Random coefficients: An Introduction

Thierry Magnac   Toulouse School of Economics

Panel Data Methods: Advanced EEE
Fall 2021

Motivation

Heterogenous treatment effects:

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

and we are interested in average treatment effects

$$E(\alpha_i) \text{ or } E(\alpha_i \mid x_i)$$

or average treatment on the treated

$$E(\alpha_i \mid D_{it} = 1) \text{ or } E(\alpha_i \mid D_{it} = 1, x_i)$$

Outline

- Exogenous random coefficients
- Correlated random coefficients
- Identification of the variance of random coefficients
- Irregular identification

References

- Hsiao and Pesaran, 2008, Random Coefficient Panel Data Models, working paper, and in *The Econometrics of Panel Data,* 2008, Springer.
- Chamberlain, 1992
- Arellano and Bonhomme, 2012
- Graham and Powell, 2012

Set up

We could start with

$$y_{it} = x_{it}\beta_{it} + \varepsilon_{it},$$

under strict exogeneity ($E(\varepsilon_{it} \mid x_i) = 0$) but there are too many parameters, β_{it}. We could restrict it into:

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

and in this case we can estimate time-series by time-series.

Yet, we might only be interested in $E(\beta_i)$. This leads to the Swamy-type random coefficient model in which (x_{it} might include a constant term)

$$\begin{aligned} \beta_i &= \beta + \alpha_i, \\ E(\alpha_i &\mid x_i) = 0, \\ E(\alpha_i\alpha_j') &= \Delta\mathbf{1}\{i = j\}, E(\alpha_i\varepsilon_{it}) = 0. \end{aligned}$$

Remark: Many variations on this: time variation in the βs given by a dynamic equation etc

Estimation

Rewrite:

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

in which even if ε_{it} is homoskedatic, the composite error term is not even if $E(\alpha_i \varepsilon_{it}) = 0$. However:

$$\begin{aligned} E(x_{it}\varepsilon_{it}) &= 0, E(x_{it}(x_{it}\alpha_i)) = E(x_{it}^2\alpha_i) = 0, \\ &\implies E(x_{it}v_{it}) = 0, \end{aligned}$$

and assumptions for OLS consistency of the estimate of β are satisfied. The only issue is heteroskedasticity (see below).
Remark: If there is some correlation between x_{it} and α_i, it means that higher orders of x_{it} need to be included in the model (see Mundlak or Chamberlain's approaches).

Efficient estimators

Improving efficiency might require strengthening the variance assumption, i.e. $V(\alpha_i \mid x_i) = \Delta$. Then:

$$V(v_i \mid x_i) = x_i \Delta x_i' + \sigma_\varepsilon^2 I_T$$

If Δ is known, a GLS estimate is more efficient than OLS or we can use Feasible GLS after estimating Δ.
Remark: Instead of using estimates of β and residuals to construct these covariances matrices, we can also estimate the model individual obervation by individual observation and get OLS estimate of $\hat{\beta}_i$. As they are consistent for $\beta + \alpha_i$ we can recover an estimate of Δ by considering the empirical variance of $V(\hat{\beta}_i)$.
Def: The mean group estimator $\hat{\beta}_{MG} = \frac{1}{N} \sum_i \hat{\beta}_i$ and can be shown to be asymptotically equivalent to Swamy's estimator.

Critique

Exogeneity of random coefficients dubious and $E(\alpha_i \mid x_i) \neq 0$. We might not be interested in formulating a model for $E(\alpha_i \mid x_i)$ becasue what we are interested in is $E(\alpha_i)$ per se (i.e. average treatment effects).

Chamberlain, 1992

The model

Single restriction:

$$E(y_i \mid x_i, \gamma_i) = d(x_i, \theta_0) + R(x_i, \theta_0)\gamma_i.$$

Parameter of interest: $\phi_0 = E(\gamma_i)$.
Remark: If $h_0(x_i) = E(\gamma_i \mid x_i)$ then $\phi_0 = E(h_0(x_i))$ and $E(y_i \mid x_i) = d(x_i, \theta_0) + R(x_i, \theta_0)h_0(x_i)$. The parameters are θ_0 and a general function $h_0(x_i)$ to be estimated jointly. Under independence, we shall have $h_0(x_i) = \phi_0$.
We will consider estimating only θ_0 and ϕ_0.
Consider $u = y - E(y \mid x)$ and $v = h_0(x) - \phi_0$. Then:

$$E(u \mid x) = 0, E(v) = 0.$$

Moment restrictions

Consider a general matrix function $B(x, \theta)$ of instruments whose properties are to be defined below and define:

$$\begin{aligned} \Psi(x, y; \theta, \phi) &= B(x, \theta)\,(y - d(x, \theta) - R(x, \theta)\,\phi) \\ &= B(x, \theta)\,(u + R(x, \theta)\,v). \end{aligned}$$

