DYNAMIC MODELS

Correlated Random Effects Panel Data Models

IZA Summer School in Labor Economics
May 13-19, 2013

Jeffrey M. Wooldridge
Michigan State University

1. Linear Models
2. General Dynamic Models
3. Estimating the APEs
4. Dynamic Probit Model
5. Other Dynamic Models
6. Unbalanced Panels
7. Tips for Applying the CRE Approach

1. Linear Models

- Dynamic linear models with unobserved effects are usually estimated by instrumental variables methods.
- CRE approaches can be used – and likely are much more efficient – at the cost of distributional assumptions.

- Common application of Arellano and Bond (1991) (and extensions):

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + c_i + u_{it}$$
$$E(u_{it}|\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, c_i) = 0,\ t = 1, 2, \ldots, T$$

- AB uses differencing and then uses IVs in the FD equation; optimal weighting matrix in GMM.

- If we assume

$$u_{it}|\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, c_i \sim Normal(0, \sigma_u^2)$$
$$c_i|\mathbf{z}_i, y_{i0} \sim Normal(\psi + \mathbf{z}_i\boldsymbol{\xi} + \xi_0 y_{i0}, \sigma_a^2)$$

then we can use MLE conditional on $(\mathbf{z}_i, y_{i0})$.

2. General Dynamic Models

- Difficult to specify and estimate models with heterogeneity if we do not assume strict exogeneity. But completely specified dynamic models can be estimated under certain assumptions.
- Even for discrete outcomes, a linear model, estimated using the Arellano and Bond approach (and extensions), is a good starting point. Coefficients can be compared with partial effects from nonlinear models.

- Binary response: Assume there is one lag of the dependent variable and all other explanatory variables are strictly exogenous: For $t = 1, \ldots, T$,

$$P(y_{it} = 1 | \mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, c_i) = P(y_{it} = 1 | \mathbf{z}_{it}, y_{i,t-1}, c_i)$$

- Specification allows us to assess the relative important of “state dependence” – that is, whether being in a certain state last period affects the probability of being in that state this period – and unobserved heterogeneity.

- The dynamic probit model with an unobserved effect is

$$P(y_{it} = 1 | \mathbf{z}_{it}, y_{i,t-1}, c_i) = \Phi(\mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + c_i).$$

- A more flexible version is

$$\Phi(\mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + y_{i,t-1}\mathbf{z}_{it}\boldsymbol{\eta} + c_i)$$

- Several approaches to dealing with the presence of c_i and the initial condition, y_{i0}.

(i) Treat the c_i as parameters to estimate (incidental parameters problem and computationally intensive).

(ii) Try to estimate δ and ρ without specifying conditional or unconditional distributions for c_i (available in some special cases). Cannot estimate partial effects.

(iii) Approximate $D(y_{i0}|c_i, \mathbf{z}_i)$ and then model $D(c_i|\mathbf{z}_i)$. Leads to $D(y_{i0}, y_{i1}, \ldots, y_{iT}|\mathbf{z}_i)$ and MLE conditional on $\mathbf{z}_i$. (This was originally proposed by Heckman, 1981.)

(iv) Model $D(c_i|\mathbf{z}_i, y_{i0})$. Leads to $D(y_{i1}, \ldots, y_{iT}|\mathbf{z}_i, y_{i0},)$ and MLE conditional on $(\mathbf{z}_i, y_{i0})$. Developed in Wooldridge (2005, Journal of Applied Econometrics).

3. Estimating the APEs

- Let $m_t(\mathbf{x}_t, \mathbf{c}, \boldsymbol{\theta}) = E(y_t|\mathbf{x}_t, \mathbf{c})$ for a scalar y_t. The average structural function is

$$ASF(\mathbf{x}_t) = E_{\mathbf{c}_i}[m_t(\mathbf{x}_t, \mathbf{c}_i, \boldsymbol{\theta})] = E_{(\mathbf{z}_i, y_{i0})}\left\{\left[\int m_t(\mathbf{x}_t, \mathbf{c}, \boldsymbol{\theta}) h(\mathbf{c}|\mathbf{z}_i, y_{i0}, \boldsymbol{\gamma}) d\mathbf{c}\right]\right\}.$$

- The term inside brackets, say $q_t(\mathbf{x}_t, \mathbf{z}_i, y_{i0}, \boldsymbol{\theta}, \boldsymbol{\gamma})$ is available, at least in principle, because $m_t(\bullet)$ and $h(\bullet)$ have been specified parametrically.
- Often, $q_t(\mathbf{x}_t, \mathbf{z}_i, y_{i0}, \boldsymbol{\theta}, \boldsymbol{\gamma})$ has a simple form, and it can be simulated if not.

