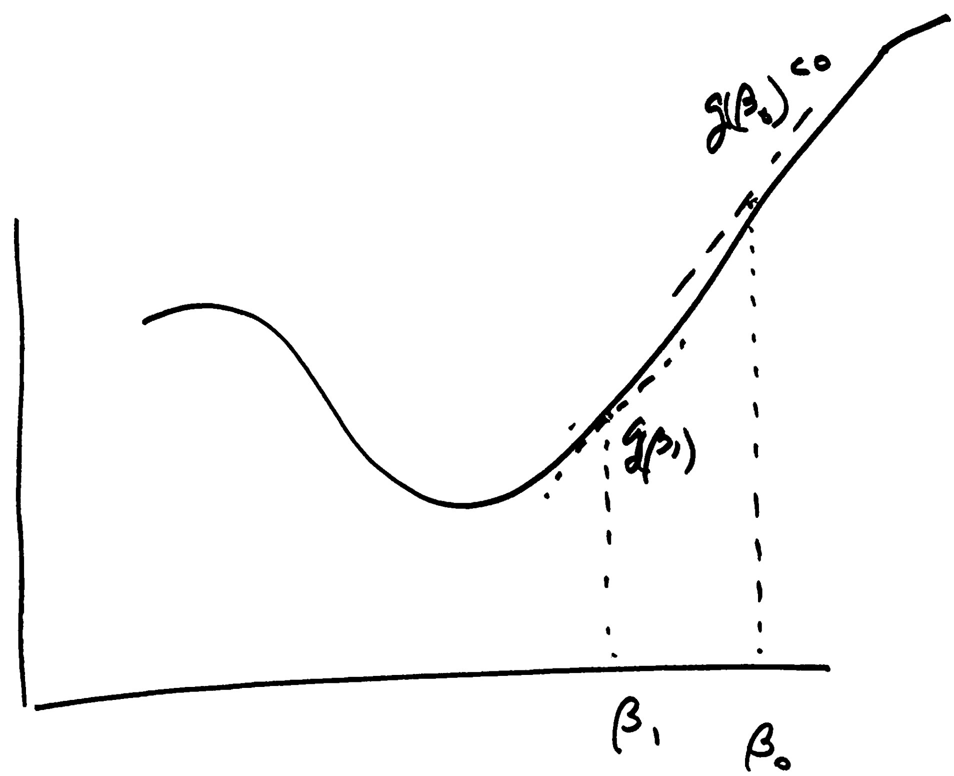
$$g_0 + H_0(\beta_{(1)} - \beta_{(0)}) = 0$$
$$g_0 + H_0\beta_{(1)} - H_0\beta_0 = 0$$
$$H_0\beta_{(1)} = H_0\beta_0 - g_0$$
$$\beta_{(1)} = \beta_0 - H_0^{-1}g_0$$

Repeat

$$\beta_{m+1} = \beta_m - H_m^{-1} g_m$$

until convergence.

The key here is that H must be pos def.



As long as H pos def, the derivative of function heads us in the correct direction. If H_0 not pos. def.; Problem!

Quasi Newton methods

$$\beta_{n+1} = \beta_n - \alpha_n D_n^{-1} g_n$$

(1) α_n is step length.
(2) D_n is pos def matrix
(3) g_n is gradient.

These methods use techniques that improve selection of α_n and chose D_n to always (or nearly so) be pos def.

In Newtons method
$\alpha = 1$
$D_n = H_n$

Stopping Rules

$$\left| Q(\beta_{n+1}) - Q(\beta_n) \right| < \varepsilon$$

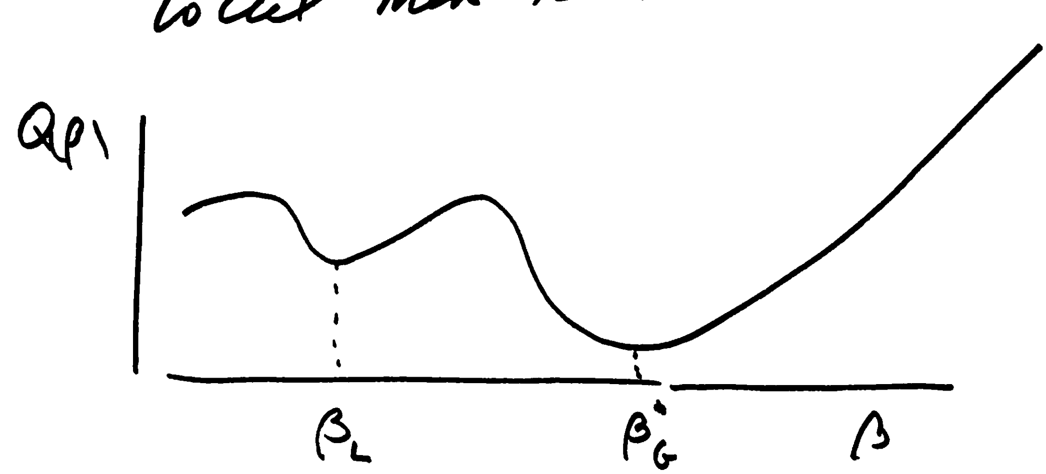
$$\left| \beta_{n+1} - \beta_n \right| < \varepsilon$$

$$g_n^T D_n^{-1} g_n < \varepsilon$$

$\therefore$ choose ε "small"

Local vs Global Minima

Most functions are NOT convex, $\therefore$ some care should be used to make sure Global and not Local Min is reached



Can depend on start values.

