*Parte Teórica (Ler em conjunto com a parte prática)

equação geral → $Y_{it} = B_0 + B_1 X_{it} + \dots B_k X_{itk} + u_{it}$

→ Tipos de dados em painel

1) Painel Verdadeiro → usa-se a mesma unidade de análise (todos os "is" são iguais), ao longo do tempo. Então por exemplo uma pesquisa de satisfação de clientes do Burguer King, se ela entrevistar os mesmos indivíduos, mas em períodos diferentes, temos um painel verdadeiro.

2) Agrupamento de cortes transversais → Os "is" não serão os mesmos entre os períodos, ou seja, duas pesquisas diferentes

→ Pooled Cross Section = Os agrupamentos de cortes transversais são muito usados para investigar o efeito do tempo, ou seja, como a relação entre as variáveis mudam no tempo, por exemplo, avaliar políticas públicas. A grande vantagem do seu uso são os aumentos dos graus de liberdade (↓ multicolinealidade), devido ao uso de várias amostras

→ estimador de diferenças e indiferenças = Sua lógica é semelhante à testes farmacêuticos, onde um grupo recebe tratamento e o outro grupo (grupo de controle) não recebe. No nosso caso temos experimentos naturais (eventos exôgeneos), que afetam a variável dependente. O estimador busca captura esse efeito, antes e depois do evento.

→ Ver exemplos: FerTiL 1, CPS78_85, KieMC, INJURY

$\rightarrow$ dummy de Tratamento

equação geral $\rightarrow \beta_0 + \underbrace{\delta_0 dx_t}_{\text{dummy de período pós evento}} + \beta_1 dT + \underbrace{\delta_1 dx \cdot dT}_{\text{dummy de interação}}$ + outros fatores

x = período do corte transversal

dummy de interação = efeito real do evento (estimador de diferenças e indiferenças)

$\rightarrow$ Painel Verdadeiro

Vantagem do painel $\rightarrow$ no agrupamento de cortes transversais, se uma das variáveis controles, tiver alguma relação com as variáveis explicativas do meu modelo, haverá à presença de endogeneidade, porque a $COV(X_n, u) \neq 0$, então estimação via MQO é viesado. Em dados em painel, dependendo da situação é possível endereçar problemas ligados a viés de omissão de variável, erro de medida e entre outros.

Quando usamos dados em painel é muito comum, encontramos uma estimação via MQO viesada, por esse fato adotamos uma equação geral expandida, que considera a presença de um efeito fixo.

equação $\rightarrow y_{it} = \beta_0 + \delta_0 dx_t + \beta_1 X_{it1} + V_{it}$

sendo:

$V_{it} = a_i + U_{it}$

$\hookrightarrow$ erro endossincrático (varia em t)

erro composto

$\hookrightarrow$ efeito fixo ou heterogeneidade não observada (não varia em t)

O efeito fixo é algo que explica y, é específica de i, todavia ele não varia no horizonte de tempo observado. Quando passamos à usar a equação expandida, considerando o efeito fixo, para evitarmos viés, usaremos a estimação de primeiras diferenças, matematicamente falando nos pegamos os cortes transversais das observações i, durante o tempo e subtrairmos uma da outra

$$y_{i2} = (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2}$$
$$y_{i1} = \beta_0 + \beta_1 x_{i1} + a_i + u_{i1}$$

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1 (x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$
$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

Usando como exemplo a base CRIME2.dTa

$$crmrte = \beta_0 + \delta_0 d87_t + \beta_1 unem_{it} + a_i + u_{it}$$

É esperado que a taxa de crimes seja explicada por mais variáveis, além do desemprego. Então há omissão de variáveis no erro, todavia, será que uma dessas variáveis pode ser considerada um efeito fixo, tipo geografia, certamente que sim, já que, ela afeta o crime, é específica de cada município, todavia não é observável e tende à não mudar no tempo

-> hipóteses do estimador de primeiras diferenças

1) Linear nos betas, cada beta representa o efeito médio de X, sobre Y.

2) Aleatoriedade de cada um dos cortes transversais do meu modelo -> exogeneidade

3) Não pode haver relação perfeita linear, entre as variáveis explicativas (multicolinealidade exata ou perfeita).

4) $Var(\Delta U_{it} | X_i) = \sigma^2$ -> homocedásticos.

5) $cov(\Delta U_{it}, \Delta U_{it} | X_i) = 0$ -> sem autocorrelação serial.

6) Controlado pelo efeito fixo ou usando um modelo onde há eliminação do efeito fixo, $E(\Delta U_{it} | X_i) = 0$, que é equivalente no caso de cortes transversais, que o valor médio dos erros condicionado a x é zero.

7) Os ΔU_{it} são normalmente distribuidos

obs: O estimador de primeiras diferenças, diminui os graus de liberdade, sendo assim, há um aumento da multicolinealidade, por esse fato existe outra alternativa.

