

24 de febrero de 2005

## TRABAJO: ELABORACIÓN Y EXPOSICIÓN, MICROECONOMETRÍA, DATOS DE PANEL

Teodosio Pérez Amaral,

[teodosio@ccee.ucm.es](mailto:teodosio@ccee.ucm.es)

91 3942380,

### 1. Objetivo

El trabajo consiste en realizar un análisis econométrico, utilizando algunas de las técnicas aprendidas en el curso, y alguna de las bases de datos de panel del libro de Wooldridge que están disponibles en la página de la asignatura y también en <http://www.msu.edu/~ec/faculty/wooldridge/book2.htm>. El trabajo hay que presentarlo en papel por escrito y exponerlo oralmente usando powerpoint o similar.

Cada uno puede elegir el trabajo que quiera, usando uno de los conjuntos de datos de Wooldridge u otros, siempre que no lo haya elegido nadie antes. Para reservar un tema basta enviarme un correo-e con el nombre del fichero de datos elegido y una descripción en pocas líneas del posible trabajo. El primero que envía el correo tiene prioridad. El profesor ha de dar su visto bueno a cada trabajo, para asegurarse de que resulta factible y útil.

### 2. Estructura del trabajo escrito.

El trabajo tendrá una estructura organizada en secciones, por ejemplo (puede variar):

1. **Resumen.** Breve descripción del problema, modelo, datos y principales conclusiones.
2. **Introducción.** Motivar el tema y explicar su interés.
3. **Presentación de datos.** Usar gráficos o estadísticos descriptivos. Mostrar solamente los que aporten algo útil para la argumentación.
4. **Modelo** teórico o empírico. Explicando el por qué de cada variable y su signo esperado. Elasticidades.
5. **Estimación e interpretación de los resultados.** Resumen del modelo favorito estimado, estimaciones puntuales, t, F, y otros. Diagnóstico del modelo favorito. Contrastes de hipótesis de interés. Restricciones. Variantes del modelo.

6. **Conclusiones.** Qué se ha aprendido. Resumir lo que antes no se sabía y gracias a esos datos, modelo y nuestro arte y esfuerzo ahora se sabe. Insistir en las principales conclusiones. Quizá algunas sugerencias para futuros trabajos.

7. **Referencias.** En formato estándar. Ej

Rodríguez, G. (2002) “Datos sincrónicos y de panel”, Revista J., Vol. 69, nº 3, 178-192.

Se puede consultar también las sugerencias de Engle en el documento “Outline for a paper in applied economics”.

Todo lo anterior se expondrá de forma breve. Se puede encontrar ayuda muy valiosa: explicación de variables y modelos, en el texto de Wooldridge, porque los datos corresponden a ejemplos del libro, en la descripción de Gorka y Paloma de las bases de datos y en el DATA SET HANDBOOK que nos envió Wooldrige. Hay que esforzarse por sintetizar, exponer ideas e ilustrar hechos usando gráficos. Idealmente el trabajo debería poder resumirse en 2 ó 3 gráficos. Las tablas y gráficos son más útiles si están incluidos en el texto.

### **3. Formato del trabajo escrito.**

El ejercicio se presentará escrito en un máximo de 10 hojas DIN A-4, con procesador de palabras, usando letra de 12 puntos por pulgada y un espacio y medio. Hay que utilizar ecuaciones, subíndices y superíndices, letras griegas, tablas, cuadros y gráficos (quizá importados de Excel).

Los apellidos y nombre del autor se escribirán en la esquina superior derecha de la primera página, en letra negrita. El trabajo se elaborará de forma individual.

Los trabajos se redactarán de forma clara y concisa, facilitando al máximo su lectura. Se evitarán párrafos largos sin puntos. Se usarán los acentos de forma correcta, así como los correctores ortográfico y gramatical. Se escribirá todo en español, sin mezcla de otros idiomas, en particular los cuadros y gráficos importados de otros programas.

