Nonlinear Statistical Models, Chapter 2, Univariate Nonlinear Regression: Special Situations

by

A. Ronald Gallant
Department of Economics
University of North Carolina
Chapel Hill NC 27599-3305 USA

References

Gallant, A. Ronald (1987) *Nonlinear Statistical Models,* Wiley, New York.

Gallant, A. Ronald (1992) *Nonlinear Regression Asymptotics,* Manuscript, Department of Economics, University of North Carolina.

Gallant, A. Ronald (1997) *Introduction to Econometric Theory,* Princeton University Press, Princeton NJ.

Topics

- Heteroskedasticity
  - Known form
  - Unknown form
- Serial Correlation
  - Known form
  - Unknown form

A General Principle (1)

Consider the situation

$$y = f(\theta) + e \quad \mathcal{E}(ee') = \sigma^2 V$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad f(\theta) = \begin{pmatrix} f(x_1,\theta) \\ f(x_2,\theta) \\ \vdots \\ f(x_n,\theta) \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

If we factor V^{-1} as $V^{-1} = P'P$, then the rotated model

$$Py = Pf(\theta) + Pe \quad \text{or} \quad \text{“}y\text{”} = \text{“}f\text{”}(\theta) + \text{“}e\text{”}$$

is of the form we just studied because

$$\begin{aligned} \mathcal{E}\left[(\text{“}e\text{”})(\text{“}e\text{”})'\right] &= \mathcal{E}Pee'P' = P\mathcal{E}(ee')P' \\ &= \sigma^2 PVP' = \sigma^2 P(P'P)^{-1}P' \\ &= \sigma^2 I \end{aligned}$$

A General Principle (2)

For this observation to have any practical importance, it is necessary for P to be a sparse matrix with rows that have a simple, repetitive pattern.

To see this, rewrite the rotated model

$$Py = Pf(\theta) + Pe \quad \text{or} \quad \text{“}y\text{”} = \text{“}f\text{”}(\theta) + \text{“}e\text{”}$$

as

$$p_t'y = p_t'f(\theta) + p_t'e \quad t = 1, \ldots, n$$

where p_t' for $t = 1, \ldots, n$ are rows of P. To fit into the framework of Chapter 1,

$$\text{“}f\text{”}(\text{“}x_t\text{”}, \theta) = p_t'f(\theta)$$

must have a simple form such as

$$\text{“}f\text{”}(\text{“}x_t\text{”}, \theta) = a_t f(x_{t-1}, \theta) + b_t f(x_t, \theta)$$

where the sequence

$$\text{“}x_t\text{”} = (x_{t-1}, x_t, a_t, b_t)$$

is a Cesaro sum generator. Our applications have this structure.

Heteroskedastic Errors, Known Form

$$y_t = f(x_t, \theta) + e_t \qquad \mathcal{E}e_t^2 = \frac{\sigma^2}{\psi^2(x_t)}$$

Rotated model

$$\psi(x_t)y_t = \psi(x_t)f(x_t, \theta) + \psi(x_t)e_t$$

Just regress

$$\text{“}y_t\text{”} = \psi(x_t)y_t \text{ on } \text{“}f\text{”}(x_t, \theta) = \psi(x_t)f(x_t, \theta)$$

i.e., weighted least squares.

Heteroskedastic Errors
Known to Within a Parameter (1)

$$y_t = f(x_t, \theta) + e_t \qquad \mathcal{E}e_t^2 = \frac{1}{\psi^2(x_t, \tau)}$$

Either (1) put the model in implicit form and apply maximum likelihood or (2) estimate τ from least squares residuals.

The first is what ought to be done if the vectors θ and τ have some elements in common.

(1) Maximum Likelihood

Implicit model: $q(y_t, x_t, \lambda) = \psi(x_t, \tau)[y_t - f(x_t, \theta)] = \text{“}e_t\text{”}$

Parameter: $\lambda = (\theta, \tau)$ with elements in common deleted.

