SYSTEMS OF EQUATIONS: SUR AND PANEL DATA, REVISITED

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. SUR, Revisited
2. Panel Data

1. SUR, REVISITED

- Consider again the SUR system, written for a random draw i as

$$y_{ig} = \mathbf{x}_{ig}\boldsymbol{\beta}_g + u_{ig}, g = 1, \ldots, G$$

where $\boldsymbol{\beta}_g$ is $K_g \times 1$. We have considered two estimators of the $\boldsymbol{\beta}_g$: OLS equation-by-equation and GLS using $\hat{\boldsymbol{\Omega}}$ as the estimated $G \times G$ variance matrix.

1.1. OLS versus SUR for Systems

- **Algebraic/Asymptotic Equivalences**:

(1) If the same regressors appear in each equation, that is, $\mathbf{x}_{ig} = \mathbf{x}_i$, $g = 1, \ldots, G$, the OLS equation-by-equation is numerically the same as FGLS for any structure of $\hat{\mathbf{\Omega}}$.

- The matrix of regressors can be written as $\mathbf{X}_i = \mathbf{I}_G \otimes \mathbf{x}_i$.
- The algebra is tedious. When the data are stacked differently – by equation, not observation – the algebra is easy, but the asymptotic analysis is unnatural.

(2) If $\hat{\mathbf{\Omega}}$ is diagonal, FGLS = OLS equation by equation for any choice of explanatory variables, $\mathbf{x}_{ig}$. Again, this is an algebraic result, which can be shown by writing

$$\left(\sum_{i=1}^{N}\mathbf{X}_i'\hat{\mathbf{\Omega}}^{-1}\mathbf{X}_i\right)^{-1} = \begin{pmatrix} \hat{\sigma}_1^2\left(\sum_{i=1}^{N}\mathbf{x}_{i1}'\mathbf{x}_{i1}\right)^{-1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \hat{\sigma}_G^2\left(\sum_{i=1}^{N}\mathbf{x}_{iG}'\mathbf{x}_{iG}\right)^{-1} \end{pmatrix}$$

$$\sum_{i=1}^{N}\mathbf{X}_i'\hat{\mathbf{\Omega}}^{-1}\mathbf{y}_i = \begin{pmatrix} \hat{\sigma}_1^{-2}\sum_{i=1}^{N}\mathbf{x}_{i1}'y_{i1} \\ \vdots \\ \hat{\sigma}_G^{-2}\sum_{i=1}^{N}\mathbf{x}_{iG}'y_{iG} \end{pmatrix}$$

$(2')$ If $\mathbf{\Omega}$ is diagonal, and $\hat{\mathbf{\Omega}} \xrightarrow{p} \mathbf{\Omega}$, then FGLS and OLS EBE are *asymptotically* equivalent. Why? By (2), OLS EBE and GLS would be identical, and we know FGLS and GLS are asymptotically equivalent:

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{GLS} - \hat{\boldsymbol{\beta}}_{SOLS}) = \mathbf{0}$$

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FGLS} - \hat{\boldsymbol{\beta}}_{GLS}) = o_p(1)$$

so

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FGLS} - \hat{\boldsymbol{\beta}}_{SOLS}) = o_p(1).$$

- Important implication of (1) and (2) [or (2$'$)]: FGLS is (asymptotically) more efficient the OLS EBE only when at least some exclusion restrictions have been made and there is some correlation in the errors across equations. Therefore, there is a tradeoff between efficiency and robustness.
- If we are interested in, say, the first equation, then $E(\mathbf{x}_{i1}'u_{i1}) = \mathbf{0}$ is sufficient for OLS on that equation to be consistent. SUR generally requires $E(\mathbf{x}_{ig}'u_{ih}) = \mathbf{0}$ for all g and h. Therefore, if we have improperly omitted, say, an explanatory variable from the second equation, the FGLS estimates of $\boldsymbol{\beta}_1$ (and $\boldsymbol{\beta}_2$) are generally inconsistent.

- FGLS gains efficiency over OLS (under system homoskedasticity) only when it is valid to use the orthogonality condition $E[(\mathbf{\Omega}^{-1}\mathbf{X}_i)'\mathbf{u}_i] = \mathbf{0}$ and some variables omitted from an equation, say, g, are assumed to be uncorrelated with at least one explanatory variable omitted from that equation.
- Can test the null hypothesis $H_0 : \sigma_{gh} = 0$, all $g \neq h$. Have $G(G-1)/2$ restrictions, with σ_g^2, $g = 1, \ldots, G$ unrestricted.
- The Breusch-Pagan test assumes normality (actually, that the first four moments of the multivariate distribution are the same as multivariate normal, with independence between $\mathbf{u}_i$ and $\mathbf{X}_i$.

• The B-P statistic uses the OLS residuals for each equation because it is a Lagrange Multiplier test, which is based on estimation under the null.

• The outcome of the B-P test is rarely in doubt: one almost always strongly rejects the null. A robust test that uses only SGLS.1 and SGLS.2 could be derived, using

$$N^{-1/2}\sum_{i=1}^{N}\check{\mathbf{u}}_i\check{\mathbf{u}}_i' = N^{-1/2}\sum_{i=1}^{N}\mathbf{u}_i\mathbf{u}_i' + o_p(1)$$

without restricting the fourth moments of $\mathbf{u}_i$.

