Another Crack at Dynamic Models in Chapter 8 (with STATA)

```
infile state year price pop popo16 cpi ndi c pimin using c:/data/baltagi/cigarettes.txt

tsset state year, yearly
gen cpc=(c*pop)/popo16
gen lcpc=L.cpc
gen lnc=log(cpc)
gen lnlc=log(lcpc)
gen lnp=log((price/cpi)*100)
gen lnpn=log((pimin/cpi)*100)
gen lny=log((ndi/cpi)*100)

* OLS (w/o time FE, following * p. 157)
reg lnc lnlc lnp lnpn lny

quietly tabulate year, gen(t)

* Within
* drop 2 year indicators, one for identification
* and the first year, b/c of the lagged dv on the RHS
xtreg lnc lnlc lnp lnpn lny t3-t30, fe i(state)

gen lnlp=L.lnp
gen lnlpn=L.lnpn
gen lnly=L.lny

* 2SLS (w/o time FE, following * p. 157)
ivreg lnc lnp lnpn lny (lnlc=lnlp lnlpn lnly)

* Within 2SLS (with time FE)
xtivreg lnc lnp lnpn lny t3-t30 (lnlc=lnlp lnlpn lnly), fe i(state)

* Arellano-Bond (with time FE)
* I got matsize error at first, so added command immediately below
set matsize 500

xtabond lnc lnp lnpn lny t3-t30, lag(1) noconstant
* this should match p. 159
xtabond lnc lnp lnpn lny t3-t30, lag(1) twostep noconstant
```

Wed Dec 7, 2005

```
. * OLS (w/o time FE, following * p. 157)
. reg lnc lnlc lnp lnpn lny

      Source |       SS       df       MS              Number of obs =    1334
-------------+------------------------------           F(  4,  1329) = 9089.46
       Model |  64.3926449     4  16.0981612           Prob > F      =  0.0000
    Residual |  2.35376627  1329  .001771081           R-squared     =  0.9647
-------------+------------------------------           Adj R-squared =  0.9646
       Total |  66.7464111  1333  .050072326           Root MSE      =  .04208

------------------------------------------------------------------------------
         lnc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnlc |   .9694942   .0061489   157.67   0.000     .9574316    .9815569
         lnp |  -.0901512   .0145795    -6.18   0.000    -.1187525   -.0615498
        lnpn |   .0240347   .0131562     1.83   0.068    -.0017746     .049844
         lny |  -.0306788   .0060279    -5.09   0.000    -.0425041   -.0188535
       _cons |   .7240693   .0729962     9.92   0.000     .5808688    .8672697
------------------------------------------------------------------------------
```

√ matches OLS row p. 157

```
. quietly tabulate year, gen(t)

.
. * Within
. * drop 2 year indicators, one for identification
. * and the first year, b/c of the lagged dv on the RHS
. xtreg lnc lnlc lnp lnpn lny t3-t30, fe i(state)

Fixed-effects (within) regression               Number of obs      =      1334
Group variable (i): state                       Number of groups   =        46

R-sq:  within  = 0.9381                         Obs per group: min =        29
       between = 0.9850                                        avg =      29.0
       overall = 0.9648                                        max =        29

                                                F(32,1256)         =    594.64
corr(u_i, Xb)  = 0.4009                         Prob > F           =    0.0000

------------------------------------------------------------------------------
         lnc |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnlc |   .8333826   .0125616     66.34   0.000     .8087387    .8580266
         lnp |   -.298598    .023601    -12.65   0.000    -.3448997   -.2522963
        lnpn |   .0340452   .0274653      1.24   0.215    -.0198377    .0879282
         lny |   .1002794   .0238499      4.20   0.000     .0534894    .1470694
          t3 |   .0312327   .0077092      4.05   0.000     .0161084     .046357
```

[SNIP]

√ matches Within row p. 157

```
. gen lnlp=L.lnp
(46 missing values generated)
. gen lnlpn=L.lnpn
(46 missing values generated)
. gen lnly=L.lny
(46 missing values generated)
.
. * 2SLS (w/o time FE, following * p. 157)
. ivreg lnc lnp lnpn lny (lnlc=lnlp lnlpn lnly)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    1334
-------------+------------------------------           F(  4,  1329) = 2398.66
       Model |  63.7233971     4  15.9308493           Prob > F      =  0.0000
    Residual |  3.02301402  1329  .002274653           R-squared     =  0.9547
-------------+------------------------------           Adj R-squared =  0.9546
       Total |  66.7464111  1333  .050072326           Root MSE      =  .04769

------------------------------------------------------------------------------
         lnc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnlc |   .8499651   .0335508    25.33   0.000     .7841468    .9157833
         lnp |  -.2050105   .0356031    -5.76   0.000    -.2748548   -.1351662
        lnpn |   .0523477   .0168147     3.11   0.002     .0193615     .085334
         lny |  -.0169926   .0077967    -2.18   0.029    -.0322878   -.0016975
       _cons |   1.601765   .2547925     6.29   0.000     1.101926    2.101605
------------------------------------------------------------------------------
Instrumented:  lnlc
Instruments:   lnp lnpn lny lnlp lnlpn lnly
------------------------------------------------------------------------------
```

