5d. Nonlinear LS

A. Colin Cameron Pravin K. Trivedi

These slides were prepared in 1999.
They cover material similar to Sections 5.8-5.9 of our subsequent book *Microeconometrics: Methods and Applications*, Cambridge University Press, 2005.

INTRODUCTION

- The nonlinear regression model specifies

$$y_i = g(\mathbf{x}_i, \boldsymbol{\beta}_0) + u_i, \quad i = 1, ..., n,$$

- where

y is a scalar dependent variable

$\mathbf{x}$ is a vector of explanatory variables

$\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters

$g(\cdot)$ is a specified function

u is an error term.

- The nonlinear least squares (NLS) estimator $\widehat{\boldsymbol{\beta}}$ minimizes the sum of squared residuals.
- In the notation used here $\widehat{\boldsymbol{\beta}}$ maximizes
$$Q_n(\boldsymbol{\beta}) = -\frac{1}{2n}S_n(\boldsymbol{\beta}) = -\frac{1}{2n}\sum_{i=1}^{n}(y_i - g(\mathbf{x}_i, \boldsymbol{\beta}))^2,$$
where $g(\cdot)$ is a specified regression function.
- The scale factor $1/2$ cancels out in the first-order conditions, see below.

- It is nonlinearity in $\boldsymbol{\beta}$ that is considered. If nonlinearity is just in $\mathbf{x}$ can do OLS with transformed regressors.
- It is assumed that the errors are additive.

EXAMPLES

- The exponential regression model
$$y = \exp(\mathbf{x}'\boldsymbol{\beta}) + u.$$
Then $\mathsf{E}[y|\mathbf{x}]$ is always positive, and effects of regressors are multiplicative rather than additive.

- Exponential regression f.o.c.
$$\frac{\partial}{\partial\boldsymbol{\beta}}\left(-\frac{1}{2n}\sum_{i=1}^{n}(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))^2\right) = \mathbf{0}$$
$$\Rightarrow -\frac{1}{2n}\sum_{i=1}^{n}(-2)\left(\frac{\partial}{\partial\boldsymbol{\beta}}\exp(\mathbf{x}_i'\boldsymbol{\beta})\right)\left(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})\right) = \mathbf{0}$$
$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\exp(\mathbf{x}_i'\boldsymbol{\beta})\left(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})\right) = \mathbf{0}.$$

- Nonlinear in $\boldsymbol{\beta}$ so no explicit solution.

- **Other Examples**
  - Regressors raised to a power

$$y = \beta_1 x_1 + \beta_2 x_2^{\beta_3} + u$$

  - Nonlinear functions from demand or production analysis, such as Cobb-Douglas production

$$y = \beta_1 x_1^{\beta_2} x_2^{\beta_3} + u.$$

  - Nonlinear restrictions on parameters

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u, \qquad \beta_3 = -\beta_2 \beta_1.$$

  - Error autoregressive of order one plus lagged dependent variable

$$y_t = \beta_1 x_t + \beta_1 y_{t-1} + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t,$$

where ε_t is i.i.d. error, which implies

$$y_t = \rho y_{t-1} + \beta_1 x_t + \rho\beta_1 y_{t-1} + \varepsilon_t.$$

MATRIX NOTATION

- It is often helpful to express the model in matrix notation.
- We have

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix},$$

where $g_i = g(\mathbf{x}_i, \boldsymbol{\beta})$.

- Or

$$\mathbf{y} = \mathbf{g} + \mathbf{u}$$

where

$\mathbf{y}$, $\mathbf{g}$ and $\mathbf{u}$ are $n \times 1$ vectors with i^{th} entries y_i, g_i and u_i.

- Then the NLS estimator minimizes

$$S_n(\boldsymbol{\beta}) = \mathbf{u}'\mathbf{u} = (\mathbf{y} - \mathbf{g})'(\mathbf{y} - \mathbf{g})$$

- For $Q_n(\boldsymbol{\beta}) = -1/2n \times S_n(\boldsymbol{\beta})$ the first-order conditons are

$$\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}(\mathbf{y} - \mathbf{g}),$$

where

$$\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \frac{\partial g_1}{\partial \beta_1} & \cdots & \frac{\partial g_n}{\partial \beta_k} \\ \vdots & & \vdots \\ \frac{\partial g_1}{\partial \beta_k} & \cdots & \frac{\partial g_n}{\partial \beta_k} \end{bmatrix}$$

DISTRIBUTION OF NLS

- The theory of extremum estimators applies directly.
- Objective function

$$Q_n(\boldsymbol{\beta}) = -\frac{1}{2n} S_n(\boldsymbol{\beta}) = -\frac{1}{2n} \sum_{i=1}^{n} (y_i - g(\mathbf{x}_i, \boldsymbol{\beta}))^2.$$

- First-order conditions

$$\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(\mathbf{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} (y_i - g(\mathbf{x}_i, \boldsymbol{\beta})) = \mathbf{0}.$$

- That is $\partial g(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$ is orthogonal to the error.

