

La enseñanza de la Econometría en el Espacio Europeo de Estudios Superiores¹

Alfonso Novales
Departamento de Economía Cuantitativa
Universidad Complutense

Agosto 2009

¹ Revisión del material preparado para la presentación en el I Simposio de Profesores de Econometría organizado en la Universidad de Granada, el 3 de julio de 2009.

La enseñanza de la Econometría en el Espacio Europeo de Estudios Superiores

1. Introducción

A lo largo de los años he experimentado profundos cambios en el modo en que desarrollo la docencia en Econometría. De hecho, durante un período largo no impartí la materia, debido a la frustración que me generaban los métodos tradicionales y mi incapacidad por percibir una alternativa interesante. Cuando la tuve, volví a dar clase en las asignaturas de Econometría I y II que existen actualmente en las licenciaturas que ahora se extinguen, sustituidas por los Grados. Mi nuevo modo de ver las cosas se basaba en: a) una enseñanza mucho más aplicada, encaminada a un análisis riguroso de la evidencia muestral disponible, más que a un amplio desarrollo de resultados analíticos, y b) un sistema de evaluación continua que motivara el estudio continuo y posibilitara mi seguimiento continuo del avance de los alumnos, pudiendo colaborar con ellos en cubrir sus limitaciones.

La aparición del Espacio Europeo de Estudios Superiores (EEES) dio cobertura a esta manera de actuar. Soy consciente de que el EEES encierra muchas otras consideraciones, pero quiero dejar claro que mi idea de EEES o lo que genéricamente denominamos Bolonia, se centra en los dos aspectos que he reseñado. En este sentido, soy partidario claro de estas instituciones docentes incorpóreas que se imponen ahora como norma. Estoy convencido de que la formación de nuestros alumnos puede mejorar muy sustancialmente y, con ello, su preparación como profesionales. Es también claro que ello conlleva para el docente un volumen de trabajo muy superior, que puede interferir con los requisitos de investigación que se imponen para su promoción, con negativas consecuencias para los profesores más jóvenes.

Adicionalmente, se trata de un método docente que es imposible aplicar con un elevado número de alumnos por grupo como, por otra parte, las mismas autoridades académicas que ahora se han lanzado alegremente a la implantación del método docente habían reconocido abiertamente con anterioridad. Esta inconsistencia, pretendiendo instaurar un método docente y de evaluación muy diferente e intensivo, en condiciones reconocidamente incompatibles con el mismo, es el mayor peligro al que se enfrenta el EEES, y puede ser la principal causa de que termine totalmente desvirtuado en sus aspectos más interesantes. Ojalá no suceda como nos advertía el Príncipe de Lampedusa: *“A veces es necesario que algo cambie para que todo siga igual”*.

Voy a dividir mi exposición en dos partes: 1) Método docente, y 2) Contenidos. Las ideas contenidas en la segunda parte tienen que ver con una discusión crítica de la utilización de métodos de inferencia estadística en ciencias sociales.²

2. Método docente

² El lector interesado puede encontrar información adicional en: *Las interrelaciones entre investigación y docencia en economía aplicada* (http://www.ucm.es/info/ecocuan/anc/charla_Zaragoza.pdf)

2.1. Situación de partida: Frustración

Toda conversación entre docentes deja evidencia clara de la frustración que genera el método docente tradicional en la Universidad, al menos en el área de Estadística y Econometría, en cualquier centro de ciencias sociales. Es importante considerar las causas de esta frustración para tratar de encontrar una solución alternativa que permita la ruptura con el tradicional estado de cosas.

Entiendo que esos factores son:

- Asistencia muy reducida de alumnos a clase y tutorías
- Muy escasa interacción entre alumnos y profesor, que suscita dudas en el profesor respecto de qué han aprendido realmente los alumnos que aprueban la asignatura
- Percepción de que se aprueba sin una verdadera comprensión de los conceptos fundamentales, lo que imposibilita la utilización en la práctica profesional posterior de lo aprendido en el curso
- Calificaciones generalmente muy bajas. Muchos alumnos aprueban tras rebajar apreciablemente el listón mínimo, ante la posibilidad de un continuo porcentaje de suspensos que no es asumible
- Aprendizaje excesivamente mecánico, con incapacidad manifiesta para resolver situaciones ligeramente distintas de las examinadas en clase
- Exámenes que en muchas ocasiones terminan siendo tipo test, ante el elevado número de alumnos, a sabiendas de que tal examen es incompatible con la posibilidad de discriminar entre una mayor o menor comprensión conceptual
- Dificultades para evaluar una enseñanza más práctica, nuevamente por el elevado número de alumnos matriculados, junto con su falta de asistencia a clase
- Dificultades para cubrir un programa teórico, por el elevado número de clases no impartidas, la necesidad de incidir en conceptos vistos en cursos anteriores pero que los alumnos no recuerdan o no asocian con el contexto en que los encuentran
- Y falta total de tiempo para estudios prácticos, como consecuencia de todo lo anterior

Todo lo cual conduce a una profunda insatisfacción con los métodos docentes que utilizamos. Pero, además, hay una cuestión que debería apremiar a los docentes en el contexto tradicional es: ¿Por qué es tan habitual aprobar sin ir a clase? Muchos alumnos aprueban sin asistir a clase ¿Por qué es así? A veces se me dice que se trata de alumnos repetidores, pero una rápida contabilidad muestra que no es así. El hecho es que una enseñanza mecánica, junto con una evaluación de curso basada en un examen final (quizá con dos pruebas parciales) y, especialmente, un posible examen tipo test, posibilitan esta situación. ¿Es tan útil una academia como la Universidad? Lo es para preparar un examen final en estas condiciones. No lo sería si la docencia universitaria fuese real, y se manifestara en un trabajo continuo y gradual a lo largo del curso. La pregunta que todo docente debe hacerse es precisamente ésta: ¿Por qué es realmente preciso que los alumnos asistan a clase? ¿Por qué no pueden aprobar la asignatura de otro modo? No tiene sentido que la obligación de asistencia a clase sea caprichosa o arbitraria, basada en el poder que da el control de las calificaciones; esto no es bueno para nadie.

