Applied Macroeconometrics

Nektarios Aslanidis

[publicacions]
urv

Applied Macroeconometrics

Nektarios Aslanidis

[publicacions]
urv

Tarragona, 2017

Publicacions de la Universitat Rovira i Virgili
Av. Catalunya, 35 - 43002 Tarragona
Tel. 977 558 474 · publicacions@urv.cat
www.publicacions.urv.cat

1st edition: May 2017
ISBN (paper): 978-84-8424-602-2
ISBN (PDF): 978-84-8424-603-9

Dipòsit legal: T 722-2017

Cite the book.

Find the book in our website.

PublicacionsURV is a member of the Unión de Editoriales Universitarias Españolas and the Xarxa Vives, which ensures the dissemination and sale of its publications nationally and internationally.

Contents

Chapter 1: Panel Data Models

1.1. Static Panel Data Models

Panel data are repeated measures on individuals (i) over time (t). A longitudinal dataset obtained by following a given sample of individual agents (or households, firms, cities, regions, countries etc) over time.

Examples:

Consumption function (data on households)

Cost function (data on firms)

Production function (data on firms)

Regress y_{it} on x_{it} for $i = 1,...,N$ and $t = 1,...,T$

id	year	yr92	yr93	yr94	DUM1	DUM2	Y	X
1	1992	1	0	0	1	0	55	70
1	1993	0	1	0	1	0	50	68
1	1994	0	0	1	1	0	66	80
2	1992	1	0	0	0	1	77	94
2	1993	0	1	0	0	1	85	100
2	1994	0	0	1	0	1	90	123
(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)

If all N individuals are observed at all time periods, then **balanced panel**. If there are missing observations, then **unbalanced panel**. Analyzing unbalanced panel data typically raises few additional issues compared with the analysis of balanced data. However, if the panel is unbalanced for reasons that are not entirely random (e.g. because firms with relatively low levels of productivity have relatively high exit rates), then we may need to take this into account when estimating the model. This can be done by means of a sample selection model. We abstract from this particular problem here.

Repeated cross sections are not the same as panel data. Repeated cross sections are obtained by sampling from the same population at different points in time. The identity of the individuals (or firms, households etc.) is not recorded, and there is no attempt to follow individuals over time. This is the key reason why pooled cross sections are different from panel data. Even with identical sample sizes, the use if a panel data set will often yield more efficient estimators than a series of independent/repeated cross-sections.

Example

$y_{it}=\lambda_t+\alpha_i+\varepsilon_{it}$ (random effects)

Suppose we are interested in the change of λ_t from one period to another. Then, the variance of the estimator $\hat{\lambda}_t-\hat{\lambda}_s\ (s\neq t)$ is given by

$$Var(\hat{\lambda}_t-\hat{\lambda}_s)=Var(\hat{\lambda}_t)+Var(\hat{\lambda}_s)-2Cov(\hat{\lambda}_t,\hat{\lambda}_s)$$

with $\hat{\lambda}_t=N^{-1}(y_{1t}+...y_{Nt})$, $t=1,..,T$

$(\hat{\lambda}_1=N^{-1}(y_{11}+...+y_{N1})$, $\hat{\lambda}_2=N^{-1}(y_{12}+...+y_{N2})$,..., $\hat{\lambda}_T=N^{-1}(y_{1T}+...+y_{NT}))$

$$Cov\left(\frac{1}{N}(y_{11}+...+y_{N1}),\frac{1}{N}(y_{12}+...+y_{N2})\right)$$

Assuming cross-sectional independence

$$= \frac{1}{N^2}\left(Cov(y_{11}, y_{12}) + \ldots + Cov(y_{N1}, y_{N2})\right) = \frac{1}{N^2} N\sigma_\alpha^2 = \frac{\sigma_\alpha^2}{N}$$

Therefore, $Cov(\hat{\lambda}_t, \hat{\lambda}_s) > 0$ in panel data but $Cov(\hat{\lambda}_t, \hat{\lambda}_s) = 0$ in repeated cross sections. Thus, if one is interested in changes from one period to another, a panel will yield more efficient estimators than a series of cross-sections.

Three specializations to general panel methods:

1. Short panels (Micro Panels): assumed, with T small and $N \to \infty$. Data on many individual units and few time periods.

2. Long panels (Macro Panels): assumed, with $T \to \infty$ and N small or $N \to \infty$. Time series data on many individual units. More common with aggregate data.

3. Dynamic models: regressors include lagged dependent variables.

Examples of Micro Panel data

- Panel Study of Income Dynamics (PSID)
 (https://psidonline.isr.umich.edu)

- The European Community Household Panel (ECHP)
(http://ec.europa.eu/eurostat/web/microdata/european-community-household-panel)

Examples of Macro Panel data

- Federal Reserve Bank of St. Louis
(https://fred.stlouisfed.org/)

- Yahoo Finance
(http://finance.yahoo.com)

- Penn World Table (PWT). Provides purchasing power parity and national income accounts converted to international prices for 188 countries over the last six decades. (httt://pwt.econ.upenn.edu)

- World Bank, World Development Indicators (WDI). Provides more than 900 indicators for 152 economies. (www.worldbank.org/data)

- International Monetary Fund (IMF), World Economic Outlook Databases & International Financial Statistics (IFS) provide more than 32000 time series covering more than 200 countries. (www.imf.org)

- Organization for Economic Co-operation and Development (OECD) (www.oecd.org)

- European Central Bank (ECB)
(http://www.ecb.int)

Consider the following panel data model

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}, \qquad (1)$$

Note that $i = 1,...,N$ denotes the individual, firm, country and so on, and $t = 1,...,T$ is the time period. The term α_i denotes unobservable individual specific effects and ε_{it} denotes the remainder disturbance assumed to be independently and identically distributed (IID).

Advantages of panel data

1. More data compared to time series or cross-sections, more variability/more informative data as variables vary over two dimensions, less collinearity among regressors, and more efficiency. Time series data suffer from multicollinearity. This is less likely in panel data since the cross-section dimension adds a lot of variability. In fact, the variation in the data can be decomposed into variation between cross sections and variation within cross sections. The former variation is usually bigger.

2. Reduces the data needs. The richness of panel data obviates the need for data on things that may be difficult or impossible to measure (unobserved heterogeneity).

Example: Wage regression

$$wage_{it} = \alpha + \beta educ_{it} + \gamma abil_i + \varepsilon_{it}$$

where $abil_i$ denotes innate ability (constant through time), which cannot be observed. Thus, run OLS

$$wage_{it} = \alpha + \beta educ_{it} + w_{it}, \quad \text{where } w_{it} = \gamma abil_i + \varepsilon_{it}$$

If innate ability is not correlated with education, then $\gamma abil_i$ is just another unobserved factor making up the residual. It is true that OLS will not be a Best Linear Unbiased Estimator (BLUE), because the error term

$w_{it} = \gamma abil_i + \varepsilon_{it}$ is serially correlated (see below). Notice that OLS would be consistent, however, and the only substantive problem with relying on OLS for this model is that the standard formula for calculating the standard errors is wrong.

However, the problem is that innate ability might be correlated with education, in which case

$$E(w_{it} / educ_{it}) \neq 0 \Rightarrow Cov(educ_{it}, w_{it}) \neq 0$$

OLS will be inconsistent (unbiased regardless of the sample size). In particular, it can be shown

$$p\lim \beta^{OLS} = \beta + \frac{Cov(educ_{it}, abil_i)}{Var(educ_{it})}$$

which shows that the OLS estimator is inconsistent unless $Cov(educ_{it}, abil_i) = 0$. If $Cov(educ_{it}, abil_i) > 0$ (positive correlation), then there is an upward bias. If the correlation is negative, we get a negative bias.

However, panel data can solve this problem by applying particular transformations to the data, which is not possible using cross-sectional data. For instance, write the model at time *t*-1

$$wage_{it-1} = \alpha + \beta educ_{it-1} + (\gamma abil_i + \varepsilon_{it-1})$$

$$wage_{it} = \alpha + \beta educ_{it} + (\gamma abil_i + \varepsilon_{it})$$

Subtracting the first from the second equation yields

$$(wage_{it} - wage_{it-1}) = \beta(educ_{it} - educ_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})$$

$$\Delta wage_{it} = \beta \Delta educ_{it} + \Delta\varepsilon_{it}$$

Innate ability has been eliminated because it does not vary through time.

Properties of $\Delta\varepsilon_{it}$

1. $E(\Delta\varepsilon_{it}) = 0$

2. $Var(\Delta\varepsilon_{it}) = Var(\varepsilon_{it} + (-\varepsilon_{it-1})) = Var(\varepsilon_{it}) + Var(-\varepsilon_{it-1}) = Var(\varepsilon_{it}) + (-1)^2 Var(\varepsilon_{it-1}) = 2\sigma_{\varepsilon}^2$

3. $Cov(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-1}) = E(\Delta\varepsilon_{it}\Delta\varepsilon_{it-1}) = E(\varepsilon_{it} - \varepsilon_{it-1})(\varepsilon_{it-1} - \varepsilon_{it-2}) = -E(\varepsilon_{it-1}^2) = -\sigma_{\varepsilon}^2$

$$Cov(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-2}) = E(\Delta\varepsilon_{it}\Delta\varepsilon_{it-2}) = E(\varepsilon_{it} - \varepsilon_{it-1})(\varepsilon_{it-2} - \varepsilon_{it-3}) = 0$$

$$Cov(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-s}) = 0, \quad s \geq 2$$

(First-order serial correlation!)

OLS will be consistent, though inefficient due to autocorrelation. This is the so-called **first-differenced** (FD) estimator.

3. Controls for parameter heterogeneity (related to the previous issue). Consider the following model:

$$wage_{it} = \alpha_i + \beta educ_{it} + \varepsilon_{it}$$

where the intercept term is specific to each individual (heterogeneous). What happens if we ignore this heterogeneity and mistakenly assume that the intercept is the same across individuals.

$$wage_{it} = (\mu - \mu) + \alpha_i + \beta educ_{it} + \varepsilon_{it}$$

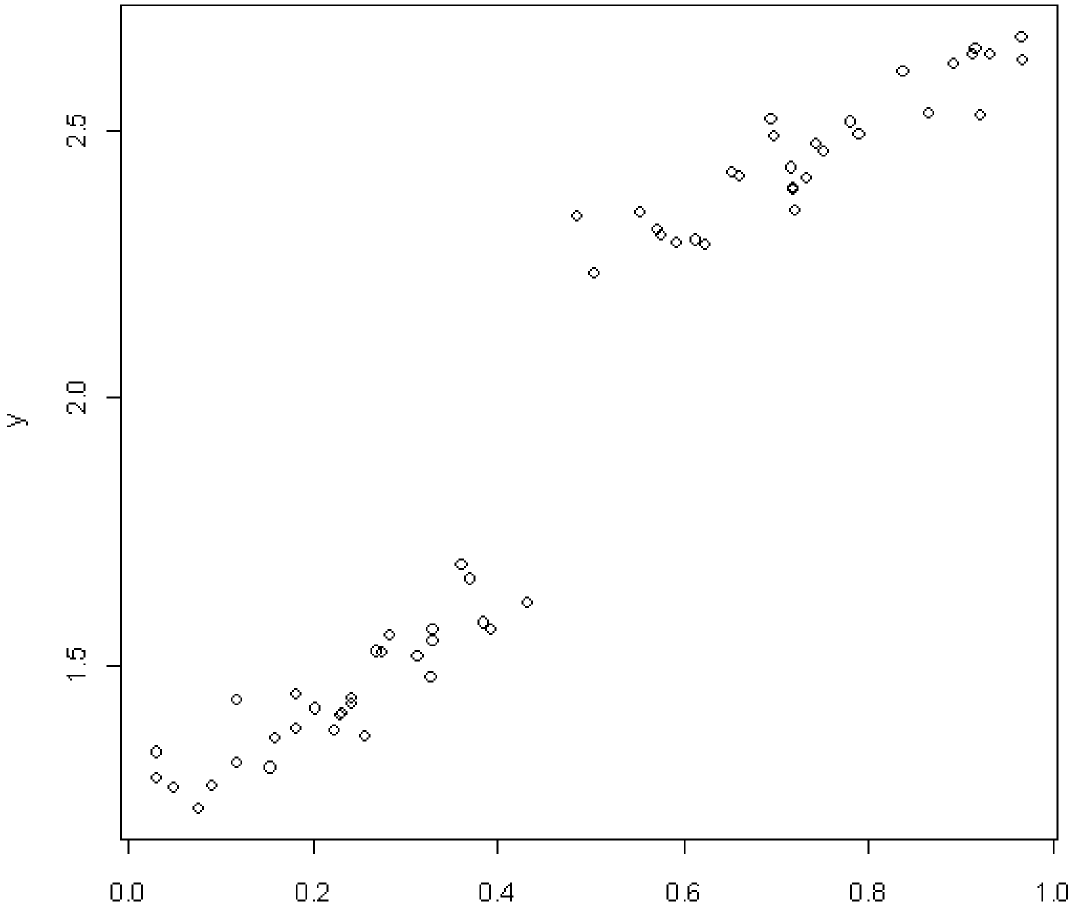
$$wage_{it} = \mu + \beta educ_{it} + w_{it}, \quad \text{where } w_{it} = \alpha_i - \mu + \varepsilon_{it}$$