CONSISTENCY

- If the model is correctly specified

$$y_i = g(\mathbf{x}_i, \boldsymbol{\beta}_0) + u_i.$$

- Then $(y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0)) = u_i$ so

$$\partial Q_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}_0} = \frac{1}{n}\sum_{i=1}^{n} \partial g_i/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}_0} \times u_i.$$

- The informal condition for consistency is that

$$\mathsf{E}\left[\partial Q_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}_0}\right] = \mathbf{0}.$$

- This holds here if $\mathsf{E}\left[(\partial g_i/\partial\boldsymbol{\beta})\times u_i\right] = \mathbf{0}$, i.e. if $\mathsf{E}\left[u_i|\mathbf{X}\right] = 0$.

- Thus consistency requires correct specification of the mean and errors uncorrelated with the regressors.

ASYMPTOTIC NORMALITY

- By Taylor series

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\left[\frac{-1}{2n}\frac{\partial^2 S_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}^+}\right]^{-1}\frac{-1}{\sqrt{2n}}\frac{\partial S_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}_0}$$

- This will yield $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathbf{A}(\boldsymbol{\beta}_0)^{-1}$ times $\mathsf{N}\left[0, \mathbf{B}(\boldsymbol{\beta}_0)\right]$.
- Need to find $\mathbf{A}(\boldsymbol{\beta}_0)$ and $\mathbf{B}(\boldsymbol{\beta}_0)$.

- Specializing

$$
\begin{aligned}
&\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= -\left[\frac{-1}{n}\left(\sum_{i=1}^{n}\frac{\partial g_i}{\partial \boldsymbol{\beta}}\frac{\partial g_j}{\partial \boldsymbol{\beta}'} - \sum_{i=1}^{n}\frac{\partial^2 g_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}(y_i - g_i)\Bigg|_{\boldsymbol{\beta}^+}\right)\right]^{-1} \\
&\quad \times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial g_i}{\partial \boldsymbol{\beta}'}u_i\Bigg|_{\boldsymbol{\beta}_0}.
\end{aligned}
$$

- For $\mathbf{A}(\boldsymbol{\beta}_0)$, the term $\left(\partial^2 g_i/\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'\right)$ drops out if $\mathsf{E}\left[u_i|\mathbf{X}\right] = 0$.

- For $\mathrm{B}(\boldsymbol{\beta}_0)$ note that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\partial g_i / \partial \boldsymbol{\beta}\right) u_i \Big|_{\boldsymbol{\beta}_0}$$

- has mean $\mathbf{0}$ if $\mathsf{E}[u_i|\mathbf{X}] = 0$ and finite variance matrix

$$\frac{1}{n} \left[\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial g_i}{\partial \boldsymbol{\beta}} \frac{\partial g_j}{\partial \boldsymbol{\beta}'} \mathsf{Cov}\left[u_i, u_j | \mathbf{X}\right] \Big|_{\boldsymbol{\beta}_0} \right].$$

PROPOSITION

- Combining yields the following proposition is more general than those presented in an introductory class.
- It permits errors to be heteroskedastic or serially correlated.
- It does not allow the errors to be correlated with regressors (see NL2SLS in this case).

Proposition: *Distribution of NLS Estimator*. *Make the assumptions:*

(i) *The model is* $y_i = g(\mathbf{x}_i, \boldsymbol{\beta}_0) + u_i$;

(ii) *In the dgp* $\mathsf{E}[u|\mathbf{x}] = 0$ *and* $\mathsf{V}[\mathbf{uu}'|\mathbf{X}] = \boldsymbol{\Omega}_0$, *where* $\boldsymbol{\Omega}_{0,ij} = \omega_{ij}$;

(iii) *The mean function* $g(\cdot)$ *satisfies* $g(\mathbf{x}, \boldsymbol{\beta}^{(1)}) = g(\mathbf{x}, \boldsymbol{\beta}^{(2)})$ *iff* $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$;

(iv) *The following matrix exists and is finite nonsingular* $\mathbf{A}(\boldsymbol{\beta}_0) = \lim \frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{n} \frac{\partial g_i}{\partial \boldsymbol{\beta}}\frac{\partial g_j}{\partial \boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}_0}\right] = \lim \frac{1}{n}\mathsf{E}\left[\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}_0}\right]$;

