

ANÁLISIS COMPARATIVO DE LAS ESTIMACIONES DE MODELOS DE PANEL EN LOS PROGRAMAS LIMDEP, TSP Y RATS*

JOAQUÍN MAUDOS

Universitat de València

EZEQUIEL URIEL

Universitat de València e IVIE

El objetivo de este trabajo es hacer un análisis comparativo de tres de los principales programas econométricos comercializados existentes para estimar con técnicas de datos de panel: LIMDEP (versión 6.0), RATS (versión 4) y TSP (versión 4.3). El análisis se va a centrar en la estimación de los modelos básicos de efectos fijos y de efectos aleatorios estáticos para paneles en los que existe el mismo número de observaciones para todos los años (balanceados). Por esta razón, no se va a comentar un programa cada vez más utilizado como es el DPD, dado que dicho programa, como el propio nombre indica (*Dynamic Panel Data*) y como ponen de manifiesto en la introducción del manual sus autores [Arellano y Bond 1988], está concebido principalmente para realizar estimaciones de modelos dinámicos con datos de panel. Además, hasta el momento este programa no se ha comercializado.

Con este objetivo, se analizarán las distintas salidas de los tres programas mencionados en un caso concreto. Estimaremos una función de producción tipo Cobb-Douglas, en la que la producción depende de tres factores productivos: capital privado (K), empleo (L) y capital público (G). Para ello, se dispone de un panel de datos para el periodo 1964-1989 y para las 17 Comunidades autónomas españolas, siendo 221 el número total de observaciones¹.

(*) Agradecemos los comentarios recibidos de un evaluador anónimo a una versión anterior de este trabajo así como los de Lorenzo Serrano. Obviamente, los posibles errores son de nuestra exclusiva responsabilidad. Asimismo, agradecemos la ayuda financiera de la DGICYT (PB94-1523).

(1) Si bien el periodo de tiempo analizado es el comprendido entre 1964 y 1989, el número de años para los que se dispone de información de las variables producción y empleo es T=13, dado que la fuente estadística utilizada (BBV) sólo ofrece información con carácter bienal.

1. ESTIMACIÓN Y CONTRASTES

Se van a examinar las estimaciones y contrastes que ofrecen los programas LIMDEP, TSP y RATS centrándonos en el modelo de efectos aleatorios ya que es en este modelo donde tienen lugar las diferencias más importantes entre programas.

La parte final de este epígrafe se dedica a los contrastes de Hausman, de autocorrelación y a la estimación de modelos con efectos temporales. En toda la exposición se va a considerar que el número de individuos es de N y que el tamaño de la muestra es la misma e igual a T para cada individuo.

1.1. Modelo de efectos aleatorios

El modelo de efectos aleatorios se puede expresar de la siguiente forma:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_K X_{Kit} + \varepsilon_{it} + u_i \quad [1]$$

donde el término de perturbación es compuesto, siendo u_i el efecto aleatorio específico del individuo i -ésimo.

La estimación por MCG del modelo [1] se obtiene como media ponderada entre las estimaciones *within-groups* y *between groups*. En la ponderación interviene la *ratio*:

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2} \quad [2]$$

donde σ_ε^2 y σ_u^2 son las varianzas de ε_{it} y u_i .

La estimación por MCG equivale a transformar la ecuación [1], restando a cada observación del regresando y de los regresores la media del individuo multiplicada por el factor $(1-\sqrt{\lambda})$, y aplicando MCO al modelo transformado.

Como λ no se conoce, no se puede aplicar el método MCG propiamente dicho. Cuando se utiliza un estimador de λ , el método aplicado es una estimación de mínimos cuadrados generalizados (EMCG). La estimación de λ requiere a su vez estimar las varianzas de σ_ε^2 y σ_u^2 . En los tres programas de ordenador considerados se utilizan distintos tipos de estimadores, lo que da lugar a estimaciones que pueden diferir de forma importante. Examinamos a continuación los procedimientos utilizados por los programas LIMDEP, TSP y RATS.