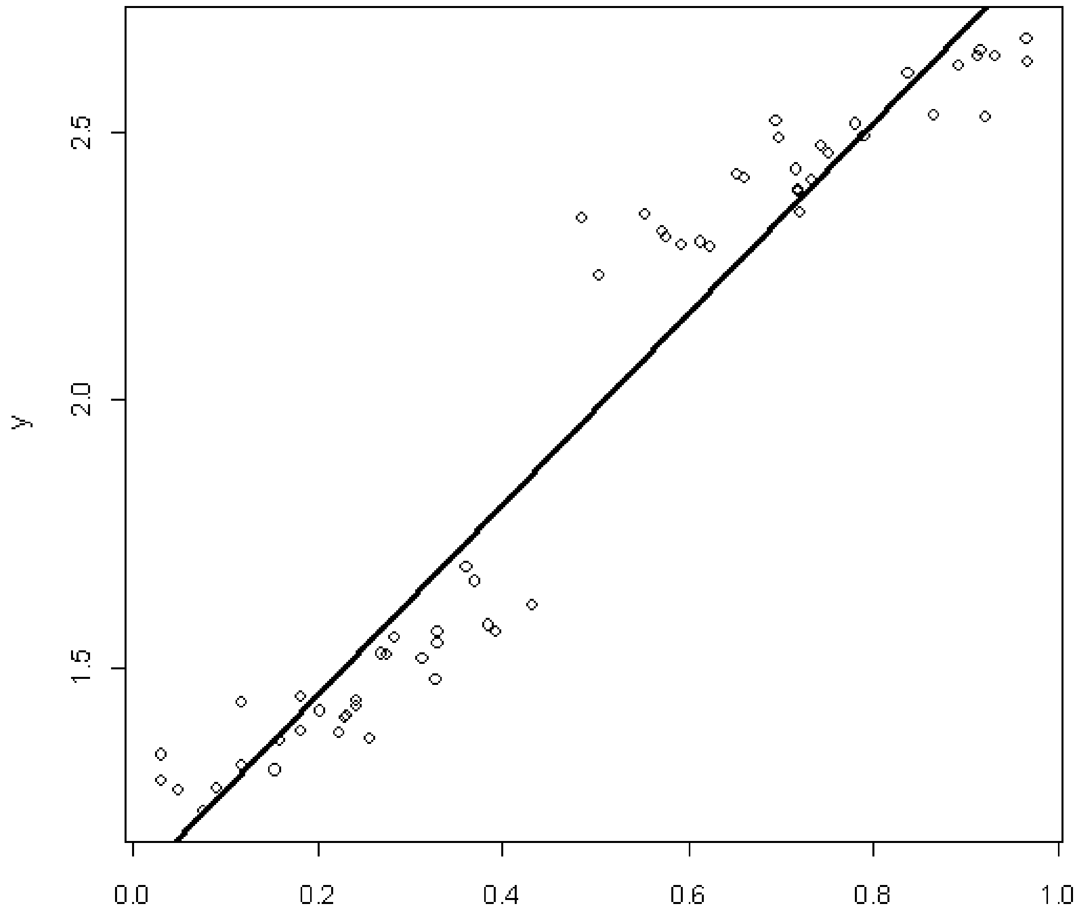
If the individual-specific intercepts are correlated with education, we will have

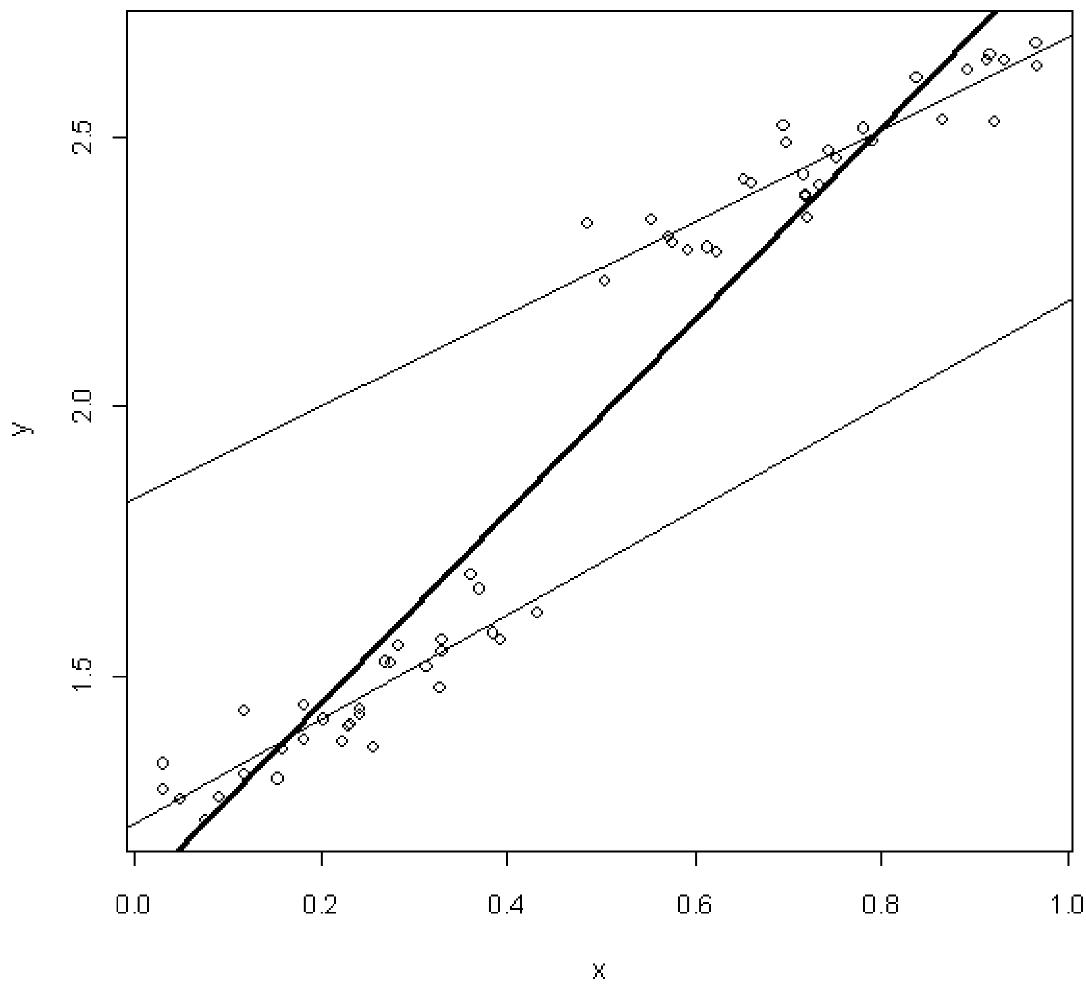
$$E(w_{it} / educ_{it}) \neq 0 \Rightarrow Cov(educ_{it}, w_{it}) \neq 0$$

Thus, OLS will be inconsistent.

-- See figures below --


y
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0


y
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0



Notice how closely related are the problems of omitted variables (individual-specific intercepts, which are time invariant) and unobserved heterogeneity (time invariant). You can always argue/set $\alpha_i - \mu = \gamma abil_i$.

1.2 The Fixed Effects ('Within') Estimator

One way to estimate the model is to assume that each α_i is a fixed/constant parameter to be estimated (just like β). The α_i thus capture the effects of those variables that are peculiar to the *i*-th individual and that are constant over time. This is called the **fixed effects** (FE) estimator. We may either allow in the model for individual-specific dummies,

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}, \qquad (\varepsilon_{it} \text{ is } IID)$$

$$y_{it} = \left(\sum_{j=1}^{N} \alpha_j d_j\right) + \beta x_{it} + \varepsilon_{it} \qquad (2)$$

We thus have a set of *N* dummies in the model. The parameters $\alpha_1, ..., \alpha_N$ and β can be estimated by OLS. It is straightforward to see how to test for whether the panel approach is really necessary at all. In other words, to test whether all of the intercept dummy variables have the same parameter,

$$H_0 : \alpha_i = \alpha_2 = ... \alpha_N \quad (N\text{-1 restrictions})$$

If this null hypothesis is not rejected, the data can simply be pooled together and standard OLS employed. If this null is rejected, however, then it is not valid to impose the restriction that the intercepts are the same over the cross-sectional units and a panel approach must be employed.

When *N* is large it may be numerically unattractive to have a regression with so many parameters to estimate. Fortunately, one can compute the estimator in a simpler way. It can be shown that exactly the same estimator for β is obtained if the regression is performed in deviations from individual means. Essentially, this implies that we eliminate the individual effects α_i first by transforming the data. To see this, note

$$\bar{y}_i = \alpha_i + \beta \bar{x}_i + \bar{\varepsilon}_i$$

where $\bar{y}_i = T^{-1}\sum_t y_{it}$ and similarly for the other variable. Consequently we can write

$$(y_{it} - \bar{y}_i) = (\alpha_i - \alpha_i) + \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$
$$(y_{it} - \bar{y}_i) = \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \qquad (3)$$

This regression involves demeaned variables and therefore does not include the individual effects α_i. So, transform the data in terms of deviations from individual-specific averages (Within Groups transformation is called because the subtraction is made within each cross-sectional unit) and remove the individual-specific (intercepts),

Both (2) and (3) can be estimated by OLS. The estimator is called **fixed effects** (FE), **least squares dummy variables** (LSDV) or **within** estimator.

The fixed effects estimator focuses on differences 'within' individuals. Put differently, it explains to what extent y_{it} differs from $\bar{y}_i$ and does not explain why $\bar{y}_i$ is different from $\bar{y}_j$. Note the assumptions about β impose that a change in x has the same (ceteris paribus) effect, whether it is a change from one period to the other or a change from one individual to the other.

The OLS estimator for β

$$\hat{\beta}^{FE} = \frac{\sum_N \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum_N \sum_t (x_{it} - \bar{x}_i)^2}$$
$$\hat{\beta}^{FE} = \left(\sum_N \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\right)^{-1} \sum_N \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)$$

(if $\hat{\beta}^{FE}$ was a vector)

Assumption 1: unobserved terms α_i can be freely correlated with x_{it}.

Assumption 2: $E(x_{it}\varepsilon_{is})=0$ for s = 1, 2, … *T* (strict exogeneity). Clearly, we cannot include y_{it-1} as a regressor.

Properties of $(\varepsilon_{it}-\bar{\varepsilon}_i)$

1. $E(\varepsilon_{it}-\bar{\varepsilon}_i)=E(\varepsilon_{it})-E(\bar{\varepsilon}_i)=E(\varepsilon_{it})-E(\frac{\varepsilon_{i1}+...+\varepsilon_{iT}}{T})=0$

2. $Var(\varepsilon_{it}-\bar{\varepsilon}_i)=Var(\varepsilon_{it})+Var(-\bar{\varepsilon}_i)=Var(\varepsilon_{it})+Var(-\frac{\varepsilon_{i1}+...+\varepsilon_{iT}}{T})$

$$=Var(\varepsilon_{it})+Var(-\frac{1}{T}(\varepsilon_{i1}+...+\varepsilon_{iT}))=Var(\varepsilon_{it})+\frac{1}{T^2}Var(\varepsilon_{i1}+...+\varepsilon_{iT})$$

$$=\sigma_\varepsilon^2+\frac{1}{T^2}T\sigma_\varepsilon^2=\sigma_\varepsilon^2+\frac{1}{T}\sigma_\varepsilon^2=\frac{(T+1)\sigma_\varepsilon^2}{T}$$

3. $Cov(\varepsilon_{it}-\bar{\varepsilon}_i,\varepsilon_{it-1}-\bar{\varepsilon}_i)=0$, since ε_{it} is *IID* across individuals and time.

Therefore, The FE estimator is unbiased and efficient.

We now see why this estimator requires strict exogeneity: the error term $\varepsilon_{it}-\bar{\varepsilon}_i=\varepsilon_{it}-\frac{\varepsilon_{i1}+...+\varepsilon_{iT}}{T}$ contains all residuals whereas the transformed explanatory variable(s) contains all values of the explanatory variable(s) $x_{it}-\bar{x}_i=x_{it}-\frac{x_{i1}+...+x_{iT}}{T}$. Hence, we need $E(x_{it}\varepsilon_{is})=0$ for s = 1, 2, … *T*; or there will be endogeneity bias if we estimate by OLS.

In the within estimator, the individual-specific intercepts can be estimated as,

$$\hat{\alpha}_i=\bar{y}_i-\hat{\beta}^{FE}\bar{x}_i, \qquad i=1,...,N$$

Note that as $T\to\infty$, the FE estimator of both $\alpha_i\ (i=1,...,N)$ and β is consistent. However, if T is fixed and $N\to\infty$ as is typical micro panels, then only the FE estimator of β is consistent. The FE estimator of α_i is inconsistent because the number of individual-specific intercepts increases to infinity as $N \to \infty$.