√ matches 2SLS row p. 157

```
. * Within 2SLS (with time FE)
. xtivreg lnc lnp lnpn lny t3-t30 (lnlc=lnlp lnlpn lnly), fe i(state)

Fixed-effects (within) IV regression        Number of obs      =        1334
Group variable: state                       Number of groups   =          46

R-sq:  within  = 0.9213                     Obs per group: min =          29
       between = 0.9369                                    avg =        29.0
       overall = 0.9230                                    max =          29

                                            Wald chi2(32)      =   2.08e+07
corr(u_i, Xb)  = 0.4579                     Prob > chi2        =     0.0000

------------------------------------------------------------------------------
         lnc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnlc |    .601629    .035302    17.04   0.000     .5324384    .6708196
         lnp |  -.4956818   .0382645   -12.95   0.000    -.5706788   -.4206847
        lnpn |  -.0159481   .0317401    -0.50   0.615    -.0781575    .0462613
         lny |   .1893821   .0296233     6.39   0.000     .1313215    .2474427
          t3 |   .0165957    .008928     1.86   0.063    -.0009029    .0340942
          t4 |    .004696   .0090059     0.52   0.602    -.0129553    .0223473
[SNIP]
```

√ matches Within-2SLS row p. 157

```
. * Arellano-Bond (with time FE)
. * I got matsize error at first, so added command immediately below
.
. set matsize 500

Current memory allocation

                    current                                 memory usage
    settable          value     description                 (1M = 1024k)
    --------------------------------------------------------------------
    set maxvar         5000     max. variables allowed           1.733M
    set memory          10M     max. data space                 10.000M
    set matsize         500     max. RHS vars in models          1.949M
                                                            -----------
                                                                13.682M

.
. xtabond lnc lnp lnpn lny t3-t30, lag(1) noconstant

Arellano-Bond dynamic panel-data estimation     Number of obs      =      1288
Group variable (i): state                       Number of groups   =        46

                                                Wald chi2(.)       =         .

Time variable (t): year                         Obs per group: min =        28
                                                               avg =        28
                                                               max =        28
```

(⇓ continued over ⇓)

```
One-step results
------------------------------------------------------------------------------
     D.lnc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnc |
         LD. |   .8428668   .0162092    52.00   0.000     .8110972    .8746363
         lnp |
         D1. |  -.3772291   .0323381   -11.67   0.000    -.4406106   -.3138475
        lnpn |
         D1. |  -.0161502   .0418921    -0.39   0.700    -.0982572    .0659568
         lny |
         D1. |   .1394493   .0364131     3.83   0.000     .0680808    .2108177
          t3 |
         D1. |    .028061   .0080258     3.50   0.000     .0123307    .0437912
          t4 |
[SNIP]...
         t30 |
         D1. |   .0149336   .0271575     0.55   0.582    -.0382941    .0681614
------------------------------------------------------------------------------
Sargan test of over-identifying restrictions:
         chi2(405) =   534.01     Prob > chi2 = 0.0000

Arellano-Bond test that average autocovariance in residuals of order 1 is 0:
         H0: no autocorrelation   z = -16.23   Pr > z = 0.0000
Arellano-Bond test that average autocovariance in residuals of order 2 is 0:
         H0: no autocorrelation   z =   2.35   Pr > z = 0.0188
```

√ matches GMM-one-step row p. 157

```
. * this should match p. 159
. xtabond lnc lnp lnpn lny t3-t30, lag(1) twostep noconstant

Arellano-Bond dynamic panel-data estimation     Number of obs      =      1288
Group variable (i): state                       Number of groups   =        46

                                                Wald chi2(.)       =         .

Time variable (t): year                         Obs per group: min =        28
                                                               avg =        28
                                                               max =        28

Two-step results
------------------------------------------------------------------------------
       D.lnc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnc |
         LD. |   .8036634   .2200741     3.65   0.000      .372326    1.235001
         lnp |
         D1. |  -.3786943   .0471325    -8.03   0.000    -.4710722   -.2863163
        lnpn |
         D1. |  -.0197169   .0495157    -0.40   0.690    -.1167659    .0773321
         lny |
         D1. |   .2391497   .2778919     0.86   0.389    -.3055084    .7838078
          t3 |
         D1. |   .0238788   .0231137     1.03   0.302    -.0214232    .0691808
          t4 |
         D1. |  -.0062309   .0287669    -0.22   0.829    -.0626129    .0501512
          t5 |
         D1. |   .0050715   .0405765     0.12   0.901     -.074457       .0846
          t6 |
         D1. |    -.00913   .0468145    -0.20   0.845    -.1008847    .0826248
          t7 |
         D1. |  -.0198448   .0546753    -0.36   0.717    -.1270063    .0873167
          t8 |
         D1. |  -.0488936   .0621244    -0.79   0.431     -.170655    .0728679
```

