Modelos Dinâmicos com Dados em Painel: revisão de literatura

Outubro, 2000

Luís David Marques
lbmarq@fep.up.pt

CEMPRE*, FACULDADE DE ECONOMIA DO PORTO
RUA DR. ROBERTO FRIAS, 4200 PORTO, PORTUGAL

ABSTRACT

Since Hsiao [1986]'s work, many surveys have appeared, providing the researcher with the necessary tools for fully enjoying the advantages of panel data in terms of both microeconometric and macroeconometric studies. This survey has this very aim and, in a sense, offers no novel contribution, beyond the fact of bringing together some new developments in estimation and specification testing in a selective manner. The estimation of dynamic models either with fixed effects or random effects in the intercept using LSDV, GLS,MLE, IV and GMM is presented, as well as binary choice models. A short review on specification tests and unit roots testing is also presented.

RESUMO

Desde o trabalho fundamental de Hsiao [1986], várias revisões de literatura têm aparecido para dotar o investigador das ferramentas necessárias a um completo aproveitamento das vantagens dos dados em painel, quer na microeconometria, quer na macroeconometria. A presente revisão de literatura tem este mesmo objectivo e, nesse sentido, não traz qualquer contributo novo, para além do facto de juntar alguns dos últimos desenvolvimentos na estimação e nos testes de especificação, de uma forma selectiva, transcrevendo-os para português. A estimação de modelos dinâmicos com efeitos fixos ou aleatórios, no termo independente, por intermédio dos estimadores LSDV, GLS, MLE, IV e GMM é apresentada, tal como a questão da estimação de modelos de escolha binária. Apresenta-se, também, um resumo de alguns testes de especificação e de testes de raízes unitárias.

Keywords: panel data, dynamic models estimation, LSDV, GLS, MLE, IV, GMM, unit root tests
Palavras-Chave: dados em painel, estimação de modelos dinâmicos, LSDV, GLS, MLE, IV, GMM e testes de raiz unitária.

* Centro de Estudos Macroeconómicos e Previsão

ÍNDICE

1. INTRODUÇÃO

O objectivo do presente texto consiste em reunir, de forma integrada, um conjunto de referências e de resultados úteis a quem está a desenvolver trabalho empírico na área dos modelos dinâmicos com dados em painel. Apesar de existirem diversos exemplos deste tipo de revisão de literatura, principalmente sob a forma de *working papers* em língua inglesa, não se encontrou exemplo algum em português. Assim, tal como foi feito por Urga [1992], pretende-se fornecer pistas de investigação, referências úteis e alguns resultados prontos a usar para quem não pode despender muito tempo, nem esforço, neste tipo de recolha. Para quem tem a estimação com dados em painel como objecto central da respectiva pesquisa, esta *survey* não substitui de forma alguma a leitura dos originais citados.

Uma das vantagens da estimação com dados em painel é a relevação da heterogeneidade individual. Assim, os dados em painel sugerem a existência de características diferenciadoras dos indivíduos, entendidos como "unidade estatística de base". Essas características podem ou não ser constantes ao longo do tempo, de tal forma que estudos temporais ou seccionais que não tenham em conta tal heterogeneidade produzirão, quase sempre, resultados fortemente enviesados.

Por outro lado, os dados em painel providenciam uma maior quantidade de informação, maior variabilidade dos dados, menor colinearidade entre as variáveis, maior número de graus de liberdade e maior eficiência na estimação. A inclusão da dimensão seccional, num estudo temporal agregado, confere uma maior variabilidade aos dados, na medida em que a utilização de dados agregados resulta em séries mais suaves do que as séries individuais que lhes servem de base. Esse aumento na variabilidade dos dados contribui para a redução da eventual colinearidade existente entre variáveis, particularmente em modelos com desfasamentos distribuídos.

Adicionalmente, os estudos com amostras longitudinais facilitam uma análise mais eficiente das dinâmicas de ajustamento: os estudos seccionais, ao não contemplarem a possibilidade de a realidade de suporte ser dinâmica, transmitem uma falsa ideia de estabilidade. Assim, a utilização de dados em painel permite conjugar a diversidade de comportamentos individuais, com a existência de dinâmicas de ajustamento, ainda que potencialmente distintas. Ou seja, permite tipificar as respostas de diferentes indivíduos a determinados acontecimentos, em diferentes momentos.

Por outro lado, a maior quantidade de informação disponível aumenta a eficiência da estimação. Ou seja, os dados em painel permitem identificar e medir efeitos que não serão pura e simplesmente detectáveis em estudos exclusivamente seccionais ou temporais, bem como construir e testar modelos comportamentais complexos, nomeadamente recorrendo a modelos com desfasamentos distribuídos com poucas restrições.

No entanto, a análise econométrica com dados em painel não está isenta de problemas, nomeadamente porque:

- aumenta o risco de se ter amostras incompletas ou com graves problemas de recolha de dados, bem como a importância dos erros de medida;
- se virmos uma população como um conjunto de decisões que se reflectem em diferentes histórias individuais (segundo uma definição de Haavelmo), estas terão que ser representadas como variáveis aleatórias idiossincráticas (i.e., específicas a cada indivíduo) e que certamente estarão correlacionadas não apenas com a variável dependente, mas também com o conjunto das variáveis explicativas, o que causa diversos problemas ao nível da identificação e estimação dos modelos;
- ocorre o chamado enviesamento de heterogeneidade, i.e., o enviesamento resultante de uma má especificação pela não consideração de uma eventual diferenciação dos coeficientes ao longo das unidades seccionais e/ou ao longo do tempo;
- surgem problemas relacionados com o enviesamento de selecção (*selectivity bias*), ou seja, erros resultantes da recolha dos dados que levam a que estes não constituam uma amostra aleatória. Inclui questões como a auto-selectividade (amostras truncadas) e ausência de resposta ou atrito (exclusão de indivíduos em sucessivas rondas devido a morte ou alteração de residência, por exemplo). Uma forma particular de enviesamento de selecção, comum nos estudos macroeconométricos, relaciona-se com a selecção das unidades individuais a utilizar no estudo. Uma selecção de acordo com um critério sistemático, como é usualmente efectuado em macroeconomia, do tipo, "países da OCDE" ou "países que aderiram a um dado regime de política económica", não garantirá a constituição de uma amostra aleatória e, dessa forma, levará a que a estimação seja genericamente inconsistente.

Na próxima secção, apresenta-se o tema da estimação de modelos uniequacionais com dados em painel, através de uma abordagem genérica dos modelos estáticos de efeitos fixos e de efeitos aleatórios, limitados ao termo independente, o que é usualmente conhecido como modelo de componentes de erro. Omitiram-se propositadamente os modelos de coeficientes variáveis (fixos ou aleatórios) mais gerais, para limitar o espectro de análise.

No terceiro capítulo, abordam-se especificamente os modelos dinâmicos, quer de efeitos fixos, quer de efeitos aleatórios e os problemas que se colocam na sua estimação, por terem variáveis dependentes desfasadas como variáveis explicativas. É ainda feita uma breve referência a uma abordagem à heterogeneidade individual alternativa aos modelos de componentes de erro e que recorre a especificações diversas para as matrizes de variâncias e covariâncias dos termos de perturbação.

Na quarta secção, resumem-se as questões relacionadas com a estimação e identificação de modelos com desfasamentos distribuídos, seguindo-se de perto a exposição que consta da secção 8.1 de Hsiao [1986].

No quinto capítulo, introduz-se a estimação de modelos dinâmicos com dados em painel através de técnicas GMM, recorrendo-se extensivamente a Mátyás [1999] e Baltagi [1995].

No sexto capítulo, apresentam-se as principais referências sobre testes de detecção da existência de raízes unitárias, com amostras em painel, questão particularmente relevante para os modelos dinâmicos.

O capítulo sete apresenta o tema da estimação de modelos dinâmicos com dados discretos, nomeadamente a estimação de modelos de escolha binária.

O capítulo oitavo resume os objectivos e conclusões de um conjunto restrito de artigos merecedores de destaque, por serem textos fundadores ou por, sendo recentes, poderem dar indicações quanto a eventuais evoluções da discussão em torno da estimação de modelos econométricos dinâmicos com dados em painel. São também apresentados diversos testes de especificação adaptados a modelos com dados em painel.

O último capítulo apresenta as principais conclusões da recolha.

2. MODELOS ESTÁTICOS: ALGUMAS CONSIDERAÇÕES

Considere-se a seguinte especificação genérica para um modelo de dados em painel:

$$y_{it} = \beta_{1it}x_{1it} + \beta_{2it}x_{2it} + \dots + \beta_{kit}x_{kit} + u_{it} \Leftrightarrow y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}_{it} + u_{it} \quad ; i = 1, \dots, N \text{ e } t = 1, \dots, T \quad (2.1)$$

em que $\boldsymbol{\beta}_{it}$ corresponde ao vector (k×1) de parâmetros desconhecidos relativos ao indivíduo i no momento t e $\mathbf{x}_{it}$ a matriz (k×1) de variáveis explicativas, cuja primeira coluna, no caso do modelo ter termo independente, será integralmente constituída por 1's.

O modelo é tão só descritivo, na medida em que diz apenas que o indivíduo i tem uma dada função de reacção específica a cada momento no tempo. Além do mais, o modelo não é estimável pois tem mais coeficientes que observações, pelo que será necessário conferir-lhe uma estrutura. Nesse sentido, pode-se recorrer aos tradicionais três tipos de pressupostos - quanto às variáveis explicativas, os termos de perturbação e a relação estatística destes com aquelas - e ao pressuposto específico aos estudos com dados em painel, quanto à "variabilidade" dos coeficientes.

Desde logo, num modelo estático, assumimos que as variáveis explicativas são independentes dos termos de perturbação. Já no que toca à questão da heterogeneidade, podemos assumir que esta reside nos coeficientes de regressão (que podem variar no tempo ou de indivíduo para indivíduo) ou na estrutura dos termos de perturbação. A escolha de uma especificação de validade universal é impossível, restando-nos escolher aquela que, face aos dados em concreto e ao tipo de problema em causa, melhor se adeque.

Podemos, então, conceber sete especificações simples, com grau crescente de heterogeneidade (entre indivíduos pois, omitiu-se a heterogeneidade temporal para não complicar a exposição), recorrendo aos seguintes pressupostos suplementares:

I. Modelo de Regressão Simples

Esta especificação mais simples (e também mais irrealista) assume que o comportamento é uniforme para todos os indivíduos e ao longo do tempo e que todas as observações são homogéneas (i.e., da mesma população), pelo que poderá ser especificado da seguinte forma:

a(I): $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}$, $\forall_{i,t}$, em que $\boldsymbol{\beta}$ é (k×1);

b(I): $u_{it} \sim$ i.i.d.$(0,\sigma^2)$.

O modelo poderá ser estimado pela aplicação de OLS à amostra longitudinal visto cumprirem-se as hipóteses clássicas do modelo de regressão linear, no que é conhecido como *pooled* OLS. No entanto, ao não dar conta de uma heterogeneidade eventualmente existente, o modelo padecerá de um grave erro de especificação e os enviesamentos serão grandes. Além disso, por ignorar a existência de heterogeneidade nos dados, a aplicação de OLS em *pool* não é verdadeiramente um método de estimação em painel.

II. Modelo de Regressão Individual

Uma forma simples de se dar conta da heterogeneidade existente é assumir que os coeficientes são constantes no tempo, mas específicos a cada indivíduo, ou seja:

a(II): $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_i$, $\forall_t$, em que $\boldsymbol{\beta}_i$ é (k×1);

b(II): $u_{it} \sim$ i.i.d.$(0,\sigma_i^2)$.

A estimação do modelo reduz-se à aplicação de OLS indivíduo a indivíduo, com as vantagens de ser fácil de calcular, tratar as diferenças individuais explicitamente e permitir que se testem diferenças comportamentais. No entanto, tem como principais defeitos o facto de produzir um número muito elevado de coeficientes (não é parcimonioso, um dos mais importantes atributos de uma boa especificação), ter pouca fiabilidade quando $N\to\infty$ mas T é fixo e não contemplar a existência de interdependência entre as decisões individuais (o que o afasta do próprio espírito dos estudos com amostras longitudinais).

III. Modelo *Seemingly Unrelated Regression* (SUR)

A hipótese de inexistência de interdependência individual é bastante irrealista uma vez que é natural que alguns factores não observáveis (incluídos nos termos de perturbação) possam afectar todos ou alguns dos indivíduos ao mesmo tempo, originando correlação contemporânea entre as perturbações de dois indivíduos, que pode ser traduzida através de uma dada estrutura não nula de covariâncias das perturbações, nomeadamente:

a(III): $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_i$, $\forall_t$, em que $\boldsymbol{\beta}_i$ é (k×1);

b(III): $E(u_{it}) = 0$, $\forall_{i,t}$

$E(u_{it}\, u_{js}) = \sigma_{ij} \Leftarrow t = s$

$E(u_{it}\, u_{js}) = 0 \Leftarrow t \neq s$

o que não é mais que o modelo SUR de Zellner. Quando T→∞, o modelo revela-se muito abrangente ao dar conta da heterogeneidade individual, ao mesmo tempo que explicita uma interdependência relativamente livre. No entanto, na maior parte dos estudos em painel, em que N é grande e T relativamente pequeno, o modelo consome demasiados graus de liberdade, tornando a sua estimação pouco eficiente, se não mesmo impossível (o número de parâmetros a estimar ascenderá a Nk+N(N+1)/2).

IV. Modelo de Efeitos Fixos (Análise de Covariância)

Uma forma de conjugar a parcimónia com a heterogeneidade e a interdependência é admitir que os coeficientes β são idênticos para todos os indivíduos, com excepção do termo independente β_{1i}, que é específico a cada indivíduo, mantendo-se a hipótese da homogeneidade das observações. Formalmente:

a(IV): $b_{kit} = b_k$, $\forall_{i,t}$, excepto para k=1, caso em que $b_{1it} = b_{1i}$

b(IV): $u_{it} \sim$ i.i.d.$(0,\sigma^2)$.

Uma forma mais simples de enunciar a especificação é fazer $b_{1it} = b_1 + a_i$, pelo que o modelo base passa a ser:

$$y_{it} = a_i + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \tag{2.2}$$

Este modelo é chamado de Análise de Covariância, um caso específico da família de Modelos de Efeitos Fixos[1], ou modelo de variáveis *dummy* individuais. O modelo é relativamente fácil de estimar, trata as diferenças individuais de forma sistemática e permite que as mesmas sejam testadas.

V. Modelo de Efeitos Aleatórios (Componentes de Variância)

Os efeitos individuais da especificação anterior resultam de uma série de factores individuais, constantes no tempo (embora esta restrição possa ser relaxada, como veremos) e não observáveis. Desta forma, talvez seja mais razoável tratá-los como se de termos de perturbação se tratassem, i.e., especificar os efeitos individuais, não de forma determinística, mas aleatória. A escolha de uma ou outra especificação pode e deve ser procurada nos pressupostos comportamentais de base. Assim, se se crê que os efeitos individuais resultam de um grande número de factores não aleatórios, a especificação com efeitos fixos é mais lógica. Para amostras de grande dimensão, o número de parâmetros a estimar, com efeitos fixos, pode ser relativamente elevado, pelo que uma especificação que relega as diferenças individuais para uma componente não sistemática, logo, não estimável, parece ser mais apropriada.

Este modelo de componentes de erro introduz a heterogeneidade individual no termo de perturbação que poderá ser dividido em duas partes: uma comum, com média nula e variância σ^2_u e uma individual, também com média zero, mas com variância σ^2_α e que se assumem independentes. Formalmente:

a(V): $\beta_{it} = \beta, \forall_{i,t}$, em que β é $(k\times1)$;

b(V): $v_{it} = \alpha_i + u_{it}$

Se se preferir, o modelo pode ser visto como um modelo em que o termo independente é aleatório, com $\beta_{1i} = \beta_1 + \alpha_i$ e $E(\alpha_i) = 0$.

[1]A notação "Efeitos Fixos" é frequentemente usada exclusivamente para este tipo de modelo, ainda que deva ser aplicada a todos os modelos em que os parâmetros (termo independente e coeficientes associados a variáveis explicativas) são "variáveis" de indivíduo para indivíduo, mas de forma não aleatória.

VI. <u>Modelo de Coeficientes Aleatórios</u>

Podemos ainda generalizar a especificação anterior estendendo a aleatoriedade a todos os coeficientes através da inclusão de uma heteroscedasticidade individual das perturbações. De forma mais clara:

a(VI): $b_{kit} = b_k + a_{ki}$, $\forall_{t,k}$ com b_k fixo e a_{ki} aleatório.

b(VI): $u_{it} \sim$ i.i.d.$(0,\sigma_i^2)$.

A estimação do modelo acaba por conduzir a apenas k parâmetros de interesse, ainda que a estimação da matriz de covariâncias dos componentes aleatórios consuma bastantes graus de liberdade. Ao contrário do modelo de componentes de erro, que necessita da estimação do modelo de covariância (possível quando $T \geq 2$), este modelo de efeitos aleatórios mais geral necessita da estimação do modelo individual (II), o que só será possível se $T \geq k$.

VII. <u>Modelo *Time Series Cross Section* (TSCS) de Kmenta</u>

Uma forma radicalmente diferente de representar a heterogeneidade latente nos dados em painel é proposta por Kmenta [1986] e defendida em Greene [1997], Capítulo 16 e consiste na consideração de estruturas alternativas para a matriz de variâncias e covariâncias dos termos de perturbação. Partindo de um modelo *standard*, temos:

$$y_{it} = \mathbf{x}'_{it} b + u_{it}$$

em que u_{it} será um termo de perturbação com média nula e matriz de variâncias $\mathbf{V}$. Será precisamente a concepção de diferentes configurações para $\mathbf{V}$ que dará origem a diferentes justificações para a heterogeneidade dos dados. No caso de $\mathbf{V}=\sigma^2\mathbf{I}_{NT}$, temos a versão *pooled* OLS do modelo em painel que será, de todas, a menos interessante. A versão mais complexa será a de admitir que

$$u_{it} = r_i u_{i,t-1} + e_{it}$$

$$E(u_{it})=0$$

$$E\left(u_{it}^2\right) = \frac{s_{ii}}{1 - r_i^2}$$

para a qual haverá um estimador GLS eficiente, desde que os regressores sejam ortogonais face aos termos de perturbação. Atendendo a que o número de parâmetros a estimar cresce com N, trata-se de uma especificação particularmente adequada para painéis relativamente longos e estreitos (N pequeno e T grande). Para os painéis típicos, com uma dimensão seccional relativamente grande face à temporal, as especificações com componentes de erro serão mais parcimoniosas, para além de serem mais robustas face a erros de especificação.

Das várias especificações de modelos de dados em painel, duas sobressaem: efeitos fixos e efeitos aleatórios. A primeira é mais apropriada para os casos em que se retiram amostras exaustivas de uma população ou quando se pretende prever o comportamento individual. A segunda é mais coincidente com a visão de Haavelmo da econometria e das populações que são objecto do seu estudo - não um conjunto de indivíduos, mas um conjunto de decisões.

Quanto à terminologia, é comum entender-se que os modelos com efeitos fixos e os com efeitos aleatórios se podem englobar numa mesma classe de modelos de componentes de erro, em que o primeiro será o caso geral (por não se assumir qualquer distribuição para os efeitos) e o segundo um caso particular, em que se admite uma dada distribuição para a heterogeneidade. De seguida proceder-se-á a uma análise mais detalhada destes dois tipos de modelos, ainda que apenas a um nível introdutório.

2.1 Modelos de Efeitos Fixos

Quando falamos em modelos de efeitos fixos, temos em mente modelos cujos coeficientes podem variar de indivíduo para indivíduo ou no tempo, ainda que permaneçam como constantes fixas, logo, não aleatórias. Se a heterogeneidade seccional e/ou temporal se evidencia apenas no termo independente, dizemos estar perante um modelo de covariância.

Uma especificação simples para um modelo linear de efeitos fixos com uma fonte de erro (i.e., que considera apenas uma heterogeneidade seccional) será a seguinte:

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \Leftrightarrow \tag{2.3}$$

$$\mathbf{y}_i = \mathbf{i}_T \alpha_i + \mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\upsilon}_i \Leftrightarrow, \ i=1,\dots,N \text{ e } t=1,\dots,T \tag{2.4}$$

$$\mathbf{Y} = \mathbf{D}_N\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\upsilon} \tag{2.5}$$

em que $\boldsymbol{\beta}$ é o vector de (k-1) coeficientes associados às variáveis explicativas (exclui-se, pois, o termo independente), $\mathbf{x}'_{it}$ a linha de (k-1) colunas relativas aos valores assumidos pelas variáveis explicativas, para o i-ésimo indivíduo, no momento t e u_{it} o termo de perturbação genérico, que se assume respeitar $E(u_{it}) = 0$ ($\forall i,t$), $E(u_{it}\,.u_{js}) = \sigma^2$, se $i=j$ e $t=s$ e $E(u_{it}\,.u_{js}) = 0$, caso contrário. A segunda representação corresponde a uma agregação do modelo para os T períodos da amostra, em que $\mathbf{y}_i$ é o vector ($T\times1$) de y_{it}, $\mathbf{i}_T$ é um vector unitário[2] coluna ($T\times1$) e $\mathbf{x}_i$ a matriz $T\times(k\text{-}1)$, cujas linhas correspondem a T observações de cada uma das variáveis explicativas (excluindo termo independente). Finalmente, a última linha corresponde à notação matricial mais condensada, em que $\mathbf{Y}$ corresponde ao vector coluna ($NT\times1$) formado a partir da agregação vertical de $\mathbf{y}_i$, a matriz $\mathbf{D}_N$ é ($NT\times N$) e resulta de $\mathbf{D}_N = \mathbf{I}_N \otimes \mathbf{i}_T$, $\boldsymbol{\alpha}$ é o vector ($N\times1$) dos termos independentes e $\mathbf{u}$ é o vector-coluna ($NT\times1$) dos termos de perturbação.

Admite-se, ainda, que $\mathbf{X}$ é não aleatória, independente de $\mathbf{u}$ e que $[\mathbf{D}_N\ \mathbf{X}]$ tem característica $(N + k - 1) < NT$ (o que é sempre possível desde que $T \geq 2$)[3].

A estimação de $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ por OLS é BLUE, se se verificarem todos os pressupostos enunciados. Os estimadores são:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}_N\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_N\mathbf{Y} \tag{2.6}$$

$$\hat{\boldsymbol{\alpha}} = (\mathbf{D}'_N\mathbf{D}_N)^{-1}\mathbf{D}'_N(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{1}{T}\mathbf{D}'_N(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \tag{2.7}$$

$$\hat{s}^2 = \frac{\mathbf{Y}'\mathbf{W}_N\mathbf{Y}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{W}_N(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{NT-N-K+1}$$
$$Var(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}'\mathbf{W}_N\mathbf{X})^{-1} \tag{2.8}$$

em que $\mathbf{W}_N = \mathbf{I}_{NT} - \mathbf{D}_N(\mathbf{D}'_N\mathbf{D}_N)^{-1}\mathbf{D}'_N = \mathbf{I}_{NT} - 1/T\,\mathbf{D}_N\mathbf{D}'_N = \mathbf{I}_{NT} - 1/T(\mathbf{I}_N \otimes \mathbf{J}_T)$ é uma matriz simétrica idempotente de ordem NT e característica (NT-N) e $\mathbf{J}_T$ uma matriz unitária de ordem T. A matriz $\mathbf{W}_n$ é um operador de desvio em relação à média, pelo que, $\mathbf{W}_N\mathbf{u}$ tem elemento genérico ($u_{it} - \hat{u}_i$), por exemplo.

Note-se que esta forma de estimação do modelo não é mais que a estimação por OLS do modelo (2.5) transformado mediante a pré-multiplicação por $\mathbf{W}_n$, ou seja, da sua

[2] Ou seja, cujos elementos são todos iguais a 1.
[3] Isto implica também que todas as colunas de $\mathbf{X}$ sejam linearmente independentes de $\mathbf{D}_N$.

estimação na forma de desvios em relação à média (pelo Teorema de Frisch-Waugh[4] e atendendo a que $\mathbf{W}_n\mathbf{D}_N = 0$):

$$\mathbf{W}_n\mathbf{Y} = \mathbf{W}_n\mathbf{X}\boldsymbol{\beta} + \mathbf{W}_n\mathbf{u} \tag{2.10}$$

a que também se chama estimação por LSDV (*Least Square Dummy Variable*)[5], sendo que o estimador de $\boldsymbol{\alpha}$ mais não é que:

$$\hat{a}_i = \sum_{t=1}^{T} y_{it} - \sum_{k=2}^{K} \hat{b}_k \bar{x}_{ki} \Leftrightarrow \hat{\boldsymbol{\alpha}} = \bar{\mathbf{y}} - \hat{\boldsymbol{\beta}}\bar{\mathbf{x}} \tag{2.11}$$

O método LSDV, na prática, elimina todos os efeitos que não variam com o tempo (como o sexo, religião, ...) e obriga a uma grande perda de graus de liberdade. Apesar de tudo, os estimadores LSDV (ou intra-grupo) são BLUE, como dissemos, desde que as perturbações sigam as hipóteses clássicas e, com $N \to \infty$ e $T \to \infty$, consistentes. No entanto, se $N \to \infty$ e T fixo, apenas o estimador LSDV de $\boldsymbol{\beta}$ permanece consistente pois o número de parâmetros α_i aumenta indefinidamente.

Sob a hipótese de normalidade das perturbações, a usual inferência estatística em amostras finitas permanece válida. Assim, um teste t passará a ter NT - N - K + 1 graus de liberdade e um teste à existência de efeitos individuais pode ser efectuado através de um simples teste de Chow com o RSS a ser retirado de uma regressão OLS sobre o modelo homogéneo (I) descrito atrás e o USS a vir da estimação do modelo (1) por LSDV. A estatística (retirada de Baltagi[1995],12) é a seguinte:

$$F_0 = \frac{(RSS - USS)/(N-1)}{USS/(NT - N - K)} \sim F(N-1, NT-N-K)$$

A estimação do modelo de efeitos fixos com heterogeneidade temporal é em tudo análoga à apresentada para os efeitos seccionais, sendo a matriz de transformação para o LSDV, agora, $\mathbf{W}_t = \mathbf{I}_{NT} - 1/N\,(\mathbf{J}_N \otimes \mathbf{I}_T)$. As propriedades dos estimadores LSDV mantêm-se, com a variante de o estimador de $\boldsymbol{\alpha}$ ser consistente desde que $N \to \infty$.

Finalmente, a estimação de um modelo com efeitos fixos individuais e temporais simultaneamente não constitui uma generalização excessivamente difícil: uma abordagem

---

[4] Greene[1997],246.

[5] De facto, pode-se distinguir esta estimação por LSDV de uma estimação OLS sobre o modelo não transformado, com inclusão de *N dummies* e que aqui se omitiu porque esta é, na maior parte dos casos, impraticável atendendo à grande dimensão da matriz a ser invertida ($NT \times [N+K]$).

sintética, mas bastante clara pode ser encontrada em Mátyás e Sevestre [1992],34-43 ou em Baltagi[1995], cap.3). O estimador LSDV ou *within* na presença de efeitos fixos seccionais e temporais opera também na base da diferenciação dos dados relativamente a médias. Considere-se um modelo tipo

$$\begin{aligned} y_{it} &= \alpha_i + \lambda_t + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \Leftrightarrow \\ \mathbf{y}_i &= \mathbf{i}_T\alpha_i + \boldsymbol{\lambda} + \mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\upsilon}_i \Leftrightarrow \\ \mathbf{Y} &= \mathbf{D}_N\boldsymbol{\alpha} + \mathbf{D}_T\boldsymbol{\lambda} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\upsilon} \end{aligned}$$

em que se torna necessário admitir duas restrições de normalização dos efeitos específicos, a saber:

$$\sum_{i=1}^{N}\alpha_i = 0 \text{ e } \sum_{t=1}^{T}\lambda_t = 0$$

por forma a que se evitem situações de multicolinearidade (vulgo "armadilha das *dummies*"). A transformação *within* será

$$(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}) = (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i - \bar{\mathbf{x}}'_t + \bar{\mathbf{x}}')\boldsymbol{\beta} + (u_{it} - \bar{u}_i - \bar{u}_t + \bar{u})$$

e a estimação prosseguirá por OLS sobre o modelo transformado, pelo que, o estimador LSDV de $\boldsymbol{\beta}$ é dado por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{\mathbf{Y}'\mathbf{W}\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{W}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{NT - N - T - K + 1}$$

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$$

em que $\mathbf{W} = \mathbf{I}_{NT} - 1/T\ \mathbf{D}_N\mathbf{D}'_N - 1/N\ \mathbf{D}_T\mathbf{D}'_T + 1/NT\ \mathbf{D}_N\mathbf{D}'_N\mathbf{D}_T\mathbf{D}'_T = \mathbf{I}_{NT} - 1/T(\mathbf{I}_N \otimes \mathbf{J}_T) - 1/N(\mathbf{I}_T \otimes \mathbf{J}_N) + 1/NT\ \mathbf{J}_N \otimes \mathbf{J}_T$. Obviamente, a transformação *within*, ao eliminar os efeitos seccionais e temporais, não permite a estimação de variáveis que não variem no tempo ou de indivíduo para indivíduo.

