Lab 6: Dynamic Panels
GECO 6281 Advanced Econometrics 1

Patrick Mokre

Fall 2020

Dynamic Panels

Panel Data is observed over time, including a lagged variable or **auto-regressive** term is an intuitive modeling choice.

$$y_t = \alpha + \phi_1 y_{t-1} + \beta x_t + \epsilon_t \text{ autoregressive process}$$
$$y_{i,t} = \alpha_i + \gamma_1 y_{i,t-1} + \beta x_{i,t} + \epsilon_{i,t} \text{ autoregressive panel data}$$

Caution: OLS with a lagged variable and serially correlated errors leads to **inconsistent estimators** (as it does in the non-panel case).

When estimating a dynamic panel using fixed effects, **first differencing** must be used rather than **mean differencing**.

Arellano-Bond instrumentalization allows for efficient FD estimation in a dynamic model. Estimated parameters are consistent for both FE and RE models.

An AR(p) panel model

$$y_{it} = \gamma_1 y_{i,t-1} + ... + \gamma_p y_{i,t-p} + \beta x_{it} + \alpha_i + \epsilon_{it}$$

3 channels of over-time correlation in y_i: **true state dependence** (directly $y_{i,t-1} \rightarrow y_{i,t}$), **observed heterogeneity** (directly through covariates $x_{i,t-1} \rightarrow x_{i,t} \rightarrow y_{i,t}$, or **unobserved heterogeneity** indirectly through α_i.

The within estimator (mean difference FE) is inconsistent with lags, as $y_{it} - \bar{y_i}$ is correlated with $\epsilon_{it} - \bar{\epsilon_i}$.

IV estimation using lags is also inconsistent, as $y_{i,t-s}$ is correlated with $\bar{\epsilon_i}$, and thus $\epsilon_{it} - \bar{\epsilon_i}$.

While first difference estimation will be inconsistent, using **appropriate lags of y_{it} as instruments** in FD estimation leads to consistent estimates.

First Difference Model

$$\Delta y_{it} = \gamma_1 \Delta y_{i,t-1} + ... + \gamma_p \Delta y_{i,t-p} + \Delta x'_{it}\beta + \Delta \epsilon_{it}$$
$$\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$$

$\Delta y_{i,t-1}$ is correlated with $\Delta \epsilon_{i,t}$, but $y_{i,t-s}$ is not $\forall s > 2$. Anderson and Hsiao (1981) propose instruments using the second lag, while Arellano and Bond (1991) showed that efficiency is increased by using even more lags, and that consistency holds under the assumption of **no serial correlation** in ϵ.

Regarding independent variables, you distinguish three categories. **Strictly exogenous covariates**, **weakly exogenous covariates** (correlated with past, but not with contemporaneous and future values of ϵ_{it}) and **temporarily endogenous covariates** (correlated with past and contemporaneous, but not future error terms).

You use instruments accordingly with past values, and can also include external instruments.

Anderson Hsiao Instrumentalization

OLS estimates in short and broad panels (little T large N) will be upward biased due to correlation of the lagged coefficient with the error term.

Fixed effect estimate for laged covariate will be downward biased by size 1/T (**“Nickell bias”**)

Anderson & Hsiao denotes a first-difference model, but instrumentalize the first difference with 2- and 3-period lag differences.

```
regress n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, cluster
estimates store OLS
xtreg n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, fe cluste
estimates store FE
ivregress 2sls D.n (D.nL1 = nL2) D.(nL2 w wL1 k kL1 kL2 ys
estimates store ahsiao1

esttab OLS FE
```

Anderson-Hsiao: Results 1

```
esttab ahsiao1

----------------------------
  (1)
D.n
----------------------------
  D.nL1              2.308
(1.17)

D.nL2              -0.224
(-1.25)

D.w                -0.810**
  (-3.10)

D.wL1               1.422
(1.21)
```

Anderson-Hsiao: Results 2

D.kL2	-0.213
(-0.89)	
D.ys	0.991*
(2.14)	
D.ysL1	-1.938
(-1.35)	
D.ysL2	0.487
(0.96)	

Anderson-Hsiao: Results 3

D.yr1979	0.0467
(1.04)	
D.yr1980	0.0761
(1.22)	
D.yr1981	0.0226
(0.41)	
D.yr1982	0.0128
(0.23)	
D.yr1983	0.00991
(0.22)	
_cons	0.0159
(0.58)	

Arellano-Bond Instrumentalization 1

In dynamic FE models, note the difference between **mean differencing** $x_{it} - \bar{x}_i$ and **first differencing** $x_{it} - x_{i(t-1)}$.

Remember: when you include serially correlated errors and/or lagged dependent (autoregressive) variables, OLS estimation of an FE model is **inconsistent**.

Arellano-Bond estimation uses a sufficient number of lags as instruments for dependent variables, which is often more efficient than OLS estimation.

You just made the step to **dynamic panel modeling**.

Arellano-Bond Instrumentalization 2

$$y_{it} = \gamma_1 y_{i(t-1)} + ... + \gamma_p y_{i(t-p)} + x'_{it}\beta + \alpha_i + \epsilon_{it}$$

Note: Both **within-estimation** and **lag instrumentalization** will be inconsistent for correlation between mean differences $y_{it} - \bar{y}_i$ or lags $y_{i(t-p)}$ and $\epsilon_{it} - \bar{\epsilon}_i$. For FD estimation, assume that ϵ_{it} is **serially uncorrelated**.

$$\Delta y_{it} = \gamma_1 \Delta y_{i(t-1)} + ... + \gamma_{p-1} \Delta y_{i(t-p)} + \Delta x'_{it}\beta + \Delta\epsilon_{it}$$

You can instrument for $\Delta y_{i(t-1)}$ using enough lags $y_{i(t-2),...,y_{i(t-s)}}$, and Δx_{it} by x_{it} themselves, if x_{it} are exogenous. If x_{it} are not exogenous, they can be instrumented by enough lags of themselves.

Arellano Bond instrumentalization in STATA

```
. xtabond lwage, lags(2) vce(robust)

Arellano-Bond dynamic panel-data estimation     Number of o
Group variable: id                              Number of g
Time variable: t
Obs per group:
  min =          4
avg =          4
max =          4

Number of instruments =     15                  Wald chi2(2
Prob > chi2         =     0.0000
One-step results
(Std. Err. adjusted for clustering on id)
---------------------------------------------------------------
  |                 Robust
lwage |       Coef.    Std. Err.       z    P>|z|       [95% Con
```