Assumed error density: $p(e, \sigma)$

Jacobian term: $J(y_t) = \frac{\partial}{\partial y} q(y_t, x_t, \lambda) = \psi(x_t, \tau)$

Log likelihood:

$$L(\lambda, \sigma) = \sum_{t=1}^{n} \log |\psi(x_t, \tau)| + \sum_{t=1}^{n} \log p[q(y_t, x_t, \lambda), \sigma]$$

Heteroskedastic Errors
Known to Within a Parameter (2)

$$y_t = f(x_t, \theta) + e_t \qquad \mathcal{E}e_t^2 = \frac{1}{\psi^2(x_t, \tau)}$$

Either (1) put the model in implicit form and apply maximum likelihood or (2) estimate τ from least squares residuals.

For (2) there are many approaches. This, in my opinion, is the best

Regress y_t on $f(x_t, \theta)$ by nonlinear least squares to get a preliminary least squares estimate $\hat{\theta}$ and residuals

$$\hat{e} = y_t - f(x_t, \hat{\theta}).$$

Compute

$$(\hat{\tau}, \hat{c}) = \underset{(t,c)}{\text{argmin}} \sum_{t=1}^{n} \left[|\hat{e}_t| - \frac{c}{\psi(x_t, \tau)} \right]^2.$$

using the optimization methods discussed in Chapter 1. Regress

$$\text{“}y_t\text{”} = \psi(x_t, \hat{\tau})y_t \text{ on } \text{“}f\text{”}(x_t, \theta) = \psi(x_t, \hat{\tau})f(x_t, \theta)$$

to get an estimate of θ . If the errors are normally distributed, then $\hat{c}$ estimates $\sqrt{\frac{2\sigma^2}{\pi}}$. If the discrepancy between this value and $\hat{c}$ is large (Wald test), you might worry that your assumption that $\mathcal{E}e_t^2 = \frac{1}{\psi^2(x_t, \tau)}$ is wrong.

Detection of Heteroskedasticity

Breusch-Pagan test:

$$\text{H}: \psi(x_t) = 1 \text{ against A}: \psi(x_t) = h(\beta' x_t)$$

Regress $\hat{e}_t^2/s^2$ on x_t, which is a linear regression, and reject H if

$$\frac{\text{SSE}(\hat{\beta})}{2}$$

exceeds the upper critical point of the chi squared distribution on $k-1$ degrees of freedom.

Plots:

Plot $|\hat{e}_t|$ against x_{it} for $i = 1, \ldots, k$.

Topics

- Heteroskedasticity
  - Known form
  - Unknown form
- Serial Correlation
  - Known form
  - Unknown form

Heteroskedasticity: Unknown Form (1)

The idea is to use the nonlinear least squares estimator and correct the variance estimate for heteroskedasticity. The correction is determine by working out the asymptotics assuming that

$$y_t = f(x_t, \theta^o) + e_t \quad \mathcal{E}e_t = 0 \quad \text{Var } e_t = \sigma_t^2$$

where σ_s^2 is not necessarily equal to σ_t^2 when $s \neq t$. The independence assumption is retained.

The first order conditions for

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\ s_n(\theta)$$

where

$$s_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} [y_t - f(x_t, \theta)]^2$$

are the same in Chapter 1.

Heteroskedasticity: Unknown Form (2)

First Order Conditions

$$\frac{\partial}{\partial \theta} s_n(\theta) = 0$$

Taylor's Expansion of FOC

$$\left[\frac{\partial^2}{\partial\theta\partial^{\theta}} s_n(\bar{\theta}_n)\right] \sqrt{n}(\hat{\theta}_n - \theta^o) = -\sqrt{n}\frac{\partial}{\partial\theta} s_n(\theta^o)$$

where $\bar{\theta}_n$ is on the line segment joining θ^o to $\hat{\theta}n$. Because $\bar{\theta}_n$ must therefore be closer to θ^o than $\hat{\theta}n$ is and $\lim_{n\to\infty} \hat{\theta}n = \theta^o$, we have $\lim_{n\to\infty} \bar{\theta}_n = \theta^o$ as well.