A consideração de perturbações não esféricas leva a que os estimadores LSDV não sejam já BLUE – os estimadores eficientes serão os GLS ou os MLE. O método GLS merece, no entanto, algumas considerações.

Podemos considerar três casos especiais de perturbações não esféricas:

1. <u>Perturbações serialmente independentes:</u> só há correlação contemporânea entre as perturbações, i.e., não há autocorrelação temporal mas os erros estão correlacionados seccionalmente (ou seja, há interdependência entre os indivíduos, tal como se admite no modelo SUR, já referido). Neste caso, com $\mathbf{A}_{ij}$ a representar a matriz de covariâncias entre os termos de perturbação do indivíduo *i* e do indivíduo

j, temos: $\mathbf{A}_{ij} = \sigma_{ij}\ \mathbf{I}_T$ e $\text{Var}(\mathbf{u}) = \sigma^2 \cdot \mathbf{\Omega} = \mathbf{A}_{ij} \otimes \mathbf{I}_T$, com $\mathbf{A}_{ii} = \mathbf{\Sigma}_i$ a serem matrizes diagonais eventualmente iguais (não escalares, no caso de heteroscedasticidade);

2. <u>Independência Individual:</u> cobre todos os casos de heteroscedasticidade e autocorrelação ao nível individual, sendo de particular interesse o caso de homoscedasticidade em bloco dado por $\mathbf{A}_{ij} = 0$ se $i \neq j$ e $\mathbf{A}_{ii} = \mathbf{A} \Rightarrow \text{Var}(\mathbf{u}) = \mathbf{I}_N \otimes \mathbf{A}$;
3. <u>Equi-correlação em bloco:</u> $\mathbf{A}_{ii} = \mathbf{A}$ e $\mathbf{A}_{ij} = \mathbf{B}$, com $\mathbf{A}$ positiva definida, $\mathbf{B}$ semi-definida positiva e $(\mathbf{A} - \mathbf{B})$ positiva definida, de tal forma que $\text{Var}(\mathbf{u}) = \mathbf{I}_N \otimes (\mathbf{A} - \mathbf{B}) + \mathbf{J}_N \otimes \mathbf{B}$.

Pode ser demonstrado que a aplicação de um GLS habitual (i.e., recorrendo a $\mathbf{\Omega}^{-1}$) sobre o modelo transformado por $\mathbf{W}$ só será equivalente (para β) à aplicação de um método GLS sobre o modelo original nos casos 1 e 3. Para uma prova e clarificação do argumento, veja-se Mátyás e Sevestre [1992], 44-5.

2.2 Modelos de Efeitos Aleatórios

Esta especificação pressupõe que o comportamento específico dos indivíduos e períodos de tempo é desconhecido, não podendo ser observado, nem medido: é parte da nossa "ignorância geral". Assim, em amostras longitudinais de grande dimensão, podemos sempre representar estes efeitos individuais ou temporais específicos sob a forma de uma variável aleatória normal.

Das vantagens dos modelos de efeitos aleatórios e para o caso específico dos modelos de componentes de erro, destacam-se:

- a sua capacidade para trabalhar com bases de dados de qualquer dimensão;
- o facto de a inferência estatística aplicável ser uma mera derivação dos testes de hipóteses usuais;
- a possibilidade de a maior parte dos problemas e dificuldades poderem ser resolvidos dentro do quadro econométrico tradicional;
- o facto de ser o modelo de dados em painel estudado com maior profundidade;
- a facilidade com que são interpretados os resultados de estimação;
- o facto de ser pouco exigente em termos de *software* econométrico.

Ao contrário do que foi visto para o modelo de efeitos fixos, a heterogeneidade não é induzida através do termo independente, logo, através de $E(y_{it})$, mas sim através da variância da variável endógena. Uma especificação geral poderá ser:

$$\begin{aligned} &\mathbf{Y} = \alpha \mathbf{i}_{NT} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \alpha \mathbf{i}_{NT} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} \otimes \mathbf{i}_T + \mathbf{v} \Leftrightarrow \\ &y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \;\text{ e }\; u_{it} = \mu_i + v_{it} \end{aligned} \quad (2.12)$$

em que μ_i (elemento genérico do vector coluna (N×1), $\boldsymbol{\mu}$) será a variável aleatória dos efeitos individuais e v_{it} o termo de perturbação geral.

Assume-se, ainda, que:

(a) μ_i é ortogonal em relação a v_{it};

(b) $E(\mu_i) = E(v_{it}) = 0$;

(c) $E(v_{it}\, v_{js}) = \sigma_v^2$ se i=j e t=s
    $E(v_{it}\, v_{js}) = 0$, caso contrário;

(d) $E(\mu_i \mu_j) = \sigma_\mu^2$ se i=j
    $E(\mu_i \mu_j) = 0$, caso contrário;

(e) $v_{it} \sim \boldsymbol{N}(0, \sigma_v^2)$ e $\mu_{it} \sim \boldsymbol{N}(0, \sigma_\mu^2)$;

(f) $\mathbf{X}$ é independente de $\boldsymbol{\mu}$ e $\mathbf{v}$.

A matriz de variâncias e covariâncias será $\boldsymbol{\Omega} = \mathbf{D}_n \times E(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{D}'_\mathbf{n} + E(\mathbf{v}\mathbf{v}') = \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_v^2\mathbf{I}_{NT} = (T\sigma_\mu^2 + \sigma_v^2)\mathbf{P} + \sigma_v^2\mathbf{W}_n$ [6], em que $\mathbf{P} = 1/T \times \mathbf{B}_n = 1/T \times (\mathbf{I}_N \times \mathbf{J}_T)$ e $\mathbf{W}_n$ retém o significado atribuído em 2.1. A estrutura assumida para as perturbações equivale a que $\boldsymbol{\Omega}$ seja uma matriz de covariâncias equi-correlacionadas em bloco que exibe correlação temporal entre as perturbações do mesmo indivíduo, pois $cov(u_{it}\, u_{js}) = \sigma_\mu^2 + \sigma_v^2$ se i = j e t=s, $cov(u_{it}\, u_{js}) = \sigma_\mu^2$ se i=j e t≠s e $cov(u_{it}\, u_{js}) = 0$, nos restantes casos.

A estimação por OLS do modelo com efeitos aleatórios, ainda que permaneça cêntrica, consistente (com as restrições já referidas[7]) e assimptoticamente normal, não é já eficiente, atendendo à configuração de $\boldsymbol{\Omega}$. O mesmo pode ser afirmado em relação ao LSDV: permanece cêntrico, consistente, mas já não será eficiente. O candidato óbvio é o

---

[6] Esta é a decomposição espectral de $\boldsymbol{\Omega}$, sendo $(T\sigma_\mu^2 + \sigma_v^2)$ e σ_v^2 as primeiras e segundas (únicas) raízes características de multiplicidade N e N(T-1), logo, $|\boldsymbol{\Omega}| = (T\sigma_\mu^2 + \sigma_v^2)^N \times (\sigma_v^2)^{N(T-1)}$.

[7] Mas já não será consistente se tivernos efeitos seccionais e temporais (Mátyás e Sevestre[1992], 51)

estimador GLS, para o que precisamos de $\mathbf{\Omega}^{-1}$ e $\mathbf{\Omega}^{-1/2}$ dadas pela decomposição de Wansbeek e Kapteyn:

$$\begin{aligned}\mathbf{\Omega}^{-1} &= \frac{1}{(T\sigma_\mu^2+\sigma_v^2)}\mathbf{P}+\frac{1}{\sigma_v^2}\mathbf{W}_n \\ \mathbf{\Omega}^{-1/2} &= \frac{1}{\sqrt{(T\sigma_\mu^2+\sigma_v^2)}}\mathbf{P}+\frac{1}{\sigma_v}\mathbf{W}_n\end{aligned} \qquad (2.13)[8]$$

O estimador GLS de $\boldsymbol{\beta}^*$ =[α $\boldsymbol{\beta}$], que mais não é que a aplicação de OLS ao modelo (2.5) pré-multiplicado por $\sigma_v\mathbf{\Omega}^{-1/2}$, é:

$$\begin{aligned}\hat{\boldsymbol{\beta}}^*_{GLS} &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{Y} \Leftrightarrow \\ \hat{\boldsymbol{\beta}}^*_{GLS} &= (\mathbf{X}'\mathbf{W}_n\mathbf{X}+\frac{\sigma_V^2}{\sigma_V^2+T\sigma_\mu^2}\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}_n\mathbf{Y}+\frac{\sigma_V^2}{\sigma_V^2+T\sigma_\mu^2}\mathbf{X}'\mathbf{P}\mathbf{Y})\end{aligned} \qquad (2.14)$$

sendo a última fórmula[9] uma média ponderada de dois estimadores: o estimador LSDV ou Intra-Grupo (*Within*) e o estimador Inter-Grupo (*Between*) que mais não é que a aplicação de OLS ao modelo (2.12) expresso em termos de médias temporais para cada indivíduo (ou pré-multiplicado por **P**). Este último argumento é mais visível se pretendermos estimar apenas $\boldsymbol{\beta}$, para o que recorremos à fórmula proposta por Baltagi [1995]:

$$\hat{\boldsymbol{\beta}}_{GLS} = \left[\mathbf{X}'\mathbf{W}_n\mathbf{X}+\frac{\sigma_V^2}{\sigma_V^2+T\sigma_\mu^2}\mathbf{X}'(\mathbf{P}-\frac{1}{NT}\mathbf{J}_{NT})\mathbf{X}\right]^{-1}\left[\mathbf{X}'\mathbf{W}_n\mathbf{Y}+\frac{\sigma_V^2}{\sigma_V^2+T\sigma_\mu^2}\mathbf{X}'(\mathbf{P}-\frac{1}{NT}\mathbf{J}_{NT})\mathbf{Y}\right] \qquad (2.15)$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{GLS} &= \left[\mathbf{X}'\mathbf{W}_n\mathbf{X}+\frac{\sigma_V^2}{\sigma_V^2+T\sigma_\mu^2}\mathbf{X}'(\mathbf{P}-\frac{1}{NT}\mathbf{J}_{NT})\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{W}_n\mathbf{X}\hat{\boldsymbol{\beta}}_W + \\ &+\left[\mathbf{X}'\mathbf{W}_n\mathbf{X}+\frac{\sigma_V^2}{\sigma_V^2+T\sigma_\mu^2}\mathbf{X}'(\mathbf{P}-\frac{1}{NT}\mathbf{J}_{NT})\mathbf{X}\right]^{-1}\left(\frac{\sigma_V^2}{\sigma_V^2+T\sigma_\mu^2}\mathbf{X}'(\mathbf{P}-\frac{1}{NT}\mathbf{J}_{NT})\mathbf{X}\right)\hat{\boldsymbol{\beta}}_B \\ \hat{\boldsymbol{\beta}}_{GLS} &= \mathbf{W}_1\hat{\boldsymbol{\beta}}_W+\mathbf{W}_2\hat{\boldsymbol{\beta}}_B = \mathbf{W}_1\hat{\boldsymbol{\beta}}_W+(1-\mathbf{W}_1)\hat{\boldsymbol{\beta}}_B\end{aligned} \qquad (2.16)$$

Note-se que se $\sigma_\mu^2 = 0$, o estimador GLS é equivalente ao OLS aplicado ao modelo em nível, o que é compreensível pois estaríamos a dar idênticos pesos às variações intra-grupo e inter-grupos; no caso de $\sigma_v^2 = 0$, o GLS equivale ao LSDV, ou seja, estaríamos a considerar apenas o estimador Intra-Grupo porque seria impossível distinguir os efeitos aleatórios dos fixos. Ou seja, podemos ver o OLS e o LSDV como dois casos extremos: o

---

[8] Considera-se que $\mathbf{\Omega}^{1/2}$ é toda a matriz que verifica $\mathbf{\Omega}=(\mathbf{\Omega}^{1/2})(\mathbf{\Omega}^{1/2})'$.

[9] De acordo com Mátyás e Sevestre[1992], 52.

OLS dá um peso excessivo à variação entre as unidades, em vez de relegar parte destas para variações aleatórias atribuíveis ao termo de perturbação μ_i variável seccionalmente; o LSDV assume que $v_{it} = 0$, logo, a única variação entre os indivíduos deve-se a um efeito constante no tempo, pelo que, a questão de saber se este é fixo ou aleatório torna-se irrelevante. Finalmente, repare-se que se $T \rightarrow \infty$, o GLS é equivalente ao LSDV. Ou seja, o GLS pode ser visto como uma combinação óptima do estimadores *Within* e *Between*.

O estimador GLS será cêntrico e, pelo Teorema de Aitken, eficiente[10], com matriz de covariâncias

$$Var(\hat{\beta}_{GLS}) = \sigma_v^2 \left[\mathbf{X'W}_n\mathbf{X} + \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\mu^2}\mathbf{X'}(\mathbf{P} - \frac{1}{NT}\mathbf{J}_{NT})\mathbf{X} \right]^{-1} \tag{2.18}$$

Sob algumas condições de regularidade (ver Mátyás e Sevestre[1992], 49), o estimador GLS é consistente desde que $N \rightarrow \infty$.

Nos casos em que desconhecemos $\mathbf{\Omega}$, temos que recorrer a um método FGLS, sendo sugeridas várias formas de estimar as componentes de variância. Os estimadores BQU (*Best Quadratic Unbiased*) são, de acordo com Baltagi [1995]:

$$\begin{aligned} \hat{\sigma}_v^2 &= \frac{\mathbf{u'Pu}}{N} \\ \hat{\sigma}_\mu^2 &= \frac{\mathbf{u'W}_n\mathbf{u}}{N(T-1)} \end{aligned} \tag{2.19}$$

podendo-se substituir **u** pelos resíduos de estimação LSDV (Amemiya [1985] sugere estes uma vez que retêm a mesma distribuição assimptótica dos estimadores BQU), ainda que nada garanta que as estimativas sejam todas não-negativas. Para não estender a exposição, não se mencionam outros estimadores, embora se possa encontrar um resumo em Baltagi [1995]. Desde que os estimadores das componentes de variância sejam consistentes, os estimadores FGLS retêm as mesmas propriedades que o GLS (ainda que apenas assimptoticamente).

[10] Note-se que se com $T \rightarrow \infty$, o GLS é equivalente ao LSDV, em amostras com séries temporais longas (o que é, infelizmente, raro acontecer nos dados em painel), a perda de eficiência motivada pela utilização do estimador LSDV não deverá ser significativa, o que permite contornar as dificuldades calculatórias do GLS, nomeadamente nos casos em que $\mathbf{\Omega}$ é desconhecida.

Um outro método de estimação utilizado é o MLE, que pode ser obtido através da maximização da função de verosimilhança condensada em relação a α e σ_v^2 dada por

$$\ln L_C = k - \frac{NT}{2}\ln\left\{\left(\mathbf{Y} - \mathbf{X}'\hat{\beta}_{MLE}\right)'\left[\mathbf{W}_n + \phi^2\left(\mathbf{P} - \bar{\mathbf{J}}_{NT}\right)\right]\left(\mathbf{Y} - \mathbf{X}'\hat{\beta}_{MLE}\right)\right\} + \frac{N}{2}\ln\phi^2 \ ;$$
$$\phi^2 = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\mu^2} \text{ e } \bar{\mathbf{J}}_{NT} = \frac{1}{NT}\mathbf{J}_{NT} \qquad (2.20)$$

A maximização de lnL_C dá os estimadores de máxima verosimilhança, concretamente:

$$\hat{\phi}^2 = \frac{\left(\mathbf{Y} - \mathbf{X}'\hat{\beta}_{MLE}\right)'\mathbf{W_n}\left(\mathbf{Y} - \mathbf{X}'\hat{\beta}_{MLE}\right)}{(T-1)\left(\mathbf{Y} - \mathbf{X}'\hat{\beta}_{MLE}\right)'\left(\mathbf{P} - \bar{\mathbf{J}}_{NT}\right)\left(\mathbf{Y} - \mathbf{X}'\hat{\beta}_{MLE}\right)}$$
$$\hat{\beta}_{MLE} = \left[\mathbf{X}'(\mathbf{W_n} + \phi^2\mathbf{P})\mathbf{X}\right]^{-1}\mathbf{X}'\left[\mathbf{W_n} + \phi^2\mathbf{P}\right]\mathbf{Y} \qquad (2.21)$$

sendo que a solução terá que ser procurada iterativamente. Os estimadores dos restantes parâmetros são baseados nas sucessivas estimativas de β e ϕ^2, nomeadamente

$$\hat{\alpha}_{MLE} = \frac{1}{NT}\sum_i\sum_t y_{it} - \sum_i\sum_t \mathbf{X}'_{it}\hat{\beta}_{MLE} = \bar{\bar{y}} - \bar{\bar{\mathbf{X}}}\hat{\beta}_{MLE}$$
$$\hat{\sigma}^2_{v,MLE} = \frac{1}{NT}\left(\mathbf{Y} - \mathbf{X}'\hat{\beta}_{MLE}\right)'\left[\mathbf{W_n} + \phi^2(\mathbf{P} - \bar{\mathbf{J}}_{NT})\right]\left(\mathbf{Y} - \mathbf{X}'\hat{\beta}_{MLE}\right) \qquad (2.22)$$

e cuja derivação pode ser encontrada em Baltagi [1995],18-9. Num estudo realizado por Breusch, verificou-se que, quando muito, havia dois máximos na função de verosimilhança, pelo que só temos que nos precaver contra um máximo local, que possivelmente será uma solução de canto (para ϕ^2=1, i.e., quando começamos com o estimador OLS, caso na segunda iteração se obtenha uma estimativa para ϕ^2 superior). Um método sugerido[11] (e que se deve a Breusch) é que se comece com, por um lado o estimador *Between* e, pelo outro, com o estimador *Within* e se verifique se as duas sequências (que são monótonas) convergem para o mesmo máximo, caso em que este será global. Os estimadores MLE assim obtidos são geralmente consistentes, ainda que para σ_μ^2 tal só seja verdade no plano semi-assimptótico (i.e., com T fixo e $N\rightarrow\infty$).

Quanto à inferência estatística, em particular quanto aos testes à existência de efeitos individuais e/ou temporais, é sugerida a utilização dos habituais testes assimptóticos, mencionando-se o teste LM por ser menos exigente em termos de cálculos.

A estimação de modelos com efeitos temporais ou com efeitos seccionais e temporais não é muito distinta da que foi apresentada apenas para os efeitos individuais, ainda que, no segundo caso, se torne notacional e computacionalmente mais pesada, motivo pelo qual foi aqui omitida. Ressalve-se apenas o caso da estimação por MLE nos modelos com efeitos duplos, em que a maximização da função de verosimilhança é de tal forma não-linear e complicada que ainda se lhe não conhecem expressões analíticas. De qualquer forma, Baltagi[1995] sugere um método iterativo semelhante ao apresentado acima. Podem ser encontrados excelentes resumos do tema, tanto em Mátyás *et al*. [1992], como em Baltagi[1995].

Finalmente, nas formulações mais simples, assume-se uma dada estrutura para a matriz de variâncias e covariâncias dos termos de perturbação (i.e., admite-se a existência de variáveis omissas que não variam temporal ou seccionalmente), o que nos permite explorar a estrutura do processo de erro, quando T é fixo (e não muito grande) e $N \to \infty$. É, no entanto, possível estimar consistentemente os parâmetros sem impor restrições sobre $\mathbf{\Omega}$ (permitindo autocorrelação ou heteroscedasticidade) através de um método proposto por Chamberlain [1984] e que consiste em tratar-se cada período t como uma equação e estimar o modelo como um sistema de T equações com dados seccionais, dado por

$$\underset{(T\times 1)}{\mathbf{y}_i} = \mathbf{i}_T \mathrm{a}_i + (\mathbf{I}_T \otimes \boldsymbol{\beta}')\mathbf{x}_i + \mathbf{u}_i \tag{2.23}$$

e que será transformado ao assumir-se que a_i e $\mathbf{x}_i$ possam estar correlacionados, fazendo-se $E^*(\mathrm{a}_i|\mathbf{x}_i) = \mathrm{m} + \mathbf{a}'\mathbf{x}_i \Rightarrow E^*(\mathbf{y}_i|\ \mathbf{x}_i) = \mathbf{i}\mathrm{m} + \boldsymbol{\Pi}\ \mathbf{x}_i$, com $\boldsymbol{\Pi} = \mathbf{I}_T \otimes \boldsymbol{\beta}' + \mathbf{i}\mathbf{a}' = f(\boldsymbol{\theta}), \boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{\alpha}']$ e $\mathbf{v}_i = \mathbf{y}_i - E^*(\mathbf{y}_i|\ \mathbf{x}_i)$ e $E^*(y|x)$ representa a projecção de y no espaço definido por x[12], para

$$\underset{(T\times 1)}{\mathbf{y}_i} = \mathbf{i}_T \mathrm{m} + (\mathbf{I}_T \otimes \mathbf{x}_i')\boldsymbol{\delta} + \mathbf{v}_i \tag{2.24}$$

e $\boldsymbol{\pi}$ é um vector coluna (KT^2-1) que contém todas as linhas de $\boldsymbol{\Pi}$ (onde não se impõe qualquer restrição) e que será estimada por recurso a OLS, de forma consistente (e assimptoticamente normal) quando T fixo. Quando a matriz $\boldsymbol{\Pi}$ é sujeita a restrições, que se especificam por $\boldsymbol{\Pi} = f(\boldsymbol{\theta})$, a estimação dos parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\alpha}$ é realizada através de um estimador de distância mínima, i.e., escolhe-se a estimativa de $\boldsymbol{\theta}$ que minimize

---

[11] Baltagi [1995],19.

[12] Ou seja, o previsor de erro quadrático médio mínimo ou de mínimos quadrados de y definido como função de x.

$\left[\hat{\boldsymbol{\pi}} - \mathrm{f}(\boldsymbol{\theta})\right]' \hat{\boldsymbol{\Omega}}' \left[\hat{\boldsymbol{\pi}} - \mathrm{f}(\boldsymbol{\theta})\right]$, em que $\hat{\boldsymbol{\Omega}}$ é um estimador consistente da matriz de variâncias de $\hat{\boldsymbol{\pi}}$ [13]. Este estimador de distância mínima é consistente (no caso semi-assimptótico $N \rightarrow \infty$) e assimptoticamente normal, não impondo qualquer restrição sobre os termos de perturbação do modelo, mas apenas que as características individuais (para cada um dos T períodos) se encontrem idêntica e independentemente distribuídas em i. Mais, o estimador de distância mínima será eficiente relativamente a todos os outros estimadores que não impõem restrições na matriz de variâncias dos termos de perturbação, mas já não o será, por exemplo, quando sabemos estar perante um modelo de efeitos aleatórios (caso em que o estimador eficiente será o GLS), ainda que permaneça consistente.

2.3 Efeitos Fixos ou Aleatórios?

Parece haver desde logo uma vantagem computacional em pressupor-se efeitos fixos e não aleatórios, ainda que tal não possa ou não deva ser aduzido como justificação. Esta deve antes ser procurada na resposta a duas questões (Hsiao *in* Mátyás e Sevestre [1992], pág. 88): (1) os objectivos do estudo em questão e (2) o contexto dos dados, a forma como foram recolhidos e a envolvente onde foram gerados.

Assim, se o que se pretende é efectuar inferência relativamente a uma população, a partir de uma amostra aleatória da mesma, os efeitos aleatórios serão a escolha apropriada. Se se pretende estudar o comportamento de uma unidade individual em concreto, então os efeitos fixos são a escolha óbvia na medida em que é indiferente considerar-se a amostra como aleatória ou não. Em particular, no caso de se estar a estudar um grupo de N países, toda a inferência terá que ser condicional em ordem ao grupo específico sob observação.

Ou seja, na generalidade dos estudos macroeconométricos, por ser impossível ver uma amostra de N países como uma selecção aleatória de uma população com dimensão tendencialmente infinita, tanto mais que representará com grande probabilidade a quase totalidade da população em estudo, torna-se evidente que a escolha acertada é a especificação com efeitos fixos, como é defendido em Judson e Owen [1996].

[13] As fórmulas podem ser encontradas em Hsiao [1986],59-60.

3. Modelos Dinâmicos Autorregressivos

3.1 Introdução

A natureza mais comum das relações económicas é dinâmica e uma das vantagens dos dados em painel, como já foi visto, é facultar uma melhor compreensão das dinâmicas de ajustamento. Estas relações dinâmicas podem ser representadas por uma variável dependente desfasada como regressor, i.e., por um modelo da forma

$$\begin{aligned} y_{it} &= \delta y_{i,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \ ; i = 1, \ldots, N \ ; t = 1, \ldots, T \\ u_{it} &= \alpha_i + v_{it} \end{aligned} \tag{3.1}$$

em que, δ[14] é um escalar, $\mathbf{x}_{it}$ o vector-coluna (K×1) de variáveis exógenas e u_{it} o termo de perturbação com apenas uma fonte de erro (individual), verificando-se que α_i e v_{it} são i.i.d. e não mutuamente correlacionadas, com média nula e variâncias σ_α^2 e σ_v^2, respectivamente.

Antes de mais, uma vez que as propriedades dos estimadores com amostras finitas são, em grande medida desconhecidas, a escolha de um método de estimação, em detrimento de outro, deverá ser feita com base nas suas propriedades assimptóticas.

A escolha entre a especificação de efeitos fixos e aleatórios acaba por ter implicações radicalmente diferentes das verificadas nos modelos estáticos, nomeadamente quanto à consistência, centricidade e eficiência dos estimadores. Assim, quando todas as variáveis explicativas são exógenas, o LSDV é BLUE, na especificação de efeitos fixos e cêntrico e consistente, na de efeitos aleatórios, ainda que não eficiente, quando T fixo. Adicionalmente, se algumas ou todas as variáveis independentes se encontram correlacionadas com os termos de perturbação, o LSDV permanece cêntrico no caso de efeitos fixos (ainda que erradique esses efeitos não observáveis da estimação), mas um GLS que não seja corrigido por essa correlação será enviesado, o que contribuiu para que se privilegiasse a primeira especificação, nestes modelos.

No entanto, quando uma dessas variáveis explicativas é endógena desfasada, o estimador LSDV (que, recorde-se, é MLE, sob o pressuposto de normalidade dos termos de perturbação) não será consistente, no caso de efeitos fixos, com T fixo e a eventual

---

[14] Não é necessário impor a restrição de estacionaridade $|\delta|<1$ desde que T seja finito, pois podemos analisar sempre as propriedades dos estimadores no plano semi-assimptótico $N\to\infty$.

consistência dos estimadores MLE e GLS (e, mesmo, a interpretação do modelo), na especificação de efeitos aleatórios, dependerá crucialmente dos pressupostos assumidos quanto à primeira observação e da forma como *T* e *N* tendem para o infinito.

Um dos problemas com a estimação de modelos dinâmicos com dados em painel, comum aos estudos convencionais[15], é a correlação existente entre um dos regressores, $y_{i,t-1}$, e o termo de perturbação, u_{it}, via α_i. Esta situação torna os estimadores OLS enviesados e não consistentes, mesmo que v_{it} não exiba autocorrelação, podendo o "enviesamento" assimptótico ser significativo. A mesma situação se verifica para os estimadores LSDV, no caso do modelo de efeitos fixos, e GLS, para o modelo de efeitos aleatórios, uma vez que as transformações operadas para eliminar α_i não eliminam a correlação entre $y_{i,t-1}$ e o termo de perturbação resultante. Assim, torna-se de crucial importância a escolha de variáveis instrumentais que assegurem a consistência e eficiência da estimação.

