

TSP: UNA REVISIÓN

JOSÉ M. LABEAGA
UNED y Fundación Empresa Pública

TSP ha sido, probablemente, el paquete más utilizado entre los economistas para la solución de sus problemas empíricos. Existen múltiples razones que pueden justificar esta amplia aceptación. El bajo coste de aprendizaje en relación a otros paquetes (y, por supuesto, a los lenguajes de programación), la facilidad de construcción de las instrucciones o de realización de cualquier tipo de transformaciones algebraicas, la amplia gama de resultados susceptibles de ser guardados tras la estimación y que pueden usarse en modelos posteriores, la facilidad de manejo de las instrucciones de operación con matrices y, en definitiva, el ser versátil y completo lo ha hecho y sigue haciéndolo atractivo.

Comentaré la versión 4.2A⁽¹⁾ centrándome especialmente en los apartados de modelos con variables dependientes cualitativas y limitadas y datos de panel. A pesar de su nombre original (Time Series Processor), el paquete ha incluido, desde sus inicios, una mezcla de métodos para datos tanto de series temporales como de corte transversal. También es posible, a partir de la versión 4.2A, estimar modelos para datos de panel. Además, no sólo incorpora instrucciones estadísticas o econométricas sino que permite la simulación de modelos económicos y experimentos de Montecarlo, predicciones tanto de variables como de modelos macroeconómicos y algunas posibilidades en temas financieros.

La versión evaluada (y también la más actual) está diseñada para funcionar en cualquier entorno informático: grandes ordenadores, miniordenadores, estaciones de trabajo y ordenadores personales y en cualquier entorno operativo: DOS, OS2, UNIX y MACINTOSH. No obstante, todos los comentarios versarán sobre la base de su utilización en ordenadores personales con sistema operativo DOS. Para este tipo de ordenadores existen dos alternativas: TSP286 y TSP386. La primera es compatible con versiones de DOS 2.0 o posterior, requiere un mínimo de 512 Kb de memoria (preferiblemente 640 Kb) y no necesita coprocesador matemático aunque es recomendable disponer de él si se desea realizar múltiples cálculos. TSP386 está pensado para su utilización en ordenadores 386 ó 486, requiere versiones del DOS 3.0 o superiores, al menos 1 Mb de memoria expandida, coprocesador matemático y disco duro con al menos 1.4 Mb libres para utilizarlo en versión normal y 3 Mb libres para utilizar memoria virtual.

La diferencia entre ambas es la capacidad de manejo de datos y, consecuentemente, de estimación de modelos. En la versión TSP386, el tamaño de los modelos que se pueden estimar está limitado por la disponibilidad de memoria RAM (en modo normal) y por la capacidad del disco duro (usando memoria

(1) Actualizada en 1991. TSP es distribuido por TSP Internacional; P.O. Box 61015, Station A. Palo Alto, CA 94306, USA. Teléfono: (415) 326-1927. Fax: (415) 328-4163. En estos momentos está disponible la versión 4.2B con la que ampliaré los comentarios en relación a las novedades incorporadas.

virtual). La versión más amplia sin uso de memoria virtual con la que el paquete es flexible, es en la actualidad la que dispone de 15 Mb de memoria expandida con capacidad de almacenamiento y proceso de hasta tres millones de números simultáneamente. La capacidad del programa se puede ampliar utilizando las posibilidades de la memoria virtual a costa de reducir sensiblemente la velocidad de proceso.

La comunicación del usuario con el programa siempre se efectúa utilizando expresiones cortas y sencillas que coinciden normalmente con la abreviatura, en inglés, del modelo que se pretende estimar o el método que se utiliza. Los modos en que TSP permite trabajar son tres: i) Modo interactivo (escribiendo las instrucciones en pantalla directamente). ii) Modo semiinteractivo o agrupado, escribiendo en pantalla un conjunto de instrucciones como si se tratara de un programa, que no se ejecutan hasta tanto no se dé la orden oportuna. iii) Proceso por lotes, que como en casi todos los paquetes o lenguajes es la ejecución que permite explotar de la mejor forma todas sus posibilidades.