Quasi Newton

DFP
BFGS } D computed based on first derivatives only and have good convergence properties.

Gauss-Newton Regressions

GNR

For simplicity, pretend we have only 1 parameter

$$y_t = x_t(\beta) + u_t$$

$$S(\beta) = \sum_{i=1}^{n} (y_t - x_t(\beta))^2$$

$$\frac{dS}{d\beta} = -2 \sum_{i=1}^{n} (y_t - x_t(\beta)) \frac{d\, x_t(\beta)}{d\beta}$$

Consider a F.O. Taylor's expansion of $x_t(\beta)$ around an initial guess.

$$x_t(\beta) \simeq x_t(\beta_0) + \frac{d\, x_t(\beta_0)}{d\beta} (\beta - \beta_0)$$

Slope $\simeq$

$$\frac{x_t(\beta) - x_t(\beta_0)}{\beta_1 - \beta_0} \simeq \frac{d\, x_t(\beta_0)}{d_1\beta)}$$

Define $\quad z_t(\beta_i) \equiv \left. \dfrac{d\, \chi_t(\beta)}{d\beta} \right|_{\beta_i}$

$$S(\beta) = \sum_{t=1}^{n} \Big(y_t - \chi_t(\beta_0) - z_t(\beta_0)(\beta - \beta_0) \Big)^2$$

let $\quad \bar{y}_t(\beta_0) = y_t - \chi_t(\beta_0) + z_t(\beta_0)\beta_0$

$$S = \sum_{t=1}^{n} \left(\bar{y}_t(\beta_0) - z_t(\beta_0)\beta \right)^2$$

Estimate β_1 with OLS Regression

$$\bar{y}(\beta_0) = z_t(\beta_0)\beta + u_t$$

$$\beta_1 = \left(Z(\beta)^T Z(\beta) \right)^{-1} Z^T(\beta)\, \bar{y}_T(\beta)$$

where $Z(\beta)_{n \times k}$ matrix with row $z_T(\beta)$

Sub for $\bar{y}$

$$\beta_{n+1} = \left[Z(\beta_n)^T Z(\beta_n) \right]^{-1} Z(\beta_n)^T \left(y_t - x_t(\beta_n) + Z(\beta_n)^T \beta_n \right)$$

$$= \left(Z(\beta_n)^T Z(\beta_n) \right)^{-1} Z(\beta_n)^T Z(\beta_n) \beta_n$$

$$+ \left(Z(\beta_n)^T Z(\beta_n) \right)^{-1} Z(\beta_n)^T \left(y - x_t(\beta_n) \right)$$

$$= \beta_n + \left(Z(\beta_n)^{-T} Z(\beta_n) \right)^{-1} \underbrace{Z(\beta_n)^T \left(y - x_t(\beta_n) \right)}_{\text{FOC.}}$$

AT FOC this term should be $= 0$

$\therefore \quad \beta_{n+1} \approx \beta_n$

GNR is b_n in

$$\beta_{n+1} = \beta_n + b_n$$

GNR b_n is OLS regression of.

$$y - x(\beta_n) = Z(\beta_n) b + u.$$

Cov estimation

$$\hat{\sigma}^2\left(Z(\beta_{n+1})^T Z(\beta_{n+1})\right)^{-1} \quad \text{with}$$

$$\hat{\sigma}^2 = SSR(\hat{\beta}_{n+1}) / n-k$$

or Robust

HCMME. $\left[Z(\beta_{n+1})^T Z(\beta_{n+1})\right]^{-1} Z(\beta_n)^T \hat{\Omega} Z(\beta_{n+1}) \left[Z(\beta_{n+1})^T Z(\beta_{n+1})\right]^{-1}$

$$\text{where} \quad \hat{\Omega} = \begin{bmatrix} \hat{u}_1^2 & & 0 \\ & \ddots & \\ 0 & & \hat{u}_n^2 \end{bmatrix}$$

More general tests:

In homoscedastic models $H_0: R\beta = r$ $H_A: R\beta \neq r$

$$F_{J,n-k} \sim \frac{(SSR_R - SSR_u)/J}{\hat{\sigma}_u^2}$$

or a Wald test based on HCMME.