(v) $n^{-1/2}\sum_{i=1}^{n}\partial g_i/\partial\boldsymbol{\beta}\times u_i|_{\beta_0}\xrightarrow{d}\mathsf{N}\left[0,\mathbf{B}(\boldsymbol{\beta}_0)\right]$ *where*

$$\mathbf{B}(\boldsymbol{\beta}_0)=\lim\frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial g_i}{\partial\boldsymbol{\beta}}\frac{\partial g_j}{\partial\boldsymbol{\beta}'}\omega_{ij}\bigg|_{\boldsymbol{\beta}_0}\right]=\lim\frac{1}{n}\mathsf{E}\left[\frac{\partial\mathbf{g}'}{\partial\boldsymbol{\beta}}\boldsymbol{\Omega}_0\frac{\partial\mathbf{g}}{\partial\boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}_0}\right].$$

Then the NLS estimator $\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}}$*, defined to be a root of the first-order conditions* $\partial n^{-1}S_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}=\mathbf{0}$*, is consistent for* $\boldsymbol{\beta}_0$ *and*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}}-\boldsymbol{\beta}_0)\xrightarrow{d}\mathsf{N}\left[0,\mathbf{A}(\boldsymbol{\beta}_0)^{-1}\mathbf{B}(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)^{-1}\right].$$

DISCUSSION

- The asymptotic distribution for $\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}}$ is

$$\mathsf{N}\left[\boldsymbol{\beta}_0, \left[\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}_0}\right]^{-1} \frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\boldsymbol{\Omega}_0\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}_0}\bigg|_{\boldsymbol{\beta}_0} \left[\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}_0}\right]^{-1}\right].$$

- Recall OLS estimator in linear model with heteroskedastic or correlated errors

$$\widehat{\boldsymbol{\beta}}_{\mathsf{OLS}} \overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}_0, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}_0\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\right],$$

- Thus replace $\mathbf{X}$ in OLS variance formula by $\partial \mathbf{g}/\partial \boldsymbol{\beta}'|_{\boldsymbol{\beta}_0}$.
- So NLS is OLS with regressors $\mathbf{x}$ replaced by $\partial g/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}_0}$, i.e. by $\partial \mathsf{E}[y|\mathbf{x}]/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}_0}$.

EXAMPLE: EXPONENTIAL REGRESSION

- Consider regression with exponential mean

$$y = \exp(\mathbf{x}'\boldsymbol{\beta}) + u.$$

- Then

$$\partial g(\mathbf{x}'\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \partial \exp(\mathbf{x}'\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \exp(\mathbf{x}'\boldsymbol{\beta})\mathbf{x}$$

- The general result for heteroskedastic errors is

$$\widehat{\boldsymbol{\beta}}_{\mathsf{EXP}} \overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}_0,\ \begin{array}{c} \left[\sum_{i=1}^n (\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2 \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \\ \times \left[\sum_{i=1}^n \omega_{ii,0}(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2 \mathbf{x}_i\mathbf{x}_i'\right] \\ \times \left[\sum_{i=1}^n (\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2 \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \end{array}\right].$$

- This depends on $\omega_{ii,0} = \mathsf{V}[u_i|\mathbf{x}_i]$ which is unknown.

VARIANCE MATRIX ESTIMATION

- Focus on variance matrix estimation when the errors have heteroskedasticity of form unknown, or only partially known.
- Then it is generally possible to adapt methods first developed for OLS to obtain consistent estimators of regression parameters and their variance matrix.
- Hence statistical inference is possible under weak distributional assumptions.

ROADMAP

NLS

- Specify a functional form for Ω and consistently estimate this. e.g. specific form of heteroskedasticity.
- Do not specify a functional form for Ω. e.g. White.

Weighted NLS

- Feasible nonlinear GLS. Specify function for Ω. Do fully efficient estimation assuming this is correct.
- Weighted NLS with working matrix. This is feasible nonlinear GLS but inference robust to misspecified Ω.

SPECIFIED $\boldsymbol{\Omega}_0 = \boldsymbol{\Omega}(\boldsymbol{\gamma}_0)$

- One approach is to assume a functional form for the variance matrix $\boldsymbol{\Omega}_0$ of the error term.
- Let
$$\boldsymbol{\Omega}_0 = \boldsymbol{\Omega}(\boldsymbol{\gamma}_0),$$
where $\boldsymbol{\gamma}_0$ is a finite-dimensional parameter vector and $\boldsymbol{\Omega}(\cdot)$ is a $n \times n$ matrix function.
- Then get a consistent estimate $\widehat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}_0$, form $\boldsymbol{\Omega}(\widehat{\boldsymbol{\gamma}})$, and evaluate earlier asymptotic results at $\boldsymbol{\Omega}(\widehat{\boldsymbol{\gamma}})$ and $\widehat{\boldsymbol{\beta}}$.