1. LA ESTRUCTURA BÁSICA DE UN PROGRAMA: POSIBILIDADES DE TSP

1.1. *Construcción y manejo de datos*

En primer lugar conviene identificar el tipo de datos (series temporales o no y su frecuencia). No hay grandes limitaciones en este apartado. Se trabaje en cualquiera de los modos mencionados, los datos se introducen de forma sencilla o se disponen dentro del propio programa. Sin embargo, lo más común será tenerlos en un fichero al que se accede con las instrucciones de lectura, identificación y definición, las opciones relativas al formato cuando sea necesario y los nombres que deseamos para las variables. Los tipos de formato de lectura posibles son: libre (datos no formateados separados por espacios en blanco) y fijo (generalmente formato FORTRAN). Como casos especiales se pueden leer ficheros en formato de TSP, de hojas de cálculo LOTUS y EXCEL o bien binario con precisión simple y doble⁽²⁾.

1.2. *La sección principal*

Primero, es posible utilizar las instrucciones de identificación de valores no disponibles, la selección de la muestra (con exclusión de valores erróneos, el control de los no disponibles o cualquier cálculo o proceso sobre submuestras), la creación y transformación de variables (con posibilidad de todos los operadores algebraicos, relacionales y lógicos, opciones específicas para la creación de variables ficticias, creación de retardos o adelantos de variables, etc), utilización de funciones (logarítmicas, trigonométricas), generación de números aleatorios (pudiéndose elegir la mayoría de distribuciones, la semilla de comienzo, la matriz de varianzas y covarianzas y el vector de medias de la distribución en el caso de las normales multivariantes) o cálculos de valores para las distribuciones estadísticas más comunes.

En los comentarios anteriores he considerado que las operaciones se realizan sobre series (variables) no sobre matrices. El complemento para operaciones con matrices también está disponible entre las instrucciones de TSP si bien las trans-

(2) Disponer los datos en este formato resulta muy útil cuando se utilizan grandes ficheros ya que el acceso a los mismos es extremadamente rápido.

formaciones sobre partes de una serie o las combinadas entre escalares, series y matrices generalmente no funcionan bien o son complicadas de programar.

A partir de la versión 4.2A la utilización de instrucciones matriciales se ha completado y simplificado. Así es posible, leer matrices de cualquier tipo (generales, simétricas, diagonales), crear matrices a partir de variables (vectores), transformar, manipular, deshacer y rehacer matrices, realizar funciones matriciales que tengan como resultado escalares (determinantes, rangos), funciones matriciales con resultados matriciales (vectores y valores propios, diagonalizaciones), todo tipo de operaciones y procesos (traspuesta, inversa, producto de Kroneker) y programar procedimientos para la selección de elementos pertenecientes a matrices.

Estadísticos descriptivos y estimación de ecuaciones lineales

En este subapartado, TSP incluye la posibilidad de describir la muestra y realizar estimaciones de modelos con cualquier método (mínimos cuadrados ordinarios y en dos etapas, mínimos cuadrados generalizados, máximo verosímiles con información limitada, variables instrumentales, modelos autoregresivos con distintas opciones para la estimación del coeficiente de correlación, modelos de retardos distribuidos con diversas opciones para la longitud del polinomio e incluso la opción de los retardos con coeficientes estocásticos). También están disponibles las opciones de realizar regresiones ponderadas y calcular los errores estándar robustos a heterocedasticidad. TSP permite identificar, estimar y predecir modelos ARIMA, así como estimar modelos ARCH y GARCH, vectores autoregresivos y modelos de regresión con parámetros cambiantes. Además, incorpora la posibilidad del test de Dickey-Fuller de raíces unitarias para los residuos.

Estimación de modelos no lineales y ecuaciones simultáneas

Probablemente las instrucciones de TSP para la estimación de modelos de regresión no lineales y ecuaciones simultáneas (lineales o no) han sido y continúan siendo los bloques mejor desarrollados y más completos del paquete. Existen dos procedimientos generales: mínimos cuadrados y máxima verosimilitud. La instrucción de mínimos cuadrados calcula el estimador de distancia mínima y el método elegido depende de la opción que se suministre al programa. Estas opciones son mínimos cuadrados no lineales uniecuacionales, de dos etapas, de tres etapas, regresiones aparentemente no relacionadas y método generalizado de los momentos que calcula el estimador de Hansen. La instrucción de máxima verosimilitud obtiene estimaciones asintóticamente eficientes para modelos no lineales con errores normales.