Se dará al trabajo un aspecto lo más profesional posible. Se preferirán esquemas y gráficos que resulten visualmente atractivos. El trabajo se presentará con sobriedad y elegancia. El uso del color, portada, tapas u otros elementos se deja al buen criterio del alumno. El trabajo se presentará impreso en papel. El trabajo se entregará en el momento de iniciar la exposición oral.

#### **4. Exposición.**

El trabajo se presentará en clase. Cada persona dispondrá de un máximo de 10 (diez) minutos para la exposición. Se usará el ordenador, video proyector y programa de presentación. Nunca transparencias en acetato u otros medios obsoletos.

Se dará prioridad a que la audiencia comprenda los puntos principales del problema y las principales conclusiones. No se insistirá en aspectos metodológicos básicos, que todos deben ya saber.

Durante la exposición tiene que quedar claro:

1. Cuál es la pregunta (el problema) que se quiere abordar. Por qué es importante.
2. Qué modelo teórico (o empírico) se va a usar.
3. De qué datos exactamente se dispone. Variables, años, posibles deficiencias, etc.
4. Qué modelo/s econométrico/s se va/n a estimar. Qué metodología econométrica se usa y por qué. Cómo se interpretan los resultados empíricos.
5. Qué se ha aprendido con el trabajo. Qué conclusiones, en términos de comportamiento humano, es razonable extraer. Insistir en las principales conclusiones aunque se hayan comentado antes.

Resulta imprescindible llevar la exposición ensayada. Para eso es muy conveniente repetirla varias veces, con público del que se pueda obtener comentarios útiles, hasta que salga bien. La exposición es importante. Estáis “vendiendo” vuestro trabajo y a vosotros mismos como profesionales. Entre otras cosas muestra lo seguro y convincente que uno puede ser hablando en público. Hay que tener pensados recursos para cuando te pillen en un renuncio, o no sepas qué contestar.

A casi nadie se le ocurre meterse en un Jacuzzi vestido de chaqué, ni presentar un telediario en rulos y bata de guatiné. Cada situación requiere una vestimenta específica. Cuando uno está trabajando de cara al público tiene que vestirse de forma adecuada. De manera que resalte lo mejor de sí mismo y el fruto de su trabajo luzca mejor

O sea que hay que vestir con el máximo glamour pero informal, a la par que sencillo y proletario. Bien peinado y rasurado. Evitando zapatillas de tenis, vaqueros, camisetas de aspecto arrastrado, etc.

También hay que procurar mostrar nuestra mejor sonrisa, cuidar dónde se coloca uno, cómo se mira al público, el volumen de voz, y dónde se ponen las manos. Es conveniente no tocarse repetidamente ninguna parte del cuerpo: nariz, oídos, rodillas...

## 5. La realización del trabajo paso a paso.

Las 12 sugerencias que siguen se pueden usar como una receta de cocina para hacer el trabajo de forma ordenada y completa.

1. Antes de comenzar el análisis empírico resulta imprescindible **conocer la muestra**, saber cuántos individuos, qué variables exactamente, años, si es incompleta cuántos y cuáles faltan, si faltan de forma aleatoria o no, posibles valores atípicos, etc. **HACER GRAFICOS DE CADA VARIABLE, Y NUBES DE PUNTOS (SCATTER PLOTS) DE LA VARIABLE DEPENDIENTE CON CADA UNA DE LAS POSIBLES EXPLICATIVAS.**

2. Como cada uno tiene ya un conjunto de datos de panel, lo primero que necesita para analizarlo es un modelo. Puede ser un **modelo teórico** sugerido por la microeconomía, por los problemas del libro, por el artículo en el que salieron publicados los resultados del trabajo original o por el sentido común -que es el menos común de los sentidos- económico. A veces tenemos que conformarnos solo con un **modelo empírico**, sugerido por los datos disponibles y la intuición del investigador, pero sin apoyo teórico.