Uma questão fundamental para analisar as propriedades assimptóticas dos estimadores é a clarificação do processo gerador da primeira observação, o que pode ser evidenciado através da representação do modelo (3.1) da seguinte forma

$$\begin{aligned} y_{it} &= \delta^{j} y_{i,0} + \sum_{j=0}^{t-1} \delta^{j} \mathbf{x}'_{it-j}\boldsymbol{\beta} + \frac{1-\delta^{t}}{1-\delta}\alpha_i + \sum_{j=0}^{t-1} \delta^{j} v_{i,t-j} \Leftrightarrow \\ y_{it} &= \delta^{j} y_{i,0} + \sum_{j=0}^{t-1} \delta^{j} \mathbf{x}'_{it-j}\boldsymbol{\beta} + \frac{1-\delta^{t}}{1-\delta}\alpha_i + \gamma_{it} \end{aligned} \tag{3.2}$$

Ou seja, cada observação da variável endógena pode ser expressa como a soma de quatro componentes: a primeira depende do valor inicial ou ponto de partida; a segunda dos valores actuais e passados das variáveis exógenas; a terceira do valor dos efeitos individuais α_i e a quarta não é mais que um processo autorregressivo, com valores iniciais fixos

$$\begin{cases} \gamma_{it} = \delta\gamma_{i,t-1} + v_{it} \\ \gamma_{i0} = 0 \end{cases} \tag{3.2′}$$

o que torna, desde logo, claro que as observações iniciais vão influenciar decisivamente as propriedades semi-assimptóticas dos estimadores (com *T* fixo), levantando-se a questão da sua eventual exogeneidade.

No entanto, como esta questão só assume relevância para a estimação com efeitos aleatórios, só será abordada no ponto 3.3.

---

[15] Aplicamos aqui a palavra convencional como sinónimo de estudos que recorram a amostras temporais ou

3.2 Estimação de Modelos com Efeitos Fixos

Como já foi visto, a escolha de uma especificação de efeitos fixos é mais apropriada quando a amostra é relativamente agregada (i.e., ao nível de sectores, regiões, países,…) e o objectivo do estudo não é a previsão do comportamento individual , bem como quando os efeitos individuais (não observáveis) não são independentes de alguma das variáveis explicativas.

Ao modelo (3.1) teremos apenas que acrescentar:

- $E(v_{it}|y_{i,t-1},\mathbf{x}_{it}) = 0$;
- $Var(v_{it}|y_{i,t-1},\mathbf{x}_{it}) = \sigma_v^2, \forall_{it}$;
- $Cov(v_{it},v_{js}|\, y_{i,t-1},\mathbf{x}_{it}) = 0$, $i\neq j$ e $t\neq s$

ou seja, os termos de perturbação são independentes das variáveis explicativas, não autocorrelacionados e homoscedásticos. Uma agregação seccional e temporal resulta no seguinte modelo condensado, análogo a (2.14):

$$\mathbf{Y} = d\mathbf{Y}_{-1} + \mathbf{D}_N\alpha + \mathbf{X}\beta + \mathbf{v} \qquad (3.3)$$

em que $\mathbf{D}_n$ permanece $\mathbf{I}_N \otimes \mathbf{i}_T$ e onde se acrescenta o vector-coluna ($NT\times 1$) $\mathbf{Y}_{-1}$ das observações desfasadas em um período de y_{it} (logo, inclui y_{i0}).

O estimador LSDV de $\theta' = [d\ \beta']$, dado por

$$\begin{bmatrix} \hat{d} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}'_{-1}\mathbf{W}_n\mathbf{Y}_{-1} & \mathbf{Y}'_{-1}\mathbf{W}_n\mathbf{X} \\ \mathbf{X}'\mathbf{W}_n\mathbf{Y}_{-1} & \mathbf{X}'\mathbf{W}_n\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}'_{-1}\mathbf{W}_n\mathbf{Y} \\ \mathbf{X}'\mathbf{W}_n\mathbf{Y} \end{bmatrix} \qquad (3.4)^{16}$$

é ainda equivalente ao MLE admitindo a normalidade de v_{it}, com y_{i0} dados e $\mathbf{W}_n\mathbf{Y} \neq 0$ (caso em que não existiria) e será consistente se $N\rightarrow\infty$ e $T\rightarrow\infty$, mas não quando T é fixo. A prova do argumento é relativamente simples, pois, invocando o Teorema de Slutsky, temos:

---

seccionais.

[16] A fórmula, retirada de Mátyás e Sevestre[1992],99, resulta da estimação por OLS do modelo (3.3) pré-multiplicado por $\mathbf{W}_n$. O estimador OLS de α, de menor interesse, é $\hat{\alpha} = \overline{\mathbf{Y}} - \hat{d}\mathbf{Y}_{-1} - \mathbf{X}\hat{\beta} \Leftrightarrow \hat{a}_i = \bar{y}_i - \hat{d}\bar{y}_{i,-1} - \bar{\mathbf{x}}_i\hat{\beta}$.

$$\underset{N\to\infty}{\text{plim}}\begin{bmatrix}\hat{\delta}\\ \hat{\beta}\end{bmatrix} = \begin{bmatrix}\delta\\ \beta\end{bmatrix} + \begin{bmatrix} \underset{N\to\infty}{\text{plim}}\dfrac{1}{NT}\mathbf{Y}'_{-1}\mathbf{W}_n\mathbf{Y}_{-1} & \underset{N\to\infty}{\text{plim}}\dfrac{1}{NT}\mathbf{Y}'_{-1}\mathbf{W}_n\mathbf{X} \\ \underset{N\to\infty}{\text{plim}}\dfrac{1}{NT}\mathbf{X}'\mathbf{W}_n\mathbf{Y}_{-1} & \underset{N\to\infty}{\text{plim}}\dfrac{1}{NT}\mathbf{X}'\mathbf{W}_n\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \underset{N\to\infty}{\text{plim}}\dfrac{1}{NT}\mathbf{Y}'_{-1}\mathbf{W}_n\mathbf{v} \\ \underset{N\to\infty}{\text{plim}}\dfrac{1}{NT}\mathbf{X}'\mathbf{W}_n\mathbf{v} \end{bmatrix}$$

e como, pela ortogonalidade das perturbações face a **X**, temos $\underset{N\to\infty}{\text{plim}}\frac{1}{NT}\mathbf{X}'\mathbf{W}_n\mathbf{v} = 0$, mas como

$$\underset{N\to\infty}{\textbf{\textit{plim}}}\frac{1}{NT}\mathbf{Y}'\mathbf{W}_n\mathbf{v} = \underset{N\to\infty}{\textbf{\textit{plim}}}\frac{1}{NT}\sum_i\sum_t(y_{i,t-1}-\bar{y}_{i,-1})(v_{i,t-1}-\bar{v}_{i,-1}) =$$
$$= E\left(\frac{1}{T}\sum_t(y_{i,t-1}-\bar{y}_{i,-1})(v_{i,t-1}-\bar{v}_{i,-1})\right) = -\frac{1}{T^2}\frac{T-1-T\delta+\delta^T}{(1-\delta)^2}\sigma_v^2 \qquad (3.5)$$

temos que (3.5) $\neq 0$, quando T fixo e quando $T\to\infty$ será zero apenas se $|\delta|<1$, como já havíamos notado. Assim, a semi-inconsistência do LSDV fica a dever-se à correlação (de ordem O(1/T)) entre ($y_{i,t-1}$ - $y_{i.-1}$) e ($v_{i,t}$ - $v_{i.}$), quando $N\to\infty$, mesmo que $y_{i,t-1}$ e $v_{i,t}$ sejam independentes, criada pela eliminação dos efeitos individuais (não observáveis) α_i do modelo transformado.

Ainda que, quando se dispõe de um painel com uma dimensão temporal grande, se espere um "enviesamento" assimptótico pequeno[17], como não é esse o caso mais frequente nas amostras longitudinais, torna-se essencial que se encontre um estimador consistente para θ, quando T é finito. Esse método é o das variáveis instrumentais (IV).

Um método IV sugerido por Balestra e Nerlove [1966] para o modelo de efeitos aleatórios (que veremos na secção seguinte) e que pode ser usado no modelo de efeitos fixos, sujeito a algumas alterações, consiste em considerar uma matriz **Z** de $m\geq$ K+1 variáveis instrumentais (p.ex., $\mathbf{x}_{it}$, $\mathbf{x}_{i,t-1}$, …) e estimar o modelo (3.3) por OLS pré-multiplicado pelo projector de **Y** em $\mathbf{Z}^*$, $P_{Z^*}=\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^*$[18], com $\mathbf{Z}^*=(\mathbf{D}_N, \mathbf{Z})$. Atendendo a que $P_{Z^*}\mathbf{X}=\mathbf{X}$ (assumindo que **X** está contido em **Z**), $P_{Z^*}\mathbf{D}_N=\mathbf{D}_N$ e que $\mathbf{W}_n$ e $\mathbf{D}_N$ são ortogonais, tal equivale à aplicação de OLS ao modelo

---

[17] Num modelo sem variáveis exógenas, o enviesamento assimptótico é, de acordo com Hsiao[1986], 74:

$$\underset{N\to\infty}{\text{plim}}(\hat{\delta}-\delta) = -\frac{1+\delta}{T-1}\left(1-\frac{1}{T}\frac{1-\delta^T}{1-\delta}\right)\left(1-\frac{2\delta}{(1-\delta)(T-1)}\left[1-\frac{1-\delta}{T(1-\delta)}\right]\right)^{-1}.$$

$$\mathbf{W}_n\mathbf{P}_{Z^*}\mathbf{Y} = \mathrm{d}\mathbf{W}_n\mathbf{P}_{Z^*}\mathbf{Y}_{-1} + \mathbf{W}_n\mathbf{X}\beta + \mathbf{W}_n\mathbf{P}_{Z^*}\mathbf{v} \qquad (3.6)$$

que, pelo Teorema de Frisch-Waugh, nos permite obter as estimativas de d e β. O resultado pode tornar-se mais eloquente se repararmos que $\mathbf{W}_n\mathrm{P}_{Z^*}=\mathbf{W}_n(\mathbf{W}_n\mathbf{Z}(\mathbf{Z}'\mathbf{W}_n\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}_n=\mathbf{P}_{WZ}$ e que o mesmo resultado pode ser obtido que em (3.6) se transformarmos o modelo (3.3) por pré-multiplicação pela matriz idempotente $\mathbf{P}_{WZ}$, que mais não é que estimar o modelo usando as variáveis instrumentos na forma centrada (i.e., como desvios em relação às suas médias), pelo que, para que estas sejam válidas, têm que ser estritamente exógenas. Atendendo a que $\mathbf{P}'_{WZ}\mathbf{P}_{WZ} = \mathbf{P}_{WZ}$, os estimadores IV são, então:

$$\begin{bmatrix} \hat{\mathrm{d}} \\ \hat{\beta} \end{bmatrix}_{IV} = \begin{bmatrix} \mathbf{Y}'_{-1}\mathbf{P}_{\mathbf{W}_N\mathbf{Z}}\mathbf{Y}_{-1} & \mathbf{Y}'_{-1}\mathbf{P}_{\mathbf{W}_N\mathbf{Z}}\mathbf{X} \\ \mathbf{X}'\mathbf{P}_{\mathbf{W}_N\mathbf{Z}}\mathbf{Y}_{-1} & \mathbf{X}'\mathbf{P}_{\mathbf{W}_N\mathbf{Z}}\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}'_{-1}\mathbf{P}_{\mathbf{W}_N\mathbf{Z}}\mathbf{Y} \\ \mathbf{X}'\mathbf{P}_{\mathbf{W}_N\mathbf{Z}}\mathbf{Y} \end{bmatrix} \qquad (3.7.)$$

Um outro estimador IV que se deve a Anderson e Hsiao [1981] poderá ser obtido se reescrevermos o modelo (3.3) nas primeiras diferenças (o que elimina α)

$$\Delta\mathbf{Y} = \mathrm{d}\Delta\mathbf{Y}_{-1} + \Delta\mathbf{X}\beta + \Delta\mathbf{v} \qquad (3.8)$$

e usar como instrumento para $\Delta\mathbf{Y}_{-1}$, alternativamente, $\Delta y_{i,t-2}=(y_{i,t-2} - y_{t-3})$ ou y_{t-2}, sendo ambos os estimadores daí resultantes consistentes, ainda que o IV baseado em y_{t-2} seja mais eficiente e menos exigente (basta que $T\geq 2$), segundo os mesmos autores.

De qualquer forma, os estimadores não serão completamente eficientes pois os termos de perturbação transformados são MA(1). Mas, se se tentar ultrapassar este problema através da pré-multiplicação do modelo nas primeiras diferenças por $\mathrm{y}^{-1/2}$, em que y é a matriz de variâncias e covariâncias dos termos de perturbação $\Delta\mathbf{v}$, obtém-se um modelo com termos de perturbação que serão combinações lineares de v_{it}, logo, correlacionados com quaisquer instrumentos não estritamente exógenos (i.e., Δy_{t-p} e Δy_{t-p}). Pode-se, no entanto, usar os valores desfasados de $\Delta\mathbf{X}$ ou $\mathrm{y}^{-1/2}\Delta\mathbf{X}$ como instrumentos.

Urga [1992] propõe que se use, em vez das primeiras diferenças, as "diferenças ortogonais" de Arellano e que consistem na transformação do modelo para desvios em relação à média dos valores futuros (em cada momento), como forma de remover os efeitos individuais sem comprometer a ortogonalidade dos termos de perturbação transformados.

Um estimador mais eficiente foi ainda proposto por Arellano e Bond [1991], sob a forma de um estimador IV generalizado (GIV) e que consiste na estimação por GLS de um modelo na forma (3.8) transformado pela pré-multiplicação por uma matriz $\mathbf{Z}$. Esta, é uma

---

[18] A sua multiplicação por $\mathbf{Y}$ dará a projecção deste no espaço definido por $\mathbf{Z}^*$, i.e., $\mathbf{Z}^*\mathbf{b}$ e $\mathbf{b}={}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^*\mathbf{Y}$.

matriz que reúne todos os instrumentos para cada valor desfasado de y_{it} ortogonais em relação a v_{it}, ou seja, $\mathbf{Z}$ =[$\mathbf{Z}'_1$,…, $\mathbf{Z}'_N$], que cumpre $E(\mathbf{Z}'_i\Delta\mathbf{v}_i)=0$[19],$\forall i$, com $\mathbf{v}'_i=(v_{i3}-v_{i2},\ldots, v_{iT}- v_{i,T-1})$. O estimador GIV de $\boldsymbol{\theta} = [\delta,\boldsymbol{\beta}]$ é

$$\hat{\boldsymbol{\theta}}_{GIV} = \left(\Delta\mathbf{X}'_*\mathbf{P}_Z\Delta\mathbf{X}_*\right)^{-1}\left(\Delta\mathbf{X}'_*\mathbf{P}_Z\Delta\mathbf{Y}\right) \tag{3.9}$$

em que $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}'\Psi^{-1}\mathbf{Z})\mathbf{Z}'$ e $\Delta\mathbf{X}^*=[\Delta\mathbf{Y}_{-1},\Delta\mathbf{X}]$. Este estimador será o mais eficiente desde que v_{it} não sejam autocorrelacionados.

Caso se não tenha qualquer informação quanto à natureza de y_{i0} e à distribuição de v_{it}, podemos recorrer a um estimador obtido pelo Método dos Momentos Generalizado (GMM) e que é idêntico a (3.9), com a diferença de $\mathbf{P_Z}$ ser agora $\mathbf{P_Z} = \mathbf{Z}\Gamma\mathbf{Z}'$ e, sendo ω_i o termo de perturbação do modelo nas primeiras diferenças, ou seja, o elemento genérico de $\Delta\mathbf{v}$, temos ainda $\Gamma = (\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_i\omega_i\omega'_i\mathbf{Z}_i)^{-1}$.

Em conclusão, ainda que nem sempre se possa comparar um estimador com outro, parece ser de recomendar o estimador IV de Balestra e Nerlove, desde que as variáveis **X** sejam estritamente exógenas e os termos de perturbação v_{it} não exibam autocorrelação. Já se **X** não for estritamente exógeno, ainda que os termos de perturbação permaneçam não autocorrelacionados, os estimadores de Arellano e Bond [1991] afiguram-se como os melhores candidatos. Se **X** não for estritamente exógeno e houver autocorrelação, Mátyás e Sevestre [1992] aconselham que se use como instrumento os valores desfasados de $\Delta\mathbf{X}$ no modelo (3.8).

O estimador sugerido em Arellano e Bond [1991], tal como a questão da filtragem com a matriz de variâncias, voltará a ser discutido no capítulo 5 relativo à estimação de modelos com dados em painel pelo método GMM.

3.3 Estimação de Modelos com Efeitos Aleatórios

Tal como no caso dos modelos estáticos, a consideração de efeitos individuais aleatórios em vez de fixos tem implicações distintas ao nível da estimação e propriedades dos estimadores. No entanto, no enquadramento específico dos modelos dinâmicos, a

[19] Por exemplo, para t=3, teríamos como instrumentos válidos y_{i0}, y_{i1}, y_{i2} e $\Delta\mathbf{x}'_i$, podendo a lista ser expandida com valores desfasados de $\Delta\mathbf{x}'_i$. A este propósito veja-se Mátyás e Sevestre[1992],100-5 e Baltagi[1995]128-

questão dos efeitos aleatórios (autocorrelacionados), ao implicar correlação entre os termos de perturbação e a variável autorregressiva tem consequências mais sérias do que nos casos precedentes.

Considere-se, então, o seguinte modelo de componentes de erro (unicamente individuais)

$$\begin{aligned} \mathbf{Y} &= \delta \mathbf{Y}_{-1} + \mathbf{X}\beta + \mathbf{u} \\ \mathbf{u} &= \alpha + \mathbf{v} \end{aligned} \tag{3.10}$$

em que E($\mathbf{u}$)=$\mathbf{0}$, Var($\mathbf{u}$) = $\mathbf{\Omega} = \sigma_v^2 \mathbf{W}_n + (\sigma_v^2 + T\sigma_\alpha^2)\mathbf{P} = \sigma_v^2(\mathbf{W}_n + 1/\phi^2 \mathbf{P})$ e $\mathbf{X}$ contém K variáveis exógenas.

É fácil ver que o estimador OLS (sem dar conta dos efeitos individuais), ao contrário do que sucedia nos modelos estáticos, não é cêntrico nem consistente, devido à correlação entre $y_{i,t-1}$ e α_i, podendo o enviesamento ser bastante significativo.

Uma abordagem bastante poderosa, proposta por Maddala, consiste em recorrer à classe geral de estimadores λ e que mais não são que a estimação do modelo (3.10) pré-multiplicado por ($\mathbf{W}_n + \sqrt{\lambda}\,\mathbf{P}$), o que resulta na seguinte formulação genérica[20]

$$\begin{bmatrix} \hat{\delta} \\ \hat{\beta} \end{bmatrix}_\lambda = \begin{bmatrix} \mathbf{Y}'_{-1}(\mathbf{W}_n + \lambda\mathbf{P})\mathbf{Y}_{-1} & \mathbf{Y}'_{-1}(\mathbf{W}_n + \lambda\mathbf{P})\mathbf{X} \\ \mathbf{X}'(\mathbf{W}_n + \lambda\mathbf{P})\mathbf{Y}_{-1} & \mathbf{X}'(\mathbf{W}_n + \lambda\mathbf{P})\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}'_{-1}(\mathbf{W}_n + \lambda\mathbf{P})\mathbf{Y} \\ \mathbf{X}'(\mathbf{W}_n + \lambda\mathbf{P})\mathbf{Y} \end{bmatrix} \tag{3.11}$$

Assim, para $\lambda \in [0;\infty[$ temos vários estimadores de $\boldsymbol{\theta}$, nomeadamente: o LSDV se $\lambda=0$; o OLS, quando $\lambda=1$; o GLS, para $\lambda=\phi^2$ e o estimador *Between* se $\lambda\to\infty$. Note-se que, quando $\delta = 0$, todos estes estimadores são consistentes, mas quase todos eles deixam de o ser, quando $\delta\neq 0$.

Para ilustrar este argumento, recorra-se, novamente, ao modelo autorregressivo (3.2), mas agora sem variáveis exógenas e com a inovação de os efeitos serem aleatórios.

$$\begin{aligned} y_{it} &= \delta y_{i,t-1} + \alpha_i + v_{it} \\ y_{it} &= \delta^t y_{i,0} + \frac{1-\delta^t}{1-\delta}\alpha_i + \sum_{j=0}^{t-1}\delta^j v_{i,t-j} \end{aligned} \tag{3.12}$$

Uma estimação por OLS do modelo (3.12) leva ao seguinte estimador:

32.

[20] Atendendo a que ($\mathbf{W}_n + \sqrt{\lambda}\,\mathbf{P})'(\mathbf{W}_n + \sqrt{\lambda}\,\mathbf{P}) = (\mathbf{W}_n + \lambda\,\mathbf{P}$).

$$\hat{\delta}_{OLS} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\, y_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^{2}} = \delta + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} (\alpha_i + v_{it})\, y_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^{2}}$$

verificando-se, de acordo com resultado apresentado em Hsiao [1986], 77

$$\operatorname*{plim}_{N\to\infty}(\hat{\delta}_{OLS} - \delta) = \operatorname*{plim}_{N\to\infty} \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} (\alpha_i + v_{it})\, y_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^{2}} > 0$$

sendo que a inclusão de variáveis exógenas reduz, mas não elimina o enviesamento, pois o estimador OLS de δ continua a sobrestimar o seu verdadeiro valor e o de β encontra-se enviesado para zero. Uma forma de determinar um estimador consistente para δ foi desenvolvida por Sevestre e Trognon, em Mátyás *et al.* [1992] e resulta num valor para λ dado por

$$\lambda^{*} = \frac{K(1-\phi^{2})}{\left(\dfrac{1-\delta^{T}}{1-\delta}\dfrac{E(y_{i0}\alpha_i)}{\sigma_{\alpha}^{2}+\sigma_{v}^{2}} + K(1-\phi^{2}+T\phi^{2}\right)}$$

$$K = \frac{(T-1-T\delta+\delta^{T})}{T(1-\delta)^{2}}$$

$$\phi^{2} = \frac{\sigma_{\alpha}^{2}}{\sigma_{\alpha}^{2}+\sigma_{v}^{2}}$$

Ou seja, se substituirmos λ por λ^* em (3.11), temos estimadores de δ e β que serão consistentes. Regra geral, $\lambda^* \neq \phi^2$, o que prova a semi-inconsistência do GLS[22], excepto no caso de E($y_{i0}\alpha_i$) = 0, pois então, $\lambda^* = \phi^2$ e os estimadores serão consistentes. Infelizmente, o método não é exequível pois λ^* é função de parâmetros desconhecidos. Sugere-se um método bi-etápico em que se estimaria, numa primeira fase, λ^* de forma consistente e utilizar-se-ia essa estimativa em (3.11), mas verificou-se que as distribuições assimptótica e finita deste estimador bi-etápico dependem crucialmente da do estimador de λ^*, o que leva a que a estimação seja, geralmente, pouco robusta. Finalmente, este princípio não é aplicável a modelos autorregressivos de ordem superior a um.

---

[21] Ver Mátyás e Sevestre [1992].

Quando o vector de perturbações é normal, a solução natural para o problema de estimação é o recurso ao MLE que, desde que nada seja assumido quanto à primeira observação, será regra geral equivalente ao OLS e, como este, não consistente.

A consistência do MLE (e também do GLS, como vimos) vai depender de forma determinante dos pressupostos admitidos quanto a y_{i0}, pelo que podemos considerar quatro casos[23] para o modelo

$$y_{it} = \delta y_{i,t-1} + \beta x_{it} + \gamma z_i + \alpha_i + v_{it} \tag{3.13}$$

em que z é um atributo que não varia no tempo (como o sexo) e x varia seccional e temporal. Temos, de acordo com Hsiao [1986]:

I. y_{i0} fixo: um indivíduo começa num dado ponto arbitrário e fixo e gradualmente aproxima-se do nível de equilíbrio dado por $(\alpha_i+\gamma z_i)/(1-\delta)+\beta\times\sum_j\gamma^j x_{i,t-j}$. Ainda que pareça razoável, o pressuposto da não aleatoriedade de y_{i0} choca com a correlação entre α_i e y_{it}: se α_i está incorporado em y_{it}, porque não há de estar em y_{i0}? Se o processo já tem alguma história, nada justifica que y_{i0} tenha sido gerado de forma diferente da de y_{it};

II. y_{i0} aleatório e $y_{i0} = \alpha_{y0} + \varepsilon_i$, em que os ε_i representam os efeitos das dotações individuais iniciais, podendo-se verificar que y_{io} seja

   a) uma variável aleatória pura, em que os impactos das dotações iniciais se vão esbatendo ao longo do tempo, logo, $\mathrm{cov}(\alpha_i,\varepsilon_i)= 0$;

   b) correlacionado com α_i, com $\mathrm{cov}(y_{i0},\alpha_i)= \varphi\sigma^2_{y0}$, de tal forma que as diferenças entre os indivíduos ficam a dever-se unicamente às diferentes dotações iniciais, sendo o seu efeito de longo prazo $\varphi\varepsilon_i/(1-\delta)$;

III. $y_{i0} = \omega_{i0} + \eta_i$, com ω_{i0} fixo e $y_{it} = \omega_{it} + \eta_i$ e $\omega_{it}=\rho\,\omega_{it-1} + u_{it}$. Ou seja, ω_{it} e η_i não estão correlacionados e ω_{it} é uma variável latente não observável que representa um dado processo dinâmico. Os indivíduos seguem um mesmo processo estocástico $\{\omega_{it}\}$ (condicionado pelas variáveis exógenas z e x) não observável e são sujeitos a choques individuais independentes (uma medida das dotações iniciais ou dos erros de medida no i-ésimo processo) η_i. Aqui, $y_{i0} = \omega_{i0} + \eta_i$ (i.e., também é afectado por

---

[22] Quando $T\to\infty$ e $N\to\infty$, o GLS é consistente pois converge para o LSDV (Hsiao[1986],88).
[23] Para uma abordagem mais completa, veja-se Hsiao[1986],78-81.

η_i, pelo que não é forçoso que y_{i0} seja o início do processo) será um ponto de partida arbitrário e y_{it} tenderá para $\eta_i+\gamma z_i/(1-\delta)+\beta\times\sum_j \gamma^j x_{i,t-j}$.

IV. $y_{i0} = \omega_{i0} + \eta_i$, com ω_{i0} aleatório e $\mathrm{cov}(y_{i0},\eta_i) \neq 0$, com as seguintes variações:

a) ω_{i0} tem as mesmas média e variância que os estados finais;

b) ω_{i0} tem a mesma média, mas variância arbitrária (não estacionário);

c) ω_{i0} tem a mesma variância, mas média diferente (i.e., é de outra população);

d) ω_{i0} tem média e variância arbitrárias.

Para os casos III e IV.d os estimadores MLE não se encontram definidos e, quando $T\rightarrow\infty$ e N fixo, os estimadores de γ e σ_α^2 não são consistentes devido à insuficiente variabilidade seccional. Os restantes estimadores MLE são, regra geral, consistentes, em qualquer dos casos semi-assimptóticos[24] e a utilização de um método MLE justifica-se quando se pretende testar hipóteses quanto ao estado inicial y_{i0}.

A estimação por MLE, maximizando a função de verosimilhança, é relativamente complicada e pode ser efectuada através de um qualquer algoritmo iterativo, como o de Newton-Raphson ou, alternativamente, através de um sistema de equações

$$\Gamma\mathbf{Y}' + \mathrm{B}\mathbf{X}' + \Phi\mathbf{Z} = \mathbf{U}'$$

$$\mathbf{Y} = \begin{bmatrix} y_{10} & y_{11} & & y_{1T} \\ y_{20} & y_{21} & & y_{2T} \\ & & & \\ y_{N0} & y_{N1} & & y_{NT} \end{bmatrix};\ \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & & x_{1T} \\ x_{21} & x_{22} & & x_{2T} \\ & & & \\ x_{N1} & x_{N2} & & x_{NT} \end{bmatrix};\ \mathbf{Z} = \begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ & \\ 1 & z_N \end{bmatrix} \qquad (3.14)$$

em que **U** é a matriz $N\times T$ de termos de perturbação, com matriz de variâncias $\mathbf{\Omega}$. A matriz $\mathbf{A}=[\mathbf{\Gamma\ B\ \Phi}]$, tal como $\mathbf{\Omega}$, poderá ser objecto de determinadas restrições, dependentes do caso que se escolher quanto a y_{i0}. A utilização de um método de mínimos quadrados tri-etápico (3SLS) deverá ser equivalente à utilização de um método MLE sujeito às referidas restrições, mas será menos eficiente que a maximização da função de verosimilhança irrestrita de (3.14).

