

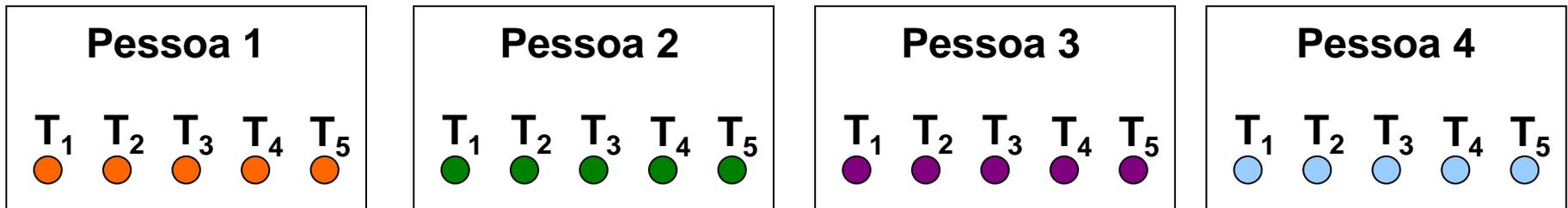
Análise de Séries de Tempo

Ernesto F. L. Amaral
Magna M. Inácio

23 de novembro de 2010
Tópicos Especiais em Teoria e Análise Política:
Problema de Desenho e Análise Empírica (DCP 859B4)

Dados de Painel

- Dados de painel geralmente se referem a um tipo particular de dados multi-níveis:
 - Uma medida no decorrer do tempo (T1, T2, T3...) ocorre para cada pessoa (ou instituições, ou países...):
 - Cada tempo é considerado uma onda (“wave”).



Dados de Painel

- Dados de painel envolvem combinação:
 - Informação sobre múltiplas causas:
 - Um componente interseccional ou transversal (“cross-sectional”).
 - Informação sobre casos no decorrer do tempo:
 - Um componente longitudinal ou de série temporal.
- Banco de dados de painel é descrito em termos de:
 - N: número de casos individuais.
 - T: número de ondas.
 - Se N é grande em relação a T, o banco é dominante transversalmente (“cross-sectionally dominant”).
 - Se T é grande em relação a N, o banco é dominante em séries de tempo (“time-series dominant”):
 - “Time-series Cross-section” (“TSCS data”): pequeno número de unidades (usualmente 10-30) e moderado T.

TSCS \neq Painel (Beck 2009)

- Estudos anteriores distinguiram entre “temporally dominated data sets” e “cross-sectionally dominated data sets” usando $T \gg N$ e $N \gg T$, respectivamente.
- Para Beck, o importante é se T é grande o suficiente, de forma que ao calcular médias no decorrer do tempo há geração de resultados estáveis.
- Geralmente, estudos de painel possuem T de um dígito, sendo 3 ondas um valor comum.
- Bancos de dados TSCS em política comparada têm em geral T 's iguais a 20 ou mais.

Terminologia de painel

- Painel significa duas coisas:
 - Painel é um tempo amplo para todos dados com séries temporais e componentes interseccionais.
 - Painel se refere especificamente a bancos com N grande e T pequeno (dominante transversalmente):
 - Ex.: uma pesquisa com 1000 pessoas em 3 pontos no tempo.
- Balanceado X Não Balanceado:
 - Dados de painel são chamados de balanceados se informação de cada pessoa é disponível para todos T 's.
 - Se há dados “missing” para alguns casos em certos pontos no tempo, os dados são não balanceados:
 - Isto é comum para muitos bancos de países ou instituições.

Exemplo de Dados de Painel

ID	Onda	Série	Teste matem.	Renda anual (\$)	Tamanho classe	Sexo
1	1	5	67	80.000	18	0
1	2	7	78	82.000	22	0
1	3	9	85	88.000	26	0
2	1	5	34	27.000	34	1
2	2	7	41	23.000	33	1

Benefícios de Dados de Painel

- Agregação (“pooling”) de casos ou tempo promove informação mais rica:
 - Quanto mais observações, melhor.
- Dados de painel são longitudinais:
 - Casos individuais são seguidos no decorrer do tempo.
 - Permite o estudo de processos dinâmicos.
 - Promove oportunidades para melhor entender relações causais.
- Modelos de painel permitem o controle de heterogeneidade individual.