• **EXAMPLE**: A two equation SUR system for hourly earnings and benefits. $N = 606$.

```
. sureg (hrearn educ exper expersq union married white male)
     (hrbens educ exper expersq union married white male), corr

Seemingly unrelated regression
----------------------------------------------------------------------
Equation          Obs  Parms        RMSE    "R-sq"       chi2        P
----------------------------------------------------------------------
hrearn            616      7    4.332039    0.1965     150.68   0.0000
hrbens            616      7    .5417217    0.3353     310.77   0.0000
----------------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hrearn       |
        educ |   .4645619   .0672265     6.91   0.000     .3328004    .5963234
       exper |  -.0530683   .0522106    -1.02   0.309    -.1553992    .0492627
     expersq |   .0033981   .0011129     3.05   0.002     .0012168    .0055794
       union |   .7685325   .3905196     1.97   0.049      .003128    1.533937
     married |   .6222725    .413202     1.51   0.132    -.1875886    1.432134
       white |   1.107492    .605861     1.83   0.068    -.0799737    2.294958
        male |   1.735931   .3939833     4.41   0.000     .9637374    2.508124
       _cons |  -3.078173   1.076508    -2.86   0.004     -5.18809   -.9682564
-------------+----------------------------------------------------------------
```

```
hrbens       |
        educ |   .0739853   .0084067     8.80   0.000     .0575085    .0904621
       exper |   .0431919   .0065289     6.62   0.000     .0303954    .0559883
     expersq |  -.0007348   .0001392    -5.28   0.000    -.0010076   -.0004621
       union |   .4442268   .0488345     9.10   0.000     .3485129    .5399406
     married |   .0889692   .0516709     1.72   0.085     -.012304    .1902424
       white |   .0866399   .0757629     1.14   0.253    -.0618527    .2351326
        male |   .2400792   .0492676     4.87   0.000     .1435164     .336642
       _cons |  -.8888685   .1346174    -6.60   0.000    -1.152714   -.6250233
------------------------------------------------------------------------------

Correlation matrix of residuals:

        hrearn  hrbens
hrearn  1.0000
hrbens  0.3022  1.0000

Breusch-Pagan test of independence: chi2(1) =    56.267, Pr = 0.0000

. test married

 ( 1)  [hrearn]married = 0
 ( 2)  [hrbens]married = 0

           chi2(  2) =    4.03
         Prob > chi2 =    0.1331
```

```
. reg hrearn educ exper expersq union married white male, robust

Linear regression                                      Number of obs =     616
                                                       F(  7,   608) =   37.08
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1965
                                                       Root MSE      =  4.3604

------------------------------------------------------------------------------
             |               Robust
      hrearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .4645619   .0764839     6.07   0.000     .3143572    .6147665
       exper |  -.0530683   .2310098    -0.23   0.818    -.5067423    .4006058
     expersq |   .0033981    .005917     0.57   0.566    -.0082221    .0150183
       union |   .7685325   .2896582     2.65   0.008     .1996804    1.337385
     married |   .6222725   .3362197     1.85   0.065    -.0380204    1.282565
       white |   1.107492   .4993442     2.22   0.027     .1268432    2.088141
        male |   1.735931   .2734542     6.35   0.000     1.198901     2.27296
       _cons |  -3.078173   .8501402    -3.62   0.000    -4.747741   -1.408605
------------------------------------------------------------------------------
```

1.2 Imposing Cross-Equation Restrictions

- Suppose a two-equation system in the population is

$$y_1 = \gamma_{10} + \gamma_{11}x_{11} + \gamma_{12}x_{12} + \alpha_1 x_{13} + \alpha_2 x_{14} + u_1$$
$$y_2 = \gamma_{20} + \gamma_{21}x_{21} + \alpha_1 x_{22} + \alpha_2 x_{23} + \gamma_{24}x_{24} + u_2$$

Let the vector of all parameters be the 8×1 vector

$$\boldsymbol{\beta} = (\gamma_{10}, \gamma_{11}, \gamma_{12}, \alpha_1, \alpha_2, \gamma_{20}, \gamma_{21}, \gamma_{24})'.$$

Then we can define the matrix of regressors as

$$
\mathbf{X}_i = \begin{pmatrix} 1 & x_{i11} & x_{i12} & x_{i13} & x_{i14} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{i22} & x_{i23} & 1 & x_{i21} & x_{i24} \end{pmatrix}
$$

• Stata has a feature that allows one to specify linear constraints on the parameters, which is more natural.

• Cross-equation restrictions arise naturally in demand systems, cost share equations, and so on.

• In share equations, where the dependent variable is a fraction, can question whether linearity seems reasonable. Almost certainly the system homoskedasticity assumption fails.

1.3. Systems with Singular Variance-Covariance Matrices.

- In expenditure and cost share systems, the G responses, if the categories are exhaustive and mutually exclusive, sum to unity.
- For firm i let s_{iK}, s_{iL}, and s_{iM} be the cost shares for capital, labor, and materials, respectively, and assume that $s_{iK} + s_{iL} + s_{iM} = 1$. A popular cost share system is

$$\begin{aligned}
s_{iK} &= \gamma_{10} + \gamma_{11}\log(p_{iK}) + \gamma_{12}\log(p_{iL}) + \gamma_{13}\log(p_{iM}) + u_{iK} \\
s_{iL} &= \gamma_{20} + \gamma_{21}\log(p_{iK}) + \gamma_{22}\log(p_{iL}) + \gamma_{23}\log(p_{iM}) + u_{iL} \\
s_{iM} &= \gamma_{30} + \gamma_{31}\log(p_{iK}) + \gamma_{32}\log(p_{iL}) + \gamma_{33}\log(p_{iM}) + u_{iM}
\end{aligned}$$

- The restriction on the sum implies

$$\gamma_{10} + \gamma_{20} + \gamma_{30} = 1,\ \gamma_{11} + \gamma_{21} + \gamma_{31} = 0,\ \gamma_{12} + \gamma_{22} + \gamma_{32} = 0$$

$$\gamma_{13} + \gamma_{23} + \gamma_{33} = 0,\ u_{iK} + u_{iL} + u_{iM} = 0$$

and that last restriction implies that $\mathbf{\Omega} = E(\mathbf{u}_i \mathbf{u}_i')$, a 3×3 matrix, has rank two, not three.

