A brief outline of STATA commands for Longitudinal data analysis

Nikhil Gupte Honfei Guo and Francesca Dominici

- **xtgee:** estimates regression coefficients using GEE method (this is not a Maximum likelihood estimation approach). It allows to specify the within-subject correlation structures. The command **xtcorr** is for use after **xtgee** and it displays the estimated matrix of within-subject correlation. Please refer to STATA help for detail explanation of the command options.

- **xtgls:** estimates regression coefficients under a linear model using generalized least squares (GLS). This command allows estimation in presence of exponential correlation, with the option, **corr(ar1)**. The default is independent correlation. More specifically:

    xtgls + ols assumes an independent correlation, estimates parameters of the covariance matrix by maximum likelihood, and regression coefficients by WLS: $\widehat{V} = \hat{\sigma}^2 I$ and $\widehat{\beta}$ is a OLS: $\widehat{\beta} = (X'X)^{-1}X'\boldsymbol{Y}$.

    xtgls + igls assumes an independent correlation by default and it is equivalent to **xtgls + ols**

    xtgls + igls + corr(ar1) assumes an exponential correlation, estimates parameters of the covariance matrix by maximum likelihood, and regression coefficients by WLS: $\widehat{V} = \widehat{V}(\hat{\rho})$ and $\widehat{\beta}$ is a WLS: $\widehat{\beta} = (X'\widehat{V}^{-1}X)^{-1}X'\widehat{V}^{-1}\boldsymbol{Y}$. Can be used for equally spaced data only.

    xtgls + igls + corr(ar1) with the option **force** approximates uneqaully spaced data to equally spaced data

- **xtreg** This command estimates linear regression models for longitudinal data. More specifically:

    - **xtreg + fe**: OLS and indipendent correlation model. This is equivalent to **xtgls + ols**
    - **xtreg + mle**: WLS with uniform correlation model. This produces a MLE estimate of ρ and a WLS estimate of β with $\widehat{V}(\hat{\rho})$
    - **xtreg + re**: estimates a linear regression model with random intercept and it is equivalent to **xtreg + mle**.
    - **xtreg + pa:** estimate the regression coefficients by GEE. It is equivalent to **xtgee + family(Guassian)+ link(Identity)+ corr(exchangeable)**

- **xtregar:** estimates a linear regression model with random intercept and under an exponential correlation model. It allows for random effects, serial correlation and measurement error. It provides maximum likelihood estimates of the parameters of the covariance matrix and WLS of the regression coefficients. Can be used for equally spaced data only.

- **prais:** is equivalent to **xtgls + igls + corr(ar1)** - eqaully spaced data only

In summary:

xtgls + ols: OLS

xtgls + igls: OLS

xtgls + igls + corr(AR1): WLS with exponential correlation

xtreg + fe: OLS

xtreg + re: WLS under uniform correlation

xtreg + mle: MLE under uniform correlation

xtregar: WLS under an exponential correlation model + random intercept + measurement error

This is just a summary of the useful STATA commands to analyze longitudinal data. Please refer to the help file of the respective STATA commands and make sure you understand before using them.