Benefícios de Dados de Painel

- Dados de painel permitem investigação de problemas que são obscuros em dados transversais:
 - Exemplo: participação feminina na força de trabalho.
 - Suponha que um banco transversal mostre que 50% das mulheres estão na força de trabalho.
 - O que isso quer dizer?
 - Há 50% de mulheres procurando emprego e 50% ficando em casa?
 - Há 100% das mulheres procurando emprego, mas muitas estão desempregadas em um tempo específico?
 - Ou algo no meio disto?
 - Sem informações longitudinais, não é possível desenvolver uma análise clara.

Problemas com Dados de Painei

- Viola pressuposto de independência de MQO:
 - Aglomeração (“clustering”) por casos.
 - Aglomeração por tempo.
 - Outras fontes? Ex.: correlação espacial.
- Para N pequeno e T grande (TSCS): “Poolability”
 - É apropriado combinar casos muito diferentes?
- Correlação serial:
 - Casos adjacentes temporalmente podem ter erro correlacionado.
- Não-estacionário para dados com T grande.
- Outros problemas: heterogeneidade.

Estratégias de Dados de Painel

- Estratégia tradicional:
 - Efeitos fixos X Efeitos aleatórios.
 - Usar teste “Hausman” para definir escolha.
- Estratégias mais recentes:
 - Distinção entre ferramentas para “painel” e “TSCS”:
 - Diferentes problemas, diferentes soluções.
 - Mais atenção para modelos dinâmicos:
 - Modelos que incluem valores defasados (de Y): estacionariedade.
- De modo geral, não há muito consenso sobre quais são as melhores práticas.

Tradição Econométrica

- Principal foco de estudos econométricos:
 - Painéis com N grande e T pequeno.
 - Preocupação com problema de variáveis omitidas:
 - Heterogeneidade não observada.
 - Heterogeneidade individual.
 - Efeitos não observados.
- Exemplo de salários individuais:
 - Geralmente modelado como função de educação e experiência.
 - Porém, indivíduos diferem em formas que são difíceis de controlar.
 - Estratégia é usar modelo que retira variabilidade individual, tal como efeitos fixos.

Tradição Econométrica

- Modelo de painel estático básico com efeitos não observados:

$$y_{it} = \alpha + \beta X_{it} + \mu_i + v_{it}$$

- Subscrito i se refere a casos e t a períodos no tempo.
 - βX = covariáveis.
 - μ_i = erro de unidade específica não observado.
 - v_{it} = erro idiossincrático (individual).
- Isso é basicamente o mesmo que os modelos de efeitos aleatórios e fixos:
 - Pode-se usar a , u ou $zeta$, ao invés de “ μ ”.
 - Pode-se usar e para erro, ao invés de “ v ”.
 - Problema: tratar μ como fixo ou como aleatório?

Efeitos Não Observados

- Modelo de painel estático básico com efeitos não observados:

$$Renda_{it} = \alpha + \beta_1 Educ_{it} + \mu_i + v_{it}$$

- O problema é o μ_i que são características individuais não observadas que não variam no tempo, e que podem causar que salários sejam altos ou baixos no decorrer do tempo:
 - Ex: Pressão dos pais para aumentar renda, gens, QI...
 - Estratégias usuais:
 - Eliminar com **primeiras diferenças**.
 - Construir modelo com **efeitos fixos** ou com **efeitos aleatórios**, se há pressuposto de que μ_i não é correlacionado com os X 's.
 - Utilizar **valores defasados** da variável dependente, os quais podem refletir o impacto de μ_i .

Efeitos Não Observados: Primeiras Diferenças

- Primeiras diferenças (duas ondas de dados):

$$Renda_{i2} - Renda_{i1} = \alpha + \beta(Educ_{i2} - Educ_{i1}) + \mu_i - \mu_i + v_{i2} - v_{i1}$$

- Expressão acima pode ser expressa como:

$$\Delta Renda_{it} = \alpha + \beta \Delta Educ_{it} + \Delta \mu_i + \Delta v_{it}$$

- Resultado é que μ_i é eliminado.
- Ao utilizar diferenças, é possível estimar coeficientes, removendo efeitos de unidade específica não observados e que não variam no tempo (μ_i).