Es importante reseñar que las instrucciones generales para la estimación de este tipo de modelos son prácticamente las mismas que para los modelos lineales por mínimos cuadrados, aunque las opciones en estos últimos son mucho más numerosas. Una ventaja, que es también relevante en términos del tiempo necesario para realizar los cálculos, es que estas rutinas utilizan derivadas analíticas en comparación a la mayoría de los paquetes estándar en los que la práctica habitual es el uso de derivadas numéricas, lo que, además, resulta útil para evitar los errores que se puedan cometer en la programación.

En el apartado de modelos de ecuaciones simultáneas se pueden obtener estimaciones por mínimos cuadrados y máxima verosimilitud con información completa tanto para modelos lineales como no lineales. De nuevo, como en el caso anterior, en el proceso de estimación se utilizan las primeras derivadas analíticas.

Asimismo, existe la posibilidad de completar las instrucciones con una variada gama de opciones.

Modelos para variables dependientes cualitativas y limitadas

A pesar de que dentro de la oferta al usuario siempre se ha contemplado una mezcla de modelos para series temporales y cortes transversales, no se ha dado a este apartado un papel preponderante. No obstante, las últimas versiones han mejorado las posibilidades de estimar modelos de estas características. Sólo tiene programados tres modelos de elección discreta, Probit y Logit binarios y Logit multinomial sin restricciones en el número de alternativas que se puede considerar. En modelos con variable dependiente limitada, están disponibles las instrucciones para los modelos Tobit y generalizado de selección de muestra.

Las posibilidades que ofrece el paquete pueden parecer limitadas pero no lo son tanto, dado que es posible la estimación de modelos mediante la instrucción de máxima verosimilitud. Así, cualquier modelo de elección discreta secuencial u ordenado, Probit bivalente (programando la distribución normal bivalente), modelo truncado, Tobit doblemente censurado, P-Tobit u otras alternativas al Tobit son susceptibles de estimación mediante la programación de la función de verosimilitud.

La estimación de los modelos anteriores se completa con múltiples opciones. De hecho, todas las disponibles para modelos no lineales también lo están para la instrucción de máxima verosimilitud. Existe la posibilidad de examinar las derivadas (numéricas y analíticas) de la función evaluadas en el vector de parámetros iniciales, especificar el método de cálculo de la matriz de varianzas y covarianzas asintótica, cambiar la aproximación del hessiano que se evalúa durante las iteraciones, el número de iteraciones, el criterio de tolerancia que determina la convergencia de los parámetros, etc. En definitiva, las limitaciones históricas para la estimación de estos modelos se han cubierto de forma bastante satisfactoria.

Modelos para datos de panel

TSP permite obtener estimaciones de modelos de regresión lineal con datos de panel, con un funcionamiento equivalente al de la instrucción de mínimos cuadrados ordinarios. Los paneles pueden ser completos (mismo número de observaciones temporales por individuo) o incompletos, con la restricción sobre la disposición de los datos relativa a que las observaciones correspondientes al mismo individuo han de ser correlativas (si no están dispuestas así en el fichero de datos, es posible realizar dicha ordenación). La información que se necesita para identificar el número de períodos por individuo se le ha de proporcionar bien mediante una variable de identificación individual bien a través de una opción que especifique el número de períodos por individuo (panel completo) u otra variable que identifique los períodos disponibles (panel incompleto).

Se pueden obtener estimadores de mínimos cuadrados ordinarios, intragrupos, entregrupos y mínimos cuadrados generalizados. No se estima de forma automática el modelo con efectos individuales y temporales (ni suponiendo que son fijos ni considerando el modelo de componentes de error con dimensión temporal), aunque siempre es posible introducir efectos fijos de tiempo creando el conjunto de dummies. Finalmente, es posible guardar las matrices de varianzas y covarianzas de todos los estimadores para la realización de contrastes de homogeneidad y especificación.