Uma especificação proposta em Mátyás *et al* [1992] para a primeira observação consiste em acrescentar a (3.13) o seguinte:

---

[24] Para um ponto de situação, consulte-se Hsiao[1986],84.

$$\begin{cases} y_{i0} = \varphi . z_i + u_{i0} & (3.15.1) \\ \alpha_i = \psi . u_{i0} + \varepsilon_i & (3.15.2) \end{cases}$$

As perturbações ($u_{i0}, \varepsilon_i, v_{i1}, \ldots, v_{iT}$) são i.i.d. normais, com média nula e variâncias σ_u^2 σ_ε^2 σ_v^2 e a função logarítmica de verosimilhança é

$$\ln L_{NT}(\delta, \beta, \gamma, \varphi, \psi, \sigma_u^2, \sigma_\varepsilon^2, \sigma_v^2) = C - \frac{N}{2}\ln|\Omega| - \frac{N}{2}\ln \sigma_u^2 - \frac{1}{2}\sum \tau_i' \Omega^{-1} \tau_i - \frac{1}{2\sigma_u^2}\sum_i u_{i0}^2 \qquad (3.16)$$

com $\mathbf{\Omega} = \sigma_v^2 \mathbf{W}_n + (\sigma_v^2 + T\sigma_\varepsilon^2)\mathbf{P}$ e τ_i um vector coluna ($T \times 1$) de elemento genérico (y_{it} - $\delta y_{i,t-1}$ - βx_{it} - γz_i - φ u_{i0}). A maximização de (3.16) com respeito aos parâmetros desconhecidos dará os estimadores MLE[25], com a particularidade de os estimadores de φ e σ_u^2 serem os estimadores OLS de (3.15.1) e de u_{i0} ser substituído pelos resíduos desta última regressão em τ_i. Os estimadores assim obtidos serão consistentes e, no caso de as perturbações serem normais, assimptoticamente eficientes, exibindo ainda um bom comportamento em amostras finitas.

Atendendo a que uma má escolha da especificação das condições iniciais poderá levar a estimadores MLE que não sejam consistentes, por não serem os mais adequados (não sendo assimptoticamente equivalentes aos verdadeiros MLE, por estarem em causa funções de verosimilhança diferentes) e porque a regra geral dos estudos é que se disponha de séries temporais relativamente curtas, torna-se imperioso que se encontre um estimador que não dependa de y_{i0}.

Os diversos métodos IV, já analisados na secção anterior, valem pelo facto de serem geralmente consistentes (desde que $N \rightarrow \infty$), independentemente de y_{i0} (ainda que pouco se possa afirmar quanto às suas eficiências relativas) e pela sua grande aplicabilidade, não havendo grandes diferenças relativamente ao modelo com efeitos fixos. Por isso, acabam por ser os mais utilizados. Para excelentes resumos de vários estimadores podem consultar-se Baltagi [1995], 126-145, Mátyás e Sevestre [1992] e Mátyás [1999].

Quando se não impõe qualquer restrição quanto à estrutura de correlações entre os termos de perturbação, um método eficiente e consistente (com T fixo) consiste em separar o modelo (3.13) nas primeiras diferenças em T-1 equações

---

[25] Ver Mátyás e Sevestre[1992], 111 para as equações normais da condição de primeira ordem.

$$\begin{cases} (y_{i2}-y_{i1})=\delta(y_{i1}-y_{i0})+\beta(x_{i2}-x_{i1})+(u_{i2}-u_{i1}) \\ (y_{i3}-y_{i2})=\delta(y_{i2}-y_{i1})+\beta(x_{i3}-x_{i2})+(u_{i3}-u_{i2}) \\ \quad\vdots \\ (y_{iT}-y_{i,T-1})=\delta(y_{i,T-1}-y_{i,T-2})+\beta(x_{iT}-x_{i,T-1})+(u_{iT}-u_{i,T-1}) \end{cases} \quad ;i=1,\ldots,N$$

e estimar o sistema por 3SLS. Como os efeitos individuais desaparecem do modelo, é indiferente estarmos perante efeitos fixos ou aleatórios. No entanto, o método exige que existam variáveis exógenas não correlacionadas com o termo de perturbação v_{it}.

Também o método de Chamberlain [1984], com recurso a um estimador de distância mínima, que vimos que era eficiente nos modelos sem restrições quanto à estrutura da matriz de variâncias e covariâncias dos termos de perturbação, pode ser aplicado no contexto dos modelos dinâmicos. O método consiste em restringir $\mathbf{\Pi}$**,** para o modelo (3.2) com apenas uma variável exógena, da seguinte forma:

$$E^*(y_{it}\,|\,y_{i0},x_{i1},\ldots,x_{iT})=\delta^t y_{i0}+\beta\sum_{j=0}^{t-1}\delta^j x_{i,t-j}+\frac{1-\delta^t}{1-\delta}\frac{E(\alpha_i y_{i0})}{Var(y_{i0})}(y_{i0}-E(y_{i0}))\Rightarrow$$

$$\begin{aligned} &\pi_{0,T}=\delta^t+\frac{1-\delta^t}{1-\delta}\frac{E(\alpha_i y_{i0})}{Var(y_{i0})} \\ &\pi_{j,t}=\beta\delta^{t-j} \quad ;\ j=1,\ldots,T \\ &\pi_{j,t}=0 \quad ;\ j=t+1,\ldots,T \\ &\pi_{T+1,t}=-\frac{1-\delta^t}{1-\delta}\frac{E(\alpha_i y_{i0})}{Var(y_{i0})}E(y_{i0}) \end{aligned} \tag{3.17}$$

em que E*($y_{it}|y_{i0},\mathbf{x}'_i$) é a projecção de y_{it}. Tal como na secção 2.3, procede-se à estimação de $\boldsymbol{\pi}$ (coluna com todas as linhas de $\mathbf{\Pi}$) na sua forma irrestrita através de uma regressão OLS de y_{it} sobre $y_{i0},\mathbf{x}'_i$. O segundo passo do método consiste em estimar através de um estimador de distância mínima os parâmetros π sujeitos às restrições expostas em (3.17)[26] e que se representam na forma matricial por $\boldsymbol{\pi}=\boldsymbol{\Gamma}(\boldsymbol{\pi})\boldsymbol{\theta}$. Assim, substituir-se-á $\boldsymbol{\pi}$ pela estimativa OLS irrestrita, pelo que se tem

$$\hat{\boldsymbol{\pi}}_{OLS}=\boldsymbol{\Gamma}(\hat{\boldsymbol{\pi}}_{OLS})\boldsymbol{\theta}+\boldsymbol{\varepsilon}\Rightarrow\hat{\boldsymbol{\theta}}_{ALS}=\left(\boldsymbol{\Gamma}(\hat{\boldsymbol{\pi}}_{OLS})'\boldsymbol{\Gamma}(\hat{\boldsymbol{\pi}}_{OLS})\right)^{-1}\boldsymbol{\Gamma}(\hat{\boldsymbol{\pi}}_{OLS})'\hat{\boldsymbol{\pi}}_{OLS}$$

em que o estimador de $\boldsymbol{\theta}$ assim obtido corresponde a um estimador de mínimos quadrados assimptóticos (ALS) e é consistente, ainda que não eficiente. Um estimador assimptoticamente eficiente teria que considerar a matriz de covariâncias de $\boldsymbol{\varepsilon}$, $\boldsymbol{\Sigma}(\mathbf{S},\boldsymbol{\theta})$,

[26] As restrições poderão e deverão ser simplificadas, de acordo com Mátyás e Sevestre [1992],113, por forma

função da matriz de covariâncias de $\hat{\boldsymbol{\pi}}_{OLS}$, **S**, e de $\boldsymbol{\theta}$. Assim, teremos um estimador FGLS mais eficiente se usarmos as estimativas OLS de **S** e ALS de $\boldsymbol{\theta}$ para estimar $\boldsymbol{\Sigma}$, com a seguinte formulação:

$$\hat{\boldsymbol{\theta}}_{FGLS} = \left(\Gamma(\hat{\boldsymbol{\pi}}_{OLS})'\hat{\Sigma}^{-1}\Gamma(\hat{\boldsymbol{\pi}}_{OLS})\right)^{-1}\Gamma(\hat{\boldsymbol{\pi}}_{OLS})'\hat{\Sigma}^{-1}\hat{\boldsymbol{\pi}}_{OLS}\,.$$

As vantagens deste método de Chamberlain [1984] prendem-se com o facto de necessitar de poucas restrições, nomeadamente quanto à dinâmica das perturbações, de poder ser aplicado a modelos com variáveis autorregressivas de ordem p>1 e de facilitar a realização de numerosos testes de especificação. No entanto, ao deixar a matriz de covariâncias de **v** irrestrita, o método leva a uma perda de eficiência, quando esta tiver uma estrutura conhecida. Por outro lado, trata-se de um método computacionalmente muito exigente, pelo que, pouco usado, ainda que esta questão seja cada vez menos relevante.

3.4 Estruturas Alternativas para a Heterogeneidade

Uma especificação alternativa aos modelos com componentes de erro é a proposta por Kmenta[27] (Mátyás e Sevestre [1992], 198-201 e Greene [1997], cap. 15) em que a heterogeneidade individual deriva, não do termo independente ou de coeficientes aleatórios, mas sim de uma particular estrutura de covariâncias para um conjunto de amostras temporais, naquilo que usualmente se considera ser um modelo da classe *time-series of cross section data* (TSCS). Este tipo de especificação é, segundo Greene [1997], particularmente indicada para estudos com painéis relativamente longos e estreitos, como é habitual surgir na macroeconometria. Admita-se, a título de exemplo e para facilitar a exposição, um estudo sobre a dinâmica da inflação para um conjunto de N países, durante T períodos e que pode ser representado da seguinte forma:

$$\begin{aligned} p_{i,t} &= b_0 + \sum_{j=1}^{p} b_j p_{i,t-j} + e_{it} \\ V = E(ee') &= \begin{bmatrix} s_{11}\Omega_{11} & s_{12}\Omega_{12} & s_{1n}\Omega_{1n} \\ s_{21}\Omega_{21} & s_{22}\Omega_{22} & s_{2n}\Omega_{2n} \\ & & \\ s_{n1}\Omega_{n1} & s_{n2}\Omega_{n2} & s_{nn}\Omega_{nn} \end{bmatrix} \end{aligned} \qquad (3.18).$$

---

a que se tornem lineares.

[27] Kmenta, Jan [1986], *Elements of Econometrics*, New York, MacMillan, citado em Baltagi *in* Mátyás e

A estrutura de covariâncias pode ser condicionada de forma a representar diversos casos de interesse, sendo o mais simples obtido para termos de perturbação homoscedásticos e não correlacionados:

$$V = \begin{bmatrix} \sigma^2\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I} & \mathbf{0} \\ & & \\ \mathbf{0} & \mathbf{0} & \sigma^2\mathbf{I} \end{bmatrix} \Leftrightarrow E(\varepsilon_{it}^2)=\sigma^2,\ \forall\ i,t \text{ e } E(\varepsilon_{it}\varepsilon_{js})=0,\ \forall\ i \neq j \text{ ou } t \neq s$$

e que equivale à estimação de (3.18) admitindo a inexistência de efeitos específicos (em que os estimadores OLS serão BLUE). Este caso é, no entanto, de longe o menos interessante de analisar na medida em que exclui toda heterogeneidade dos dados e a única vantagem de se trabalhar com um painel será a de reunir mais observações, aumentando os graus de liberdade ao dispor da inferência estatística (na prática faz-se o *pooling* dos dados e ignora-se a dimensão temporal ou seccional da heterogeneidade).

Um caso bem mais interessante será o de incluir heteroscedasticidade seccional ou em grupo no modelo, i.e., admitir homoscedasticidade dos termos de perturbação para um dado indivíduo ao longo do tempo mas que a cada indivíduo corresponde uma variância diferente. Ou seja, admite-se que:

$$V = \begin{bmatrix} \sigma_1^2\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I} & \mathbf{0} \\ & & \\ \mathbf{0} & \mathbf{0} & \sigma_n^2\mathbf{I} \end{bmatrix}$$

o que se pode ficar a dever a uma grande variação nos níveis médios de inflação entre os diversos países, em que alguns deles registaram episódios de hiperinflação mesmo. Esta hipótese pode ser introduzida no modelo como fonte adicional de heterogeneidade, para além dos efeitos fixos. Nesta situação, torna-se necessária a estimação da matriz de variâncias e covariâncias, pelo estimador de White com correcção por grupo, para uma estimação robusta, ainda que não eficiente, por OLS. Uma alternativa mais eficiente será a estimação do modelo através de um estimador GLS se **V** for conhecida ou FGLS, caso **V** seja desconhecida.

Finalmente, uma outra hipótese de heterogeneidade será a resultante de uma estrutura de correlações entre as séries individuais (mas não dentro de cada uma das séries), a que se chama de correlação seccional (Greene [1997], pág. 658) e que, neste

---

Sevestre [1992].

caso, pode representar a existência de relações económicas estreitas entre dois ou mais países da amostra que originam alguma simetria na resposta a certos choques aleatórios sobre as respectivas dinâmicas inflacionárias. Neste caso, a matriz de covariâncias será

$$\mathbf{V} = E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} & \cdots & \sigma_{1n}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} & \cdots & \sigma_{2n}\mathbf{I} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1}\mathbf{I} & \sigma_{n2}\mathbf{I} & \cdots & \sigma_{nn}\mathbf{I} \end{bmatrix}$$

em que se admite a ausência de autocorrelação entre os termos de perturbação[28]. O pressuposto de ausência de autocorrelação torna-se essencial para garantir a consistência dos estimadores de mínimos quadrados (ordinários ou generalizados) pelo que a sua não verificação levaria a uma estimação por IV ou GMM por os regressores deixarem de ser ortogonais, face aos termos de perturbação. Em Davidson e Mackinnon [1993] e Greene [1997] sugere-se o método mais eficiente de Hatanaka que avança como instrumento para a variável dependente desfasada, os valores desfasados das restantes variáveis exógenas para a obtenção de estimadores consistentes para β e ρ e a implementação de um procedimento FGLS. Para os modelos puramente autorregressivos, tal torna-se impraticável na medida em que não existirão variáveis verdadeiramente exógenas, o que, segundo Hamilton [1994], faz com que um método FGLS iterado não corresponda já a MLE[29].

Tal como no caso dos modelos com componentes de erro, os modelos de Kmenta têm a vantagem de trabalharem com amostras relativamente grandes e com bastante heterogeneidade e ao permitirem uma correlação serial, encontram uma fundamentação macroeconómica pertinente. No entanto, Baltagi em Mátyás e Sevestre [1992] alerta para o facto de este tipo de especificação ser menos robusto face a eventuais erros de especificação que a usual abordagem dos componentes de erro.

---

[28] Por uma questão de constância na terminologia, estabelece-se que um termo de perturbação se diz autocorrelacionado quando $E(u_{it}.u_{i,t-s}) \neq 0,\ s \neq 0$ e serialmente correlacionada quando $E(u_{it}.u_{jt}) \neq 0,\ i \neq j$.

[29] Uma alternativa apontada por Hamilton [1994] seria a estimação por MLE o que exigiria um conhecimento, que se poderá não ter, sobre as distribuições dos termos de perturbação e sobre a especificação do processo estocástico gerador da autocorrelação. Mesmo se se conhecer a estrutura ARMA dos termos de perturbação, caso esta não seja suficientemente simples, a aplicação de um método MLE ou FGLS tornar-se-á impraticável do ponto de vista computacional.

4. Estimação de Modelos com Desfasamentos Distribuídos[30]

Devido à existência de rigidez de ordem técnica, institucional ou psicológica, é frequente o comportamento não se adaptar imediatamente a alterações nas variáveis explicativas, sendo tal adaptação, na maior parte dos casos, progressiva. Assim, essa adaptação pode ser racionalizada através de diversos mecanismos teóricos de ajustamento, havendo a alternativa de estimarmos y através de variáveis dependentes desfasadas, juntamente com variáveis exógenas, ou através de um modelo de desfasamentos distribuídos, em que o valor actual de y será uma função de uma cadeia de valores contemporâneos e passados de um vector de variáveis independentes. Ainda que estes dois modelos possam ser reciprocamente transformados um no outro, é sempre necessário determinar com algum rigor, não apenas o momento a partir do qual se começam a sentir os efeitos na variável dependente de uma inovação no vector de variáveis independentes, mas também determinar a partir de que momento se encontram completamente repercutidos em y os referidos efeitos.

A determinação da dimensão da estrutura de desfasamento, ainda que um tema merecedor de atenção, não é propriamente uma idiossincrasia do estudo de dados em painel, aplicando-se os usuais critérios utilizados nas séries temporais ou nos estudos *cross-section*: critérios de Schwarz ou Akaike, testes LR, ...

Já a questão da estimação de um modelo com uma dada dimensão de desfasamento sem impor restrições de identificação, como os desfasamentos geométricos de Koyck ou os polinomiais de Almon, que dificilmente poderá ser evitada nos estudos temporais com desfasamentos importantes, devido ao excessivo número de parâmetros a estimar (e que inviabiliza um qualquer estimador consistente, mesmo com $T \rightarrow \infty$, pois o número de parâmetros a estimar tende também para infinito) justifica uma referência detalhada nos estudos com dados em painel.

Assim, uma vez que as amostras temporais são frequentemente de reduzida dimensão e dizem respeito a variáveis provavelmente bastante correlacionadas umas com as outras, é geralmente impossível obter estimativas, com qualidade, para os parâmetros dos diversos desfasamentos sem impor as referidas restrições. Mas, se dispusermos de N

---

[30]Esta secção foi quase exclusivamente baseada em Hsiao[1986], Secção 8.1, sujeita a algumas simplificações e adaptações.

séries temporais, em princípio será mais fácil estimar pelo menos alguns dos parâmetros, uma vez que passamos a dispor de N vezes mais informação. Por outro lado, ao juntarmos a dimensão seccional, permitindo diferenças nas características individuais, estamos muito provavelmente a reduzir a importância do mais que provável problema de multicolinearidade entre x_t, x_{t-1} ... x_{t-p} ..., sempre presente neste tipo de especificação.

Assim, Hsiao [1986] sugere que para a estimação do seguinte modelo

$$y_t = \mu + \sum_{\tau=0}^{\infty} \beta_\tau x_{t-\tau} + u_t \ , \ t = 1,\dots,T \qquad (4.1)$$

se recorra ao modelo transformado

$$y_{it} = \mu + \sum_{\tau=0}^{t-1} \beta_\tau x_{i,t-\tau} + b_{it} + u_{it} \ , \ t = 1,\dots,T; \ i = 1,\dots,N \qquad (4.2)$$

onde $b_{it} = \sum_{\tau=0}^{\infty} \beta_{t+\tau} x_{i,-\tau}$ será a contribuição dos valores pré-amostrais não observados de x.

Para a estimação do modelo são necessários alguns pressupostos, nomeadamente:

0. o modelo é dado por

$$y_{it} = \alpha_i^* + \sum_{\tau=0}^{\infty} \beta_\tau x_{i,t-\tau} + u_{it} \ , \ t = 1,\dots,T; \ i = 1,\dots,N,$$

com u_{it} independente de x_{is} e i.i.d., com valor esperado nulo e variância σ_α;

1. $E(\beta_{\tau i}) = \beta_\tau$;
2. $\tilde{\beta}_{\tau i} = \beta_{\tau i} - \beta_\tau$, $\xi_{it} = \sum_{\tau=0}^{\infty} \tilde{\beta}_{\tau i} x_{i,t-\tau}$ e $\xi_i' = (\xi_{i1},\dots,\xi_{iT}) \Rightarrow E^*(\xi_i|\mathbf{x}_i) = 0$;
3. $E^*(\alpha_i^*|\mathbf{x}_i) = \mu + \mathbf{a}'\mathbf{x}_i$.

em que E^* representa o previsor de erro quadrático mínimo da primeira variável na segunda (ou projector) e $\mathbf{x}_i$ o vector de todas as observações de x_{it}. Admite-se ainda que existem p + 1 observações para antes da primeira observação disponível para y e que $E(\mathbf{x}_i'\mathbf{x}_i) = \Sigma_{xx}$ positiva definida, com $\mathbf{x}_i = [x_{i,-p},\dots,x_{i,T}]$ a ser um vector aleatório de dimensão 1 x (p + 1 + T). O segundo pressuposto pode ser aproximado pela hipótese de as diferenças entre os coeficientes individuais $\beta_{\tau I}$ serem independentes de $\mathbf{x}_i$, ainda que os efeitos individuais α_i^* possam ser correlacionadas com este. De acordo com Hsiao [1986], os pressupostos 1-3 garantem a identificação de E({β_τ}) quando N$\rightarrow\infty$ e T$\rightarrow\infty$. Se T é fixo, o modelo deverá ser reescrito da seguinte forma:

$$y_{it} = \alpha_i^* + \sum_{\tau=0}^{t+p} \beta_\tau x_{i,t-\tau} + b_{it} + u_{it} \ , \ t = 1,\ldots,T; \ i = 1,\ldots,N \qquad (4.3)$$

com o termo de perturbação agregado $\hat{u}_{it}$ = ξ_{it} + u_{it} satisfazendo E($\hat{\mathbf{u}}_i|\mathbf{x}_i$) = 0 e o termo dos coeficientes individuais residuais (i.e., truncados) a ser dado por $b_{it} = \sum_{\tau=p+1}^{\infty} \beta_{t+\tau} x_{i,-\tau}$.

Na generalidade dos casos, b_{it} estará correlacionado com as variáveis explicativas, pelo que não será possível obter estimadores consistentes, a menos que se introduzam restrições adicionais sobre os coeficientes de desfasamento ou sobre o processo estocástico gerador dos valores de *x*, válido não apenas para os valores amostrais, mas também para os pré-amostrais.

Podemos, então, distinguir duas estratégias de identificação que podem ser resumidas da seguinte forma:

i. <u>usando a estrutura do processo gerador da variável exógena</u>

Como pretendemos estimar pelo menos alguns dos parâmetros β_τ, com τ = 1, ... e de forma irrestrita (i.e., sem necessitarmos de racionalizarmos {β_τ} através de um qualquer método de Koyck ou de Almon), vamos considerar que um coeficiente de desfasamento β_τ se encontra identificado se for possível calculá-lo através da matriz de coeficientes $\mathbf{\Pi}$ (de dimensão T x (T + *p* + 1)) obtida através do previsor de erro quadrático médio mínimo ou projecção de $\mathbf{y}_i$ em $\mathbf{x}_i$: $E^*(\mathbf{y}_i|\ \mathbf{x}_i) = \mathbf{I}\mu + \mathbf{\Pi}\mathbf{x}_i$, em que $\mathbf{y}_i$ é o vector das T observações de *y* para o i-ésimo indivíduo. Esta matriz $\mathbf{\Pi}$ contém informação sobre a combinação dos coeficientes de desfasamento relevantes **B** e sobre as projecções de α_i^* e b_{it} em $\mathbf{x}_i$, que terá que ser separada para se proceder à identificação dos referidos coeficientes.

Uma forma simples de o fazer será reescrever a projecção de $\mathbf{y}_i$ em $\mathbf{x}_i$ e α_i^* da seguinte forma:

$$E^*(\mathbf{y}_i|\mathbf{x}_i,\alpha_i^*) = [\mathbf{B}+\mathbf{W}]\mathbf{x}_i + [\mathbf{i}+\mathbf{c}]\alpha_i^* \qquad (4.4)$$

em que **B** será a matriz T x (T + *p* + 1) que conterá T + *p* coeficientes β e **W** e **c** serão definidos através de

$$E^*(\mathbf{b}_i|\mathbf{x}_i,\alpha_i^*) = \mathbf{W}\mathbf{x}_i + \mathbf{c}\alpha_i^* \qquad (4.5)$$

Ora, Π será então definida por $\Pi = \mathbf{B} + [\mathbf{W} + (\mathbf{I} + \mathbf{c})\mathbf{a}']$[31], com $\mathbf{a}'$ obtida através de $E^*(\alpha_i^*|\mathbf{x}_i) = \mu + \mathbf{a}'\mathbf{x}_i$. Hsiao[1986] demonstra que para se identificar $\mathbf{B}$ bastam apenas algumas restrições sobre $\mathbf{W}$, sugerindo, para α_i^* independente de $\mathbf{x}_i$, que se admita um processo autorregressivo de ordem q para x, na medida em que é possível obterem-se estimadores consistentes para os (T + p - q + 1) primeiros parâmetros de desfasamento, com expressão genérica

$$E^*(y_{it}|\mathbf{x}_i) = \mu + \sum_{\tau=0}^{t+p-q} \beta_\tau x_{t-\tau} + \sum_{j=1}^{q} \phi_{t,j-p-1} x_{i,j-p-1} \tag{4.6}$$

com $\phi_{t,\,j-p-1} = \beta_{t+p+1-j} + w_{t,\,j-p-1}$, t = 1, ..., T e j = 1, ..., q e $q = p + 2$ (por simplificação). Os coeficientes β de cada uma das equações são estimativas dos primeiros ($t + p - q + 1$) coeficientes de desfasamento do modelo, desde que t seja suficientemente grande, i.e., desde que $(t + p - q + 1) > 0$.

Ou seja, desde que se assuma 1., 2. e 3., juntamente com um dado comportamento para os valores pré-amostrais de x, neste caso, que tenham sido gerados por um processo AR(q)[32] e que se disponha de um número suficiente de observações (i.e., que T + p seja grande face a q), será sempre possível estimar os primeiros coeficientes de desfasamento. É evidente que, dada a pequena dimensão temporal dos estudos em painel, não será possível obter muita informação (irrestrita) sobre os coeficientes mais afastados, mas será sempre possível obter estimativas para os primeiros e, a partir daí, desenhar uma dada estrutura de desfasamento que restrinja o número de parâmetros a estimar.

ii. <u>usando restrições sobre a estrutura dos coeficientes de desfasamento</u>

Sempre que se disponha de alguma informação teórica sobre que valores os parâmetros β_τ podem ou não assumir, ou sobre as suas ordens de grandeza, etc..., pode-se reduzir o número de parâmetros a estimar, numa generalização e extensão das técnicas utilizadas nos modelos de desfasamentos distribuídos convencionais, recorrendo-se às potencialidades dos estudos em painel. Assim, em vez de assumir um dado processo gerador para x (como o AR(q) usado na secção anterior) vamos assumir que os parâmetros β_τ se comportam de uma dada forma conhecida *a priori*.

---

[31] Para mais pormenores, consultar Hsiao[1986], 185.

[32] Hsiao[1986] fornece uma fórmula mais geral (p.186), que contempla o caso de α_i^* e $\mathbf{x}_i$ serem correlacionados, mas que por ser notacionalmente bastante complexa, foi aqui omitida.

Uma forma sugerida por Hsiao [1986] é considerar-se que há k parâmetros β iniciais irrestritos e que os restantes seguem um comportamento de natureza autorregressiva, nomeadamente:

$$\begin{cases} \beta_\tau \Leftarrow \tau \leq k \\ \sum_{j=1}^{J} \delta_j \beta_{\tau - j} \Leftarrow \tau > k \end{cases}$$

em que se assegura que β_τ decline geometricamente após os primeiros k desfasamentos[33]. A solução geral da equação às diferenças

$$\beta_\tau - \delta_1 \beta_{-1} - \quad - \delta_J \beta_{\tau - J} = 0$$

é dada por $\beta_\tau = \sum_{j=1}^{J} A_j \lambda_j^\tau$, em que A_j são constantes dadas pelas condições iniciais da equação e $\{\lambda_j\}$ as suas raízes características, e permite-nos escrever o termo residual b_{it} da seguinte forma:

$$b_{it} = \sum_{\tau = p+1}^{\infty} \left(\sum_{j=1}^{J} A_j \lambda_j^{t+\tau}\right) x_{i,-\tau} = \sum_{j=1}^{J} \lambda_j^{t} b_{ij} \,.$$

Ou seja, podemos representar o termo residual b_{it} como uma combinação linear das condições iniciais não observáveis (b_{i1}, ..., b_{iJ}), pelo que, se juntarmos este pressuposto aos três já estabelecidos, o modelo de desfasamentos distribuídos reduz-se a um sistema de T regressões com J+1 coeficientes não observados (α_i^*; b_{i1}, ..., b_{iJ}) e em que os últimos decaem geometricamente. No caso particular de J=1, temos o modelo de Koyck, em que $b_{it} = \delta\, b_{i,t-1}$ que, após algumas transformações (que aqui se omitem para não tornar a exposição demasiado pesada, mas que podem ser consultados em Hsiao[1986], 190-1), pode ser identificado a partir de $\mathbf{\Pi} = \mathbf{B}^* + \delta^*\mathbf{w}^* + \mathbf{ia}'$, em que $\mathbf{B}^*$ é a matriz (T×(T+k)) que contém os k +1 parâmetros β (de β_0 a β_k) irrestritos e os restantes T – 1 parâmetros obtidos a partir de $\beta_k \times \delta^\tau$, τ = 1, ..., T-1. A matriz δ^* é $[1,\delta,...,\delta^{T-1}]$ e $\mathbf{w}^*$ o vector de coeficientes obtido da projecção de b_i em $\mathbf{x}_i$ e exige-se que T ≥ 3 por forma a que $\mathbf{\Pi}$ identifique os parâmetros β_τ e δ[34].