Asymptotics of RHS

$$-\sqrt{n}\frac{\partial}{\partial\theta}s_n(\theta^o) = \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}f(x_t,\theta^o)\,e_t$$

Mean: $\mathcal{E}\left[-\sqrt{n}\frac{\partial}{\partial\theta}s_n(\theta^o)\right] = 0$

Variance:

$$\begin{aligned}\mathcal{I}_n &= \text{Var}\left[-\sqrt{n}\frac{\partial}{\partial\theta}s_n(\theta^o)\right]\\ &= \frac{4}{n}\sum_{t=1}^{n}\sigma_t^2\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]'\end{aligned}$$

Limiting Variance: (**These are assumptions**)

$$\begin{aligned}\mathcal{I} &= \lim_{n\to\infty}\mathcal{I}_n\\ \mathcal{I}_n &= \lim_{n\to\infty}\hat{\mathcal{I}}_n\end{aligned}$$

where

$$\hat{\mathcal{I}}_n = \frac{4}{n}\sum_{t=1}^{n}\left[y_t - f(x_t,\hat{\theta})\right]^2\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]'$$

Central Limit Theorem:

$$-\sqrt{n}s_n(\theta^o) \xrightarrow{\mathcal{L}} N_p(0,\mathcal{I})$$

Asymptotics of LHS (same as before)

$$\begin{aligned}\mathcal{J}_n &= \left[\frac{\partial^2}{\partial\theta\partial^\theta}s_n(\bar{\theta}_n)\right]\\ &= \frac{2}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta}f(x_t,\bar{\theta}_n)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\bar{\theta}_n)\right]'\\ &\quad + \frac{2}{n}\sum_{t=1}^{n}e_t\left[\frac{\partial^2}{\partial\theta\partial\theta'}f(x_t,\bar{\theta}_n)\right]\end{aligned}$$

A consequence of the uniform strong law of large numbers is that a joint limit can be computed as an iterated limit; i.e.

$$\lim_{n\to\infty}\max_{\theta\in\Theta}|g_n(\theta)-g(\theta)| = 0 \;\;\&\;\; \lim_{n\to\infty}\bar{\theta}_n = \theta^o \;\;\Rightarrow\;\; \lim_{n\to\infty}g_n(\bar{\theta}_n) = g(\theta^o)$$

Therefore:

$$\begin{aligned}\mathcal{J} &= \lim_{n\to\infty}\mathcal{J}_n\\ &= 2\int\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]'d\mu(x)\\ &\quad + 2\int e\,dP(e)\int\frac{\partial^2}{\partial\theta\partial\theta'}f(x_t,\theta^o)\,d\mu(x)\\ &= 2Q\end{aligned}$$

LHS & RHS Combined

Slutsky's Theorem:

$$\mathcal{J}_n\sqrt{n}(\hat{\theta}_n - \theta^o) = -\sqrt{n}s_n(\theta^o)$$

$$-\sqrt{n}s_n(\theta^o) \xrightarrow{\mathcal{L}} N_p(0,\mathcal{I})$$

$$\mathcal{J} = \lim_{n\to\infty}\mathcal{J}_n$$

imply

$$\sqrt{n}(\hat{\theta}_n - \theta^o) \xrightarrow{\mathcal{L}} N_p(0,\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}).$$

That is all there is. In the heteroskedastic case the matrix $\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$ cannot be reduced to a simpler form.

Heteroskedasticity: Unknown Form (3)

To summarize, use the nls estimator

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\text{argmin}}\,[y - f(\theta)]'\,[y - f(\theta)]$$

and estimate the variance-covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta^o)$ by

$$\hat{V} = \hat{\mathcal{J}}^{-1}\hat{\mathcal{I}}\hat{\mathcal{J}}^{-1}$$

using

$$\hat{\mathcal{J}} = \frac{2}{n}\hat{F}'\hat{F}$$

$$\hat{\mathcal{I}}_n = \frac{4}{n}\sum_{t=1}^{n}\hat{e}_t^2\left[\frac{\partial}{\partial\theta}f(x_t,\hat{\theta}_n)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\hat{\theta}_n)\right]'$$

where

$$\hat{F} = \frac{\partial}{\partial\theta'}f(\hat{\theta}_n) \quad \hat{e} = y - f(\hat{\theta}_n)$$

Heteroskedasticity: Unknown Form, Tests

$$\text{H}: h(\theta^o) = 0 \text{ against A}: h(\theta^o) \neq 0$$

The proof that the likelihood ratio test follows the chi squared distribution requires $\mathcal{I}$ to equal $\mathcal{J}$ to within a scalar multiple. Therefore the likelihood ratio test cannot be used.