- For heteroskedasticity specify $\mathsf{V}[u_i|\mathbf{x}_i] = \exp(\mathbf{z}_i'\boldsymbol{\gamma})$, where $\mathbf{z}_i$ is composed of just a few key components of $\mathbf{x}_i$.
- Then
  - NLS of y on $g(\mathbf{x}, \boldsymbol{\beta})$ gives $\widehat{\boldsymbol{\beta}}$ and hence $\widehat{u}_i$.
  - NLS of $\widehat{u}_i^2$ on $\exp(\mathbf{z}_i'\boldsymbol{\gamma})$ gives $\widehat{\boldsymbol{\gamma}}$.
  - Form $\boldsymbol{\Omega}(\widehat{\boldsymbol{\gamma}}) = \mathsf{Diag}[\exp(\mathbf{z}_i'\widehat{\boldsymbol{\gamma}})]$.
- Evaluate variance matrices at $\boldsymbol{\Omega}(\widehat{\boldsymbol{\gamma}})$ and $\widehat{\boldsymbol{\beta}}$.

- For homoskedastic and uncorrelated errors get simple introductory text result.
- Then $\mathbf{\Omega}_0 = \sigma_0^2\mathbf{I}$.
- Simplification occurs as $\mathbf{B}(\boldsymbol{\beta}_0) = \sigma_0^2\mathbf{A}(\boldsymbol{\beta}_0)$,

$$\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}} \overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}_0, \sigma_0^2\left[\left.\frac{\partial\mathbf{g}'}{\partial\boldsymbol{\beta}}\frac{\partial\mathbf{g}}{\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0}\right]^{-1}\right],$$

- σ_0^2 is consistently estimated by $s^2 = n^{-1}(\mathbf{y}-\widehat{\mathbf{g}})'(\mathbf{y}-\widehat{\mathbf{g}})$.
- Aside: NLS is then efficient among estimators using only the first two moments.

UNSPECIFIED $\boldsymbol{\Omega}_0$

- In practice the variance matrix of errors may be unknown.
- For heteroskedasticity of unknown functional form, the error variance matrix is $\boldsymbol{\Omega}_0 = \mathsf{Diag}[\mathsf{E}[u_i^2|\mathbf{x}_i]]$.
- Without further structure it is not possible to obtain a consistent estimate of $\boldsymbol{\Omega}_0$, since the number of entries in $\boldsymbol{\Omega}_0$ equals the sample size n. As $n \to \infty$ the number of error variances to estimate also goes to infinity.

- White (1980a,b) gave conditions under which

$$\left[\frac{1}{n}\left.\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}} \widehat{\boldsymbol{\Omega}} \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\widehat{\boldsymbol{\beta}}}\right] \xrightarrow{p} \lim \frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{n} \frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\boldsymbol{\Omega}_0 \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0}\right],$$

for the obvious candidate

$$\widehat{\boldsymbol{\Omega}} = \mathsf{Diag}[(y_i - g(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}))^2],$$

where $\widehat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$.

- This leads to White's heteroskedastic-consistent estimate of the variance matrix of the NLS estimator

$$\widehat{\mathsf{V}}[\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}}] = \left[\left.\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}} \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\widehat{\boldsymbol{\beta}}}\right]^{-1} \left[\left.\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}} \widehat{\boldsymbol{\Omega}} \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\widehat{\boldsymbol{\beta}}}\right] \left[\left.\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}} \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\widehat{\boldsymbol{\beta}}}\right]^{-1}.$$

- This estimate works because it leads to a consistent estimate $\widehat{\mathrm{B}}(\boldsymbol{\beta}_0)$ of the $q \times q$ matrix $\mathrm{B}(\boldsymbol{\beta}_0)$.
- Note that $\widehat{\Omega}$ is clearly not consistent for the $n \times n$ matrix $\boldsymbol{\Omega}_0$ since $(y_i - g(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}))^2$ is not consistent for $\mathsf{E}[u_i^2|\mathbf{x}_i]$.
- Aside: Since $\widehat{\boldsymbol{\beta}}$ is consistent, $(y_i - g(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}))^2$ behaves asymptotically as $u_i^2 = (y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0))^2$, a random variable, not as its mean $\mathsf{E}[u_i^2|\mathbf{x}_i]$.

EXAMPLE: EXPONENTIAL REGRESSION

- For the model with exponential mean and heteroskedastic errors we already have obtained

$$\widehat{\boldsymbol{\beta}}_{\mathsf{EXP}} \overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}_0,\ \begin{array}{c} \left[\sum_{i=1}^{n}(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \\ \times\left[\sum_{i=1}^{n}\omega_{ii}(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i'\right] \\ \times\left[\sum_{i=1}^{n}(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \end{array}\right],$$

- A consistent estimate using White's estimator replaces $\boldsymbol{\beta}_0$ by $\widehat{\boldsymbol{\beta}}$ and ω_{ii} by $\widehat{u}_i^2$, where $\widehat{u}_i = y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})$.