3. Idealmente, el modelo teórico nos facilitaría escribir una **especificación econométrica**, que pudiésemos estimar. Sabríamos la/s variable/s dependiente/s, las variables explicativas, el signo esperado de cada derivada parcial y quizá el tamaño aproximado de cada parámetro. Además, podríamos tener una idea de si puede haber simultaneidad u otros problemas y comprobarlo luego.

4. Puede que sepamos o tengamos una idea de si el modelo está en **niveles**, en diferencias, tasas de crecimiento, en logaritmos, tiene forma semilogarítmica, etc. Si no, tendremos que comprobarlo, teniendo cuidado de no transformar las variables ficticias. Hay que pensar

bien y decidir si el modelo hay que estimarlo en términos per capita, o hay alguna otra **variable de escala** que resulte razonable.

5. A continuación podemos **estimar** el modelo de la manera que resulte apropiada, quizá por efectos fijos, luego variables, luego hacer el contraste de Hausman, etc. O bien por variables instrumentales. **Es fundamental interpretar los coeficientes, comentar su signo, la significatividad individual y conjunta de las variables explicativas y calcular y comentar las elasticidades.**

6. También hay que comprobar si se dan o no, y justificar en cualquier caso la conclusión, los **problemas típicos** que hemos visto en el curso: errores de medida en las variables explicativas, modelos dinámicos, ecuaciones simultáneas, paneles incompletos, etc. En cada caso habrá que estimar de forma adecuada. La comprobación se hará mediante argumentos teóricos y el sentido común económico.

Si no se estima un modelo dinámico habrá que **justificar por qué**. Si no se estima un modelo con posibles errores de medida, lo mismo. En caso de que no haya simultaneidad en una ecuación habrá que argumentar por qué no la hay. En caso de que la justificación no sea sólida, habrá que presentar, además, la estimación por variables instrumentales del modelo. En el caso de paneles incompletos, habrá que comparar siempre la estimación del panel completo con el incompleto.

7. Luego hay que comprobar si se da o no la **homogeneidad de coeficientes** entre los distintos grupos de individuos. Por ejemplo, homogeneidad de constantes entre individuos usando los contrastes F explicados en clase. También comprobar la **estabilidad (homogeneidad) espacial** de todos los coeficientes entre, por ejemplo 2 (o más) grupos de la muestra: norte-sur, este-oeste, hombres-mujeres, etc. Dependiendo de cada caso los grupos serán unos u otros. La información para segmentar la muestra de esta forma se hallaría en la propia base de datos. Se usarían contrastes F tipo Chow o con variables ficticias.

Los contrastes de Chow comparan la suma residual (SR) del modelo restringido, para toda la muestra, con la suma de las sumas residuales de los modelos sin restringir en cada una de las (por ejemplo 2) submuestras ( $SR_1$  y  $SR_2$ ). Si son aproximadamente iguales, no habrá cambio estructural (habría homogeneidad de coeficientes) y si no, no la habría. Esto también se podría hacer, de forma más flexible utilizando variables ficticias, si las sumas residuales no están disponibles.

8. También conviene comprobar la **estabilidad (homogeneidad) temporal** de todos los coeficientes dividiendo la muestra en dos periodos (con suerte con un criterio razonable, si no, a la mitad) y aplicando un contraste tipo Chow a las dos submuestras o usando variables ficticias.

La **estabilidad temporal** de los coeficientes se comprueba también corriendo regresiones de corte transversal para cada una de las fechas de la muestra. Los resultados en términos de coeficientes estimados, por ejemplo para cada año, se presentarán en un cuadro resumen.

En el cuadro resumen, los años están en el eje vertical y los coeficientes estimados (con su desviación típica estimada debajo) para cada año en el eje horizontal. Cada columna de coeficientes estimados lleva encima el nombre de la variable a la que corresponden esos coeficientes y desviaciones típicas estimadas. El contraste se realiza de forma visual (no formal), considerando también las desviaciones típicas.