The Wald test is essentially $\hat{h} = h(\hat{\theta}_n)$ divided by its standard error. This can still be done:

$$W = n\hat{h}'(\hat{H}\hat{V}\hat{H}')^{-1}\hat{h}$$

where $\hat{H} = (\partial/\partial\theta')h(\hat{\theta}_n)$.

The Lagrange multiplier test is the G-N downhill direction $\tilde{D} = (\tilde{F}'\tilde{F})^{-1}\tilde{F}'[y - f(\tilde{\theta}_n)]$ divided by its standard error:

$$R = n\tilde{D}'\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-1}\tilde{H}\tilde{D}$$

where $\tilde{\theta}_n = \underset{h(\theta)=0}{\text{argmin}}\,[y - f(\theta)]'\,[y - f(\theta)]$

In both cases, reject when the statistic exceeds upper critical point of the chi squared distribution on q degrees freedom.

Topics

- Heteroskedasticity
  - Known form
  - Unknown form
- Serial Correlation
  - Known form
  - Unknown form

Serial Correlation: Known Form (1)

$$y_t = f(x_t, \theta^o) + u_t \quad \mathcal{E}u_t = 0$$

If the errors u_t are stationary, a standard assumption, then

$$\mathcal{E}u_t u_{t+h} = \gamma(h).$$

That is, the covariances only depend on the distance in time between errors, not on their position in time.

Written in vector form, the model is

$$y = f(\theta^o) + u \quad \mathcal{E}u = 0 \quad \mathcal{E}uu' = \Gamma_n$$

where

$$\Gamma_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & & \gamma(n-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & & \gamma(n-3) \\ \vdots & & & \ddots & \\ \gamma(n-1) & \gamma(n-2) & \gamma(n-3) & & \gamma(0) \end{pmatrix}$$

Serial Correlation: Known Form (2)

If $\gamma(h)$ declines to zero at a geometric rate, a standard assumption, then the variance matrix of an autoregressive model of order q can approximate Γ_n to within arbitrary accuracy. The autoregressive model has a very convenient factorization: $(\Gamma_n)^{-1} = P'P$

AR-1

$$u_t + a u_{t-1} = e_t \quad \mathcal{E}e_t = 0 \quad \mathcal{E}e_t^2 = \sigma^2$$

Yule-Walker Equations:

$$\mathcal{E}u_t u_t + a\mathcal{E}u_t u_{t-1} = \mathcal{E}u_t e_t$$

$$\mathcal{E}u_{t-1} u_t + a\mathcal{E}u_{t-1} u_{t-1} = \mathcal{E}u_{t-1} e_t$$

that is,

$$\gamma(0) + a\gamma(1) = \sigma^2$$

$$\gamma(1) + a\gamma(0) = 0$$

AR-1 Transformation:

$$P = \begin{pmatrix} \sigma/\sqrt{\gamma(0)} & 0 & & \dots & 0 \\ a & 1 & & & \\ & a & 1 & & \\ & & & \ddots & \\ & & & a & 1 \end{pmatrix}$$

Rotated model:

$$\frac{\sigma y_1}{\sqrt{\gamma(0)}} = \frac{\sigma f(x_1, \theta)}{\sqrt{\gamma(0)}} + e_1 \quad t = 1$$

$$a y_t + y_{t-1} = a f(x_t, \theta) + f(x_{t-1}, \theta) + e_t \quad t = 2, \dots, n$$

AR-q

$$u_t + a_1 u_{t-1} + a_2 u_{t-2} + \cdots + a_q u_{t-q} = e_t \quad \mathcal{E}e_t^2 = \sigma^2$$

Yule-Walker Equations:

$$\Gamma_{q+1} \begin{pmatrix} 1 \\ a_1 \\ \vdots \\ a_q \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Solution:

$$\begin{aligned} a &= -\Gamma_q^{-1} \gamma_q \\ \sigma^2 &= \gamma(0) + a'\gamma_q \end{aligned}$$

where

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_q \end{pmatrix} \qquad \gamma_q = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(q) \end{pmatrix}$$