FEASIBLE NONLINEAR GLS

- When the error variance matrix can be consistently estimated, by specifying a functional form $\boldsymbol{\Omega}(\boldsymbol{\gamma})$ and consistently estimating $\widehat{\gamma}$, one would actually go further and implement feasible nonlinear GLS.
- The estimator $\widehat{\boldsymbol{\beta}}_{\mathsf{FNLGLS}}$ maximizes
$$Q_n(\boldsymbol{\beta}) = -\frac{1}{n}(\mathbf{y}-\mathbf{g})'\boldsymbol{\Omega}(\widehat{\boldsymbol{\gamma}})^{-1}(\mathbf{y}-\mathbf{g}).$$
- It can be shown that provided $\boldsymbol{\Omega} = \boldsymbol{\Omega}_0(\boldsymbol{\gamma})$
$$\widehat{\boldsymbol{\beta}}_{\mathsf{FNLGLS}} \overset{a}{\sim} \mathsf{N}\left(\boldsymbol{\beta}_0, \left[\left.\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0} \boldsymbol{\Omega}(\boldsymbol{\gamma}_0)^{-1} \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\widehat{\mathbf{v}}}\right]^{-1}\right).$$

WORKING MATRIX

- In practice a chosen functional form $\boldsymbol{\Omega}(\boldsymbol{\gamma})$ may be a reasonable approximation for $\boldsymbol{\Omega}_0$, certainly better than the NLS implicit choice of $\boldsymbol{\Omega}_0 = \sigma_0^2 \mathbf{I}$.
- Start with a simple model for heteroskedasticity, such as $\mathsf{V}[u_i|\mathbf{x}_i] = \exp(\mathbf{z}_i'\boldsymbol{\gamma})$ where $\mathbf{z}_i$ is composed of just a few key components of $\mathbf{x}_i$ and do weighted NLS.
- Such simple assumed specification for the variance matrix is called a working matrix.
- Present results robust to misspecification of $\boldsymbol{\Omega}_0$

WEIGHTED NONLINEAR LEAST SQUARES

- The weighted nonlinear least squares (WNLS) estimator $\widehat{\boldsymbol{\beta}}_{\mathsf{WNLS}}$ with symmetric weighting matrix $\widehat{\mathbf{V}}$ minimizes

$$Q_n(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{g})'\widehat{\mathbf{V}}(\mathbf{y} - \mathbf{g}).$$

- Here $\widehat{\mathbf{V}}$ is the inverse of the working matrix: $\widehat{\mathbf{V}} = \boldsymbol{\Omega}(\widehat{\boldsymbol{\gamma}})^{-1}$.
- Assume $\mathbf{V}_0 = \operatorname{plim} \widehat{\mathbf{V}}$ exists and is nonstochastic, in addition to the usual assumptions for NLS.

- Then

$$\widehat{\boldsymbol{\beta}}_{\mathsf{WNLS}} \overset{a}{\sim} \mathsf{N}\left(\boldsymbol{\beta}_0,\ \left[\left.\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0} \mathbf{V}_0 \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0}\right]^{-1} \left[\left.\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0} \mathbf{V}_0 \boldsymbol{\Omega}_0 \mathbf{V}_0 \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0}\right] \left[\left.\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0} \mathbf{V}_0 \left.\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0}\right]^{-1}.\right)$$

- This result is essentially same as for linear WLS with $\mathbf{X}$ replaced by $\partial \mathbf{g}'/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}_0}$.

- Aside: Recall linear WLS. $\widehat{\boldsymbol{\beta}}_{\mathsf{WLS}}$ estimator minimizes
$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$
where $\mathbf{V}$ is an $n \times n$ symmetric weighting matrix.
Then
$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\mathsf{WLS}} &= (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{y} \\ &= (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}[\mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u}] \quad \text{if } \mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u}. \\ &= \boldsymbol{\beta}_0 + (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{u} \\ &\overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}_0, (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\boldsymbol{\Omega}_0\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-\mathbf{1}}\right].\end{aligned}$$
if $\mathbf{u}|\mathbf{X} \sim [\mathbf{0}, \boldsymbol{\Omega}_0]$.