9. La introducción de **restricciones** entre los coeficientes puede ayudar a mejorar la eficiencia de la estimación de manera muy notable. El alumno planteará al menos una restricción razonable entre los coeficientes y en caso de no ser rechazada por los datos (usando un contraste F o similar dependiendo del caso) proceder a su incorporación al modelo. Posiblemente mediante una simple sustitución o eliminación de una variable.

10. **Nuevas ficticias.** En ocasiones hay suficiente información cualitativa en la muestra, o ficheros .des .raw o en los problemas del libro o en la descripción de los datos que está en la página de la asignatura, o en el DATA SET HANDBOOK que envió Jeff Wooldridge como para que el alumno pueda construir y utilizar con facilidad alguna variable ficticia (cuantitativa) que tenga relevancia para el análisis de los datos de panel utilizados, y que pueda ayudar a mejorar el modelo econométrico.

Un ejemplo típico es el caso de una variable ficticia para un solo momento del tiempo (ej: un año) que puede servir para recoger un hecho especial, extraordinario, o atípico pero común a todos los individuos, ocurrido ese año (por ejemplo un ciclón). La ficticia valdría 1 para ese año para todos los individuos y 0 para todos los demás años de todos los individuos.

Si el hecho a recoger mediante esa ficticia afectase solamente a una parte de la muestra (submuestra 1), entonces la ficticia sería igual que en el caso anterior para esa submuestra 1 y tomaría el valor 0 para todos los períodos e individuos que no pertenezcan a esa submuestra 1.

11. Conviene trabajar de **forma ordenada**, guardando en hojas numeradas y fechadas las impresiones de todo lo que se hace y escribiendo sobre las impresiones, a mano, los comentarios con respecto a qué se hace, por qué y qué se concluye. Puede resultar práctico ir escribiendo el texto del trabajo en papel simultáneamente a ir realizando las estimaciones. Posteriormente se puede extraer del fichero del texto del trabajo la presentación de Powerpoint a base de recorta y pega.

12. No es necesario que cada alumno haga con su modelo todo lo que se sugiere en este epígrafe. En ocasiones, debido a limitaciones de los propios datos o la estructura del trabajo, algunas de las sugerencias resultarán inviables, dificultosas o irrelevantes. En general, se espera que se hagan **al menos un 50% de las sugerencias** contenidas en este epígrafe.

## **6. Evaluación del curso.**

El trabajo en papel cuenta un 25% de la nota del curso, la exposición otro 25% y un 50% los exámenes escritos. Las intervenciones en clase cuentan positivamente. Si se ha aprobado por curso, no hay que presentarse al examen final.

## **7. Ayuda por parte del profesor.**

El trabajo no es una batalla entre alumnos y profesor. Al contrario, se trata de que el profesor ayude a los alumnos todo lo que pueda, para que les salga un buen trabajo. No pedir ayuda no es un mérito. Y hay que pedirla pronto, sin perder tiempo por un atasco. Se puede preguntar antes o después de clase, en horas de consulta o por correo-e.

## **8. Otros.**

### **A. Bullshitting.**

Si durante una exposición una pregunta o sugerencia del público revela una carencia, error, incongruencia o contradicción en el trabajo que se está exponiendo, el ponente no se puede quedar callado, azorado o dubitativo. Por el contrario, tiene que tener preparado y utilizar algún recurso que le permita salir del apuro, y quizá pensarlo mejor más adelante.

- a. Me alegra que me haga esa pregunta, que es muy interesante y que paso a explicar... (y luego explicarlo como pueda...)
- b. Esa es una magnífica sugerencia que procedo a anotar para incorporarla en la próxima versión...
- c. Esa es una pregunta muy interesante que no había pensado desde ese punto de vista y que voy a considerar con mucha atención...
- d. Esa pregunta prefiero contestarla de forma privada después de terminar la exposición porque llevaría mucho tiempo hacerlo ahora porque nos desviaría del tema principal... Y luego salir corriendo.