t9						
D1.	-.0030379	.0759863	-0.04	0.968	-.1519683	.1458925
t10						
D1.	-.0003904	.0858245	-0.00	0.996	-.1686034	.1678226
t11						
D1.	-.0454377	.1008626	-0.45	0.652	-.2431248	.1522494
t12						
D1.	-.0501865	.0908816	-0.55	0.581	-.2283112	.1279382
t13						
D1.	-.0722568	.0858978	-0.84	0.400	-.2406133	.0960997
t14						
D1.	-.0627064	.0905982	-0.69	0.489	-.2402758	.1148629
t15						
D1.	-.0932763	.0961624	-0.97	0.332	-.2817512	.0951985
t16						
D1.	-.0902554	.1064234	-0.85	0.396	-.2988414	.1183307
t17						
D1.	-.1328357	.1049618	-1.27	0.206	-.338557	.0728856
t18						
D1.	-.1338162	.0982222	-1.36	0.173	-.3263282	.0586958
t19						
D1.	-.1496917	.0984574	-1.52	0.128	-.3426647	.0432814
t20						
D1.	-.1514073	.0964368	-1.57	0.116	-.3404198	.0376053
t21						
D1.	-.1336293	.1054678	-1.27	0.205	-.3403424	.0730839
t22						
D1.	-.1251314	.1242345	-1.01	0.314	-.3686267	.1183638
t23						
D1.	-.0931677	.1372333	-0.68	0.497	-.3621401	.1758047
t24						
D1.	-.0954667	.1482346	-0.64	0.520	-.3860011	.1950677
t25						
D1.	-.1008324	.1580258	-0.64	0.523	-.4105572	.2088924
t26						
D1.	-.1019861	.1690787	-0.60	0.546	-.4333742	.2294021

```
         t27 |
         D1. |  -.1031526   .1812312    -0.57   0.569    -.4583591     .252054
         t28 |
         D1. |  -.1002765   .1922681    -0.52   0.602    -.4771151    .2765621
         t29 |
         D1. |  -.0833889   .1990996    -0.42   0.675    -.4736169     .306839
         t30 |
         D1. |  -.0441254   .2174354    -0.20   0.839     -.470291    .3820401
------------------------------------------------------------------------------
Warning: Arellano and Bond recommend using one-step results for
         inference on coefficients

Sargan test of over-identifying restrictions:
         chi2(405) =    15.40     Prob > chi2 = 1.0000

Arellano-Bond test that average autocovariance in residuals of order 1 is 0:
         H0: no autocorrelation   z =  -3.53   Pr > z = 0.0004
Arellano-Bond test that average autocovariance in residuals of order 2 is 0:
         H0: no autocorrelation   z =   1.63   Pr > z = 0.1027
```

√ matches GMM-two-step row p. 157

SUNDRY COMMENTS ON (EVER-POPULAR) BECK AND KATZ METHOD

- Beck and Katz suggest estimating models by OLS with PCSEs and lagged dependent variable
- PCSEs don't accommodate heterogeneity and require large T and small N (if N bigger than T, you get singular covariance matrix)
- Arellano already identified robust SEs long before (extension of White)—truly robust to heterogeneity in DGP
- large T and small N → SURE with cross-equation restrictions, FE and RE are *very* similar in that situation and can be quite different from OLS
- Should focus on problem of unobserved heterogeneity that cause estimates to be biased

RE estimation is based on the structure of the variance-covariance matrix of the residuals (e.g. $\varepsilon_{it} = \mu_i + \upsilon_{it}$ in the usual one-way case), $\Omega_{NT\times NT}$ given the usual simplifying assumptions of homoscedasticity and no serial correlation, except as implied by the variance components themselves. In the one-way (spatial effects only) case,

$$E(uu') = \Omega = \sigma_\mu^2(I_N \otimes J_T) + \sigma_v^2(I_N \otimes I_T) \qquad \text{with } \mathrm{cov}(u_{it},u_{js}) = \sigma_\mu^2 + \sigma_v^2 \text{ for } i=j,\ t=s$$
$$\text{and } = \sigma_\mu^2 \text{ (not 0) for } i=j,\ t\neq s$$

$$\text{In turn, } \rho = \mathrm{corr}(u_{it},u_{js}) = 1 \text{ for for } i=j,\ t=s$$
$$\text{and } = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_v^2) \text{ for } i=j,\ t\neq s$$

Devise weighting matrices based on Ω^{-1} to transform the data matrices such that the transformed Ω is spherical. Then, use estimates of the variance components σ_μ^2, σ_λ^2, and σ_v^2 to do FGLS.