- For heteroskedasticity of unknown form we use White (1980b) result using $u_i = y_i - g(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{\mathsf{NWLS}})$ to form $\widehat{\boldsymbol{\Omega}} = \mathsf{Diag}[\widehat{u}_i^2]$ as estimate for $\boldsymbol{\Omega}_0$.
- Note that we cannot go one step further and use $\widehat{\mathbf{V}} = \widehat{\boldsymbol{\Omega}}^{-1}$ where $\widehat{\boldsymbol{\Omega}} = \mathsf{Diag}[\widehat{u}_i^2]$. Here $\operatorname{plim} \widehat{\boldsymbol{\Omega}} = \mathsf{Diag}[u_i^2]$ is stochastic and so to is $\mathbf{V}_0 = \operatorname{plim} \widehat{\boldsymbol{\Omega}}^{-1}$, leading to inconsistency.
- For estimation as efficient as GLS without specifying $\boldsymbol{\Omega}_0$ see semi-paramteric regression.

EXAMPLE: EXPONENTIAL REGRESSION

- For model with exponential mean specify a working model of heteroskedasticity.
- Specify $\mathsf{V}[u_i|\mathbf{x}_i] = \exp(\mathbf{z}_i'\boldsymbol{\gamma})$ where $\mathbf{z}_i$ is a specified function of $\mathbf{x}_i$ such as selected subcomponents of $\mathbf{x}_i$.
- Then
  - First estimate $\widehat{\boldsymbol{\beta}}$ by NLS regression of y_i on $\exp(\mathbf{x}_i'\boldsymbol{\beta})$.
  - Then estimate $\widehat{\boldsymbol{\gamma}}$ by NLS regression of $(y_i - g(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}))^2$ on $\exp(\mathbf{z}_i'\boldsymbol{\gamma})$.
  - Finally calculate $\widetilde{\omega}_{ii} = \exp(\mathbf{z}_i'\widetilde{\boldsymbol{\gamma}})$.

- Since $\widehat{\mathbf{V}} = \mathsf{Diag}[\widetilde{\omega}_{ii}]$ the objective function $Q_n(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{g})'\widehat{\mathbf{V}}(\mathbf{y} - \mathbf{g})$ simplifies and the weighted NLS estimator $\widehat{\beta}_{\mathsf{WEXP}}$ estimator minimizes

$$Q_n(y, \boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))^2}{\exp(\mathbf{z}_i'\widetilde{\boldsymbol{\gamma}})}.$$

- This can clearly be minimized by usual NLS estimation in the transformed regression of $y_i / \exp(\mathbf{z}_i'\widetilde{\boldsymbol{\gamma}})^{1/2}$ on $\exp(\mathbf{x}_i'\boldsymbol{\beta}) / \exp(\mathbf{z}_i'\widetilde{\boldsymbol{\gamma}})^{1/2}$.
- Then letting $\omega_{ii,0} = \mathsf{V}[u_i|\mathbf{x}_i]$ denote the true (unknown) variance and $\omega_{ii} = \exp(\mathbf{z}_i, \boldsymbol{\gamma})$ denote the assumed (work-

ing) error variance

$$\widehat{\boldsymbol{\beta}}_{\mathsf{WEXP}} \overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}_0,\ \begin{array}{c} \left[\sum_{i=1}^{n}\frac{1}{\omega_{ii}}(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \\ \times\left[\sum_{i=1}^{n}\frac{1}{\omega_{ii}}\omega_{ii,0}(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i'\right] \\ \times\left[\sum_{i=1}^{n}\frac{1}{\omega_{ii}}(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \end{array}\right],$$

- The variance matrix of $\widehat{\boldsymbol{\beta}}_{\mathsf{WEXP}}$ is estimated by replacing $\boldsymbol{\beta}_0$ by $\widehat{\boldsymbol{\beta}}_{\mathsf{WEXP}}$, $\omega_{ii} = \exp(\mathbf{z}_i, \widetilde{\boldsymbol{\gamma}})$, and $\omega_{ii,0}$ by $\widehat{u}_i^2$, where $\widehat{u}_i = y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{\mathsf{WEXP}})$.
- This estimator is robust to misspecification of ω_{ii}.

- If one is prepared to do inference assuming that $\omega_{ii,0} = \exp(\mathbf{z}_i'\boldsymbol{\gamma})$ is correctly specified then the estimator is in fact the feasible nonlinear GLS estimator.
- The preceding result simplifies (because $\omega_{ii} = \omega_{ii,0}$) to

$$\widehat{\boldsymbol{\beta}}_{\mathsf{WEXP}} \overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}_0, \left[\sum_{i=1}^{n} \frac{1}{(\exp(\mathbf{z}_i'\boldsymbol{\gamma}))^2}(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2 \mathbf{x}_i\mathbf{x}_i'\right]^{-1}\right].$$

- This estimator of the variance matrix is generally not used as it is not robust to misspecification of ω_{ii}.