$E\Psi(x, y; \theta_0, \phi_0) = 0$ if $B(x, \theta_0)R(x, \theta_0)$ is non stochastic and does not depend on x. For any such function:

$$\begin{cases} E(B(x, \theta_0)u) = 0, \\ E(B(x, \theta_0)R(x, \theta_0)v) = 0. \end{cases}$$

Remark: Dimensions of θ and ϕ should be defined more rigorously.

Instruments

Let $\Phi^{-1}(x)$ a $T \times T$ matrix function so that $P_\Phi R = I$ for P_Φ defined as

$$P_\Phi = (R'\Phi^{-1}R)^{-1}R'\Phi^{-1}$$

and $M_\Phi R = 0$ for M_Φ defined as

$$M_\Phi = \Phi^{-1} - \Phi^{-1}RP_\Phi.$$

We can then choose :

$$B = \begin{pmatrix} A'(x,\theta)M_\Phi \\ P_\Phi \end{pmatrix}.$$

A NLS argument suggest to choose:

$$A(x,\theta) = \frac{\partial d(x,\theta)}{\partial \theta'} + \sum_{j=1}^{q} a_j(x)\frac{\partial R(x,\theta)}{\partial \theta'},$$

in which the "best" $a_j(x_i)$ are $h_0(x_i)$.

Method of moments

The empirical counterpart of $E\Psi(x, y; \theta_0, \phi_0) = 0$, maintaining the hypothesis of good specification, is:

$$\sum_{i=1}^{n} \Psi(x_i, y_i; \hat{\theta}, \hat{\phi}) = 0,$$

that is:

$$\begin{aligned} \sum_{i=1}^{n} A'(x_i, \hat{\theta}) M_\Phi(x_i, \hat{\theta})(y_i - d(x_i, \hat{\theta})) &= 0, \\ P_\phi(x_i, \hat{\theta})(y_i - d(x_i, \hat{\theta}) - R(x_i, \hat{\theta})\hat{\gamma}_i) &= 0. \end{aligned}$$

so that $\hat{\gamma}_i(\hat{\theta}) = P_\phi(x_i, \hat{\theta})(y_i - d(x_i, \hat{\theta}))$ and $\hat{\phi} = \frac{1}{N}\sum_{i=1}^{n} \hat{\gamma}_i(\hat{\theta})$.

Properties

As $\hat{\theta}$ is the solution of a moment condition that does not depend on the nuisance parameters γ_i or the nuisance function $h_0(x_i)$, $\hat{\theta}$ is consistent and asymptotically normal. Then as $\hat{\gamma}_i(\hat{\theta})$ is a LS estimator, it is unbiased and its empirical mean $\hat{\phi}$ is consistent at rate $\sqrt{n}$.

Remark: If $R(x_i, \theta)$ does not depend on θ, the concentrated least-squares problem

$$\min_{\theta,\gamma_1,..,\gamma_N} \sum_{i=1}^{n} (y_i - d - R\gamma_i)'\Phi^{-1}(y_i - d - R\gamma_i)$$

yields:

$$\min_{\theta,\gamma_1,..,\gamma_N} \sum_{i=1}^{n} (y_i - d)'M_\Phi(y_i - d)$$

in which M_Φ does not depend on θ and thus there is no incidental parameter issue.

Efficiency bounds

The estimator such that:

$$\begin{aligned} a_j(x) &= h_{0,j}(x), \\ \Phi(x) &= V(y_i - d_i - R\gamma_i), \end{aligned}$$

is semi-parametrically efficient.
Note that:

$$V(y_i \mid x_i) = V(w_i \mid x_i) + R(x_i, \theta_0) Var(\gamma_i \mid x_i) R'(x_i, \theta_0).$$