Dynamic-model estimators start with what Ω should look like
e.g. simplest case, y_{it} is a function of x_{it} and also y_{it-1}, but not x_{it-1}, y_{it-2}, etc. and otherwise the usual simplifying assumptions about homoscedasticity and no-serial-correlation obtain

B&K (1995) targets the Parks (1967) method, which is essentially using OLS residuals to estimate Ω and then estimating a model with FGLS. But the key problem with Parks's method is too many parameters.

assumptions:

- variation across units in residual variance component: $E(u_{it}^2)=E(u_{is}^2) \forall\, t \neq s$
- contemporaneous (but not lagged) spatial correlation: $E(u_{it},u_{jt}) = \sigma_{ij} \;\forall\, i \neq j$

  $= 0 \quad \forall\, i \neq j,\ t \neq s$

- $E(u_{it},u_{is})=\rho$ or $E(u_{it},u_{is})=\rho_i$

$$u_{it} = \rho_i\, u_{it-1} + v_{it} \qquad w/\ v_{it} \sim IID(0,\sigma_v^2)$$

(with only T observations to estimate each ρ_i, it might be infeasible to allow rho to vary across units)

$$\mathbf{\Omega} = \begin{bmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \Sigma \end{bmatrix} = \mathbf{\Sigma} \otimes \mathbf{I}_T \quad \text{with} \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \cdots & \sigma_N^2 \end{bmatrix}$$

so N(N-1)/2 off-diag. parameters to estimate

Griliches (1999) on Maddala:

"GS has an unassuming and quiet way. But he also has something that is close to perfect pitch in econometrics: when he sees work that is 'off,' that somehow strikes the wrong note, it bothers him, and his irritation often produces pearls of papers. While much of his work is constructive, much is also critical of many current fads in econometrics."

Maddala (1999 *Pol Anal*) on Beck & Katz:

"[SUR and the Parks method] involve the estimation of so many parameters that they have fallen into disrepute and are almost never used. The introduction of lagged dependent variables creates problems with OLS estimation if the errors are serially correlated. This is why instrumental variables estimation is the preferred method in the estimation of dynamic panel data models. This area is recent and is exploding, because there are many thorny problems to be tackled…The procedure of using OLS and reporting the 'panel corrected' standard errors is sweeping the problems under the rug….Their solution can also be categorized in 'what not to do' if there are lagged dependent variables."

- B&K are right about problems w/ Parks & Kmenta methods, but OLS with LDV and PCSE is no magic bullet
- pay more attention to pooling and specification, don't use LDV w/o good reason

Wilson & Butler "A Lot More To Do: Promise and Peril of Panel Data" (2004 ms)

0. Static $Y_t = \alpha + \beta_0 X_t + u_t$

1. AR(·) $Y_t = \alpha + \beta_0 X_t + u_t \quad , \quad u_t = \rho u_{t-1} + e_t$

2. DL(·) $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + u_t$

(BAK) 3. LDV $Y_t = \alpha + \beta_0 X_t + \gamma_1 Y_{t-1} + u_t$

4. ARDL(·) $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \gamma_1 Y_{t-1} + u_t$

5. FD $Y_t - Y_{t-1} = \beta_0 (X_t - X_{t-1}) + u_t$

Assuming 3 given DGP of 1 introduces Bias

see Achen (2000) "Why Lagged Dep. Vars Can Suppress the Explanatory Power of other ind. Vars"

also see Green, Kim, and Yoon (2001 *IO*)

Beck and Katz

```
. tsset firm year, yearly
       panel variable:  firm, 1 to 10
        time variable:  year, 1935 to 1954

. xtpcse i f c

Linear regression, correlated panels corrected standard errors (PCSEs)

Group variable:   firm                        Number of obs      =       200
Time variable:    year                        Number of groups   =        10
Panels:           correlated (balanced)       Obs per group: min =        20
Autocorrelation:  no autocorrelation                         avg =        20
                                                             max =        20
Estimated covariances      =        55        R-squared          =    0.8124
Estimated autocorrelations =         0        Wald chi2(2)       =    637.41
Estimated coefficients     =         3        Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |           Panel-corrected
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .1155622   .0072124    16.02   0.000      .101426    .1296983
           c |   .2306785   .0278862     8.27   0.000     .1760225    .2853345
       _cons |  -42.71437   6.780965    -6.30   0.000    -56.00482   -29.42392
------------------------------------------------------------------------------
```