- Can drop any of the equations. Make it the last one, and impose the restrictions on the parameters. Can write

$$s_{iK} = \gamma_{10} + \gamma_{11} \log(p_{iK}/p_{iM}) + \gamma_{12} \log(p_{iL}/p_{iM}) + u_{iK}$$

$$s_{iL} = \gamma_{20} + \gamma_{12} \log(p_{iK}/p_{iM}) + \gamma_{22} \log(p_{iL}/p_{iM}) + u_{iL}$$

• This two-equation system has a cross equation restriction, too. But the singularity in the variance matrix is gone, so can apply FGLS with

$\boldsymbol{\beta} = (\gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{20}, \gamma_{22})$

$$\mathbf{X}_i = \begin{pmatrix} 1 & \log(p_{iK}/p_{iM}) & \log(p_{iL}/p_{iM}) & 0 & 0 \\ 0 & 0 & \log(p_{iK}/p_{iM}) & 1 & \log(p_{iL}/p_{iM}) \end{pmatrix}.$$

• Can add firm characteristics to the share equations without essential change.

• Current interesting question: What nonlinear systems are consistent with production theory that respect the fractional nature of the shares?

2. PANEL DATA

- Write for a random draw i as

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it},\ t = 1,\ldots,T. \tag{2.1}$$

- Remember, $\mathbf{x}_{it}$ can contain all kinds of explanatory variables, including time period dummies and variables that do not change over time.

2.1. Assumptions for Pooled OLS (POLS).

Assumption POLS.1 (Contemporaneous Exogeneity):

$$E(\mathbf{x}_{it}'u_{it}) = \mathbf{0},\ t = 1,\ldots,T. \tag{2.2}$$

• Remember, POLS.1 allows for lagged dependent variables as well as other non-strictly exogenous regressors.

Assumption POLS.2 (Rank Condition):

$$rank\left[\sum_{t=1}^{T} E(\mathbf{x}_{it}'\mathbf{x}_{it})\right] = K. \tag{2.3}$$

- Under POLS.1 and POLS.2 the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\beta}}_{POLS} - \boldsymbol{\beta})$ is

$$\left[\sum_{t=1}^{T} E(\mathbf{x}_{it}'\mathbf{x}_{it})\right]^{-1}\left[\sum_{t=1}^{T}\sum_{s=1}^{T} E(u_{it}u_{is}\mathbf{x}_{it}'\mathbf{x}_{is})\right]\left[\sum_{t=1}^{T} E(\mathbf{x}_{it}'\mathbf{x}_{it})\right]^{-1}. \tag{2.4}$$

- This expression simplifies if we appropriately restrict the conditional variances and covariances.

Assumption POLS.3 (Homoskedasticity and No Serial Correlation):

$$
\begin{aligned}
\text{(a) } E(u_{it}^2 \mathbf{x}_{it}' \mathbf{x}_{it}) &= E(u_{it}^2) E(\mathbf{x}_{it}' \mathbf{x}_{it}) && (2.5)\\
&= \sigma^2 E(\mathbf{x}_{it}' \mathbf{x}_{it}), \text{ where } \sigma^2 = E(u_{it}^2), \text{ all } t \\
\text{(b) } E(u_{it} u_{is} \mathbf{x}_{it}' \mathbf{x}_{is}) &= \mathbf{0}, \text{ all } t \neq s. && (2.6)
\end{aligned}
$$

- Under POLS.3, (2.4) becomes

$$
\sigma^2 \left[\sum_{t=1}^{T} E(\mathbf{x}_{it}' \mathbf{x}_{it}) \right]^{-1}.
$$

• POLS.3 implies that the "usual" asymptotic variance matrix estimator of $\hat{\boldsymbol{\beta}}_{POLS}$ is valid:

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{POLS}) = \hat{\sigma}^2\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \tag{2.7}$$

$$\hat{\sigma}^2 = (NT-K)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2 = SSR/(NT-K). \tag{2.8}$$

• Can use the usual t and F statistics as approximately valid for large N.

• Without POLS.3, generally need fully robust variance matrix. That is, robust to arbitrary heteroskedasticity and serial correlation (unconditional or conditional):

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{POLS}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{u}_{it}\hat{u}_{is}\mathbf{x}_{it}'\mathbf{x}_{is}\right) \cdot \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1} \tag{2.9}$$

• This estimator in Stata is computed using a "cluster" option, where each unit i is a cluster of T time series observations.

• If we maintain the no serial correlation part of POLS.3, that is, $E(u_{it}u_{is}\mathbf{x}_{it}'\mathbf{x}_{is}) = \mathbf{0}$, all $t \neq s$, then a heteroskedasticity-robust form is valid:

$$\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^{2}\mathbf{x}_{it}'\mathbf{x}_{is}\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}$$

• In Stata, this estimator is obtained with a “robust” option, but its robustness is limited to heteroskedasticity, not serial correlation.

