

PROBLEMAS DE ESTIMACIÓN Y CONTRASTE EN LOS MODELOS DE DIFERENCIAS EN DIFERENCIAS

Mayo 2006

nº 13

José Vicéns Otero

Los modelos de diferencias en diferencias (DD) son posiblemente la especificación econométrica más utilizada para analizar el efecto o impacto de un cambio sobre un sistema y dados sus elevados niveles de aplicación con datos individuales puede considerarse un método perteneciente al contexto de la microeconometría. En términos estadísticos puede expresarse como la medición del efecto o impacto de una variable de tipo discreto sobre una variable dependiente continua, siendo la variable discreta el cambio y la variable continua la variable resultado. Sus aplicaciones son múltiples en el contexto de experimentos naturales o cuasiexperimentos. Este tipo de experimentos se diferencian de los experimentos habituales en que el cambio es exógeno y el diseño o elección de los grupos de control y tratamiento son consecuencia del cambio, a diferencia de la experimentación habitual donde los grupos de control y tratamiento se diseñan y escogen aleatoriamente. Sus aplicaciones en el análisis económico han sido múltiples, pero desgraciadamente la mayoría de los análisis se han fijado en los resultados y sus consecuencias, pasando por alto, salvo honrosísimas excepciones, ciertos problemas que afectan directamente a los resultados econométricos de la estimación. En el presente trabajo se desarrollan y plantean soluciones para los tres problemas econométricos más importantes en este tipo de procesos: Endogeneidad, Correlación Intragrupo y Autocorrelación.

Edita:

Instituto L.R.Klein – Centro Gauss
Facultad de CC.EE. y EE.
Universidad Autónoma de Madrid
28049 Madrid
Teléfono y Fax: 914974191
Correo Electrónico: klein.gauss@uam.es
Página Web: www.uam.es/klein/gauss

ISSN 1696-5035

Depósito Legal: M-30165-2003

© Todos los derechos reservados. Queda prohibida la reproducción total o parcial de esta publicación sin la previa autorización escrita del editor.

I.- INTRODUCCIÓN

Los modelos de diferencias en diferencias (DD) son posiblemente la especificación econométrica más utilizada para analizar el efecto o impacto de un cambio sobre un sistema. En muchas de las aplicaciones con este tipo de modelos el cambio suele ser la aplicación de una nueva ley o la modificación de una situación previa, analizándose los efectos de este cambio sobre una variable de interés. En términos estadísticos puede expresarse como la medición del efecto o impacto de una variable de tipo discreto sobre una variable dependiente continua, siendo la variable discreta el cambio y la variable continua la variable resultado. La variable discreta suele tener dos estados con los que se recoge el cambio estudiado, y se la denomina variable de tratamiento.

Otra forma alternativa de describir la aplicación de un modelo de diferencias en diferencias es la de modelizar una experimentación en la que existe un grupo de observaciones que son afectadas por la variable tratamiento y otro grupo de observaciones, denominadas grupo de control, que no están afectadas por ella. Con el estimador de diferencias en diferencias se compara la diferencia entre el antes y después del cambio entre el grupo tratamiento y el grupo control para determinar el impacto neto. Sus similitudes con los análisis de la varianza y covarianza del análisis experimental son evidentes, si bien las propiedades del estimador de diferencias en diferencias ha recibido una gran atención y desarrollo en el análisis econométrico.

En la literatura especificada podemos encontrar referencias tanto como modelo, como estimador de diferencias en diferencias y son múltiples las aplicaciones en el contexto de experimentos naturales o cuasiexperimentos. Este tipo de experimentos se diferencian de los experimentos habituales en que el cambio es exógeno y el diseño o elección de los grupos de control y tratamiento son consecuencia del cambio, a diferencia de la experimentación habitual donde los grupos de control y tratamiento se diseñan y escogen aleatoriamente. Evidentemente en el contexto del análisis económico lo habitual será encontrarnos con experimentos naturales o situaciones en las que se ha adoptado una política económica o llevado a cabo un cambio en el sistema y el analista dispone de los resultados a posteriori, sin poder diseñar la fuente de variación y los grupos afectados.

Entre los múltiples investigadores que han utilizado esta aproximación metodológica, cabe destacar las de Card (1990) sobre impacto de la inmigración en los salarios de los nativos, Card y Krueger (1995) sobre salario mínimo y empleo, Gruber y Madrina (1994) y Gruber (2000) sobre seguros y oferta

laboral, Buchmueller y Valletta (1996) (1999) sobre seguros, movilidad y oferta laboral, Dean (2001) sobre impuestos, Grogger y Willis (2000) sobre consumo de drogas y criminalidad, Lawrence y Kessler (2001) sobre legislación y salarios en el sector de la construcción, Neumark y Wascher (2001) sobre salario mínimo y formación y la de Ashenfelter y Greenstone (2001) sobre límites de velocidad. Un análisis de los muchos trabajos y aplicaciones realizados con este tipo de modelos de los últimos años, llevan a considerarlos como el instrumento más popular para evaluar los efectos de leyes o intervenciones.