• $ASF(\mathbf{x}_t)$ is consistently estimated by

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1}\sum_{i=1}^{N} q_t(\mathbf{x}_t, \mathbf{z}_i, y_{i0}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}}).$$

Partial derivatives and differences with respect to elements of $\mathbf{x}_t$ (which, remember, can include y_{t-1}) can be computed.

• With large N and small T, the panel data bootstrap (resampling all time periods from the cross-sectional units) can be used for standard errors and inference.

4. Dynamic Probit Model

- A linear model, estimated using the Arellano and Bond approach (and extensions), is a good starting point. Coefficients can be compared with partial effects from nonlinear models.

• Dynamic probit model leads to computationally simple estimators (logit is more difficult):

$$P(y_{it} = 1|\mathbf{z}_{it}, y_{i,t-1}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i),$$

where $\mathbf{x}_{it}$ is a function of $(\mathbf{z}_{it}, y_{i,t-1})$.

• A simple analysis is obtained from

$$c_i|\mathbf{z}_i, y_{i0} \sim Normal(\psi + \mathbf{z}_i\boldsymbol{\xi} + \xi_0 y_{i0}, \sigma_a^2)$$

or

$$c_i|\mathbf{z}_i, y_{i0} \sim Normal(\psi + \mathbf{x}_{i0}\boldsymbol{\eta}, \sigma_a^2)$$

where $\mathbf{x}_{i0}$ is a function of $(\mathbf{z}_i, y_{i0})$.

• We have

$$P(y_{it} = 1|\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, a_i) = \Phi(\mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + \psi + \mathbf{z}_i\boldsymbol{\xi} + \xi_0 y_{i0} + a_i),$$

where $a_i \equiv c_i - \psi - \mathbf{z}_i\boldsymbol{\xi} - \xi_0 y_{i0}$.

• Allows us to characterize $D(y_{i1}, \ldots, y_{iT}|\mathbf{z}_i, y_{i0})$ after "integrating out" a_i.

• In what follows $\mathbf{z}_i$ can contain initial-period values $\mathbf{z}_{i0}$ if they are available. Not needed but might help make distributional assumption closer to being true.

• Turns out the likelihood function has the same form as when the $\{\mathbf{x}_{it}\}$ are strictly exogenous. We can use standard random effects probit software, where the explanatory variables in time t are $(1, \mathbf{z}_{it}, y_{i,t-1}, \mathbf{z}_i, y_{i0},)$.

- Easily get the average partial effects, too:

$$\widehat{ASF}(\mathbf{z}_t, y_{t-1}) = N^{-1}\sum_{i=1}^{N}\Phi(\mathbf{z}_t\hat{\boldsymbol{\delta}}_a + \hat{\rho}_a y_{t-1} + \hat{\psi}_a + \mathbf{z}_i\hat{\boldsymbol{\xi}}_a + \hat{\xi}_{a0} y_{i0})$$

and take differences or derivatives with respect to elements of $(\mathbf{z}_t, y_{t-1})$.

- The scaled coefficients are, for example, $\hat{\boldsymbol{\delta}}_a = \hat{\boldsymbol{\delta}}(1 + \hat{\sigma}_a^2)^{-1/2}$, and $\hat{\sigma}_a^2$ is obtained from the random effects output.

• Let $\mathbf{x}_{i0} \equiv (\mathbf{z}_i, y_{i0},)$. Then the (unconditional) first two moments of c_i are easily estimated:

$$\hat{\mu}_c = \hat{\psi} + \bar{\mathbf{z}}\hat{\boldsymbol{\xi}} + \hat{\xi}_0 \bar{y}_0$$

$$\hat{\sigma}_c^2 = \hat{\boldsymbol{\lambda}}' \left(N^{-1} \sum_{i=1}^{N} (\mathbf{x}_{i0} - \bar{\mathbf{x}}_0)'(\mathbf{x}_{i0} - \bar{\mathbf{x}}_0) \right) \hat{\boldsymbol{\lambda}} + \hat{\sigma}_a^2$$

where $\hat{\boldsymbol{\lambda}} = (\hat{\boldsymbol{\xi}}', \hat{\xi}_0)'$.