EXAMPLE: Relationsip Between Air Fares and Concentration Ratio.

$N = 1,149$, $T = 4$.

```
. use airfare

. tab year

1997, 1998, |
 1999, 2000 |      Freq.     Percent        Cum.
------------+-----------------------------------
       1997 |      1,149       25.00       25.00
       1998 |      1,149       25.00       50.00
       1999 |      1,149       25.00       75.00
       2000 |      1,149       25.00      100.00
------------+-----------------------------------
      Total |      4,596      100.00\pagebreak

. des fare concen

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
fare            int    %9.0g                  avg. one-way fare, $
concen          float  %9.0g                  market share, largest carrier
```

```
. list id year fare concen dist in 1/16

     +------------------------------------+
     | id   year   fare   concen   dist |
     |------------------------------------|
  1. |  1   1997    106    .8386    528 |
  2. |  1   1998    106    .8133    528 |
  3. |  1   1999    113    .8262    528 |
  4. |  1   2000    123    .8612    528 |
  5. |  2   1997    104    .5798    861 |
     |------------------------------------|
  6. |  2   1998    105    .5817    861 |
  7. |  2   1999    115    .7319    861 |
  8. |  2   2000    129    .5386    861 |
  9. |  3   1997    207     .818    852 |
 10. |  3   1998    188    .8172    852 |
     |------------------------------------|
 11. |  3   1999    229    .7998    852 |
 12. |  3   2000    247    .7097    852 |
 13. |  4   1997    243    .4604    724 |
 14. |  4   1998    226    .4614    724 |
 15. |  4   1999    229    .4334    724 |
     |------------------------------------|
 16. |  4   2000    176    .3716    724 |
     +------------------------------------+
```

```
. sum fare concen

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        fare |      4596    178.7968    74.88151         37        522
      concen |      4596    .6101149     .196435      .1605          1

. reg lfare concen ldist ldistsq y98 y99 y00

      Source |       SS       df       MS              Number of obs =    4596
-------------+------------------------------           F(  6,  4589) =  523.18
       Model |  355.453858     6  59.2423096           Prob > F      =  0.0000
    Residual |  519.640516  4589  .113236112           R-squared     =  0.4062
-------------+------------------------------           Adj R-squared =  0.4054
       Total |  875.094374  4595  .190444913           Root MSE      =  .33651

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203   .0300691    11.98   0.000     .3011705    .4190702
       ldist |  -.9016004    .128273    -7.03   0.000    -1.153077   -.6501235
     ldistsq |   .1030196   .0097255    10.59   0.000     .0839529    .1220863
         y98 |   .0211244   .0140419     1.50   0.133    -.0064046    .0486533
         y99 |   .0378496   .0140413     2.70   0.007      .010322    .0653772
         y00 |     .09987   .0140432     7.11   0.000     .0723385    .1274015
       _cons |   6.209258   .4206247    14.76   0.000     5.384631    7.033884
------------------------------------------------------------------------------
```

```
. reg lfare concen ldist ldistsq y98 y99 y00, robust

Linear regression                                      Number of obs =    4596
                                                       F(  6,  4589) =  558.39
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4062
                                                       Root MSE      =  .33651

------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203   .0318147    11.32   0.000     .2977482    .4224925
       ldist |  -.9016004   .1406543    -6.41   0.000    -1.177351   -.6258503
     ldistsq |   .1030196   .0104402     9.87   0.000     .0825518    .1234875
         y98 |   .0211244   .0141734     1.49   0.136    -.0066623     .048911
         y99 |   .0378496   .0144012     2.63   0.009     .0096162    .0660829
         y00 |     .09987   .0143821     6.94   0.000     .0716742    .1280658
       _cons |   6.209258   .4711359    13.18   0.000     5.285605    7.132911
------------------------------------------------------------------------------
```

```
. reg lfare concen ldist ldistsq y98 y99 y00, cluster(id)

Linear regression                                      Number of obs =    4596
                                                       F(  6,  1148) =  205.63
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4062
                                                       Root MSE      =  .33651

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203    .058556     6.15   0.000     .2452315    .4750092
       ldist |  -.9016004   .2719464    -3.32   0.001    -1.435168   -.3680328
     ldistsq |   .1030196   .0201602     5.11   0.000     .0634647    .1425745
         y98 |   .0211244   .0041474     5.09   0.000     .0129871    .0292617
         y99 |   .0378496   .0051795     7.31   0.000     .0276872     .048012
         y00 |     .09987   .0056469    17.69   0.000     .0887906    .1109493
       _cons |   6.209258   .9117551     6.81   0.000     4.420364    7.998151
------------------------------------------------------------------------------
```

2.2. Dynamic Completeness and Time Series Persistence

- In an important case, there can be no serial correlation in the errors in the sense that POLS.3(b) must hold. If $\mathbf{x}_t$ has been chosen such that

$$E(y_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1) = E(y_t|\mathbf{x}_t) \tag{2.10}$$

then the errors $\{u_t\}$ can have no serial correlation, and neither can $\{\mathbf{x}_t' u_t : t = 1, \ldots, T\}$.