(also see page 399 of STATA 7 manual Su-Z for output above, and below)

```
. xtgls i f c, p(correlated)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation;   no autocorrelation

Estimated covariances      =        55          Number of obs      =        200
Estimated autocorrelations =         0          Number of groups   =         10
Estimated coefficients     =         3          Time periods       =         20
                                                Wald chi2(2)       =    3738.07
Log likelihood             = -879.4274          Prob > chi2        =     0.0000

------------------------------------------------------------------------------
           i |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .1127515   .0022364     50.42   0.000       .1083683    .1171347
           c |   .2231176   .0057363     38.90   0.000       .2118746    .2343605
       _cons |  -39.84382   1.717563    -23.20   0.000      -43.21018   -36.47746
------------------------------------------------------------------------------

. xtgls i f c, p(h)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   no autocorrelation

Estimated covariances      =        10          Number of obs      =        200
Estimated autocorrelations =         0          Number of groups   =         10
Estimated coefficients     =         3          Time periods       =         20
                                                Wald chi2(2)       =     669.69
Log likelihood             = -1037.152          Prob > chi2        =     0.0000
```

```
------------------------------------------------------------------------------
           i |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .1116328   .0049823    22.41   0.000      .1018676     .121398
           c |   .1537718   .0125707    12.23   0.000      .1291336      .17841
       _cons |  -21.44348   3.901219    -5.50   0.000     -29.08973   -13.79723
------------------------------------------------------------------------------

. xtgls i f c, p(het) corr(ar1)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   common AR(1) coefficient for all panels  (0.9261)

Estimated covariances      =        10          Number of obs      =       200
Estimated autocorrelations =         1          Number of groups   =        10
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    107.43
Log likelihood             = -856.2668          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           i |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .0715306   .0087269     8.20   0.000      .0544262     .088635
           c |   .1405652   .0314945     4.46   0.000      .0788371    .2022933
       _cons |  -1.979683   6.781349    -0.29   0.770     -15.27088    11.31152
------------------------------------------------------------------------------

.
```

```
. xtgee i f c, robust

Iteration 1: tolerance = .33946478
Iteration 2: tolerance = .00362671
Iteration 3: tolerance = .00001665
Iteration 4: tolerance = 7.455e-08

GEE population-averaged model                   Number of obs      =       200
Group variable:                       firm      Number of groups   =        10
Link:                             identity      Obs per group: min =        20
Family:                           Gaussian                     avg =      20.0
Correlation:                  exchangeable                     max =        20
                                                Wald chi2(2)       =     71.77
Scale parameter:                  9203.121      Prob > chi2        =    0.0000

                             (standard errors adjusted for clustering on firm)
------------------------------------------------------------------------------
             |             Semi-robust
           i |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .1097626   .0136213     8.06   0.000     .0830655    .1364598
           c |    .307942   .0548629     5.61   0.000     .2004128    .4154712
       _cons |   -57.7672   24.55615    -2.35   0.019    -105.8964   -9.638037
------------------------------------------------------------------------------
```

Grunfeld Again and Again, Part 1

	OLS	**avg(10×OLS) unweighted**	**Between**	**Within (oneway)**	**R.E. (Am) (oneway)**
f (market value)	0.116 (0.006)	0.091 (0.042)	0.135 (0.029)	.110 (0.012)	0.110 (0.010)
c (capital stock)	0.231 (0.025)	0.205 (0.067)	0.032 (0.191)	0.310 (0.017)	0.308 (0.017)
constant	-42.7 (9.5)	-21.4 (58.9)	-8.53 (47.5)	-58.7 (12.5)	-7.92 (4.00)
firm indicators	--	--	--	yes	--
year indicators	--	--	--	--	--
i_(t-1) lagged investment	--	--	--	--	--

Grunfeld Again and Again, Part 2

	OLS	**B&K PCSE**	**FE PCSE**	**GLS (robust)**	**FGLS het,AR(1)**
f (market value)	0.116 (0.006)	0.116 (0.007)	0.110 (0.018)	0.110 (0.014)	0.072 (0.009)
c (capital stock)	0.231 (0.025)	0.231 (0.028)	0.310 (0.025)	0.308 (0.055)	0.141 (0.031)
constant	-42.7 (9.5)	-42.7 (6.8)	--	-57.8 (24.6)	-2.00 (6.8)
firm indicators	--	--	yes	--	--
year indicators	--	--		--	--
i_(t-1) lagged investment	--	--		--	--

Grunfeld Again and Again, Part 3

	ArBond robust	**A&H FDIV**	**B&K PCSE**	**Within-2**	**R.E. (Am) two-way**
f (market value)	0.106 (0.013)	0.092 (0.014)	0.117 (0.008)	0.118 (0.014)	0.112 (0.011)
c (capital stock)	0.139 (0.031)	0.671 (0.511)	0.220 (0.037)	0.358 (0.023)	0.325 (0.019)
constant	-1.53 (1.36)	0	0	0	-8.00 (3.87)
firm indicators	--	--	--	yes	--
year indicators	--	yes	yes	yes	--
i_(t-1) lagged investment	0.668 (0.080)	-0.601 (0.916)	--	--	--