• The unconditional density can be estimated as

$$\hat{f}_{c_i}(c) = N^{-1} \sum_{i=1}^{N} \phi[(c - \hat{\psi} - \mathbf{z}_i\hat{\boldsymbol{\xi}} - \hat{\xi}_0 y_{i0})/\hat{\sigma}_a)]/\hat{\sigma}_a.$$

EXAMPLE: (Dynamic Married Women's Labor Force Participation)

$$\begin{aligned} P(lfp_{it} &= 1|kids_{it}, lhinc_{it}, lfp_{i,t-1}, c_i) \\ &= \Phi(\alpha_t + \delta_1 kids_{it} + \delta_2 lhinc_{it} + \rho lfp_{i,t-1} + c_i) \\ c_i|\mathbf{z}_i, lfp_{i0} &\sim Normal(\psi + \xi_0 lfp_{i0} + \mathbf{z}_i\boldsymbol{\xi}, \sigma_a^2) \end{aligned}$$

- To get a measure of the magnitude of state dependence, estimate

$$E_{c_i}[\Phi(\alpha_t + \delta_1 kids_t + \delta_2 lhinc_t + \rho + c_i) - \Phi(\alpha_t + \delta_1 kids_t + \delta_2 lhinc_t + c_i)]$$

and put in interesting values for $kids_t$ and $lhinc_t$, or average those out in the sample.