1.1.1. Estimación de σ_ε^2 y σ_u^2 en el programa LIMDEP

En el programa LIMDEP se utilizan tres estimadores alternativos según los casos. En el primer procedimiento puede ocurrir que la estimación de σ_u^2 sea negativa. Cuando esto ocurre se aplica el segundo procedimiento en el que también puede plantearse el mismo problema, en cuyo caso se aplica el tercer procedimiento en el que, por construcción, no puede darse ese supuesto. El motivo de que en los dos primeros procedimientos puedan producirse estimaciones negativas de σ_u^2 se debe a que la estimación se obtiene como diferencia entre estimaciones de dos varianzas, calculadas por métodos diferentes.

Para los dos primeros procedimientos se toma como base de la estimación la varianza del término de perturbación del modelo *between-group*, que viene dada por,

$$\text{var}(\bar{\varepsilon}_i + u_i) = \frac{\sigma_\varepsilon^2}{T} + \sigma_u^2 \quad [3]$$

La estimación de σ_ε^2 en los tres procedimientos, de acuerdo con el manual del LIMDEP, viene dada por

$$\hat{\sigma}_\varepsilon^2 = \frac{\text{SCRW}}{N \times T - N - K - 1} \quad [4]$$

donde SCRW es la suma de cuadrados de los residuos obtenidos al estimar por MCO el modelo *within-group*. En la estimación anterior el programa LIMDEP no aplica en el denominador de [4] los grados de libertad correctos, que deberían ser $N \times T - N - K$, ya que el número de restricciones del modelo intragrupos es $N + K$, aunque sólo se estiman K parámetros debido a que previamente se toman desviaciones con respecto a las medias de cada individuo.

Como se ha indicado, la estimación de σ_u^2 en los dos primeros procedimientos se obtiene de forma residual. Por ello es necesario estimar el primer miembro de [3].

En el primer procedimiento, la estimación de σ_u^2 se obtiene a partir de la Suma de Cuadrados de los Residuos *Between-group* (SCRB), es decir,

$$\text{SCRB} = \sum_{i=1}^N \hat{v}_i^{2*} = \sum_{i=1}^N [Y_i - (\hat{\alpha}^b + \hat{\beta}_1^b X_{1i} + \hat{\beta}_2^b X_{2i} + \dots + \hat{\beta}_K^b X_{Ki})]^2 \quad [5]$$

Por analogía con [3] se obtiene el siguiente estimador:

$$\hat{\sigma}_u^{2*} = \frac{\text{SCRB}}{N - K} - \frac{\hat{\sigma}_\varepsilon^2}{T} \quad [6]$$

En el segundo procedimiento se utiliza la SCR obtenida al sustituir en el segundo miembro de [5] las estimaciones *between-group*, designadas mediante el superíndice b , por estimaciones de β obtenidas por MCO aplicados al modelo [1], al que designaremos mediante el superíndice t , es decir,

$$\sum_{i=1}^N \hat{v}_i^{2**} = \sum_{i=1}^N [Y_i - (\hat{\alpha}^t + \hat{\beta}_1^t X_{1i} + \hat{\beta}_2^t X_{2i} + \dots + \hat{\beta}_K^t X_{Ki})]^2 \quad [7]$$

La estimación de σ_u^2 en el segundo procedimiento es, por tanto, la siguiente:

$$\hat{\sigma}_u^{2**} = \frac{\sum_{i=1}^N \hat{v}_i^{2**}}{N - K} - \frac{\hat{\sigma}_\varepsilon^2}{T} \quad [8]$$

Como se ha indicado, si [6] resulta negativa se aplica el segundo procedimiento. Si también [8] resulta negativa, entonces se aplica el tercer procedimiento. En este úl-

timo, las u_i se estiman mediante los correspondientes efectos fijos *centrados*, es decir, mediante las desviaciones de los efectos fijos respecto a la media de los efectos. Así pues, la media de los efectos fijos se define de la siguiente forma:

$$\bar{\hat{\alpha}} = \frac{\sum_{i=1}^N \hat{\alpha}_i}{N} \quad [9]$$

La estimación de la varianza σ_u^2 en este tercer procedimiento viene dada por:

$$\hat{\sigma}_u^{2***} = \frac{\sum_{i=1}^N (\hat{\alpha}_i - \bar{\hat{\alpha}})^2}{N} \quad [10]$$

