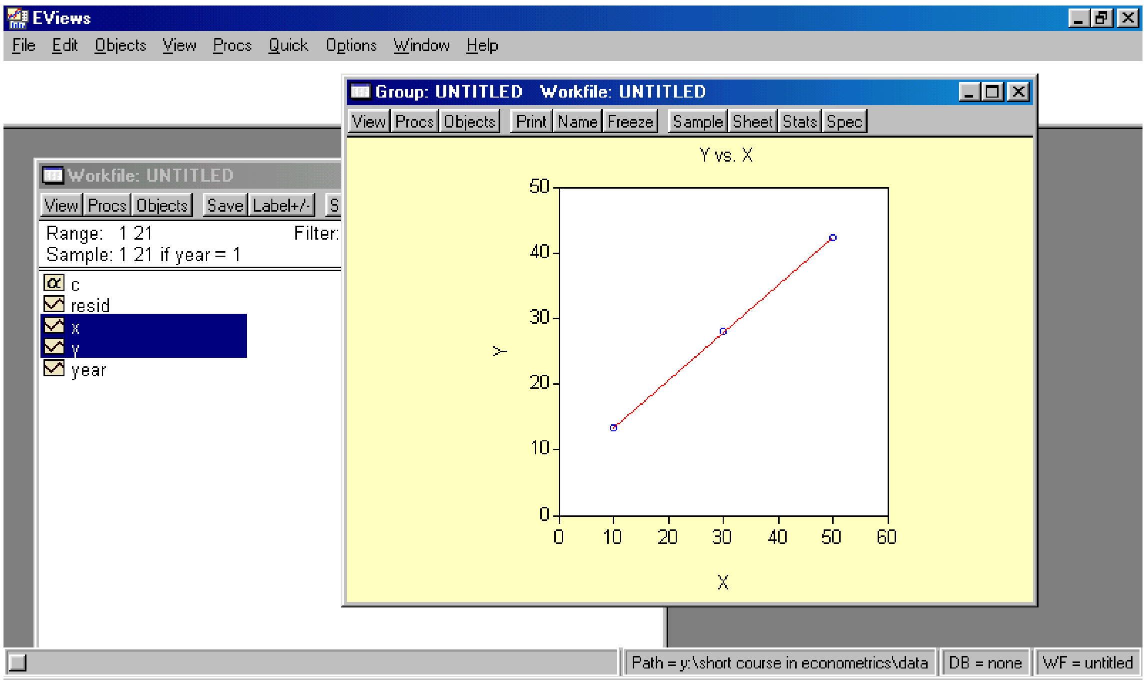
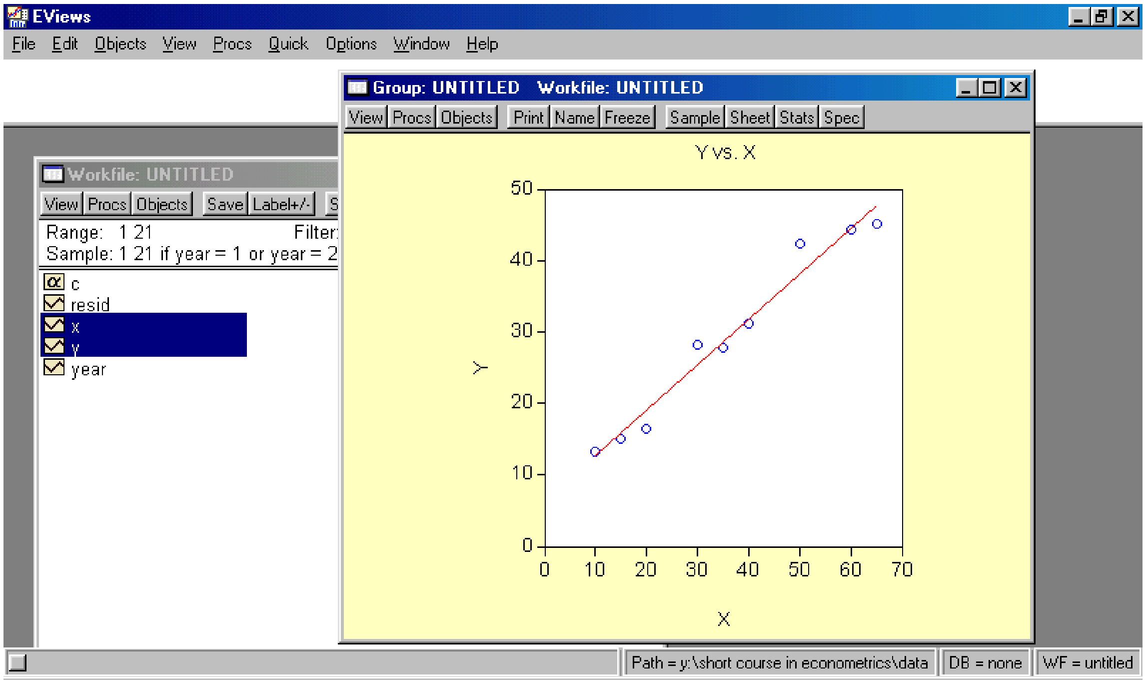
Dữ liệu dạng bảng