The covariance matrix for $\hat{\beta}^{FE}$ (vector)

$$Var(\hat{\beta}^{FE}) = \sigma_\varepsilon^2 \left(\sum_N \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1}$$

with

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)} \sum_N \sum_t (y_{it} - \hat{\alpha}_i - x'_{it} \hat{\beta}^{FE})^2$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)} \sum_N \sum_t (y_{it} - \bar{y}_i - (x_{it} - \bar{x}_i)' \hat{\beta}^{FE})^2$$

It is possible to apply the usual degrees of freedom correction in which case the number of explanatory variables is subtracted from the denominator. How many degrees of freedom? *NT-N-k* where *k* is the number of explanatory variables. Note the **least squares dummy variables** (LSDV) method estimates *N+k* parameters, or put differently, the **within** estimator uses a further *N* degrees of freedom in constructing the demeaned variables (we constructed *N* individual means).

Under weak regularity conditions, the fixed effects estimator is asymptotically normal, so standard inference can be applied.

The **within** estimator regression will give identical parameters and standard errors as would have been obtained directly from the LSDV regression, but without the hassle of estimating so many parameters. The disadvantage of **within** estimator regression, however, is that we lose the ability to determine the influences of all of the variables that affect the dependent variable but do not vary over time. For example, consider

$$pol_{it} = \alpha_i + \beta GDP_{it} + \gamma Ineq_i + \varepsilon_{it} ,$$

Averaging over time

$$\overline{pol}_i = \alpha_i + \beta \overline{GDP}_{it} + \gamma Ineq_i + \bar{\varepsilon}_i$$

Consequently we can write

$$(pol_{it} - \overline{pol}_i) = (\alpha_i - \alpha_i) + \beta(GDP_{it} - \overline{GDP}_i) + \gamma(Ineq_i - Ineq_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

$$(pol_{it} - \overline{pol}_i) = \beta(GDP_{it} - \overline{GDP}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

1.3 The Between Estimator

An alternative to the within estimator (fixed effects) would be to simply run a cross-sectional regression on the time-averaged data, which is know as **between estimator**,

$$\bar{y}_i = \alpha_i + \beta \bar{x}_i + \bar{\varepsilon}_i , \; i = 1, \ldots, N$$

An advantage of the **between estimator** over the **within estimator** is that this averaging often reduces the effect of measurement error in the variables on the estimation process

1.4 The First-Differenced (FD) Estimator

Another way to estimate the model is to use the first-differenced estimator

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta \varepsilon_{it}$$

Clearly this removes the individual fixed effect, and so we can obtain consistent estimates of β by estimating the equation in first differences by OLS.

Assumption 1: unobserved terms α_i can be freely correlated with x_{it} .

Assumption 2: $E(x_{it}\varepsilon_{is})=0$ for s = *t*, *t*-1. This is a weaker form of strict exogeneity than what is required for fixed-effects (FE), in the sense that $E(x_{it}\varepsilon_{it-2})=0$; for example, is not required). Thus, if there is feedback from ε_{is} to x_{it} that takes more than two periods, FD will be consistent whereas FE will not (hence weaker form of strict exogeneity).

You now see why this estimator requires exogeneity: the error term contains ε_{it} and ε_{it-1}, whereas the vector of transformed explanatory variable(s) contains x_{it} and x_{it-1} : Hence, we need $E(x_{it}\varepsilon_{is})=0$ for s = *t*, *t*-1; or there will be endogeneity bias if we estimate by OLS.

Important: FE versus FD.

So, FE and FD are two alternative ways of removing the fixed effect. Which method should we use? In general, FD is consistent but inefficient (due to autocorrelation).

(i) The FD and FE estimators are the same if *T*=2 (i.e. we have only two time periods).

Proof

FD: $(y_{it}-y_{it-1})=\beta(x_{it}-x_{it-1})+(\varepsilon_{it}-\varepsilon_{it-1})$

Note that there is just one cross-section!

T=2 $(y_{i2}-y_{i1})=\beta(x_{i2}-x_{i1})+(\varepsilon_{i2}-\varepsilon_{i1})$

We cannot have autocorrelation. Thus, OLS is consistent and efficient.

FE: $(y_{it}-\bar{y}_i)=\beta(x_{it}-\bar{x}_i)+(\varepsilon_{it}-\bar{\varepsilon}_i)$

T=1 $(y_{i1}-\frac{y_{i1}+y_{i2}}{2})=\beta(x_{i1}-\frac{x_{i1}+x_{i2}}{2})+(\varepsilon_{i1}-\frac{\varepsilon_{i1}+\varepsilon_{i2}}{2})$

$$(\frac{y_{i1}-y_{i2}}{2})=\beta(\frac{x_{i1}-x_{i2}}{2})+(\frac{\varepsilon_{i1}-\varepsilon_{i2}}{2})$$

T=2 $(y_{i2}-\frac{y_{i1}+y_{i2}}{2})=\beta(x_{i2}-\frac{x_{i1}+x_{i2}}{2})+(\varepsilon_{i2}-\frac{\varepsilon_{i1}+\varepsilon_{i2}}{2})$

$$(\frac{y_{i2}-y_{i1}}{2})=\beta(\frac{x_{i2}-x_{i1}}{2})+(\frac{\varepsilon_{i2}-\varepsilon_{i1}}{2})$$

So, one of the 2 cross-sections is redundant.

(ii) However, for T>2, the FD and FE estimators are NOT the same.

Under "classical assumptions", i.e. $\varepsilon_{it} \sim IID(0,\sigma_{\varepsilon}^{2})$, the FE estimator will be more efficient than the FD estimator (as in this case the FD residual ε_{it} will exhibit negative serial correlation, $E(\Delta\varepsilon_{it}\Delta\varepsilon_{it-1})=-\sigma_{\varepsilon}^{2}$).

Under the null hypothesis that the model is correctly specified, FE and FD will differ only because of sampling error. Hence, if FE and FD are significantly different - so that the differences in the estimates cannot be attributed to sampling error - we should worry about the validity of the strict exogeneity assumption.

Note that strict exogeneity rules out feedback from past ε_{is} shocks to current x_{it}. One implication of this is that FE and FD will not yield consistent estimates if the model contains lagged dependent variables (dynamics models). In this case, we may be able to use instruments to get consistent estimates.

1.5 An extension of the Fixed Effects Estimator

Consider the a fixed effects model with a two-way error component

$$y_{it} = \alpha_i + \lambda_t + \beta x_{it} + \varepsilon_{it}, \quad (\varepsilon_{it} \text{ is } IID)$$

$$y_{it} = (\textstyle\sum_{j=1}^{N} \alpha_j d_j^{id}) + (\sum_{t=1}^{T} \lambda_t d_t^{time}) + \beta x_{it} + \varepsilon_{it}$$

Note that λ_t denotes is individual-invariant and accounts for any time-specific effect that is not included in the regression. For example, it could account for strike year effects that disrupt production, oil price effects, macroeconomics and financial crisis effects, etc.

However, the number of parameters to be estimated now would be $k+N+T$, and the within transformation in this two-way model would be more complex.

1.6 The Pooled OLS Estimator

Consider

$$y_{it} = \beta x_{it} + (\alpha_i + \varepsilon_{it})$$

where I have put $(\alpha_i + \varepsilon_{it})$ within parentheses to emphasize that these terms are unobserved and are will not be estimated separately.

Assumption 1: unobserved terms α_i are uncorrelated with x_{it}.

Assumption 2: $E(x_{it}\varepsilon_{it}) = 0$ (contemporaneously uncorrelated). This is an even weaker form of strict exogeneity than what is required for FD and FE estimators in the sense that $E(x_{it}\varepsilon_{it-1}) = 0$; for example, is not required). Clearly under these assumptions, $w_{it}^{OLS} = \alpha_i + \varepsilon_{it}$ will be uncorrelated with x_{it}, implying we can estimate β consistently using OLS. In this context we refer to this as the **Pooled OLS (POLS) estimator**.

1.7 Random effects

Another way to estimate the model is to assume that each α_i is a random draw from a common distribution with a finite mean and finite variance (i.e. random factors *IID* distributed over individuals). Re-write,

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}$$

$$y_{it} = \alpha_i - u_i + \beta x_{it} + u_i + \varepsilon_{it}$$

$$y_{it} = (\alpha_i - u_i) + \beta x_{it} + (u_i + \varepsilon_{it})$$

$$y_{it} = \alpha + \beta x_{it} + w_{it}$$

where $\alpha_i = \alpha + u_i$ (thus, $\alpha = \alpha_i - u_i$) and $w_{it} = u_i + \varepsilon_{it}$

$$u_i \sim IID(0, \sigma_u^2), \qquad \varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2), \quad E(u_i \varepsilon_{it}) = 0$$

and u_i measures the random deviation of each individual's intercept term from the 'global' intercept term α.

Assumption 1: unobserved terms u_i are uncorrelated with x_{it}.

Assumption 2: $E(x_{it}\varepsilon_{is}) = 0$ for s = 1, 2, … T (strict exogeneity).

Note that this combines the strongest assumption underlying FE estimation (strict exogeneity) with the strongest assumption underlying POLS estimation (no correlation between unobserved effects and the explanatory variables).

There are no dummy variables to capture heterogeneity in the cross-sectional dimension. Instead, this occurs via the u_i terms.

Note: Under the above assumptions:

1) POLS will be consistent but inefficient because of omitted random effects problem u_i or because the composite error term $(u_i+\varepsilon_{it})$ is autocorrelated. Explores both the within and between dimension of the data.
2) FE will be consistent but inefficient due the fact that it explores only the within dimension of the data.
3) FD will be consistent but inefficient due to autocorrelation.

Properties of composite error $w_{it}=u_i+\varepsilon_{it}$

1. $E(w_{it}) = E(u_i) + E(\varepsilon_{it}) = 0 \qquad \forall i, \forall t$

2. $Var(w_{it}) = Var(u_i+\varepsilon_{it}) = Var(u_i) + Var(\varepsilon_{it}) = \sigma_u^2 + \sigma_\varepsilon^2 = \sigma_w^2 \qquad \forall i, \forall t$

3. $Cov(w_{it}, w_{it-1}) = E(w_{it} w_{it-1}) = E(u_i+\varepsilon_{it})(u_i+\varepsilon_{it-1}) = E(u_i^2 + u_i \varepsilon_{it-1} + \varepsilon_{it} u_i + \varepsilon_{it}\varepsilon_{it-1})$

$$= E(u_i^2) + E(u_i \varepsilon_{it-1}) + E(\varepsilon_{it} u_i) + E(\varepsilon_{it}\varepsilon_{it-1})$$

$$= \sigma_u^2 + 0 + 0 + 0 = \sigma_u^2 \qquad \forall i, \forall t$$

$$Cov(w_{it}, w_{it-s}) = E(w_{it} w_{it-s}) = E(u_i+\varepsilon_{it})(u_i+\varepsilon_{it-s}) = \sigma_u^2, \qquad s \geq 1 \quad \forall i, \forall t$$

(Higher-order serial correlation!)

That is, the correlation of the error terms over time is attributed to the individual effects u_i.

Also note that if σ_u^2 is high relative to σ_ε^2 the serial correlation in the error terms will be high. As a result the conventional estimator of the covariance matrix for the OLS estimator will not be correct.

Thus, the composite error is serially correlated, which implies that the optimal (most efficient) estimator should be a **Generalized Least Squares (GLS)** estimator. This is the so-called **random effects** (RE) estimator for panel data.

Derivation of the GLS-random effects estimator

(based on Hsiao C. (1986), Analysis of panel data, Cambridge University Press)

For individual i all errors can be stacked as

$$u_i \iota_T + \varepsilon_i$$

where $\iota_T = (1,1,\ldots,1)'$ of dimension T and $\varepsilon_i = (\varepsilon_{i1},\ldots,\varepsilon_{iT})'$

$$Var(u_i \iota_T + \varepsilon_i) = \Omega = \sigma_u^2 \iota_T \iota'_T + \sigma_\varepsilon^2 I_T$$

For each individual i we transform the data by premultiplying $y_i = (y_{i1},\ldots,y_{iT})'$ by

$$\Omega^{-1} = \sigma_\varepsilon^{-2}\left[I_T - \frac{\sigma_u^2}{\sigma_\varepsilon^2 + T\sigma_u^2}\iota_T \iota'_T \right] = \sigma_\varepsilon^{-2}\left[\left(I_T - \frac{1}{T}\iota_T \iota'_T \right) + \frac{\sigma_u^2}{\sigma_\varepsilon^2 + T\sigma_u^2}\frac{1}{T}\iota_T \iota'_T \right]$$

Thus, the GLS estimator is given by

$$\hat{\beta}^{GLS} = \left(\sum_N \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' + \frac{\sigma_u^2}{\sigma_\varepsilon^2 + T\sigma_u^2} T \sum_N (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \right)^{-1}$$
$$\times \left(\sum_N \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) + \frac{\sigma_u^2}{\sigma_\varepsilon^2 + T\sigma_u^2} T \sum_N (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y}) \right)$$

where $\bar{x} = \frac{\sum_{NT} x_{it}}{NT}$ is the overall sample average.