Insatisfacción que no es independiente del método de evaluación que tradicionalmente hemos utilizado, basado en una prueba única. Uno o dos parciales a

mitad de curso no solucionan nada, y pueden incluso empeorar la situación. La cuestión es ¿Tiene sentido un examen final después de 15 semanas de docencia? ¿Tiene sentido estar 15 semanas subido a una tarima sin saber cómo progresan los alumnos o, lo que es peor, con la sospecha fundada de que no están estudiando, sino acumulando apuntes hasta la última semana de curso? ¿Tiene sentido tal esfuerzo docente sabiendo que el único estudio se realiza para superar un examen final y, dado la premura con que se lleva a cabo, el alumno nunca va a asimilar los conceptos? En definitiva, los métodos de evaluación deben y pueden jugar un papel fundamental en el aprendizaje del alumno; descubrir este aspecto central fue, en mi experiencia, definitivo para los cambios introducidos.

Se produce también entre muchos docentes una insatisfacción con los contenidos que impartimos, que no sabemos muy bien cómo se enmarcan con el resto de las materias, y que no entendemos cómo se relacionan con el trabajo empírico que realizan los profesionales que hace unos pocos años pasaron por las aulas. Personalmente, considero que el análisis empírico de profesionales e incluso de investigadores, en ocasiones, deja bastante que desear en términos de rigor estadístico, hasta el punto de poder sesgar totalmente las conclusiones que obtiene. Siendo así, un objetivo docente central debe ser que cuando los alumnos ocupen puestos profesionales o de investigación, no incurran en tales deficiencias.

2.2. El EEES como alternativa

En este contexto, como antes dije, el EEES viene a dar cobertura institucional a un tipo de actuaciones que considero definitivas:

- Poner el énfasis en aprender, no en aprobar. El alumno que sigue el curso aprueba porque ha aprendido.
- El trabajo y el aprendizaje deben ser continuados. Solo esto garantiza la comprensión y asimilación de conceptos y posibilita su rigurosa utilización práctica.
- Involucrar al alumno en su proceso de aprendizaje. El profesor no puede, por sí solo, lograr que el alumno aprenda. Puede y debe posibilitar que esto suceda, pero siempre sólo si el alumno está decidido a hacerlo.
- Por tanto, el alumno es responsable de su proceso de aprendizaje y, en consecuencia, de su aprobado. Pero lo importante es que desaparece la incertidumbre que vincula a ambos procesos. El alumno aprueba porque ha decidido aprender y ha seguido el programa de trabajo establecido por el profesor.
- Importancia de la programación global del material: coherencia entre desarrollo de conceptos teóricos y ejercicios prácticos. Esta es la principal responsabilidad del docente: estructurar las tareas aplicadas y las discusiones de clase de modo que si el alumno trabaja continuamente, aprenda el contenido propio de la materia.
- Finalmente, he de insistir en que el sistema de evaluación que se diseñe es clave, y debe concebirse como encaminado a) a discriminar lo antes posible a aquellos alumnos que no quieren o no estén en condiciones de realizar el esfuerzo preciso para aprender los conceptos de la asignatura, b) detectar las carencias y limitaciones analíticas o conceptuales de los alumnos que lo intentan, de modo que se puedan organizar sesiones de tutoría sobre temas específicos, o diseñar ejercicios con una finalidad concreta, c) calificar de manera diferenciada a los

alumnos que siguen el programa de trabajo establecido. Es muy importante que el alumno entienda este punto

La conclusión fundamental es que posiblemente los alumnos tengan carencias de conocimientos y pocos deseos de estudiar, pero hemos de reconocer que el sistema tradicional no favorece precisamente lo contrario. En definitiva, los profesores podemos hacer mucho más de lo que hacíamos en un sistema docente tradicional.

2.3. *Estrategia docente*

Ya he expuesto una buena parte de los aspectos que debe incorporar la estrategia docente. Siempre es difícil señalar un objetivo fundamental para un curso; los docentes somos dados a pensar que cada asignatura es de total trascendencia, todos los temas del programa son imprescindibles, y que hay muchos aspectos que cada asignatura aporta a la formación de los alumnos. Por el contrario, creo que es útil pensar en un único objetivo que nos facilite centrar nuestra atención a lo largo de todo el curso. Si tuviera que elegir tal objetivo para los cursos básicos de Econometría de las Licenciaturas y futuros Grados, diría que éste debe ser: *Aprender a hacer una interpretación rigurosa de modelos estimados de distinta naturaleza*. Conocer cuáles son los tipos de cuestiones que cada modelo está diseñado para responder, familiarizarse con la naturaleza de los datos económicos, aprender a explotar toda la información que contienen de modo descriptivo, expresar con precisión y concisión las preguntas objeto de análisis; todas estas y muchas otras que pudieran añadirse, son cuestiones encaminadas a lograr el objetivo fundamental que he mencionado.

En línea con el desarrollo por parte del profesor de un esquema gradual de aprendizaje, quiero enfatizar la importancia de hacer explícitos los conceptos y resultados a aprender cada semana. Es enormemente motivador para el alumno que éste centre sus ideas cada semana alrededor de dos ideas, y que perciba cómo estas ideas se enmarcan dentro del contenido global del curso. Es difícil para el profesor mantener esta disciplina, pero creo que es su tarea más importante.