AR-q

AR-q Transformation:

$$P = \left(\begin{array}{ccccccc|c} & & & \sigma P_q & & & & 0 \\ \hline a_q & a_{q-1} & \dots & a_1 & 1 & & & \\ & a_q & a_{q-1} & \dots & a_1 & 1 & & \\ & & & \ddots & & & & \\ & & & a_q & a_{q-1} & \dots & a_1 & 1 \end{array} \right)$$

where

$$(\Gamma_q)^{-1} = P_q' P_q$$

Rotated Model:

$$P_q \begin{pmatrix} y_1 \\ \vdots \\ y_q \end{pmatrix} = P_q \begin{pmatrix} f(x_1, \theta) \\ \vdots \\ f(x_q, \theta) \end{pmatrix} + e_t$$

$$y_t + a' \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-q} \end{pmatrix} = f(x_t, \theta) + a' \begin{pmatrix} f(x_{t-1}, \theta) \\ \vdots \\ f(x_{t-q}, \theta) \end{pmatrix} + e_t$$

AR Transformations

All one needs to compute the transformation are estimates of $\gamma(h)$. These can be estimated as follows:

Regress y_t on $f(x_t, \theta)$ by nonlinear least squares to get a preliminary least squares estimate $\hat{\theta}$ and residuals

$$\hat{u} = y_t - f(x_t, \hat{\theta}).$$

Estimate $\gamma(h)$ by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} \hat{u}_t \hat{u}_{t+h}$$

for $h = 0, 1, \dots, q$.

Use upward t-testing or BIC to determine q.

These are rotated models so the methods of Chapter 1 apply.

Reference: Gallant, A. Ronald, and J. Jeffery Goebel (1976), "Nonlinear Regression with Auto-correlated Errors," *Journal of the American Statistical Association* 71, 961–967.

Topics

- Heteroskedasticity
  - Known form
  - Unknown form
- Serial Correlation
  - Known form
  - Unknown form

Serial Correlation: Unknown Form (1)

$$y_t = f(x_t, \theta^o) + e_t \quad \mathcal{E}e_t = 0 \quad \mathcal{E}e_t e_{t+h} = \gamma(t, h)$$

The correlations depend on the separation in time and the position in time, and can be heteroskedastic as well.

Regularity conditions are usually stated in terms of mixing coefficients such as the strong mixing coefficient

$$\alpha_h = \max_t \max_{A,B} |P(A \cap B) - P(A)P(B)|$$

where A is an event that depends only on the past, namely $(\cdots, e_{t-1}, e_t)$, and B depends only on the future, namely $(e_{t+h}, e_{t+h+1}, \ldots)$. Notice the gap h between the past and future.

The relation between covariances and mixing coefficients is as follows: If $\mathcal{E}u_t^r \le B^r$, then

$$|\gamma(t,h)| \le 8B(\alpha_h)^{(r-2)/r}$$

The typical rate to get a strong law and a central limit theorem is

$$\alpha_h = h^{-r/(r-2)-\epsilon}$$

for some $\epsilon > 0$. Notice that this is slower that the geometric rate on covariances implied by the AR assumption used earlier.

Serial Correlation: Unknown Form (2)

As with heteroskedasticity of unknown form, the idea is to use the nonlinear least squares estimator and correct the variance estimate for both serial correlation and heteroskedasticity. The correction is determined by working out the asymptotics assuming that

$$y_t = f(x_t, \theta^o) + e_t \quad \mathcal{E}e_t = 0 \quad \mathcal{E}e_t e_{t+h} = \gamma(t, h).$$

where both x_t and e_t satisfy mixing conditions.

The first order conditions for

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\ s_n(\theta)$$

where

$$s_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} [y_t - f(x_t, \theta)]^2$$

are the same in Chapter 1.

Serial Correlation: Unknown Form (3)

First Order Conditions

$$\frac{\partial}{\partial \theta} s_n(\theta) = 0$$

Taylor's Expansion of FOC

$$\left[\frac{\partial^2}{\partial \theta \partial \theta} s_n(\bar{\theta}_n)\right] \sqrt{n}(\hat{\theta}_n - \theta^o) = -\sqrt{n} \frac{\partial}{\partial \theta} s_n(\theta^o)$$

where $\bar{\theta}_n$ is on the line segment joining θ^o to $\hat{\theta}n$. Because $\bar{\theta}_n$ must therefore be closer to θ^o than $\hat{\theta}n$ is and $\lim_{n\to\infty} \hat{\theta}n = \theta^o$, we have $\lim_{n\to\infty} \bar{\theta}_n = \theta^o$ as well.