When $T \to \infty$ the term $\frac{\sigma_u^2}{\sigma_\varepsilon^2 + T\sigma_u^2} \to 0$

$$\hat{\beta}^{GLS} = \left(\sum_N \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\right)^{-1} \sum_N \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) = \hat{\beta}^{FE}$$

It can also be derived

$$\hat{\beta}^{GLS} = \Delta\hat{\beta}^B + (I - \Delta)\hat{\beta}^{FE}$$

where $\hat{\beta}^B = \left(\sum_N (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\right)^{-1} \sum_N (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})$ is the **between estimator** for β. It is the OLS estimator in the model for the individual means

$$\bar{y}_i = \alpha + \beta\bar{x}_i + (u_i + \bar{\varepsilon}_i),\ \ i = 1,\ldots,N$$

where Δ is the weighting matrix that is proportional to the inverse of the covariance matrix of $\hat{\beta}^B$. Thus, the GLS estimator is a matrix-weighted average of the between estimator and the within (fixed-effects) estimator, where the weight depends upon the relative variances of the two estimators.

The between estimator ignores any information within individuals. The GLS estimator, under Assumptions 1-2, is the optimal combination of the within and between estimators, and is therefore more efficient than either of these two estimators.

The RE estimator involves (as any other GLS estimator) running OLS on a "suitably transformed" model. The term "suitably transformed" means that the transformed model has serial uncorrelated errors. Therefore, OLS is the best linear unbiased estimator (BLUE) in this case. Averaging over time, in terms of unit means,

$$\bar{y}_i = \alpha + \beta \bar{x}_i + \bar{w}_i$$

Multiply by θ, where $\theta = 1 - \sqrt{\dfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2}}$

$$\theta \bar{y}_i = \theta\alpha + \theta\beta \bar{x}_i + \theta \bar{w}_i$$

Subtract this equation from the initial one. The transformed model is given by

$$(y_{it} - \theta \bar{y}_i) = \alpha(1-\theta) + \beta(x_{it} - \theta \bar{x}_i) + (w_{it} - \theta \bar{w}_i)$$

It can be shown

$$Cov\left(w_{it} - \theta \bar{w}_i, w_{it-1} - \theta \bar{w}_i\right) = E\left(u_i(1-\theta) + \varepsilon_{it} - \theta \bar{\varepsilon}_i\right)\left(u_i(1-\theta) + \varepsilon_{it-1} - \theta \bar{\varepsilon}_i\right) = 0$$

Note that (i) if $T \to \infty$, then $\dfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2} \to 0$ and $\theta \to 1$ and the RE (GLS) estimator tends to the **fixed effects** (FE) estimator (micro panel versus macro panel).

The above equation is very interesting because it involves quasi-demeaned data on each variable. In other words, rather than subtracting the entire individual mean (which is what the fixed effects does), the transformation subtracts only some fraction of the mean, as defined by θ. Notice that this implies that unobserved heterogeneity (as reflected by the individual-specific time-invariant effects) is not fully eliminated because

$$(w_{it} - \theta \overline{w}_i) = u_i(1-\theta) + (\varepsilon_{it} - \theta \overline{\varepsilon}_i)$$

As usual, GLS is unfeasible because we do not know the parameter θ. So, θ has to be estimated first. This involves estimating σ_u^2 and σ_ε^2. One way to do that, the simplest perhaps, is to use POLS in the first stage to obtain estimates of the composite residual $\hat{w}_{it}$ and its variance $\hat{\sigma}_w^2$. Based on this, we can calculate σ_u^2 as the covariance between $\hat{w}_{it}$ and $\hat{w}_{it-1}$ (for instance), and then calculate

$$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_w^2 - \hat{\sigma}_u^2 \;\; (\sigma_w^2 = \sigma_\varepsilon^2 + \sigma_u^2)$$

We can then plug $\hat{\sigma}_\varepsilon^2, \hat{\sigma}_u^2$ into the formula for θ

$$\hat{\theta} = 1 - \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2 + T\hat{\sigma}_u^2}}$$

Then, estimate the transformed equation.

$$(y_{it} - \hat{\theta}\overline{y}_i) = \alpha(1-\hat{\theta}) + \beta(x_{it} - \hat{\theta}\overline{x}_i) + (w_{it} - \hat{\theta}\overline{w}_i)$$

This is the **Feasible Generalized Least Squares (FGLS) estimator.**

Also, another consistent estimator of σ_ε^2 is obtained from the within residuals

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)} \sum_N \sum_t (y_{it} - \hat{\alpha}_i - x'_{it}\hat{\beta}^{FE})^2$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)} \sum_N \sum_t (y_{it} - \overline{y}_i - (x_{it} - \overline{x}_i)'\hat{\beta}^{FE})^2$$

Under weak regularity conditions, the random effects estimator is asymptotically normal with covariance matrix given by

$$Var(\hat{\beta}^{GLS}) = \sigma_\varepsilon^2 \left(\sum_N \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' + \frac{\sigma_u^2}{\sigma_\varepsilon^2 + T\sigma_u^2} T \sum_N (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \right)^{-1}$$

As long as $\frac{\sigma_u^2}{\sigma_\varepsilon^2 + T\sigma_u^2} > 0$, the random effects estimator is more efficient than the fixed effects estimator ($T \sum_N (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$ is positive definite). The gain in efficiency is due to using the between variation in the data $(\bar{x}_i - \bar{x})$. The covariance matrix is routinely estimated by the OLS expressions in the transformed model given above.

1.8 Fixed Effects or Random Effects

- Testing for non-zero correlation between the unobserved (individual) effect and the regressor(s): FE versus RE. The RE estimator requires that the individual effect must be uncorrelated with the regressors for it to be consistent. If this assumption is not tenable, the FE estimator should be used. In the present context, the FE estimator is consistent regardless of whether α_i is or is not correlated with x_{it}, while the RE requires this correlation to be zero in order to be consistent. Strict exogeneity is assumed for both models.

The Hausman statistic is computed as

$$H = (\hat{\beta}^{FE} - \hat{\beta}^{RE})'[Var(\hat{\beta}^{FE}) - Var(\hat{\beta}^{RE})]^{-1}(\hat{\beta}^{FE} - \hat{\beta}^{RE})$$

using matrix notation. Note that because the random effects estimator is efficient under the null

$$Var(\hat{\beta}^{FE} - \hat{\beta}^{RE}) = Var(\hat{\beta}^{FE}) - Var(\hat{\beta}^{RE})$$

Under the null hypothesis,

$$p\lim(\hat{\beta}^{FE} - \hat{\beta}^{RE}) \to 0$$

this test statistic follows a chi-squared distribution with *M* degrees of freedom, where *M* is the number of time explanatory variables in the model. In the case of a single slope parameter, the Hausman statistic is given by

$$H = \frac{(\hat{\beta}^{FE} - \hat{\beta}^{RE})^2}{Var(\hat{\beta}^{FE}) - Var(\hat{\beta}^{RE})} \sim \chi_1^2$$

Failing to reject the null hypothesis implies that the individual effects are uncorrelated with the explanatory variable(s). Thus, we may decide to use the RE model in the analysis on the grounds that this model is efficient. The null hypothesis is that both models are consistent, and a statistically significant difference is therefore interpreted as evidence against the RE model.

Also, in practice when computing the covariance matrix

$$Var(\hat{\beta}^{FE} - \hat{\beta}^{RE}) = Var(\hat{\beta}^{FE}) - Var(\hat{\beta}^{RE})$$

may not be positive definite in finite samples, such that the inverse cannot be computed.

- Is the key explanatory variable constant over time? In this case, the FE estimator may not so appropriate because the within transformation will eliminate this variable.

$$w\_im_{it} = \alpha_i + \beta ed_{it} + \gamma ed\_h_i + \varepsilon_{it}$$

$$\overline{w}\_im_i = \alpha_i + \beta \overline{ed}_i + \gamma ed\_h_i + \overline{\varepsilon}_i$$

$$(w\_im_{it} - \overline{w}\_im_i) = \beta(ed_{it} - \overline{ed}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

On the other hand, the RE estimator can control as many time-constant variables as possible.

- It is often argued that the RE model is more appropriate when the cross sections in the sample can be thought of a having been randomly selected from one population, but a FE model is more plausible when the cross sections effectively are the whole population (e.g., stocks traded on a particular exchange).

- Since there are fewer parameters to be estimated with the RE model (no dummy or within transformation to perform) and thus degrees of freedom are saved, the RE has an advantage.

- Are inferences made conditional on the effects that are in the sample or unconditional?

The FE estimator implies that inferences are made ‘conditional upon the effects of the model’. This means that we can only speak about those individuals included in the sample. That is, it essentially considers the distribution of y_{it} given α_i, the the fixed effects can be estimated. This makes sense intuitively if the individuals are 'one of a kind' and cannot be viewed as a random draw from the same underlying distribution (e.g., countries, large companies, etc). Inferences are with respect to the effects that are in the sample.

On the other hand, the RE estimator implies that inferences are made ‘unconditionally’. Basically, this is because in this model there is an implicit assumption that all individual effects come from a common distribution. Thus, the nature of the effect of any individual not included in the sample can be predicted. In fact, this question is related to the size of *N*. If *N* is small, the FE may be preferred, otherwise, the RE model is more sensible.

Thus, the random effects method allows one to make inference with respect to the population characteristics. One way to formalized this is the random effects model says

$$E(y_{it}|x_{it}) = \beta x_{it}$$

while for the fixed effects

$$E(y_{it}|x_{it},\alpha_i) = \alpha_i + \beta x_{it}$$

The parameter β in the two conditional expectations is the same only if $E(\alpha_i|x_{it}) = 0$. So, the reason why one may prefer fixed effects is that some interest lies in the alphas, which is the case if the number of individuals is relatively small and of a specific nature.

1.9 Mean Group (MG) estimator

Consider the following model

$$y_{it} = \alpha_i + \beta_i x_{it} + \varepsilon_{it}$$

Assumption: Parameter heterogeneity can be freely correlated with x_{it}.

Suppose we are interested in the average effect across individuals (the mean impact of α_i and β_i on y_{it}). The Mean Group (MG) estimator estimates the individual-specific time series by standard OLS and then averages these coefficients over individuals.

$$\beta^{MG} = \frac{1}{N}\sum_{i=1}^{N} \beta_i^{OLS}$$

The MG estimator is consistent and asymptotically normal for $N \to \infty$.

The variance of the MG estimator is given by

$$Var(\beta^{MG}) = \frac{1}{N(N-1)}\sum_{i=1}^{N} (\beta_i^{OLS} - \beta^{MG})^2$$

Standard inference applies.

The advantage of MG estimator is that we do not calculate the variances of the estimates for each individual (in this case, we would need to account for cross-sectional dependence, if there is any). Instead, we compute the variance of the estimates over individuals. A further advantage of MG is that we can accommodate unbalanced panels.

1.10 Dynamics Panel Data Models

An autoregressive panel data model, $AR(1)$

$$y_{it} = \gamma y_{it-1} + \alpha_i + \varepsilon_{it}, \qquad |\gamma| < 1$$

The fixed effects estimator for γ

$$\hat{\gamma}^{FE} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)(y_{it-1} - \bar{y}_{i,-1}) / NT}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - \bar{y}_{i,-1})^2 / NT}$$

where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i,-1} = T^{-1}\sum_{t=1}^{T} y_{it-1}$. Substitute $AR(1)$ into the estimator yields

$$\hat{\gamma}^{FE} = \gamma + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{it} - \bar{\varepsilon}_i)(y_{it-1} - \bar{y}_{i,-1}) / NT}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - \bar{y}_{i,-1})^2 / NT}$$

It can be shown

$$p\lim_{N\to\infty}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{it} - \bar{\varepsilon}_i)(y_{it-1} - \bar{y}_{i,-1})\right) = -\frac{\sigma_\varepsilon^2}{T^2}\frac{(T-1) - T\gamma + \gamma^T}{(1-\gamma)^2} \neq 0$$

For fixed T and $N \to \infty$, the fixed effects estimator is biased and inconsistent!

Example

» SIGMA=1;

» T=5;

» G=0.2;

```
» BIAS=-((T-1)-T*G+G^T)/((T^2)*(1-G)^2);
» BIAS;
```
-0.18752000

More persistent process

```
» G=0.8;
» BIAS=-((T-1)-T*G+G^T)/((T^2)*(1-G)^2);
» BIAS;
```
-0.32768000

Larger T

```
» T=100;
» BIAS=-((T-1)-T*G+G^T)/((T^2)*(1-G)^2);
» BIAS;
```
-0.047500000

Note inconsistency is not caused by anything we assumed about the alphas. The problem is that $Cov\left((\varepsilon_{it}-\bar{\varepsilon}_i),(y_{it-1}-\bar{y}_{i,-1})\right)\neq 0$.

However, if $T\rightarrow\infty$, then $-\frac{\sigma_\varepsilon^2}{T^2}\frac{(T-1)-T\gamma+\gamma^T}{(1-\gamma)^2}\rightarrow 0$! So, fixed effects is consistent when $N,T\rightarrow\infty$.