Efeitos Não Observados: Efeitos Fixos

- Comece com o mesmo modelo básico:

$$y_{it} = \alpha + \beta X_{it} + \mu_i + v_{it}$$

- Realize uma transformação interna (“within”):
 - Centralize tudo em torno da média da unidade específica.
 - Isto retira o efeito de μ .

$$y_{it} - \bar{y}_i = \beta(X_{it} - \bar{X}_i) + (\mu_i - \bar{\mu}_i) + (v_{it} - \bar{v}_i)$$

- Mesmo que usar variáveis “dummies” em cada caso:
 - Estimar μ como um efeito fixo.
 - No Stata: xtreg renda educ, fe

Efeitos Fixos (EF) X Primeiras Diferenças (PD)

- EF e PD são abordagens muito similares:
 - Para bancos com duas ondas, resultados são idênticos.
 - Para bancos com três ondas, resultados podem diferir.
- Qual é melhor para N grande e T pequeno?
 - EF é mais eficiente se não há correlação serial.
 - PD é mais eficiente se há muita correlação serial.
- Qual é melhor para N pequeno e T grande (TSCS)?
 - Maior probabilidade de haver problemas com não-estacionariedade (mudança ao longo do tempo).
 - EF e PD podem gerar resultados enviesados.

Efeitos Fixos (EF) X

Primeiras Diferenças (PD)

- EF e PD podem ter estimativas enviesadas se variáveis não forem estritamente exógenas:
 - Variáveis independentes devem ser não correlacionadas com erros passados, presentes e futuros.
 - Mesmo a inclusão de variável dependente defasada pode não resolver problema.
 - Viés em modelos EF é menor com T grande.
 - Viés em modelos PD não é menor com T grande.
- É difícil escolher entre EF e PD quando eles estimam resultados substancialmente diferentes:
 - É interessante mostrar os dois resultados e tentar determinar porque eles diferem.

Efeitos Fixos (EF) X Primeiras Diferenças (PD)

- EF e PD não conseguem estimar efeitos de variáveis que não mudam no tempo:
 - Podem ter problemas se variáveis mudam raramente.
- EF e PD não são tão eficientes como modelos que incluem variabilidade “entre” (*between*) casos.
- EF e PD são muito sensíveis a erros de medida:
 - Variabilidade “intra” (*within*) casos é geralmente pequena e pode ser contaminada por erro.

Efeitos Aleatórios

- Melhora na eficiência ao custo de adição de pressupostos:
 - Principal pressuposto é que efeito de unidade específica não observado (μ) não é correlacionado com X 's.
 - Se requisito não é atendido, resultados são enviesados.
 - Variáveis X 's omitidas geralmente induzem correlação entre outras variáveis X 's e o efeito não observado.
 - Se principal objetivo da análise de painel é evitar efeitos não observados, modelo de efeitos fixos é opção segura.
 - No Stata:
 - `xtreg renda educ, re` (estimador linear: GLS)
 - `xtreg renda educ, mle` (estimador de máxima verossimilhança: ML)

Efeitos Aleatórios

- Modelo de efeitos aleatórios (EA) é híbrido, ou seja, é um modelo intermediário ao MQO e EF:
 - Se T é grande, modelo EA é mais similar ao EF.
 - Se efeito não observado tem variância pequena, EA é mais similar ao MQO.
 - Se efeito não observado é grande (casos diferem muito), resultados serão mais similares ao EF.
- Estimadores de efeitos aleatórios lineares (GLS) são utilizados para N grande:
 - Suas propriedades para N pequeno e T grande não foram muito estudados.
 - Beck e Katz não recomendam estes estimadores, e indicam estimador ML para efeitos aleatórios.

Quando Usar Efeitos Aleatórios?