Asymptotics of RHS

$$-\sqrt{n}\frac{\partial}{\partial\theta}s_n(\theta^o) = \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}f(x_t,\theta^o)\,e_t$$

Mean: $\mathcal{E}\left[-\sqrt{n}\frac{\partial}{\partial\theta}s_n(\theta^o)\right] = 0$

Variance:

$$\begin{aligned}\mathcal{I}_n &= \text{Var}\left[-\sqrt{n}\frac{\partial}{\partial\theta}s_n(\theta^o)\right]\\ &= \frac{4}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\mathcal{E}\left\{e_s e_t\left[\frac{\partial}{\partial\theta}f(x_s,\theta^o)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]'\right\}\end{aligned}$$

Limiting Variance: (**This is an assumption**)

$$\mathcal{I} = \lim_{n\to\infty}\mathcal{I}_n$$

Central Limit Theorem:

$$-\sqrt{n}s_n(\theta^o) \xrightarrow{\mathcal{L}} N_p(0,\mathcal{I})$$

Asymptotics of LHS (same as before)

$$\begin{aligned}\mathcal{J}_n &= \left[\frac{\partial^2}{\partial\theta\partial\theta'}s_n(\bar\theta_n)\right]\\ &= \frac{2}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta}f(x_t,\bar\theta_n)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\bar\theta_n)\right]'\\ &\quad + \frac{2}{n}\sum_{t=1}^{n}e_t\left[\frac{\partial^2}{\partial\theta\partial\theta'}f(x_t,\bar\theta_n)\right]\end{aligned}$$

A consequence of the uniform strong law of large numbers is that a joint limit can be computed as an iterated limit; i.e.

$$\lim_{n\to\infty}\max_{\theta\in\Theta}|g_n(\theta)-g(\theta)| = 0 \ \& \ \lim_{n\to\infty}\bar\theta_n = \theta^o \ \Rightarrow \ \lim_{n\to\infty}g_n(\bar\theta_n) = g(\theta^o)$$

Therefore:

$$\begin{aligned}\mathcal{J} &= \lim_{n\to\infty}\mathcal{J}_n\\ &= 2\int\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]' d\mu(x)\\ &\quad + 2\int e\,dP(e)\int\frac{\partial^2}{\partial\theta\partial\theta'}f(x_t,\theta^o)\,d\mu(x)\\ &= 2Q\end{aligned}$$

LHS & RHS Combined

Slutsky's Theorem:

$$\mathcal{J}_n\sqrt{n}(\hat\theta_n-\theta^o) = -\sqrt{n}s_n(\theta^o)$$

$$-\sqrt{n}s_n(\theta^o) \xrightarrow{\mathcal{L}} N_p(0,\mathcal{I})$$

$$\mathcal{J} = \lim_{n\to\infty}\mathcal{J}_n$$

imply

$$\sqrt{n}(\hat\theta_n-\theta^o) \xrightarrow{\mathcal{L}} N_p(0,\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}).$$

As for the heteroskedastic case, that is all there is. The matrix $\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$ cannot be reduced to a simpler form.

Estimation of $\mathcal{I}$ from NLS Residuals (1)

Rewrite the variance by grouping terms that are equidistant in time:

$$\begin{aligned}\mathcal{I}_n &= \frac{4}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\mathcal{E}\left\{e_s e_t\left[\frac{\partial}{\partial\theta}f(x_s,\theta^o)\right]\left[\frac{\partial}{\partial\theta}f(x_t,\theta^o)\right]'\right\}\\ &= \sum_{\tau=-(n-1)}^{n-1}\mathcal{I}_{n\tau}\end{aligned}$$

where

$$\mathcal{I}_{n\tau} = \begin{cases}\dfrac{4}{n}\displaystyle\sum_{t=\tau+1}^{n}\mathcal{E}e_t e_{t-\tau}\frac{\partial}{\partial\theta}f(x_t,\theta^o)\frac{\partial}{\partial\theta'}f(x_{t-\tau},\theta^o) & \tau\ge 0\\ \mathcal{I}'_{n,-\tau} & \tau<0\end{cases}$$

This looks like the formula for the variance of a spectral density at the zero frequency. Results from the spectral density literature can be applied.

