Nonlinear mixed-effects models using Stata

Yulia Marchenko

Executive Director of Statistics
StataCorp LP

2017 German Stata Users Group meeting

- Nonlinear mixed-effects models (NLMEMs)
  - mixed effects = fixed effects + random effects
- Nonlinear multilevel models
- Nonlinear hierarchical models

NLMEMs are popular in studies of biological and agricultural growth processes, population pharmacokinetics, bioassays, and more. For example, NLMEMs have been used to model drug absorption in the body, intensity of earthquakes, and growth of plants.

- Nonlinear regression:

$$y = \frac{1}{\beta_1 + \beta_2 x + \beta_3 x^2} + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$.

- Let, e.g., β_1 vary randomly across G groups:

$$\beta_1 = \beta_{1j} = b_1 + u_j,\ j = 1, 2, \ldots, G$$

where $u_j \sim N(0, \sigma_u^2)$.

- Variance components: error variance σ^2 and between-group variance σ_u^2.
- Coefficients β_2 and β_3 can also be group-specific.

- Alternatively, consider a linear mixed-effects model:

$$y_{ij} = \beta_1 + \beta_2 x_{ij} + \beta_3 x_{ij}^2 + u_j + \epsilon_{ij}$$

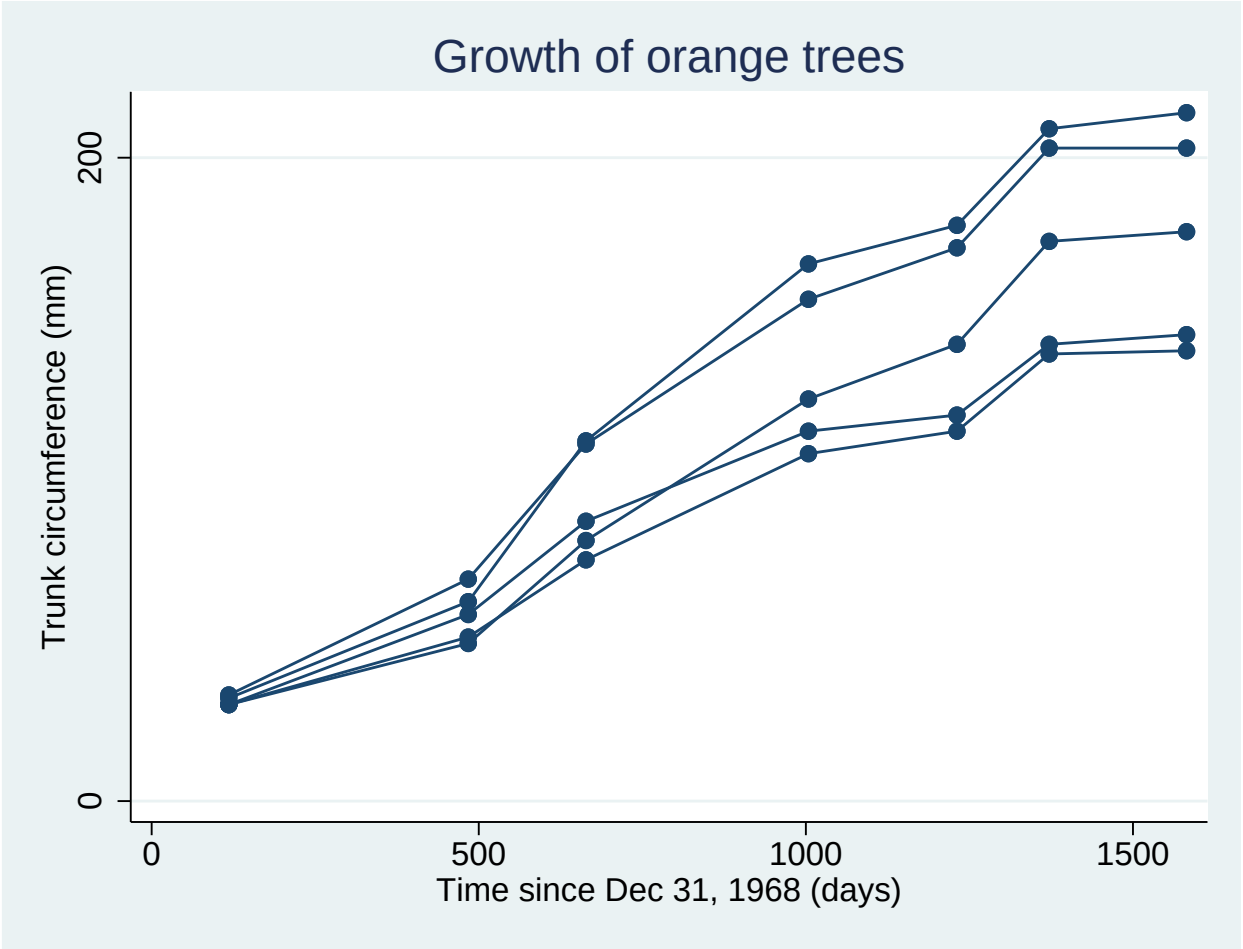
where $\epsilon_{ij} \sim N(0, \sigma^2)$ and $u_j \sim N(0, \sigma_u^2)$.