1.1.2. Estimación de σ_e^2 y σ_u^2 en el programa TSP

Para la estimación de σ_e^2 y σ_u^2 , el programa TSP toma como referencia la varianza del término de perturbación del modelo [1]. Esta varianza es igual a:

$$\text{var}(\varepsilon_{it} + u_i) = \sigma_e^2 + \sigma_u^2 \quad [11]$$

Para la estimación de la varianza correspondiente al primer miembro de [11] se utiliza la suma de cuadrados de los residuos obtenidos al estimar por MCO el modelo [1]. A esta suma la denominaremos SCRT, debido a que esta estimación se suele denominar *total o pooled*.

La estimación de σ_e^2 se obtiene, como en el programa LIMDEP, a partir de la suma de los cuadrados de los residuos (SCRW) obtenidos al estimar por MCO el modelo *within-group*.

En el TSP se utilizan dos procedimientos alternativos para estimar σ_e^2 y σ_u^2 , denominados de pequeñas y grandes muestras, respectivamente. En ambos procedimientos se utilizan los estadísticos SCRT y SCRW. Cuando la fórmula de pequeñas muestras produce un resultado negativo en la estimación de σ_u^2 , el TSP aplica automáticamente el procedimiento de grandes muestras.

En muestras pequeñas, el estimador de σ_e^2 viene dado, según el manual del TSP, por:

$$\hat{\sigma}_e^{2p} = \frac{\text{SCRW}}{N \times T - N - K} \quad [12]$$

Como puede verse, los grados de libertad utilizados en [12] son los correctos a diferencia de lo que ocurre en el programa LIMDEP.

Sin embargo, en la salida ofrecida por la versión 4.2 del TSP correspondiente al modelo de efectos aleatorios *VWITH (variance of Uit)*, en lugar de [12] se aplica, según hemos podido comprobar en todos los casos que hemos analizado, la siguiente fórmula:

$$\hat{\sigma}_e^{2p} = \frac{\text{SCRW}}{N \times T - N - K - 2} \quad [12\text{bis}]$$

No hemos encontrado ninguna explicación plausible de la reducción en dos grados de libertad en la estimación de esta varianza. No obstante, en la nueva versión 4.3 del TSP se aplica [12].

El estimador de σ_u^2 en muestras pequeñas viene dado por,

$$\hat{\sigma}_u^{2p} = \frac{\text{SCRT}}{N \times T - K - 1} - \frac{\text{SCRW}}{N \times T - K - N} \quad [13]$$

Como consecuencia de la corrección por grados de libertad de la SCRW, la fórmula [13] puede dar lugar a un resultado negativo. En ese caso, el TSP propone los siguientes estimadores que no son insesgados, pero sí consistentes. Por ello, se les denomina estimadores de muestras grandes.

En muestras grandes, el estimador de σ_ε^2 viene dado por

$$\hat{\sigma}_\varepsilon^{2G} = \frac{\text{SCRW}}{N \times T} \quad [14]$$

En muestras grandes, el estimador de σ_u^2 viene dado por

$$\hat{\sigma}_u^{2G} = \frac{\text{SCRT} - \text{SCRW}}{N \times T} \quad [15]$$

1.1.3. Estimación de $s2\varepsilon$ y $s2u$ en el programa RATS

Como ya se ha indicado, en el programa RATS el usuario debe introducir un conjunto de instrucciones para obtener la estimación del modelo de efectos aleatorios. Concretamente, y siguiendo el ejemplo 12.2 del Manual del usuario de RATS para la estimación de σ_ε^2 y σ_u^2 , se realiza la descomposición de la varianza de los residuos obtenidos en la estimación por OLS del modelo [1], es decir, en la estimación *pooled* o *total*.