- Data from LFP.DTA.
- The APE from dynamic probit with heterogeneity is about $.260$ ($se = .026$). If we ignore the heterogeneity, estimated APE is $.837$ ($se = .005$); standard errors from 500 panel bootstrap replications.
- Linear model estimates: $.382$ ($se = .020$) with heterogeneity, $.851$ ($se = .004$) without.

```
. * Start with a linear model estimated by Arellano and Bond:

. xtabond lfp kids lhinc per3 per4 per5

Arellano-Bond dynamic panel-data estimation  Number of obs         =     16989
Group variable: id                           Number of groups      =      5663
Time variable: period
                                             Obs per group:    min =         3
                                                               avg =         3
                                                               max =         3

Number of instruments =     12               Wald chi2(6)          =    378.77
                                             Prob > chi2           =    0.0000
One-step results
------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lfp_L1. |   .3818295   .0201399    18.96   0.000     .3423559    .4213031
        kids |  -.0130903   .0091827    -1.43   0.154     -.031088    .0049075
       lhinc |  -.0058375   .0053704    -1.09   0.277    -.0163633    .0046882
        per3 |  -.0053284   .0039777    -1.34   0.180    -.0131245    .0024677
        per4 |  -.0038833   .0039916    -0.97   0.331    -.0117067      .00394
        per5 |  -.0090286   .0039853    -2.27   0.023    -.0168396   -.0012176
       _cons |   .4848731   .0458581    10.57   0.000      .394993    .5747533
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).lfp
        Standard: D.kids D.lhinc D.per3 D.per4 D.per5
Instruments for level equation
        Standard: _cons
```

```
. * Accounting for heterogeneity is important, even in the linear
. * approximation. Without heterogeneity, the estimated state dependence is
. * much higher:

. reg lfp l.lfp kids lhinc per3 per4 per5, robust

Linear regression                                      Number of obs =   22652
                                                       F(  6, 22645) = 7938.78
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.7207
                                                       Root MSE      =  .24664

------------------------------------------------------------------------------
             |               Robust
         lfp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lfp |
         L1. |   .8510015   .0039478   215.57   0.000     .8432637    .8587394
             |
        kids |  -.0021431   .0014379    -1.49   0.136    -.0049615    .0006754
       lhinc |  -.0071892   .0025648    -2.80   0.005    -.0122164   -.0021619
        per3 |  -.0036044   .0047215    -0.76   0.445    -.0128588      .00565
        per4 |   .0010464   .0046287     0.23   0.821    -.0080262     .010119
        per5 |  -.0036555   .0045471    -0.80   0.421    -.0125681    .0052571
       _cons |    .157911   .0210127     7.52   0.000     .1167247    .1990972
------------------------------------------------------------------------------
```

```
. * Generate variables needed for dynamic probit.

tsset id period

* Lagged dependent variable:
bysort id (period): gen lfp_1 = L.lfp
* Put initial condition in periods 2-5:
by id: gen lfp1 = lfp[1]
* Create kids variables for periods 2-5:
forv i=2/5 {
by id: gen kids‘i’ = kids[‘i’]
}
* Create lhinc variables for periods 2-5:
forv i=2/5 {
by id: gen lhinc‘i’ = lhinc[‘i’]
}
```

```
. * Now include initial condition, leads and lags, and other
. * time-constant variables in RE probit
.
. xtprobit lfp lfp_1 lfp1 kids kids2-kids5 lhinc lhinc2-lhinc5 educ
            black age agesq per3-per5, re


Random-effects probit regression                Number of obs      =     22652
Group variable (i): id                          Number of groups   =      5663

Random effects u_i ~Gaussian                   Obs per group: min =          4
                                                              avg =        4.0
                                                              max =          4

                                                Wald chi2(19)      =   4091.17
Log likelihood  = -5028.9785                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfp_1 |   1.541288    .066803    23.07   0.000     1.410357     1.67222
        lfp1 |   2.530053   .1565322    16.16   0.000     2.223256    2.836851
        kids |  -.1455379   .0787386    -1.85   0.065    -.2998626    .0087868
       kids2 |   .3236282   .0968499     3.34   0.001      .133806    .5134504
       kids3 |   .1072842   .1235197     0.87   0.385    -.1348099    .3493784
       kids4 |     .01792   .1275595     0.14   0.888    -.2320921    .2679322
       kids5 |  -.3912412   .1058482    -3.70   0.000    -.5986998   -.1837825
       lhinc |  -.0748846   .0508406    -1.47   0.141    -.1745304    .0247612
      lhinc2 |  -.0232267   .0590167    -0.39   0.694    -.1388973    .0924438
      lhinc3 |   -.083386   .0626056    -1.33   0.183    -.2060908    .0393188
      lhinc4 |  -.0862979    .060961    -1.42   0.157    -.2057793    .0331835
      lhinc5 |   .0627793   .0592742     1.06   0.290     -.053396    .1789547
        educ |    .049906   .0100314     4.97   0.000     .0302447    .0695672
       black |   .1316009   .0982941     1.34   0.181     -.061052    .3242539
         age |   .1278946   .0193999     6.59   0.000     .0898715    .1659177
       agesq |  -.0016882     .00024    -7.03   0.000    -.0021586   -.0012177
        per3 |  -.0560723   .0458349    -1.22   0.221    -.1459071    .0337625
        per4 |   -.029532   .0463746    -0.64   0.524    -.1204245    .0613605
```

```
        per5 |  -.0784793   .0464923    -1.69   0.091    -.1696025     .012644
       _cons |  -2.946082   .4367068    -6.75   0.000    -3.802011   -2.090152
-------------+----------------------------------------------------------------
    /lnsig2u |   .0982792   .1225532                     -.1419206     .338479
-------------+----------------------------------------------------------------
     sigma_u |   1.050367   .0643629                      .9314989    1.184404
         rho |     .52455   .0305644                      .4645793     .583821
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =   160.73 Prob >= chibar2 = 0.000

. predict xdh, xb
(5663 missing values generated)

. gen xd0 = xdh - _b[lfp_1]*lfp_1
(5663 missing values generated)

. gen xd1 = xd0 + _b[lfp_1]
(5663 missing values generated)

. gen xd0a = xd0/sqrt(1 + (1.050367)^2)
(5663 missing values generated)

. gen xd1a = xd1/sqrt(1 + (1.050367)^2)
(5663 missing values generated)
```

```
. gen PHI0 = norm(xd0a)
(5663 missing values generated)

. gen PHI1 = norm(xd1a)
(5663 missing values generated)

. gen pelfp_1 = PHI1 - PHI0
(5663 missing values generated)

. sum pelfp_1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     pelfp_1 |     22652    .2591284    .0551711   .0675151   .4047995

. * .259 is the average probability of being in the labor force in
. * period t, given participation in t-1. This is somewhat lower than
. * the linear model estimate, .382.
```

```
. * A nonlinear model without heterogeneity gives a much larger
. * estimate:

. probit lfp lfp_1 kids lhinc educ black age agesq per3-per5

Probit regression                                 Number of obs   =      22652
                                                  LR chi2(10)     =   17744.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -5332.5289                       Pseudo R2       =     0.6246

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfp_1 |   2.875679   .0269811   106.58   0.000     2.822797    2.928561
        kids |   -.060792    .012217    -4.98   0.000    -.0847368   -.0368472
       lhinc |  -.1143176   .0211668    -5.40   0.000    -.1558037   -.0728315
        educ |   .0291868   .0052362     5.57   0.000     .0189241    .0394495
       black |   .0792495   .0536694     1.48   0.140    -.0259406    .1844395
         age |    .084403   .0099983     8.44   0.000     .0648067    .1039993
       agesq |  -.0010991   .0001236    -8.90   0.000    -.0013413    -.000857
        per3 |  -.0340795   .0369385    -0.92   0.356    -.1064777    .0383187
        per4 |   .0022816   .0371729     0.06   0.951    -.0705759    .0751391
        per5 |  -.0304156   .0371518    -0.82   0.413    -.1032318    .0424006
       _cons |  -2.170796   .2219074    -9.78   0.000    -2.605727   -1.735866
------------------------------------------------------------------------------
```

```
. predict xdph, xb
(5663 missing values generated)

. gen xdp0 = xdph - _b[lfp_1]*lfp_1
(5663 missing values generated)

. gen xdp1 = xdp0 + _b[lfp_1]
(5663 missing values generated)

. gen PHI0p = norm(xdp0)
(5663 missing values generated)

. gen PHI1p = norm(xdp1)
(5663 missing values generated)

. gen pelfp_1p = PHI1p - PHI0p
(5663 missing values generated)

. sum pelfp_1p

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    pelfp_1p |     22652    .8373056     .012207   .6019558   .8495204

. * Without accounting for heterogeneity, the average state dependence
. * is much larger: .837 versus .259.

. * The .837 estimate is pretty close to the dynamic linear model without
. * heterogeneity, .851.
```