A ello debe ayudar que los ejercicios prácticos se diseñen con una finalidad concreta, que es conveniente que incluso se describa en el enunciado del ejercicio. De este modo, el alumno ve que, lejos de ser un capricho, el ejercicio forma parte imprescindible del proceso de aprendizaje en el que se ha comprometido de acuerdo con el profesor. Es asimismo posible diseñar partes de un ejercicio de modo que el alumno deba leer algo de material docente, incluso antes de discutirlo en clase. Este es otro aspecto de suma utilidad; nunca va a aprender más el alumno que cuando se discuten en clase conceptos, prácticas y resultados que él ha trabajado previamente

2.4. *Métodos de evaluación*

El método de evaluación tiene una importancia fundamental, y sólo un procedimiento de evaluación continua es consistente con un aprendizaje sólido que, inexcusablemente, debe basarse en un estudio continuado. La evaluación debe utilizar ejercicios con datos, hechos en grupos reducidos (3 alumnos máximo), a ser posible con visitas esporádicas a las sesiones de tutoría, programadas por el profesor. En conjunción con tales ejercicios, pueden hacerse tests rápidos en clase. Estos pueden ser tests de 5 preguntas que requieran una rápida contestación, de una o dos líneas, a realizar en un total de 15 minutos. Las preguntas pueden ser relativas a aspectos muy concretos de los ejercicios ya entregados, que permitan saber si los alumnos individualmente, han

aprendido los procedimientos adecuados. No hay que plantearse que las preguntas sean exhaustivas del material discutido. No se trata de esto, sino de que los alumnos sientan que hay una supervisión real, y de que el profesor pueda hacer un seguimiento del trabajo y del aprendizaje de la materia por parte de los alumnos. Pensando en hacer un test casi todas las semanas y un ejercicio con datos cada dos semanas, aunque teniendo en cuenta la imposibilidad material de ser muy estrictos con tal programación, es posible finalizar un semestre de 15 semanas con 7 ejercicios prácticos y cerca de 10 tests de clase, información más que suficiente para evaluar el aprendizaje del alumno durante el curso. En mi práctica, he encontrado que proporcionar de manera continuada una relación ordenada de calificaciones numéricas, de modo que cada alumno pueda saber qué posición ocupa en el grupo, es muy estimulante tanto en casos de alumnos que están en la parte alta como de alumnos que están en la parte baja de dicho ranking.

Lo anterior sugiere, como estoy convencido, de que el examen final es prescindible. Desde luego, no puede confiarse la calificación del curso a tal examen. Además, con la cantidad de información que he reseñado, es difícil imaginar qué más va a aportar un examen final. Lo que es peor. El tipo de preparación de tal examen, muy intenso y durante un breve espacio de tiempo no es compatible con un aprendizaje de la materia. En los grupos que he impartido del modo que explico posteriormente no se celebran exámenes finales en la primera convocatoria.

Son, por el contrario, el diseño de trabajos empíricos y de discusiones en clase los que deben garantizar el aprendizaje. Los ejercicios pueden diseñarse de modo que los alumnos tengan que estudiar parte del material teórico que no es preciso explicar en clase en detalle. En todo caso, es siempre preferible explicar el material en clase después que el alumno lo ha utilizado. Entonces es cuando mejor lo aprende.

En un esquema de trabajo de estas características, el alumno decide si puede o no seguir el ritmo de trabajo. Esto es más razonable que el sistema hasta ahora habitual de exámenes finales. Apenas existirían los suspensos; pueden tener calificación de No Presentado, porque dejan el curso para otra ocasión o lo sustituyen por otro, pero el Suspenso tras asistir a clase durante el curso no tiene sentido. Los alumnos siguen los cursos para los que se sienten capacitados, y cuando están en condiciones de realizar el trabajo oportuno.

Por último, y no es una cuestión menor, quiero incidir en que copiar no debe considerarse un problema serio. Es un problema, pero el alumno que copia no finaliza el curso. No es posible, porque tarde o temprano va a quedar en evidencia. Incluso los ejercicios empíricos pueden ser comunes en parte con años anteriores, pues no puede generarse tanto material nuevo todos los años. Tampoco es necesario; también el profesor debe mentalizarse de que el objetivo es que el alumno aprenda, no tanto que pasa un único examen, con el posible problema de la copia. En todo caso, en los tests de clase se van a hacer preguntas al respecto, o en las tutorías, y los alumnos enseguida se dan cuenta de ello.

Una solución que ha funcionado hasta ahora correctamente es mi compromiso de que el alumno recibe como calificación a final del curso el promedio de las calificaciones de los tests de clase. El resto de la información que pueda acumular el profesor que, como he dicho antes, es mucha, a través de ejercicios prácticos en equipo, cuestiones en clase, participación en foros de discusión, etc., puede utilizarse aunque únicamente para mejorar la calificación que resultaría del criterio expuesto. La realidad es que el promedio de los tests proporciona una calificación muy reducida; por ello es conveniente establecer desde el comienzo que, sin asumir ningún compromiso al respecto, se valorará positivamente la mejoría que se haya producido durante el curso. Se trata de ver una evolución gradual en el aprendizaje, y que el resultado final sea

bueno, no necesariamente que el resultado haya sido bueno desde el inicio. De hecho, acumulo ya bastante experiencia de alumnos que han comenzado con limitaciones conceptuales importantes (en algunos casos, alumnos que provenían de Escuelas Universitarias), pero que han progresado a lo largo del curso para finalizar con calificaciones altas, claramente por encima de la media del grupo

2.5. Uso exhaustivo de una herramienta de Campus Virtual

Una herramienta de este tipo es central en el tipo de proceso de aprendizaje que he descrito. Esta herramienta debe contener, desde el comienzo: a) Descripción general del curso, b) Método docente: clases de discusión teórico-práctica, clases en aula informática, c) Procedimiento de evaluación, d) Distribución de lecciones por semanas. En el campus Virtual, el profesor deja disponibles a los alumnos material docente de todo tipo: a) apuntes de clase, b) presentaciones Power Point con conceptos principales y cuestiones de interpretación y aplicación, c) soluciones a los tests propuestos en clase, d) ejercicios aplicados propuestos, e) solución a dichos ejercicios, inmediatamente después de su entrega, f) calificaciones provisionales para cada alumno, dentro del ranking general el curso, g) utilización del foro de discusión para discusión de dudas entre alumnos. Este material se va actualizando y completando durante el curso, por lo que el Campus Virtual es una herramienta muy activa