Arellano & Bonhomme, 2012

Set up

$$\underset{[T,1]}{y_i} = \underset{[T,K]}{Z_i}\ \delta + \underset{[T,q]}{X_i}\ \gamma_i + v_i$$

under mean independence: $E(v_i \mid W_i, \gamma_i) = 0$ $(W_i = (Z_i, X_i))$.
Remark: Strict exogeneity is necessary to identify moments of γ_i.
Rem 2: $q < T$ so that $Q_i = I - X_i(X_i' X_i)^{-} X_i$ has rank at least equal to $T - q$ and the equation above can be rewritten:

$$Q_i y_i = Q_i Z_i \delta + Q_i v_i.$$

The assumption that the rank of the matrix $\tilde{Z} = (Q_1 Z_1, ., Q_N Z_N)$ is equal to K is a generic assumption so that δ is identified.
Rem 3: We retain a population $\mathbb{S}$ of individuals i such that $\det(X_i' X_i) \neq 0$ i.e. X_i has full rank q. See Graham & Powell blow for the general case

Estimates and moments

Let:

$$H_i = (X_i' X_i)^{-1} X_i$$

The least square estimate is:

$$\hat{\gamma}_i = (X_i' X_i)^{-1} X_i' (y_i - Z_i \delta) = \gamma_i + H_i v_i$$

and the main equation is equivalent to:

$$\begin{cases} Q_i (y_i - Z_i \delta) = Q_i v_i \text{ (orthogonal to } X_i) \\ \hat{\gamma}_i - \gamma_i = H_i v_i \text{ (complement)} \end{cases}$$

Identification of parameters and first moments

Using the exogeneity assumption:

$$E(Q_i(y_i - Z_i\delta) \mid W_i) = 0$$

and δ is identified if $E(Z_i' Q_i Z_i)$ has rank K.
Furthermore,

$$E(\hat{\gamma}_i - \gamma_i \mid W_i) = 0,$$

so that $E(\gamma_i \mid W_i)$ is identified because $E(\hat{\gamma}_i \mid W_i)$ is identified if δ is identified.

Identification of second moments

Let $\Omega_i = V(v_i \mid W_i)$. We then have:

$$E((y_i - Z_i\delta)(y_i - Z_i\delta) \mid W_i) = X_i' E(\gamma_i \gamma_i' \mid W_i) X_i + \Omega_i.$$

Rem: $E(\gamma_i \gamma_i' \mid W_i)$ cannot be identified if no restriction on Ω_i.
General assumption: $vec(\Omega_i) = \omega(W_i; \theta)$ in which θ is a finite dimensional parameter whose dimension is less than $T(T+1)/2$.
Example: if Ω_i is diagonal, with variances $\sigma_i^2 \omega_t^2$ then $\omega(W_i; \theta)$ is a T^2 vector with 0 everywhere except at positions 1, $T+2$, ., T^2 that we can also write as, $\theta_i = (\sigma_i^2 \omega_1^2, .., \sigma_i^2 \omega_T^2)$:

$$\omega(W_i; \theta) = S\theta$$

in which S is a selection matrix of dimension $T^2 \times T$.

Identification condition

The covariance matrix, $\Omega_i = V(v_i \mid W_i)$, cannot be estimated directly since only residuals $\hat{v}_i = Q_i v_i$ of the regression of $y_i - z_i\delta$ on x_i can be constructed. We have:

$$V(\hat{v}_i \mid W_i) = Q_i V(v_i \mid W_i) Q_i'$$

and as a consequence:

$$\begin{aligned} vec(V(\hat{v}_i \quad \mid \quad W_i)) &= vec(Q_i V(v_i \mid W_i) Q_i') \\ &= (Q_i \otimes Q_i') vec(V(v_i \mid W_i)) \\ &= (Q_i \otimes Q_i)\omega(W_i;\theta) = M_i \omega(W_i;\theta) \end{aligned}$$

Condition: $rank(M_i \frac{\partial \omega(W_i;\theta)}{\partial \theta'}) = \dim(\theta)$.
Example: If $\omega(W_i;\theta) = S\theta$ then $rank(M_i S) = \dim(\theta)$.

Variance of random coefficients

As $\hat{\gamma}_i - \gamma_i = H_i v_i$ and $E(v_i \mid W_i, \gamma_i) = 0$, we have by the variance decomposition identity :

$$\begin{aligned} V(\hat{\gamma}_i \quad \mid \quad W_i) &= V(E(\hat{\gamma}_i \mid W_i, \gamma_i)) + E(V(\hat{\gamma}_i \mid W_i, \gamma_i)) \\ &= V(\gamma_i \mid W_i) + H_i V(v_i \mid W_i) H_i' \end{aligned}$$

the sum of the true variance of the random coefficients and the noise component due to estimation (the incidental parameter issue!) which depends on the identification of Ω_i above.
Consequence: Any correction of the biases in the variance of random coefficients needs an assumption on the variance of idiosyncratic errors.
Remark: True also for other characteristics of the distribution of random effects (see Jochmans and Weidner below).