- Certain extensions should not be too difficult, for example,

$$c_i|\mathbf{z}_i, y_{i0} \sim Normal(\psi + \mathbf{x}_{i0}\boldsymbol{\eta}, \sigma_a^2 \exp(\mathbf{x}_{i0}\boldsymbol{\omega}))$$

- Like a random effects probit model with heteroskedasticity.

5. Other Dynamic Models

Tobit

- As with probit, there is a way to estimate dynamic Tobit models using standard RE software.

- Assume

$$y_{it} = \max(0, \mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + c_i + u_{it}), t = 1, \ldots, T$$

and

$$u_{it}|(\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, c_i) \sim Normal(0, \sigma_u^2), t = 1, \ldots, T.$$

- Makes sense only for corner solutions, not for truly censored data.
- Not clear how lagged y should appear. Could define a dummy variable $w_{it} = 1[y_{it} = 1]$ and use, say, $\rho_1(1 - w_{i,t-1}) + \rho_2 y_{i,t-1}$. Can also interact these with the $\mathbf{z}_{it}$.
- Can replace $\mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1}$ with $\mathbf{x}_{it}\boldsymbol{\beta}$ where $\mathbf{x}_{it}$ is any function of $(\mathbf{z}_{it}, y_{i,t-1})$.

- A simple analysis is obtained from

$$c_i|\mathbf{z}_i, y_{i0} \sim Normal(\psi + \mathbf{z}_i\boldsymbol{\xi} + \xi_0 y_{i0}, \sigma_a^2)$$

or by letting this depend more flexibly on the initial value, y_{i0}. Then we have

$$y_{it} = \max(0, \mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + \psi + \mathbf{z}_i\boldsymbol{\xi} + \xi_0 y_{i0} + a_i + u_{it}).$$

- The log-likelihood takes the same form as the RE Tobit model with strictly exogenous variables, even though the explanatory variables are not strictly exogenous.

• As in the probit and ordered probit cases, a pooled Tobit analysis, with the same set of explanatory variables, is not consistent for the parameters.

• The APEs are again easy to compute:

$$N^{-1}\sum_{i=1}^{N} m(\mathbf{z}_t\hat{\boldsymbol{\delta}} + \hat{\rho}y_{t-1} + \hat{\psi} + \mathbf{z}_i\hat{\boldsymbol{\xi}} + \hat{\xi}_0 y_{i0}, \hat{\sigma}_a^2 + \hat{\sigma}_u^2),$$

where all estimates are from the MLE procedure.

- For a continuous variable, the scale factor is

$$N^{-1}\sum_{i=1}^{N}\Phi[(\mathbf{z}_t\hat{\boldsymbol{\delta}} + \hat{\rho}y_{t-1} + \hat{\psi} + \mathbf{z}_i\hat{\boldsymbol{\xi}} + \hat{\xi}_0 y_{i0})(\hat{\sigma}_a^2 + \hat{\sigma}_u^2)^{-1/2}],$$

and one can further average across $(\mathbf{z}_t, y_{t-1})$.

