Nonlinear models

General specification testing:

- For $E(Y|X)$:
  - RESET test
  - Chow test
  - P test for nonnested hypotheses

- For $Pr(Y|X)$:
  - Information Matrix text, usually very hard to implement
  - More common: tests designed specifically to particular models

Nonlinear models

General specification testing:

RESET test:

- Implementation:
  - Estimate the original model:

$$Pr(Y|X) = F(\beta_0 + \beta_1 x_1 + \cdots \beta_k x_k)$$

  - Generate the variables $(X\hat{\beta})^2, (X\hat{\beta})^3, (X\hat{\beta})^4, \ldots$
  - Add the generated variables to the original model and estimate the following auxiliary model:

$$Pr(Y|X) = F\left[\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma_1 (X\hat{\beta})^2 + \gamma_2 (X\hat{\beta})^3 + \gamma_3 (X\hat{\beta})^4 + \cdots\right]$$

  - Apply a LR / Wald test for the significance of the added variables:

    $H_0: \gamma_1 = \gamma_2 = \gamma_3 = \cdots = 0$ (suitable model functional form)

    H_1: No H_0 (unsuitable model functional form)

Nonlinear models

General specification testing:

Chow test for structural breaks:

- Context:
  - Two groups of individuals / firms / ...: G_A, G_B
  - It is suspected that the behaviour of the two groups in which regards the dependent variable may have different determinants
- Implementation:
  - Generate the dummy variable $D = \begin{cases} 1 & \text{if the individual belongs to } G_A \\ 0 & \text{if the individual belongs to } G_B \end{cases}$
  - Estimate the original model 'duplicated':

$$Pr(Y|X) = F(\theta_0 + \theta_1 X_1 + \cdots + \theta_k X_k + \gamma_0 D + \gamma_1 D X_1 + \cdots + \gamma_k D X_k)$$

  - Apply a LR / Wald test for the significance of the variables including D:

    $H_0: \gamma_1 = \cdots = \gamma_k = 0$ (no structural break)

    H_1: Não H_0 (with a structural break)

General specification testing:

P Test for Nonnested Hypotheses:

- Usefulness:
  - Most nonlinear models are based on specifications for $E(Y|X)$ that are nonnested: one specification is not a particular case of the other
  - The *P* test checks nonnested specifications, one against the other, with the test being calculated twice, with each model under H_0:

$$H_0: E(Y|X) = G(X\beta);\ H_1: E(Y|X) = S(X\beta)$$
$$H_0': E(Y|X) = S(X\beta);\ H_1': E(Y|X) = G(X\beta)$$

- Possibles outcomes from the test:
  - G is better than S, since H_0' is rejected but H_0 is not
  - S is better than G, since H_0 is rejected but H_0' is not
  - No model is suitable, since both H_0 and H_0' are rejected
  - Both models are suitable, since neither H_0 nor H_0' are rejected

Nonlinear models

General specification testing:

- Implementation:
  - $H_0: E(Y|X) = G(X\beta); H_1: E(Y|X) = S(X\beta)$:
    - Estimate the null model and get $G\left(X\hat{\beta}\right)$
    - Estimate the alternative model and get $S\left(X\hat{\beta}\right)$
    - Estimate the linear artificial model:
    $$Y - G\left(X\hat{\beta}\right) = \left[g\left(X\hat{\beta}\right)X\right]\theta + \gamma\left[S\left(X\hat{\beta}\right) - G\left(X\hat{\beta}\right)\right]$$
    - t test for:
    $$H_0: \gamma = 0 \Leftrightarrow E(Y|X) = G(X\beta)$$
  - $H_0': E(Y|X) = S(X\beta); H_1': E(Y|X) = G(X\beta)$
    - Similar procedures but inverting the roles of the two models

Nonlinear models

Partial Effects:

- Linear models:
  - Model: $E(Y|X) = X\beta$
  - Effects: $\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \beta_j$

- Nonlinear models:
  - Model:
    - $E(Y|X) = G(X\beta)$
    - $Pr(Y|X) = F(X\beta)$
  - Effects: $\Delta X_j = 1 \Longrightarrow$
    - $\Delta E(Y|X) = \frac{\partial E(Y|X)}{\partial X_j} = \frac{\partial G(X\beta)}{\partial X_j} = \beta_j \frac{\partial G(X\beta)}{\partial X\beta} = \beta_j g(x_i'\beta)$
    - $\Delta Pr(Y|X) = \frac{\partial P\ \ (Y|X)}{\partial X_j} = \frac{\partial F(X\beta)}{\partial X_j} = \beta_j \frac{\partial F(X\beta)}{\partial X\beta} = \beta_j f(x_i'\beta)$