La descomposición del análisis de la varianza anterior se puede expresar de la siguiente forma:

$$\text{SCRT} = \text{SCRb} + \text{SCRw} \quad [16]$$

donde SCRb es la suma de cuadrados de los residuos *between-group* (INDIV) y SCRw es la suma de cuadrados de los residuos *within-group* (ERROR).

Las instrucciones subsiguientes del ejemplo 12.2 del manual del usuario del RATS, implica que las estimaciones de σ_ε^2 y σ_u^2 vienen dadas respectivamente por

$$\hat{\sigma}_\varepsilon^2 = \frac{\text{SCRw}}{N \times T - N} \quad [17]$$

$$\hat{\sigma}_u^2 = \frac{\text{SCRb}}{N - 1} \quad [18]$$

En la estimación propuesta por el RATS para σ_ε^2 no se tiene en cuenta que la SCR está sometida a K restricciones y, consecuentemente, en el denominador de [17] se debe deducir también K.

Aunque los estimadores son sesgados en muestras pequeñas, la ventaja que tiene el método propuesto por RATS es que siempre se obtienen estimaciones de σ_u^2 no negativas.

En el cuadro 1 se muestran los coeficientes estimados y sus estadísticos-t en los programas LIMDEP, RATS y TSP correspondientes a la estimación del modelo de efectos aleatorios del ejemplo utilizado.

A la vista de los resultados se aprecian las importantes diferencias existente entre los tres programas (las diferencias llegan a ser del 37% en la estimación de la elasticidad asociada al capital público). Si bien en el ejemplo que estamos manejando las diferencias no afectan a la significatividad estadística de los parámetros estimados, puede haber casos en los que alguna variable, en función del programa utilizado, deje de ser significativa y de ahí las consecuencias para la investigación que se lleve a cabo.

1.2. Contraste de Hausman

En principio, los efectos individuales pueden considerarse siempre aleatorios. Ahora bien, para que la estimación por MCG permita obtener estimaciones consistentes de los parámetros es preciso que los efectos aleatorios no estén correlacionados con las variables explicativas. El contraste propuesto por Hausman para determinar dicha correlación está construido de la siguiente forma:

$$W = [\hat{\beta}^W - \hat{\beta}^{EMCG}]' \hat{\Sigma}^{-1} [\hat{\beta}^W - \hat{\beta}^{EMCG}] \quad [19]$$

donde $\hat{\Sigma}^{-1}$ es la diferencia entre la matriz de covarianzas de β^W y la matriz de covarianzas de β^{EMCG} .

La matriz $\hat{\Sigma}^{-1}$ es el estimador de una matriz que es definida positiva.

Cuadro 1

	LIMDEP	RATS	TSP
C. Privado (K)	0,3868 (11,843)	0,4167 (11,743)	0,3917 (12,129)
Empleo (L)	0,5244 (12,348)	0,5478 (12,424)	0,5287 (12,714)
C. Público (G)	0,1373 (3,972)	0,0862 (2,344)	0,1289 (3,776)
Tendencia (t)	0,0185 (7,901)	0,0198 (7,580)	0,0187 (8,033)
Constante (a)	-36,737 (-8,037)	-39,329 (-7,725)	-37,177 (-8,175)

Entre paréntesis, la *ratio*-t

Sin embargo, suele ocurrir con mucha frecuencia que la matriz que calculan los programas TSP y LIMDEP, programas que ofrecen el contraste de Hausman en la salida estándar, no gozan de la propiedad de ser definidas positivas. Cuando se presenta esta situación, que como hemos dicho es frecuente, el tratamiento dado por dichos programas es diferente.

Nuestra experiencia muestra que el programa LIMDEP asigna un valor 0 en el contraste de Hausman en el caso en el que exista al menos un elemento negativo en la diagonal principal de la matriz obtenida como diferencia de las matrices de varianzas-covarianzas de la estimación del modelo de efectos fijos y de efectos aleatorios, y sin que se diga expresamente en el manual.