- In the nonlinear mixed-effects model

$$y_{ij} = \frac{1}{\beta_1 + \beta_2 x_{ij} + \beta_3 x_{ij}^2 + u_j} + \epsilon_{ij}$$

all coefficients and random intercept u_j enter nonlinearly.

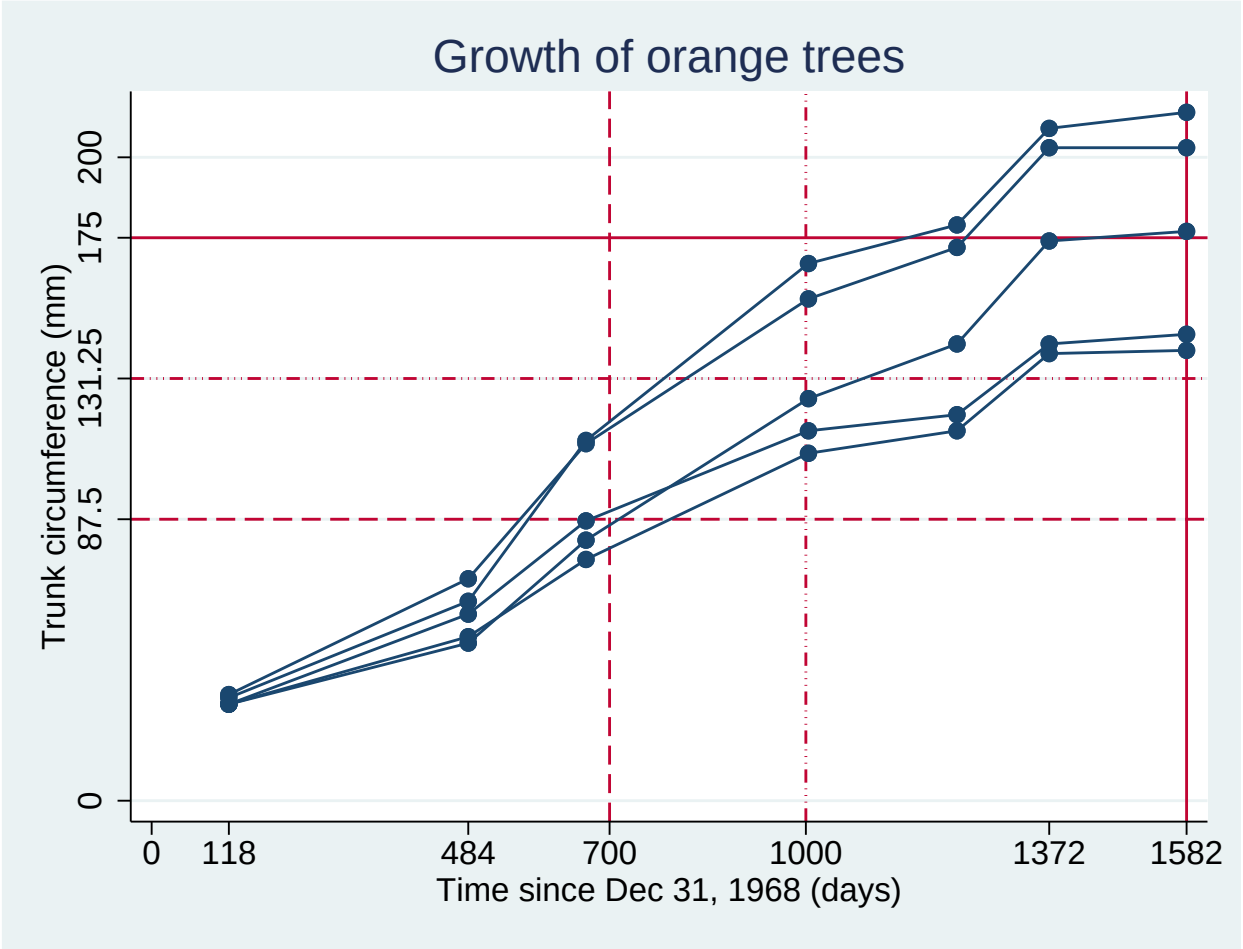
```
. webuse orange
(Growth of orange trees (Draper and Smith, 1998))
. twoway connected circumf age, connect(L) title(Growth of orange trees)
```



- Consider the following nonlinear growth model:

$$\texttt{circumf}_{ij} = \frac{\beta_1}{1 + \exp\left\{-\left(\text{age}_{ij} - \beta_2\right)/\beta_3\right\}} + \epsilon_{ij}$$

  where $\epsilon_{ij} \sim N(0, \sigma^2)$.
- β_1 is the average asymptotic trunk circumference of trees as age $\to \infty$.
- β_2 estimates the age at which a tree attains half of β_1.
- β_3 represents the number of days it takes for a tree to grow from 50% to about 73% of its average asymptotic trunk circumference β_1.



- $\beta_1 \approx 175$ mm, $\beta_2 \approx 700$ days, and $\beta_3 \approx 1{,}000 - 700 = 300$ days.
- Notice that the variability between trees increases with age.