Take first difference and calculate

$$y_{it}-y_{it-1}=\gamma(y_{it-1}-y_{it-2})+(\varepsilon_{it}-\varepsilon_{it-1})$$

OLS is not consistent since $Cov(y_{it-1}, \varepsilon_{it-1}) \neq 0$ even when $T \to \infty$. This transformed model suggests IV estimation. Given $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$ (no autocorrelation), for example, use the instrument y_{it-2} as

$Cov\big((y_{it-1} - y_{it-2}), y_{it-2}\big) \neq 0$ (relevant)

$Cov\big((\varepsilon_{it} - \varepsilon_{it-1}), y_{it-2}\big) = 0$ (exogenous)

Thus, the IV estimator

$$\hat{\gamma}^{IV} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-2}(y_{it} - y_{it-1})}{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-2}(y_{it-1} - y_{it-2})}$$

A necessary condition for consistency

$$p\lim\left(\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}(\varepsilon_{it} - \varepsilon_{it-1}) y_{it-2}\right) = 0$$

for either $N \to \infty$ or $N, T \to \infty$.

An alternative estimator uses the instrument $(y_{it-2} - y_{it-3})$ as

$Cov\big((y_{it-1} - y_{it-2}), (y_{it-2} - y_{it-3})\big) \neq 0$ (relevant)

$Cov\big((\varepsilon_{it} - \varepsilon_{it-1}), (y_{it-2} - y_{it-3})\big) = 0$ (exogenous)

$$\hat{\gamma}^{IV} = \frac{\sum_{i=1}^{N}\sum_{t=3}^{T} (y_{it-2} - y_{it-3})(y_{it} - y_{it-1})}{\sum_{i=1}^{N}\sum_{t=3}^{T} (y_{it-2} - y_{it-3})(y_{it-1} - y_{it-2})}$$

which is consistent if

$$p\lim \frac{1}{N(T-2)}\sum_{N}\sum_{t}(\varepsilon_{it} - \varepsilon_{it-1})(y_{it-2} - y_{it-3}) = 0$$

$$p\lim\left(\frac{1}{N(T-2)}\sum_{i=1}^{N}\sum_{t=3}^{T}(\varepsilon_{it}-\varepsilon_{it-1})(y_{it-2}-y_{it-3})\right)=0$$

Which estimator do we use?

Use both adopting a GMM.

Note $p\lim\left(\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}(\varepsilon_{it}-\varepsilon_{it-1})y_{it-2}\right)=E\big((\varepsilon_{it}-\varepsilon_{it-1})y_{it-2}\big)=0$

$$p\lim\left(\frac{1}{N(T-2)}\sum_{i=1}^{N}\sum_{t=3}^{T}(\varepsilon_{it}-\varepsilon_{it-1})(y_{it-2}-y_{it-3})\right)=E\big((\varepsilon_{it}-\varepsilon_{it-1})(y_{it-2}-y_{it-3})\big)=0$$

are moment conditions. Both IV estimators impose one moment condition in estimation. Generally, imposing moment conditions increases the efficiency of the estimation.

Arellano and Bond (1991, *Review of Economic Studies*).

Example $T=4$

In period 2 $E\big((\varepsilon_{i2}-\varepsilon_{i1})y_{i0}\big)=0$

In period 3 $E\big((\varepsilon_{i3}-\varepsilon_{i2})y_{i1}\big)=0$, $E\big((\varepsilon_{i3}-\varepsilon_{i2})y_{i0}\big)=0$

In period 4 $E\big((\varepsilon_{i4}-\varepsilon_{i3})y_{i2}\big)=0$, $E\big((\varepsilon_{i4}-\varepsilon_{i3})y_{i1}\big)=0$, $E\big((\varepsilon_{i4}-\varepsilon_{i3})y_{i0}\big)=0$

GMM estimator

Define the vector of transformed error terms

$$\Delta\varepsilon_i=\begin{pmatrix}\varepsilon_{i2}-\varepsilon_{i1}\\ \ldots\\ \varepsilon_{iT}-\varepsilon_{iT-1}\end{pmatrix}$$

and the matrix of instruments

$$Z_i = \begin{pmatrix} (y_{i0}) & 0 \ldots & & \cdots 0 \\ 0 & (y_{i0}, y_{i1}) & & \cdots 0 \\ \vdots & & & \\ 0 \cdots 0 \cdots 0 & & & (y_{i0}, \ldots, y_{iT-2}) \end{pmatrix}$$

each row contains the instruments that are valid for a given period. Thus, we write compactly

$$E\left(Z'_i \Delta\varepsilon_i\right) = 0$$

$$E\left(Z'_i (\Delta y_i - \gamma \Delta y_{i,-1})\right) = 0$$

…

It can be shown that the GMM estimator is consistent and asymptotically normal.

An autoregressive panel data model with exogenous variables

$$y_{it} = \beta x_{it} + \gamma y_{it-1} + \alpha_i + \varepsilon_{it}$$

Use GMM. Take first difference and calculate

$$y_{it} - y_{it-1} = \beta(x_{it} - x_{it-1}) + \gamma(y_{it-1} - y_{it-2}) + (\varepsilon_{it} - \varepsilon_{it-1})$$

$$\Delta y_{it} = \beta \Delta x_{it} + \gamma \Delta y_{it-1} + \Delta\varepsilon_{it}$$

If x_{it} is strictly exogenous, we have

$$E(\Delta x_{it} \Delta\varepsilon_{it}) = 0, \ \forall t$$

and the matrix of instruments

$$Z_i = \begin{pmatrix} (y_{i0}, \Delta x_{i2}) \; 0 \; \ldots & & \cdots 0 \\ 0 & (y_{i0}, y_{i1}, \Delta x_{i3}) & \cdots 0 \\ \vdots & & \\ 0 \cdots 0 \cdots 0 & (y_{i0}, \ldots, y_{iT-2}, , \Delta x_{iT}) & \end{pmatrix}$$

…

Chapter 2: VAR Models

Since Sims (1980) critique of traditional macroeconometric modeling, vector autoregressive (*VAR*) models are widely used in macroeconomics. In the traditional approach the typical question asked is ‘What is the optimal response by the monetary authority to movements in macroeconomic variables to achieve given targets?’ Sims argued that a *VAR* model is an unrestricted model that treats all variables as endogenous “without restrictions based on supposed a priori knowledge” derived from theory.

2.1 Bivariate Structural Model

Let y_t, z_t endogenous in bivariate first order *structural* *VAR*(1)

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{yt} \quad (1)$$

$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt} \quad (2)$$

assumptions (a) y_t, z_t stationary processes (b) ε_{yt}, ε_{zt} white noise processes $\varepsilon_{yt} \sim WN(0, \sigma_y^2)$, $\varepsilon_{zt} \sim WN(0, \sigma_z^2)$ (c) ε_{yt} and ε_{zt} are uncorrelated.

There are feedback effects between y_t and z_t

Time lag effects

$\gamma_{12} \rightarrow$ time lag effect of z_{t-1} on y_t

$\gamma_{21} \rightarrow$ time lag effect of y_{t-1} on z_t

Contemporaneous effects

$-b_{12} \rightarrow$ contemporaneous effect of z_t on y_t

$-b_{21} \rightarrow$ contemporaneous effect of y_t on z_t

Derive reduced-form VAR(1)

(1), (2) $\Rightarrow$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} - \begin{bmatrix} 0 & b_{12} \\ b_{21} & 0 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

or

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

where

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix},\ x_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix},\ \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix},\ \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix},\ \varepsilon_t = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

Premultiply by B^{-1}

$$x_t = A_0 + A_1 x_{t-1} + e_t$$

where $A_0 = B^{-1}\Gamma_0,\ A_1 = B^{-1}\Gamma_1,\ e_t = B^{-1}\varepsilon_t$

Let $A_0 \equiv \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix},\ A_1 \equiv \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \Rightarrow$

$$y_t = a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t}$$
$$z_t = a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t}$$

but

$$B^{-1} = \frac{1}{\det(B)}\begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} = \frac{1}{1-b_{12}b_{21}}\begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}$$

transformed errors (reduced-form errors)

$$e_t = B^{-1}\varepsilon_t = \frac{1}{1-b_{12}b_{21}}\begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}\begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

$$= \frac{1}{1-b_{12}b_{21}}\begin{bmatrix} \varepsilon_{yt} - b_{12}\varepsilon_{zt} \\ -b_{21}\varepsilon_{yt} + \varepsilon_{zt} \end{bmatrix}$$

so

$$e_{1t} = \frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1-b_{12}b_{21}}, \quad e_{2t} = \frac{\varepsilon_{zt} - b_{21}\varepsilon_{yt}}{1-b_{12}b_{21}}$$

Properties of reduced-form errors

Mean

$$E(e_{1t}) = \frac{1}{1-b_{12}b_{21}}E(\varepsilon_{yt} - b_{12}\varepsilon_{zt}) = 0, \text{ since } E(\varepsilon_{yt}) = E(\varepsilon_{zt}) = 0$$

Variance

$$Var(e_{1t}) = E(e_{1t} - E(e_{1t}))^2 = E\left(\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right)^2$$

$$= \frac{1}{(1-b_{12}b_{21})^2} E(\varepsilon_{yt}^2 - 2b_{12}\varepsilon_{yt}\varepsilon_{zt} + (b_{12}\varepsilon_{zt})^2)$$

$$= \frac{1}{(1-b_{12}b_{21})^2} (E(\varepsilon_{yt}^2) - 2b_{12}E(\varepsilon_{yt}\varepsilon_{zt}) + b_{12}^2 E(\varepsilon_{zt}^2))$$

$$= \frac{\sigma_y^2 + b_{12}^2\sigma_z^2}{(1-b_{12}b_{21})^2}$$

since $Cov(\varepsilon_{yt}, \varepsilon_{zt}) = E(\varepsilon_{yt} - E(\varepsilon_{yt}))(\varepsilon_{zt} - E(\varepsilon_{zt}))$

$$= E(\varepsilon_{yt}\varepsilon_{zt}) = 0$$

Autocovariance

$$Cov(e_{1t}, e_{1,t-i}) = E(e_{1t} - E(e_{1t}))(e_{1,t-i} - E(e_{1,t-i}))$$

$$= E(e_{1t}e_{1,t-i}) = E\left(\frac{(\varepsilon_{yt} - b_{12}\varepsilon_{zt})(\varepsilon_{yt-i} - b_{12}\varepsilon_{zt-i})}{(1-b_{12}b_{21})^2}\right)$$

$$= \frac{1}{(1-b_{12}b_{21})^2} E(\varepsilon_{yt}\varepsilon_{yt-i} - b_{12}\varepsilon_{yt}\varepsilon_{zt-i} - b_{12}\varepsilon_{zt}\varepsilon_{yt-i} + b_{12}^2\varepsilon_{zt}\varepsilon_{zt-i})$$

$$= 0 \qquad \text{for } i \neq 0$$

since

$$E(\varepsilon_{yt}\varepsilon_{yt-i}) = -b_{12}E(\varepsilon_{yt}\varepsilon_{zt-i}) = -b_{12}E(\varepsilon_{zt}\varepsilon_{yt-i}) = b_{12}^2 E(\varepsilon_{zt}\varepsilon_{zt-i}) = 0$$

$$\Rightarrow e_{1t} \sim WN(0, \frac{\sigma_y^2 + b_{12}^2\sigma_z^2}{(1-b_{12}b_{21})^2}), \quad e_{2t} \sim WN(0, \frac{\sigma_z^2 + b_{21}^2\sigma_y^2}{(1-b_{12}b_{21})^2})$$

Cross-correlation (*covariance*)

$$Cov(e_{1t}, e_{2t}) = E(e_{1t} - E(e_{1t}))(e_{2t} - E(e_{2t}))$$

$$= E\left(\frac{(\varepsilon_{yt} - b_{12}\varepsilon_{zt})(\varepsilon_{zt} - b_{21}\varepsilon_{yt})}{(1-b_{12}b_{21})^2}\right)$$

$$= \frac{1}{(1-b_{12}b_{21})^2} E(\varepsilon_{yt}\varepsilon_{zt} - b_{21}\varepsilon_{yt}^2 - b_{12}\varepsilon_{zt}^2 + b_{12}b_{21}\varepsilon_{zt}\varepsilon_{yt})$$

$$= \frac{-b_{21}\sigma_y^2 - b_{12}\sigma_z^2}{(1-b_{12}b_{21})^2}$$

The errors in the reduced-form equations are correlated!

Only when $b_{12} = b_{21} = 0$, there is no correlation.