Contraste de hipótesis

Las alternativas en este apartado son múltiples tanto en los contrastes que de forma automática genera el paquete, como en las posibilidades de programación. Además, no sólo se pueden obtener los estadísticos mediante las opciones disponibles en cada uno de los modelos (por ejemplo, posibilidad de obtener hasta cinco diagnósticos para correlación serial de orden uno, contrastes de autocorrelación de orden superior, contraste de raíces unitarias en los residuos, tests de homocedasticidad, etc) sino que existe una instrucción específica susceptible de ser utilizada para contrastar cualquier tipo de restricción tanto en modelos lineales como no lineales (utilizando un método de linearización analítica de las funciones alrededor de los valores estimados de los parámetros y usando el método delta para la evaluación de las matrices de varianzas y covarianzas correspondientes).

Otras posibilidades

El paquete incorpora opciones útiles para otros campos de análisis. Por ejemplo, se pueden resolver y simular modelos de ecuaciones simultáneas sin restricciones en relación al número de ecuaciones (muy útil, por ejemplo, para modelos macroeconómicos). También se pueden calcular series de capital a partir de series de inversión bruta y ratios de depreciación, calcular índices divisa de precios y cantidades a partir de series de los mismos o corregir la estacionalidad de las series por el método de medias móviles.

1.3. Los resultados

Existe una instrucción general que permite controlar la ejecución y los resultados. Sin embargo, la potencialidad del paquete la constituyen principalmente las opciones asociadas a cada método de estimación. Por citar un ejemplo, la instrucción de mínimos cuadrados permite desde solicitar el mínimo nivel de información en los resultados (parámetros, errores estándar y estadísticos t) hasta completar un cuadro con un resumen amplio que incluya el estadístico h de Durbin, contrastes de Multiplicadores de Lagrange de autocorrelación, diagnósticos de estabilidad de los coeficientes, estadísticos de normalidad u otros. Finalmente, es posible en cualquier punto del programa guardar variables o matrices en ficheros externos con todos los formatos y opciones mencionados en las instrucciones de lectura.

2. LÍMITES EN EL USO DE TSP

En este epígrafe voy a tratar de subrayar algunos de los límites que conlleva la utilización de TSP y las mejoras que, desde mi punto de vista, se podrían incorporar. En cuanto al manejo de los datos, el tratamiento de las observaciones erróneas debe ser mejorado y creo que debiera darse la posibilidad de leer variables de tipo carácter. Debiera ser posible disponer de mayores posibilidades para la frecuencia de los datos (por ejemplo, semanal y diaria). Los usuarios acostumbrados a trabajar en modo interactivo o con menús echarán en falta mejoras en estos apartados aunque no parece que sea ésta la dirección a seguir por los autores.

Las instrucciones de operación con matrices se han mejorado en las últimas versiones. Sin embargo, creo que un aspecto importante en este contexto que no está bien logrado es la posibilidad de seleccionar partes de una matriz (por ejemplo, para realizar contrastes sobre subconjuntos de parámetros). Aunque es posible la selección de la submatriz necesaria que se puede realizar bien mediante la

asignación de subíndices bien por medio de operaciones de deshacer y rehacer, ello no es inmediato. Éste es un aspecto claramente mejorable. Incluso sería importante la posibilidad de utilizar matrices en la construcción de ecuaciones (lo que permitiría programar algunos estimadores de forma mucho más sencilla) y dentro de instrucciones de estimación como la de máxima verosimilitud.

Las operaciones con partes de una serie o entre matrices, variables y escalares, así como las instrucciones de control son susceptibles de programar pero de forma bastante complicada y, desde mi punto de vista, este apartado también es manifiestamente mejorable.

En el tratamiento de series temporales existen algunas limitaciones como la imposibilidad de realizar análisis espectral, estimar funciones de transferencia, u otros ajustes por estacionalidad. Se echan en falta más estadísticos de raíces unitarias, ampliación de los métodos para la estimación de modelos AR1, tests de cointegración a la Johansen, errores estándar a la Newey-West, etc.