---

[33] Nomeadamente, impondo que as raízes, reais e distintas, da equação característica $1 - \sum_{j=1}^{J} \delta_j L^j = 0$ caiam fora do círculo unitário. Ver Hsiao[1986], 189.

[34] De facto, o que se verifica é que o jacobiano de $\mathbf{\Pi}$ (relativamente aos parâmetros desconhecidos) respeita a condição de característica, desde que T ≥ 3. É possível verificar-se, ainda, que para J>1 a condição de

A estimação de um modelo de desfasamentos distribuídos com base em amostras de dados em painel de curta duração pode ser feita através da decomposição do modelo base (4.2) no seguinte sistema de *T* equações reduzidas:

$$\underset{(T\times 1)}{\mathbf{y}_i} = \mathbf{i}\mathrm{m} + \left[\mathrm{I}_{\mathrm{T}} \otimes \mathbf{x}_i'\right]\boldsymbol{\pi} + \mathbf{v}_i \;,\; i = 1, \quad , N \qquad (4.7)$$

em que $\boldsymbol{\pi}' = [\boldsymbol{\pi}'_1, \ldots , \boldsymbol{\pi}'_T]$ e $\boldsymbol{\pi}'_j$ é a j-ésima linha da matriz $\boldsymbol{\Pi}$ (o que torna $\boldsymbol{\pi}$ uma matriz de dimensão $(T^2\times 1)$) e $\mathbf{v}_i$ não está correlacionado com $\mathbf{x}_i$. Sob as usuais condições de regularidade, o estimador LS de $\boldsymbol{\pi}$ é consistente e assimptoticamente normal, sendo aplicáveis os usuais testes assimptóticos.

---

característica se verifica para $T \geq J+2$, de acordo com Hsiao [1986].

5. ESTIMAÇÃO GMM DE MODELOS DINÂMICOS COM DADOS EM PAINEL

Como se viu, um problema que ocorre frequentemente na estimação de modelos dinâmicos com dados em painel é o da perda da consistência dos estimadores convencionais, pelo menos quando $N \to \infty$. O estimador *within*, por exemplo, torna-se inconsistente, com T fixo, porque a transformação *within* origina uma correlação de ordem $1/T$ entre a variável dependente desfasada e o termo de perturbação, como se viu no capítulo 3. Também notada nesse capítulo foi a solução popular proposta por Anderson e Hsiao [1981] de transformar o modelo para as primeiras diferenças (removendo os efeitos individuais) e usar a variável dependente desfasada em dois períodos como instrumento para o termo autorregressivo. Arellano e Bond [1991] propõem um conjunto mais alargado de instrumentos, que incluiriam recursivamente todos os valores passados de y_{it} (disponíveis para cada momento). No entanto, como nota Mátyás [1999], a viabilidade dos valores passados da variável dependente como instrumentos exigirá um conjunto de restrições sobre as covariâncias entre o termo de perturbação aleatório, o efeito individual e a observação inicial da variável dependente.

O objectivo do GMM será, então, o de encontrar um estimador consistente com um mínimo de restrições sobre os momentos. A organização deste capítulo será então a seguinte: numa primeira parte discute-se a aplicação do GMM aos dados em painel de forma genérica e, numa segunda parte, especificamente o GMM com modelos dinâmicos.

5.1 Estimação GMM com Regressores Exógenos

Considere-se o modelo genérico com efeitos seccionais, apenas:

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, \quad u_{it} = \alpha_i + \varepsilon_{it} \Leftrightarrow \\ \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i, \quad \mathbf{u}_i = \mathbf{i}_T \otimes \alpha_i + \varepsilon_i \end{aligned} \tag{5.1}$$

O objectivo do método GMM será, então, a estimação dos parâmetros de um modelo especificando um mínimo de condições de momentos, não necessitando, pois, da especificação completa das distribuições das variáveis aleatórias usadas. A lógica do GMM é a idêntica à do mais prosaico método dos momentos (MM): fazer equivaler os momentos populacionais às suas contrapartes amostrais, resolvendo as equações que daí resultarem, desde que identificadas. O MM exige, para ser exequível, que as equações resultantes das condições de momentos sejam exactamente identificadas, i.e., que o número de parâmetros as estimar (dimensão do vector β) seja igual ao número de restrições impostas sobre os momentos.

Nos casos em que os parâmetros do modelo estão sobre-identificados pelas condições de momentos pode, então, recorrer-se ao método GMM para se obter estimadores genericamente consistentes. Assim, as q condições de momentos $g_i = E(f(\mathbf{X}_i\beta)) = 0$ representam q equações com $k>p$ incógnitas, de tal forma que as equações com os momentos amostrais

$$f_N\left(\hat{b}_N\right) = 0$$

não terão uma solução. O melhor que se poderá fazer é encontrar uma estimativa para b_N que aproximará *f(.)* o mais possível de zero, ou seja, um estimador (GMM) que respeite

$$\hat{b}_N = \boldsymbol{arg}\, min_{b} Q_N(b) = f_N\left(\hat{b}_N\right)' \mathbf{V}_N\, f_N\left(\hat{b}_N\right) = 0$$

em que $\mathbf{V}_N$ é uma matriz de ponderação estocástica definida positiva e Q_N é estritamente positiva no caso sobre-identificado. Sob certas condições de regularidade relativamente gerais, os estimadores assim obtidos serão consistentes[35] e assimptoticamente normais[36]. Para que os estimadores GMM sejam assimptoticamente eficientes, haverá que escolher apropriadamente a matriz de ponderação $\mathbf{V}_N$, no que poderá ser empregue um processo bi-etápico, partindo-se de um estimador consistente de b. Infelizmente, como o estimador obtido dependerá do ponto de partida (ainda que assimptoticamente equivalente), há que encontrar uma forma de garantir estimadores com bons comportamentos em amostras finitas, para o que se poderá usar, por exemplo, processos iterativos. Para se encontrar um estimador GMM mais eficiente assimptoticamente, ter-se-á que considerar, de igual forma, o conjunto das condições de momentos, não sendo verdadeiro que a um maior conjunto de condições de ortogonalidade (ou seja, um maior conjunto de informação) corresponda um estimador GMM mais eficiente, como nota Mátyás [1999]. Assim, apenas na medida em que uma condição adicional contribua com nova informação, para além da já contida nas

---

[35] Para que os estimadores GMM sejam fracamente consistentes (i.e., que convirjam em probabilidade para b) é necessário que g_i exista, seja finita e uma aplicação bijectiva; que os q momentos amostrais convirjam uniformemente em probabilidade para os seus contrapartes populacionais e que $\mathbf{V}_N$ convirja em probabilidade para uma matriz não aleatória definida positiva. Ver Mátyás [1999], por exemplo.

[36] Para o que será necessário impor, para além da consistência: que $f(\mathbf{X}_N, b)$ seja continuamente diferenciável em ordem a b, num espaço compacto **B**; que o Jacobiano (em ordem a b, logo, uma matriz $q \cdot p$) das condições de momentos convirja em probabilidade (para uma qualquer sequência b_N^* que convirja em probabilidade para o verdadeiro vector b) para uma sequência de matrizes de igual dimensão e que não

condições anteriores é que a sua inclusão aumentará a eficiência do estimador GMM. Esta condição é particularmente importante para modelos dinâmicos com dados em painel (Mátyás [1999]).

Sobre este tema, devem-se consultar, entre outros, Amemiya [1985], Davidson e MacKinnon [1993], Greene [1997] e Mátyás [1999].

Relativamente ao modelos (5.1), uma estimação por GMM começará por impor uma condição de ortogonalidade entre um conjunto de $(T\cdot p)$ instrumentos $\mathbf{Z}_i$ e os termos de perturbação $\mathbf{u}_i$, dada por $E(\mathbf{Z}_i'\mathbf{u}_i) = 0$ e uma condição de identificação, dada por $car[E(\mathbf{Z}_i'\mathbf{u}_i)] = k$. Existindo esses instrumentos, pode-se obter um estimador GMM de β pela minimização da função objectivo $N(\mathbf{Y}\text{-}\mathbf{X}\boldsymbol{\beta})'\mathbf{Z}(\mathbf{V}_N)^{-1}\mathbf{Z}'(\mathbf{Y}\text{-}\mathbf{X}\boldsymbol{\beta})$ e em que $\mathbf{V}_N$ pode ser qualquer estimador consistente de $\mathbf{V}= E(\mathbf{Z}_i'\mathbf{u}_i\,\mathbf{u}_i'\mathbf{Z}_i)$. O estimador GMM será:

$$\hat{b}_{GMM} = \left(\mathbf{X}'\mathbf{Z}(\mathbf{V}_N)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{V}_N)^{-1}\mathbf{Z}'\mathbf{Y} \tag{5.2}$$

e em que, como sugere Mátyás [1999], se pode obter $\mathbf{V}_N$ a partir de um estimador inicial consistente de $\boldsymbol{\beta}$, como um estimador 2SLS. Assim, sendo $\mathbf{e}_i$ os resíduos dessa estimação:

$$V_N = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i$$

Se se impuser que $E(\mathbf{u}_i\,\mathbf{u}_i'|\mathbf{Z}_i) = \boldsymbol{\Sigma}$, com $\boldsymbol{\Omega}=\sigma_m^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_v^2\mathbf{I}_{NT} = \mathrm{I}_N \otimes \boldsymbol{\Sigma}$, como se vem fazendo desde a secção 2.2, o estimador GMM será equivalente ao estimador 3SLS dado por (5.2), mas com $\mathbf{V}_N = \mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}$. Aliás, segundo Mátyás [1999], para que essa equivalência entre 3SLS e GMM se verifique, basta que seja válido o pressuposto de inexistência de heteroscedasticidade condicional (NCH – *No Conditional Heteroskedasticity*) relativamente aos instrumentos, i.e., que $E(\mathbf{Z}_i'\mathbf{u}_i\,\mathbf{u}_i'\mathbf{Z}_i) = E(\mathbf{Z}_i'S\,\mathbf{Z}_i)$, o que é menos restritivo que $E(\mathbf{u}_i\,\mathbf{u}_i'|\mathbf{Z}_i) = \boldsymbol{\Sigma}$. Esta condição NCH permite igualmente a utilização do estimador GIV (referido na secção 3.2 a propósito do estimador de Arellano e Bond [1991]) como um estimador GMM que e de forma análoga a (3.9), pode ser definido como

$$\hat{b}_{GIV} = \left(\mathbf{X}'\Omega^{-1}\mathbf{Z}\left(\mathbf{Z}'\Omega^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Z}\left(\mathbf{Z}'\Omega^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\Omega^{-1}\mathbf{Y}$$

o que mais não é que aplicar o método das variáveis instrumentais ao modelo (5.1) pré-multiplicado por $\boldsymbol{\Sigma}^{-½}$, tendo como instrumentos $\boldsymbol{\Sigma}^{-½}\mathbf{Z}_i$ (ou seja, depois de filtrar modelo e instrumentos por $\boldsymbol{\Sigma}^{-½}$). Um método regra geral equivalente, segundo Mátyás [1999] é o das

---

dependam de b.

variáveis instrumentais filtradas (FIV) em que o processo de filtragem por $\mathbf{\Sigma}^{-½}$ se aplica apenas a (5.1) e que consiste em:

$$\hat{\beta}_{FIV} = \left(\mathbf{X}'\Omega^{-1/2}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\Omega^{-1/2}\mathbf{X}\right)^{-1}\mathbf{X}'\Omega^{-1/2}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\Omega^{-1/2}\mathbf{Y}$$

e que, mais uma vez, é GMM, desde que vigore a condição NCH. Mas Mátyás [1999] mostra que, regra geral, a filtragem (que é o equivalente à aplicação de GLS no contexto do OLS) não melhora a eficiência assimptótica do GMM e, assim, deve ser considerada como irrelevante, no contexto da estimação, quer com efeitos fixos, quer com efeitos aleatórios, desde que se explorem todas as condições de momentos. O mesmo acaba por ser válido para outro tipo de instrumentos com filtragem, como o estimador FF (*Forward Filtering*) de Keane e Runkle [1992] e que aqui se omite para não alongar em demasia a exposição.

De qualquer forma, pode ser demonstrado que, regra geral, os estimadores 3SLS, FIV e FF não são superiores a um qualquer estimador GMM que faça uso de todas as condições de momentos, que tem contra si "apenas" a questão de uma eventual inexequibilidade. Um estimador GMM, que use todas as condições de momentos (lineares) e uma matriz de ponderação irrestrita, com p instrumentos fracamente exógenos e uma amostra de dimensão T, conta com $p\cdot(1+2+\ldots+T)=p\cdot(T+1)T/2$ instrumentos (Mátyás [1999]).

Questão interessante, referida em Mátyás [1999], é a da determinação de uma fronteira de eficiência para os estimadores GMM, i.e., se existirá ou não um número máximo de condições de momentos a partir do qual não será possível aumentar a eficiência do estimador. A questão é, por exemplo, analisada por Koenker e Machado [1997] que, para modelos GMM lineares com heteroscedasticidade genérica, demonstram que basta que o número de condições de momentos ou instrumentos cresça a um ritmo inferior a $N^{1/3}$, relacionando dessa forma a dimensão da amostra com o número de instrumentos.

De entre os estimadores GMM utilizados para regressores estritamente exógenos, merecem destaque em termos de eficiência os apresentados em Hausman e Taylor [1981] Breusch, Myzon e Schmidt [1989] e em Amemiya e Macurdy [1986] e que passarão a ser notados por H-T, BMS e A-M, respectivamente. Considere-se, para esse efeito, o seguinte modelo:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}\mathbf{Z}_i + u_{it}\,, \quad u_{it} = \alpha_i + \varepsilon_{it} \tag{5.3}$$

em que $\mathbf{Z}_i$ é uma matriz de g variáveis explicativas constantes no tempo.

Hausman e Taylor [1981] sugerem um estimador 2SLS eficiente para a estimação dos vectores de parâmetros β e λ associados às matrizes de variáveis explicativas $\mathbf{X}_{it}$ e $\mathbf{Z}_{i}$, em que as duas matrizes podem incluir variáveis ortogonais ($\mathbf{X}_{1it}$ e $\mathbf{Z}_{1i}$) e outras correlacionadas com os efeitos específicos a_i ($\mathbf{X}_{2it}$ e $\mathbf{Z}_{2i}$). Em (5-3), as matrizes $\mathbf{X}_{it}$, $\mathbf{Z}_{i}$, $\mathbf{X}_{1it}$, $\mathbf{X}_{2it}$, $\mathbf{Z}_{1i}$ e $\mathbf{Z}_{2i}$ têm dimensão ($k\cdot 1$), ($g\cdot 1$), ($k_1\cdot 1$), ($k_2\cdot 1$), ($g_1\cdot 1$) e ($g_2\cdot 1$), respectivamente.

Neste contexto, de eventual correlação entre uma ou várias variáveis explicativas e os efeitos específicos, sabe-se que os estimadores OLS, assim como os GLS, não são consistentes e que os estimadores *within* serão consistentes, mesmo que ineficientes e não permitirão a estimação dos parâmetros contidos em λ.

Uma solução proposta por H-T será a estimação por OLS do modelo (5-3) pré-multiplicado por $\mathbf{\Omega}^{-1/2}$, ou seja, aplicando um operador das primeiras diferenças generalizadas (1-(1-f)L), com f dado por (2.9)[37] e transformando as variáveis $\mathbf{X}$ e $\mathbf{Z}$ através da respectiva projecção ortogonal num espaço $\mathbf{A}$ de variáveis instrumentais $\mathbf{X}_{1it}$, $\mathbf{Z}_{1i}$ e respectivos vectores de médias. Na prática, tal significa, num primeiro passo, regredir $\mathbf{Z}_{2i}$ em $\mathbf{X}_{1it}$ e $\mathbf{Z}_{1i}$ (com N observações) e $\mathbf{X}_{2i.}$ (ou seja, as médias amostrais para cada indivíduo) em $\mathbf{X}_{1i.}$ e $\mathbf{Z}_{1i.}$ (mais uma vez, as médias amostrais), obtendo-se $\hat{\mathbf{Z}}_{2i}$ e $\hat{\mathbf{X}}_{2i.}$. Num segundo passo, substitui-se em (5.3), $\mathbf{X}_{2it}$ e $\mathbf{Z}_{2i}$ por, respectivamente, $\hat{\mathbf{Z}}_{2i}$ e $\hat{\mathbf{X}}_{2it} = \mathbf{X}_{2it} - \mathbf{X}_{2i.} + \hat{\mathbf{X}}_{2i.}$ e estima-se por OLS a equação daí resultante, previamente transformada por (1-(1-f)L), em que L é usual operador de desfasamento. H-T chegam ainda a um resultado importante quanto à identificação dos parâmetros β e λ: há uma condição necessária de ordem, segundo a qual os dois vectores serão identificados se o número de variáveis exógenas $\mathbf{X}_1$ for igual ou superior ao número de variáveis pré-determinadas $\mathbf{Z}_2$[38] ($k_1 \geq g_2$).

Usando pressupostos mais restritivos quanto à exogeneidade face aos efeitos seccionais das variáveis $\mathbf{X}_{1it}$ e $\mathbf{Z}_{1i}$ do que H-T (que apenas admite ausência de correlação entre as médias destas variáveis e os efeitos α_i), Amemiya e MaCurdy [1986] chegam a um estimador IV mais eficiente. Assim, assumem que estas variáveis não são correlacionadas com a_i para todo o t, o que aumenta o leque de variáveis instrumentais a incluir: não usam

---

[37] Ou seja, $\mathrm{f} = \sqrt{\mathrm{s}_v^2 \big/ \left(\mathrm{s}_v^2 + T\mathrm{s}_{\mathrm{m}}^2\right)}$.

[38] Haverá uma condição de necessária e suficiente de característica, que consta de Hausman e Taylor [1981], 1385.

apenas as médias de $\mathbf{X}_1$ e $\mathbf{X}_2$, mas também todos os valores desfasados do primeiro vector de variáveis, recursivamente.

Assumindo um pressuposto de exogeneidade ainda mais restritivo, Breusch, Myzon e Schmidt [1989] deduzem um estimador IV com maior eficiência assimptótica. Para tal, impõem que os efeitos seccionais, para além de respeitarem as condições impostas por H-T e A-M, não exibirão correlação com os desvios em relação às médias seccionais das variáveis $\mathbf{X}_2$, i.e., ainda que α_i e $\mathbf{X_2}$ possam estar correlacionadas, a sua covariância é constante (ou seja, impõem uma condição de estacionaridade). Assim, BMZ acrescentam os desvios em relação às respectivas médias de $\mathbf{X}_2$ ao lote de instrumentos propostos por A-M, também de forma recursiva (ou seja, em T, teremos mais T-1 instrumentos, i.e., os desvios até T-1). De acordo com Mátyás [1999], qualquer um destes estimadores, H-T, A-M e BMZ, faz parte da classe dos estimadores FIV (variáveis instrumentais filtrados) e será equivalente ao estimador *within*, no caso exactamente identificado pela definição de H-T ($k_1 = g_2$), ou mais eficiente no caso sobre-identificado ($k_1 > g_2$).

Mátyás [1999] refere alguns estimadores que poderão ser mais eficientes que os estimadores H-T, A-M e BMZ, ao reconhecerem mais condições de momentos que as utilizadas por estes autores, citando para o efeito os estimadores de Ahn e Schmidt (AS) e Arellano e Bover (AB)[39] e que na prática significarão a estimação do modelo (5.3) usando como instrumentos, para além dos propostos H-T, A-M ou BMZ, as primeiras diferenças disponíveis até ao momento das variáveis explicativas não correlacionadas com os efeitos individuais, garantindo mais $(T\text{-}1)(Tk_1+g_1)\text{-}k_1$ instrumentos. Apesar disso, Mátyás [1999] sustenta que, do ponto de vista prático, os estimadores anteriores constituem boas alternativas ao GMM.

5.2 Estimação GMM e Modelos Dinâmicos

Os primeiros estimadores que podem ser referenciados como estimadores GMM para modelos dinâmicos com dados em painel são os já referidos AB e AS e que são sumariamente expostos em Mátyás [1999] e, com maior detalhe, em Baltagi [1995].

Considere-se uma especificação mais simples que a do modelo (3-13):

---

[39] Ver Ahn e Schmidt [1995], "Efficient Estimation of Dynamic Panel Data", *Journal of Econometrics*, 68, 5-27 e Arellano e Bover [1995], "Another Look at Instrumental Variables Estimation of Error-Components Models", *Journal of Econometrics*, 68, 29-51, citados em Mátyás [1999] e também em Baltagi [1995], sob a

$$y_{it} = \delta y_{i,t-1} + u_{it}, \text{ com } u_{it} = \alpha_i + v_{it} \qquad (5.4)$$

em que os termos de perturbação α e v têm ambos médias nulas e em que se considera que

1. $E(y_{i0})=0$;
2. $E(y_{i0} \cdot v_{it})=0, \ \forall t$;
3. $E(\alpha_i \cdot v_{it})=0, \ \forall t$;
4. os termos de perturbação v_{it} não estão autocorrelacionados, para todo o i.

Destas condições, relativamente comuns na literatura dos modelos dinâmicos em painel, como nota Mátyás [1999], pode-se retirar as $T(T-1)/2$ condições de momentos

$$E(y_{is}\Delta u_{it})=0, \text{ para } t=2,...,T \text{ e } s=0,1,...,T-2 \qquad (5.5)$$

em que $\Delta u_{it} = v_{it} - v_{i,t-1}$ e que foram usadas em diversos textos relativos à estimação por IV, como Anderson e Hsiao [1981], Holtz-Eakin, Newey e Rosen [1988] e Arellano e Bond [1991]. Ahn e Schmidt [1997], retomando o trabalho desenvolvido no seu artigo de 1995, verificam que as hipóteses 2-4 impõem mais $T-2$ condições de momentos que as usadas pelos outros autores e que a hipótese 1., relativa a y_{i0}, é desnecessariamente restritiva. As condições de momentos adicionais são:

$$E(u_{iT}\Delta u_{it}) = 0, \ t=2,\ldots,T-1 \qquad (5.6)$$

As condições (5.5) e (5.6) vêm directamente dos pressupostos de os v_{it} não serem correlacionados entre si, nem com os efeitos específicos, nem tão pouco com y_{i0}, não necessitando da hipótese de homoscedasticidade. Para além disso, os autores demonstram que estas seriam a totalidade das condições de momentos relevantes mesmo para os casos em que essas correlações não são nulas, desde que sejam constantes (no que é um pressuposto de estacionaridade). As condições podem e devem ainda ser vistas sob a forma de restrições lineares sobre os elementos da matriz Λ de covariâncias de u_{it}, $t=0,1,\ldots,T$:

$$\Lambda = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \\ u_{i0} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1T} & \lambda_{10} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2T} & \lambda_{20} \\ \vdots & \vdots & & \vdots & \vdots \\ \lambda_{T1} & \lambda_{T2} & \cdots & \lambda_{T0} & \lambda_{T0} \\ \lambda_{10} & \lambda_{20} & \cdots & \lambda_{T0} & \lambda_{00} \end{bmatrix}$$

de tal forma que todos os elementos das últimas linha e coluna (λ_{i0}) são iguais entre si ($T-1$ restrições) e que todos os restantes elementos fora da diagonal principal são também todos

---

forma de *working papers*.

iguais entre si (*T(T-1)/2-1* condições)[40]. No entanto, como as condições emergentes de (5.6) são não lineares em *u*, a utilização deste método GMM terá que levar a uma ponderação entre os ganhos de eficiência permitidos e o acréscimo exigido na complexidade computacional pela sua implementação.

No caso dos termos *v* serem homoscedásticos, como consta de Ahn e Schmidt [1997], há que considerar um novo grupo de (*T-1*) novas condições de momentos dadas genericamente por:

$$E\left(\bar{u}_i u_{it}\right) = 0,\, t = 2,...,T \tag{5.7}$$

Para além disso, como notam Ahn e Schmidt [1997], as condições (5.6) podem ser substituídas por um idêntico número de condições lineares dadas por

$$E(y_{i,t-2}\Delta u_{i,t-1} - y_{i,t-1}\Delta u_{it}) = 0, \quad t = 3,\ldots,T \tag{5.8}$$

Como refere Mátyás [1999], os ganhos de eficiência permitidos pelo pressuposto de homoscedasticidade acabam por ser reduzidos, pelo que a condição pode ser dispensada, tendo como contrapartida uma especificação mais robusta.

Várias outros pressupostos podem ser acrescentados ou combinados, como é efectuado em Ahn e Schmidt [1997], dando origem a diferentes conjuntos de condições de momentos. Uma condição que vale a pena analisar é a de estacionaridade, como proposta em Breusch, Mizon e Schmidt [1989], já aqui discutida e cujas consequências em termos de uma estrutura GMM são analisadas em Ahn e Schmidt [1997]. Na linha do que foi feito por Arellano e Bover, Ahn e Schmidt[41] impõem que $cov(\alpha_i, y_{it})$ seja constante, o que leva a uma nova restrição:

$\sigma_{0\alpha} = \sigma_{\alpha\alpha}/(1-\delta)$ ou de forma equivalente, $\lambda_{0t} = \lambda_{ts}/(1-\delta)$, para $t,s=1\ldots,T;\ t \neq s$

e que resulta num conjunto *T(T-1)/2 + (2T-2)* de condições de momentos lineares em δ dadas por (5.5) e, ainda,

$$E(u_{iT}\Delta y_{it}) = 0,\ t = 1,\ldots,T-1$$

$$E(u_{it}y_{it} - u_{i,t-1}y_{t-1}) = 0,\ t=2,\ldots T$$

Ahn e Schmidt [1997] analisam ainda o caso do pressuposto de estacionaridade em covariância da série $y_{i0},\ldots,y_{iT}$ e que leva a que as condições de momentos necessárias à estimação GMM sejam as contidas em (5.5) e ainda um única condição não linear adicional:

[40] Isto é, iguais a, respectivamente, $\sigma_{0\alpha}$ e σ_α^2.
[41] Artigos citados em Baltagi [1995] e Mátyás [1999].

$$E[y_{it}^{2} + y_{i1}\Delta u_{i2}/(1-\delta^{2}) - u_{i,2}u_{i1}/(1-\delta^{2})] = 0$$

A partir do modelo (5.4), pode-se definir um estimador GMM eficiente da seguinte forma:

- usam-se as condições de momentos (5.5), escrevendo-as na forma matricial como $E(\mathbf{A}_i' u_i(\delta)) = 0$, com

$$\mathbf{A}_i = \begin{bmatrix} -y_{i0} & 0 & 0 & 0 \\ y_{i0} & -(y_{i0}, y_{i1}) & 0 & 0 \\ 0 & (y_{i0}, y_{i1}) & (y_{i0}, y_{i1}, y_{i2}) & 0 \\ & & & \\ 0 & 0 & 0 & -(y_{i0}, y_{i1}, \ldots, y_{iT-2}) \\ 0 & 0 & 0 & (y_{i0}, y_{i1}, \ldots, y_{iT-2}) \end{bmatrix}$$

- as condições (5.8) podem ser escritas também da mesma forma, i.e., como $E(\mathbf{B}_i' u_i(\delta)) = 0$, em que

$$\mathbf{B}_i = \begin{bmatrix} -y_{i1} & 0 & 0 & 0 \\ (y_{i1} + y_{i2}) & -y_{i2} & 0 & 0 \\ -y_{i2} & (y_{i2} + y_{i3}) & -y_{i3} & 0 \\ & & & \\ 0 & 0 & 0 & (y_{i,T-2} + y_{i,T-1}) \\ 0 & 0 & 0 & -y_{i,T-1} \end{bmatrix}$$

- as l condições de momentos (5.6), (5.7) (quadráticas em δ) e (5.8) podem ser divididas em l_1 condições lineares[42] e l_2 condições não lineares em δ;
- as l_1 condições lineares podem ser representadas recorrendo a S_i, uma matriz $T \cdot l_1$ com todas as colunas de $\mathbf{A}_i$ e $\mathbf{B}_i$, que reúne todos os instrumentos lineares disponíveis, da seguinte forma

$$E(f_i(\delta)) = 0, \text{ com } f_i(\delta) = S_i' u_i(\delta) = f_{1i} + f_{2i}\delta, \text{ em que } f_{1i} = S_i' y_i \text{ e } f_{2i} = -S_i' y_{i,-1}$$

- as restantes l_2 condições não lineares poderão ser representadas como uma função quadrática, tal que $E(g_i(\delta))=0$, com

$$g_i(\delta) = g_{1i} + g_{2i}\delta, + (I_{l_2} \otimes \delta')g_{3i}\delta$$

- um estimador eficiente pode ser obtido através de um método GMM baseado em todas as condições de momentos definidas por

[42] O número de condições lineares, l_1, terá que ser pelo menos igual ao número de parâmetros a estimar (neste caso, 1) para que estes sejam identificados.

$$E[m_i(\delta)] = E\begin{bmatrix} f_i(\delta) \\ g_i(\delta) \end{bmatrix} = 0$$

- defina-se $m_N = N^{-1}\sum_i m_i(\delta)$[43] e $M_N = \partial m_N / \partial \delta = [\partial f_N / \partial \delta, \partial g_N / \partial \delta] = [F_N, G_N] = [f_{2N}\ ,\ g2i + 2(I_{l_2} \otimes \delta')g_{3i}]$ e, ainda, $M = plim\ M_N = [plimF_N\ ,\ plimG_N]$ e, finalmente, como matriz de ponderação óptima

$$\Omega = \begin{bmatrix} \Omega_{ff} & \Omega_{fg} \\ \Omega_{gf} & \Omega_{gg} \end{bmatrix} = \lim_{N \to \infty} cov\left(\sqrt{N} \sum_i m_i \right)$$

  e que é estimada de forma consistente por $\hat{\Omega} = N^{-1} \sum_i m_i(\hat{\delta}) m_i(\hat{\delta})'$, nomeadamente a partir das condições lineares f_i;

- o estimador GMM, d_{GMM}, resultará da minimização de $Nm_N(\delta)'\hat{\Omega}^{-1} m_N(\delta)$ e terá como matriz de covariâncias assimptótica $N^{1/2}(d_{GMM} - \delta) = [M'\Omega^{-1}M]^{-1}$;
- a validade das condições de momentos $E[m_i(\delta)] = 0$ pode ser testada através de um teste convencional de sobre-identificação, recorrendo à estatística $J_N = Nm_N(d_{GMM})'m_N(d_{GMM})$, com distribuição $\chi^2(l-1)$[44], sob a hipótese nula de todas as condições serem simultaneamente válidas;
- no caso de todas as condições de momentos serem lineares, o estimador GMM eficiente (quando comparado com IV, por exemplo) tem uma solução algébrica fechada dada por $d_f = -[f_{2N}'\hat{\Omega}^2_{ff}\ f_{2N}]^{-1}\ f_{2N}'\hat{\Omega}^2_{ff}\ f_{1N}$.
- para o caso de o modelo a estimar conter também variáveis exógenas, $\boldsymbol{X}_i$ (como um termo independente), haveria que considerar um conjunto de condições de momentos $E[\boldsymbol{X}_i' u_i(\xi)] = 0$, em que $\xi = [\delta\ \beta']'$ reúne o conjunto dos parâmetros a estimar. A matriz S_i incluiria também as colunas de $\boldsymbol{X}_i$ e a estimação prosseguiria de igual forma, com ξ a ser usado em vez de δ e $\boldsymbol{W}_i = [\boldsymbol{y}_{i,-1}\ \boldsymbol{X}_i]$ (que inclui todas as variáveis explicativas do modelo, endógenas desfasadas ou exógenas) em vez de $\boldsymbol{y}_{i,-1}$.