- Se estiver certo que efeito não observado unitário (μ) não está correlacionado com os X's:
 - Seja por razões teóricas ou porque você possui controles de boa qualidade (“dummies” para áreas sub-nacionais).
- Seu principal foco é nas variáveis que são constantes ao longo do tempo:
 - EF não é uma opção.
- Quando seu foco é na variabilidade entre casos:
 - E/ou há muito pouca variabilidade intra casos.
- Quanto o teste “Hausman” indicar que há resultados similares.

Teste de Especificação de Hausman

- Ferramenta para avaliar ajuste de modelos de efeitos fixos e efeitos aleatórios.
- Modelos de efeitos fixos e aleatórios são consistentes se são especificados corretamente.
- Porém, algumas violações causam inconsistência nos modelos de efeitos aleatórios:
 - Se variáveis X 's estão correlacionadas com erro aleatório.
- Se modelos não estimam mesmos resultados, efeitos aleatórios são enviesados:
 - Se resultados forem similares, use modelo mais eficiente (efeitos aleatórios).
 - Se resultados divergirem, chances são que modelos de efeitos aleatórios sejam enviesados. Use efeitos fixos.

Teste de Especificação de Hausman

- Estime ambos modelos de efeitos fixos e aleatórios.
- Salve as estimativas de cada um.
- Utilize o teste de Hausman.
- Exemplo:
 - `xtreg var1 var2 var3, i(groupid) fe`
 - `estimates store fixed`
 - `xtreg var1 var2 var3, i(groupid) re`
 - `estimates store random`
 - `hausman fixed random`

Teste de Especificação de Hausman

- Exemplo sobre atitudes ambientais: EF x EA

```
. hausman fixed random
```

Coeficientes de comparação direta...

	---- Coefficients ----		(b-B)	sqrt(diag(V_b-V_B))
	(b)	(B)	Difference	S.E.
	fixed	random		
age	-.0038917	-.0038709	-.0000207	.0000297
male	.0979514	.0978732	.0000783	.0004277
dmar	.0024493	.0030441	-.0005948	.0007222
demp	-.0733992	-.0737466	.0003475	.0007303
educ	.0856092	.0857407	-.0001314	.0002993
incomere1	.0088841	.0090308	-.0001467	.0002885
ses	.1318295	.131528	.0003015	.0004153

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\begin{aligned} \text{chi2}(7) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 2.70 \\ \text{Prob}>\text{chi2} &= 0.9116 \end{aligned}$$

Valores de p não significativos indicam que modelos levam a resultados similares.

Problemas com Teste de Hausman

- Falha em rejeitar H_0 significa:
 - EF e EA são similares (bom resultado).
 - EF e EA são similares, mas estimativas de EF são imprecisas, por causa de dados de baixa qualidade.
- Esteja atento para “significância estatística” e “significância prática”:
 - Com amostra grande, teste de Hausman pode falhar mesmo que estimativas de EF e EA sejam similares.
 - Se diferenças entre modelos são pequenas, pode-se argumentar que estimativas de EA são apropriadas.

Abordagem de Modelos de Equações Estruturais (MEE)

- Modelos EF e EA podem ser estimados com MEE:
 - Além de estimar modelos individualmente no Stata, é possível explorar módulo GLLAMM.
- Benefícios:
 - Como EF e EA são aninhados, é possível comparar estatísticas de ajuste (BIC...).
 - Permite flexibilidade, tais como:
 - Variável dependente defasada.
 - Relaxa pressuposto de que efeitos de covariáveis ou variâncias são constantes entre ondas dos dados.
 - Flexibilidade na correlação entre efeitos não observados e X's.