- Let's incorporate the between-tree variability into the model.
- Consider the following two-level nonlinear growth model (Pinheiro and Bates 2000):

$$\texttt{circumf}_{ij} = \frac{\beta_1 + u_{1j}}{1 + \exp\left\{-\left(\text{age}_{ij} - \beta_2\right)/\beta_3\right\}} + \epsilon_{ij}$$

where $u_{1j} \sim N(0, \sigma^2_{u_1})$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

- We use `menl` to fit the model.

```
. menl circumf = ({b1}+{U1[tree]})/(1+exp(-(age-{b2})/{b3}))
Mixed-effects ML nonlinear regression           Number of obs     =         35
Group variable: tree                            Number of groups  =          5
                                                Obs per group:
                                                              min =          7
                                                              avg =        7.0
                                                              max =          7
Linearization log likelihood = -131.58458
```

circumf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
/b1	191.049	16.15403	11.83	0.000	159.3877	222.7103
/b2	722.556	35.15082	20.56	0.000	653.6616	791.4503
/b3	344.1624	27.14739	12.68	0.000	290.9545	397.3703

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
tree: Identity				
var(U1)	991.1514	639.4636	279.8776	3510.038
var(Residual)	61.56371	15.89568	37.11466	102.1184

- Similarly, we can let β_2 and β_3 vary across trees.
- We use a more convenient multistage formulation:

$$\texttt{circumf}_{ij} = \frac{\beta_{1j}}{1+\exp\left\{-\left(\text{age}_{ij}-\beta_{2j}\right)/\beta_{3j}\right\}} + \epsilon_{ij}$$

where

$$\begin{aligned}\beta_{1j} &= b_1 + u_{1j}\\ \beta_{2j} &= b_2 + u_{2j}\\ \beta_{3j} &= b_3 + u_{3j}\end{aligned}$$

and where $u_{1j} \sim N(0, \sigma^2_{u_1})$, $u_{2j} \sim N(0, \sigma^2_{u_2})$ and $u_{3j} \sim N(0, \sigma^2_{u_3})$.

```
. menl circumf = ({beta1:})/(1+exp(-(age-{beta2:})/{beta3:})),
> define(beta1:{b1}+{U1[tree]})
> define(beta2:{b2}+{U2[tree]})
> define(beta3:{b3}+{U3[tree]})
```

Mixed-effects ML nonlinear regression — Number of obs = 35
Group variable: tree — Number of groups = 5
Obs per group:
min = 7
avg = 7.0
max = 7

Linearization log likelihood = -131.55076

beta1: {b1}+{U1[tree]}
beta2: {b2}+{U2[tree]}
beta3: {b3}+{U3[tree]}

circumf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
/b1	191.1332	15.96228	11.97	0.000	159.8477	222.4187
/b2	722.7144	34.94627	20.68	0.000	654.2209	791.2078
/b3	345.2863	27.70935	12.46	0.000	290.977	399.5956

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
tree: Independent				
var(U1)	970.67	665.4967	253.2113	3721.004
var(U2)	140.9707	2669.433	1.07e-14	1.85e+18
var(U3)	248.5962	1397.996	.0040617	1.52e+07
var(Residual)	59.43549	18.44102	32.35519	109.1812

- With only five trees, the previous model is already too rich for these data.
- Otherwise, we could have considered a more complicated covariance structure for the random effects:

$$(u_{1j}, u_{2j}, u_{3j}) \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}),\ \mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

- Or assuming dependence only between some random effects such as u_{1j} and u_{2j}:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}$$

- And variations of the above.

- For example,

```
. menl circumf = ({beta1:})/(1+exp(-(age-{beta2:})/{beta3:})),
> define(beta1:{b1}+{U1[tree]})
> define(beta2:{b2}+{U2[tree]})
> define(beta3:{b3}+{U3[tree]})
> covariance(U1 U2 U3, unstructured)
```

- The above is also equivalent to:

```
. menl ..., ... covariance(U*, unstructured)
```

- Or, assuming correlation between only `U1` and `U2`

```
. menl ..., ... covariance(U1 U2, unstructured)
```

- `menl` provides flexible modeling of within-group error structures (or residual covariance structures).
- Use option `resvariance()` to model error heteroskedasticity as a linear, power, or exponential function of other covariates or of predicted values.
- Use option `rescorrelation()` to model the dependence of the within-group errors as, e.g., AR or MA processes.
- Combine `resvariance()` and `rescorrelation()` to build flexible residual covariance structures.

- Continuing with growth processes, consider the growth of soybean plants.
- Variable `weight` records an average leaf weight per plant in grams.
- Variable `time` records the number of days after planting at which plants were weighed.
- The data are obtained from 48 plots.

```
. webuse soybean
(Growth of soybean plants (Davidian and Giltinan, 1995))
```

- Consider the following growth model:

$$\texttt{weight}_{ij} = \frac{\phi_{1j}}{1+\exp\left\{-\left(\texttt{time}_{ij}-\phi_{2j}\right)/\phi_{3j}\right\}} + \epsilon_{ij}$$

where

$$\begin{aligned}\phi_{1j} &= b_1 + u_{1j}\\ \phi_{2j} &= b_2 + u_{2j}\\ \phi_{3j} &= b_3 + u_{3j}\end{aligned}$$

and where $(u_{1j}, u_{2j}, u_{3j}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{\Sigma})$ with

$$\mathbf{\Sigma} = \begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{12} & \sigma_{22} & \sigma_{23}\\ \sigma_{13} & \sigma_{23} & \sigma_{33}\end{bmatrix}$$

and $\epsilon_{ij} \sim N(0, \sigma^2)$.