- When (2.10) holds, we say the model is **dynamically complete** (in its mean). In the context of the linear model, it is the same as

$$E(u_t|\mathbf{x}_t, u_{t-1}, \mathbf{x}_{t-1}, \ldots, u_1, \mathbf{x}_1) = 0. \tag{2.11}$$

- Let $s < t$, so that $(\mathbf{x}_t, u_s, \mathbf{x}_s) \subset (\mathbf{x}_t, u_{t-1}, \mathbf{x}_{t-1}, \ldots, u_1, \mathbf{x}_1)$. By iterated expectations,

$$
\begin{aligned}
E(u_t u_s \mathbf{x}_t' \mathbf{x}_s) &= E[E(u_t u_s \mathbf{x}_t' \mathbf{x}_s | \mathbf{x}_t, u_s, \mathbf{x}_s)] \\
&= E[E(u_t | \mathbf{x}_t, u_s, \mathbf{x}_s) u_s \mathbf{x}_t' \mathbf{x}_s] \\
&= E[0 \cdot u_s \mathbf{x}_t' \mathbf{x}_s]
\end{aligned}
$$

because $E(u_t | \mathbf{x}_t, u_s, \mathbf{x}_s) = 0$ under dynamic completeness.

- A weaker sufficient conditon for POLS.3(b) is

$$
E(u_t u_s | \mathbf{x}_t, \mathbf{x}_s) = 0, \text{ all } t \neq s. \tag{2.12}
$$

- $E(u_t u_s) = 0$ for $t \neq s$ is not enough with random regressors.

- Dynamic completeness (DC) is a very strong assumption in static models. Suppose

$$y_t = \eta_t + \mathbf{z}_t\boldsymbol{\gamma} + u_t, \tag{2.13}$$

where $\mathbf{z}_t$ is dated contemporaneously with y_t. DC requires

$$E(y_t|\mathbf{z}_t, y_{t-1}, \mathbf{z}_{t-1}, \ldots, y_1, \mathbf{z}_1) = E(y_t|\mathbf{z}_t), \tag{2.14}$$

that is, once $\mathbf{z}_t$ is controlled for, neither past values of y or $\mathbf{z}$ help to predict y_t.

- Also for finite distributed lags, say

$$y_t = \eta_t + \mathbf{z}_t\boldsymbol{\gamma}_0 + \mathbf{z}_{t-1}\boldsymbol{\gamma}_1 + \mathbf{z}_{t-2}\boldsymbol{\gamma}_2 + u_t, \tag{2.15}$$

it may be reasonable to assume the distributed lag dynamics are correct:

$$E(y_t|\mathbf{z}_t, \mathbf{z}_{t-1}, \mathbf{z}_{t-2}) = E(y_t|\mathbf{z}_t, \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \ldots, \mathbf{z}_1). \tag{2.16}$$

But dynamic completeness as stated in (2.10) requires much more: no lagged outcomes on y help to predict y_t:

$$E(y_t|\mathbf{z}_t, y_{t-1}, \mathbf{z}_{t-1}, \ldots, y_1, \mathbf{z}_1) = E(y_t|\mathbf{z}_t, \mathbf{z}_{t-1}, \mathbf{z}_{t-2}). \tag{2.17}$$

• One way to interpret the presence of serial correlation in the errors of panel data models is that the model has misspecified dynamics. However, we may not want a model to satisfy the DC assumption

$$E(y_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1) = E(y_t|\mathbf{x}_t)$$

We may be happy estimating, say, $E(y_t|\mathbf{z}_t)$, $E(y_t|\mathbf{z}_t, \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \ldots, \mathbf{z}_1)$, or even $E(y_t|\mathbf{z}_t, y_{t-1})$ without having any of these represent the fully dynamic conditional mean.

• Remember that the presence of serial correlation is entirely different from strict exogeneity. Strict exogeneity always fails in models with a lagged dependent variable.

- An important point for inference is that all statistics we have discussed are valid for large N and small T without restricting the time series dependence in the data. So, for example, suppose our model is

$$y_t = \eta_t + \mathbf{z}_t\boldsymbol{\gamma} + \rho y_{t-1} + u_t, \ t = 1, \ldots, T. \tag{2.18}$$

If this were a pure time series problem, we would need to worry about the time series properties of $\{\mathbf{z}_t\}$, and we would have to use different inference methods if $\rho \geq 1$. But with a large cross section and small T, the statistical properties of the estimators are invariant to the time series properties of the series.

2.3. Testing for Serial Correlation and Heteroskedasticity

- Recall under strict exogeneity that we can ignore estimation of $\boldsymbol{\beta}$ – in this case, estimation by POLS – in testing assumptions about the unconditional variance-covariance matrix.

- Therefore, testing for serial correlation, or for constant variances across time, is straightforward.

- Consider testing for AR(1) serial correlation:

$$u_t = \rho u_{t-1} + e_t \tag{2.19}$$

$$H_0 : \rho = 0 \tag{2.20}$$

- The natural steps for testing (2.20) (with strictly exogenous regressors) are (1) Run pooled OLS of y_{it} on $\mathbf{x}_{it}$, $t = 1, \ldots, T$; $i = 1, \ldots, N$, and obtain the POLS residuals (and one lag). (2) Run the POLS regression $\hat{u}_{it}$ on $\hat{u}_{i,t-1}$; $t = 2, \ldots, T$; $i = 1, \ldots, N$, and use either the usual t statistic or that made robust to heteroskedasticity.