Quiero resaltar dos de los aspectos anteriores, por la importancia que tienen en una motivación adecuada de los alumnos. Por un lado, que las soluciones a los ejercicios se hagan disponibles inmediatamente después de su entrega. La lectura de tales soluciones, cuya redacción debe ser mucho más detallada de lo que es exigible al alumno, es posiblemente el mejor método de aprendizaje. Es importante que el alumno se convenza de que los ejercicios están diseñados para aprender, no para examinarle. Por eso las soluciones que se les proponen son muchos más detalladas de lo que se espera de ellos. Hay que asegurarse de su lectura, lo que puede incentivarse incluyendo en los tests de clase alguna cuestión relativa a aspectos expuestos en dichas soluciones.

El segundo aspecto es la utilización del Foro de discusión que suele incluirse en una herramienta de Campus virtual. Hay que animar a los alumnos a que formulen sus dudas al resto de sus compañeros, incluso en el período de realización de los ejercicios. El profesor lee las preguntas y las respuestas, y entra a participar en el Foro de discusión cuando considera que es necesario, quizá porque las respuestas que se están cruzando son incorrectas, o no son completas. Recomiendo que se consideren tanto las preguntas como las respuestas para mejorar la calificación del curso

2.6. Principales dificultades para alumnos y profesores

Conocer cuáles son las principales dificultades con que se encuentran los protagonistas de esta historia, profesores y alumnos, motivará el diseño e implantación de soluciones a las mismas, por lo que voy a intentar establecer tal relación. Creo que las principales dificultades a que se enfrenta el alumno tienen que ver con:

- Dificultad para ajustarse a un nuevo modelo de formación
- Dificultad del alumno para tratar de comprender los conceptos teóricos en el tiempo de clase, lo que sugiere la conveniencia de diseñar ejercicios cuidadosamente, que le obliguen a entrar en dichos conceptos y sus propiedades e implicaciones
- Escasa costumbre de atender en clase
- Hábito de un aprendizaje excesivamente mecánico

- Dificultades en redactar
- Dificultades de expresión en público
- Excesivo número de materias
- Enseñanza poco coordinada entre materias

mientras que las principales dificultades para el profesor se centran en:

- Un número excesivo de alumnos
- El número de horas de trabajo: preparación de material, corrección de ejercicios y tests de clase, aumenta muy notablemente
- El mantenimiento y actualización de materiales en herramientas de Campus Virtual
- Hacer comprender al alumno la metodología de trabajo y aprendizaje, la importancia de su continuidad, y el protagonismo del alumno en asumir la responsabilidad del mismo
- La escasa sensibilidad (al menos actualmente) de la dirección académica por posibilitar las condiciones precisas para este tipo de enfoque docente
- En definitiva, la importante cantidad de trabajo que, como dije al inicio, puede interferir con las exigencias de investigación que recibe.

En definitiva, estamos ante la posibilidad de implantar un proceso docente innovador, dirigido exclusivamente al aprendizaje del alumno, que le permita entender y asimilar los conceptos, y familiarizarse con una utilización rigurosa de los mismos, lo que le resultará útil en su práctica profesional. Es un planteamiento muy distinto del tradicional, cuyos mayores retos estriban en la frecuente inercia de las instituciones académicas, de los alumnos y de los propios docentes, y en que las condiciones de trabajo son inadecuadas, especialmente en lo concerniente al número de alumnos.

3. Contenidos

Paso en esta sección a comentar sobre las implicaciones que los planteamientos docentes expuestos antes tienen sobre la organización de contenidos de la materia. Para ello, propongo un principio básico: *La evidencia empírica relativa a la cuestión que se estudia puede obtenerse mediante una variedad de procedimientos gráficos y estadísticos, paramétricos y no paramétricos. Utilicemos dicha variedad de métodos. Centrémonos en la cuestión, no en la metodología.*

3.1. Enfoque conceptual

- Desde un punto de vista general, considero que un buen ejercicio empírico comienza con un resumen detallado de la información muestral, utilizando exhaustivamente procedimientos gráficos y tests estadísticos no paramétricos que, como es bien sabido, tienen propiedades válidas en muestras finitas y robustas a incumplimientos de los supuestos de Normalidad y otros en los que habitualmente se basan otros contrastes.
- En segundo lugar, es importante dedicar un tiempo a analizar la situación de precisión de la estimación, evaluando, en la medida de lo posible, cuáles de los parámetros se han estimado con aceptable precisión, y cuáles no. Por supuesto, ésta no es una cuestión que se resuelva mediante el uso de valores p o introduciendo reglas de validez absoluta sobre la magnitud de los estadísticos t .

- Para este tipo de cuestiones, es interesante enfatizar los ejercicios de simulación, estimación y generación de distribuciones de probabilidad de estimadores. Se trata de simular datos muestrales bajo un determinado diseño experimental que recoja las características estadísticas que queremos considerar: heterocedasticidad, colinealidad, etc., obtener el estimador, y guardar el resultado obtenido con un número elevado de este tipo de muestras, para posteriormente caracterizar las propiedades de la distribución empírica del estimador: su histograma de frecuencias, su posible Normalidad, su media y varianza, etc.
- Solemos olvidar, y haciéndolo podemos caer en un error importante, que el contenido informativo o capacidad explicativa de una variable en un modelo que incluye varios factores explicativos es siempre condicional al que incorporan las restantes variables del modelo. Por supuesto, la colinealidad puede hacer que una variable con capacidad explicativa resulte no significativa al incorporarla a un modelo, debido a que otras variables incorporan bastante información en común con ella. Este principio básico explica también la aparente paradoja que frecuentemente inquieta a alumnos e investigadores acerca de cómo una variable parece tener capacidad explicativa en un modelo, pero no en otro.