Pooled Intercept, Varying Slopes

```
. reg i f  _IfirXf_2 _IfirXf_3 _IfirXf_4 _IfirXf_5 _IfirXf_6 _IfirXf_7 _IfirXf_8 _IfirXf_9
_IfirXf_10 c _I
> firXc_2  _IfirXc_3 _IfirXc_4 _IfirXc_5 _IfirXc_6 _IfirXc_7 _IfirXc_8 _IfirXc_9
_IfirXc_10
```

Source	SS	df	MS
Model	9017883.96	20	450894.198
Residual	342059.957	179	1910.94948
Total	9359943.92	199	47034.8941

Number of obs	=	200
F(20, 179)	=	235.95
Prob > F	=	0.0000
R-squared	=	0.9635
Adj R-squared	=	0.9594
Root MSE	=	43.714

i	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
f	.0872213	.0046097	18.92	0.000	.0781249	.0963177
_IfirXf_2	.0703704	.0116173	6.06	0.000	.047446	.0932949
_IfirXf_3	-.0593291	.0103507	-5.73	0.000	-.0797542	-.0389039
_IfirXf_4	.0000671	.026621	0.00	0.998	-.0524643	.0525984
_IfirXf_5	.2429773	.233709	1.04	0.300	-.218202	.7041566
_IfirXf_6	.0634017	.1377529	0.46	0.646	-.2084268	.3352301
_IfirXf_7	.0421774	.167553	0.25	0.802	-.2884559	.3728107
_IfirXf_8	-.0132334	.03838	-0.34	0.731	-.0889688	.062502
_IfirXf_9	.0007175	.1243338	0.01	0.995	-.2446311	.2460661
_IfirXf_10	.0831719	.3224964	0.26	0.797	-.553212	.7195557
c	.3825272	.0172113	22.23	0.000	.3485641	.4164903
_IfirXc_2	.0020079	.0661753	0.03	0.976	-.1285762	.132592
_IfirXc_3	-.2303208	.0430791	-5.35	0.000	-.315329	-.1453126
_IfirXc_4	-.0676998	.0962194	-0.70	0.483	-.2575702	.1221705
_IfirXc_5	-.3924791	.1066863	-3.68	0.000	-.6030038	-.1819544
_IfirXc_6	-.3404627	.4978163	-0.68	0.495	-1.322806	.6418809
_IfirXc_7	-.2529759	.07327	-3.45	0.001	-.3975599	-.1083918
_IfirXc_8	-.3226458	.2256228	-1.43	0.154	-.7678685	.1225769
_IfirXc_9	-.2979863	.1340945	-2.22	0.028	-.5625958	-.0333768

```
  _IfirXc_10 |   .2402446   2.985162     0.08   0.936    -5.650392    6.130881
       _cons |  -12.8802   12.96132    -0.99   0.322    -38.45684    12.69644
------------------------------------------------------------------------------
```

Pooled Slopes, Varying Intercepts

```
. reg i f c f1 f2 f3 f4 f5 f6 f7 f8 f9 f10, nocons

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F( 12,   188) =  391.97
       Model |   13097228    12  1091435.66            Prob > F      =  0.0000
    Residual | 523478.114   188  2784.45805            R-squared     =  0.9616
-------------+------------------------------           Adj R-squared =  0.9591
       Total | 13620706.1   200  68103.5304            Root MSE      =  52.768

------------------------------------------------------------------------------
           i |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .1101238   .0118567     9.29   0.000     .0867345    .1335131
           c |   .3100653   .0173545    17.87   0.000     .2758308    .3442999
          f1 |  -70.29669   49.70796    -1.41   0.159    -168.3537    27.76034
          f2 |   101.9058   24.93832     4.09   0.000     52.71093    151.1007
          f3 |  -235.5718   24.43162    -9.64   0.000    -283.7672   -187.3765
          f4 |  -27.80929   14.07775    -1.98   0.050    -55.57995    -.038631
          f5 |  -114.6168   14.16543    -8.09   0.000    -142.5604   -86.67319
          f6 |  -23.16129   12.66874    -1.83   0.069    -48.15244    1.829856
          f7 |  -66.55347   12.84297    -5.18   0.000    -91.88833   -41.21862
          f8 |  -57.54565   13.99315    -4.11   0.000    -85.14941    -29.9419
          f9 |  -87.22227   12.89189    -6.77   0.000    -112.6536   -61.79091
         f10 |  -6.567843   11.82689    -0.56   0.579    -29.89831    16.76262
------------------------------------------------------------------------------
```