GMM estimation of common parameters

Let A_i the $[T-q, T]$ orthogonal decomposition of $Q_i = I - X_i(X_i'X_i)^{-}X_i$. From Chamberlain (1992), we have that optimal GMM is using the following moments:

$$\begin{cases} E(Z_i'A_i'(A_iV_iA_i')^{-1}A_i(y_i - Z_i\delta)) = 0, \\ E((X_i'V_i^{-1}X_i)^{-1}X_i'V_i^{-1}(y_i - Z_i\delta - X_i\gamma)) = 0, \end{cases}$$

in which $V_i = Var(Y_i \mid W_i)$.

Rem: First step, replace V_i by any conformable and invertible matrix Ψ_i (*e.g.* I_T) and compute the method of moments estimator (which is also the OLS estimator here) for δ.

Rem 2: For the semi-parametric efficient estimate of δ, estimating $(A_iV_iA_i')$ is in order. As it is equal to $(A_iE(v_iv_i')A_i')$ and $A_iV_i = A_i(y_i - Z_i\delta)$, we can estimate it as

$$\frac{1}{N}\sum A_i\hat{v}_i\hat{v}_i'A_i,$$

in which $A_i\hat{v}_i$ is derived from an initial estimate, $\hat{\delta}$, and replace

Average individual parameters

We proceed likewise and get the weighted mean-group estimate:

$$\hat{\gamma} = \frac{1}{N} \sum_i (X_i' \Psi_i^{-1} X_i)^{-1} X_i' \Psi_i^{-1} (y_i - Z_i \hat{\delta}).$$

If $(X_i' \Psi_i^{-1} X_i)^{-1} X_i' \Psi_i^{-1} = (X_i' V_i^{-1} X_i)^{-1} X_i' V_i^{-1} (y_i - Z_i \delta)$ then $\hat{\gamma}$ is semi-parametrically efficient.

Remark: Individual estimates $\hat{\gamma}_i$ can be regressed on F_i and those regression estimates are consistent. As it depends on pre-estimated $\hat{\delta}$, however, their covariance matrix should be corrected.

Variance estimates

Assume as above:

$$vec(V(v_i \mid W_i)) = \omega(W_i;\theta) = S\theta$$

in which S is a selection matrix of dimension $T^2 \times m$ (*e.g.* time-heteroskedastic variances but no autocorrelation or generally MA variances). As

$$vec(V(\hat{v}_i \mid W_i)) = M_i S\theta,$$

$$\hat{\omega}(W_i;\theta) = \frac{1}{N}\sum_i S(M_i S)^{-} M_i \hat{v}_i \otimes \hat{v}_i.$$

Example: Under homoskedaticity:

$$\hat{\sigma}^2 = \frac{1}{N(T-q)}\sum_i (y_i - Z_i\hat{\delta})' Q_i (y_i - Z_i\hat{\delta})$$

Variance of individual coefficients

Use:

$$V(\hat{\gamma}_i \mid W_i) = V(\gamma_i \mid W_i) + H_i V(v_i \mid W_i) H_i',$$

to write (using $vec(ABC) = (C' \otimes A)vec(B)$):

$$\begin{aligned} vec(\hat{V}(\gamma_i \mid W_i)) &= \frac{1}{N}\sum_i (\hat{\gamma}_i - \hat{\gamma}) \otimes (\hat{\gamma}_i - \hat{\gamma}) \\ &- \frac{1}{N}\sum_i H_i \otimes H_i S(M_i S)^- M_i \hat{v}_i \otimes \hat{v}_i. \end{aligned}$$

Rem: It might not be positive definite (**Discussion**).
Rem 2: If homoskedastic then:

$$vec(\hat{V}(\gamma_i \mid W_i)) = \frac{1}{N}\sum_i (\hat{\gamma}_i - \hat{\gamma}) \otimes (\hat{\gamma}_i - \hat{\gamma}) - \frac{\hat{\sigma}^2}{N}\sum_i (X_i' X_i)^{-1}.$$