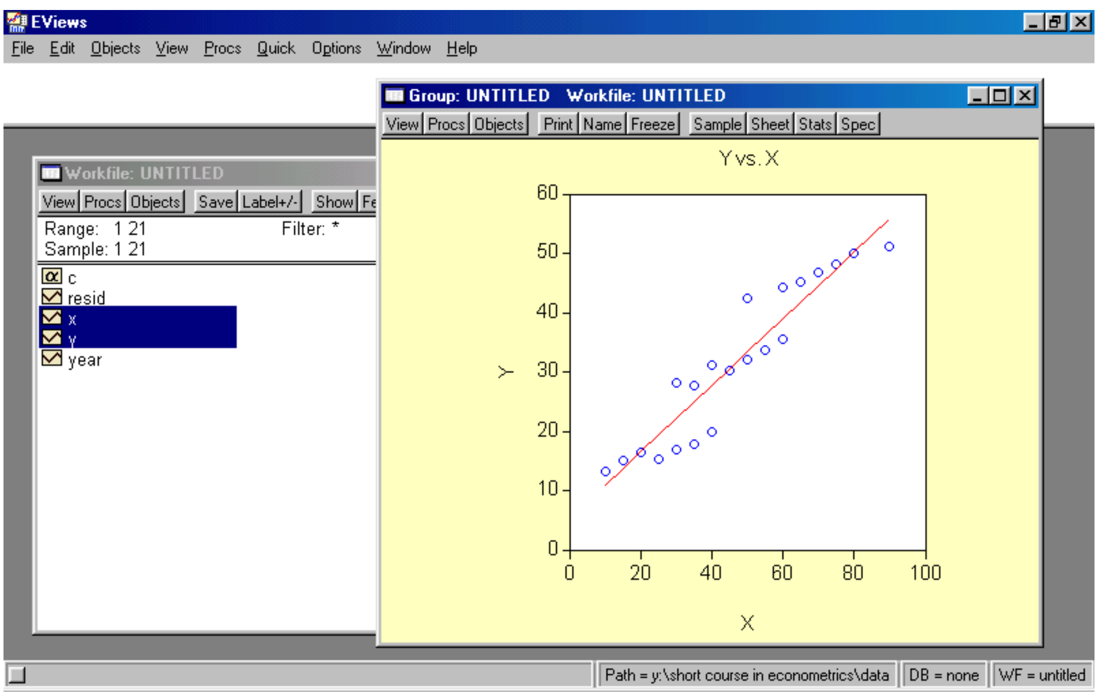
Giả sử chúng ta có dữ liệu về sản lượng và nhân công của 3 hãng đối với 1 năm cho trước và lược đồ tương quan điểm rời rạc trông giống thế này :



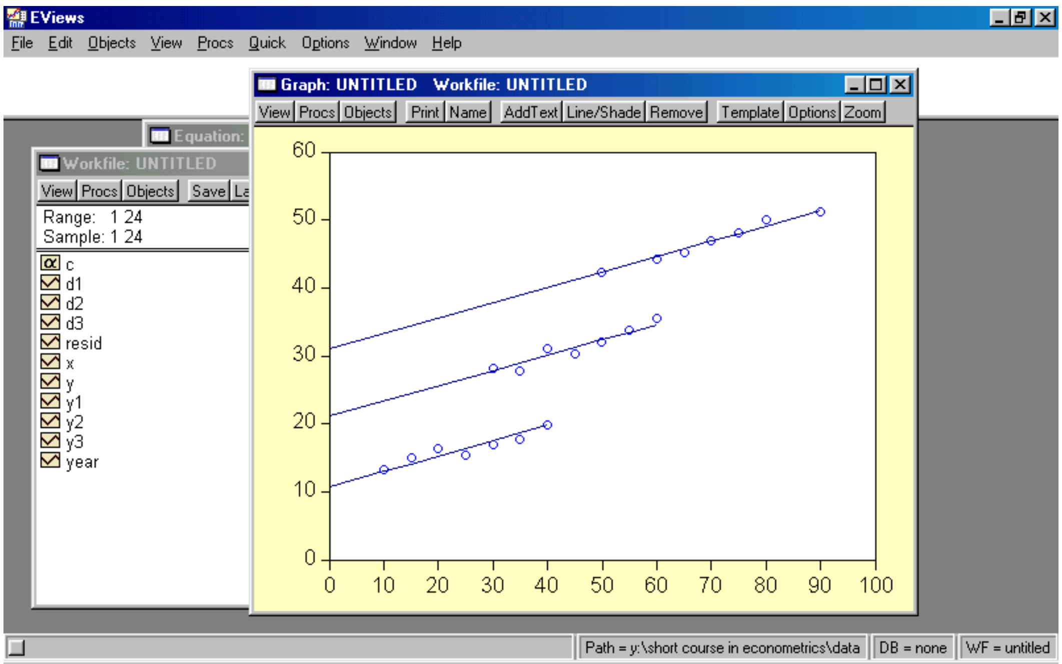
Khi thời gian trôi qua, chúng ta thu thập được thêm dữ liệu; lược đồ tương quan điểm rời rạc sau đó chỉ ra đường hồi qui của chúng ta sau khi chúng ta đã thu thập dữ liệu cho 3 năm.



Sau 7 năm trôi qua, chúng ta có một phép hồi qui trông như thế này :



Dường như rõ ràng là đường hồi qui này không đại diện chuẩn xác cho mối quan hệ giữ a biến giải thích và biến phụ thuộc đối với một hã ng cụ thể. Trên thực tế, tập hợp các phép hồi qui sau đây dường như phù hợp hơn :



Rõ ràng, chúng ta có thể trình bày dữ liệu cho ba hã ng này với một phép hồi qui bao gồm một cặp biến giả, nhưng chúng ta sẽ thấy rằng đối tượng GỘP CHUNG (POOL object) của EViews, được thiết kế để xử lý dữ liệu dạng bảng, linh động và mạnh hơn nhiều so với bất cứ điều gì chúng ta có thể làm một cách dễ dàng với các biến giả.

Các điểm mạnh của dữ liệu dạng bảng : ước lượng các tham số khi chiều chéo rất ngắn, loại bỏ thiên lệch do bỏ sót biến trong nhữ ng trường hợp nhất định, đặc trưng phong phú của cấu trúc tích sai sai số.