Variance-covariance matrix of the errors

$$\Sigma = E(e_t e'_t) = E\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}\begin{pmatrix} e_{1t} & e_{2t} \end{pmatrix} = \begin{bmatrix} E(e_{1t}^2) & E(e_{1t}e_{2t}) \\ E(e_{2t}e_{1t}) & E(e_{2t}^2) \end{bmatrix}$$

$$= \begin{bmatrix} Var(e_{1t}) & Cov(e_{1t}, e_{2t}) \\ Cov(e_{2t}, e_{1t}) & Var(e_{2t}) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

2.2 Multivariate Structural Model

Consider *K*-dimensional time series vector $x_t = (y_{1t}, ..., y_{Kt})'$ generated by reduced-form *VAR*(1)

$$\begin{aligned}
y_{1t} &= a_{10} + a_{11} y_{1t-1} + a_{12} y_{2t-1} + ... + a_{1K} y_{Kt-1} + e_{1t} \\
y_{2t} &= a_{20} + a_{21} y_{1t-1} + a_{22} y_{2t-1} + ... + a_{2K} y_{Kt-1} + e_{2t} \\
&. \\
&. \\
&. \\
y_{Kt} &= a_{K0} + a_{K1} y_{1t-1} + a_{K2} y_{2t-1} + ... + a_{KK} y_{Kt-1} + e_{Kt}
\end{aligned}$$

$$\Rightarrow \begin{bmatrix} y_{1t} \\ y_{2t} \\ \dots \\ y_{Kt} \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \\ \dots \\ a_{K0} \end{bmatrix} + \begin{bmatrix} a_{11} a_{12} \dots a_{1K} \\ a_{21} a_{22} \dots a_{2K} \\ \dots \\ a_{K1} a_{K2} \dots a_{KK} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ \dots \\ y_{Kt-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ \dots \\ e_{Kt} \end{bmatrix}$$

$$x_t = A_0 + A_1 x_{t-1} + e_t$$

where $x_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \dots \\ y_{Kt} \end{bmatrix}$, $A_0 = \begin{bmatrix} a_{10} \\ a_{20} \\ \dots \\ a_{K0} \end{bmatrix}$, $A_1 = \begin{bmatrix} a_{11} a_{12} \dots a_{1K} \\ a_{21} a_{22} \dots a_{2K} \\ \dots \\ a_{K1} a_{K2} \dots a_{KK} \end{bmatrix}$, $e_t = \begin{bmatrix} e_{1t} \\ e_{2t} \\ \dots \\ e_{Kt} \end{bmatrix}$

$$e_t \sim WN(0, \Sigma)$$

$E[e_t] = 0$, $E[e_t e'_t] = \Sigma$, and $E[e_t e'_s] = 0$ for $s \neq t$ where variance-covariance matrix Σ is time-invariant, symmetric, non-singular

$$\Sigma = E(e_t e'_t) = E\begin{bmatrix} e_{1t} \\ e_{2t} \\ \cdots \\ e_{Kt} \end{bmatrix}\begin{bmatrix} e_{1t} & e_{2t} & \ldots & e_{Kt} \end{bmatrix}$$

$$= \begin{bmatrix} E(e_{1t}^2) & \ldots & E(e_{1t}e_{Kt}) \\ E(e_{2t}e_{1t}) & \ldots & E(e_{2t}e_{Kt}) \\ \cdots & & \\ E(e_{Kt}e_{1t}) & \ldots & E(e_{Kt}^2) \end{bmatrix} = \begin{bmatrix} \sigma_1^{\ 2} & \ldots & \sigma_{1K} \\ \sigma_{21} & \ldots & \sigma_{2K} \\ \cdots & & \\ \sigma_{K1} & \ldots & \sigma_K^2 \end{bmatrix}$$

More generally, the *structural VAR*(p)

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \ldots + \Gamma_p x_{t-p} + \varepsilon_t$$

where B is the $(K \times K)$ matrix of contemporaneous (structural) effects, Γ_j $(j = 1,\ldots,p)$ are $(K \times K)$ matrices of (structural) lagged coefficients, and ε_t is the vector of structural errors.

Reduced-form *VAR*(p)

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \ldots + A_p x_{t-p} + e_t$$

where $A_j = B^{-1}\Gamma_j$ $(j = 1,\dots,p)$, $(K \times K)$ coefficient matrices

$e_t = B^{-1}\varepsilon_t$

$$\Sigma_\varepsilon = E(\varepsilon_t \varepsilon'_t) = E\begin{bmatrix} \varepsilon_{y_{1t}} \\ \varepsilon_{y_{2t}} \\ \dots \\ \varepsilon_{y_{Kt}} \end{bmatrix}\begin{bmatrix} \varepsilon_{y_{1t}} & \varepsilon_{y_{2t}} & \dots & \varepsilon_{y_{Kt}} \end{bmatrix}$$

$$= \begin{bmatrix} E(\varepsilon^2_{y_{1t}}) & \dots & E(\varepsilon_{y_{1t}}\varepsilon_{y_{Kt}}) \\ E(\varepsilon_{y_{2t}}\varepsilon_{y_{1t}}) & \dots & E(\varepsilon_{y_{2t}}\varepsilon_{y_{Kt}}) \\ \dots & & \\ E(\varepsilon_{y_{Kt}}\varepsilon_{y_{1t}}) & \dots & E(\varepsilon^2_{y_{Kt}}) \end{bmatrix} = \begin{bmatrix} \sigma^2_{y_1} & \dots & 0 \\ 0 & \dots & 0 \\ \dots & & \\ 0 & \dots & \sigma^2_{y_K} \end{bmatrix}$$

$$\Sigma_\varepsilon = E(\varepsilon_t \varepsilon'_t) \rightarrow \text{diagonal}$$

$$\Sigma = E(e_t e'_t) = E(B^{-1}\varepsilon_t (B^{-1}\varepsilon_t)')$$
$$= E(B^{-1}\varepsilon_t \varepsilon'_t (B^{-1})')$$

$$= B^{-1}E(\varepsilon_t\varepsilon'_t)(B^{-1})' \rightarrow \text{non-diagonal}$$

$$= B^{-1}\Sigma_\varepsilon(B^{-1})'$$

$$= B^{-1}I_2(B^{-1})', \quad \text{if } \Sigma_\varepsilon = I_2 \text{ identity matrix}$$

$$\Sigma = B^{-1}(B^{-1})'$$

or

$$\Sigma = Var(e_t) = Var(B^{-1}\varepsilon_t)$$

$$= B^{-1}Var(\varepsilon_t)(B^{-1})'$$

$$= B^{-1}\Sigma_\varepsilon(B^{-1})'$$

$$= B^{-1}I_2(B^{-1})', \quad \text{if } \Sigma_\varepsilon = I_2 \text{ identity matrix}$$

$$\Sigma = B^{-1}(B^{-1})'$$

2.3 Stationarity

Vector version of weak stationarity

Mean

$E(x_t) = \mu$

where $\mu \equiv (\mu_1, \mu_2, ..., \mu_K)'$ independent of t

Variance-covariance

$$E\left[(x_t - \mu)(x_t - \mu)'\right] = \Sigma_x$$

where Σ_x $(K \times K)$ independent of t

Conditions for stationarity

Consider univariate $AR(p)$ (for illustration)

$$y_t = \delta + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t$$

$$\Rightarrow \phi(L) y_t = \delta + u_t$$

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$$

factorise

$$= (1 - \lambda_1 z) \times (1 - \lambda_2 z) \times \ldots \times (1 - \lambda_p z)$$

roots

$$z_i = \frac{1}{\lambda_i}, \; i = 1, \ldots, p$$

stationarity and stability requires inverse of roots of pth order polynomial to lie inside unit circle

$$|\lambda_i| < 1 \text{ (or } |z_i| > 1), \; i = 1, \ldots, p$$

Infinite Moving Average (*IMA*) or *Wold representation*

Let $\phi(L)^{-1} = f(L) = \sum_{j=0}^{\infty} f_j L^j$ then

$$y_t = \phi(L)^{-1}(\delta + u_t)$$

$$\Rightarrow y_t = \phi(1)^{-1}\delta + \sum_{j=0}^{\infty} f_j u_{t-j}$$

$$\Rightarrow y_t = \frac{\delta}{1-\phi_1-\ldots-\phi_p} + \sum_{j=0}^{\infty} f_j u_{t-j}$$

Bivariate $VAR(1)$

$$x_t = A_0 + A_1 x_{t-1} + e_t$$

$$\Rightarrow x_t - A_1 x_{t-1} = A_0 + e_t$$

$$\Rightarrow x_t - A_1 L x_t = A_0 + e_t$$

$$\Rightarrow (I_2 - A_1 L) x_t = A_0 + e_t$$

$$(I_2 - A_1 L)^{-1} \text{ ?}$$

pre-multiply by $(I_2 - A_1 L)^{-1}$

$$I_2 - A_1 L = \begin{bmatrix} 1-a_{11}L & -a_{12}L \\ -a_{21}L & 1-a_{22}L \end{bmatrix}$$

Determinantal equation

$$\det(I_2 - A_1 L)$$
$$= (1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2$$
$$= 1 - (a_{11} + a_{22})L + (a_{11}a_{22} - a_{12}a_{21})L^2$$
$$= (1 - \lambda_1 z) \times (1 - \lambda_2 z)$$

roots

$$z_1 = \frac{1}{\lambda_1},\ z_2 = \frac{1}{\lambda_2}$$

stationarity and stability requires inverse of the roots of 2nd order polynomial to lie inside unit circle

$$|\lambda_1| < 1,\ |\lambda_2| < 1\ \text{(or } |z_1| > 1,\ |z_2| > 1)$$

If one of the two roots is one then both y_t, z_t are non-stationary

2.4 Identification

Consider a bivariate structural $VAR(1)$

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{yt}$$

$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt}$$

The structural system is not directly estimable by OLS since

$$Cov(z_t, \varepsilon_{yt}) \neq 0 \text{ and } Cov(y_t, \varepsilon_{zt}) \neq 0$$

implies biased and inconsistent estimates!

Consider the reduced-form $VAR(1)$

$$y_t = a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t}$$
$$z_t = a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t}$$

here OLS is applicable.

Recover all information present in structural model?

Structural model is underidentified. Why?

However, if we set $b_{21} = 0 \Rightarrow$

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{yt}$$

$$z_t = b_{20} + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt}$$

or

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} - \begin{bmatrix} 0 & b_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

let $B = \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix} \Rightarrow B^{-1} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}$

Premultiply by B^{-1}

$\Rightarrow$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

or

$$y_t = (b_{10} - b_{12} b_{20}) + (\gamma_{11} - b_{12}\gamma_{21}) y_{t-1} + (\gamma_{12} - b_{12}\gamma_{22}) z_{t-1} + (\varepsilon_{yt} - b_{12}\varepsilon_{zt})$$

$$z_t = b_{20} + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt}$$

or

$$y_t = a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t}$$

$$z_t = a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t}$$

where $a_{10}=b_{10}-b_{12}b_{20}$, $a_{11}=\gamma_{11}-b_{12}\gamma_{21}$, $a_{12}=\gamma_{12}-b_{12}\gamma_{22}$

$a_{20}=b_{20}$, $a_{21}=\gamma_{21}$, $a_{22}=\gamma_{22}$, $e_{1t}=\varepsilon_{yt}-b_{12}\varepsilon_{zt}$, $e_{2t}=\varepsilon_{zt}$

notice that y_t is affected by both ε_{yt} and ε_{zt} whereas z_t is affected only by ε_{zt} ⇒ causal ordering

Variance-covariance matrix of the errors

$$\Sigma = E(e_t e'_t) = E\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}\begin{pmatrix} e_{1t} & e_{2t} \end{pmatrix} = \begin{bmatrix} E(e_{1t}^2) & E(e_{1t}e_{2t}) \\ E(e_{2t}e_{1t}) & E(e_{2t}^2) \end{bmatrix}$$

$$= \begin{bmatrix} (\sigma_y^2 + b_{12}^2\sigma_z^2) & (-b_{12}\sigma_z^2) \\ (-b_{12}\sigma_z^2) & \sigma_z^2 \end{bmatrix}$$

Orthogonalize residuals using *Choleski decomposition*

2.5 Estimation

Consider the reduced-form $VAR(p)$

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \ldots + A_p x_{t-p} + e_t$$

Under conditional normality,

$$x_t | x_{t-1}, \ldots, x_{-p+1} \sim N(A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \ldots + A_p x_{t-p}, \Sigma)$$

More compactly, let

$$z_{t\,(Kp+1)\times 1} = \begin{pmatrix} 1 \\ x_{t-1} \\ . \\ . \\ . \\ x_{t-p} \end{pmatrix}, \; \Pi'_{(K\times(Kp+1))} = \left(A_0, A_1, \ldots, A_p\right)$$

The jth row of Π' have the parameters of the jth equation in the *VAR*.

Thus, write

$$x_t | x_{t-1}, \ldots, x_{-p+1} \sim N(\Pi' z_t, \Sigma)$$

Derive the log-likelihood function

$$LLF(\theta) = -\frac{TK}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_{t=1}^{T}(x_t - \Pi' z_t)'\Sigma^{-1}(x_t - \Pi' z_t)'$$

The first order conditions (FOCs)

$$\hat{\Pi}'^{MLE} = \left[\sum_{t=1}^{T} x_t z'_t\right]\left[\sum_{t=1}^{T} z_t z'_t\right]^{-1}$$

The jth row of $\hat{\Pi}'$ is

$$\hat{\pi}'^{MLE}_j = \left[\sum_{t=1}^{T} x_{jt} z'_t\right]\left[\sum_{t=1}^{T} z_t z'_t\right]^{-1},\ j = 1,\ldots,K$$

which amounts to equation-by-equation OLS. The MLE for error variance-covariance matrix

$$\hat{\Sigma}^{MLE} = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_t \hat{e}'_t$$

$$\hat{\sigma}^{2\,MLE}_i = \frac{1}{T}\sum_{t=1}^{T} \hat{e}^2_{it}$$ for variances

$$\hat{\sigma}^{2\,MLE}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_{it} \hat{e}_{jt}$$ for covariances

2.6 Model selection criteria

How do we choose lag order?