Por esta misma razón, un coeficiente individual no mide casi nunca el efecto sobre Y de un cambio unitario en X , a pesar de la frecuencia con que se interpretan de este modo. En presencia de colinealidad, situación que debemos entender como habitual en Economía, cuando una variable experimenta una variación unitaria, cabe esperar que las restantes variables explicativas del modelo también varíen, y es este efecto total el que sería interesante medir. Un coeficiente habitual admite la interpretación mencionada bajo el supuesto *ceteris paribus*, es decir, si ninguna otra variable cambia. Esta interpretación es correcta, sólo que proporciona la respuesta a una pregunta generalmente poco interesante, debido a la correlación entre variables explicativas.

- Una vez estimado un modelo, es importante resumir las características de la estimación resultante examinando los valores ajustados de la variable dependiente sobre determinadas particiones de los rangos muestrales de las X . Se trata de calcular el valor medio de la variable dependiente sobre una partición de las variables explicativas, es decir, calcular su esperanza condicional sobre determinadas regiones del espacio muestral de las variables explicativas. Más que examinar propiedades teóricas de los coeficientes o de los parámetros estimados, es importante analizar cómo varía el valor implicado por el modelo estimado para la variable dependiente cuando las variables explicativas toman valores que el investigador considere interesantes.
- Tampoco debe olvidarse que las inferencias que con un modelo estimado puedan hacerse, deben llevarse a cabo dentro del rango muestral observado. Si la variable X toma valores en la muestra entre 0 y 20, por ejemplo, no debe utilizarse un modelo estimado para inferir cuál será el comportamiento de Y para $X=100$.
- Por último, parece importante evitar un excesivo resumen de la información muestral. En muchas ocasiones disponemos de información muestral muy rica, con un número grande de observaciones sobre un número relativamente alto de variables. Sin embargo, tras estimar los modelos que se han considerado adecuados, la inferencia se efectúa examinando un número muy reducido de estadísticos, como el ajuste global de los modelos alternativos a través de los valores del logaritmo de la verosimilitud respectiva. O se selecciona un modelo preferido, y se discute la relevancia relativa de las variables explicativas utilizando unos pocos estadísticos, cuando disponíamos quizá de miles de datos. Este aspecto está relacionado con la

conveniencia de examinar el ajuste del modelo de manera detallada, a lo largo de toda la muestra, sin reducirlo a una única medida global, al que luego me referiré.

3.2. Precisión y potencia

Un segundo bloque de comentarios se refiere a la relación entre precisión en la estimación y potencia en el contraste. Es necesario introducir el concepto de precisión estadística en estimación, aunque debe ser conocido de los cursos de Estadística. Basta considerar un concepto relativamente informal, vinculando la precisión al inverso de la desviación típica de un estimador, aunque las correlaciones entre parámetros deben jugar un papel en la evaluación de este concepto. Y, como comenté antes, una vez estimado un modelo, es preciso evaluar la situación de precisión

La idea general es que una reducida precisión conduce a conjuntos de confianza (intervalo de confianza, en el caso de un solo parámetro) muy amplios. Bajo el esquema Neyman-Pearson de contrastación de hipótesis, una hipótesis nula no se rechaza si el valor numérico considerado bajo la hipótesis nula cae dentro de la región de confianza. Lo cual, bajo una precisión reducida, sucederá para muchas hipótesis nulas que pudieran plantearse. La consecuencia es que todo un amplio conjunto de hipótesis paramétricas no son rechazadas o, equivalentemente, que la potencia del contraste es baja. En definitiva, una reducida precisión en la estimación conduce a una pérdida de potencia en la contrastación de hipótesis.

Creo que es importante adoptar el principio de no proceder a la contrastación de una hipótesis paramétrica excepto si los parámetros involucrados se han estimado con suficiente precisión. Así debe enseñarse a los alumnos, y es la razón por la cual es importante evaluar la situación de precisión de cada parámetro una vez estimado un modelo. Esta idea se extiende fácilmente a la precisión de una combinación lineal de parámetros (por ejemplo, su suma o diferencia, entre muchas otras posibilidades), cuando es su valor numérico lo que define la hipótesis nula. También se generaliza sin mucha dificultad al caso de varias restricciones. El Apéndice incluye lo que considero que son ejemplos interesantes y, posiblemente sorprendentes, en relación con las conexiones entre precisión y potencia.

Esta asociación entre precisión y potencia hace que el investigador deba ser especialmente cuidadoso al elaborar sus conclusiones cuando habiendo contrastado una (o varias) hipótesis no la rechaza. Si no se cuenta con suficiente precisión en la estimación, no rechazar la hipótesis nula no significa mucho. En particular, en modo alguno puede interpretarse en el sentido de haber encontrado evidencia a favor de la hipótesis nula, error de interpretación muy habitual.

Este serio problema se evitaría si se presentase, acompañando a cada contraste, la función de potencia del mismo como, por otra parte, enseña la teoría Estadística más básica. Es bastante incomprensible el olvido en que la función de potencia ha caído entre investigadores en las ciencias sociales. Una función de potencia que se eleva rápidamente hacia 1.0 cuando nos desviamos del valor paramétrico recogido en la hipótesis nula refleja la precisión en la estimación y justifica que el no rechazo de la hipótesis nula pueda interpretarse como evidencia muestral favorable a la hipótesis nula, lo que no siempre puede hacerse, como hemos dicho. Por último, conviene puntualizar que examinar el p-valor del contraste no es suficiente, y no aporta la misma información. Adicionalmente, la interpretación del p-valor no está exenta de errores: recordemos que el p-valor no es la probabilidad de que la hipótesis nula sea cierta, dada la muestra, sino más bien una indicación de la probabilidad de obtener una muestra de

las características de la que disponemos (en lo relativo a la hipótesis nula) si la hipótesis nula es cierta.