Efeito Aleatório com MEE

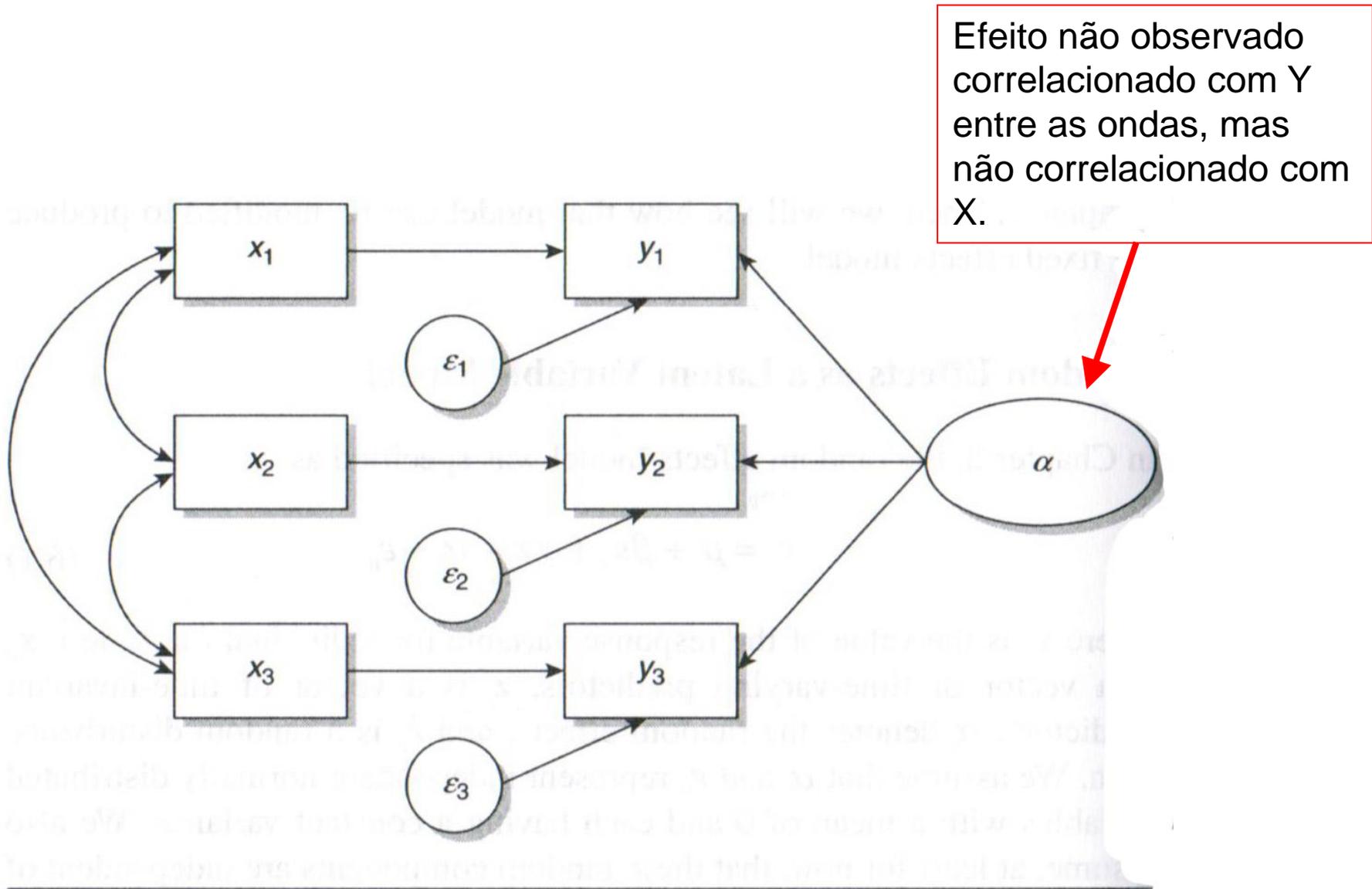


Figure 6.1 Path Diagram of a Random Effects Model for Three Periods

Efeito Fixo com MEE

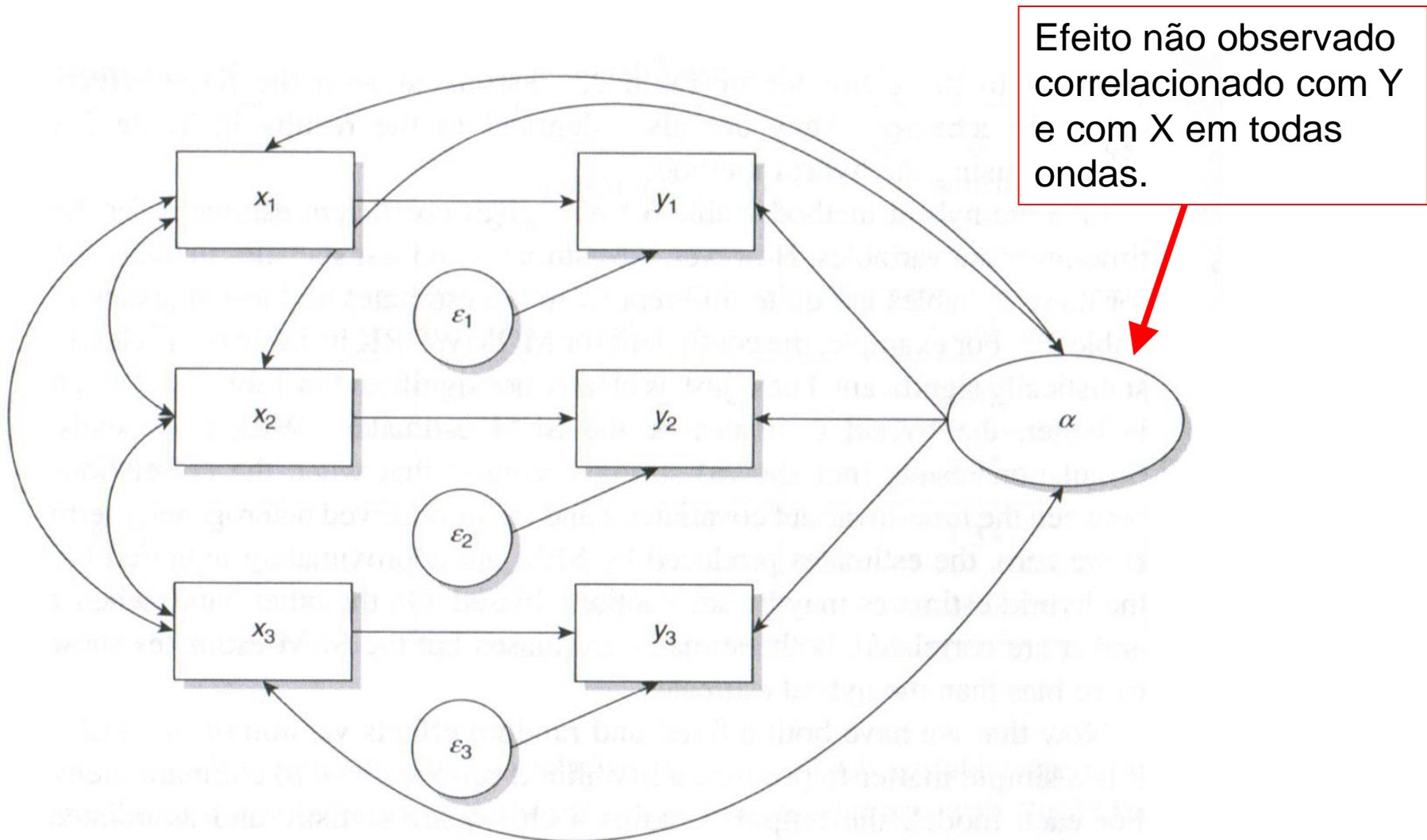


Figure 6.2 Path Diagram for a Fixed Effects Model for Three Periods

Correlação Serial

- O que fazer quando há erro correlacionado em ondas próximas nos dados?
- Por enquanto, foi discutido erro correlacionado com efeito não observado (dentro de cada caso).
- Estratégias:
 - Use um modelo que considera correlação serial.
 - No Stata, “xtregar” e “xtgee” para EF e EA.
 - Desenvolva um modelo dinâmico:
 - Permite modelar os padrões de correlação sobre Y.
 - É possível incluir a variável dependente ou variáveis independentes defasadas.
 - São necessários outros modelos, pois EF e EA se tornam enviesados.