• Extensions along the lines of allowing heteroskedasticity in $D(c_i|y_{i0},\mathbf{z}_i)$, flexible conditional means, and even more flexible distributions, seem worth exploring.

• Honoré (1993) shows how to estimate $\boldsymbol{\delta}$ and ρ without distributional assumptions for c_i or u_{it}. Partial effects at different values of $(\mathbf{z}_t, y_{t-1})$ are not available, and y_{t-1} must appear in linear, additive form.

Count Data

- There are transformation methods for estimating

$$E(y_{it}|\mathbf{x}_{it},\ldots,\mathbf{x}_{i1},c_i) = E(y_{it}|\mathbf{x}_{it},c_i) = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta}).$$

where we have assumed sequential exogeneity.

- Can use a differencing-like transformation:

$$y_{it} - y_{i,t+1}\left[\frac{\exp(\mathbf{x}_{it}\boldsymbol{\beta})}{\exp(\mathbf{x}_{i,t+1}\boldsymbol{\beta})}\right] = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta})r_{it} - c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta})r_{i,t+1}$$
$$= c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta})(r_{it} - r_{i,t+1})$$

where

$$E(r_{it}|\mathbf{x}_{it},\ldots,\mathbf{x}_{i1},c_i) = 1$$

and then

$$E\{[y_{it} - y_{i,t+1}\exp((\mathbf{x}_{it} - \mathbf{x}_{i,t+1})\boldsymbol{\beta})|\mathbf{x}_{it},\ldots,\mathbf{x}_{i1}] = 0.$$

- If violation of strict exogeneity is due only to a lagged dependent variable, can use a conditional MLE approach. For example, suppose

$$D(y_{it}|\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i1}, y_{i0}, c_i) = Poisson[c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta})]$$

where, say, $\mathbf{x}_{it}$ can be any function of $(\mathbf{z}_{it}, y_{i,t-1})$. (Adding lags of $\mathbf{z}_{it}$, or further lags of y_{it}, is relatively straightforward with several time periods.). This assumption implies correct dynamics as well as strict exogeneity of $\{\mathbf{z}_{it} : t = 1, \ldots, T\}$.

• As usual, the presence of dynamics and heterogeneity in nonlinear models raises an "initial conditions" problem. A simple solution is to model the dependence between c_i and $(\mathbf{z}_i, y_{i0})$:

$$
\begin{aligned}
c_i &= \exp(\psi + \mathbf{z}_i\boldsymbol{\gamma} + \xi y_{i0})a_i \\
D(a_i|\mathbf{z}_i, y_{i0}) &= Gamma(\delta, \delta)
\end{aligned}
$$

where $E(a_i) = 1$ and $\delta = 1/\eta^2 = 1/Var(a_i)$.

- As shown in Wooldridge (2005, Journal of Applied Econometrics), the resulting likelihood function is identical to the Poisson RE likelihood with explanatory variables

$$(\mathbf{z}_{it}, y_{i,t-1}, \mathbf{z}_i, y_{i0})$$

in the case $\mathbf{x}_{it} = (\mathbf{z}_{it}, y_{i,t-1})$.

- So, to implement the method, generate $\mathbf{z}_i$ and y_{i0} so that they appear on every line (time period) of data for each i.
- Could estimate a dynamic patents-R&D relationship using this approach.

6. Unbalanced Panels

- Unbalanced panels are difficult to deal with in dynamic models, even if we assume the sample selection is appropriate exogenous.
- Pure attrition is easiest to deal with because of its sequential nature.
- Assume that we have a random sample at time $t = 1$, which means we observe $(\mathbf{y}_{i1}, \mathbf{z}_{i1}, \mathbf{y}_{i0})$ for all units. Some units leave the sample starting at $t = 2$ (and never return). And so on.

- Let T_i be the last time period for unit i, a random variable.
- Assume that attrition is exogenous in the sense that

$$D(\mathbf{y}_{it}|\mathbf{z}_i,\mathbf{y}_{i,t-1},\ldots,\mathbf{y}_{i0},\mathbf{c}_i,T_i) = D(\mathbf{y}_{it}|\mathbf{z}_{it},\mathbf{y}_{i,t-1},\mathbf{c}_i,) \text{ for } t = 2,\ldots,.T.$$

This combines correct dynamics, strict exogeneity of $\{\mathbf{z}_{it}\}$, and the sense in which attrition is exogenous.