Một tập hợp dữ liệu dạng bảng bao gồm1 tập hợp các quan sát chuỗi thời gian trên một tập hợp các đơn vị chéo. Nếu mỗi một trong N chuỗi thời gian đều có cùng chiều dài T, thì NxT quan sát tạo nên một bảng cân đối. Các bảng không cân đối tuy không đòi hỏi các phương pháp phân tích đặc biệt, nhưng ký hiệu hơi phức tạp hơn so với các bảng cân đối. Đa số phần mềm dành cho dữ liệu dạng bảng đều có thể dễ dàng xử lý các bảng không cân đối. Chúng ta sẽ tập trung vào các bảng cân đối. EViews đòi hỏi sự trình bày của dữ liệu dạng bảng phù hợp với một bảng cân đối nhưng dữ liệu điền vào không cần phải cân đối.

Dễ nhất là nghĩ về các tập hợp dữ liệu dạng bảng như những vectơ chuỗi thời gian được xếp lại theo phương thẳng đứng, mặc dù đây chỉ là một trong nhiều cách mà dữ liệu dạng bảng có thể được trình bày trong các POOL objects trên EViews . Ví dụ, nếu Y là biến phụ thuộc trong một mô hình dữ liệu dạng bảng, chúng ta có thể hình dung các thành phần của nó được bố trí như sau :

$$\begin{bmatrix} Y_{1,1} \\ Y_{1,2} \\ \vdots \\ Y_{1,T} \\ Y_{2,1} \\ Y_{2,2} \\ \vdots \\ Y_{2,T} \\ \vdots \\ Y_{N,1} \\ Y_{N,2} \\ \vdots \\ Y_{N,T} \end{bmatrix}$$

Vectơ dữ liệu này bao gồm NxT thành phần. Số nhỏ phiá dưới đầu tiên là chỉ số cho các đơn vị chéo; Số thứ hai là chỉ số cho các thời đoạn.

Các quan sát trên các biến hồi qui và những biểu tượng cho các thành phần nhiễu ngẫu nhiên không thể quan sát có thể được bố trí tương tự.

Các mô hình tuyến tính đối với dữ liệu dạng bảng

Thành phần cơ bản để xây dựng một mô hình tuyến tính cho dữ liệu dạng bảng là phương trình này:

$$Y_{it} = \alpha_i + \beta_{i1} X_{1it} + \beta_{i2} X_{2it} + \cdots + \beta_{iK} X_{Kit} + \varepsilon_{it}$$

Đặc trưng mô hình này cho phép tung độ gốc hồi qui và các hệ số hồi qui riêng phần biến thiên giữ a các đơn vị chéo nhưng không theo thời gian. Đây là lớp các mô hình mà POOL object trên EViews ước lượng dễ dàng nhất. Hã y ghi nhận rằng chúng ta không giả định rằng $X_{1it} = 1$. Nhữ ng tung độ gốc này được thể hiện riêng biệt bởi các biểu tượng α_i.

Phương trình này cho phép nhà nghiên cứu trình bày nhiều mô hình đa dạng phụ thuộc vào nhữ ng giả định được đưa ra về các tung độ gốc hồi qui, các hệ số hồi qui riêng phần và các thành phần nhiễu ngẫu nhiên. Sự phong phú của tập hợp dữ liệu dạng bảng cho phép có sự đa dạng phong phú của đặc trưng mô hình và đòi hỏi xử lý tinh vi các thành phần nhiễu ngẫu nhiên. Sự phân loại các mô hình cho dữ liệu dạng bảng phản ánh các lựa chọn giữ a các khiá cạnh :

Các giả định thay thế đối với tung độ gốc

1. Không: $\alpha_i = 0$ đối với mọi i.

2. Chung : $\alpha_i = \alpha$ đối với mọi i.

3. Các tác động cố định : α_i khác nhau giữ a các đơn vị chéo và $E[\alpha_i \varepsilon_{it}] = 0$

4. Các tác động ngẫu nhiên : các đơn vị chéo có các tung độ gốc khác nhau là nhữ ng nhận biết (realizations) của một biến ngẫu nhiên :

$$\alpha_i = \alpha + v_i \text{ và } E[v_i \varepsilon_{it}] = 0 \text{ và } E[v_i X_{kit}] = 0$$

Các giả định thay thế đối với nhữ ng hệ số hồi qui riêng phần

1. Chung: ít nhất là một số hệ số hồi qui riêng phần là chung giữ a các đơn vị chéo.

2. Chéo và cụ thể : ít nhất là một số hệ số hồi qui riêng phần khác nhau đối với các đơn vị chéo khác nhau

Các giả định về những thành phần sai số ngẫu nhiên

Các thành phần sai số có thể đã làm phức tạp thêm các cấu trúc phương sai và tích sai. Hãy xét một mô hình có các tác động cố định cho sản lượng của các hãng, trong đó chúng ta giả định rằng tất cả mọi hãng đều có độ co giãn của sản lượng theo vốn như nhau và độ co giãn của sản lượng theo nhân công như nhau , nhưng chúng có các tung độ gốc khác nhau, có lẽ là vì chúng khác nhau về tần suất gián đoạn nguồn cung cấp điện cho chúng.

$$\log(Y_{1t}) = \alpha_1 + \beta_1 \log(K_{1t}) + \beta_2 \log(L_{1t}) + \varepsilon_{1t}$$

$$\log(Y_{2t}) = \alpha_2 + \beta_1 \log(K_{2t}) + \beta_2 \log(L_{2t}) + \varepsilon_{2t}$$

$$\vdots$$

$$\log(Y_{Nt}) = \alpha_N + \beta_1 \log(K_{Nt}) + \beta_2 \log(L_{Nt}) + \varepsilon_{Nt}$$

✓ Các thành phần sai số ngẫu nhiên có thể có phương sai không đồng nhất giữa các hãng và phương sai đồng nhất trong phạm vi các hãng :

$E\left[\varepsilon_{it}^2\right] = \sigma_i^2$ Ghi nhận là không có chỉ số t trên phương sai.