1st Way

Adding lags reduces the determinant of the variance-covariance matrix of the reduced-form errors $|\Sigma|$, but also leads to loss of degrees of freedom (*df*)

Model selection criteria trade off reduction of $|\Sigma|$ for a more parsimonious model

Akaike Information Criterion (AIC)

$$AIC = \log|\Sigma| + \frac{2}{T}N$$

Schwarz Bayesian Criterion (SBC)

$$SBC = \log|\Sigma| + \frac{\log(T)}{T}N$$

where $N \rightarrow$ total number of estimated parameters

$N=K(p+1)$, and T (fixed)

Minimization

SBC marginal cost of adding regressors greater than AIC

2nd Way

Conduct a series of Likelihood Ratio (LR) tests.

2.7 Impulse response analysis

VAR models concentrate on shocks. First the relevant shocks are identified, and the response of the system to shocks is described by analysis impulse responses (the propagation mechanism of the shocks).

Consider bivariate reduced-form *VAR*(1)

$$x_t = A_1 x_{t-1} + e_t$$

backward iteration implies

$$x_t = A_1 (A_1 x_{t-2} + e_{t-1}) + e_t$$

after *n* iterations

$$= \sum_{i=0}^{n} A_1^i e_{t-i} + A_1^{n+1} x_{t-n-1}$$

as $n \to \infty$

$$x_t = \sum_{i=0}^{\infty} A_1^i e_{t-i}\text{ , since } A_1^{n+1} x_{t-n-1} \to 0$$

Infinite Moving Average (*IMA*) or *Wold representation*

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix}$$

but

$$e_t = B^{-1} \varepsilon_t = \frac{1}{1 - b_{12} b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} y_t \\ z_t \end{bmatrix} = [1/(1-b_{12}b_{21})]\sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

if we let $\phi_i = [A_1^i/(1-b_{12}b_{21})] \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}$

Structural Infinite Moving Average (*IMA*) *representation*

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \phi_{11}(0) & \phi_{12}(0) \\ \phi_{21}(0) & \phi_{22}(0) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} + \begin{bmatrix} \phi_{11}(1) & \phi_{12}(1) \\ \phi_{21}(1) & \phi_{22}(1) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-1} \\ \varepsilon_{zt-1} \end{bmatrix} + \ldots$$

$$y_t = \phi_{11}(0)\varepsilon_{yt} + \phi_{12}(0)\varepsilon_{zt} + \phi_{11}(1)\varepsilon_{yt-1} + \phi_{12}(1)\varepsilon_{zt-1} + \ldots$$

$$z_t = \phi_{21}(0)\varepsilon_{yt} + \phi_{22}(0)\varepsilon_{zt} + \phi_{21}(1)\varepsilon_{yt-1} + \phi_{22}(1)\varepsilon_{zt-1} + \ldots$$

Impulses hitting the system

$$\begin{bmatrix} \phi_{11}(0) & \phi_{12}(0) \\ \phi_{21}(0) & \phi_{22}(0) \end{bmatrix} \rightarrow \text{matrix of impact multipliers}$$

$\phi_{12}(0) \rightarrow$ instantaneous effect of ε_{zt} on y_t

$\phi_{21}(0) \rightarrow$ instantaneous effect of ε_{yt} on z_t

$\phi_{11}(0) \rightarrow$ instantaneous effect of ε_{yt} on y_t

$\phi_{22}(0) \rightarrow$ instantaneous effect of ε_{zt} on z_t

$$\begin{bmatrix} \phi_{11}(1) & \phi_{12}(1) \\ \phi_{21}(1) & \phi_{22}(1) \end{bmatrix} \rightarrow$$

$\phi_{11}(1) \rightarrow$ 1-period effect of ε_{yt-1} on y_t

$\phi_{12}(1) \rightarrow$ 1-period effect of ε_{zt-1} on y_t

$\phi_{21}(1) \rightarrow$ 1-period effect of ε_{yt-1} on z_t

$\phi_{22}(1) \rightarrow$ 1-period effect of ε_{zt-1} on z_t

$\phi_{12}(n) \rightarrow$ n-period effect of ε_{zt-n} on y_t (or ε_{zt} on y_{t+n})

$\sum_{i=0}^{n} \phi_{12}(i) \rightarrow$ accumulated effect of ε_{zt} on $\{y_t\}$ after n periods

$\phi_{21}(n) \rightarrow$ n-period effect of ε_{yt-1} on z_t (or ε_{yt} on z_{t+n})

$\sum_{i=0}^{n}\phi_{21}(i) \rightarrow$ accumulated effect of ε_{yt} on $\{z_t\}$ after *n* periods

Therefore

$\sum_{i=0}^{\infty}\phi_{jk}(i) \rightarrow$ long-run multipliers

$\phi_{jk}(i)$ versus $i \rightarrow$ impulse response functions

$$\lim_{i\to\infty}\phi_{jk}(i)=0,\ \ j,k=1,2$$

No structural shock should have long-run impact. If the variables are stationary then shocks have transitory effects.

In his famous article Sims (1980) proposed the following identification strategy. To identify the shocks use *Choleski decomposition* in the structural model, $b_{21}=0 \Rightarrow$

$$e_t = B^{-1}\varepsilon_t = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

$\Rightarrow e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt},\ e_{2t} = \varepsilon_{zt}$

Structural Infinite Moving Average (IMA) representation

$$\Rightarrow \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

$\Rightarrow$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-1} \\ \varepsilon_{zt-1} \end{bmatrix} + \ldots$$

if we let $\phi_i = A_1^i \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \phi_{11}(0) & \phi_{12}(0) \\ 0 & \phi_{22}(0) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} + \begin{bmatrix} \phi_{11}(1) & \phi_{12}(1) \\ \phi_{21}(1) & \phi_{22}(1) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-1} \\ \varepsilon_{zt-1} \end{bmatrix} + \ldots$$

$$y_t = \phi_{11}(0)\varepsilon_{yt} + \phi_{12}(0)\varepsilon_{zt} + \phi_{11}(1)\varepsilon_{yt-1} + \phi_{12}(1)\varepsilon_{zt-1} + \ldots$$

$$z_t = \phi_{22}(0)\varepsilon_{zt} + \phi_{21}(1)\varepsilon_{yt-1} + \phi_{22}(1)\varepsilon_{zt-1} + \ldots$$

Asymmetry $\rightarrow$ z_t prior to y_t (causal ordering)

Example of calculation of impulse response functions (IFRs)

- Set $x_{t-1} = \ldots = x_{t-p} = 0$
- Set $\varepsilon_{jt} = 1$ and $\varepsilon_{kt} = 0$ for $k \neq j$
- Simulate the system for dates t, $t+1$, $t+2, \ldots,$ $t+n$

Assume VAR(1)

$$y_t = 0.7 y_{t-1} + 0.2 z_{t-1} + e_{1t}$$

$$z_t = 0.2 y_{t-1} + 0.7 z_{t-1} + e_{2t}$$

where $a_{10} = a_{20} = 0$ (for simplicity) and the reduced-form errors are given

$$e_{1t} = \varepsilon_{yt} + 0.8 \varepsilon_{zt},\ b_{12} = -0.8$$

$$e_{2t} = \varepsilon_{zt}$$

Asymmetry $\rightarrow$ z_t prior to y_t (causal ordering)

At period t

set $\varepsilon_{zt} = 1$, $\varepsilon_{yt} = 0$ and $y_{t-1} = z_{t-1} = 0$

$\Rightarrow$

$$e_{1t} = 0 + 0.8(1) = 0.8$$

$$e_{2t} = 1$$

$$y_t = 0.7(0) + 0.2(0) + 0.8 = 0.8$$

$$z_t = 0.2(0) + 0.7(0) + 1 = 1$$

At period $t+1$

set $\varepsilon_{zt+1} = 0$, $\varepsilon_{yt+1} = 0$

$\Rightarrow$

$$y_{t+1} = 0.7(0.8) + 0.2(1) + 0 = 0.76$$

$$(y_{t+1} = 0.7y_t + 0.2z_t)$$

$$z_{t+1} = 0.2(0.8) + 0.7(1) + 0 = 0.86$$

$$(z_{t+1} = 0.2y_t + 0.7z_t)$$

At period $t+2$

set $\varepsilon_{zt+2} = 0$, $\varepsilon_{yt+2} = 0$

$\Rightarrow$

$$y_{t+2} = 0.7(0.76) + 0.2(0.86) + 0 = 0.704$$

$$(y_{t+2} = 0.7y_{t+1} + 0.2z_{t+1})$$

$$z_{t+2} = 0.2(0.76) + 0.7(0.86) + 0 = 0.754$$

$$(z_{t+2} = 0.2y_{t+1} + 0.7z_{t+1})$$

At period $t+3$

set $\varepsilon_{zt+3} = 0$, $\varepsilon_{yt+3} = 0$

$\Rightarrow$

$$y_{t+3} = 0.7(0.704) + 0.2(0.754) + 0 = 0.6436$$

$$(y_{t+3} = 0.7y_{t+2} + 0.2z_{t+2})$$

$$z_{t+3} = 0.2(0.704) + 0.7(0.754) + 0 = 0.6686$$

$$(z_{t+3} = 0.2y_{t+2} + 0.7z_{t+2})$$

At period $t+4$

set $\varepsilon_{zt+4} = 0$, $\varepsilon_{yt+4} = 0$

$\Rightarrow$

$$y_{t+4} = 0.7(0.6436) + 0.2(0.6686) + 0 = 0.584$$

$$(y_{t+4} = 0.7y_{t+3} + 0.2z_{t+3})$$

$$z_{t+4} = 0.2(0.6436) + 0.7(0.6686) + 0 = 0.597$$

$$(z_{t+4} = 0.7y_{t+3} + 0.2z_{t+3})$$

Stationarity assures the impulse responses ultimately decay

$$\lim_{i\to\infty} y_{t+i} = 0$$

$$\lim_{i\to\infty} z_{t+i} = 0$$

Similarly, a shock on the other variable

At period t

set $\varepsilon_{yt} = 1$, $\varepsilon_{zt} = 0$ and $y_{t-1} = z_{t-1} = 0$

$\Rightarrow$

$$e_{1t} = 1 + 0.8(0) = 1$$

$$e_{2t} = 0$$

$$y_t = 0.7(0) + 0.2(0) + 1 = 1$$

$$z_t = 0.2(0) + 0.7(0) + 0 = 0$$

At period $t+1$

set $\varepsilon_{yt+1} = 0$, $\varepsilon_{zt+1} = 0$

$\Rightarrow$

$$y_{t+1} = 0.7(1) + 0.2(0) + 0 = 0.7$$

$$(y_{t+1} = 0.7y_t + 0.2z_t)$$

$$z_{t+1} = 0.2(1) + 0.7(0) + 0 = 0.2$$

$$(z_{t+1} = 0.7y_t + 0.2z_t)$$

At period $t+2$

set $\varepsilon_{yt+2} = 0$, $\varepsilon_{zt+2} = 0$

$\Rightarrow$

$$y_{t+2} = 0.7(0.7) + 0.2(0.2) + 0 = 0.53$$

$$(y_{t+2} = 0.7y_{t+1} + 0.2z_{t+1})$$

$$z_{t+2} = 0.2(0.7) + 0.7(0.2) + 0 = 0.28$$

$$(z_{t+2} = 0.7y_{t+1} + 0.2z_{t+1})$$

At period $t+3$

set $\varepsilon_{yt+3} = 0$, $\varepsilon_{zt+3} = 0$

$\Rightarrow$

$$y_{t+3} = 0.7(0.53) + 0.2(0.28) + 0 = 0.43$$

$$(y_{t+3} = 0.7y_{t+2} + 0.2z_{t+2})$$

$$z_{t+3} = 0.2(0.53) + 0.7(0.28) + 0 = 0.3$$

$$(z_{t+3} = 0.7y_{t+2} + 0.2z_{t+2})$$

At period $t+4$

set $\varepsilon_{yt+4} = 0$, $\varepsilon_{zt+4} = 0$

$\Rightarrow$

$$y_{t+4} = 0.7(0.43) + 0.2(0.30) + 0 = 0.36$$

$$(y_{t+4} = 0.7y_{t+3} + 0.2z_{t+3})$$

$$z_{t+4} = 0.2(0.43) + 0.7(0.30) + 0 = 0.3$$

$$(z_{t+4} = 0.7y_{t+3} + 0.2z_{t+3})$$

At period $t+5$

set $\varepsilon_{yt+5} = 0$, $\varepsilon_{zt+5} = 0$

$\Rightarrow$

$$y_{t+5} = 0.7(0.36) + 0.2(0.30) + 0 = 0.31$$

$$(y_{t+5} = 0.7y_{t+4} + 0.2z_{t+4})$$

$$z_{t+5} = 0.2(0.36) + 0.7(0.30) + 0 = 0.28$$

$$(z_{t+5} = 0.7y_{t+4} + 0.2z_{t+4})$$

At period $t+6$

set $\varepsilon_{yt+6} = 0$, $\varepsilon_{zt+6} = 0$

$\Rightarrow$

$$y_{t+6} = 0.7(0.31) + 0.2(0.28) + 0 = 0.25$$

$$(y_{t+6} = 0.7y_{t+5} + 0.2z_{t+5})$$

$$z_{t+6} = 0.2(0.31) + 0.7(0.28) + 0 = 0.26$$

$$(y_{t+6} = 0.7y_{t+5} + 0.2z_{t+5})$$

...