Pooled Slopes & Intercept

```
. reg i f c

      Source |       SS       df       MS              Number of obs =     200
```

```
-------------+------------------------------           F(  2,   197) =  426.58
       Model |  7604093.48     2  3802046.74           Prob > F      =  0.0000
    Residual |  1755850.43   197  8912.94636           R-squared     =  0.8124
-------------+------------------------------           Adj R-squared =  0.8105
       Total |  9359943.92   199  47034.8941           Root MSE      =  94.408

------------------------------------------------------------------------------
           i |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .1155622   .0058357    19.80   0.000     .1040537    .1270706
           c |   .2306785   .0254758     9.05   0.000     .1804382    .2809188
       _cons |  -42.71437   9.511676    -4.49   0.000    -61.47215   -23.95659
------------------------------------------------------------------------------
```

Varying Slopes & Intercept

```
. reg i f  _IfirXf_2 _IfirXf_3 _IfirXf_4 _IfirXf_5 _IfirXf_6 _IfirXf_7 _IfirXf_8 _IfirXf_9
_IfirXf_10 c _I
> firXc_2  _IfirXc_3 _IfirXc_4 _IfirXc_5 _IfirXc_6 _IfirXc_7 _IfirXc_8 _IfirXc_9
_IfirXc_10 f1 f2 f3 f4 f5
>  f6 f7 f8 f9 f10, nocons

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F( 30,   170) =  232.02
       Model |  13295977.5    30  443199.251           Prob > F      =  0.0000
    Residual |  324728.552   170  1910.16795           R-squared     =  0.9762
-------------+------------------------------           Adj R-squared =  0.9720
       Total |  13620706.1   200  68103.5304           Root MSE      =  43.705

------------------------------------------------------------------------------
           i |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .1192808    .012302     9.70   0.000     .0949965    .1435651
   _IfirXf_2 |   .0555752   .0358072     1.55   0.123    -.0151088    .1262592
   _IfirXf_3 |  -.0927296   .0273253    -3.39   0.001    -.1466703    -.038789
   _IfirXf_4 |   -.041333   .0668818    -0.62   0.537    -.1733589    .0906928
```

_IfirXf_5	.0430969	.2752797	0.16	0.876	-.5003098	.5865035
_IfirXf_6	.012174	.1690119	0.07	0.943	-.3214583	.3458064
_IfirXf_7	-.0317536	.304842	-0.10	0.917	-.6335169	.5700096
_IfirXf_8	-.0663867	.0683301	-0.97	0.333	-.2012714	.068498
_IfirXf_9	-.0438929	.163553	-0.27	0.789	-.3667492	.2789634
_IfirXf_10	-.1147074	1.093767	-0.10	0.917	-2.273823	2.044408
c	.3714448	.0176537	21.04	0.000	.3365961	.4062935
_IfirXc_2	.0181971	.0668942	0.27	0.786	-.1138532	.1502473
_IfirXc_3	-.2197509	.0439884	-5.00	0.000	-.3065847	-.1329171
_IfirXc_4	-.0557266	.0964656	-0.58	0.564	-.2461514	.1346982
_IfirXc_5	-.3683431	.1073683	-3.43	0.001	-.5802899	-.1563962
_IfirXc_6	-.2860706	.5426892	-0.53	0.599	-1.357348	.7852069
_IfirXc_7	-.2476634	.0811476	-3.05	0.003	-.4078501	-.0874767
_IfirXc_8	-.2790383	.240715	-1.16	0.248	-.7542138	.1961372
_IfirXc_9	-.2893412	.1356119	-2.13	0.034	-.5570415	-.021641
_IfirXc_10	.0659244	3.204899	0.02	0.984	-6.260599	6.392448
f1	-149.7824	50.40091	-2.97	0.003	-249.2746	-50.29017
f2	-49.19831	67.10986	-0.73	0.465	-181.6743	83.27768
f3	-9.956308	49.17834	-0.20	0.840	-107.0352	87.12256
f4	-6.189955	44.45563	-0.14	0.889	-93.94611	81.5662
f5	22.70711	33.13412	0.69	0.494	-42.70019	88.11442
f6	-8.685544	24.57788	-0.35	0.724	-57.20269	39.8316
f7	-4.499533	52.39817	-0.09	0.932	-107.9344	98.93533
f8	-.5093876	34.30018	-0.01	0.988	-68.21852	67.19974
f9	-7.72284	44.95757	-0.17	0.864	-96.46984	81.02416
f10	.1615187	83.1752	0.00	0.998	-164.0277	164.3508