El esfuerzo que los autores han realizado a lo largo de muchos años en los apartados de modelos no lineales y ecuaciones simultáneas ha hecho que no sea necesaria la introducción de muchas mejoras en las actualizaciones. De hecho, la versión 4.2B no incluye cambios importantes y tampoco se contemplan en la versión 4.3. Está en la mente de los autores la mejora de la opción del método generalizado de momentos y la posibilidad de inclusión de segundas derivadas analíticas en la instrucción de máxima verosimilitud.

En el epígrafe de modelos para variables dependientes cualitativas o limitadas se debe seguir avanzando en la misma dirección que con la versión actualizada. A pesar de que es posible estimar prácticamente cualquier modelo utilizando la instrucción de máxima verosimilitud, se debería disponer de forma automática de la distribución normal bivariante (que, así como la programación de otros modelos, sólo aparece entre los ejemplos en la versión 4.2B) y tampoco supondría excesivas complicaciones la incorporación de modelos de elección discreta ordenados, modelos Logit anidados, modelos de Poisson, algunas alternativas de modelos de duración, algún modelo simultáneo con variable dependiente cualitativa o limitada, posibilidad de disponer de contraste de especificación o creación de los residuos generalizados.

Ya se ha comentado en la sección anterior que TSP es, probablemente, el mejor paquete para la estimación de modelos con datos de panel⁽³⁾ tanto por las posibilidades de obtener los parámetros bajo diversas hipótesis como por los resultados que son guardados una vez se han estimado los modelos. La versión 4.2B mejora estas posibilidades en un aspecto importante que es la alternativa de programar modelos para datos de panel utilizando la opción de estimación del método generalizado de los momentos. Queda, sin embargo, mucho camino por recorrer. Por citar algunas posibilidades de ampliación, se pueden proponer más contrastes de especificación, procedimientos de estimación por variables instrumentales con efectos fijos y cálculo de errores estándar robustos⁽⁴⁾ o estimación de ecuaciones simultáneas.

(3) Véase la comparación que realiza MacKie-Mason (1992).

(4) Los interesados en la estimación de modelos lineales por el método generalizado de los momentos con datos de panel disponen de un excelente programa, Dynamic Panel Data, escrito en Gauss por Arellano y Bond (1988).

Otro de los aspectos manifiestamente mejorables de TSP lo constituye el apartado relativo a los contrastes que automáticamente realiza tras la estimación de los modelos. Se podría incluir algún contraste de normalidad en modelos lineales adicional al de multiplicadores de Lagrange, más posibilidades de contrastar no linealidad, estadísticos adicionales de raíces unitarias como el de Phillips-Perron, algún diagnóstico de homocedasticidad y normalidad en modelos con variable dependiente cualitativa o limitada o algún test en el tema de datos de panel.

Finalmente, parece oportuno señalar ampliaciones en línea con procedimientos semiparamétricos o no paramétricos. Existen algunos paquetes como LIMDEP y lenguajes como GAUSS que han seguido esta dirección en las últimas versiones. TSP parece haberse quedado al margen en las actualizaciones más recientes, aunque, se ha previsto su inclusión en la versión 4.3, siempre en función de los deseos que los usuarios manifiesten.

3. COMENTARIOS FINALES: COMPARACIONES Y VALORACIONES

El aprendizaje de TSP es sencillo y se ve facilitado por una documentación y una descripción de errores excelentes que se pueden comparar de forma ventajosa con la mayoría de los paquetes existentes. Además, la documentación cuenta con las referencias más importantes para cada uno de los modelos, lo que resulta muy útil en no pocas ocasiones. No es posible, sin embargo, utilizar TSP en forma de menú y su uso interactivo tampoco resulta satisfactorio (PC-GIVE o MICROFIT son más completos en este apartado). Además, solamente dispone de un editor insuficiente para trabajar en modo interactivo y la posibilidad de ayudas es muy limitada.