Estimation of $\mathcal{I}$ from NLS Residuals (2)

Use residuals

$$\hat{e}_t = y_t - f(x_t, \hat{\theta}_n)$$

from the nonlinear least squares estimate

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\, [y - f(\theta)]' \, [y - f(\theta)]$$

to compute

$$\mathcal{I}_n = \sum_{\tau=-l(n)}^{l(n)} w\left(\frac{\tau}{l(n)}\right) \hat{\mathcal{I}}_{n\tau}$$

where $l(n) = n^{1/5}$ and

$$\hat{\mathcal{I}}_{n\tau} = \begin{cases} \dfrac{4}{n} \displaystyle\sum_{t=\tau+1}^{n} \hat{e}_t \hat{e}_{t-\tau} \frac{\partial}{\partial \theta} f(x_t, \hat{\theta}) \frac{\partial}{\partial \theta'} f(x_{t-\tau}, \hat{\theta}) & \tau \geq 0 \\ \hat{\mathcal{I}}'_{n,-\tau} & \tau < 0 \end{cases}$$

$$w(v) = \begin{cases} 1 - 6|v|^2 + 6|v|^3 & 0 \leq |v| \leq \frac{1}{2} \\ 2(1 - |v|)^3 & \frac{1}{2} \leq |v| \leq 1 \end{cases}$$

Estimation of $\mathcal{I}$ from NLS Residuals (3)

In the previous transparency, the truncation of the sum at $n^{1/5}$ is to avoid summing n^2 items. The overall divisor is n, so the sum would never converge. This introduces bias due to the neglected terms $\mathcal{I}_{n\tau}$ for $|\tau| > n^{1/5}$, but this bias is small because $\mathcal{I}_{n\tau}$ declines to zero at a polynomial rate due to the mixing assumptions that were used to get asymptotic normality.

The truncation of the sum at $n^{1/5}$ would destroy the positive definiteness of $\mathcal{I}_n$ were not for the weight function $w(v)$. That is the reason for its presence. The weight function shown is called the Parzen window, which is the recommended choice in the spectral density literature.

Serial Correlation: Unknown Form (4)

To summarize, use the nls estimator

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\, [y - f(\theta)]' \, [y - f(\theta)]$$

and estimate the variance-covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta^o)$ by

$$\hat{V} = \hat{\mathcal{J}}^{-1} \hat{\mathcal{I}} \hat{\mathcal{J}}^{-1}$$

using

$$\hat{\mathcal{J}} = \frac{2}{n} \hat{F}' \hat{F}$$

and $\hat{\mathcal{I}}_n$ as defined above, where

$$\hat{F} = \frac{\partial}{\partial \theta'} f(\hat{\theta}_n).$$

Serial Correlation: Unknown Form, Tests

$$\text{H} : h(\theta^o) = 0 \text{ against A} : h(\theta^o) \neq 0$$

The proof that the likelihood ratio test follows the chi squared distribution requires $\mathcal{I}$ to equal $\mathcal{J}$ to within a scalar multiple. Therefore the likelihood ratio test cannot be used.

The Wald test is essentially $\hat{h} = h(\hat{\theta}_n)$ divided by its standard error. This can still be done:

$$W = n \hat{h}' (\hat{H} \hat{V} \hat{H}')^{-1} \hat{h}$$

where $\hat{H} = (\partial / \partial \theta') h(\hat{\theta}_n)$.

The Lagrange multiplier test is the G-N downhill direction $\tilde{D} = (\tilde{F}' \tilde{F})^{-1} \tilde{F}' [y - f(\tilde{\theta}_n)]$ divided by its standard error:

$$R = n \tilde{D}' \tilde{H}' (\tilde{H} \tilde{V} \tilde{H}')^{-1} \tilde{H} \tilde{D}$$

where $\tilde{\theta}_n = \underset{h(\theta)=0}{\text{argmin}}\, [y - f(\theta)]' \, [y - f(\theta)]$

In both cases, reject when the statistic exceeds upper critical point of the chi squared distribution on q degrees freedom.

Topics

- Heteroskedasticity
  - Known form
  - Unknown form
- Serial Correlation
  - Known form
  - Unknown form