✓ Nếu các hãng trải qua các cơn sốc ngẫu nhiên tương tự một cách đồng thời, thì các thành phần sai số ngẫu nhiên có thể bộc lộ mối tương quan cùng thời chéo giữa các phương trình :

$E\left[\varepsilon_{it}\varepsilon_{jt}\right] = \sigma_{ij}$

✓ Mỗi hãng có thể có một thành phần sai số ngẫu nhiên tự tương quan; sự tự tương quan này có thể là chung giữa các hãng, hoặc là nó có thể khác biệt giữa các hãng :

$\varepsilon_{it} = \rho\,\varepsilon_{i(t-1)} + \xi_{it}$ hoặc $\varepsilon_{it} = \rho_i\,\varepsilon_{i(t-1)} + \xi_{it}$

✓ Hãy ghi nhận rằng sự kết hợp của giả định thứ 2 và thứ 3 cho phép các sai số ngẫu nhiên của một hãng tự tương quan với các sai số ngẫu nhiên trong quá khứ của các hãng khác.

EViews cung cấp những công cụ rất thuận tiên để ước lượng tất cả những trường hợp này.

Các khái niệm ước lượng

Kỹ thuật ước lượng phù hợp được xác định bởi cấu hình của các hệ số hồi qui và các giả định về thành phần nhiễu ngẫu nhiên. Những trường hợp đơn giản nhất xuất hiện khi các giả định cổ điển (không có tính không đồng nhất của phương sai, không có tương quan cùng thời chéo giữa các phương trình, và không có tự tương quan) gắn liền với hành vi của ε_{it}.

Trong những trường hợp này, OLS thu được các hàm ước lượng không chệch tuyến tính tốt nhất cho một loạt mô hình khác nhau . Mô hình đơn giản nhất sẽ là mô hình mà trong đó tất cả mọi đơn vị chéo đều có chung tung độ gốc và các hệ số hồi qui riêng phần:

$$Y_{i,t} = \alpha + \beta_1 X_{1,i,t} + \beta_2 X_{2,i,t} + \cdots + \beta_K X_{K,i,t} + \varepsilon_{i,t}$$

Đây là một mô hình có giới hạn cao và giả thuyết rằng tất cả mọi đơn vị chéo đều có chung tung độ gốc thường bị bác bỏ.

Phép ước lượng trong phạm vi

Đặc trưng mô hình thường gặp nhất đối với dữ liệu dạng bảng có thể là mô hình có các tác động cố định. Các hệ số hồi qui riêng phần được giả định là chung cho các đơn vị chéo, nhưng các tung độ gốc hồi qui được coi là khác nhau giữa các đơn vị chéo. Những phương trình này trông như thế này :

$$Y_{1,t} = \alpha_1 + \beta_1 X_{1,1,t} + \beta_2 X_{2,1,t} + \cdots + \beta_K X_{K,1,t} + \varepsilon_{1,t}$$

$$Y_{2,t} = \alpha_2 + \beta_1 X_{1,2,t} + \beta_2 X_{2,2,t} + \cdots + \beta_K X_{K,2,t} + \varepsilon_{2,t}$$

$$\vdots$$

$$Y_{N,t} = \alpha_N + \beta_1 X_{1,N,t} + \beta_2 X_{2,N,t} + \cdots + \beta_K X_{K,N,t} + \varepsilon_{N,t}$$

Tập hợp các phương trình này có thể được ước lượng bởi OLS với một tập hợp phù hợp các biến giả. Nếu một nhà nghiên cứu muốn sử dụng tất cả những biến giả này , thì việc bỏ đi tung độ gốc hồi qui sẽ cho phép cô ta tránh được bẫy biến giả.

Mặt khác, nếu bảng này là một bảng lớn, thì có thể chứng tỏ rằng xây dựng tập hợp các biến giả cần thiết là hoàn toàn nhàm chán. Phương pháp "trong phạm vi" cho phép chúng ta ước lượng các hệ số hồi qui riêng phần mà không phải xác định bất cứ biến giả nào, và nó cũng cho phép chúng ta minh hoạ những trường hợp mà trong đó dữ liệu dạng bảng tạo điều kiện cho chúng ta tránh được thiên lệch do bỏ biến.

Giả sử rằng kèm theo mỗi đơn vị chéo là một biến không thể quan sát nào đó không thay đổi theo thời gian. Ví dụ , xét một phương trình nhằm liên kết điểm số với số giờ học. Chúng ta có thể thu thập dữ liệu cho một tập hơn các sinh viên qua nhiều học kỳ và chạy một phép hồi qui , nhưng một biến mà chúng ta không thể dễ dàng quan sát là trí thông minh.

Hãy viết tung độ gốc của từng phương trình sao cho nó trông giống như thế này:

$$\alpha_i = \alpha_i' + \gamma_i Z_i$$

Hãy luôn nhớ rằng Z_i không thay đổi theo thời gian nên $Z_i = Z_{it}$.

Hãy xác định các giá trị trung bình chéo cụ thể của dữ liệu như sau :

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$$

Rõ ràng là $\bar{z}_i = \frac{1}{T}\sum_{t=1}^{T} Z_{it} = Z_i$

Bây giờ, hã y chuyển lại mô hình này dưới dạng các độ lệch khỏi các giá trị trung bình chéo:

$$Y_{i,t} - \bar{y}_i = (\alpha_i - \alpha_i) + \gamma_i (Z_i - \bar{z}_i) + \beta_1 (X_{1,i,t} - \bar{x}_{1,i}) + \cdots + \beta_K (X_{K,i,t} - \bar{x}_{K,i}) + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$$

Thật thú vị, nhữ ng thành phần kèm theo tung độ gốc và kèm theo biến không thể quan sát biến mất :

$$Y_{i,t} - \bar{y}_i = \beta_1 (X_{1,i,t} - \bar{x}_{1,i}) + \cdots + \beta_K (X_{K,i,t} - \bar{x}_{K,i}) + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$$

OLS được thực hiện trên dữ liệu dưới dạng độ lệch được gọi là phép ước lượng "trong phạm vi" vì toàn bộ biến thiên giữ a các đơn vị chéo đã được trừ ra ngoài : Chỉ có biến thiên màvới nó các hệ số hồi qui riêng phần được ước lượng là biến thiên trong phạm vi các đơn vị chéo theo thời gian.