De um modo geral, Ahn e Schmidt [1997] mostram que há ganhos significativos a alcançar pela introdução de condições de momentos não lineares (que conduzem inevitavelmente a métodos de estimação iterativos) e desenvolvem um conjunto de testes

---

[43] Com f_N, g_N, f_{1N}, f_{2N}, g_{1N}, g_{2N} e g_{3N} definidos de igual forma.

[44] $\chi^2(l-k)$ no caso de haver k regressores, exógenos ou pré-determinados.

para aferir a validade das referidas condições pela comparação entre os estimadores GMM completos e estimadores GMM linearizados, de mais simples implementação.

Note-se, ainda, que o primeiro dos pressupostos, relativo a $E(y_{i0})$, faz com que as condições de momentos de primeira ordem ($E(u_{it}=0)$, com $u_{it}=a_i+v_{it}$) cessem de ser informativas, na medida em que deixam de identificar univocamente d, pelo que sempre seria possível obter estimadores GMM mais eficientes se se não restringisse y_{i0}.

Nesse sentido, temos Crépon, Kramarz e Trognon [1997] que, admitindo condições iniciais irrestritas para y, mostram como a eliminação de parâmetros redundantes (*nuisance parameters*)[45] pode não ter qualquer efeito em termos de perda de eficiência. Para tal, basta que se substituam os parâmetros redundantes pelas respectivas contrapartes empíricas, nas condições de ortogonalidade.

Admita-se que o vector de parâmetros a estimar, num modelo autorregressivo com regressores exógenos, dado por θ, pode ser dividido num vector de parâmetros relevantes, γ e num vector de parâmetros redundantes, ζ. Crépon *et al*. [1997] mostram que os cálculos exigidos pela estimação GMM, baseada na totalidade das condições de ortogonalidade, podem ser grandemente simplificados. Ao contrário do GMM completo (em que os parâmetros redundantes, ζ, têm que ser expressos em função dos parâmetros relevantes, γ), no GMM abreviado, as condições de momentos só necessitam de ser expressas com relação aos parâmetros γ, na medida em que os parâmetros ζ são substituídos por estimativas obtidas por um estimador etápico consistente. O procedimento acaba por ser análogo ao usado na estimação por MLE, quando se opta pela concentração da função de verosimilhança, ainda que a equivalência assimptótica entre o procedimento GMM completo e este abreviado só seja verdadeira caso se verifiquem certas condições e que constam de Crepon *et al.* [1997] (proposição 2).

5.3 Conclusões

Uma primeira conclusão que se pode retirar da questão da estimação IV ou GMM (que, como se viu, não é exactamente coincidente, nomeadamente quando as condições de momentos não são todas lineares) é a de que ela tem, de facto, interesse quando se está a

---

[45] O conjunto de parâmetros a estimar pode ser dividido em duas categorias: uma de parâmetros com interesse económico, como o coeficiente associado ao termo AR(1) do modelo (5.4); outra de parâmetros

estimar modelos dinâmicos com amostras em painel com uma dimensão temporal pequena. Isto porque, com um grande número de observações ao longo do tempo para um número comparativamente pequeno de indivíduos (ou unidades seccionais), os estimadores *within* são consistentes, mesmo que não completamente eficientes. Depois, como é mostrado em Arellano e Bond [1991], entre outros, a quantidade de condições de momentos disponíveis aumenta quadraticamente em T, o que torna o peso computacional do GMM, para um T elevado, demasiado oneroso.

Para a estimação de modelos lineares, o estimador GMM frequentemente assume a forma de um estimador IV, o que é particularmente verdadeiro se for válido o pressuposto de inexistência de heteroscedasticidade condicional (NCH). A diferença entre o GMM e a forma convencional de aplicação das variáveis instrumentais tem que ver com a completa identificação das condições de momentos impostas pelos pressupostos do modelo subjacente que o GMM implica, em que algumas dessas condições de momentos poderão ser redundantes, logo, não contribuindo para a eficiência do estimador.

No caso concreto dos modelos dinâmicos, a condição NCH é necessariamente insustentável, caso se opte pela consideração da totalidade das condições de momentos suscitadas pelos pressupostos de exogeneidade. Assim, na estimação de modelos dinâmicos com dados em painel, o estimador GMM eficiente não corresponderá a um estimador IV, tanto mais que muitas das condições em que se baseia são não lineares, dando lugar a procedimentos de estimação iterativos.

A literatura GMM relativa à estimação de modelos com dados em painel está a atravessar um período de rápida expansão, pelo que esta revisão estará necessariamente incompleta. Para além disso, todos os tópicos relacionados com a estimação de modelos não-lineares (dinâmicos) com GMM foi propositadamente deixada de fora, sob pena de se alongar indefinidamente a discussão. Uma boa revisão sobre o tema pode ser encontrada em Mátyás [1999], (capítulo 9).

---

sem significado económico relevante e cuja estimação dispensamos.

6. RAÍZES UNITÁRIAS E DADOS EM PAINEL

Desde o trabalho fundamental de Balestra e Nerlove [1966] que os modelos dinâmicos com dados em painel têm revelado uma crescente importância na análise econométrica, na medida em que permitem uma abordagem mais abrangente de fenómenos de ajustamento que não podem ser vistos de forma isolada. No entanto, devido à pequena dimensão temporal da maioria dos estudos com dados em painel (essencialmente na área da microeconometria), a maior parte dos modelos contem dinâmicas homogéneas e pouca atenção foi dada, até ao momento, a modelos com dinâmicas de ajustamento heterogéneas do ponto de vista individual e temporal. Com o avanço da macroeconometria para a área das aplicações com painéis, a dimensão temporal destes tende a aumentar e novos problemas se colocam, nomeadamente os relacionados com temas até então típicos das séries temporais, como as raízes unitárias e a cointegração. A disponibilidade de painéis[46] com dimensões temporais alargadas põe em causa o pressuposto tradicional da homogeneidade dinâmica e suscitam também novas abordagens à estimação (por exemplo, devido à inconsistência dos estimadores por agregação (*pooled estimators*) em modelos dinâmicos heterogéneos).

A literatura relativa às questões da estacionariedade das séries económicas é muito extensa para estudos com amostras temporais mas ainda é relativamente incipiente no que toca à sua relevância em amostras em painel. Adicionalmente, muita da ênfase tem sido dada à possibilidade de os dados em painel poderem ser usados como uma alternativa, mais poderosa, aos habituais testes de raízes unitárias[47]. Estes, regra geral, padecem de reduzida potência ou são apenas válidos em circunstâncias excessivamente restritivas para serem de utilidade prática. Uma forma de aumentar a potência desses testes será o aumento da dimensão da amostra temporal, o que causa problemas como o aumento da possibilidade de ocorrerem quebras de estruturas que sugiram, erradamente, raízes unitárias. A inclusão de uma dimensão seccional à amostra, em paralelo com a temporal, é uma forma de aumentar a sua dimensão sem o risco de quebras de estrutura, desde que se modelize de forma conveniente a heterogeneidade individual.

---

[46] Em boa verdade, os painéis usados em estudos macroeconométricos são aquilo a que a literatura chama de "pseudo" painéis (Im, Pesaran e Shin [1997]) ou "*cohorts*" (bandos) na medida em que resultam da agregação de N séries temporais individuais e não da recolha simultânea de T observações para N indivíduos.
[47] Heimonen [1999] é um exemplo, com uma aplicação à hipótese das paridades de poder de compra.

Diversos artigos exploram a utilização de painéis para testar a integração de variáveis económicas, como alternativa à simples extensão de amostras temporais (Maddala e Kim [1998]), de que se destacam os artigos de Levin e Lin [1992] e [1993] e o de Im, Pesaran e Shin [1997].

Levin e Lin [1993] usam como hipótese nula a existência de resíduos integrados para cada indivíduo, contra a hipótese alternativa de todos os indivíduos terem resíduos estacionários. Assim, a hipótese nula impõe uma restrição transversal entre as diversas equações individuais[48] (*cross-equation*), sobre os coeficientes de autocorrelação parcial de primeira ordem. Levin e Lin [1993] desenvolvem um teste de raiz unitária que pode incorporar efeitos individuais e temporais específicos no termo independente e nos restantes coeficientes e com uma estrutura de variâncias e covariâncias dos termos de perturbação sem restrições seccionais: a única restrição será a existência ou não de uma raiz unitária. A revelação deste trabalho acaba por ser a normalidade assimptótica das estatísticas dos testes de raiz unitária e a sua super-consistência na dimensão temporal, i.e., que a convergência das estatísticas é mais rápida para $T \to \infty$ do que para $N \to \infty$. Mais, no mesmo artigo, os autores recomendam os testes de estacionariedade específicos a dados em painel para amostras de média dimensão, i.e., com $10<N<250$ e $25<T<250$, enquanto que para amostras de grande dimensão temporal se recomendam os testes *Augmented Dickey-Fuller* (ADF) tradicionais separadamente para cada série individual. Para painéis com grande dimensão seccional, quando comparada com a temporal, os autores recomendam os testes tradicionais da literatura de painéis, como o de Holtz-Eakin, Newey e Rosen [1988], que se verá adiante. Assim, de acordo com Levin e Lin [1993], desde que se utilizem estimadores apropriados para a média e variância, os valores críticos da distribuição normal reduzida poderão servir como boas aproximações para os testes de raiz unitária convencionais, com melhores resultados que os obtidos para cada série temporal isoladamente.

Maddala e Kim [1998] criticam, no entanto, os testes de Levin e Lin [1992] por admitirem que o coeficiente associado ao termo autorregressivo não apresenta heterogeneidade individual, ou seja, por terem como hipótese nula a não estacionariedade simultânea de todos os processos geradores de dados individuais e, por isso, consideram preferível o procedimento proposto por Im, Pesaran e Shin [1997]. O teste proposto em

---

[48] Intui-se que Levin e Lin [1993] adoptam a versão de Chamberlain [1984] da estimação com dados em

Levin e Lin [1993], no entanto, já admite que ocorra heterogeneidade no coeficiente associado à variável dependente desfasada, sendo o caso com maior interesse o que recorre ao seguinte modelo:

$$\Delta y_{it} = \alpha_{0i} + \alpha_{1i} t + \gamma_i y_{i,t-1} + u_{it}$$

$$u_{it} = \sum_{j=0}^{\infty} \rho_{it} u_{it} + \varepsilon_{it}$$

em que se admite que o termo de perturbação *u* segue um processo ARMA invertível e que os processo geradores de dados individuais são mutuamente independentes ou apenas dependentes através de um efeito temporal específico agregado (δ_t). O modelo admite, então, heterogeneidade individual ao nível do termo independente e do *trend* temporal. O teste consiste nos seguintes passos:

(i) remover os efeitos temporais agregados através da transformação do modelo para a forma de diferenças em relação à média seccional ($y^*_{it} = y_{it} - y_{.t}$);

(ii) estimar *N* regressões separadamente do tipo $y_{it} = \alpha_{10i} + \alpha_{11i} t + \sum_{l=1}^{p_i} \gamma_{1il} \Delta y_{i,t-1} + u_{1it}$ e $\Delta y_{it} = \alpha_{20i} + \alpha_{21i} t + \sum_{l=1}^{p_i} \gamma_{2il} \Delta y_{i,t-1} + u_{2it}$, obter os respectivos resíduos, e_{it} e v_{it} e, de seguida, estimar $e_{it} = \gamma_i v_{i,t-1} + \varepsilon_{it}$;

(iii) estimar o rácio dos desvios padrão de curto prazo relativamente aos de longo prazo para cada indivíduo e calcular o rácio médio para o painel dados por $\hat{s}_i = \frac{\hat{\sigma}_{y_i}}{\hat{\sigma}_{e_i}}$ e $\hat{S} = \frac{1}{N} \sum_{i=1}^{N} \hat{s}_i$, em que $\hat{\sigma}^2_{y_i} = \frac{1}{T-1} \sum_{t=2}^{T} \Delta y^2_{it} + 2 \sum_{l=1}^{\bar{k}} \frac{L}{\bar{k}+1} \left(\frac{1}{T-1} \sum_{t=2+l}^{T} \Delta y_{it} \Delta y_{i,t-l} \right)$[49] e $\hat{\sigma}^2_{e_i} = \frac{1}{T - p_i - 1} \sum_{t=p_i+2}^{T} \left(e_{it} - \hat{\gamma}_i v_{i,t-1} \right)^2$;

(iv) calcular a estatística do teste:

$$t^*_{\gamma} = \frac{\hat{\gamma} / RSE(\hat{\gamma}) - N\tilde{T} \cdot \hat{S} \cdot RSE(\hat{\gamma}) \mu^*_{m\tilde{T}} / \hat{\sigma}^2_{\varepsilon}}{\sigma^*_{m\tilde{T}}} \overset{a}{\sim} N(0,1) \qquad (6.1)$$

onde,

painel mediante um sistema de *n* equações individuais.

[49] A dimensão óptima do *truncation lag k* é dada em anexo por Levin e Lin [1993].

$$\hat{\gamma}=\frac{\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}\tilde{e}_{it}\tilde{v}_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}v_{i,t-1}^{2}}\ ,\ \hat{\sigma}_{e_i}^{2}=\frac{1}{N\tilde{T}}\sum_{i=1}^{N}\sum_{t=p_i+2}^{T}\left(\tilde{e}_{it}-\hat{\gamma}\tilde{v}_{i,t-1}\right)^{2}\ ,\ RSE(\hat{\gamma})=\hat{\sigma}_{\varepsilon}\Big/\sqrt{\sum_{i=1}^{N}\sum_{t=p_i+2}^{T}\tilde{v}_{i,t-1}^{2}}\ ,$$

$$\tilde{e}_{i,t}=\frac{e_{it}}{\hat{\sigma}_{e_i}}\ ,\ \tilde{v}_{i,t}=\frac{v_{it}}{\hat{\sigma}_{e_i}}\ ,\ \tilde{T}=T-\frac{1}{N}\sum_{i=1}^{N}p_i-1$$

e em que os valores para os parâmetros com * em sobrescrito, obtidos por simulações de Monte Carlo, podem ser encontrados em anexo ao artigo referido. A convergência em distribuição da estatística (6.1) é entendida para $T\rightarrow\infty$ apenas devido ao já referido resultado de super-consistência.

A principal deficiência que se pode encontrar neste teste de Levin e Lin [1993] acaba por ser, não como dizem Maddala e Kim [1998], o de não considerarem heterogeneidade individual ao nível dos coeficientes associados aos termos autorregressivos, que se viu ter em conta nesta versão revista[50], mas sim o de restringirem de forma excessiva a estrutura de correlações ao nível seccional, renegando a própria natureza dos dados em painel. Nesse sentido, o teste de Levin e Lin não pode ser considerado como um verdadeiro teste de raízes unitárias com dados em painel.

O teste de Im, Pesaran e Shin [1997], IPS, tem como hipótese nula a não estacionariedade das séries temporais individuais e recorre a um procedimento assente na média das estatísticas LM (multiplicador de Lagrange) para cada grupo presente no painel. O teste é menos restritivo que o de Levin e Lin [1993] em termos de heterogeneidade das dinâmicas e das estrutura de variâncias inter-grupos e em termos de velocidade de convergência das dimensões temporal e seccional (exige apenas que $N/T\rightarrow k\neq 0$ e não $N/T\rightarrow 0$, como em Levin e Lin [1993]). Adicionalmente, o teste IPS permanece consistente mesmo que a hipótese alternativa admita que uma parcela das séries individuais exibe raízes unitárias. Segundo os seus autores, o teste IPS é relativamente robusto perante diferentes configurações em termos de *trends* e de estruturas de variâncias e covariâncias dos termos de perturbação.

O teste IPS, na sua versão base, parte de um modelo típico de um teste Dickey-Fuller (DF):

$$\Delta y_{it}=\alpha_i+\beta_i y_{i,t-1}+\varepsilon_{i,t}$$

[50] Maddala e Kim [1998] provavelmente referiam-se à primeira versão do teste em Levin e Lin [1992].

e tem como hipóteses nula e alternativa, respectivamente:

$$H_0\!:\beta_i=0\ ,\ \forall\ i$$

$$H_1\!:\beta_i<0\ ,\ i=1,\dots N_1$$

$$\beta_i=0\ ,\ i=N_1+1,\dots,N$$

ou seja, em que se admite como alternativa que nem todas os processos geradores de dados individuais sejam estacionários.

A estatística do teste é construída com base nas estatísticas LM para os testes individuais de $\beta_i=0$ e que são:

$$LM_{iT}=\frac{T\Delta\mathbf{y}_i'\mathbf{P}_i\Delta\mathbf{y}_i}{\Delta\mathbf{y}_i'\mathbf{M}_T\Delta\mathbf{y}_i}$$

em que se tem como matrizes idempotentes $\mathbf{M}_T=\mathbf{I}_T-\mathbf{i}_T\left(\mathbf{i}_T'\mathbf{i}_T\right)^{-1}\mathbf{i}_T'$ e $\mathbf{P}=\mathbf{M}_T\mathbf{y}_{i,-1}\left(\mathbf{y}_{i,-1}'\mathbf{M}_T\mathbf{y}_{i,-1}\right)^{-1}\mathbf{y}_{i,-1}'\mathbf{M}_T$.

A média das estatísticas, dada por $\overline{LM}=\frac{1}{N}\sum_{i=1}^{N}LM_{iT}$ servirá de base à estatística normalizada

$$\Gamma_{\overline{LM}}=\frac{\sqrt{N}\left(\overline{LM}_{NT}-E(\eta_T)\right)}{\sqrt{Var(\eta_T)}}\overset{a}{\sim}N(0,1) \qquad (6.2)^{51}$$

em que (η_T) e $Var(\eta_T)$ são parâmetros fornecidos em Im, Pesaran e Shin [1997] e obtidos mediante simulações de Monte Carlo.

Para painéis com dinâmicas heterogéneas e erros autocorrelacionados, de longe o caso mais interessante, o teste IPS é apenas ligeiramente mais elaborado, partindo um modelo AR(p_i+1), o que leva a que as N equações ADF[52] possam ser escritas da seguinte forma:

$$\Delta y_{it}=\alpha_i+\beta_i y_{i,t-1}+\sum_{j=1}^{p_i}\rho_{ij}\Delta y_{i,t-j}+\varepsilon_{it}\ ;\ i=1,\dots,N;\ t=1,\dots,T$$

ou na forma matricial equivalente:

$$\Delta\mathbf{y}_i=\beta_i\mathbf{y}_i+\mathbf{Q}_i\boldsymbol{\Gamma}_i+\boldsymbol{\varepsilon}_i\ ;$$

$$\boldsymbol{\Gamma}=\left[\alpha_i\quad\rho_{i1}\quad\rho_{i2}\quad\cdots\quad\rho_{ip_i}\right]'\ e\ \mathbf{Q}_i=\left[\mathbf{i}_T\quad\Delta\mathbf{y}_{i,-1}\quad\Delta\mathbf{y}_{i,-2}\quad\cdots\quad\Delta\mathbf{y}_{i,-p_i}\right]'$$

---

[51] A convergência em distribuição é suficiente para $N\to\infty$, com $T\geq 3$.

[52] O teste ADF deverá ser usado quando os termos de perturbação da regressão DF exibem autocorrelação.

Sob a hipótese da independência dos termos de perturbação idiossincráticos, ε_{it}, que terão que ser estacionários na variância (ainda que possam ser seccionalmente heteroscedásticos) e conhecidos os pontos de partida y_{i0} (estocásticos ou determinísticos), tem-se que a estatística definida em (6.2) converge fracamente[53] para uma distribuição normal reduzida (Teorema 4-1 em Im, Pesaran e Shin [1997]). Para calcular (6.2) deve agora usar-se:

$$LM_{iT} = \frac{T\Delta\mathbf{y}_i'\mathbf{P}_i^*\Delta\mathbf{y}_i}{\Delta\mathbf{y}_i'\mathbf{M}_{Q_i}\Delta\mathbf{y}_i}$$

em que se tem como matrizes idempotentes de ordem ($T\cdot T$) e característica (T-p_i-1) $\mathbf{M}_{Q_i} = \mathbf{I}_T - \mathbf{Q}_i(\mathbf{Q}_i'\mathbf{Q}_i)^{-1}\mathbf{Q}_i'$ e $\mathbf{P}_i^* = \mathbf{M}_{Q_i}\mathbf{y}_{i,-1}(\mathbf{y}_{i,-1}'\mathbf{M}_{Q_i}\mathbf{y}_{i,-1})^{-1}\mathbf{y}_{i,-1}'\mathbf{M}_{Q_i}$.

Os mesmos procedimentos (DF e ADF) podem ser aplicados a um modelo com efeitos temporais específicos, denunciando um certo grau de dependência entre os grupos, desde que sejam aplicados às respectivas equações expressas em desvios face às médias seccionais[54] (o que cria uma correlação cruzada entre os erros). Os testes passam, no entanto, a ser estritamente válidos apenas assimptoticamente (independentemente do pressuposto da normalidade). A estatística Γ_{LM} mantém-se, mas LM é dado por:

$$\overline{\overline{LM}}_{iT} = \frac{T\left(\Delta\tilde{\mathbf{y}}_i'\mathbf{M}_T\tilde{\mathbf{y}}_{i,-1}\right)^2}{\left(\Delta\tilde{\mathbf{y}}_i'\mathbf{M}_T\Delta\tilde{\mathbf{y}}_i\right)\left(\tilde{\mathbf{y}}_{i,-1}'\mathbf{M}_T\tilde{\mathbf{y}}_{i,-1}\right)}$$

onde as variáveis estão expressas na forma de desvios face às T médias seccionais e $\mathbf{M}_T$ mantém a mesma composição.

Segundo Im, Pesaran e Shin [1997], o teste revela-se relativamente robusto face a potenciais más especificações das componentes temporais, comparando com os testes correntes em estudos de séries temporais e a estruturas alternativas para as correlações entre os termos de perturbação, não sendo apenas válidos para modelos de efeitos temporais com heterogeneidade seccional (tipo $\lambda_i\cdot\theta_t$).

Um outro teste que é sugerido por Maddala e Kim [1998] é o teste de Fisher[55] e que consiste numa combinação da evidência fornecida por diversos testes independentes. O

---

[53] Convergência fraca é sinónimo de convergência em probabilidade (Amemiya [1985]).

[54] Ter-se-ia que trabalhar com $\Delta\tilde{y}_{it} = \Delta y_{it} - \frac{1}{N}\sum_{j=1}^{N}\Delta y_{jt}$, etc.…

[55] O teste deve-se a Fisher, R.A. [1932]. *Statiscal Methods for Research Workers*, 2nd Edition, Wiley: New

teste consiste na realização de N testes de raiz unitária para cada um dos indivíduos no painel e obter o respectivo *p-value*, sendo a estatística $P_\lambda = -2\Sigma ln(p_i)$ com distribuição exacta $\chi^2(2N)$ usada para efeitos de teste da hipótese nula de raízes unitárias. Este teste, também conhecido por teste P_λ de Pearson e que é detalhado em Maddala [1977] (pág. 47-8)[56] tem a vantagem de ser relativamente simples na sua implementação, ainda que exija a determinação dos *p-values* por simulações de Monte Carlo.

Finalmente, para painéis curtos é de usar o teste de Holtz-Eakin, Newey e Rosen [1988], H-ENR, que consiste na estimação de um modelo autorregressivo (com ou sem variáveis exógenas) através de um método GMM. Holzt-Eakin *et al.* [1988] sugerem que se estime, por 2SLS, um modelo autorregressivo como (5-4) nas primeiras diferenças, usando como instrumentos para $\Delta y_{i,t-1}$, o vector $\mathbf{Z}_t = [1, y_{i,t-2}, y_{i,t-3}, \ldots, y_{i,1}]$. Se os efeitos específicos variarem no tempo, i.e., se forem do tipo $\lambda_t \alpha_i$, haverá que proceder a uma quasi-diferenciação do modelo, usando em vez das primeiras diferenças: $y_{i,t} - (\lambda_t/\lambda_{t-1})y_{i,t-1}$. Holtz-Eakin *et al.* [1988] defendem que se use uma versão 3SLS, onde se efectua uma correcção GLS. O teste H-ENR pode ser usado para testar a estacionaridade dos parâmetros, a existência de raízes unitárias ou explosivas na equação autorregressiva, a dimensão do *truncation lag* ou o sentido da causalidade à Granger, num contexto VAR. Qualquer um destes testes pode ser visto como um teste a restrições lineares sobre os parâmetros do modelo, reunidos num vector θ $(k \times 1)$. A hipótese nula do teste é

$$H_0: \theta = \mathbf{H}\gamma + \mathbf{G}$$

em que **H** e **G** são matrizes de restrições lineares e γ o vector de parâmetros $(g \times 1)$ com restrições (de dimensão necessariamente inferior à de θ). O teste necessita que se estime o modelo sem restrições e com as restrições dadas por H_0, retendo-se as somas dos quadrados dos resíduos para ambos os caso, ou seja, Q e Q_r. Sob H_0 e com $N \to \infty$:

$$L = Q_r - Q \sim \chi^2(J) \text{ e } J = (k-g) - (z-k)$$

em que z é o número de instrumentos usados para estimar θ (Holtz-Eakin *et al.* [1988], 1381).

Um problema comum a qualquer teste de estacionariedade é o da selecção da dimensão do desfasamento, i.e., da escolha do *truncation lag*, questão já amplamente

York, citado em Maddala e Kim [1998]. Este mesmo teste poderá ser usado em contextos totalmente diferentes, como se propõe, na secção 8.3, para testar uma estrutura particular de heteroscedasticidade.