- We use the following specification of `menl`:

```
. menl weight = {phi1:}/(1+exp(-(time-{phi2:})/{phi3:})),
> define(phi1: U1[plot], xb)
> define(phi2: U2[plot], xb)
> define(phi3: U3[plot], xb)
> covariance(U1 U2 U3, unstructured)
```

- Option

```
define(phi1: U1[plot], xb)
```

  is essentially a shortcut for

```
define(phi1: {b1}+{U1[plot]})
```

- The above shortcut is useful to specify linear combinations.

- Estimates of regression coefficients:

```
Mixed-effects ML nonlinear regression          Number of obs     =        412
Group variable: plot                           Number of groups  =         48
                                               Obs per group:
                                                             min =          8
                                                             avg =        8.6
                                                             max =         10

Linearization log likelihood = -739.83445
        phi1:  U1[plot], xb
        phi2:  U2[plot], xb
        phi3:  U3[plot], xb
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
phi1						
_cons	19.25314	.8031811	23.97	0.000	17.67893	20.82734
phi2						
_cons	55.01999	.7272491	75.65	0.000	53.59461	56.44537
phi3						
_cons	8.403468	.3152551	26.66	0.000	7.78558	9.021357

- Estimates of variance components:

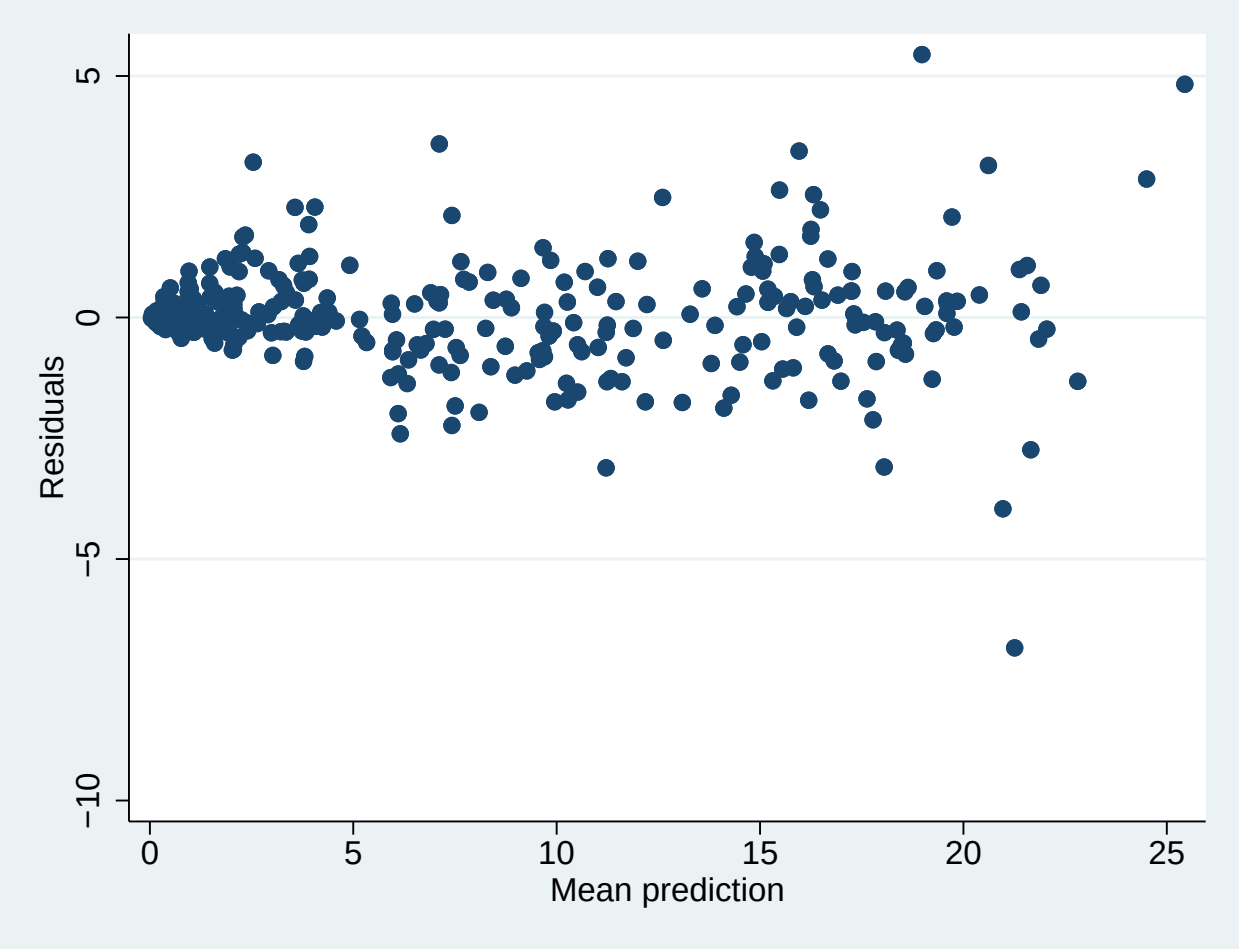
Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
plot: Unstructured				
var(U1)	27.05081	6.776516	16.55561	44.19929
var(U2)	17.61605	5.317899	9.748766	31.83227
var(U3)	1.972036	.9849825	.7409021	5.248904
cov(U1,U2)	15.73304	5.413365	5.123042	26.34304
cov(U1,U3)	5.193819	2.165586	.9493488	9.438289
cov(U2,U3)	5.649306	2.049458	1.632442	9.66617
var(Residual)	1.262237	.1111686	1.062119	1.50006

- Store estimation results for later comparison