En la versión 4.2 del programa TSP, el valor del contraste de Hausman depende en muchas ocasiones del orden en el que se introduzcan los regresores en la estimación. Dado que el orden de introducción de los regresores no debería afectar a los cálculos que se realizan en el contraste de Hausman, parece desprenderse que los algoritmos utilizados por el TSP para invertir la matriz Σ^{-1} presentan alguna anomalía. No obstante, en la nueva versión 4.3, dicho problema ha sido resuelto.

1.3. Autocorrelación

1.3.1. Autocorrelación en el LIMDEP

Con técnicas de panel, el LIMDEP ofrece un tratamiento de la autocorrelación de orden uno. En este caso, la estimación de los modelos se hace dos etapas. En una primera etapa, el modelo se estima ignorando la existencia de autocorrelación con el objetivo de estimar el parámetro autoregresivo de los residuos. La estimación se realiza o bien en el modelo de efectos fijos o bien en el modelo de efectos aleatorios. En una segunda etapa, se estima nuevamente el modelo mediante la transformación de Cochrane-Orcutt.

1.3.2. Autocorrelación en el TSP

El programa TSP, en el contexto del tratamiento genérico de la autocorrelación de orden uno, realiza una estimación *maximoverosímil* cuando se dispone de un panel de datos. Este tratamiento de la autocorrelación al margen de los modelos de panel, conlleva que, a pesar de haber utilizado previamente las instrucciones necesarias para estratificar la muestra, sea necesario particionar nuevamente la muestra con objeto de que el programa tenga conocimiento cuando comienzan y terminan los datos correspondientes a cada individuo de tal forma que en el tratamiento de la autocorrelación no mezcle información de distintos individuos.

El problema que presenta este tratamiento es que, estrictamente hablando, no corresponde a técnicas de panel ya que no tiene en cuenta la existencia del efecto individual, por lo que es un tratamiento de la autocorrelación en un *pool* de datos. No obstante, y en el contexto del modelo de efectos fijos, este inconveniente puede ser salvado introduciendo las correspondientes variables *dummies* individuales.

1.4. Estimación con efectos temporales

La estimación de los modelos de efectos fijos y aleatorios con efectos temporales tan sólo está expresamente implementada en el LIMDEP. Concretamente, LIMDEP estima el siguiente modelo de efectos fijos:

$$Y_{it} = \alpha_0 + \alpha_i + \gamma_t + \beta X_{it} + \varepsilon_{it} \quad [20]$$

imponiendo, para evitar la multicolinealidad perfecta, la siguiente restricción:

$$\sum_i \alpha_i = \sum_t \gamma_t = 0 \quad [21]$$

El modelo de efectos aleatorios adopta la siguiente especificación:

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} + u_i + w_t \quad [22]$$

En TSP y RATS, la estimación con efectos temporales tan sólo requiere su introducción como regresores adicionales. Sin embargo, la anomalía que presenta la versión 4.2 del TSP es que la estimación no es invariante al efecto temporal excluido, si bien en la nueva versión 4.3 ya no aparece dicha anomalía.

2. CONCLUSIONES

El objetivo de este trabajo ha sido analizar el tratamiento de los modelos estáticos de datos de panel realizado por los programas LIMDEP, TSP y RATS. Dicho análisis comparativo ha permitido detectar las siguientes anomalías²:

a) En el contraste de la hipótesis nula de igualdad de los coeficientes de los efectos fijos, en el estadístico F correspondiente, los grados de libertad del numerador deben ser N-1. Si bien el programa LIMDEP aplica correctamente dichos grados de libertad en este contraste, el programa TSP utiliza erróneamente N grados de libertad.

b) En la versión 4.2 del programa TSP, el contraste del modelo de efectos fijos *versus* aleatorio (contraste de Hausman) depende en muchas ocasiones del orden en el que se introduzcan los regresores en la estimación, si bien en la nueva versión del programa (4.3), ya no aparecerá dicho problema.