Ad hoc restrictions for some varying slopes & intercepts

```
. reg i f  _IfirXf_3 c _IfirXc_3 _IfirXc_5 _IfirXc_7 _IfirXc_9 f1 f2 f4 f6 f8, nocons

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F( 12,   188) =  580.63
       Model |  13262843.4    12  1105236.95           Prob > F      =  0.0000
    Residual |  357862.678   188  1903.52488           R-squared     =  0.9737
-------------+------------------------------           Adj R-squared =  0.9720
       Total |  13620706.1   200  68103.5304           Root MSE      =  43.629

------------------------------------------------------------------------------
           i |      Coef.   Std. Err.      t    P>|t|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |     .11374   .0107487    10.58   0.000      .0925363    .1349436
   _IfirXf_3 |  -.0917551   .0142069    -6.46   0.000     -.1197805   -.0637297
           c |    .368109   .0165488    22.24   0.000      .3354638    .4007542
   _IfirXc_3 |  -.2181603   .0426321    -5.12   0.000      -.302259   -.1340616
   _IfirXc_5 |  -.3021013   .0240904   -12.54   0.000     -.3496234   -.2545791
   _IfirXc_7 |   -.267541   .0326215    -8.20   0.000     -.3318922   -.2031899
   _IfirXc_9 |  -.3506112   .0350371   -10.01   0.000     -.4197276   -.2814948
          f1 |  -123.6061   44.31642    -2.79   0.006     -211.0275   -36.18476
          f2 |   77.66093   21.95567     3.54   0.001       34.3498    120.9721
          f4 |  -37.35355   11.92762    -3.13   0.002     -60.88272   -13.82439
          f6 |  -30.73267   10.58462    -2.90   0.004     -51.61256   -9.852785
          f8 |  -64.94263    11.8607    -5.48   0.000     -88.33979   -41.54547
------------------------------------------------------------------------------
```

F=((357,862-324,729)/18)/(324,728/170)=0.9

$$F = \frac{(SSR_R - SSR_U)/r}{SSR_U / df}$$

$SSR_0 = 342{,}059.957$	($K = 21$)	α, β_i	
$SSR_1 = 523{,}478.114$	($K = 12$)	α_i, β	FE
$SSR_2 = 324{,}728.552$	($K = 30$)	α_i, β_i	10x OLS
$SSR_3 = 1{,}755{,}850.43$	($K = 3$)	α, β	OLS

$$F_1 = \frac{(1{,}755{,}850.43 - 324{,}728.55)/27}{324{,}728.55/170} = 27.8$$

$$F_2 = \frac{(342{,}059.96 - 324{,}728.55)/9}{324{,}728.55/170} = 1.01$$

$$F_3 = \frac{(523{,}478.11 - 324{,}728.55)/18}{324{,}728.55/170} = 5.78$$

$$F_4 = \frac{(357{,}862.68 - 324{,}728.55)/18}{324{,}728.55/170} = 0.9$$

(ad hoc inductive model)

WHAT ARE THE ISSUES IN ESTIMATING PD MODELS?

A. Dynamic or Static?
- include lagged d.v. on right? if so, how?
- guide: theory, ideally
- tests: unsettled

TS rule of thumb: you need ~50 observations

PD with smallish T contain multiple instances of TS,
so more variation to exploit than size of T suggests
BUT ability to lag and difference is limited by small T
and ability to test rival forms of dynamics (AR, DL, MA,...) is limited

B. 1.Serial correlation?

This is not the same question as whether the data-generating process is dynamic, though serially correlated errors are a form of dynamics. MC studies show that if the Data Generating Process has equation dynamics and the models do not, there are spurious error dynamics. The converse is true as well.

B.2.Heteroscedasticity?

In cross-sectional models, "robust std errors" have caught on in a huge way. Poli-sci seems to be moving towards a norm that one should always compute robust SEs, without any explanation of why they are necessary. Though it is true that WLS is a lot of trouble (comparatively) in the XS world, when it is the SEs that we are worried about, robust SEs don't always make sense. Extension of the robust idea to PD models is not always straight-forward (though STATA allows the "robust" option on some of the `xt` procedures).

C. Poolability

- of intercepts…
  - over time? over space?
- of slopes…
  - over time? over space?

We can't test all of the above (at once). We impose some constraints by theory or according to panel dimensions. OLS on the whole NT sample is the ultimate pool (almost always rejected by data).

tests: depends on whether effects assumed to be random or fixed
(and complicated by dynamic specification)

Ongoing debate (especially in dynamic models) whether heterogeneous coefficients are necessary.

D. Balance

Lack of balance in the panel at minimum complicates computations.

Severe imbalance wreaks havoc—known results for balanced panels do not carry over to unbalanced panels.

Asymptotics complicated by variation in T across units (or in N across waves).

There are a variety of measures of balance to do quick checks on whether the unbalance is "severe" or not.

E. Specification

Not uniquely a PD problem, but a generic issue that is probably the most fundamental problem for most of our problems much of the time, whatever the data structure.

F. Spatial auto-correlation

New rapidly growing area—very complicated likelihood functions, but software emerging.

"Finally, it should be noted that although panel data offer many advantages, they are not panacea. The power of panel data analysis depends critically on the compatibility of the assumptions of statistical tools with the data generating process. Otherwise, misleading inference will follow."

(Cheng Hsiao)