Nonlinear models

Partial Effects:

- Partial effects may be compared across different models, but the values of β cannot
- However, because $\frac{\partial G(X\beta)}{\partial X\beta} > 0$ and $\frac{\partial F(X\beta)}{\partial X\beta} > 0$:
  - The sign of the partial effect is given by the sign of β_j
  - Testing the statistical significance of the partial effect is equivalent to test for $H_0: \beta_j = 0$

Nonlinear models

- To calculate the magnitude of the partial effects, there are three main alternatives:
  - Average marginal effect: calculate the partial effects for each individual in the sample and then obtain the mean of those effects
  - Marginal effect at mean: replace x by its sample means
  - Marginal effect at a representative value: replace x by specific values

Stata
(after estimating the model)

margins, dydx($varlist$) at(...)

margins, dydx($varlist$) atmeans

margins, dydx($varlist$)

Models for Nonnegative Outcomes

Continuous Outcomes and Count Data

Log-Linear and Exponential Regression Models

Poisson and Negative Binomial Models

Panel Data Models

- Nonnegative outcomes can be:
  - Continuous: $Y \in [0, +\infty[$
    - Examples: prices, wages,…
  - Discrete (counts): $Y \in \{0,1,2,3, \dots\}$
    - Examples: patents applied for by a firm in a year, times someone is arrested in a year,...

- Linear regression models are not the most suitable option because:
  - May generate negative predictions for the dependent variable
  - At least close to the lower bound of Y, it does not make sense to assume constant partial effects

Log-linear regression model:

$$ln(Y_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i$$

- Assumption: $E(u_i|x) = 0$
- With this transformation, the dependent variable becomes unbounded: $Y \in]0, +\infty[\Longrightarrow \ln(Y) \in]-\infty, +\infty[$
- However, two new problems arise:
  - The log-linear model is not defined for $Y = 0$; adding a small constant value to Y or dropping zeros are not in general good solutions
  - Prediction is more interesting in the original scale, $\widehat{Y}_i$, and not in the logarithmic scale, $\widehat{ln(Y_i)}$; the log-linear model gives the latter directly but retransforming it to the original scale requires additional assumptions and calculations and/or the application of relatively complex methods

Exponential regression model:

$$Y = exp(x'\beta + u)$$
$$E(Y|X) = exp(x'\beta)$$

- Assumption: $E(e^u|x) = 1$
- Advantages:
  - $\widehat{Y}_i$ is always nonnegative
  - Predictions are obtained directly in the original scale, without requiring any retransformations
- Partial effects:

$$\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \beta_j exp(x'\beta)$$

  - The sign of the effect is given by the sign of β_j
  - β_j can be interpreted as a semi-elasticity, since:

$$100\beta_j = 100\frac{\Delta E(Y|X)}{E(Y|X)}, \text{i.e. } \Delta X_j = 1 \Longrightarrow \%\Delta E(Y|X) = 100\beta_j\%$$

- Assumptions and estimation methods according to the type of nonnegative outcome:
  - Continuous response:
    - Assumption: only $E(Y|X)$; estimation: QML
  - Count data - two alternatives:
    - Assumption: only $E(Y|X)$; estimation: QML
    - Assumption: $E(Y|X)$ and $Pr(Y = j|X)$; estimation: ML

- Three main distribution functions are used as basis for QML and/or ML estimation:
  - Poisson
  - Negative Binomial 1
  - Negative Binomial 2

Poisson regression model:

$$Y_i \sim Poisson(\lambda_i) \Longrightarrow Pr(Y_i = y|x_i) = \frac{e^{-\lambda_i}\lambda_i{}^y}{y!}$$

where $\lambda_i = E(Y|X) = exp(x'\beta)$

- Estimation methods: ML (only count data) or QML, since the Poisson distribution belongs to the linear exponential family

- By definition, $E(Y|X) = Var(Y|X)$ (equidispersion), which may be a strong assumption is some empirical applications

Stata
ML: poisson $YX_1 \ldots X_k$
QML: poisson $YX_1 \ldots X_k$, robust

Negative binomial regression models:

- Two variants, both allowing for overdispersion ($\delta > 0$):
  - NEGBIN1: $Var(Y|X) = (1 + \delta)E(Y|X)$ - ML estimation
  - NEGBIN2: $Var(Y|X) = [1 + \delta E(Y|X)]E(Y|X)$ - it belongs to the linear exponential family, enabling estimation by both ML (only count data) and QML

Stata
NEGBIN1: nbreg $Y X_1$ … X_k, dispersion(constant)
NEGBIN2 (ML): nbreg $Y X_1$ … X_k, dispersion(mean)
NEGBIN2 (QML): nbreg $Y X_1$ … X_k, dispersion(mean) robust

- Overdispersion test:

$H_0: \delta = 0$ (Poisson model)

$H_1: \delta \neq 0$ (Negative Binomial 1 or 2 model)

Base model:

- Continuous / count data:

$$E(Y_{it}|x_{it}, \alpha_i) = exp(\gamma_i + x'_{it}\beta) = \alpha_i exp(x'_{it}\beta)$$

- Count data:

$$Pr(Y_{it} = y|x_{it}, \alpha_i) = \frac{e^{-\lambda_{it}}{\lambda_{it}}^{y}}{y!}$$

$$\lambda_i = E(Y_{it}|x_{it}, \alpha_i) = \alpha_i exp(x'_{it}\beta)$$

Pooled estimator:

- Based on the cross-sectional assumption $E(Y_{it}|x_{it}) = exp(x'_{it}\beta)$
- Produces consistent estimators only if $E(\alpha_i|x_{it}) = 1$
- Does not require the Poisson distributional assumption
- Using a robust vce controls for both overdispersion and time dependence

Stata
poisson $Y X_1 \ldots X_k$, vce(cluster *clustvar*)

Random Effects Poisson Estimator:

- Assumptions:
  - $Y_{it} \sim Poisson(\lambda_{it})$
  - $\lambda_i = E(Y_{it}|x_{it}, \alpha_i) = \alpha_i exp(x'_{it}\beta)$
  - $log(\alpha_i) = \gamma_i \sim Gamma(1, \eta)$

- Resulting model:
  - NEGBIN2-type model
  - Estimation method: ML
  - $E(Y_{it}|x_{it}) = exp(x'_{it}\beta)$, which implies that the Pooled estimator is consistent under random effects of this type

Stata
xtpoisson $Y X_1 \dots X_k$, re

- Alternative model: assumes $log(\alpha_i) = \gamma_i \sim N(0, \sigma^2)$ and produces $(Y_{it}|x_{it}) = exp(x'_{it}\beta)$ but has no close form solution

Fixed Effects Estimators:

- Fixed effects Poisson estimator (three equivalent versions):
  - Pooled estimator with individual effects
  - Estimator conditional on $\sum_{t=1}^{T} Y_{it}$, with $\sum_{t=1}^{T} Y_{it} \neq 0$
  - Quasi mean-differenced GMM estimator (Hausman, Hall and Griliches, 1984)

- Quasi-differences GMM estimator:
  - Chamberlain (1992)
  - Wooldridge (1997)

Do not require the Poisson distributional assumption

Fixed effects Poisson estimator:

- May be derived using the three equivalent versions
- Pooled estimator with individual effects:
  - Adds individual dummies, associated to the γ_i's
  - As in linear models, β is consistently estimated even in short panels (no incidental parameters problem)
- The quasi mean-differenced GMM estimator is based on the following moment condition:

$$E\left(Y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{Y}_i \middle| x_{it}\right) = 0,$$

where $\lambda_{it} = exp(x_{it}'\beta)$ and $\bar{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T}\lambda_{it}$

- Requires strictly exogenous explanatory variables

Stata
xtpoisson $Y X_1 \ldots X_k$, fe

Quasi-differences GMM estimator :

- Chamberlain (1992):

$$E\left(\frac{\lambda_{i,t-1}}{\lambda_{i,t}}Y_{it} - Y_{i,t-1}\middle| x_{it}\right) = 0$$

- Wooldridge (1997):

$$E\left(\frac{Y_{it}}{\lambda_{it}} - \frac{Y_{i,t-1}}{\lambda_{i,t-1}}\middle| x_{it}\right) = 0$$

- In both cases the explanatory variables do not need to be strictly exogenous, so these estimators are particularly useful in dynamic models