Modelos de Painel Dinâmicos

- O que fazer quando temos um processo dinâmico?
- Exemplo de democracia em países (Baltagi 2008).
- Poderíamos usar um modelo como este:

$$y_{it} = y_{it-1} + \beta X_{it} + \mu_i + v_{it}$$

- Y de um período anterior é incluído como uma variável independente.
- Problema é que estimadores de EF e EA são enviesados:
 - Variável Y com tempo defasado é correlacionada com erro.
 - EF é enviesado para T pequeno. Resultado é melhor para T maior (pelo menos 30).
 - EA é enviesado.

Modelos de Painel Dinâmicos

- Se há problema entre erro e variável defasada Y , é possível tentar calcular uma nova versão desta variável defasada.
- É possível utilizar variáveis defasadas que voltem mais no tempo.
- Variáveis podem ser utilizadas como instrumentos, uma espécie de *proxy* para a variável defasada Y .
- Alguns comandos podem ser explorados no Stata (“xtabond”, “xtdpdsys” e “xtdpd”).

Modelos de Painel Dinâmicos

- É importante pensar sobre processos dinâmicos:
 - Quanto tempo leva para que sua variável de interesse se manifeste?
 - Quais variáveis defasadas faz sentido incluir?
 - Com bancos de dados grandes, podemos simplesmente incluir muitas variáveis defasadas.
 - Com bancos de dados pequenos, é importante pensar sobre estes processos antes de montar o modelo.

Modelos de Painel com Variáveis Instrumentais

- Estimador de painel com variável instrumental tradicional:

$$y_{it} = \beta X_{it} + \gamma Z_{it} + \mu_i + v_{it}$$

- βX : covariáveis exógenas.
- γZ : covariáveis endógenas (podem se relacionar com v_{it}).
- μ_i : erro de unidade específica não observada.
- v_{it} : erro idiossincrático (individual).
- Trate μ_i como aleatório, fixo ou use diferença para excluí-lo.
- Use valores originais ou defasados de X e valores defasados de Z como instrumentos (quando apropriados), em estimações de dois estágios de y_{it} .
- Este procedimento funciona se Z defasado é exógeno.

“Time Series Cross Section Data”³⁴ (TSCS)

- N pequeno e T grande.
- Exemplo são variáveis econômicas para países industrializados.
- Geralmente entre 10 e 30 países (N).
- Geralmente entre 30 e 40 anos de dados (T).
- Beck (2001) sugere que não há um valor mínimo para T, mas ficaria desconfiado se $T < 10$.
- N grande não é requerido, mas não é prejudicial.

Dados TSCS: “OLS PCSE”

- Procedimentos anteriores:
 - “Feasible Generalized Least Squares (FGLS)” = “Generalized Least Squares (GLS)” com matriz de covariância:
 - Problema de subestimação do erro padrão.
- Nova proposta sugere uso de regressão MQO:
 - “Ordinary least squares (OLS) with panel-corrected standard errors (PCSE)” (Beck e Katz 1995...).
 - Erros padrão são corrigidos com painel para levar em consideração heteroscedasticidade do painel.
 - Com efeitos fixos para corrigir heterogeneidade unitária.
 - Com variável dependente defasada no modelo para corrigir a correlação serial.

Observações Finais

- Estratégias com dados de painel são ensinadas como necessidade de “fixar” o modelo:
 - Como “fixamos” efeitos não observados?
 - Como “fixamos” correlação dinâmica e serial?
 - Ao “fixar” tais componentes, há mudança na modelagem.
 - Modelo de efeito fixo (intra) é um olhar diferente sobre os dados, comparado ao modelo MQO.
 - Objetivo deve ser de aprender a “fixar”, mas devemos nos concentrar na interpretação dos modelos.

Observações Finais

- Há muito debate na literatura:
 - Há diferentes argumentos sobre qual “fix” é o melhor.
 - Não tenha medo de críticas, mas espere que diferentes pessoas terão diferentes pontos de vista.
- Questão mais importante é testar vários modelos:
 - Se resultados são robustos, há indício de bom ajuste.
 - Se não, diferenças entre modelos ajudarão a entender o que está acontecendo com dados.
 - É normal chegar em modelos em que nossos resultados desapareçam, ao seguir sugestões de revisores. :)
 - **A teoria deve guiar a escolha do modelo!!!**