Panel data estimation

$$y_{it} = x_{it}\beta + z_{it}\gamma + \theta_i + \epsilon_{it}$$

Introduction

A general model based on panel data would look like

$$y_{it} = x_{it}\beta + \alpha y_{i,t-1} + u_{it} = x_{it}\beta + \alpha y_{i,t-1} + \theta_i + \epsilon_{it}$$

or can be written as

$$\Delta y_{it} = x_{it}\beta + (\alpha - 1)y_{i,t-1} + \theta_i + \epsilon_{it}$$

A common feature of the estimators of interests are the following:

1. A linear model
2. Small T and large N
3. (One left hand side variable y_{it} that is dynamic, depending on its own lagged value)
4. Regressors x_{it} that are not strictly exogenous, meaning that they are correlated with current and past errors.
5. Individual effects θ_i.
6. Heteroskedasticity ε_{it} and serial/auto-correlation within individuals but not across them.
7. The only instruments are internal.

Name	Paper	Moment Conditions
Difference/Transformed GMM	Arellano and Bond (1991)	$E[y_{i,t-2}(\Delta y_{it} - \alpha \Delta y_{i,t-1} - \beta \Delta x_{it})]$
		$E[x_{i,t-2}(\Delta y_{it} - \alpha y_{i,t-1} - \beta \Delta x_{it})]$
System GMM	Arellano and Bover (1995) Blundell and Bond (1998)	$E[\Delta y_{i,t-1}(y_{it} - \alpha y_{i,t-1} - \beta x_{it})]$
		$E[\Delta x_{i,t-1}(y_{it} - \alpha y_{i,t-1} - \beta x_{it})]$

Example

Labor employment in firms: The hiring decision is costly. We expect firms to adjust their labor with delay to changes in other input (capital), input prices (wages), and demand for output. > The process of adjustment to changes in these factors may depend both on the > 1. passage of time, which indicates **lagged versions of these factors** as regressors, and on the > 2. difference between equilibrium employment level and the previous year's actual level, which argues for a dynamic model, in which **lags of the dependent variable** are also regressors.

```r
data("EmplUK", package = "plm")

## Arellano and Bond (1991), table 4 col. b
z1 <- pgmm(log(emp) ~ lag(log(emp), 1:2) + lag(log(wage), 0:1)
           + log(capital) + lag(log(output), 0:1) | lag(log(emp), 2:99),
            data = EmplUK, effect = "twoways", model = "twosteps")
summary(z1, robust = FALSE)

## Blundell and Bond (1998) table 4 (cf. DPD for OX p. 12 col. 4)
z2 <- pgmm(log(emp) ~ lag(log(emp), 1)+ lag(log(wage), 0:1) +
           lag(log(capital), 0:1) | lag(log(emp), 2:99) +
           lag(log(wage), 2:99) + lag(log(capital), 2:99),
```

```r
           data = EmplUK, effect = "twoways", model = "onestep",
           transformation = "ld")
summary(z2, robust = TRUE)
```

Standard errors

Estimation

If we have homoskedastic error term ϵ_{it}, then the WG estimation is consistent and efficient. Otherwise, first difference and use GLS on the differenced data.

Inference

The view of WG as a dummy-variable least-squares regression may suggest using a traditional (cross-sectional) 'White'-type variance formula to deal with heteroskedasticity.