Các hệ số hồi qui riêng phần chéo cụ thể

Vượt ra ngoài mô hình có các tác động cố định sang các mô hình mà trong đó một số hệ số hồi qui riêng phần là chéo cụ thể có thể được thực hiện dễ dàng bởi việc tạo ra các thành phần biến giả/tương tác đối với các biến mà các hệ số của chúng biến thiên giữ a các đơn vị chéo. Làm việc này một cách máy móc thì rất nhàm chán nếu nhiều biến giả chéo có liên quan; Tuy nhiên, EViews làm điều này rất dễ dàng.

Các phép tổng quát hoá cấu trúc tích sai của sai số

Nếu những giả định cổ điển đối với các thành phần sai số ngẫu nhiên không đúng, thì các hàm ước lượng OLS vẫn là không chệch nhưng không hiệu quả. Tệ hơn nữa, ma trận phương sai - tích sai của hệ số ước lượng là chệch và không nhất quán, vì thế các trị thống kê t và F không có hiệu lực.

Như đã nêu trên đây, trong các mô hình dữ liệu dạng bảng, một loạt giả định đa dạng về cấu trúc của tích sai sai số ngẫu nhiên là hiện thực hơn so với các giả định cổ điển . Trong những tập hợp các giả định thay thế này, kỹ thuật ước lượng phù hợp là bình phương tối thiểu tổng quát khả thi (FGLS).

Trong phần này chúng ta tập trung vào đặc trưng của ma trận đồng phương sai của sai số.

EViews và dữ liệu dạng bảng

POOL objects trên EViews hoạt động trên các biến có các tên gọi đặc biệt gồm2 phần. Phần thứ nhất là tên của biến này, và phần thứ hai của tên gọi là yếu tố xác định đơn vị chéo chỉ đơn vị chéo mà biến này thuộc về.

Nói chung, tôi bắt đầu các yếu tố xác định đơn vị chéo bằng dấu gạch nối thấp để cho các tên đầy đủ của biến dễ đọc hơn.

Ví dụ : Tôi muốn làm việc với một tập hơn dữ liệu dạng bảng về Mỹ , Canada, và Mexico. Những biến mà tôi muốn sử dụng là GDP, dân số (Population), và dòng trao đổi mậu dịch (Trade Flows).

Phần đầu tiên của tên biến:

GDP
POP
TRA

Phần thư hai của tên biến (Yếu tố xác định đơn vị chéo)

_USA
_CAN
_MEX

Các biến

GDP_USA	GDP_CAN	GDP_MEX
POP_USA	POP_CAN	POP_MEX
TRA_USA	TRA_CAN	TRA_MEX

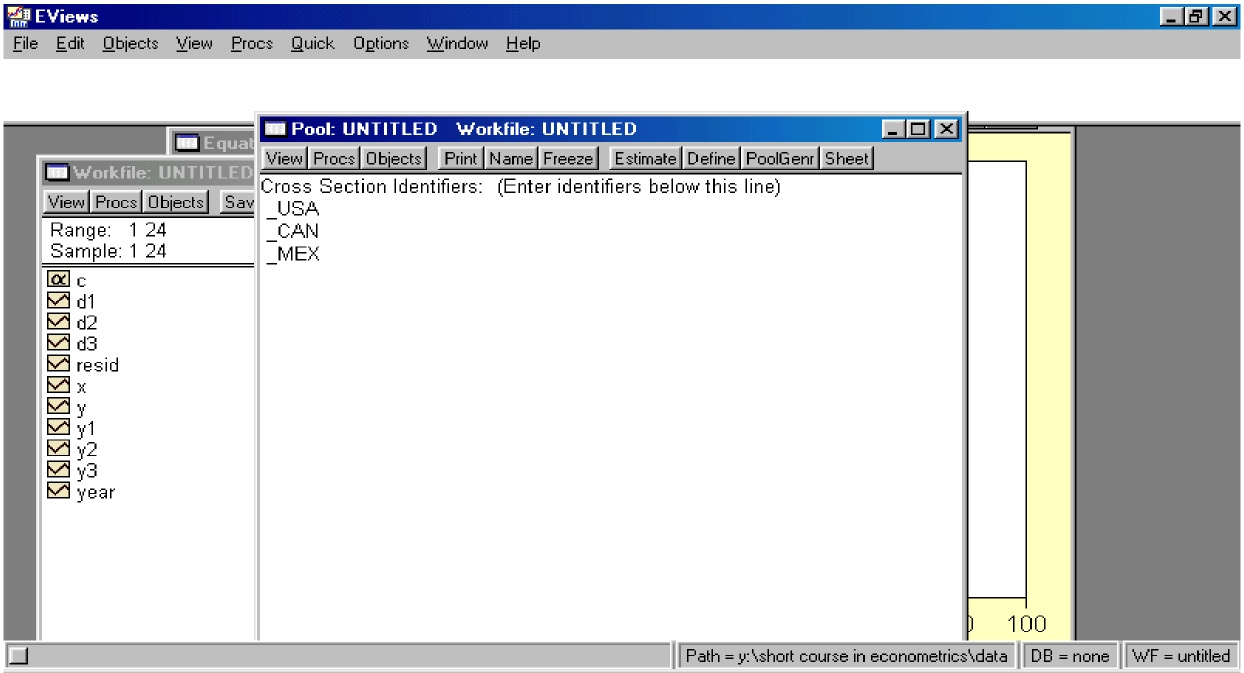
Vì vậy, có 9 biến trong workfile của tôi trên EViews. Dễ dàng thấy rằng các tập hợp dữ liệu dạng bảng có thể nhanh chóng trở nên rất lớn . Ví dụ, Thầy Randy đang làm việc với một bảng của 61 tỉnh tại Việt nam mà đối với chúng Thầy có chuỗi thời gian 10 năm trên 8 biến có liên quan tới nông nghiệp: workfile của Thầy có (61 x 8) = 488 biến.