```
. estimates store nohet
```

- Residuals-versus-fitted plot

```
. predict fitweight, yhat
. predict res, residuals
. scatter res fitweight
```



- Davidian and Giltinan (1995) proposed to model heteroskedasticity (the error variance) in this example as a power function of the mean:

$$\mathrm{Var}(\epsilon_{ij}) = \sigma^2(\widehat{\texttt{weight}}_{ij})^{2\delta}$$

  where $\widehat{\texttt{weight}}_{ij}$ denotes predicted mean weight values.

- The corresponding `menl` specification is

```
. menl weight = {phi1:}/(1+exp(-(time-{phi2:})/{phi3:})),
> define(phi1: U1[plot], xb)
> define(phi2: U2[plot], xb)
> define(phi3: U3[plot], xb)
> covariance(U1 U2 U3, unstructured)
> resvariance(power _yhat, noconstant)
```

- Estimates of regression coefficients:

```
Mixed-effects ML nonlinear regression          Number of obs     =        412
Group variable: plot                           Number of groups  =         48
                                               Obs per group:
                                                             min =          8
                                                             avg =        8.6
                                                             max =         10

Linearization log likelihood = -357.55571
        phi1:  U1[plot], xb
        phi2:  U2[plot], xb
        phi3:  U3[plot], xb
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
phi1						
_cons	16.9422	.6060387	27.96	0.000	15.75439	18.13002
phi2						
_cons	51.77667	.462577	111.93	0.000	50.87004	52.68331
phi3						
_cons	7.540957	.0963157	78.29	0.000	7.352182	7.729732

- Estimates of variance components:

plot: Unstructured				
var(U1)	11.47264	2.747485	7.174911	18.34469
var(U2)	3.014802	1.278198	1.313322	6.920641
var(U3)	.1017371	.0442522	.0433746	.2386292
cov(U1,U2)	.5324789	.131718	.2743164	.7906415
cov(U1,U3)	.9081537	.2459849	.4260321	1.390275
cov(U2,U3)	.340901	.1091677	.1269363	.5548658
Residual variance:				
Power _yhat				
sigma2	.0496757	.0043236	.0418849	.0589156
delta	.9376681	.0253201	.8880416	.9872945

- Store estimation results for comparison

```
. estimates store het
```

- Likelihood-ratio test:

```
. lrtest het nohet
Likelihood-ratio test                                 LR chi2(1)  =     764.56
(Assumption: nohet nested in het)                     Prob > chi2 =     0.0000
```

- Information criteria:

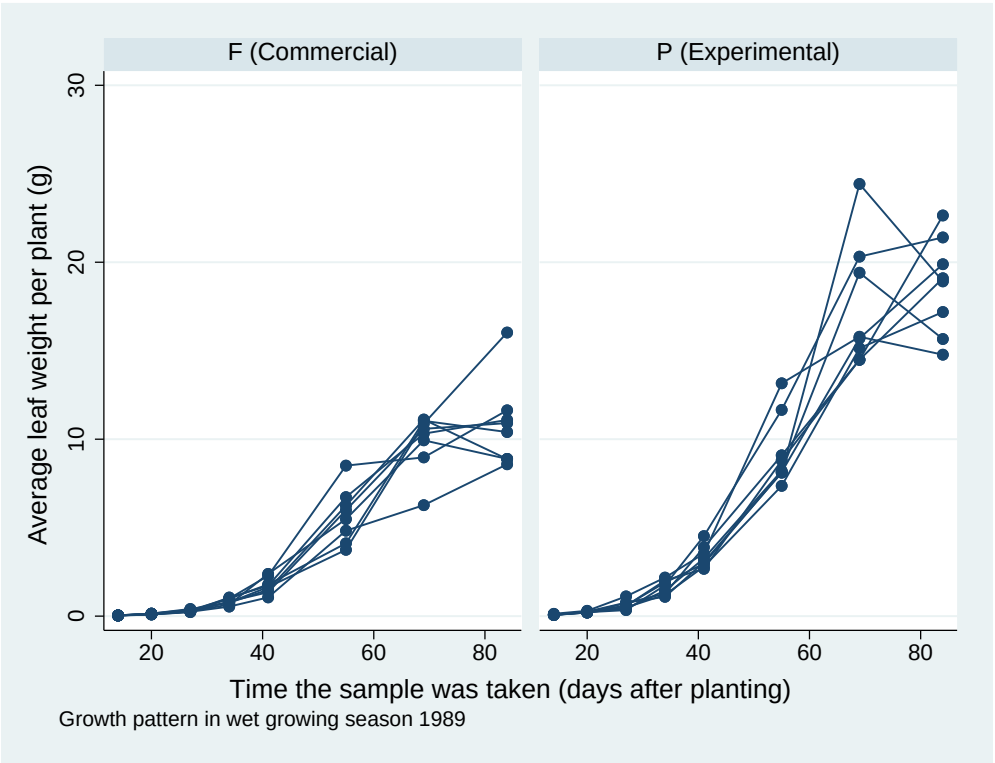
```
. estimates stats het nohet
Akaike´s information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
het	412	.	-357.5557	11	737.1114	781.3427
nohet	412	.	-739.8344	10	1499.669	1539.879

Note: N=Obs used in calculating BIC; see [R] BIC note.

- A heteroskedastic model fits data better.

- The actual objective of the soybean study was to compare the growth patterns of two genotypes of soybean plants in three types of growing seasons.
- Genotypes, `variety`: commercial variety `F` and experimental variety `P`
- Growing seasons, `year`: dry (1988), wet (1989), and normal (1990).



Growth pattern in wet growing season 1989

- We can include the main effects of genotypes and of years, and their interaction in the equation for the asymptotic rate:

$$\phi_{1j} = b_1 + \boldsymbol{\beta}_G^\top \mathbf{G} + \boldsymbol{\beta}_Y^\top \mathbf{Y} + \cdots + u_{1j}$$

- `menl` specification:

```
. menl weight = {phi1:}/(1+exp(-(time-{phi2:})/{phi3:})),
> define(phi1: i.variety##i.year U1[plot])
> define(phi2: U2[plot], xb)
> define(phi3: U3[plot], xb)
> covariance(U1 U2 U3, unstructured)
> resvariance(power _yhat, noconstant)
```

- We can also let the coefficients for, e.g., genotypes vary across plots:

$$\phi_{1j} = b_1 + \beta_G^\top \mathbf{G} + \beta_Y^\top \mathbf{Y} + \cdots + u_{1j} + \mathtt{F} \times f_{1j} + \mathtt{P} \times p_{1j}$$

  where `F` and `P` are genotype indicators and $f_{1j} \sim N(0, \sigma_F^2)$ and $p_{1j} \sim N(0, \sigma_P^2)$.
- `menl` specification:

```
. menl weight = {phi1:}/(1+exp(-(time-{phi2:})/{phi3:})),
> define(phi1: i.variety##i.year U1[plot] 1.variety#F1[plot] 2.variety#P1[plot])
> define(phi2: U2[plot], xb)
> define(phi3: U3[plot], xb)
> covariance(U1 U2 U3, unstructured)
> resvariance(power _yhat, noconstant)
```

- The `i.` operator is not allowed with factor variables when specifying random coefficients because a distinct name is required for each random coefficient.

- Constant elasticity of substitution (CES) production function is used in macroeconomic modeling to model the production process as a function of inputs such as capital and labor.
- It introduces and estimates the CES parameter, which makes it a flexible modeling tool.
- Elasticity of substitution (ES) measures how easy it is to substitute one input such as capital for another such as labor. And constant ES does not depend on input values.
- Other common production functions such as Cobb-Douglas and Leontief can be viewed as special cases of the CES production function. For example, Cobb-Douglas function assumes that ES is 1.

- Consider fictional data on log(production) from the 50 U.S. states plus D.C. divided into 9 regions over the period of 1990 to 2017.
- We wish to fit the CES production function

$$\ln P_{ijt} = \beta_0 - \frac{1}{\rho}\ln\{\delta K_{ijt}^{-\rho} + (1-\delta)L_{ijt}^{-\rho}\} + \epsilon_{ijt}$$

  where $\epsilon_{ijt} \sim N(0, \sigma^2)$.
- $\ln P_{ijt}$, K_{ijt}, and L_{ijt} are log(production), capital, and labor of state j within region i in year t.
- Parameters: log-factor productivity β_0, share δ, and ρ is related to the elasticity of substitution $\sigma = 1/(1+\rho)$.

- We suspect that δ may be affected by regions and states-within-regions:

$$\delta = \delta_{ij} = \delta_0 + u_{1i} + u_{2ij}$$

where $u_{1i} \sim N(0, \sigma^2_{u_1})$ and $u_{2ij} \sim N(0, \sigma^2_{u_2})$. u_2's are nested within u_1's.

```
. menl lnprod = {b0}-1/{rho}*ln({delta:}*capital^(-{rho})+(1-{delta:})*labor^(-{rho})),
> define(delta: {delta0} + {U1[region]} + {U2[region>state]})
Mixed-effects ML nonlinear regression           Number of obs     =      1,377
```

Path	No. of Groups	Observations per Group Minimum	Average	Maximum
region	9	108	153.0	216
region>state	51	27	27.0	27

```
Linearization log likelihood =  1094.2223
      delta:  {delta0}+{U1[region]}+{U2[region>state]}
```

lnprod	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
/b0	3.49166	.0040189	868.82	0.000	3.483783	3.499537
/delta0	.3439896	.0490629	7.01	0.000	.2478281	.4401511
/rho	1.109318	.0272072	40.77	0.000	1.055993	1.162644

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
region: Identity				
var(U1)	.0199948	.0102071	.0073517	.0543809
region>state: Identity				
var(U2)	.0073329	.001642	.004728	.0113729
var(Residual)	.0102169	.0003967	.0094681	.0110248

There is some variability between regions and states-within-regions in the estimates of the share parameter.

- We can predict the share parameter for each region:

```
. predict (delta = {delta:}), relevel(region)
. list region delta if region[_n]!=region[_n-1], sep(0)
```

	region	delta
1.	New England	.2699136
163.	Mid Atlantic	.1453616
271.	E North Central	.6366224
406.	W North Central	.3761043
595.	South Atlantic	.3879336
811.	E South Central	.344411
919.	W South Central	.17091
1027.	Mountain	.4102365
1243.	Pacific	.3544133

- We can use `nlcom` to estimate the ES:

```
. nlcom (sigma: 1/(1+_b[/rho]))
      sigma:  1/(1+_b[/rho])
```

lnprod	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sigma	.4740868	.006115	77.53	0.000	.4621015	.4860721

- The estimated ES is 0.47, which is less than one meaning that the capital and labor are not very good substitutes in this example. If the labor price increases, substituting capital for labor will not offset the increase in the total expenditure on labor.

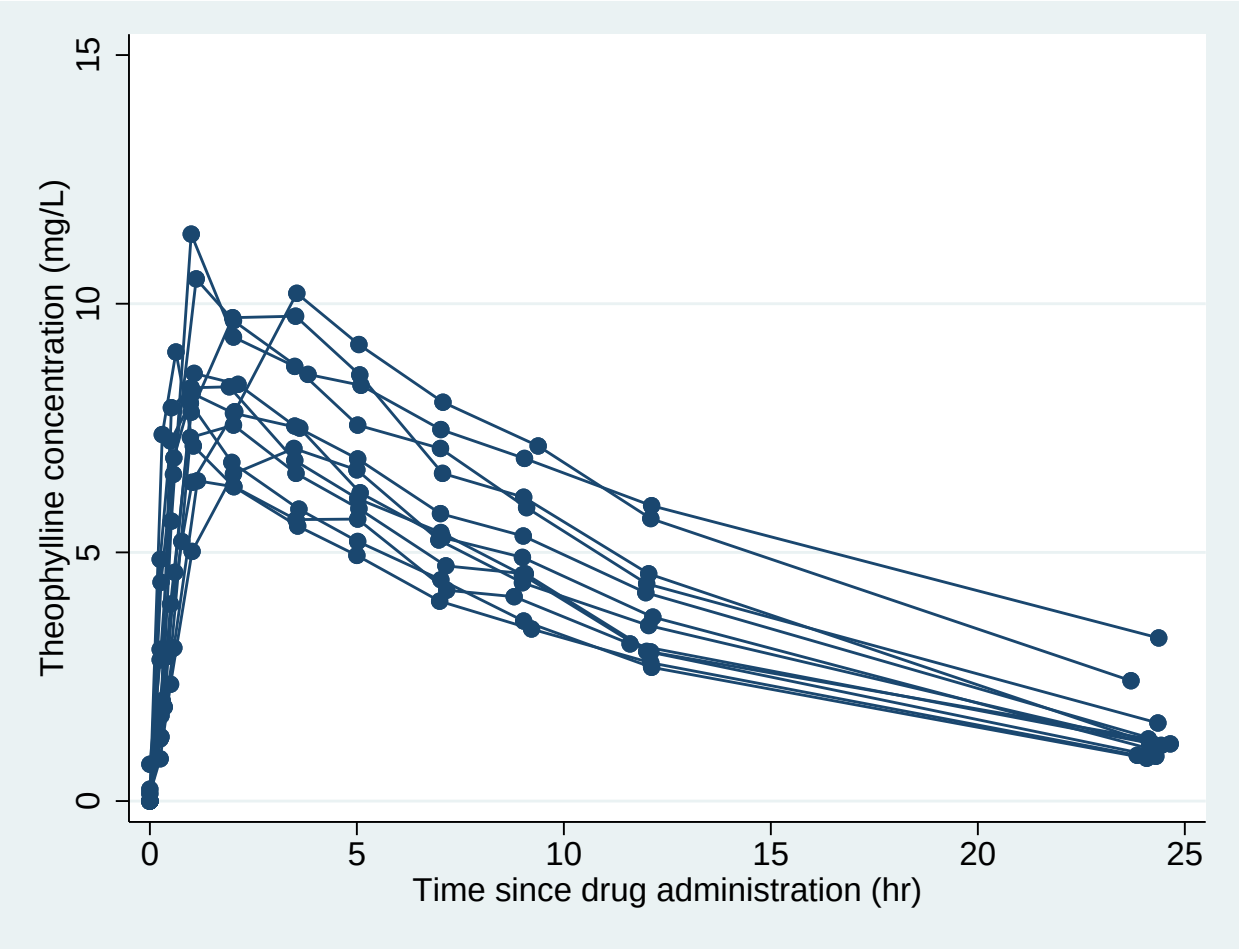
- Pharmacokinetics (PKs) studies the distribution of drugs within the body and is often referred to as the study of "what the body does with a drug".

- It models drug output based on drug input by summarizing concentration-time measurements, while accounting for patient-specific characteristics.

- Consider data on the antiasthmatic agent theophylline (Boeckmann, Sheiner, and Beal [1994] 2011).
- The drug was administered orally to 12 subjects with the dosage (mg/kg) given on a per weight basis.
- Serum concentrations (in mg/L) were obtained at 11 time points per subject over 25 hours following administration.

- Concentration-time profiles of 12 subjects:

```
. webuse theoph
(Theophylline kinetics (Boeckmann et al., [1994] 2011))
. twoway connected conc time, connect(L)
```



- The concentration rises rapidly initially and then decays exponentially.
- In PKs, such pattern is often described by a so-called one-compartment open model with first-order absorption and elimination. (Body is viewed as one "blood compartment".)
- This model is used for drugs that distribute relatively rapidly throughout the body, which is a reasonable assumption for the kinetics of theophylline after oral administration.

- One-compartment open model for theophylline kinetics:

$$\texttt{conc}_{ij} = \frac{\texttt{dose}_j k_e k_{a_j}}{\mathrm{Cl}_j \left(k_{a_j} - k_e\right)} \left\{\exp\left(-k_e \texttt{time}_{ij}\right) - \exp\left(-k_{a_j} \texttt{time}_{ij}\right)\right\} + \epsilon_{ij}$$

  for $i = 1, \ldots, 11$ and $j = 1, \ldots, 12$.

- Parameters: elimination rate constant k_e, and, for each subject j, absorption rate constant k_{a_j} and clearance Cl_j.

- Elimination rate constant describes the rate at which a drug is removed from the body. It is defined as the fraction of drug in the body eliminated per unit time.

- Absorption rate constant describes the rate at which a drug is absorbed by the body.

- Clearance measures the rate at which a drug is cleared from the plasma. It is defined as the volume of plasma cleared of drug per unit time.

- All parameters must be positive, and clearance and absorption rate constant are allowed to vary among subjects:

$$\begin{aligned} \mathrm{Cl}_j &= \exp\left(\beta_0 + u_{0j}\right) \\ k_{a_j} &= \exp\left(\beta_1 + u_{1j}\right) \\ k_e &= \exp\left(\beta_2\right) \end{aligned}$$

  where $u_{0j} \sim N(0, \sigma^2_{u_0})$ and $u_{1j} \sim N(0, \sigma^2_{u_1})$.

- Heteroskedasticity, often present in PK data, is modeled using the power function plus a constant.

$$\mathrm{Var}\left(\epsilon_{ij}\right) = \sigma^2\{(\widehat{\texttt{conc}}_{ij})^{\delta} + c\}^2$$

  Adding a constant avoids the variance of zero at `time` $= 0$, because the concentration is zero at that time.

```
. menl conc =(dose*{ke:}*{ka:}/({cl:}*({ka:}-{ke:})))*(exp(-{ke:}*time)-exp(-{ka:}*time)),
> define(cl: exp({b0}+{U0[subject]}))
> define(ka: exp({b1}+{U1[subject]}))
> define(ke: exp({b2}))
> resvariance(power _yhat) reml
Mixed-effects REML nonlinear regression          Number of obs     =        132
Group variable: subject                          Number of groups  =         12
                                                 Obs per group:
                                                               min =         11
                                                               avg =       11.0
                                                               max =         11
Linear. log restricted-likelihood = -172.44384
          cl:  exp({b0}+{U0[subject]})
          ka:  exp({b1}+{U1[subject]})
          ke:  exp({b2})
```

conc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
/b0	-3.227295	.0619113	-52.13	0.000	-3.348639	-3.105951
/b1	.4354519	.2072387	2.10	0.036	.0292716	.8416322
/b2	-2.453743	.0517991	-47.37	0.000	-2.555267	-2.352218

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
subject: Independent				
var(U0)	.0316416	.014531	.0128634	.0778326
var(U1)	.4500585	.2228206	.1705476	1.187661
Residual variance:				
Power _yhat				
sigma2	.1015759	.086535	.0191263	.5394491
delta	.3106636	.2466547	-.1727707	.7940979
_cons	.7150935	.3745256	.2561837	1.996063

- In the previous `menl` model, we used restricted maximum likelihood estimation (REML) via option `reml` instead of the default maximum likelihood (ML) estimation to account for a moderate number of subjects.

- We specified **nonlinear** functions of model parameters in the `define()` options.

- `menl` fits NLMEMs; see **[ME] menl**.
- `menl` implements the Lindstrom–Bates method, which is based on the linearization of the nonlinear mean function with respect to fixed and random effects.
- `menl` supports ML and REML estimation and provides flexible random-effects and residual covariance structures.
- `menl` supports single-stage and multistage specifications.
- You can predict random effects and their standard errors, group-specific nonlinear parameters, and more after estimation; see **[ME] menl postestimation**.
- NLMEMs are known to be sensitive to initial values. `menl` provides default, but for some models you may need to specify your own. Use option `initial()`.
- NLMEMs are known to have difficulty converging or converging to a local maximum. Trying different initial values may help.

Boeckmann, A. J., L. B. Sheiner, and S. L. Beal. [1994] 2011. *NONMEM Users Guide, Part V: Introductory Guide*. San Francisco: Regents of the University of California. https://nonmem.iconplc.com/nonmem720/guides/v.pdf.

Davidian, M., and D. M. Giltinan. 1995. *Nonlinear Models for Repeated Measurement Data*. Boca Raton, FL: Chapman & Hall/CRC.

Pinheiro, J. C., and D. M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.