3.3. Significación estadística

Uno de los principales defectos en la interpretación de un modelo estimado consiste en la errónea identificación entre la significación estadística de un coeficiente con la capacidad explicativa, el contenido informativo o la relevancia económica de la variable a la que acompaña. No hay ninguna razón para esta identificación. De hecho, convendría mantener en mente que *no existe el concepto de variable significativa (o no significativa)*, contra las muchas afirmaciones que se leen en trabajos empíricos, en sentido positivo o negativo, acerca de la significación de una variable.

En la relación entre significación estadística de un coeficiente y la capacidad explicativa o contenido informativo de una variable, se pueden producir dos tipos de situaciones que inducirían a error si se identifican ambos conceptos.

Por un lado, recordemos la relación entre el estadístico t y la precisión en la estimación: el estadístico t se obtiene dividiendo el grado de incumplimiento de una restricción (u holgura) por la precisión en la estimación de la misma. Por tanto, una reducida precisión puede conducir a un valor numérico del estadístico t inferior al umbral crítico (habitualmente 2,0 en muestras grandes) incluso si el grado de incumplimiento (la holgura) de la restricción es asimismo grande. En el caso de un contraste de significación estadística de un coeficiente³, esto significa que podemos tener una estimación numéricamente alejada de cero a efectos de las razones que motivan el contraste y, a pesar de ello, no rechazar la hipótesis nula de ausencia de significación debido a una reducida precisión en la estimación del parámetro. Concluiríamos así equivocadamente, bajo la identificación conceptual que he criticado antes, la irrelevancia de la variable explicativa en cuestión, su ausencia de capacidad explicativa o de contenido informativo.

Pero también puede darse la situación contraria. Tamaños muestrales muy elevados (miles de observaciones) pueden generar estadísticos t espuriamente altos, incluso si el incumplimiento de la hipótesis nula es mínimo. En el caso de un contraste de significación, el parámetro estimado podría ser prácticamente nulo a efectos de la motivación que subyace al contraste y, sin embargo, rechazar la hipótesis nula, dado que la desviación típica de la estimación podría ser muy reducida (ver el segundo tipo de ejemplos en el Apéndice).

Conviene comentar que hay caminos alternativos que hay que explotar. Esencialmente, debemos recuperar la idea de evaluar un modelo estimado mediante el examen individualizado y detallado de los residuos que genera. Por ejemplo, en el caso de contrastar la capacidad explicativa de una determinada variable, podríamos decir que: *Una variable X contiene información sobre Y , adicional a la proporcionada por las variables Z si los residuos de Y sobre Z son diferentes de los residuos de Y sobre $\{X, Z\}$.* Por supuesto que esto requiere decidir si dos conjuntos de residuos son diferentes entre sí, cuestión nada sencilla. Pero esta situación es preferible a la alternativa habitual, por distintas razones. Por ejemplo, podríamos detectar que los dos conjuntos de residuos difieren tan sólo en un número reducido de observaciones muestrales. Pero quizá esas observaciones comparten una característica que nos ayuda a detectar algún error de especificación del modelo. Lo que es más importante, si esto sucede en un subconjunto muestral pequeño, seguramente la diferencia no aflorará en los habituales estadísticos de

³ ¿Por qué habría que hacer este tipo de contrastes de significación?

ajuste global, como el logaritmo de la función de verosimilitud, que es la base de los contrastes habituales. La comparación de dos conjuntos de residuos es una comparación entre conjuntos ordenados y, lamentablemente, no se dispone de muchos contrastes para este tipo de situaciones, que deberían constituir una interesante línea de investigación. Pero esto no debe utilizarse para negar este camino: el investigador debe aportar su capacidad para comparar los errores de ajuste de dos modelos, imponiendo la restricción y sin imponerla, y discutir el grado en que ambos conjuntos de residuos difieren. Después de todo, si el análisis de relaciones económicas se consistiese en presionar unas pocas teclas y comparar algunos de los valores numéricos resultantes con tablas de valores críticos ¿para qué sería preciso un analista de datos?

3.4. Comparando capacidad explicativa

Estas observaciones tienen implicaciones importantes respecto de una práctica habitual: la comparación de la capacidad explicativa o contenido informativo de dos variables.⁴ Por supuesto que no pueden utilizarse los valores numéricos de los coeficientes estimados a este fin. Pero, contra una práctica muy extendida, tampoco los estadísticos t de los coeficientes que acompañan a dichas variables indican nada a este respecto. Esto es consecuencia de los comentarios que antes hice acerca de cómo un estadístico t consolida, de modo inseparable, dos cuestiones: a) el grado de incumplimiento de una restricción, y b) la precisión con que se estimó dicha restricción. Por la misma razón, no puede utilizarse el estadístico t obtenido para el coeficiente de una misma variable en dos submuestras, para discutir el aumento o disminución de capacidad explicativa de dicha variable entre dichas submuestras.

Las críticas vertidas al uso de los estadísticos t para evaluar el contenido informativo de las variables explicativas hace que tampoco tenga mucho sentido incluir los habituales asteriscos (de uno a tres) en función del p -valor del coeficiente estimado. Ni el p -valor es un indicador apropiado de capacidad explicativa ni, por las razones expuestas, puede resumirse los p -valores, ni los estadísticos t de los distintos coeficientes en un número de asteriscos que sugiera la mayor o menor relevancia de unas variables frente a otras.

Ni tampoco tienen sentido frases del tipo: “*La variable X no es significativa, aunque su coeficiente tiene el signo correcto*” ó su dual: “*El coeficiente de la variable X tiene el signo contrario al sugerido por la teoría aunque no es significativamente distinto de cero*”⁵

3.5. Resultados interesantes

Existe todo un conjunto de resultados, habituales en los cursos de Econometría, y que creo que quedan excesivamente diluidos entre el contenido global de la asignatura. Una revisión rápida sugiere considerar:

- la verdadera expresión de la matriz de covarianzas del estimador MCO, que debería enseñarse como la que se obtiene cuando el término de error del modelo no tiene matriz de covarianzas diagonal. Por otro lado, más que plantear la autocorrelación, la heterocedasticidad y la colinealidad como problemas que el investigador puede encontrar en ocasiones, la naturaleza de los datos económicos sugiere que estas son

⁴ Si se quiere utilizar el coeficiente estimado para analizar la relevancia de una variable explicativa, hay que tener en cuenta la variabilidad muestral de la variable explicativa.