O respuestas similares. Hay que responder con seguridad, mirando a los ojos del que ha preguntado. Quizá la respuesta c es la más honesta.

Casi todo el mundo en la sala se va a dar cuenta de la situación, así que no se debe intentar engañar a nadie, pero si salir airoso. En ocasiones es buena idea hablar con la persona que ha planteado la pregunta después de la exposición e intentar aclarar posturas.

El conjunto de respuestas anterior se enmarcan dentro de la categoría general de actividades que se llaman en inglés “bullshitting”.

### **B. Pensar con profundidad (think hard).**

El profesor Robert F. Engle de la Universidad de California en San Diego enseñaba Econometría Aplicada el pasado Siglo XX, cuando la mayoría de ustedes usaba pañales, e insistía en que la actividad más útil para modelizar era pensar duro, en profundidad -think hard- y... se rascaba la cabeza. Robert F. Engle ha recibido el premio Nobel de Economía 2003, conjuntamente con Clive Granger, también de la Universidad de California en San Diego.

Esa sugerencia sigue siendo válida. Frente a la vorágine de los datos, el ordenador, las salidas impresas y las instrucciones, hay que parar y replantearse de vez en cuando, solamente con un folio en blanco delante, si el modelo que se está usando tiene sentido económico, si hay variables posiblemente omitidas, si los coeficientes tienen tamaños y signos razonables, si los contrastes y restricciones que se plantean tienen sentido, etc. Todo desde el punto de vista del sentido común y la teoría económica relevante, usando un modelo simple.

### C. Estrategias de especificación.

En general, parece buena idea empezar con un modelo teórico relevante, escribir una especificación empírica estimable con los datos de los que disponemos y estimarla a continuación.

Para hacer eso, hay dos estrategias opuestas.

a. **KISS.** La primera estrategia sugiere ir de lo particular a lo general. Parte de un modelo lo más simple posible y lo va complicando solamente en la medida en que el modelo no explique las regularidades empíricas de los datos que se quieren modelizar. Esta es una estrategia de modelización muy utilizada en todo el mundo, acorde con la experiencia científica y con la experiencia de la vida cotidiana. El acrónimo KISS significa, en realidad, *keep it simple stupid!* Esto en modo alguno es un insulto, sino un reconocimiento de la ignorancia del investigador. Para que nadie se sienta aludido, también se interpreta como *keep it sophisticatedly simple!*

**b. Metodología de la LSE (London School of Economics) o de Hendry o de lo general a lo particular, o GETS (General to Specific).**

Este enfoque parte del extremo opuesto del anterior, considerando primero un modelo lo más general posible (difícil de precisar!) y luego lo va reduciendo a un modelo más simple a base de incorporar restricciones que los datos no rechacen basados en un conjunto de diagnósticos hoy en día estándar.

Arnold Zellner, famoso econométra bayesiano de Chicago, dice con frecuencia que nunca vió un modelo útil obtenido por la metodología de Hendry (LSE). Obviamente un comentario un poco malvado.

### D. Selección de variables.

En cuanto a qué variables explicativas o transformaciones o interacciones de las mismas usar en un modelo econométrico, se podrían distinguir también dos enfoques extremos.

a. El primero **guiado por la teoría**, en el que se usan solamente las variables tal y como las sugiere la teoría económica, sin considerar otras, ni posibles transformaciones, interacciones o formas funcionales más flexibles.

b. El segundo sería una estrategia de **modelización basada en los datos** (data driven) en la que las variables se incluirían en el modelo según la relación entre los datos de esa variable en la muestra con los de la dependiente. Esta estrategia de modelización se puede implementar de lo **general a lo particular**, empezando por un modelo muy amplio y eliminando variables que no contribuyan apreciablemente al modelo según algún criterio, y también de lo **particular a lo general** incluyendo las variables una a una, de mayor a menor poder de contribución a según algún otro criterio.