• The heteroskedasticity-robust form is robust to changes in the unconditional variance of u_{it} as well as dynamic heteroskedasticit in $Var(u_{it}|u_{i,t-1})$. (ARCH)

• If the $\mathbf{x}_{it}$ are not strictly exogenous, the test needs to be adjusted. Can use the (heteroskedasticity-robust) t statistic for $\hat{\rho}$ from the regression

$$\hat{u}_{it} \text{ on } \hat{u}_{i,t-1}, \mathbf{x}_{it}, \; t = 2,\ldots,T; i = 1,\ldots,N. \tag{2.21}$$

• In effect, this accounts for the possibility that $u_{i,t-1}$ is correlated with $\mathbf{x}_{it}$, which must happen when $\mathbf{x}_{it}$ contains lagged dependent variables but could happen other times, too.

• To test for constant variance over time, regress the squared OLS residuals on time period dummies:

$$\hat{u}_{it}^2 \text{ on } 1, d2_t, \ldots, dT_t,\ t = 1, \ldots, T;\ i = 1, \ldots, N \tag{2.22}$$

and use a joint F test.

• Can show under $E(y_{it}|\mathbf{x}_{it}) = \mathbf{x}_{it}\boldsymbol{\beta}$ that the asymptotic distribution of the test statistic does not depend on that of the POLS estimator, $\hat{\boldsymbol{\beta}}_{POLS}$. But, there might be serial correlation in the squared errors (ARCH), or the fourth moment of u_{it} might not line up with the normal, so the joint test from (2.21) should be made "cluster robust."

- Can add functions of $\mathbf{x}_{it}$ to the regression in (2.21), too, such as the POLS fitted values and their squares.
- What do we do with the information? If we reject constant variances, maybe just use fully robust inference. But it could be used as motivation for GLS.

AIRFARE data:

```
. predict uh, resid

. sort id year

. gen uh_1 = uh[_n-1] if year > 1997
(1149 missing values generated)

. reg uh uh_1

      Source |       SS       df       MS              Number of obs =    3447
-------------+------------------------------           F(  1,  3445) =21752.96
       Model |    322.9744     1    322.9744           Prob > F      =  0.0000
    Residual |   51.149214  3445  .014847377           R-squared     =  0.8633
-------------+------------------------------           Adj R-squared =  0.8632
       Total |  374.123614  3446  .108567503           Root MSE      =  .12185

------------------------------------------------------------------------------
          uh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uh_1 |   .9072729   .0061515   147.49   0.000      .895212    .9193338
       _cons |  -1.33e-10   .0020754    -0.00   1.000    -.0040692    .0040692
------------------------------------------------------------------------------
```

```
. reg uh uh_1, robust

Linear regression                                      Number of obs =    3447
                                                       F(  1,  3445) =16321.87
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8633
                                                       Root MSE      =  .12185

------------------------------------------------------------------------------
             |               Robust
          uh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uh_1 |   .9072729   .0071015   127.76   0.000     .8933492    .9211966
       _cons |  -1.33e-10   .0020754    -0.00   1.000    -.0040692    .0040692
------------------------------------------------------------------------------
. * Very strong serial correlation. Using heteroskedasticity-robust version
. * does not change the outcome; with rhohat = .907, this should not
. * be surprising.
```

```
. reg uhsq y98 y99 y00

      Source |       SS       df       MS              Number of obs =    4596
-------------+------------------------------           F(  3,  4592) =    6.18
       Model |  .32784636     3  .10928212           Prob > F      =  0.0003
    Residual |  81.1617978  4592  .017674608           R-squared     =  0.0040
-------------+------------------------------           Adj R-squared =  0.0034
       Total |  81.4896441  4595  .017734417           Root MSE      =  .13295

------------------------------------------------------------------------------
        uhsq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         y98 |  -.0232182   .0055466    -4.19   0.000    -.0340923   -.0123441
         y99 |  -.0152361   .0055466    -2.75   0.006    -.0261101    -.004362
         y00 |  -.0158774   .0055466    -2.86   0.004    -.0267515   -.0050033
       _cons |   .1266466   .0039221    32.29   0.000     .1189574    .1343357
------------------------------------------------------------------------------

. * The F test, with p-value = .0003, assumes no serial correlation in the
. * squared errors and also that the fourth moments are constant over time.
. * Nevertheless, the rejection is strong.
```

```
. reg uhsq y98 y99 y00, cluster(id)

Linear regression                                      Number of obs =    4596
                                                       F(  3,  1148) =   35.42
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0040
                                                       Root MSE      =  .13295

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
        uhsq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         y98 |  -.0232182   .0024718    -9.39   0.000     -.028068   -.0183684
         y99 |  -.0152361   .0032039    -4.76   0.000    -.0215222     -.00895
         y00 |  -.0158774   .0032599    -4.87   0.000    -.0222734   -.0094814
       _cons |   .1266466   .0045367    27.92   0.000     .1177453    .1355478
------------------------------------------------------------------------------

. * The above F test (given p = .0000) is robust to serial correlation in the
. * squared errors and nonconstant fourth moments of the errors. Its rejection
. * is even stronger than the nonrobust test.
. * Later, we will use the estimated variances for the different time periods:

. predict sigsqh
(option xb assumed; fitted values)
```

2.4. FGLS with Strictly Exogenous Regressors

- If we detect serial correlation or heteroskedasticity, it is tempting to use a FGLS method to try to improve over POLS in the model

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it},\ t = 1, \ldots, T.$$

(Remember, with large N we can do valid inference with POLS. So, this is an efficiency issue.)

- If we use FGLS to account for serial correlation, strict exogeneity is key. If we make adjustments just for heteroskedasticity, contemporaneous exogeneity sufficies provided our estimated variance functions depend only on elements of $\mathbf{x}_{it}$.