La potencialidad de un paquete se ve completada cuando permite la programación de procedimientos y el control de su ejecución. Ambos aspectos están incluidos (con algunas deficiencias) en TSP y se están mejorando en las nuevas versiones. Debe quedar claro, por otra parte, que no está pensado como un paquete para extraer muchas posibilidades en la construcción y manejo de gráficos. En este punto es claramente inferior a otros paquetes como RATS o MICROFIT o lenguajes como GAUSS.

La lectura, transformación y procesamiento de datos es extremadamente flexible con TSP. Permite que los ficheros contengan datos en la mayoría de los formatos, ordenados por variables o por observaciones. La misma flexibilidad se puede aprovechar para la escritura de los ficheros de resultados. La transformación de variables resulta muy sencilla y no tiene prácticamente límites en cuanto a operaciones, transformaciones especiales, funciones o distribuciones estadísticas pero sí muchas deficiencias en la combinación de operaciones con matrices, variables y escalares. Salvando estas deficiencias, TSP está, desde mi punto de vista, a la altura de los mejores como RATS, SAS/ETS o LIMDEP.

El tercer punto a resaltar consiste en comparar las posibilidades estadístico-económicas y es el fundamental en la decisión sobre la elección del paquete. En aplicaciones de series temporales, TSP ofrece casi todas las posibilidades (aunque con ciertas peculiaridades): AR, ARCH, GARCH, VAR, modelos ARIMA con posibilidades de identificación, estimación y predicción. En relación a modelos integrados y cointegrados realiza automáticamente contrastes de raíces unitarias (Dickey-Fuller) para los residuos y calcula los puntos críticos del test

de cointegración Engle-Granger. Permite, además, la estimación y el contraste del orden de integración y otras hipótesis.

En aplicaciones de modelos para variable dependiente cualitativa o limitada sólo existen cuatro modelos programados que se obtienen de forma automática (Probit, Logit binomial y multinomial, Tobit y generalizado de selección de muestra). No obstante, la instrucción de máxima verosimilitud permite estimar casi cualquier modelo. Finalmente, la posibilidad de estimar modelos con datos de panel hace de TSP un programa muy completo. La versión 4.2B lo completa aún más ya que es posible, con límites, programar modelos de variables instrumentales por el método generalizado de los momentos.

Comparando globalmente todas las posibilidades de estimación y contraste de hipótesis, TSP es, en conjunto y a mi juicio, el paquete más completo. Sin embargo, se puede argumentar que RATS es mejor para series temporales y concretamente para problemas de integración y cointegración y LIMDEP es más completo para modelos con variable dependiente cualitativa o limitada. Por otra parte, TSP es claramente superior para modelos no lineales, sistemas de ecuaciones y datos de panel.

Los contrastes que de forma automática realiza al mismo tiempo que la estimación de los diversos modelos, se completan con las posibilidades que permite una instrucción específica para contrastar cualquier tipo de restricciones y la flexibilidad para programarlos por medio de las instrucciones de matrices cuyo uso es sencillo. Mi opinión es que TSP está en este apartado al nivel de la mayoría de los paquetes conocidos como LIMDEP, RATS y SAS/ETS.

La posibilidad de realizar simulaciones, predicciones, componentes principales, control exhaustivo de los ficheros de salida o uso de los resultados en modelos posteriores, terminan de completar un paquete que por su sencillez, flexibilidad y posibilidades sigue siendo un buen candidato en la elección de los economistas aplicados. Sin embargo, de los comentarios anteriores se desprende que la combinación de varios paquetes es tal vez la mejor elección.

No siquiera terminar esta nota sin realizar una mención expresa a la absoluta disposición de los autores de TSP para recibir cualquier crítica, sugerencia o comentario que les permita corregir errores en la programación de alguna instrucción o mejorar cualquiera de los apartados del paquete.



REFERENCIAS BIBLIOGRÁFICAS

- Arellano, M. y S. Bond (1988): *Dynamic Panel Data Estimation Using DPD - A Guide for Users*, Institute for Fiscal Studies, WP número 88/15.
- MacKie-Mason, J.K. (1992): "Econometric Software: A User's View", *Journal of Economic Perspectives*, 6, págs. 165-187.