Stationarity assures the impulse responses ultimately decay

$$\lim_{i\to\infty} y_{t+i} = 0$$

$$\lim_{i\to\infty} z_{t+i} = 0$$

2.8 Impulse response analysis: Sensitivity analysis

Does the assumed causal ordering affect the structural inferences?

If Σ close to diagonal $\rightarrow$ B close to diagonal (identity)

$\Rightarrow$ the ordering does not matter

$\Rightarrow$ the importance of ordering depends on

$$\rho = \frac{Cov(e_{1t}, e_{2t})}{\sqrt{Var(e_{1t})Var(e_{2t})}} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

$H_0 : \Sigma \; diagonal$

$$\hat{\rho} = \frac{Cov(\hat{e}_{1t}, \hat{e}_{2t})}{\sqrt{Var(\hat{e}_{1t})Var(\hat{e}_{2t})}} = \frac{\hat{\sigma}_{12}}{\hat{\sigma}_1 \hat{\sigma}_2}$$

$$LM = T\hat{\rho}^2 \overset{a}{\sim} \chi_1^2$$

2.9 Examples from macro VARs

The VAR models of the monetary transmission mechanism are not estimated to give advice on the best monetary policy. Rather they are estimated to provide empirical evidence on the response of macroeconomic variables to monetary policy impulses.

It is interesting to see how the specification of the standard *VAR* model has developed over time. Initially models were estimated on a rather limited set of variables, i.e. prices, output (real activity) and money (monetary policy). The underlying structural model is specified as follows

$$p_t = b_{10} + \gamma_{11} p_{t-1} + \gamma_{12} y_{t-1} + \gamma_{13} m_{t-1} + \varepsilon_{pt}$$

$$y_t = b_{20} - b_{21} p_t + \gamma_{21} p_{t-1} + \gamma_{22} y_{t-1} + \gamma_{23} m_{t-1} + \varepsilon_{yt}$$

$$m_t = b_{30} - b_{31} p_t - b_{32} y_t + \gamma_{31} p_{t-1} + \gamma_{32} y_{t-1} + \gamma_{33} m_{t-1} + \varepsilon_{mt}$$

p_t contemporaneous independent of y_t, m_t

y_t contemporaneous independent of m_t

This is a just-identification scheme, where the identification of structural shocks depends on the ordering of variables. It corresponds to a recursive economic structure, with the most endogenous variable ordered last.

Causal ordering $\Rightarrow\ p_t \rightarrow y_t \rightarrow m_t$

Intuitively, inflation shock (supply shock) $\rightarrow$ output $\rightarrow$ monetary policy

or

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \Sigma_\varepsilon)$$

where

$$B = \begin{bmatrix} 1 & 0 & 0 \\ b_{21} & 1 & 0 \\ b_{31} & b_{32} & 1 \end{bmatrix}, \ x_t = \begin{bmatrix} p_t \\ y_t \\ m_t \end{bmatrix}, \ \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \end{bmatrix},$$

$$\Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix}, \ \varepsilon_t = \begin{bmatrix} \varepsilon_{p_t} \\ \varepsilon_{y_t} \\ \varepsilon_{m_t} \end{bmatrix}$$

Identification is Choleski-type with money ordered last.

This *VAR* model can be extended to include short-term interest rates just before money as a penultimate variable in the Choleski identification. The idea is to see the robustness of the above results after identifying the part of money, which is endogenous to the interest rate. More specifically, the underlying structural model is specified as follows

$$p_t = b_{10} + \gamma_{11} p_{t-1} + \gamma_{12} y_{t-1} + \gamma_{13} i_{t-1} + \gamma_{14} m_{t-1} + \varepsilon_{pt}$$

$$y_t = b_{20} - b_{21} p_t + \gamma_{21} p_{t-1} + \gamma_{22} y_{t-1} + \gamma_{23} i_{t-1} + \gamma_{24} m_{t-1} + \varepsilon_{yt}$$

$$i_t = b_{30} - b_{31} p_t - b_{32} y_t + \gamma_{31} p_{t-1} + \gamma_{32} y_{t-1} + \gamma_{33} i_{t-1} + \gamma_{34} m_{t-1} + \varepsilon_{it}$$

$$m_t = b_{40} - b_{41} p_t - b_{42} y_t - b_{43} i_t + \gamma_{41} p_{t-1} + \gamma_{42} y_{t-1} + \gamma_{43} i_{t-1} + \gamma_{44} m_{t-1} + \varepsilon_{mt}$$

p_t contemporaneous independent of y_t, i_t, m_t

y_t contemporaneous independent of i_t, m_t

i_t contemporaneous independent of m_t

Causal ordering $\Rightarrow$ $p_t \rightarrow y_t \rightarrow i_t \rightarrow m_t$

or

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \Sigma_\varepsilon)$$

where

$$B=\begin{bmatrix} 1 & 0 & 0 & 0 \\ b_{21} & 1 & 0 & 0 \\ b_{31} & b_{32} & 1 & 0 \\ b_{41} & b_{42} & b_{43} & 1 \end{bmatrix},\ x_t=\begin{bmatrix} p_t \\ y_t \\ i_t \\ m_t \end{bmatrix},\ \Gamma_0=\begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \\ b_{40} \end{bmatrix},$$

$$\Gamma_1=\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} \end{bmatrix},\ \varepsilon_t=\begin{bmatrix} \varepsilon_{p_t} \\ \varepsilon_{y_t} \\ \varepsilon_{i_t} \\ \varepsilon_{m_t} \end{bmatrix}$$

Some evidence from the literature

After a contractionary monetary policy shocks, plausible models of the monetary transmission mechanism should be consistent at least with the following evidence on price, output and interest rates: (i) price level initially responds very little, (ii) interest rates initially rise, and (iii) output initially falls , with a *j*-shaped response, with a zero long-run effect of the monetary impulse.

Having identified the ‘monetary rule’ by proposing an explicit solution to the problem of the endogeneity of money, the VAR method focuses on deviation from the rule. Deviations from the rule are obtained either by changing the systematic component of monetary policy or by considering exogenous shocks, which leave monetary policy unaltered. In the former case the deviation from the rule is obtained by changing some parameters in the *B* matrix describing the simultaneous relations among variables, while in the latter case the parameters of the matrix *B* are not changed. Consider for example the case of interest rate targeting. The first type of deviations is obtained by modifying the response of the Central Bank’s interest rate to macroeconomic conditions (fluctuations in output and prices), while the second type of deviations is obtained by considering an exogenous shock which does not change the response of the monetary policy-maker to macroeconomic conditions. VAR modeling has focused on simulating shocks, leaving the systematic component of monetary policy unchanged.

Focusing on the shocks is important since only when the Central Bank deviates from its rules it becomes possible to collect interesting information on the response of macroeconomic variables to monetary policy impulses (shocks)--the best opportunity to detect the response of macroeconomic variables to monetary policy impulses unexpected by the market.

Often there are difficulties with interpreting shocks to interest rates as monetary policy shocks. The response of prices to an innovation (error) in interest rates gives rise to the ‘price puzzle’—prices increase significantly after an interest rate hike. The ‘price puzzle’ may be due to mis-specification of the VAR model. Suppose monetary policy reacts to *expected* inflation, then we have an omitted variable from the VAR positively related to inflation and interest rates. Such omission makes the VAR mis-specified and (partly?) explains the positive relation prices and interest rates observed in the impulse response functions.

2.10 Forecasting

Consider

$$VAR(1):\ x_t = A_0 + A_1 x_{t-1} + e_t$$

The model 1-period ahead

$$x_{t+1} = \hat{A}_0 + \hat{A}_1 x_t + e_{t+1}$$

Produce 1-period ahead forecast

$x^f_{t+1} = \hat{A}_0 + \hat{A}_1 x_t$, since the forecast for e_{t+1} is (on average) zero

The model 2-periods ahead

$$x_{t+2} = \hat{A}_0 + \hat{A}_1 x_{t+1} + e_{t+2}$$

Produce 2-periods ahead forecast

$x^f_{t+2} = \hat{A}_0 + \hat{A}_1 x^f_{t+1}$, since the forecast for e_{t+2} is (on average) zero

The model 3-periods ahead

$$x_{t+3} = \hat{A}_0 + \hat{A}_1 x_{t+2} + e_{t+3}$$

Produce 3-periods ahead forecast

$$x^f_{t+3} = \hat{A}_0 + \hat{A}_1 x^f_{t+2}, \qquad \text{since the forecast for } e_{t+3} \text{ is zero}$$

Consider

$$VAR(3):\ x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + e_t$$

The model 1-period ahead

$$x_{t+1} = \hat{A}_0 + \hat{A}_1 x_t + \hat{A}_2 x_{t-1} + \hat{A}_3 x_{t-2} + e_{t+1}$$

Produce 1-period ahead forecast

$$x^f_{t+1} = \hat{A}_0 + \hat{A}_1 x_t + \hat{A}_2 x_{t-1} + \hat{A}_3 x_{t-2}, \quad \text{since the forecast for } e_{t+1}$$

is (on average) zero

The model 2-periods ahead

$$x_{t+2} = \hat{A}_0 + \hat{A}_1 x_{t+1} + \hat{A}_2 x_t + \hat{A}_3 x_{t-1} + e_{t+2}$$

Produce 2-periods ahead forecast

$$x^f_{t+2} = \hat{A}_0 + \hat{A}_1 x^f_{t+1} + \hat{A}_2 x_t + \hat{A}_3 x_{t-1}, \quad \text{since the forecast for } e_{t+2}$$

is (on average) zero

The model 3-periods ahead

$$x_{t+3} = \hat{A}_0 + \hat{A}_1 x_{t+2} + \hat{A}_2 x_{t+1} + \hat{A}_3 x_t + e_{t+3}$$

Produce 3-periods ahead forecast

$x^f_{t+3} = \hat{A}_0 + \hat{A}_1 x^f_{t+2} + \hat{A}_2 x^f_{t+1} + \hat{A}_3 x_t$, since the forecast for e_{t+3} is zero

Forecast uncertainty

In large samples, $x^f_{t+n} \sim N(x_{t+n}, Var(x^f_{t+n}))$. We can constrict confidence intervals

$$x^f_{t+n} \pm 1.96 * se(x^f_{t+n})$$

-- An example using Gretl --

The iterated forecast method versus the multiperiod forecast method

So far, we looked at the iterated forecast method. Another way to obtain forecast is by using the multiperiod forecast method.

Multiperiod forecasts

Consider

$$VAR(3):\ x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + e_t$$

The model 1-period ahead

$$x_{t+1} = \hat{A}_0 + \hat{A}_1 x_t + \hat{A}_2 x_{t-1} + \hat{A}_3 x_{t-2} + e_{t+1}$$

Produce 1-period ahead forecast

$x^f_{t+1} = \hat{A}_0 + \hat{A}_1 x_t + \hat{A}_2 x_{t-1} + \hat{A}_3 x_{t-2}$, since the forecast for e_{t+1} is (on average) zero

The model 2-periods ahead

$$x_{t+2} = \hat{A}_0 + \hat{A}_1 x_t + \hat{A}_2 x_{t-1} + \hat{A}_3 x_{t-2} + e_{t+2}$$

Produce 2-periods ahead forecast

$x^f_{t+2} = \hat{A}_0 + \hat{A}_1 x_t + \hat{A}_2 x_{t-1} + \hat{A}_3 x_{t-2}$, since the forecast for e_{t+2} is (on average) zero

The model 3-periods ahead

$$x_{t+3} = \hat{A}_0 + \hat{A}_1 x_t + \hat{A}_2 x_{t-1} + \hat{A}_3 x_{t-2} + e_{t+3}$$

Produce 3-periods ahead forecast

$x^f_{t+3} = \hat{A}_0 + \hat{A}_1 x_t + \hat{A}_2 x_{t-1} + \hat{A}_3 x_{t-2}$, since the forecast for e_{t+3} is zero

If the model is correctly specified, the iterated method is more precise. Iterating can lead to biased forecasts. Otherwise, the multiperiod forecast method is preferred.

This book treats econometric methods for analysis of applied econometrics with a particular focus on applications in macroeconomics. Topics include macroeconomic data, panel data models, unobserved heterogeneity, model comparison, endogeneity, dynamic econometric models, vector autoregressions, forecast evaluation, structural identification. The books provides undergraduate students with the necessary knowledge to be able to undertake econometric analysis in modern macroeconomic research.