⁵ Ejercicio: ¿Cuántos errores de los hasta ahora mencionados hemos incorporado en estas frases?

las situaciones habituales, salvo que en algún caso podamos encontrar que la presencia de una de estas situaciones es irrelevante. Pero en todo caso, con la matriz de covarianzas indicada nunca cometemos error, mientras que con la matriz más sencilla, cometemos errores si existe autocorrelación y/o heterocedasticidad, sesgando el desarrollo de los contrastes de hipótesis a que tan aficionados somos.

- El resultado de la regresión particionada es muy útil conceptualmente, ayudando a pensar el papel que juega la ausencia de correlación entre variables explicativas en la estimación de efectos individuales. Esta observación está relacionada con el comentario que antes hicimos acerca de la interpretación de coeficientes individuales.
- La expresión para la varianza de un coeficiente estimado en una regresión múltiple ilustra perfectamente la relación entre colinealidad y pérdida de precisión, con todas las implicaciones que ello tiene de cara a la contrastación de hipótesis y la interpretación de coeficientes individuales, pro las razones antes expuestas
- El análisis de los coeficientes de correlación y determinación parcial en una regresión múltiple permite entrar en el análisis comparativo del contenido informativo relativo de dos o más variables explicativas
- El punto anterior está íntimamente relacionado con el tratamiento de la colinealidad, cuestión nada sencilla de realizar. Conviene distinguir claramente entre efectos parciales y efectos totales, o entre efectos directos e indirectos. Solo en ausencia de correlación entre variables coinciden unos con otros. La regresión simple estima sin sesgo el efecto total, mientras que la regresión múltiple estima sin sesgo el efecto parcial, siempre que el modelo no omita variables que aportan información sobre las ya existentes. Esta observación es importante hasta el punto de cuestionar la habitual presentación del llamado *sesgo de variables omitidas*. No existe tal sesgo, sino una inapropiada descripción del tipo de efecto que se quiere medir.
- El hecho de que la naturaleza no experimental de la Economía nos obligue en la mayoría de los casos a trabajar con una única muestra, debería llevarnos a enfatizar las propiedades de consistencia (propiedades de una única muestra), frente a la de ausencia de sesgo.⁶ No es muy difícil probar la inconsistencia de mínimos cuadrados utilizando sumatorios bajo las hipótesis adecuadas. Por esto también, creo que es preferible utilizar supuestos del tipo $E(u/X) = 0$, $E(u^2/X) = \sigma^2$ que del tipo: $E(u) = 0$, $E(u^2) = \sigma^2$. De hecho, recomendaría claramente la utilización de sumatorios y la demostración de propiedades para el modelo de regresión de una sola variable. Entender estas demostraciones es suficiente para entender las situaciones en modelos más complejos, sin necesidad de entrar en demostraciones formales con notación matricial.
- Obtener las expresiones analíticas del sesgo de inconsistencia en presencia de: a) simultaneidad, b) errores en variables es un ejercicio formativo en muchos aspectos, c) modelos dinámicos con autocorrelación, que recomiendo para alumnos que ya tienen una cierta formación en Econometría

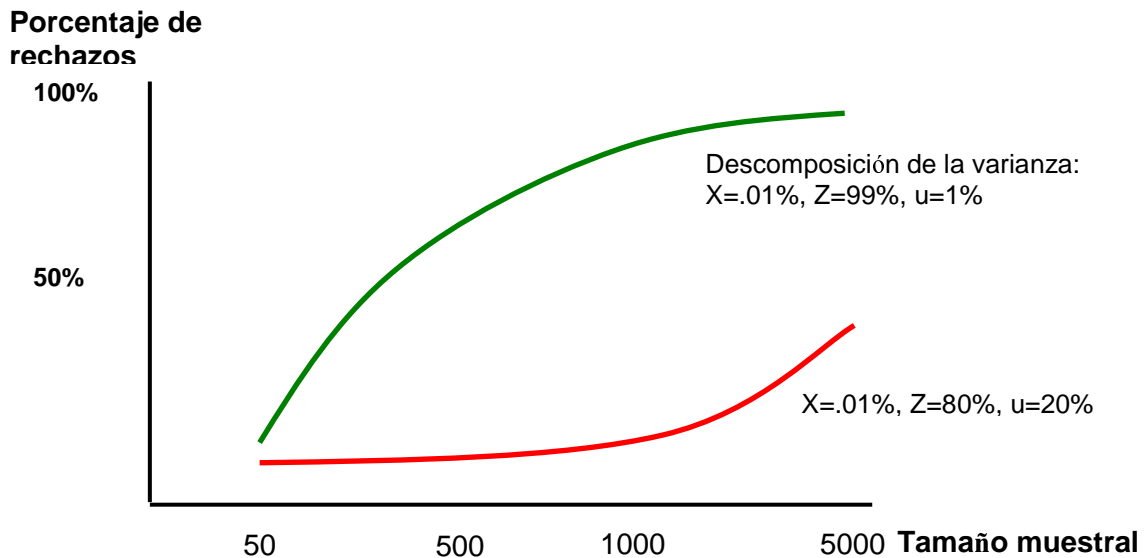
⁶ ¿De qué sirve la ausencia de sesgo, por sí sola, cuando se dispone de una única muestra?