-> estimador de efeito fixo = Anteriormente apresentamos o chamado efeito fixo e usamos o estimador de primeiras diferenças, que é a versão do estimador de diferenças e indiferenças para dados em painel. Porém, quando há um efeito fixo não observável, há uma outra alternativa, o estimador de efeito fixo. Se de fato temos um efeito fixo, é esperado que o valor médio da constante em i é exatamente igual a constante, ou seja, a média de a_i no tempo é igual a_i, então eu posso fazer uma transformação de subtrair da média, chamada de transformação intra-grupo (WITHIN)

* painel balanceado -> é quando se tem a informação completa para todos os "is" em todos os "t" períodos

$$y_{it} = B_1 x_{it} + a_i + u_{it}$$
$$\bar{y}_{it} = B_1 \bar{x}_{it} + a_i + \bar{u}_{it}$$
$\downarrow$ Transformação
$$y_{it} - \bar{y}_{it} = B_1 (x_{it} - \bar{x}_i) + u_{it} - \bar{u}_{it}$$
$$\ddot{y}_{it} = B_1 \ddot{x}_{it} + \ddot{u}_{it}$$

-> aplico MQO -> Não será viesado

* em vez de tomar a diferença entre dois periodos, eu tomo a diferença em relação à sua média

-> LSDV (Least Squares Dummy Variable) = O estimador de efeito fixo é idêntico ao método de incluir uma dummy ou intercepto para cada i, a diferença que o LSDV mostra explicitamente o efeito fixo de cada i. A vantagem é que você pode conhecer o efeito fixo individual, já a desvantagem está num possível excesso de dummies, o que reduz os graus de liberdade

-> comparação de estimadores
- até T = 2 os resultados são idênticos.
- a escolha é feita pela facilidade do econometrista.
- Fe é mais fácil de implementar em paineis de dados não balanceados, do que primeiras diferenças
- tamanho da amostra é importante para definir a facilidade.

-> atrito = É quando há falta de informação em algum período da amostra, ou seja, um painel desbalanceado, o que impede a generalização dos resultados, porque se as variáveis retiradas, estiverem correlacionadas com o termo de erro endossincrático, teremos o problema de seleção amostral. (solução = estimador de heckman)

→ ESTIMADOR EFEITO ALEATÓRIO = Nesse caso haverá a presença do "a_i", todavia ele não estará correlacionado com nenhuma das variáveis explicativas "X". Sendo assim MQO não será viesado, ou seja, ele é consistente, mas não será eficiente, porque há autocorrelação serial no erro composto (o erro de um período será levado para o outro, pois "a_i" não é eliminado), sendo assim realizamos uma transformação de mínimos quadrados generalizados (GLS). Na realidade seria uma transformação "quase na média", porque nesse caso conhecemos a $\mathrm{cor}(V_{it}, V_{it-1})$ e essa correlação está associada a variância do termo de erro endossincrático e a variancia de "a_i", então poderiamos ponderar os "X_s" e o "Y", pela estrutura de mudança da variância dos erros, dependendo de "X", esse seria nosso estimador de efeito aleatório ou quase na média.

*MATEMATICAMENTE:

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1(x_{it1} - \lambda\bar{x}_{i1}) + \dots$$
$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_{ai}^2)]^{1/2}$$

λ → ponderação da estrutura de correlação serial devido aos efeitos fixos presentes no MQO. Sendo assim ocorreria, em vez de transforma dentro do grupo, cada "Y" e cada "X" de it em relação à média de i, ele faz uma transformação quase na média, ponderado pelo λ (ESTIMA O λ).

* Casos:

1) $\lambda = 1$

$\lambda = 1 - \underbrace{[\sigma_u^2/(\sigma_u^2 + T\sigma_{ai}^2)]}_{\text{Média}}{}^{1/2}$

→ variância de a_i é muito grande, então o efeito fixo é relevante

Nesse caso de $\lambda = 1$, eu estaria subtraindo da média, sendo assim eu estaria usando o estimador de efeito fixo. Quando σ_{ai}^2 tem uma variância muito alta, significa que existe uma diferença muito grande entre os a_is, então o efeito fixo é relevante.

2) $\lambda = 0$

$Y_{it} = \beta_0 + \beta_1 X_{i1}$

→ então:

$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_{ai}^2)]^{1/2}$

$\lambda = 1 - 1$

$\lambda = 0$

→ variância de a_i é muito pequena, então não há efeito fixo, que diferencia as unidades de análise

Então o estimador de GLS, no contexto de efeitos aleatórios é uma situação intermediaria

*Resumindo:

→ quanto mais próximo de $\lambda = 0$, MQO é preferível

→ quanto mais próximo de $\lambda = 1$, FE é preferível