[56] C.f. Maddala [1977], 48,*n*.

explorada no âmbito das séries temporais e que, para já, suscita as seguintes linhas de orientação:

- o número de termos desfasados no teste ADF, por exemplo, deve ser o suficiente para garantir a ausência de autocorrelação nos resíduos da regressão de suporte do referido teste;
- a existência de termos MA negativos e de dimensão assinalável tende a exigir um aumento do número de termos autorregressivos;
- ao invés de se adoptar uma regra determinística comum para a selecção do número de desfasamentos, deve-se fazer recair a escolha em critérios assentes nos próprios dados (Maddala e Kim [1998]).

De qualquer forma, segundo Heimonen [1999], a escolha da dimensão do desfasamento a utilizar em teste de raízes unitárias com dados em painel será de menor importância, na medida em que a inferência parece largamente independente do número de termos autorregressivos. Este autor e Maddala e Kim [1998], por exemplo, sugerem o seguinte critério, dito de Schwert, para dados trimestrais e mensais, respectivamente[57]:

$$l_4 = \boldsymbol{int}\left(4\left(\frac{T}{100}\right)^{1/4} \right)$$

$$l_{12} = \boldsymbol{int}\left(12\left(\frac{T}{100}\right)^{1/12} \right)$$

No entanto, Im, Pesaran e Shin [1997] previnem contra os efeitos prejudiciais de uma subestimação da dimensão do desfasamento a usar no seu teste IPS, pelo que aconselham a utilização de um número considerável de termos autorregressivos, para o que conviria uma regra do tipo *general-to-specific* de Hendry.

Como resumo, pode-se prescrever a seguinte linha de acção:

- para painéis curtos, deve-se usar o teste de Holtz-Eakin, Newey e Rosen [1988];
- para painéis de dimensão intermédia, pode-se escolher entre o teste de Levin e Lin [1993] e o teste de Im, Pesaran e Shin [1997], dando preferência ao primeiro, quando a dimensão temporal domina e ao segundo, no caso inverso;
- para painéis longos ($T>250$), não é de excluir a utilização dos testes ADF, DF ou Phillips-Perron, num enquadramento típico de *time series*.

---

[57] O critério deve-se a Schwert, G.W. [1989], “Tests for unit roots: a Monte Carlo investigation”, *Journal of Business and Economic Statistics*, 7, 147-59.

7. DADOS DISCRETOS

7.1 Modelos Estáticos de Escolha Binária: uma introdução

Como se viu, os modelos desenvolvidos para analisar dados seccionais cometem invariavelmente o pecado original de ignorar as diferenças individuais, tratando os efeitos individuais de forma agregada, apenas, relegando os efeitos provenientes de variáveis omitidas para acontecimentos puramente aleatórios. Os dados em painel, pelo seu conteúdo relativo às dinâmicas individuais intertemporais, permitem separar um modelo de comportamentos individuais de um modelo do comportamento médio de um grupo de indivíduos. Assim, podemos considerar as usuais fontes de heterogeneidade – seccional e temporal, ou ambas – com distintas naturezas, fixa e aleatória. A maior parte dos resultados apresentados neste capítulo vêm de Hsiao [1986], secção 7.4.

Considere-se, então, o seguinte modelo genérico:

$$\begin{aligned} y^*_{it} &= \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it} \\ y_{it} &= \begin{cases} 1 \Leftarrow y^*_{it} > 0 \\ 0 \Leftarrow y^*_{it} \leq 0 \end{cases} \end{aligned} \tag{7.1}$$

Para não complicar a exposição, opta-se por abordar apenas os modelos com efeitos individuais, quer fixos, quer aleatórios, que será capturada pelo termo de perturbação $u_{it} = \alpha_i + v_{it}$. No caso dos efeitos fixos, temos $\mathrm{Var}(u_{it}|\alpha_i) = \mathrm{Var}(v_{it}) = \sigma_v^2$ e, no caso de efeitos aleatórios, temos $\mathrm{E}(\alpha_i) = \mathrm{E}(v_{it}) = 0$ e α_i e v_{it} independentes. Admite-se, ainda, que u_{it} terá uma distribuição F, que pode ser a logística (Λ) ou a normal reduzida (Φ).

Assim, a existência destes componentes não observáveis permite que indivíduos homogéneos, i.e., com as mesmas características observáveis (expressas pela sua função de reacção), sejam heterogéneos nas suas respostas.

i. Modelo de Efeitos Fixos

Tal como para os modelos convencionais, admite-se que os efeitos individuais α_i são não aleatórios, pelo que temos $\mathrm{Var}(u_{it}|\alpha_i) = \mathrm{Var}(v_{it}) = \sigma_v^2 = 1$ (por normalização). A estimação de $\theta = [\alpha_i\ ;\ \boldsymbol{\beta}]$ a partir de $\mathrm{Prob}(y_{it} = 1) = \mathrm{F}(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)$ é efectuada por MLE, que será consistente quando $T\rightarrow\infty$. No entanto, não é este o caso mais comum nos estudos com dados em painel, em que T é usualmente pequeno e, no caso de T fixo, o estimador MLE de a_i não é consistente, mesmo que $N \to \infty$: é o recorrente problema do parâmetro acidental (por oposição aos parâmetros estruturais β que não variam de indivíduo), cuja estimação não terá significado se for apenas avaliada pelas suas propriedades assimptóticas.

Resta, pois, estimar β: o problema que se coloca neste enquadramento é que a estimação deste não é independente da estimação de a_i, ou seja, a inconsistência de $\hat{a}_{i,MLE}$ é transmitida para $\hat{\beta}_{MLE}$, mesmo que $N \to \infty$, mas T fixo.[58]

Cumpre então encontrar condições para a existência de um estimador consistente de β. Um princípio geral será o de Neyman-Scott, segundo o qual é preciso encontrar funções $\Psi(\mathbf{y}_1,\ldots,\mathbf{y}_N \,|\beta)$ independentes de α e que convirjam em probabilidade para zero (com $N \to \infty$), quando β é avaliado no seu valor real. Assim, $\hat{\beta}$ é obtido resolvendo $\Psi(\mathbf{y}_1,\ldots,\mathbf{y}_N \,|\beta) = 0$, sendo consistente, sob determinadas condições de regularidade. Uma forma expedita de encontrar tal função para os modelos probit e logit é determinar

$$\mathrm{f}(\mathbf{y}_i \mid \beta, \tau_i) = \frac{\mathrm{f}(\mathrm{y}_i \mid \beta, \alpha_i)}{\mathrm{g}(\tau_i \mid \beta, \alpha_i)} \text{ , com } \mathrm{g}(\tau_i \mid \beta, \alpha_i) > 0 \qquad (7.2)$$

que não depende de a_i e, a partir daí, construir a função de densidade conjunta

$$\Psi(\mathbf{y}_1, \ldots, \mathbf{y}_N \mid \beta, \tau_1, \ldots, \tau_N) = \prod_{i=1}^{N} \mathrm{f}(\mathbf{y}_i \mid \beta, \tau_i) \qquad (7.3)$$

cuja maximização resultará num estimador consistente de β. A demonstração da existência de um estimador consistente de β pode ser encontrada em Hsiao [1986],162-3 para uma especificação logit. No entanto, o mesmo autor alerta para o facto de não parecer existir um estimador consistente para o modelo probit. Assim, atendendo ao facto de ser possível encontrar diversos estimadores consistentes de β no caso logit, através de algumas simples transformações (que passam pela maximização de uma função de verosimilhança condicionada por uma qualquer estatística suficiente mínima de y_{it}, como Σy_{it}) e de, nos modelos univariados, os resultados do logit não diferirem muito dos do probit[59], tudo aponta para que se favoreça a especificação logit em detrimento da probit.

---

[58] Veja-se uma prova em Hsiao[1986],160.

[59] Podendo ser convertidos pela relação de Amemiya: $\hat{\beta}_L \cong 1.6\hat{\beta}_P$

Já no que toca aos modelos de regressão múltipla, os resultados podem variar muito, consoante sejam obtidos por logit ou probit, daí que não seja despicienda a questão da escolha da especificação. Ainda que a evidência não seja muito sólida, os estudos de Monte Carlo realizados para o modelo probit apontam para um enviesamento relativamente pequeno (sempre inferior a 10%) e que será tanto menor, quanto menor a escala de variação dos efeitos fixos (Hsiao [1986]).

ii. <u>Modelo de Efeitos Aleatórios</u>

Assume-se um modelo em que os efeitos individuais a_i são aleatórios, mas independentes de $\mathbf{x}_i$, tendo uma distribuição **H**, condicionada por um conjunto finito de parâmetros δ. A função logarítmica de verosimilhança será:

$$\ln L = \sum_{i=1}^{N} \ln \int \prod \mathrm{F}(\beta' \mathbf{x}_{it} + a)^{y_{it}} \left[1 - \mathrm{F}(\beta' \mathbf{x}_{it} + a)\right]^{1-y_{it}} \mathrm{d}\,\mathrm{H}(a \mid \delta) \qquad (7.4)$$

em que F(.) é a distribuição marginal de y em a. Como ln$L$ é uma função de um conjunto $[\beta', \delta']$ finito de parâmetros, sob algumas condições de regularidade, a sua maximização resultará em estimadores consistentes de β e δ, quando $N \to \infty$.

Um modelo mais geral considerará uma eventual correlação entre a_i e $\mathbf{x}_i$ (pelo que o procedimento anterior, ao omitir a_i, não será consistente), fazendo $a_i = \alpha' \mathbf{x}_i + h_i$, com $\alpha' = (\alpha'_1, \ldots, \alpha'_T)$ e $\mathbf{x}'_i = (\mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{iT})$, sendo $\mathrm{E}(a_i | \mathbf{x}_i)$ linear e h_i independente de $\mathbf{x}_i$ e com função de distribuição $\mathbf{H}^*$. A função de verosimilhança é

$$\ln L = \sum_{i=1}^{N} \ln \int \prod \mathrm{F}(\beta' \mathbf{x}_{it} + \alpha' \mathbf{x}_i + h)^{y_{it}} \left[1 - \mathrm{F}(\beta' \mathbf{x}_{it} + \alpha' \mathbf{x}_i + h)\right]^{1-y_{it}} \mathrm{d}\,\mathrm{H}^*(h) \qquad (7.5)$$

e, caso se pressuponha que **F** é a normal reduzida e $\mathbf{H}^*$ uma distribuição normal de média zero e variância σ^2_h, resultará numa especificação probit multivariada: $y_{it} = 1 \Leftarrow \beta' \mathbf{x}_{it} + \alpha' \mathbf{x}_i + h_i + u_{it} > 0$ e $(\mathbf{u}_i + \mathbf{i} h_i) \sim \mathbf{N}(\mathbf{0}, \mathbf{I}_T + \sigma^2_h \mathbf{ii}')$. Os estimadores resultantes serão consistentes.

7.2 Modelos Dinâmicos de Escolha Binária

Admite-se, agora, que a probabilidade de um indivíduo alterar ou manter um dado estado não é independente da sua experiência passada. Um modelo de análise das relações intertemporais entre variáveis discretas pode ser especificado da seguinte forma:

$$\begin{aligned} & y_{it}^{*} = \boldsymbol{\beta}'\mathbf{x}_{it} + \sum_{p=1}^{t-1} \gamma_p y_{i,t-p} + \phi \sum_{s=1}^{t-1} \prod_{p=1}^{s} y_{i,t-p} + u_{it} \\ & \begin{cases} y_{it} = 1 \Leftarrow y_{it}^{*} > 0 \\ y_{it} = 0 \Leftarrow y_{it}^{*} \leq 0 \end{cases} \end{aligned} \tag{7.6}$$

em que u_{it} é independente de $\mathbf{x}_{it}$ e independentemente distribuído sobre i, com matriz de variâncias e covariâncias intertemporal $E(u_i u'_i) = \boldsymbol{\Omega}$ e γ_p o coeficiente de resposta de y^{*}_{it} à experiência ocorrida p períodos atrás. A potencial heterogeneidade dos indivíduos é permitida através de $u_{it} = \alpha_i + v_{it}$.

Os modelos de escolha binária dinâmicos suscitam, de imediato, duas questões:

- como tratar as observações iniciais y_{i0};
- como distinguir uma real dependência intertemporal das decisões de uma dependência meramente espúria.

A primeira questão importa para a derivação de estimadores consistentes e a segunda vai retirar a sua importância do facto de um dado acontecimento observado poder ter ocorrido devido a uma dada experiência passada que alterou um comportamento individual ou devido a uma sucessão de perturbações não observáveis e correlacionadas ao longo do tempo (i.e., a um puro acaso) ou, ainda, devido a uma combinação das duas.

i. <u>Condições iniciais (ou história pré-amostral):</u>

Para simplificar a exposição, considere-se um modelo AR(1) com termo independente, i.e., que $\boldsymbol{\beta}'\mathbf{x}_{it} = \beta_0$, $p = 1$ e $\phi = 0$, sobre o qual se assume, alternativamente: (1) y_{i0} é exógeno; (2) o processo está em equilíbrio.

(1) implica que a função de probabilidade conjunta de $y'_i = \{y_{i1},\ldots,y_{iT}\}$, condicionada por α_i, seja

$$\prod_{t=1}^{T} F(y_{it} \mid y_{i,t-1}, \alpha_i) = \prod_{t=1}^{T} \Phi(\beta_0 + \gamma y_{i,t-1} + \alpha_i)(2y_{it} - 1) \tag{7.7}$$

em que Φ é a função de distribuição de probabilidade normal.

(2) implica que a probabilidade marginal limite de y_{it} = 1 e a função de probabilidade conjunta de $\mathbf{y}'_i$, condicionadas por α_i, sejam, respectivamente

$$
\begin{aligned}
&P_i = \frac{\Phi(\beta_0 + \alpha_i)}{1 - \Phi(\beta_0 + \gamma + \alpha_i) + \Phi(\beta_0 + \alpha_i)} \text{ e} \\
&\prod_{t=1}^{T} \mathrm{F}(y_{it} \mid y_{i,t-1}, \alpha_i) = \prod_{t=1}^{T} \left\{\Phi(\beta_0 + \gamma y_{i,t-1} + \alpha_i)(2y_{it} - 1)\right\} P_i^{y_{i0}} (1 - P_i)^{1-y_{i0}}
\end{aligned}
\tag{7.8}
$$

Caso α_i seja aleatório com distribuição $\mathbf{G}(\alpha)$, as funções de verosimilhança sob as hipóteses (1) e (2) são, respectivamente, segundo Hsiao [1986]:

$$
\begin{aligned}
&L = \prod_{i=1}^{N} \int \prod_{t=1}^{T} \left\{\Phi(\beta_0 + \gamma y_{i,t-1} + \alpha)(2y_{it} - 1)\right\} d\,\mathrm{G}(\alpha) \text{ e} \\
&L = \prod_{i=1}^{N} \int \prod_{t=1}^{T} \left\{\Phi(\beta_0 + \gamma y_{i,t-1} + \alpha)(2y_{it} - 1)\right\} P_i^{y_{i0}} (1 - P_i)^{1-y_{i0}}\, d\,\mathrm{G}(\alpha)
\end{aligned}
\tag{7.9}
$$

Quando α_i é aleatório, os estimadores MLE de β_0, γ e σ_α^2 são consistentes, desde que $N \to \infty$. Na especificação com efeitos fixos, os estimadores MLE só serão consistentes no caso semi-assimptótico $T \to \infty$, sendo o enviesamento grande, quando T é fixo, de acordo com a evidência empírica existente[60].

No entanto, o pressuposto de que y_{i0} seja exógeno não é válido se as perturbações forem autocorrelacionadas e/ou o processo em curso se tiver iniciado antes do início da amostra. Também o pressuposto do estado de equilíbrio é inaceitável quando o processo estocástico é influenciado por variáveis exógenas temporais.

Com y_{i0} endógeno, as funções de verosimilhança para os modelos de efeitos fixos e aleatórios são, respectivamente:

$$
\begin{aligned}
&L = \prod_{i=1}^{N} \prod_{t=1}^{T} \left\{\Phi(\beta_0 + \gamma y_{i,t-1} + \alpha_i)(2y_{it} - 1)\right\} \mathrm{f}(y_{i0} \mid \alpha_i) \text{ e} \\
&L = \prod_{i=1}^{N} \int \prod_{t=1}^{T} \left\{\Phi(\beta_0 + \gamma y_{i,t-1} + \alpha)(2y_{it} - 1)\right\} \mathrm{f}(y_{i0} \mid \alpha)\, d\,\mathrm{G}(\alpha)
\end{aligned}
\tag{7.10}
$$

em que $\mathrm{f}(y_{i0}|\alpha_i)$ é a probabilidade marginal de y_{i0} condicionada por α_i. O problema é que a maximização de qualquer uma das referidas funções de verosimilhança é relativamente complexa, pelo que foi proposta uma aproximação (que Hsiao[1986] atribui a Heckman), com resultados bastantes aceitáveis e que consiste em:

---

[60] Hsiao [1986], pág. 170

1. aproximar a probabilidade de y_{i0} através de um modelo probit $y^*_{i0} = \mathbf{Q}(\mathbf{x}_{it}) + e_{it}$, com $y_{i0} = 1$ se $y^*_{i0} > 0$, $y_{i0} = 0$ se $y^*_{i0} \leq 0$ e $e_{it} \sim N(0,1)$;
2. permitir que e_{it} seja livremente correlacionado com u_{it}, $t = 1,\ldots,T$;

   estimar o modelo (7.6) sem restrições entre os parâmetros do modelo, na forma

   y_{i0}

ii. Dependência intertemporal:

A segunda questão, da dependência intertemporal com sentido causal vs. uma

que as experiências passadas afectam de forma determinante as experiências futuras porque exercem um efeito comportamental: um indivíduo que tenha experimentado uma

desempregado no futuro do que um indivíduo idêntico em tudo o resto, mas que não tenha estado desempregado. A segunda diz-nos que tal relação é apenas aparente porque a

proxy para uma heterogeneidade que radica em

situação de dependência temporal, uma vez que um indivíduo experimentou um dado acontecimento, as suas preferências alteram-se de tal forma que se comportará de forma

mesmo acontecimento no passado. No caso de dependência espúria, a experiência não tem qualquer efeito na probabilidade de um indivíduo experimentar um acontecimento no futuro, mas, na medida em que os indivíduos revelam uma heterogeneidade que não pode ser representa através de variáveis observáveis, a inclusão da variável dependente desfasada como variável explicativa (i.e., a experiência passada) será uma *proxy* para essas variáveis seccionais não observáveis.

A importância prática desta questão é maior do que pode parecer à primeira vista e pode ser exemplificada da seguinte forma: perante um problema de política de emprego, se considerarmos válida a dependência intertemporal, estamos a sancionar uma política activa de diminuição do desemprego no curto prazo pois, as políticas conjunturais de estímulo da procura agregada provocarão uma redução da taxa de desemprego de longo prazo; já se considerarmos que a dependência é espúria, as políticas de curto prazo não terão qualquer efeito sobre a taxa de desemprego estrutural. Assim, torna-se crucial testar a hipótese de dependência intertemporal real vs. espúria, o que será apesar de tudo complicado, quando existe heterogeneidade (efeitos individuais). De facto, se o modelo contiver efeitos individuais aleatórios, mesmo que não haja dependência intertemporal, teremos P(y_{it} | $\mathbf{x}_{it}$, $y_{i,t-1}$) ≠ P(y_{it} | $\mathbf{x}_{it}$). Só no caso de não existir heterogeneidade é que a inexistência de dependência será equivalente a P(y_{it} | $_{it}$, $_{i,t-1}$) = P($_{it}$ | $_{it}$).

[1995],187, consiste em adoptar o modelo P($_{it}$ = 1 | $\mathbf{x}_{it}$, $y_{i,t-1}$) = F($\mathbf{x}'_{it}\boldsymbol{\beta}$+ $\gamma y_{i,t-1}$) e testar H_0: $\gamma = 0$ usando os habituais testes LM. Se não rejeitarmos H_0, ignora-se a questão da heterogeneidade e prossegue-se com a estimação pelos métodos convencionalmente utilizados com amostras seccionais. Se rejeitarmos H_o, temos que apurar se estamos perante dependência intertemporal real ou heterogeneidade: uma forma de o fazer é prosseguir com o teste, mas condicionado pelos efeitos individuais. Ou seja, desde que os termos de perturbação v_{it} não estejam autocorrelacionados (no modelo condicionado por a_i), pode-se testar $\gamma = 0$, no modelo P(y_{it} = 1 | $\mathbf{x}_{it}$, $y_{i,t-1}$,a_i) = F($\mathbf{x}'_{it}\boldsymbol{\beta}$+ $\gamma y_{i,t-1}$+a_i). Quando subsiste autocorrelação no modelo condicionado pelos efeitos individuais, sugere-se um método avançado por Chamberlain [1984] e que consiste em testar, para o modelo condicionado pelos efeitos individuais a_i:

H_0: P(y_{it} = 1 | $\mathbf{x}_{it}$, $y_{i,t-1}$,a_i) = F($\mathbf{x}'_{it}\boldsymbol{\beta}+\Sigma_p\mathbf{x}'_{t-p}\boldsymbol{\delta}_p$ + $\gamma y_{i,t-1}$+a_i) ⇒ ausência de dependência intertemporal;

H_1: P(y_{it} = 1 | $\mathbf{x}_{it}$, $y_{i,t-1}$,a_i) ≠ F($\mathbf{x}'_{it}\boldsymbol{\beta}+\Sigma_p\mathbf{x}'_{t-p}\boldsymbol{\delta}_p$ + $\gamma y_{i,t-1}$+a_i) ⇒ existência de dependência intertemporal.

Ainda que seja uma forma eficiente de distinguir a existência de dependência intertemporal da heterogeneidade e autocorrelação, o teste não permite avançar quanto às diferentes formas dos referidos fenómenos, o que exigiria especificações mais pormenorizadas.

8. Miscelânea de Temas

O objectivo da presente secção é apresentar uma pequena revisão, não apenas de alguns dos artigos mais importantes, mesmo fundadores, da análise econométrica de dados em painel, mas também de alguns artigos mais recentes e que poderão fornecer pistas quanto à evolução futura do tema. O capítulo termina com um apanhado breve de alguns testes de especificação existentes para modelos que usem dados em painel.

8.1 ALGUNS ARTIGOS FUNDAMENTAIS

Em Balestra e Nerlove [1966], artigo já citado acima, é desenvolvido um modelo dinâmico para a procura de gás natural, discutida a sua especificação e avançados alguns métodos de estimação. Trata-se de um artigo fundamental e, de certa forma, pioneiro na sua abordagem dos métodos de estimação IV em modelos com efeitos aleatórios. Os autores começam por justificar a utilização de um modelo dinâmico para a especificação de uma equação de procura de gás natural pois, esta depende do *stock* de equipamento consumidor de gás. Assim sendo, propõem uma função de procura de gás em que o consumo total de gás é a variável dependente e usam como variáveis explicativas: o preço relativo do gás, a variável dependente desfasada, a população e o rendimento per capita.

Confrontados com resultados de estimação OLS não conformes com a teoria económica, nomeadamente quanto ao parâmetro associado ao termo autorregressivo (aproximadamente 1), os autores levantam a questão de tal ser resultado de uma má especificação pela omissão de uma eventual heterogeneidade seccional, cujo efeito seria captado pela variável dependente desfasada. Adoptando uma especificação com efeitos aleatórios, Balestra e Nerlove [1966] notaram que os estimadores GLS ou FGLS convencionais não eram aplicáveis na presença da variável autorregressiva.

Assim, os autores sugerem um método IV que use como variáveis instrumentais as variáveis exógenas desfasadas, uma vez que estas cumprem o princípio de Theil para as variáveis instrumentais, segundo o qual devem ser escolhidas as combinações lineares de variáveis exógenas mais fortemente correlacionadas com a variável pré-determinada que vão substituir.

Nerlove [1971], constatando por um lado a importância de tirar partido das vantagens dos dados em painel no estudo das dinâmicas comportamentais e, por outro, o problema colocado à estimação pela pouca profundidade seccional da maior parte das amostras temporais, estuda o enviesamento dos métodos OLS, LSDV, GLS, FGLS, IV e MLE (que evidencia uma tendência para fornecer soluções de canto, levando a um estimador inconsistente da matriz de covariâncias dos termos de perturbação). Apesar de tudo e com base em estudos de Monte Carlo, o autor dá preferência ao método FGLS (com d estimado por OLS) pelo seu pequeno enviesamento e menor erro quadrático médio e chama a atenção para os problemas do MLE.

Nickell [1981] deriva a expressão analítica do enviesamento assimptótico do estimador LSDV no modelo dinâmico de efeitos fixos, quando *T* é fixo e conclui que este pode ser muito significativo, ainda que a introdução de variáveis exógenas o diminua, sem nunca o eliminar. No entanto, o enviesamento diminui e tende para zero, quando *T* aumenta, mas será tanto maior quanto mais próximo da unidade estiver o parâmetro associado à variável dependente desfasada.

Bhargava e Sargan [1983] advogam a utilização de métodos de estimação de equações simultâneas nos modelos dinâmicos com efeitos aleatórios, especialmente os métodos de máxima verosimilhança com informação limitada (LIML), comparando as diversas hipóteses para a primeira observação (exógena ou endógena).

Anderson e Hsiao [1982] exploram a importância das hipóteses relativas às primeiras observações, não apenas para as propriedades assimptóticas dos estimadores, mas também para a própria interpretação das estimativas. Outra questão abordada é da dependência intertemporal vs. autocorrelação e avançam com um método de estimação MLE de modelos dinâmicos com autocorrelação. Como os métodos MLE são relativamente difíceis de implementar, os autores sugerem um método GLS (ou FGLS), que lhe é assimptoticamente equivalente.

8.2 TEMAS RECENTES SOBRE ESTIMAÇÃO COM DADOS EM PAINEL

Em termos de literatura mais recente, citam-se desde logo quatro artigos (*working papers*) que, admite-se, possam indiciar futuros desenvolvimentos.

Assim, Judson e Owen [1996] exploram os problemas específicos à estimação de modelos macroeconómicos com dados em painel, tipicamente com *N* pequeno e *T* grande e numa especificação com efeitos fixos, concluindo que o melhor método de estimação varia muito consoante a dimensão do painel. De acordo com as simulações realizadas, as autoras avançam que o enviesamento do LSDV não é de forma alguma irrelevante quando *N* é fixo: o aumento da dimensão temporal parece reduzir pouco o enviesamento.

O estudo justifica ainda a escolha de uma especificação com efeitos fixos por duas ordens de razões: (1) porque se os efeitos individuais representam variáveis omitidas, é muito provável que estas se encontrem correlacionadas com os restantes regressores; (2) como os painéis na macroeconomia, apesar da pequena dimensão seccional, incluem muito provavelmente a maior parte dos indivíduos da população, a hipótese de serem uma amostra aleatória de uma população de países muito maior não tem cabimento. O objectivo acaba por ser a comparação de diversos estimadores disponíveis para esta especificação – LSDV, Anderson-Hsiao, Arellano-Bond e OLS – concluindo-se que:

- o estimador OLS apresenta-se fortemente enviesado, particularmente para o parâmetro associado à variável dependente desfasada (ainda que, ao contrário dos restantes estimadores, apresente um enviesamento significativo para β), mesmo quando *T* aumenta;
- o LSDV comporta-se exactamente da maneira prevista por Nickell [1981], ainda que não aumente de eficiência à medida que *T* aumenta. Assim, as autoras defendem a utilização de estimadores com melhor relação enviesamento-eficiência;
- é sugerido um método LSDV corrigido pelo enviesamento (através de uma fórmula análoga à apresentada por Nickell [1981]) e que se apresenta como o mais eficiente;
- O estimador Anderson-Hsiao é considerado o melhor, do ponto de vista do enviesamento médio, ainda que seja pouco eficiente para amostras com *T* pequeno.

Como a eficiência do estimador Anderson-Hsiao aumenta sensivelmente com *T* e devido à sua simplicidade, o estudo sugere que este seja aplicado para amostras grandes (T>10), preferindo-se o LSDV corrigido para amostras pequenas. Os estimadores GMM, por serem relativamente exigentes e não darem provas de grande eficiência, são preteridos.

Granger e Huang [1997] discutem a avaliação e comparação da qualidade das diversas especificações de modelos com amostras longitudinais através de técnicas usadas em séries temporais. A metodologia consiste em comparar as previsões (tanto no sentido temporal – *forecasting* – como no seccional – *prediction*) com os dados reais: um método simples consiste num teste de hipóteses H_0: $b_0=0; b_1=1$ sobre a regressão de

$y_{it}=b_0+b_1E^*(y_{it})+w_{it}$, em que $E^*(y_{it})$ é a previsão de y_{it} (*prediction* e *forecast*) e deve ser w_{it} *white noise*.