Sau khi Anh/Chị đã đặt tên cho tất cả các biến của mình và đã thu dữ liệu vào trong một workfile trên EViews, Anh/Chị sẵn sàng tạo ra POOL object.

Anh/Chị làm việc này bằng cách nhấp theo thứ tự sau:

Objects / New Object / Pool

Một cửa sổ sẽ mở ra với khoảng trống dành cho Anh/Chị liệt kê các yếu tố xác định đơn vị chéo của mình :

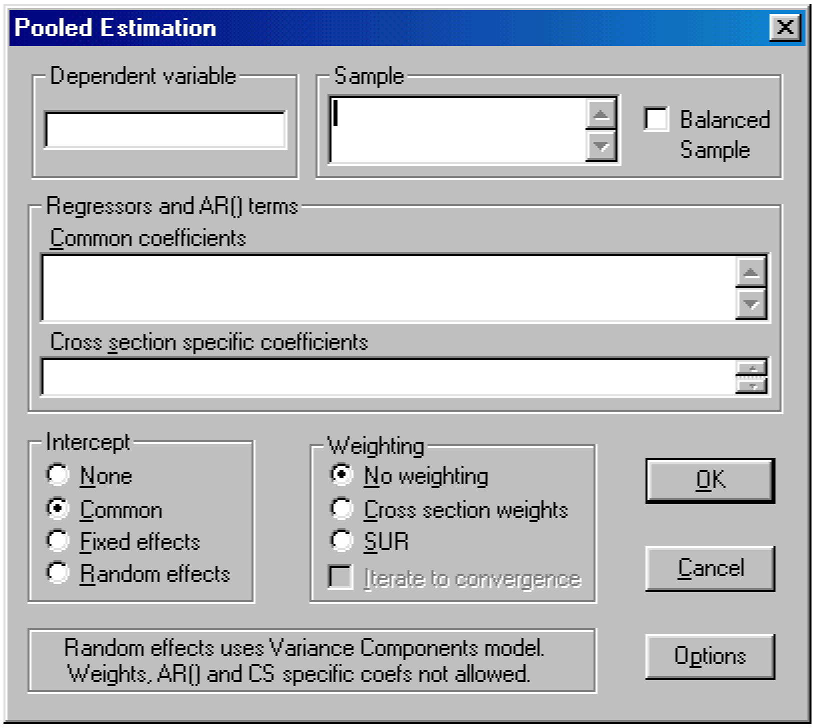


Trong POOL object Anh/Chị nêu các biến bằng tên ở phần thứ nhất cùng các dấu hỏi chấm. Do đó, nếu tôi đánh máy một lệnh sử dụng GDP? , thì EViews sử dụng tất cả ba chuỗi GDP đối với Mỹ, Canada, và Mexico.

Hãy ghi nhận nút PoolGenr. PoolGenr được sử dụng để tạo ra các biến mới phù hợp với các nguyên tắc tương tự như những nguyên tắc đối với Genr thông thường. Ví dụ, nếu tôi muốn tạo ra GDP đầu người cho cả ba quốc gia trong POOL của tôi, thì tôi sẽ nhấp PoolGenr và sau đó đánh máy phương trình :

GDPPC? = GDP? / POP?

Để ước lượng, EViews có một cửa sổ trong đó người sử dụng xác định phương trình và các giả định liên quan tới thành phần nhiễu ngẫu nhiên. Cửa sổ này được chỉ ra ở đây:



Chúng ta sẽ mô tả từng thành phần của cửa sổ xác định này.

Biến phụ thuộc

Biến phụ thuộc sẽ được đánh máy vào phù hợp với tên của nó và dấu chấm hỏi. Ví dụ , Anh/Chị có thể sử dụng GDPPC?

Các hệ số chung

Trong vị trí này, Anh/Chị hãy liệt kê tất cả các biến giải thích mà Anh/Chị giả định là có cùng hệ số hồi qui riêng phần đối với mỗi đơn vị chéo. Hãy sử dụng dạng VAR? Anh/Chị có thể sử dụng các đặc trưng AR(p) nếu Anh/Chị muốn lập mô hình tự hồi qui. Hãy luôn nhớ rằng tập hợp dữ liệu dạng bảng của Anh/Chị nên có chiều (dimension) chuỗi thời gian khá dài để nhận được các hàm ước lượng có thể tin cậy cho các hệ số tự tương quan.

Các hệ số chéo cụ thể

Trong cửa sổ này, Anh/Chị đánh máy những tên gọi của tất các các biến gỉai thích mà Anh/Chị giả định là có các giá trị hệ số hồi qui riêng phần khác nhau đối với những đơn vị chéo khác nhau. Hãy sử dụng dạng VAR?

Tung độ gốc

Ở đây Anh/Chị xác định xem liệu mô hình của Anh/Chị có phải là

Không có tung độ gốc ...	trường hợp này hiếm .
Có tung độ gốc chung ...	trường hợp này không thông thường .
Có các tác động cố định ...	đặc trưng điển hình .
Có các tác động ngẫu nhiên ...	đặc trưng này không thường được sử dụng vì nó yêu cầu các giả định mạnh khó thỏa mãn trên thực tế.

Thêm trọng số

Ở đây, việc thêm trọng số là nói tới "bình phương tối thiểu có trọng số khả thi".