COEFFICIENT INTERPRETATION

- We are particularly interested in $\partial \mathsf{E}[y|\mathbf{x}]/\partial \mathbf{x}$.
- For general regression function $g(\mathbf{x}, \boldsymbol{\beta})$ it is customary to present one of the following estimates.
  - The response of the individual with average characteristics: $\partial \mathsf{E}[y|\mathbf{x}]/\partial \mathbf{x}|_{\bar{\mathbf{x}}}$.
  - The average response of all individuals in the sample: $\sum_{i=1}^{n} \partial \mathsf{E}[y_i|\mathbf{x}_i]/\partial \mathbf{x}_i$.
  - The response of a representative individual with characteristics $\mathbf{x} = \mathbf{x}^*$: $\partial \mathsf{E}[y|\mathbf{x}]/\partial \mathbf{x}|_{\mathbf{x}^*}$. e.g. a female

with twelve years of schooling etc.

- In the linear regression model, $\mathsf{E}[y|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta} \Rightarrow \partial\mathsf{E}[y|\mathbf{x}]/\partial\mathbf{x} = \boldsymbol{\beta}$ so these three measures are all the same.
- For nonlinear regression models, however, these three measures differ.

SINGLE-INDEX MODEL

- For general $g(\mathbf{x}, \boldsymbol{\beta})$ the coefficients $\boldsymbol{\beta}$ are difficult to interpret.
- Interpretation possible for single-index model
$$\mathsf{E}[y|\mathbf{x}] = g(\mathbf{x}'\boldsymbol{\beta}).$$
- Then nonlinearity is of the mild form that the mean is a nonlinear function of a linear combination of the regressors and parameters.

- For single-index model
$$\frac{\partial \mathsf{E}[y|\mathbf{x}]}{\partial x_j} = \frac{\partial g(\mathbf{x}'\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\boldsymbol{\beta}_j,$$
- The relative effects of changes in regressors are given by the ratio of the coefficients since
$$\frac{\partial \mathsf{E}[y|\mathbf{x}]/\partial x_j}{\partial \mathsf{E}[y|\mathbf{x}]/\partial x_k} = \frac{\beta_j}{\beta_k}.$$
- And for $g(\cdot)$ monotonic it follows that the signs of the coefficients give the signs of the effects.
- Single index models are particularly advantageous due to their simple interpretation.

EXAMPLE: EXPONENTIAL REGRESSION

- Then $\mathsf{E}[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta})$.
- This is a single-index model.
- Thus if, say, $\widehat{\beta}_j = 0.4$ and $\widehat{\beta}_k = 0.8$ the impact on $\mathsf{E}[y|\mathbf{x}]$ of a one unit change in the k^{th} regressor is twice that of a one-unit change in the j^{th} regressor.
- Furthermore $\partial\mathsf{E}[y|\mathbf{x}]/\partial\mathbf{x} = \exp(\mathbf{x}'\boldsymbol{\beta}) \times \boldsymbol{\beta} = \mathbf{E}[y|\mathbf{x}] \times \boldsymbol{\beta}$.
- So the parameters can be interpreted as semi-elasticities.
- Thus $\widehat{\beta}_j = 0.4$ implies that a one unit increase in the j^{th} regressor leads to a 40% increase in $\mathsf{E}[y|\mathbf{x}]$.

DATA EXAMPLE: EXPONENTIAL REGRESSION

$$
\begin{aligned}
y &\sim \text{exponential}(\exp(\alpha + \beta x + \gamma z)) \\
& x, z \text{ correlated normal} \\
\text{R-squared} &= .2 \text{ for regression}
\end{aligned}
$$

Point and interval estimates.

Misspecified variables.

Omitted variables.

TIME SERIES

- Preceding cross-section results can be adapted to nonlinear time series

$$y_t = g(x_t, \boldsymbol{\beta}) + u_t, \quad t = 1, ..., T.$$

- There are several ways to proceed depending on whether or not
  - regressors x_t include lagged values of y, such as y_{t-1}
  - u_t is serially correlated.

- If u_t is serially correlated use ARMA error model
  $$u_t = \rho_1 u_{t-1} + \cdots + \rho_p u_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q},$$
  where ε_t is iid with mean 0 and variance σ^2.

TIME SERIES METHODS

- 1. Include in the regressors sufficient lags of y that the error is serially uncorrelated.
  Do NLS with i.i.d. errors.
  Simplest NLS results for iid error apply.
- 2. Specify ARMA model for correlated errors. Regressors that may or may not include lags of y.
  Do nonlinear GLS.
  In simplest case of AR errors can estimate $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$ by using Cochrane-Orcutt transformation to obtain model with i.i.d. error, and estimate this transformed model

by nonlinear GLS.

