10: Seemingly Unrelated Regressions (SUR)

GECO 6281 Advanced Econometrics 1 (Lab)

Patrick Mokre

Fall 2020

Recapitulation

- OLS and GLS
- Discrete Choice Modeling: LOGIT, PROBIT, TOBIT and Heckman's Selection Model
- Panel Data
- Instrumental Variable Regression
- ARDL Modeling
- Anderson-Hsiao, Arellano-Bond, Arellano-Bover/Blundell-Bond instrumentalization

Seemingly Unrelated Regressions

- One has a number of individual regressions, with different dependent and independent variables.
- However, the relationships captured in these regressions are correlated with each other.
- This correlation will materialize in the error terms.
- Estimating the system of regressions in a feasible generalized linear regression (fGLS), using the different variance-covariance matrix is **more efficient** than using stacked OLS.
- If one wanted to use the independent variables from some regression as the dependent variable in another, **simultaneous regresion modeling** is the appropriate generalization.

SUR 2

- The most intuitive simple example: $i \in N$ individuals, $t \in T$ periods, $k \in K$ covariates
- Actually, the covariates **do not have to be the same**, however, it is illustrative
- using SUR modeling makes sense if there is (a) a reasonable number of time periods and (b) the researcher believes the coefficients vary between individuals, and this difference matters.

There is N regressions

$$\begin{array}{ccccccc} y_{it} & = & X'_{it} & & \beta_i + \epsilon_{it}, & & i = 1, ..., N \\ T \times 1 & = & T \times K & K \times 1 & T \times 1 & & \end{array}$$

Gauss-Markov conditions are in general satisified, especially:

$$E(\epsilon_i \mid X_i) = 0$$
$$E(\epsilon_i \epsilon_i' \mid X) = \sigma^2 I_i$$

But:

$$E(\epsilon_{ir} \epsilon_{is}) = 0 \quad \forall r \neq s$$
$$E(\epsilon_{it} \epsilon_{jt}) = \omega_{ij} \neq 0 \quad \text{for some } i, j$$

SUR 3

To understand the GLS method, one stacks the equations in a **matrix of matrices**.

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

Figure 1: Stacked Regressions in Matrix notation

SUR 4

- Remember that $E(\epsilon_{it}\epsilon_{jt} \mid X) = \omega_{ij}$ and define a $N \times N$ matrix $\Sigma = \omega_{ij}$.
- The most popular estimation method is a **2 Step GLS**.
- In Step 1, run all N regressions individually. Use the residuals to estimate $\hat{\Sigma}$: $\hat{\omega}_{ij} = \frac{1}{T}\epsilon_i'\epsilon_j$
- In Step 2 run a GLS regression with a variance matrix $\hat{\Omega} = \hat{\Sigma} \otimes I_T$
- $\hat{\beta}_{GLS} = (X'(\hat{\Sigma}^{-1} \otimes I_T)X)^{-1}X'(\hat{\Sigma}^{-1} \otimes I_T)y$
- Alternatively, SUR can be estimated using maximum likelihood or iterative GLS.
- SUR is equivalent to OLS if Σ is diagonal (there is no covariance between the error terms).
- When each equation has the same covariates, the estimators are numerically equivalent to OLS.

SUR 5: Breusch-Pagan Test for Independence of equations

Breusch and Pagan (1980) present a Lagrange Multiplier test for independence of the regressions.

For N observations in M equations, $r_{m,l}$ denotes the estimated correlation between equation residuals.

$$\lambda = N \sum_{m=1}^{M} \sum_{l=1}^{m-1} r^2 ml$$

The test staistic is χ^2 distributed with $\frac{M(M-1}{2}$ degrees of freedom.

On a sidenote, the R^2 can be used to compare the explanatory power gain between nested models, but is in general not well-defined for GLS.

Slope heterogeneity

(Pesaran 2015: Chapter 29.4.2)

- With sufficiently large T (time periods), **heterogeneous slopes** can be measured for each individual.
- Sometimes the differential between slopes is the key information in a dataset, e.g. when analyzing wage differentials, international power relations, effects of political frameworks.
- At the same time, slope heterogeneity can be designed such that **common features** are not neglected.
- We need to assume stationary dependent and independent variables (for now).
- Most general ("descriptive") formulation: $y_{it} = x_{it}'\beta_{it} + u_{it}$.

Slope heterogeneity 2

Assume that β_{it} depends on a common value β as well as a random variable η_{it}, drawn from a distribution whose parameters **do not vary over N and T**.

$$\beta_{it} = \beta + \eta_{it}$$
$$E(\eta_i) = 0, E(\eta_i x'_{it}) = 0$$
$$E(\eta_i \eta_i) = \Omega_\eta, E(\eta_i \eta_j) = 0 \quad \forall i \neq j$$

Hsiao's example is most intuitive: $\beta_{it} = \beta + \eta_i + \lambda_t$.