Không thêm trọng số ... không có tính không đồng nhất của phương sai trong phương trình cụ thể.
Các trọng số chéo ... WLS khả thi để chính xác hoá cho phù hợp với tính không đống nhất của phương sai trong phương trình cụ thể.

SUR ... xét tới tương quan cùng thời chéo giữa các phương trình của các sai số và tính không đồng nhất của phương sai trong phương trình nhất định. Để sử dụng điều này, chiều chuỗi thời gian phải lớn hơn chiều chéo (T > N).

Lặp tới hội tụ ... làm cho chương trình tính các phần dư mới dựa trên các hàm ước lượng hệ số GLS khả thi, sau đó cập nhật những hàm ước lượng hệ số GLS khả thi này; tính các phần dư mới dựa trên những hàm ước lượng hệ số GLS mới, sau đó cập nhật những hàm ước lượng hệ số GLS khả thi này , v.v...

Các lựa chọn

Ở đó chỉ có một lựa chọn : HCCM của White có thể được đưa vào nếu Anh/Chị chọn không thêm trọng số hay các trọng số chéo.

Kiểm định giả thuyết

Trong các mô hình dữ liệu dạng bảng, chúng ta quan tâm tới việc kiểm định hai loại giả thuyết: giả thuyết về các phương sai và tích sai của các thành phần sai số ngẫu nhiên và giả thuyết về các hệ số hồi qui.

Một chút nghệ thuật có liên quan, nhưng qui trình từ tổng quát với đơn giản cho ta một hướng dẫn tốt.

Trước khi kiểm định giả thuyết về các hệ số hồi qui, điều quan trọng là có một đặc trưng tốt cho ma trận tích sai của sai số để các trị thống kê kiểm định đối với các hệ số hồi qui là đáng tin cậy.

Kiểm định giả thuyết về ma trận tích sai của sai số

Điều hữu ích là hãy nghĩ về các ma trận tích sai của sai số *có giới hạn* và *không có giới hạn* Một ma trận tích sai của sai số là một ma trận vuông với các phương sai sai số của các phương trình chéo riêng nằm dọc theo đường chéo và với các tích sai sai số cùng thời trên các thành phần ngoài đường chéo. Tất cả mọi ma trận tích sai đều đối xứng, nên nếu chúng ta xác định một ma trận tích sai sai số cho một mô hình dạng bảng với 5 đơn vị chéo thì ta có một ma trận (5 x 5) với 5 đơn vị đường chéo và 10 đơn vị ngoài đường chéo :

$$\begin{array}{ccccc}
\sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\
 & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\
 & & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\
 & & & \sigma_4^2 & \sigma_{45} \\
 & & & & \sigma_5^2
\end{array}$$

Nếu chúng ta nhấp vào nút dành cho SUR, EViews sẽ ước lượng tất cả những tham số này. Mặt khác, nếu chúng ta tin rằng những đơn vị chéo không có bất cứ tích sai sai số đồng thời chéo giữa các phương trình nào, thì chúng ta sẽ nhấp vào nút dành cho Cross-Section Weighting và EViews sẽ áp đặt các giới hạn zero lên tất cả các thành phần ngoài đường chéo của ma trận này. Chỉ có các thành phần trên đường chéo là sẽ được ước lượng :

$$\begin{array}{ccccc}
\sigma_1^2 & 0 & 0 & 0 & 0 \\
 & \sigma_2^2 & 0 & 0 & 0 \\
 & & \sigma_3^2 & 0 & 0 \\
 & & & \sigma_4^2 & 0 \\
 & & & & \sigma_5^2
\end{array}$$

Mô hình thứ hai bao gồm việc áp đặt 10 giới hạn.

Cuối cùng, nếu chúng ta giả định rằng các nhiễu ngẫu nhiên của chúng ta không có tính không đồng nhất của phương sai chéo, thì chúng ta nhấp vào nút No Weighting và EViews sẽ chỉ ước lượng một thành phần trên đường chéo thay cho năm thành phần: 4 giới hạn sẽ được áp đặt.

$$\begin{matrix} \sigma^2 & 0 & 0 & 0 & 0 \\ & \sigma^2 & 0 & 0 & 0 \\ & & \sigma^2 & 0 & 0 \\ & & & \sigma^2 & 0 \\ & & & & \sigma^2 \end{matrix}$$

Việc kiểm định những giới hạn này dễ dàng được thực hiện bởi phương tiện là một kiểm định có tên *kiểm định tỉ lệ tương đồng*. Trong các cuộc thám hiểm ước lượng trước đây, Anh/Chị có thể đã ghi nhận một trị thống kê được gọi là *Logarit-Tương đồng,* có mặt trong kết quả của EViews. Đây là một hàm ước lượng cho xác suất kết hợp của mẫu đã quan sát, cho trước các ước lượng điểm của tham số. Như vậy, nó là một số bị giới hạn bởi 0 và 1.

Tất cả các phương pháp ước lượng của chúng ta là nhằm tối đa hoá logarit-tương đồng này. Trong nhiều ứng dụng, việc tối đa hoá logarit-tương đồng dẫn tới hàm ước lượng y như là phương pháp bình phương tối thiểu thực hiện, nhưng yêu cầu phần toán học nặng hơn, vì thế chúng ta theo phương pháp bình phương tối thiểu.

Mối quan tâm của chúng ta ở đây là mức độ mà việc áp đặt các giới hạn lên ma trận tích sai sai số làm giảm trị thống kê logarit-tương đồng.

Nếu chúng ta hình thành một tỉ lệ giữa mức độ tương đồng của mô hình có giới hạn $\hat{L}_R$ và mức độ tương đồng của mô hình không có giới hạn $\hat{L}_U$, thì chúng ta kỳ vọng rằng tỉ lệ này sẽ nhỏ hơn 1 vì mức tương đồng tối đa tuỳ thuộc vào một giới hạn có thể không lớn hơn mức tương đồng tối đa của mô hình không có giới hạn.