- So, we might estimate a model for $Var(u_{it}|\mathbf{x}_{it})$ without taking a stand on the fully dynamic mean, which means we allow for the possibility of serial correlation. If we use weighted least squares, with weights $1/\hat{h}(\mathbf{x}_{it})$ (the estimated variance function), we should use the fully robust variance matrix for two reasons: (i) There is likely to be serial correlation; (ii) Our model for $Var(u_{it}|\mathbf{x}_{it})$ might be wrong. If we can rule out serial correlation, it suffices to use the "robust" option for WLS.

- Under strict exogeneity, we might use a simple AR(1) correction.
- A panel version of Prais-Winsten method uses the OLS residuals in the regression

$$\hat{u}_{it} \text{ on } \hat{u}_{i,t-1}; t = 2, \ldots, T; i = 1, \ldots, N \tag{2.23}$$

to get $\hat{\rho}$. Then, define the quasi-differenced data, $\tilde{y}_{it} = y_{it} - \hat{\rho} y_{i,t-1}$ for $t \geq 2$, $\tilde{y}_{i1} = (1 - \hat{\rho}^2)^{1/2} y_{i1}$ (and similarly for $\tilde{\mathbf{x}}_{it}$). Finally, use pooled OLS to get FGLS:

$$\tilde{y}_{it} \text{ on } \tilde{\mathbf{x}}_{it}, t = 1, \ldots; i = 1, \ldots, N. \tag{2.24}$$

- Why might we use a fully robust ("cluster" robust) variance matrix in (2.24)? Not to account for the first-stage estimation of ρ. That does not affect the large-sample distribution of $\hat{\boldsymbol{\beta}}_{FGLS}$; it is as if we know ρ. Instead, it is because (i) The AR(1) model might be wrong and/or (ii) The system homoskedasticity assumption fails.
- The FGLS estimator *might* be more efficient that POLS even if the AR(1) model is not quite right. Maybe it accounts for "enough" of the serial correlation. But we should make our inference robust.

• The previous comment is the motivaton in the **generalized estimating equations** (**GEE**) literature. GEE is essentially FGLS (and certainly asymptotically equivalent to it) recognizing that our chosen variance matrix – such as the homoskedasticity AR(1) – might be incorrect. It also allows for unrestricted system heteroskedasticity in conducting inference.

• In the AR(1) case, not too hard to implement (2.24) “by hand,” but the “xtgee” command in Stata is convenient. If the “robust” option is not included, the inference is the same as FGLS under SGLS.3. With the “robust” option, the inference is robust to incorrect restrictions on $\mathbf{\Omega}$ – if any are imposed – along with system heteroskedasticity.

- Tradeoff between efficiency and consistency: The POLS estimator only requires

$$E(\mathbf{x}_{it}'u_{it}) = \mathbf{0}$$

while Prais-Winsten (and other methods that exploit serial correlation in estimation) effectively requires

$$E(\mathbf{x}_{i,t-1}'u_{it}) = \mathbf{0}, E(\mathbf{x}_{it}'u_{it}) = \mathbf{0}, E(\mathbf{x}_{i,t+1}'u_{it}) = \mathbf{0}$$

(except with some fluke cancellations).

• Can use other forms for $\mathbf{\Omega}$, too, or leave it fully unrestricted (attractive with large N, small T, to estimate $T(T+1)/2$ separate elements). In Stata, use an "unstructured" option in xtgee.

• Even with $\hat{\mathbf{\Omega}}$ unrestricted (with $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ strictly exogenous), still can use a fully robust variance matrix estimator to account for possible system heteroskedasticity.