- 3. Do not include lags of y as regressors and let error be serially correlated.
  Do NLS even though the error is serially correlated.
  Fet adjusted standard errors that are correct in presence of serial correlation. This is the analog of White.
- Aside: A fourth possible approach gives inconsistent estimates. Include lags of y and do NLS with error that is serially correlated. Then $\mathbf{x}_t$ and u_t are correlated violating the assumption that $\mathsf{E}[u|\mathbf{x}] = 0$.
  Instead the second approach needs to be taken.

TIME SERIES: UNSPECIFIED Ω

- White and Domowitz (1984) considered heteroskedasticity and serial correlation of unknown functional form.
- Restrict serial correlation to at most lag length l periods. Also allow heteroskedasticity.
  Then

$$\begin{aligned}\boldsymbol{\Omega}_{0,st} &= \mathsf{E}[u_t u_s] && |s-t| \leq l \\ &= 0 && |s-t| > l.\end{aligned}$$

- Then one can use the general result for NLS with $\widehat{\mathbf{\Omega}}$ that has st^{th} entry

$$\begin{aligned} \widehat{\mathbf{\Omega}}_{st} &= \widehat{u}_s \widehat{u}_t \quad |s-t| \leq l \\ &= 0 \qquad\ \ |s-t| > l, \end{aligned}$$

where $\widehat{u}_t = y_t - g(\mathbf{x}_t, \widehat{\boldsymbol{\beta}})$ is the NLS residual.

- Then $\widehat{\mathsf{V}}[\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}}]$ is that given earlier where the consistent estimate of $\mathrm{B}(\boldsymbol{\beta}_0)$ is

$$\widehat{\mathrm{B}}(\boldsymbol{\beta}_0) = \frac{1}{n}\left[\sum_{t=1}^{n} \frac{\partial g_t}{\partial \boldsymbol{\beta}} \frac{\partial g_t}{\partial \boldsymbol{\beta}'}\bigg|_{\widehat{\boldsymbol{\beta}}} \widehat{u}_t^2 + \sum_{\tau=1}^{l} \sum_{t=\tau+1}^{n} \frac{\partial g_t}{\partial \boldsymbol{\beta}} \frac{\partial g_{t-\tau}}{\partial \boldsymbol{\beta}'}\bigg|_{\widehat{\boldsymbol{\beta}}} \widehat{u}_t \widehat{u}_{t-\tau}\right].$$

- Again this estimates works because it leads to a consistent estimate of the $q \times q$ matrix $\mathrm{B}(\boldsymbol{\beta}_0)$, even though $\widehat{\boldsymbol{\Omega}}$ is not consistent for the $n \times n$ matrix $\boldsymbol{\Omega}_0$.
- This general result specializes to results for heteroskedasticity and for serial correlation.

EXAMPLE: EXPONENTIAL REGRESSION

- Dgp is an exponential density with exponential mean.

$$y_i|\mathbf{x}_i \sim \text{exponential}(\exp(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}))$$
$$(x_{2i}, x_{3i}) \sim \mathsf{N}[0.1, 0.1; 0.1^2, 0.1^2, 0.005]$$
$$(\beta_1, \beta_2, \beta_3) = (-2, 2, 2).$$
$$i = 1, ..., 200.$$

- For the joint normal the means, variances and co-variance are respectively given. The implied squared correlation between x_2 and x_3 is 0.25.

- For the particular sample of 200 observations drawn here
  - sample mean of y is 0.21
  - sample standard deviation of y is 0.22.

- For the exponential the conditional mean and variance are

$$\begin{aligned} \mathsf{E}[y|\mathbf{x}] &= \exp(\beta_1 + \beta_2 x_2 + \beta_3 x_3) \\ \mathsf{V}[y|\mathbf{x}] &= (\mathsf{E}[y|\mathbf{x}])^2 . \end{aligned}$$

- It follows that
  - OLS with mean $\beta_1 + \beta_2 x_2 + \beta_3 x_3$ is inconsistent.
  - NLS with mean $\exp(\beta_1 + \beta_2 x_2 + \beta_3 x_3)$ is consistent but is inefficient with standard errors that should be adjusted for heteroskedasticity.
  - MLE is consistent and efficient.

$Variable$	$Estimator$		
	OLS	NLS	MLE
ONE	0.14	-1.84	-2.68
	(10.3)	(-15.0)	(-19.9)
$x1$	0.48	2.00	3.23
	(3.2)	(2.6)	(3.1)
$x2$	0.20	0.60	1.27
	(1.5)	(0.9)	(1.3)
R^2	$.08$	$.09$	

- The coefficients are relatively imprecisely estimated for this sample with $R^2 \simeq 0.09$.
- In particular the slope parameters for NLS and MLE are generally different from their theoretical values of 2.0.
- Though this difference is not statistically significant at 5% (t-ratios are given).