Finalmente, Kao e Chiang [1997] investigam as questões relacionadas com a não-estacionaridade de séries temporais usadas em estudos com dados em painel, nomeadamente as formas de analisar as propriedades das séries com raízes unitárias, os problemas de regressões espúrias e a cointegração em amostras longitudinais. Os autores sublinham que existe pouco trabalho teórico sobre o tema da não-estacionaridade, em geral e sobre a questão da cointegração, em particular. Numa análise das propriedades de três estimadores num modelo com cointegração – OLS, OLS modificado (pelo enviesamento, de forma análoga a Judson e Owen[1996]) e OLS dinâmico (i.e., estimação por OLS do modelo base aumentado com as primeiras diferenças passadas e futuras da variável exógena) – os autores concluem pela superioridade deste último em termos de centricidade.

Kao [1999] estuda a questão das regressões espúrias com dados em painel e analisa as propriedades assimptóticas do estimador *within*, bem como de outras estatísticas convencionais. Assim, Kao [1999] conclui que a distribuição assimptótica do estimador LSDV é diferente da de estimadores equivalentes no contexto das *time series* com regressões espúrias, pelo que os testes de cointegração convencionais terão que ser adaptados ao contexto dos dados em painel. Outra referência a consultar, dentro do tema da cointegração com dados em painel, é Larsson e Lyhagen [1999].

Um tema que também tem suscitado numerosos artigos é o da estimação de modelos dinâmicos recorrendo aos estimadores *between* e à chamada técnica da *pooled mean regression* proposta por Pesaran e Smith[61] e que de certa forma se relaciona com a questão da cointegração. Já Baltagi e Griffin [1984] haviam discutido a questão da estimação de multiplicadores de curto e de longo prazo, segundo a noção de que o recurso a amostras temporais tenderia a captar os primeiros e as seccionais os segundos. Baltagi e Griffin [1984] sugeriram que se usasse o estimador *within* ou o GLS (consoante se assumissem efeitos fixos ou aleatórios) para estimar os efeitos de curto prazo e os estimadores OLS ou *between* para os efeitos de longo prazo, caso os resultados produzidos

[61] Pesaran, M. Hashem e Ron Smith [1993]. “Alternative Approaches to Estimating Long-Run Energy Demand Elasticities: An Application to Asian Developing Countries”, University of Cambridge Department of Applied Economics Working Paper: 9308.

pelos primeiros se afastem consideravelmente dos fornecidos pelos estimador *between*. Para uma actualização desta discussão, veja-se Pirotte [1999].

Baltagi e Griffin [1997] efectuam uma comparação entre estimadores que assumem a homogeneidade dos parâmetros (i.e., a *poolability* dos dados) e outros que consideram que os parâmetros são heterogéneos, pelo que estimam cada indivíduo separadamente e depois recorrem a médias, eventualmente ponderadas pelas respectivas matrizes de variâncias (*shrinkage estimators*). Os resultados tendem a confirmar a superioridade dos estimadores convencionais, em que se contavam os LSDV, GLS e 2SLS. Além disso, Baltagi e Griffin [1997] notam que, para uma dimensão temporal considerável, a implementação dos estimadores LSDV ou GLS pode ser preferível à utilização de estimadores IV, sobretudo quando o conteúdo informativo dos instrumentos é pequeno.

Pesaran, Shin e Smith [1998] propõem uma variante do estimador de Grupos de Médias (MG), para painéis de dimensões consideráveis, quer em N, quer em T. Em vez de estimar N regressões e calcular as médias dos coeficientes, sugerem um meio termo entre esta metodologia e os estimadores com o pressuposto de homogeneidade. Assim, Pesaran *et al.* [1998] restringem os parâmetros de longo prazo a serem idênticos, enquanto que os de curto prazo podem variar de indivíduo para indivíduo. A estimação, por MLE, requer que se usem algoritmos iterativos, nomeadamente do tipo Newton-Raphson, podendo ser aplicado o mesmo procedimento, independentemente da ordem de integração dos regressores. Estes estimadores *Pooled Mean Group* (PMG) são promissores, tanto mais que podem ser justificados de ponto de vista económico, sobretudo na macroeconometria, ainda que a sua utilização (tal como a do estimador LSDV) requeira uma dimensão temporal da amostra considerável. O estimador PMG permite que as dinâmicas de ajustamento individuais sejam diferenciadas (tal como as estruturas de variâncias e covariâncias), mantendo uma coesão no sistema para o longo prazo, ainda que nem sempre as estimativas de curto prazo sejam plausíveis.

8.3 ALGUNS TESTES DE ESPECIFICAÇÃO

Para aferir a robustez dos resultados de qualquer trabalho empírico, devem ser efectuados testes à existência de autocorrelação e heteroscedasticidade dos termos de perturbação aleatórios, para o que se sugerem alguns procedimentos. Para o efeito, considere-se o seguinte modelo:

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it}\text{, com } u_{it} \sim N(0,\sigma^2_{it}). \qquad (8.1)$$

em que u_{it} tem média nula, variância σ^2_{it}, não necessariamente constante e covariâncias não necessariamente nulas. Para simplificar a apresentação, admite-se que os efeitos específicos são apenas de dimensão seccional (ainda que a inclusão de efeitos temporais não constitua uma complicação intransponível) e são tratados como fixos.

– **testes à autocorrelação:**

Para testar a ausência de autocorrelação dos termos de perturbação recomendam-se alguns testes:

i) **teste LM de Breusch-Godfrey** sobre os resíduos de estimação por LSDV ou *Within* (quando consistentes) com um *truncation lag p* a ser definido pelo critério de Schwert e que consiste na estimação de uma equação auxiliar com os resíduos da equação principal como variável dependente e com as variáveis explicativas do modelo original e os resíduos do mesmo modelo desfasados, seguida de um teste à significância destes últimos mediante estatística $nR^2 \sim \chi^2(p)$. O teste tem a vantagem de ser relativamente robusto face a especificações alternativas para o processo de correlação, testando processos estocásticos AR(p) e MA(p), sendo o seu principal inconveniente o facto de a sua potência depender crucialmente da escolha do *truncation lag*. Para mais pormenores do teste consulte-se, por exemplo, Davidson e MacKinnon [1993], 357-360. Ainda assim e por o teste não ter sido concebido para dados em painel, aconselha prudência na sua utilização, complementando-o com testes alternativos.

ii) **teste LM5 de Baltagi [1995]** (pág. 93) que testa a hipótese nula de ausência de autocorrelação AR(1) ou MA(1) num modelo com efeitos fixos através de um teste de multiplicador de Lagrange com estatística

$$LM5 = NT^2/(T-1)\cdot\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}_{-1}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \sim N(0,1) \text{ para } T \to \infty.$$

em que $\hat{\mathbf{u}}$ representa um vector de resíduos de estimação *within*, admitindo efeitos fixos (pelo que o teste não será apropriado para testar, por exemplo, a existência de autocorrelação em modelos *pooled OLS*). O teste tem a vantagem de se comportar relativamente bem para outras especificações ARMA para as perturbações e por ser aplicável para modelos dinâmicos ou outros modelos com regressores não estritamente exógenos.

Adicionalmente, como a transformação *within* elimina os efeitos individuais (ou temporais), sejam estes tidos como fixos ou aleatórios, o teste também pode ser usado para testar a existência de autocorrelação da primeira ordem em modelos com efeitos aleatórios (Baltagi [1995], pág. 94), ainda que não seja possível efectuar esta extensão no caso de modelos dinâmicos[62] (por a transformação *within* não gerar estimadores consistentes, nessa situação).

iii) **Teste de Durbin-Watson com dados em painel:** para um modelo com efeitos fixos e com regressores não aleatórios, é possível usar-se o teste de Durbin-Watson para testar-se a existência de autocorrelação gerada por um termo AR(1). A estatística é idêntica à usada para séries temporais, com a diferença de os resíduos usados serem os *within* e não os OLS, ou seja

$$DW_p = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{u}_{it} - \hat{u}_{i,t-1}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2} \tag{8.2}$$

Bhargava, Franzini e Narendranathan [1982] apresentam os limites superiores e inferiores para o valor crítico da estatística.

iv) **Teste $\mathbf{m_2}$ de Arellano e Bond [1991] para processos MA(2):** o teste parte de um estimador GMM para um modelo autorregressivo com efeitos aleatórios (com N grande e T pequeno) equivalente a (3-9) e genericamente dado por

$$\hat{\boldsymbol{\theta}} = \left(\Delta\mathbf{X}'\mathbf{Z}\mathbf{A}_\mathbf{N}\mathbf{Z}'\Delta\mathbf{X}\right)^{-1}\left(\Delta\mathbf{X}'\mathbf{Z}\mathbf{A}_\mathbf{N}\mathbf{Z}'\Delta\mathbf{Y}\right) \tag{8.3}$$

em que $\mathbf{Z}$ é uma matriz de instrumentos, $\Delta\mathbf{X}$ a matriz de variáveis explicativas nas primeiras diferenças ou nas diferenças ortogonais e $\mathbf{A_N} = 1/N\sum_i(\mathbf{z}_i'\ddot{\mathbf{u}}_i\,\ddot{\mathbf{u}}_i'\mathbf{z}_i)$ uma matriz de ponderação obtida a partir dos resíduos de estimação ($\ddot{\mathbf{u}}$) resultantes de um estimador inicial consistente. Com os resíduos de estimação GMM, $\hat{\mathbf{u}}_i$, calcula-se a estatística do teste, para a hipótese nula de ausência de autocorrelação da segunda ordem:

$$m_2 = \frac{\hat{\mathbf{u}}_{-2}\hat{\mathbf{u}}_*}{\hat{\mathbf{u}}^{1/2}} \overset{a}{\sim} N(0,1)$$

em que $\hat{\mathbf{u}}_*$ é um vector dos resíduos "aparado" para que a sua dimensão coincida com a de $\hat{\mathbf{u}}_{-2}$. Para pormenores sobre o teste, consulte-se Arellano e Bond [1991], 282.

[62] Para modelos estáticos, com efeitos aleatórios, Baltagi aconselha uma outra estatística, dita LM4 (pág. 91-2) ou o uso do teste de Durbin-Watson.

Baltagi [1995], Capítulo 5, constitui uma boa fonte para exemplos adicionais de testes à existência da autocorrelação em circunstâncias diversas, nomeadamente quando se pretende testar simultaneamente a hipótese de autocorrelação e da existência de efeitos aleatórios ou quando se pretende escolher entre um processo AR(1) ou MA(1).

– **testes à heteroscedasticidade:**

O teste de White com termos cruzados será uma proposta óbvia para testar a existência de heteroscedasticidade genérica, segunda a qual, a cada observação corresponde uma variância distinta. Assim, testa-se a hipótese nula de homoscedasticidade contra a alternativa de heteroscedasticidade genérica, incidindo o teste sobre os resíduos de estimação *Within*. Mais uma vez, um dos problemas do teste é o de ter uma reduzida potência, a que se soma o facto de não ter sido concebido especificamente para dados em painel.

Pode ainda testar-se a hipótese nula de homoscedasticidade com hipótese alternativa de heteroscedasticidade em bloco, i.e., em que podemos ter termos de perturbação homoscedásticos para cada i, mas com variâncias diferentes de grupo para grupo. Formalmente, a hipótese nula será

$$H_0: \sigma^2_{i,t}=\sigma^2_i\ , \ \forall\ t=1,\dots T,\ i=1,\dots,N$$

com hipótese alternativa de heteroscedasticidade ao nível das observações individuais:

$$H_1: \exists\ \sigma^2_{i,t}=\sigma^2_i\ ;\ i=1\dots,N; t=1,\dots,T$$

Uma sugestão para o teste será o de estimar o modelo com efeitos fixos e recolher os resíduos para cada um dos indivíduos, correndo N regressões auxiliares do teste de White. Em concreto, estima-se o modelo (8.1), por exemplo e adicionalmente as seguintes N regressões auxiliares:

$$e_{it}^2=\gamma_0+\gamma_1\overline{y_{i,t-1}}+\gamma_2\overline{y_{i,t-1}}^2+\lambda\,\overline{\mathbf{x}}_{it}+\theta\,\overline{\mathbf{x}}_{it}\overline{\mathbf{x}}_{it}'+\zeta_{it}\quad ,\ i=1,\dots,N$$

em que a barra significa a variável expressa na forma de diferença face à respectiva média seccional e λ e θ são $(1\times k)$, omitindo-se os termos cruzados, para simplificar. Para cada uma das regressões calcula-se a estatística do teste e o respectivo *p-value*, dados por:

$$TR^2 \sim \chi^2(J-1)$$

em que J é o número de parâmetros em cada regressão. Com os N *p-values* dos referidos testes individuais, procede-se a um teste de Maddala-Fisher para testes independentes (Maddala [1977], pág. 47-8 e Maddala e Kim [1998]) que se baseia na estatística

$$P_\lambda = -2\sum_{i=1}^{N}\ln(p_i) \sim \chi^2(2N)$$

e que, segundo Maddala [1977] se trata de uma distribuição exacta para um teste de hipóteses independentes. Maddala e Kim [1998] sugerem este teste para testar a hipótese nula de estacionariedade das séries com dados em painel como alternativa aos testes de Levin e Lin [1993] e Im, Pesaran e Shin [1996][63]. A sua aplicação a este contexto de heteroscedasticidade foi implementada em Marques [2000]. Ainda assim, a potência do teste, segundo Maddala [1977] dependerá da potência dos teste individuais subjacentes, que para o teste geral de White não será particularmente elevada, ainda que a consistência esteja assegurada, desde que N fixo e $T \to \infty$.

Caso se rejeite a hipótese nula de homoscedasticidade, justifica-se a correcção da matriz de covariâncias dos estimadores (*within*, por exemplo) pelo método consistente de White, com ou sem correcção de grupo, dependendo da hipótese alternativa prever homoscedasticidade intra-grupo, ou não, respectivamente.

Um outro teste proposto por Baltagi [1995] é o já conhecido teste de Bartlett em que se verifica sob a hipótese nula de homoscedasticidade

$$u = \sum_{i=1}^{N}\left(\frac{\frac{1}{T}\sum_{t=1}^{T}\left(y_{it}-\bar{y}_i\right)^2}{\frac{1}{NT}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\left(y_{it}-\bar{y}_i\right)^2\right]}\right)^{T/2} \sim \chi^2(N-1) \qquad (8.4)$$

contra a hipótese alternativa de heteroscedasticidade em bloco. Pode e deve-se usar a estatística de Bartlett modificada, dada por:

$$M = \frac{(T-N)\ln\hat{\sigma}^2 - \sum_{i=1}^{N}(T-1)\ln\hat{\sigma}_i^2}{1+[1/3(N-1)][N/(T-1)-1/(T-N)]} \sim \chi(N-1) \qquad (8.5)$$

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(y_{it}-\bar{y}_i\right)^2 \quad e \quad \hat{\sigma}^2 = \frac{N(T-1)}{T-N}\hat{\sigma}_i^2$$

[63] Cuja aplicação vem dificultada pela indisponibilidade de valores exactos para os *p-values* dos testes ADF convencionais, como já se referiu no capítulo 6.

e que, tal como a anterior, pode ser alterada para acomodar painéis não balanceados (isto é, com dimensão temporal variável). Para mais pormenores sobre o teste consulte-se Judge *et al.* [1985], pág. 447-8 ou Gujarati [1996].

A escolha de uma dada especificação para um modelo, tema não abordado ao longo deste trabalho, mas que constitui matéria para futuras actualizações, irá depender da resposta prévia (ao trabalho de estimação) a uma séria de pressupostos de índole teórica e não empírica. Assim, há que ter em conta: a existência ou não heterogeneidade; o estarmos perante modelos de efeitos fixos ou aleatórios; as perturbações não-específicas (v_{it}) serem ou não esféricas e as variáveis exógenas (**X**) serem ou não estritamente exógenas.

A escolha pode ser assistida, nalguns casos, pelo recurso ao teste de especificação de Hausman (ou outro nele baseado), por exemplo: para testarmos a hipótese de autocorrelação dos termos de perturbação genéricos, podemos testar H_0 comparando os resultados de uma estimação MLE (consistente sob H_0, mas não sob H_1) com os obtidos com um estimador IV, p.ex., o Balestra-Nerlove (consistente sob ambas as hipóteses). Sobre isto, veja-se, entre outros, Baltagi [1995] ou Mátyás e Sevestre [1992].

Em Metcalf [1997] podem encontrar-se mais alguns testes de especificação para a estimação de modelos com dados em painel usando variáveis instrumentais.

9. CONCLUSÃO

Um dos primeiros passos na escolha do método de estimação com dados em painel é a definição da forma como é introduzida a heterogeneidade no modelos. Regra geral esta é apresentada através de termos independentes variáveis de indivíduo para indivíduo ou ao longo do tempo (ou em ambas as dimensões), sob a forma de componentes de erro e que poderá ser dividida em dois grandes tipos: efeitos fixos e efeitos aleatórios.

O primeiro tipo admite que os efeitos específicos α_i e λ_t se encontram correlacionados com as variáveis explicativas e que, dessa forma, podem ser entendidos como *N* e *T* constantes separadas. Assim, a estimação e inferência estatística terão que ser feitas condicionadas por α_i e λ_t.[64] Admitindo-se que esses efeitos individuais e temporais representam variáveis omissas, torna-se altamente provável que tais efeitos se encontrem correlacionados com os regressores (Judson e Owen [1996]).

Já os efeitos aleatórios admitem que os efeitos específicos são apropriadamente descritos por variáveis aleatórias não correlacionadas com os regressores (o que será uma dificuldade, particularmente para modelos autorregressivos) e com distribuições possivelmente do tipo $\alpha_i \sim N(0,\sigma_\alpha^2)$ e $\lambda_t \sim N(0,\sigma_\lambda^2)$. Ou seja, os modelos com efeitos aleatórios podem ser vistos como um caso particular dos modelos com efeitos fixos, na medida em que assumem determinadas distribuições para os efeitos, permitindo a inferência incondicional.

Se se admitirem efeitos fixos, no caso genérico de regressores não estocásticos, os estimadores LSDV serão BLUE, desde que $u_{it} \sim N(0,\sigma^2)$ e assimptoticamente eficientes se os mesmo termos de perturbação forem *i.i.d.* e de qualquer forma cêntricos e consistentes para estruturas de covariâncias mais complexas. No caso concreto dos modelos dinâmicos, em que entre os regressores se encontram termos desfasados da variável endógena, os estimadores *Within* permanecerão consistentes para $T \to \infty$ desde que, obviamente, os termos de perturbação u_{it} não se apresentem autocorrelacionados na dimensão temporal.

Já com efeitos aleatórios, existem estimadores GLS eficientes face aos estimadores *within* no caso de regressores não estocásticos, mas não serão sequer consistentes com

[64] Ou seja, os efeitos fixos são na verdade resultado de variáveis aleatórias mas que, dado estarmos a condicionar a estimação por α_i e λ_t, podem ser tratadas como constantes (Judge *et al.* [1985], pág. 527).

variáveis endógenas desfasadas, pelo que a única saída será a estimação dos modelos através de variáveis instrumentais.

Para evitar o problema da perda de consistência, em ambas as situações, tem-se como alternativa genericamente consistente, após a diferenciação do modelo, a utilização de estimadores de variáveis instrumentais, IV. A eficiência destes estimadores é questionável, dependendo do conteúdo informativo dos instrumentos, que poderá não ser muito, em particular se o número de regressores exógenos for diminuto ou nulo.

Uma alternativa plausível, quer ao LSDV, quer ao IV, é a estimação por GMM, conforme foi discutido no capítulo 5. No entanto, haverá aqui que efectuar um balanço crítico relativamente ao *trade-off* existente entre aos ganhos de eficiência que estes estimadores permitem (preservando a consistência) e o acréscimo de complexidade e de carga computacional que exigem.

Como regra geral, para modelos dinâmicos com efeitos fixos, com painéis longos, pode optar-se pelo estimador *within*, enquanto que para painéis curtos se deve preferir a estimação GMM.

Os modelos podem ser generalizados por forma a que heterogeneidade se revele não apenas no termo independente, mas também nos próprios parâmetros associados às variáveis explicativas. No entanto, por economia de espaço, optou-se por não abordar essa possibilidade nesta revisão de literatura. Para além deste tópico, muitos outros foram preteridos ou abordados apenas superficialmente, como as questões da estimação por recurso a especificações alternativas para as estruturas de variâncias, os estimadores PMG ou MG, a existência de relações de cointegração em modelos em painel e, acima de tudo, a estimação GMM, que poderia e deveria ter merecido um maior desenvolvimento.

Para quem necessitar de um maior aprofundamento das matérias aqui expostas, referências fundamentais serão Hsiao [1986], Baltagi [1995] e Mátyás e Sevestre [1992] (existe uma edição mais recente, de 1995, com melhorias substanciais). Para temas como o GMM ou as raízes unitárias, sugerem-se Mátyás [1999] e Maddala e Kim [1998], respectivamente. Finalmente, para quem desejar uma referência rápida à estimação de modelos dinâmicos com dados em painel, propõe-se Urga [1992], ainda que complementado com referências mais recentes.

10. REFERÊNCIAS

[1] Amemyia, Takeshi [1985]. *Advanced Econometrics*, Oxford: Basil Blackwell.

[2] Amemyia, Takeshi e Thomas E. MaCurdy [1986]. "Instrumental-Variable Estimation of An Error-Components Model", *Econometrica*, Vol. 54, No. 4, 869-880.

[3] Anderson, T.W. e Hsiao, Cheng [1981]. "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, 18:47-82.

[4] Ahn, Seung C. e Peter Schmidt [1997]. "Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation", *Journal of Econometrics*, 76, 309-321.

[5] Arellano, Manuel e Stephen Bond [1991]. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58: 277-297.

[6] Balestra, Pietro e Marc Nerlove [1966]. "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: the Demand for Natural Gas", *Econometrica*, 34, No.3, 585-612.

[7] Baltagi, Badi H. [1995]. *Econometric Analysis of Panel Data*, New York: John Wiley & Sons.

[8] Baltagi, Badi H. e James Griffin [1984]. "Short and long run effects in pooled models", *International Economic Revi*ew, 25, No. 3, 631-645.

[9] Baltagi, Badi H. e James Griffin [1997]. "Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline", *Jounal of Econometrics*, 77, 303-327.

[10] Bhargava, Franzini e Narendranathan [1982]. "Serial correlation and fixed effects model", *Review of Economic Studies* 49: 129-140.

[11] Bhargava, A. e Sargan, J.D. [1983]. "Estimating Dynamic Random Effects Models from Panel Data covering Short Time Periods", *Econometrica*, 51:1635-59.

[12] Breusch, Trevor S., Grayham E. Mizon e Peter Schmidt [1989]. "Efficient Estimation Using Panel Data", *Econometrica*, Vol. 57, No. 3, 695-700.

[13] Chamberlain, Gary [1984]. "Panel Data" in Zvi Griliches e Michael Intriligator, editores, *Handbook of Econometrics*, Vol. II, 1247-1318. Amesterdam: Elsevier Science Publishers BV.

[14] Crepon, Bruno, Francis Kamarz e Alain Trognon [1997]. "Parameters of interest, nuisance parameters and orthogonality conditions: an application to autoregressive error component models", *Journal of Econometrics*, 82, 135-156.

[15] Davidson, Russell e James G. MacKinnon [1993]. *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

[16] Granger, Clive W.J. e Huang, Ling-Ling [1997]. “Evaluation of Panel Data Models: Some Suggestions From Time Series”, *University of California in San Diego - Department of Economics Discussion Paper 97-10*.

[17] Gujarati, Damodar [1995]. *Basic Econometrics*, 5th Ed., New York: McGraw-Hill

[18] Greene, William H. [1997]. *Econometric Analysis*, 3rd Ed. (International Edition), Upper Saddle River: Prentice Hall.

[19] Hamilton, James D. [1994]. *Time Series Analysis*, Princeton, New Jersey: Princeton University Press.

[20] Hausman, J.A. e W.E. Taylor [1981]. "Panel Data and Unobservable Individual Effects", *Econometrica*, 49, no. 6: 1377-98.

[21] Heimonen, Kari [1999]. “Stationarity of the European real exchange rates – evidence from panel data”, *Applied Economics*, 1999, 31: 673-77.

[22] Holtz-Eakin, Douglas, Whitney Newey e Harvey S. Rosen [1988]. “Estimating Vector Autoregressions with Panel Data”, *Econometrica*, vol. 56, no. 6, 1371-95.

[23] Hsiao, Cheng [1986]. *Analysis of Panel Data*. Cambridge: Cambridge University Press.

[24] Im, Kyung So, M. Hashem Pesaran e Yongcheol Shin [1997]. “Testing for Unit Roots in Heterogeneous Panels”, *Department of Applied Economics Discussion Paper*, University of Cambridge s/n.

[25] Jonhston, Jack e John DiNardo [1997]. *Econometric Methods*, 4ª Edição. New York: McGraw-Hill Companies, Inc..

[26] Judge, George, W.E. Griffiths, R. C. Hill, H. Lüktepohl e Tsoung-Chao Lee [1985]. *The Theory and Practice of Econometrics*, 2ª Edição, New York: John Wiley and Sons.

[27] Judson, Ruth e Owen, Ann [1996]. “Estimating Dynamic Panel Data Models: A Pratical Guide for Macroeconomists”, *Federal Reserve System Working Paper*.

[28] Kao, Chihwa [1999]. “Spurious regression and residual-based tests for cointegration in panel data”, *Journal of Econometrics*, 90, 1-44.

[29] Kao, Chihwa e Chiang, Min-Hsien [1997]. “On the Estimation and Inference of a Cointegrated Regression in Panel Data”, *Syracuse University Working Paper*.

[30] Keane, Michael P. e David E, Runkle [1992]. “On the Estimation of Panel Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous”, *Journal of Business and Economic Statistics*, Vol. 10, No. 1, 18-22.

[31] Kmenta, Jan [1986]. *Elements of Econometrics*. New York: Macmillan.

[32] Koenker, Roger e José A.F. Machado [1997]. “GMM inference when the number of moment conditions is large”, *Working Paper* da University of Illinois em Champaign.

[33] Larsson, Rolf e Johan Lyhagen [1999]. “Likelihood-Based Inference in Multivariate Panel Cointegration Models”, *Stockholm School of Economics Working Paper Series in Economics and Finance*, no. 331.

[34] Levin, Andrew e Chen.Fu Lin [1992]. “Unit Root Tests in Panel Data: asymptotic and finite sample properties”, *University of California in San Diego Discussion Paper* 92-23.

[35] Levin, Andrew e Chen-Fu Lin [1993]. “Unit Root Tests in Panel Data: new results”, *University of California in San Diego Discussion Paper* 93-56.

[36] Maddala, G.S. [1977]. *Econometrics*. New York: McGraw-Hill.

[37] Maddala, G.S. e In-Moo Kim [1998]. *Unit Roots, Cointegration, and Structural Change*. Cambridge: Cambridge University Press.

[38] Marques, Luís David [2000]. *Inflation Targeting: Conceitos e Evidência de um Novo Regime de Objectivos Intermédios em Política Monetária*, Tese de Mestrado em Economia, Faculdade de Economia do Porto.

[39] Mátyás, Lázló e Sevestre, Patrick, eds. [1992]. *The Econometrics of Panel Data: Handbook of Theory and Applications*, Dordrecht: Kluwer Academic Publishers.

[40] Mátyás, Lázló [1999]. *Generalized Method of Moments*. Cambridge: Cambridge University Press.

[41] Metcalf, Gilbert E. [1997]. “Specification testing in panel data with instrumental variables”, *Journal of Econometrics*, 71, 291-307.

[42] Nerlove, Marc [1971]. “Further Evidence on the Estimation of Dynamic Relations from a Time Series of Cross Sections”, *Econometrica*, 39:359-382.

[43] Nickell, S. [1981]. “Biases in Dynamic Models with Fixed Effects”, *Econometrica*, 49:1399-416.

[44] Pesaran, M. Hashem, Yongcheol Shin e Ron P. Smith [1998]. “Pooled Mean Group Estimation of Dynamic Heterogeneous Panels”, *Department of Applied Economics, University of Cambridge Working Paper* s/n., Novembro 1998.

[45] Pirotte, Alain [1999]. “Convergence of the static estimation toward the long run effects of dynamic panel data models”, *Economic Letters*, 63, 151-158

[46] Urga, Giovanni [1992]. “The Econometrics of Panel Data: a selective introduction”, texto não publicado da London Business School.