c. Seguramente una estrategia razonable consistiría en combinar las dos metodologías extremas anteriores, en las que la teoría sugiere las posibles variables relevantes y los datos cuáles de ellas lo son en la muestra y qué transformaciones (ratios, interacciones...) pueden ser aconsejables para esos datos.

e. Una estrategia mixta ha sido propuesta recientemente por Pérez Amaral, Gallo y White. Se trata de un método de construcción de modelos y selección de variables conocido por RETINA, en el trabajo: *A Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA)*, publicado en el Oxford Bulletin of Economics and Statistics, 2003. Este método ha sido implementado en un programa en GAUSS y está disponible para su uso con datos de corte transversal en la versión más reciente de Massimo Marinucci, RETINA for Windows.

Este trabajo está motivado por la búsqueda de modelos flexibles pero sencillos e interpretables que permitan aproximar fenómenos económicos y no económicos de manera apropiada.

En él nos centramos en el aspecto esencial de la modelización estadística de fenómenos económicos: en primer lugar la selección de variables estadísticamente relevantes, en segundo lugar la selección de una forma funcional que aproxime de manera adecuada el fenómeno que se intenta explicar, y por último la utilización de un criterio para la selección de modelos que esté basado en la **capacidad predictiva del modelo fuera de la muestra**.

El procedimiento resultante, RETINA, es computacionalmente simple, tiene mejores propiedades de muestras finitas que los procedimientos alternativos y da lugar a modelos interpretables económicamente.

## 9. Diagnosis.

Una vez que se ha especificado y estimado un modelo econométrico de cualquier tipo, con cualquier tipo de datos y a cuya especificación se ha llegado de la manera que sea, hay que proceder al análisis de la calidad y utilidad del modelo. Esta fase del trabajo empírico se llama diagnosis y en ella se analizan de forma sistemática y exhaustiva los supuestos en los que se fundamentan las propiedades de los estimadores del modelo y el uso que podemos hacer del mismo.

Para concretar y simplificar voy a escribir aquí debajo los supuestos que se utilizan habitualmente para la estimación e inferencia de un modelo econométrico y la forma habitual de comprobarlos en la fase de diagnosis. En caso de otros tipos de modelos o de datos, habría que modificar de manera apropiada la diagnosis.

Modelo Lineal General:  $y = X\beta + u$

### SUPUESTOS HABITUALES

#### S1. El modelo es estocástico.

No se contrasta. Es suficientemente general.

#### S2. El modelo está correctamente especificado.

Contrastes de variables omitidas y de variables irrelevantes, t, F,

Wald, Multiplicadores de Lagrange, Cociente de las Verosimilitudes.

Contrastes de simultaneidad, Hausman.

Homogeneidad de coeficientes, cambio estructural, Chow, F, Wald.

#### S3. $E(u) = 0$ .

No se contrasta. Siempre que haya una constante entre los regresores,

se cumple.

**S4.**  $V(\mathbf{u}) = \sigma_u^2 I$

Autocorrelación, Durbin-Watson, Breusch-Godfrey, acf y pacf residuales.

Heteroscedasticidad, Breusch-Pagan, White.

**S5. Linealidad en los parámetros.**

Contrastes de no linealidades y forma funcional. F, t, Wald.

**S6'. Las variables explicativas X son estocásticas.**

Endogeneidad, simultaneidad.

Contraste de que no están correlacionadas con el error. Hausman. (visto en S2).

**S7. Ausencia de multicolinealidad.**

**Perfecta:** matriz singular, no se puede estimar.

**Aproximada:** detectarla mediante  $R^2$  de regresiones auxiliares. Número de condición.

**Los contrastes se deben hacer en el orden en el que han sido expuestos. Grosso modo, primero los que están relacionados con la consistencia de los estimadores y luego los que lo están con la eficiencia.**