AIRFARE EXAMPLE: Use weighted least squares to account for different variances over time, but make inference robust to serial correlation (and other kinds of heteroskedasticity). Then, use FGLS with an AR(1) model, with and without robust inference.

```
. * Use weighted least squares to adjust for different unconditional
. * variances over time. First, nonrobust inference.

. reg lfare concen ldist ldistsq y98 y99 y00 [w = 1/sigsqh]
(analytic weights assumed)
(sum of wgt is   4.0868e+04)

      Source |       SS       df       MS              Number of obs =    4596
-------------+------------------------------           F(  6,  4589) =  523.85
       Model |  354.010324     6  59.0017206           Prob > F      =  0.0000
    Residual |  516.866129  4589  .112631538           R-squared     =  0.4065
-------------+------------------------------           Adj R-squared =  0.4057
       Total |  870.876453  4595  .189526976           Root MSE      =  .33561

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3592068   .0300054    11.97   0.000     .3003817    .4180319
       ldist |  -.9008375   .1279271    -7.04   0.000    -1.151636   -.6500389
     ldistsq |   .1028932   .0096992    10.61   0.000     .0838781    .1219083
         y98 |   .0211325   .0141639     1.49   0.136    -.0066355    .0489006
         y99 |   .0378426   .0144068     2.63   0.009     .0095984    .0660868
         y00 |     .09986   .0143893     6.94   0.000     .0716501    .1280698
       _cons |   6.210433    .419516    14.80   0.000      5.38798    7.032886
------------------------------------------------------------------------------
```

```
. * Make inference robust to any serial correlation and additional
. * heteroskedasticity.

. reg lfare concen ldist ldistsq y98 y99 y00 [w = 1/sigsqh], cluster(id)
(analytic weights assumed)
(sum of wgt is   4.0868e+04)

Linear regression                                      Number of obs =    4596
                                                       F(  6,  1148) =  205.89
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4065
                                                       Root MSE      =  .33561

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3592068   .0584782     6.14   0.000     .2444707    .4739428
       ldist |  -.9008375   .2710967    -3.32   0.001    -1.432738   -.3689368
     ldistsq |   .1028932   .0200969     5.12   0.000     .0634624     .142324
         y98 |   .0211325   .0041453     5.10   0.000     .0129994    .0292657
         y99 |   .0378426    .005181     7.30   0.000     .0276773    .0480079
         y00 |     .09986   .0056486    17.68   0.000     .0887772    .1109427
       _cons |   6.210433   .9088932     6.83   0.000     4.427155    7.993711
------------------------------------------------------------------------------
```

```
. xtgee lfare concen ldist ldistsq y98 y99 y00, corr(ar1)

GEE population-averaged model                   Number of obs      =      4596
Group and time vars:                 id year    Number of groups   =      1149
Link:                               identity    Obs per group: min =         4
Family:                             Gaussian                   avg =       4.0
Correlation:                           AR(1)                   max =         4
                                                Wald chi2(6)       =   1157.88
Scale parameter:                     .1136252   Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2173983   .0279859     7.77   0.000     .1625469    .2722497
       ldist |  -.9000279   .2408907    -3.74   0.000    -1.372165   -.4278908
     ldistsq |   .1009652   .0182148     5.54   0.000     .0652649    .1366655
         y98 |   .0223992   .0041045     5.46   0.000     .0143545    .0304439
         y99 |   .0367543   .0056737     6.48   0.000     .0256341    .0478746
         y00 |   .0983042   .0068041    14.45   0.000     .0849684    .1116399
       _cons |   6.379169   .7915448     8.06   0.000      4.82777    7.930569
------------------------------------------------------------------------------
```

```
. xtgee lfare concen ldist ldistsq y98 y99 y00, corr(ar1) robust

GEE population-averaged model                   Number of obs      =      4596
Group and time vars:                 id year    Number of groups   =      1149
Link:                               identity    Obs per group: min =         4
Family:                             Gaussian                   avg =       4.0
Correlation:                           AR(1)                   max =         4
                                                Wald chi2(6)       =   1200.79
Scale parameter:                    .1136252    Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2173983   .0371709     5.85   0.000     .1445446     .290252
       ldist |  -.9000279   .2817608    -3.19   0.001    -1.452269    -.347787
     ldistsq |   .1009652   .0208502     4.84   0.000     .0600995    .1418309
         y98 |   .0223992   .0041428     5.41   0.000     .0142795    .0305189
         y99 |   .0367543   .0051472     7.14   0.000      .026666    .0468427
         y00 |   .0983042   .0055529    17.70   0.000     .0874207    .1091877
       _cons |   6.379169   .9472529     6.73   0.000     4.522588    8.235751
------------------------------------------------------------------------------
```

```
. xtgee lfare concen ldist ldistsq y98 y99 y00, corr(uns)

GEE population-averaged model                   Number of obs      =      4596
Group and time vars:                 id year    Number of groups   =      1149
Link:                               identity    Obs per group: min =         4
Family:                             Gaussian                   avg =       4.0
Correlation:                    unstructured                   max =         4
                                                Wald chi2(6)       =   1321.99
Scale parameter:                    .1135142    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2364893   .0249533     9.48   0.000     .1875817    .2853969
       ldist |  -.8806104   .2457398    -3.58   0.000    -1.362252   -.3989693
     ldistsq |   .0992803   .0185775     5.34   0.000     .0628691    .1356915
         y98 |   .0222287    .003546     6.27   0.000     .0152787    .0291786
         y99 |   .0369008   .0040047     9.21   0.000     .0290518    .0447499
         y00 |   .0985136   .0046874    21.02   0.000     .0893264    .1077008
       _cons |   6.313734   .8076273     7.82   0.000     4.730813    7.896654
------------------------------------------------------------------------------

. * The above estimates allow omega to be unrestricted, but maintain
. * system homoskedasticity for inference.
```

```
. xtgee lfare concen ldist ldistsq y98 y99 y00, corr(uns) robust

GEE population-averaged model                   Number of obs      =      4596
Group and time vars:                 id year    Number of groups   =      1149
Link:                               identity    Obs per group: min =         4
Family:                             Gaussian                   avg =       4.0
Correlation:                    unstructured                   max =         4
                                                Wald chi2(6)       =   1246.97
Scale parameter:                    .1135142    Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2364893   .0406545     5.82   0.000     .1568079    .3161706
       ldist |  -.8806104     .26696    -3.30   0.001    -1.403842   -.3573785
     ldistsq |   .0992803   .0197484     5.03   0.000     .0605741    .1379866
         y98 |   .0222287   .0041432     5.37   0.000     .0141082    .0303492
         y99 |   .0369008   .0051386     7.18   0.000     .0268293    .0469724
         y00 |   .0985136   .0055411    17.78   0.000     .0876533     .109374
       _cons |   6.313734   .8977898     7.03   0.000     4.554098     8.07337
------------------------------------------------------------------------------

. * The above inference is robust to system heteroskedasticity. The fully robust
. * confidence standard error for concen is quite a bit larger than the
. * nonrobust one: about .041 versus .025.
```

- When “robust” is used as an option, Stata labels the standard errors “semi-robust.” For linear models, there is no distinction between fully robust and semi-robust. But for certain kinds of nonlinear models, one distinguishes between standard errors that allow misspecification of the conditional mean – given fully robust standard errors – and those that only allow misspecification of the conditional variance – which are dubbed “semi-robust.”