Apéndice.- La relación entre precisión y potencia

Para motivar la discusión sobre los habituales excesos en la contrastación de hipótesis, suelo sugerir considerar el modelo: $y_i = \alpha + \beta x_i + u_i$, en el que se quiere contrastar la hipótesis nula: $H_0: \beta=1$ versus: $H_1: \beta \neq 1$, habiendo obtenido una estimación numérica $\hat{\beta} = 0,50(0,30)$, donde la cifra en paréntesis es la desviación típica. Con esta estimación, los criterios habituales conducen a no rechazar la hipótesis nula, con un p-valor de 0,106. ¿Es este proceder razonable? Imaginemos que se trata del Banco Central Europeo tratando de analizar cuánto debe elevar sus tipos de intervención para frenar en un 1% la inflación de la zona euro. A la vista de lo expuesto, ¿debería elevarlos en un 1% o en un 2%? Desde luego, no es lo mismo, en términos de su posible impacto sobre la evolución de la economía. Evidentemente, estamos en una situación de reducida precisión y la única afirmación sensata sería: “*La información muestral no permite pronunciarse sobre esta cuestión con la precisión que sería necesaria*”. Con la estimación numérica mencionada, el investigador debería cuestionarse seriamente que el coeficiente sea unitario. Quien aún no esté convencido, debe notar que, en este contexto, tampoco rechazaría la hipótesis nula $H_0: \beta=0$ versus: $H_1: \beta \neq 0$. ¿Qué aspecto tendría la función de potencia de este contraste?

El segundo ejemplo tiene que ver con el hecho de que en Economía, la mayoría de los contrastes debieran ser unilaterales, contra la práctica habitual, ya que contamos con suficiente información para establecer una hipótesis alternativa de ese tipo. En ese caso, hemos de recordar que para rechazar una hipótesis nula en el enfoque Neuman-Pearson, deben darse siempre dos condiciones: 1) que la evidencia muestral sea contraria a la hipótesis nula, y 2) que sea favorable a la hipótesis alternativa. Nadie ignora la primera, pero ignoramos frecuentemente la segunda. En un contraste bilateral no hay problema, porque ambas condiciones son equivalentes, pero en un contraste unilateral la situación es muy distinta. Consideremos el contraste: $H_0: \beta=1$ versus: $H_1: \beta < 1$ en el mismo modelo anterior, con una estimación numérica: 1,05 (0,05). La aplicación de la metodología habitual de contraste conducirá a no rechazar H_0 . Esto se debe a que si bien se cumple la primera condición, no se cumple la segunda. Sorprendentemente, este tipo de situaciones surgen incluso en trabajos de investigación y, lo que es peor, se interpreta en el sentido de rechazar la hipótesis nula *en favor* de la hipótesis alternativa. En este caso, la simple observación de la estimación numérica sugiere que no debe procederse al contraste. No rechazaríamos la hipótesis nula ni siquiera con una estimación de $\hat{\beta}=1,65$, con independencia de la desviación típica estimada, lo que muestra lo anómalo de la situación.

Un segundo tipo de problemas surgen cuando un elevado tamaño muestral conduce a una alta precisión, tendiendo a observarse entonces valores numéricos del estadístico t muy reducidos. El gráfico siguiente presenta resultados obtenidos en un ejercicio de simulación fácilmente reproducible con alumnos. Consideremos el modelo $y = \beta_1 x + \beta_2 z + u$, $Corr(x, z) = Corr(x, u) = Corr(z, u) = 0$ que presenta tres factores determinantes de la variable dependiente: las dos variables explicativas, y el término de error. Los tres factores se toman incorrelacionados entre sí, por construcción, por un lado, para que la colinealidad no sea un condicionante de los resultados; por otro, para no vulnerar la consistencia del estimador de mínimos cuadrados. El ejercicio consiste en extraer un número elevado de muestras con unas mismas características, estimar el modelo con cada una de ellas, y llevar a cabo, con las estimaciones resultantes, el contraste de la hipótesis nula: $H_0: \beta_1 = 0$. Queremos ver cómo el



porcentaje de rechazos de la hipótesis cambia con el tamaño muestral T .

Al ser los tres factores incorrelacionados, podemos descomponer la varianza de Y como suma de las contribuciones de las varianzas de X , Z y u . La contribución

porcentual de X a la varianza de Y es: $\frac{\beta_1 \sigma_x^2}{\beta_1 \sigma_x^2 + \beta_2 \sigma_y^2 + \sigma_u^2}$, e igualmente se calculan los

otros dos términos. Así se puede controlar en el diseño experimental una mayor o menor relevancia de cada una de las dos variables explicativas. En nuestro caso, los dos coeficientes fueron elegidos iguales a 1, y se modificaron las varianzas de X , Z y u para obtener los dos escenarios deseados.

En el primer escenario (línea superior), la variable dependiente es esencialmente igual a Z , excepto en un 1%, que es causado por el término de error. La contribución de X es despreciable. En el segundo escenario, la relevancia de Z es inferior, ganando importancia el término de error, no observable. La variable X continúa siendo irrelevante. En consecuencia, en ambos casos querríamos no rechazar la hipótesis nula. Los porcentajes de rechazo recogidos por las dos curvas muestran cómo el porcentaje de rechazos, que es efectivamente reducido para tamaños muestrales pequeños, crece con el tamaño de la muestra. Esto resulta a primera vista sorprendente: pensaríamos que con mayor tamaño muestral tenemos mayor precisión y, con ella, mayor capacidad de discriminar hipótesis ciertas de hipótesis falsas. No es así. Predomina la reducción que en la desviación típica se produce al aumentar el tamaño muestral. Disminuye con ello el denominador del estadístico t , por lo que se llega a rechazar la hipótesis nula incluso aunque cabe pensar que la estimación del coeficiente no será, en términos absolutos, muy diferente de cero. En el segundo escenario, la mayor importancia del término de error reduce la precisión de la estimación y, con ello, la disminución en desviación típica estimada, y el estadístico t no aumenta tanto. Aunque el porcentaje de rechazos se eleva con el tamaño muestral, los porcentajes son inferiores a los del primer escenario. En todo caso, una imagen preocupante, de gran impacto pedagógico.