- Rules out shocks affect attrition. Allows T_i to depend on the path of the exogenous covariates, $\{\mathbf{z}_{it}\}$, and on the unobserved heterogeneity, $\mathbf{c}_i$.

• Now we need to model the distribution of $\mathbf{c}_i$ given $(\mathbf{z}_{i1},\ldots,\mathbf{z}_{iT_i},\mathbf{y}_{i0})$ for the different possible outcomes on T_i.

• The log likelihood conditional on $(\mathbf{z}_i^{T_i},\mathbf{y}_{i0})$, where $\mathbf{z}_i^{T_i} = (\mathbf{z}_{i1},\ldots,\mathbf{z}_{i,Ti})$ is

$$\log\left\{\int\left[\prod_{t=1}^{T_i} f_t(\mathbf{y}_{it}|\mathbf{z}_{it},\mathbf{y}_{i,t-1},\mathbf{c};\boldsymbol{\theta})\right]h_{T_i}(\mathbf{c}|\mathbf{z}_{i,T_i},\mathbf{y}_{i0};\boldsymbol{\gamma}_{T_i})d\mathbf{c}\right\}$$

where we should allow the conditional density of $\mathbf{c}_i$ to change with T_i.

• We can pool the estimation of $\boldsymbol{\theta}$ and the $\boldsymbol{\gamma}_R$, $R = 2,\ldots,T$, across R. The APEs now average across the estimated models for each T_i.

• Alternatively, we could estimate separate $\boldsymbol{\theta}_R$ for each R compute the APEs for each r, and then average across $R = 2,\ldots,T$. So, the average structural function would be estimated as

$$\widehat{ASF}(\mathbf{x}_t) = (T-1)^{-1}\sum_{R=2}^{T}\left[N^{-1}\sum_{i=1}^{N} q_t(\mathbf{x}_t, \mathbf{z}_i, \mathbf{y}_{i0}, \hat{\boldsymbol{\theta}}_R, \hat{\boldsymbol{\gamma}}_R)\right].$$

- Estimating separate models is computationally attractive, too, because in many cases we can estimate a simple model for each T_i.
- For example, in a dynamic probit model, we should allow the coefficients in the mean of the heterogeneity distribution to change with T_i, and we should allow the variance to (at least) depend on T_i, too.

• Can extend this to different starting times, too. Suppose unit i enters the sample at time S_i and stays for each time period. Now the conditional heterogeneity distribution should be

$$h_{S_i,T_i}(\mathbf{c}|\mathbf{z}_{i,S_i,T_i},\mathbf{y}_{i,S_i};\boldsymbol{\gamma}_{S_i,T_i})$$

We perform (if possible) a different estimation for all (S_i, T_i) combinations.

• Even better is to pool and impose common $\boldsymbol{\theta}$.

• For the APEs we now average across (S_i, T_i), too.

7. Tips for Applying the CRE Approach

- Suppose we observe $(\mathbf{y}_{i0}, \mathbf{z}_{i0}, \mathbf{y}_{i1}, \mathbf{z}_{i1} \ldots, \mathbf{y}_{iT}, \mathbf{z}_{iT})$ and the model with heterogeneity includes $(\mathbf{z}_{it}, \mathbf{y}_{i,t-1})$. Conditioning on $\mathbf{z}_{i0}$ in $D(\mathbf{c}_i | \mathbf{z}_{i0}, \mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}, \mathbf{y}_{i0})$ is optional. What is required is for us to have $D(\mathbf{c}_i | \mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}, \mathbf{y}_{i0})$.
- The CRE approach with $\mathbf{y}_{i0}$ is often implemented using $D(\mathbf{c}_i | \bar{\mathbf{z}}_i, \mathbf{y}_{i0})$. This conserves on parameters but is generally not a good idea.

- Not clear to handle general pattern of missing data. Using $\bar{\mathbf{z}}_i$ cannot be justified.
- If all units are observed at $t = 0$, might make the assumption

$$D(\mathbf{c}_i | \mathbf{z}_{i0}, \mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}, \mathbf{y}_{i0}) = D(\mathbf{c}_i | \mathbf{z}_{i0}, \mathbf{y}_{i0})$$