Heterogenous Slopes in STATA: Mean Groups

Markus Eberhardt 2012: "Estimating panel time-series models with heterogeneous slopes."

$$y_{it} = \beta_i x_{it} + u_{it}$$
$$u_{it} = \alpha_{1i} + \lambda_i f_t + \epsilon_{it}$$
$$x_{it} = \alpha_{2i} + \lambda_i f_t + \gamma_i g_t + e_{it}$$

Only y and x are observed, all factors in u are unobserved, and ϵ_{it} is the error term.

The principle idea is to estimate N OLS regressions, the find a weighted average of the coefficients.

Heterogenous Slopes in STATA: Mean Groups 2

- Pesaran and Smith 1995: No cross-sectional dependency, but common linear trend
- Pesaran 2006: Cross-Sectional Dependencies and Unobservables in x_{it} (e.g. productivity shocks). Cross-Section Averaged Parameters cannot be interpreted meaningfully, but allow for consistent estimation of coefficients for observed variables.
- Eberhardt and Teal: 3 Step procedure, allows for estimation of coefficients for unobservables (important for production functions, *total factor productivity*)

ARDL with Heterogenous Slopes 1

(https://www.stata.com/meeting/switzerland18/slides/switzerland18_Ditzen.pdf)

Reminder of ARDL setup to test for long-run relationship, including Error Correction Model for short-run adjustment processes.

- Level-ARDL

$$y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=0}^{q} \beta_j' x_{t-j} + v_t$$

- Conditional ECM:

$$\Delta y_t = \alpha_0 + \alpha_1 t + \alpha_2 (y_{t-1} - \theta x_t) + \sum_{i=1}^{p-1} \psi_{yi} \Delta y_{t-i} + \sum_{i=0}^{q-1} \psi_{xi}' \Delta x_{t-i} + u_t$$

- $\theta = (\sum_{j=0}^{q} \beta_j)/\alpha_2$ and $\alpha_2 = 1 - \sum_{i=1}^{p} \phi_i$
- Pesaran's Common Correlated Effect (CCE) estimation: $\beta_i = \beta + v_i, \quad v_i \sim IID(0, \Omega_v)$ and $x_{it} = \gamma_i' f_t + u_{it}$ where u_{it} is allowed ot be serially correlated, where $\hat{\beta}_{MG} = 1/N \sum \hat{\beta}_i$.

ARDL with Heterogenous Slopes 2

Setup with individual slopes, a common factor and a heterogenous “factor loading”, including mean group estimation.

$$
\begin{aligned}
y_{it} &= \lambda_i y_{it-1} + \beta_i x_{it} + u_{it} \\
u_{it} &= \gamma_i' f_t + \varepsilon_{it} \\
\hat{\beta}_{MG} &= \frac{1}{N} \sum \beta_i, \quad \hat{\lambda}_{MG} = \frac{1}{N} \sum \lambda_i
\end{aligned}
$$

Note: For consistent estimation of both $\hat{\beta}_i$ and $\hat{\beta}_{MG}$, large N and T are necessary. If the common unobserved factor f_t is left out, the *omitted variable bias* can be substantial.

Individual fixed effects can be added, but are not the crucial point in the methodology.

$$
\begin{aligned}
\beta_i &= \beta + v_i, \quad v_i \sim IID(0, \Omega_\beta) \\
\lambda_i &= \lambda + \zeta_i, \quad \zeta_i \sim IID(0, \Omega_\zeta)
\end{aligned}
$$

ARDL with Heterogenous Slope 3

Formulation of a special error correction model (**Dynamic Common Correlated Effects** Estimation):

$$\Delta y_{it} = \phi_i(y_{it-1} - \theta_i x_{it}) - \sum_{j=1}^{p} \lambda_{ji} \Delta_j y_{it-j} - \sum_{j=0}^{q} \beta'_{ji} \Delta_j x_{it} + \sum_{j=0}^{r} \gamma_{ij} \bar{z_{it}} + u_{it}$$

Where

- θ the **long-run coefficients** from the OLS level regression.
- Δ_j the lag-length denomination, i.e. $\Delta_3 x_{it} = x_{it} - x_{it-3}$.
- $\hat{\phi}_i = (1 - \sum_{j=1}^{p} \hat{\lambda}_{ij})$.
- $\bar{z}_t = (\bar{y}_t, \bar{x}_t)$, the cross-sectional averages.

You want to install the `moremata`, `xtmg` and `xtdcce2` packages from SSC.