c) Nuestra experiencia muestra que el programa LIMDEP asigna un valor 0 en el contraste de Hausman en el caso en el que exista al menos un elemento negativo en la diagonal principal de la matriz obtenida como diferencia de las matrices de varianzas-covarianzas de la estimación del modelo de efectos fijos y de efectos aleatorios, sin que se diga expresamente en el manual.

d) En la estimación del modelo de efectos fijos, el estadístico R² corregido difiere entre el programa RATS y TSP, a pesar de que ambos programas realizan la estimación en desviaciones con respecto a las medias individuales. Ello es debido a que el programa TSP, al contrario que el RATS, considera que los efectos fijos son regresores, por lo que los grados de libertad que utiliza son menores y el R² ajustado también menor.

(2) En relación a las anomalías detectadas, hemos recabado la opinión de los servicios técnicos tanto de LIMDEP como de TSP. De TSP hemos recibido contestación en la que nos confirman las 4 anomalías por nosotros detectadas en la versión 4.2. Como nosotros hemos comprobado, 3 de las anomalías han sido corregidas en la versión 4.3, pero no la cuarta. Sobre esta última, nos agradecen el haberla detectado y reportado, y Clint Cummins nos ha indicado que se corregirá en la próxima versión del programa.

Respecto del programa LIMDEP, el profesor Greene está de acuerdo en que quizás sería conveniente hacer una llamada de atención al usuario cuando el programa reporta un valor cero en el test de Hausman como consecuencia de que la diferencia de matrices varianzas-covarianzas no sea definida positiva. Igualmente, nos confirma la existencia de un error en el manual en relación a los grados de libertad de la estimación de σ^2_ε , si bien dicho error no afecta al programa.

e) En la estimación de la varianza del término de error del modelo de efectos aleatorios, el manual del LIMDEP da como grados de libertad $N \times T - N - K - 1$, siendo $N \times T - N - K$ los grados de libertad correctos. No obstante, dicho error no afecta al programa.

f) Asimismo, la estimación propuesta por RATS para esta misma varianza considera $N \times T - N$ grados de libertad sin tener en cuenta las K restricciones a las que está sometida la suma de cuadros de los residuos utilizada en la estimación.

g) En la versión 4.2 del programa TSP, la estimación de los modelos introduciendo efectos temporales es sensible al efecto temporal excluido. No obstante, en la nueva versión 4.3 del programa ya no aparece dicho problema.

h) Si bien la varianza estimada del término de error (u_{it}) del modelo de efectos aleatorios realizada en la versión 4.2 del TSP se obtiene a partir de la estimación del modelo de efectos fijos, la salida correspondiente a dicha estimación reportada en este último modelo no coincide con la utilizada. Concretamente, ambas estimaciones difieren sistemáticamente en dos grados de libertad. Dicha anomalía ha sido corregida en la nueva versión del programa (4.3).

Como evaluación global de los programas podemos decir que de los tres analizados, el RATS es el que requiere mayores tareas de programación. LIMDEP es el programa que ofrece un mayor número de contrastes, si bien el contraste de Hausman en muchas ocasiones carece de validez dada la estrategia que adopta. Así, es necesario advertir de esta estrategia a los usuarios de este programa, ya que un valor del contraste igual a cero no implica necesariamente el no rechazo de la hipótesis de ausencia de correlación entre los efectos individuales y las variables explicativas.

Por último decir que la comparación de las versiones 4.2 y 4.3 del programa TSP muestra cómo casi todas las anomalías que presentaba la vieja versión del programa ya no aparecen en la nueva.



REFERENCIAS BIBLIOGRÁFICAS

Arellano, M., y S. Bond (1988): "Dynamic Panel Data Estimation Using DPD. A Guide for Users".

Limdep (1991): "User's Manual and Reference Guide", Version 6.0, William H. Green.

TSP (1990): "TSP User's Guide", Version 4.2, TSP International.

TSP (1995): "TSP User's Guide", Version 4.3, TSP International.

Rats (1992): "User's Manual", Version 4, by Thomas A. Doan.

Fecha de recepción del original: mayo, 1995

Versión final: agosto, 1996