Xác định tỉ lệ tương đồng : $\ell = \dfrac{\hat{L}_R}{\hat{L}_U}$. Khi đó $0 \le \ell \le 1$

Nếu mô hình có giới hạn không khác mô hình không có giới hạn một cách có ý nghĩa thì chúng ta kỳ vọng tỉ lệ tương đồng sẽ gần bằng 1. Lý thuyết phân phối xác suất của tỉ lệ tương đồng hơi phiền hà. Tuy nhiên, điều được biết rõ là phân phối xác suất của $-2 \times \ell$ là tiệm cận với Khi bình phương, vì thế bất cứ ứng dụng nào với một kích cỡ mẫu đủ lớn chúng ta đều có thể sử dụng:

$$-2 \times \ell = -2 \times \left(\log\left(\hat{L}_R\right) - \log\left(\hat{L}_U\right)\right) \text{ approx} \sim \chi^2_q$$ trong đó q là số các giới hạn.

Với giả thuyết "không" này, chúng ta kỳ vọng $-2 \times \ell$ sẽ gần bằng zero; Chúng ta bác bỏ giả thuyết "không" nếu giá trị tính được của trị thống kê tỉ lệ tương đồng này vượt giá trị tới hạn phù hợp hoặc nếu giá trị p của kiểm định nhỏ hơn mức ý nghĩa đã lựa chọn trước.

Mô hình giữ lại

Trong khi kiểm định giả thuyết về các giới hạn trên ma trận tích sai sai số, đặc trưng nào đó của mô hình hồi qui cho dữ liệu dạng bảng phải được giữ lại. Điều được khuyến cáo là mô hình đã giữ lại là "tổng quát" theo nghĩa mà chúng ta đã sử dụng từ này trong mô tả chiến lược lập mô hình từ "tổng quát tới đơn giản ".

Kiểm định các giới hạn trên mô hình cho dữ liệu dạng bảng

Sau khi một đặc trưng có cơ sở cho cấu trúc tích sai sai số đã được xác lập, các kiểm định kèm theo chiến lược lập mô hình từ tổng quát tới đơn giản có thể được thực hiện . Những kiểm định này có thể là các kiểm định Wald hoặc các kiểm định t thông thường trên các hệ số riêng.

Hãy luôn nhớ rằng khi các trọng số chéo hay các phương pháp SUR hay bất cứ đặc trưng AR(p) nào được sử dụng, tất cả những kết quả này đều có cơ sở xấp xỉ, nên các trị thống kê t có phân phối xác suất xấp xỉ chuẩn chuẩn hoá và các trị thống kê F của Wald xấp xỉ với Khi bình phương.

Mô hình không có giới hạn

Mô hình hoàn toàn không có giới hạn là mô hình này:

$$Y_{it} = \alpha_i + \beta_{i1} X_{1it} + \beta_{i2} X_{2it} + \cdots + \beta_{iK} X_{Kit} + \varepsilon_{it}$$

Trong mô hình này, các tung độ gốc và các hệ số hồi qui riêng phần biến thiên giữ a các đơn vị chéo. Nếu dù là việc không thêm trọng số hay lựa chọn các trọng số chéo được lựa chọn cho cấu trúc tích sai của sai số , thì nhữ ng kết quả này đều sẽ đúng y như khi áp dụng OLS cho dữ liệu đối với từng đơn vị chéo.

Nếu lựa chọn SUR được chọn , thì hiệu quả sẽ được nâng cao bởi sủ dụng thông tin có chứa trong các tích sai sai số chéo giữ a các phương trình. Hã y nhớ rằng (T > N) là cần thiết để sử dụng lựa chọn này.

Các mô hình có giới hạn một phần

Trong nhiều tập hợp dữ liệu dạng bảng, chiều chuỗi thời gian là hoàn toàn ngắn, nên ước lượng mô hình mà trong đó tất cả các tham số đều biến thiên giữ a các đơn vị chéo là không mang tính thực tiễn. Trong trường hợp này, mô hình khả thi tổng quát nhất là mô hình có các tác động cố định : chỉ có các tung độ gốc biến thiên giữ a các đơn vị chéo; các hệ số hồi qui riêng phần như nhau đối với tất cả mọi đơn vị chéo.

Tất nhiên, có các mô hình mà trong đó một số hệ số hồi qui riêng phần như nhau giữ a các đơn vị chéo trong khi các hệ số khác lại biến thiên.

Mô hình có giới hạn

Mô hình giới hạn nhất là mô hình trong đó các tung độ gốc và các hệ số hồi qui riêng phần như nhau đối với tất cả mọi đơn vị chéo.

Việc kiểm định các giới hạn mô hình có thể được thực hiện thông qua kiểm định hệ số Wald hay thông qua kiểm định tỉ lệ tương đồng. Hai phương pháp này là xấp xỉ tương đương, mặc dù chúng có thể cho các kết quả khác nhau đối với một mẫu hữ u hạn cụ thể.

Khi Anh/Chị đi theo chiến lược xây dựng mô hình từ tổng quát tới đơn giản,việc hợp lý là kiểm tra lại cấu trúc tích sai của sai số khi Anh/Chị áp đặt các giới hạn lên nhữ ng hệ số hồi qui riêng phần và các tung độ gốc của mô hình. Mặc dù Anh/Chị có thể không bác bỏ được giả thuyết đại diện cho các giới hạn mà Anh/Chị áp đặt, thì giả thuyết đó cũ ng có thể không phải là đúng *một cách hoàn hảo*, và điều đó có thể tác động lên như ng hàm ước lượng và như ng kiểm định của các tích sai sai số.

Bài tập 16 dẫn dắt Anh/Chị trải qua toàn bộ quá trình này đối với một bảng bao gồm 5 hã ng trong 30 năm.