II.- PLANTEAMIENTO DEL MODELO

Para entender el origen del modelo y su denominación como modelo de diferencias en diferencias vamos a partir del caso más simple en el que solo existe un grupo de individuos al que se le somete a un tratamiento. El grupo tendrá un conjunto de observaciones o resultados antes del tratamiento, una por individuo, y otro conjunto después del tratamiento, siendo la formulación del modelo

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \varepsilon_{it}$$

donde Y_{it} es la variable resultado analizada para el individuo i en el momento t , con $i = 1, 2, \dots, nt$; $t = 0, 1$ indicando 0 el momento de tiempo antes del tratamiento y 1 el momento de tiempo después del tratamiento. La variable X_{it} es una variable ficticia con valor 1 si el individuo i ha recibido el tratamiento y 0 en otra situación, es decir valdrá 1 para $t=1$ y 0 para $t=0$. ε_{it} es un término de error con las hipótesis habituales de la perturbación aleatoria y β_1 es el parámetro de interés que nos recogerá el efecto o impacto del tratamiento sobre la variable resultado Y_{it} . Un estimador insesgado de β es

$$\hat{\beta}_1 = \bar{y}_{.1} - \bar{y}_{.0}$$

recogiendo la diferencia de la medias de las observaciones que han experimentado el tratamiento y las que no lo han experimentado. Este modelo, que en la literatura se denomina modelo o diseño en diferencias, es el mismo que el denominado modelo ANOVA en el contexto del diseño experimental cuando se considera un único factor y este tiene solamente dos niveles. El modelo podemos estimarlo por mínimos cuadrados ordinario (MCO) para todas

las observaciones y poder así contrastar la hipótesis de influencia del tratamiento, es decir $H_0: \beta_1 = 0$, con los contrastes t habituales.

El modelo en diferencias es muy limitado ya que supone que los dos grupos de observaciones (y), correspondientes a los dos momentos del tiempo considerados, mostrarían medias iguales si no se hubiera experimentado el tratamiento. De esta forma al haber existido tratamiento, los cambios registrados en la variable (y) se asignan exclusivamente a este. Esta hipótesis es difícil de mantener ya que en experimentos no controlados o difícilmente controlables, como ocurre en el contexto de las ciencias sociales, actuarán otras variables además del tratamiento o efecto que queremos medir sobre la variable resultado, de forma tal que no podemos aislar lo que pertenece a la variable tratamiento y lo que pertenece a otras variables. El simple caso del transcurso del tiempo, entre los dos momentos analizados, seguramente afectará a la variable que se analiza, especialmente si se trata de una variable económica.

Para solucionar el problema anterior se acude a un procedimiento simple como es tener, además del grupo objetivo al que se le aplica el tratamiento, un grupo de control que no recibe el tratamiento. Este grupo de control nos permitirá medir el impacto de otras variables que afectan al grupo objetivo pero que son distintas del tratamiento estudiado. A este tipo de diseño experimental o modelo es al que se denomina de diferencias en diferencias, tomando la forma general siguiente

$$Y_{ijt} = \beta_0 + \beta_1 X_{it} + \beta_2 X_{ij} + \beta_3 X_{ijt} + \varepsilon_{ijt} \quad [2]$$

donde Y_{ijt} recoge la observación del individuo i del grupo j en el momento t , indicando el nuevo subíndice j el grupo, con $j=0$ para el grupo control y $j=1$ para el grupo tratamiento. La variable X_{it} es una variable ficticia con valor 1 si la observación es del momento posterior al tratamiento y 0 en cualquier otro caso, X_{ij} es una variable ficticia con valor 1 si la observación pertenece al grupo tratamiento y 0 en cualquier otro caso y X_{ijt} es otra variable ficticia con valor 1 cuando la observación pertenezca al grupo tratamiento y en el momento posterior al mismo. El parámetro de interés es el que recoge la diferencia entre el antes y el después del tratamiento, así como la diferencia del grupo tratado con el grupo de control, es decir el parámetro β_3 , y este parámetro puede estimarse mediante:

$$\beta_3 = \bar{y}_{.11} - \bar{y}_{.10} - (\bar{y}_{.01} - \bar{y}_{.00})$$

indicando las medias para todos los elementos i . Así, el estimador se obtiene calculando la diferencia entre antes y después del tratamiento para el grupo objetivo y eliminando de esta diferencia lo que se debe a otras causas y que se recoge mediante la diferencia entre antes y después para el grupo de control. Esta forma de calcular el parámetro y que está en la esencia del modelo es lo que hace que a este tipo de diseños se les denomine “diferencias en diferencias” (DD). Los parámetros pueden estimarse por MCO, al igual que el caso anterior, lo que nos permitirá realizar los habituales contrastes.

Es fácil observar como la variable ficticia X_{ijt} es igual al producto de las variables ficticias X_{ij} y X_{it} y que el parámetro β_1 recoge la diferencia entre el antes y el después del tratamiento y que el parámetro β_2 recoge las diferencias entre el grupo de tratamiento y el de control. Desde el punto de vista de un ANOVA nos encontramos con dos factores, tiempo y grupo, con dos niveles cada uno, y una interacción recogida por la variable X_{ijt} .

III.- ALGUNAS EXTENSIONES DEL MODELO DD

El modelo DD plantea en sus aplicaciones algunas extensiones con las que el investigador o analista puede resolver de forma fácil ciertos problemas y que se resumen en los siguientes puntos:

a) La existencia de otros factores externos. Puede ocurrir que existan factores distintos del tratamiento y que estos influyan en las diferencias del grupo objetivo antes y después y las diferencias con el grupo de control. A este hecho se le suele denominar interacciones omitidas y distintas a la interacción analizada y se presenta cuando un factor externo influye de forma diferente al grupo objetivo y al grupo de control. Si es esta la situación debe procederse a una nueva especificación del modelo que incluya más de un tratamiento. Supongamos la existencia de dos tratamientos (t y k) con lo que tendríamos dos grupos (j), dos momentos de tiempo (t), un nuevo impacto o tratamiento (k) y unos individuos por grupo (i); el modelo tendría la forma

$$Y_{ijkt} = \beta_0 + \beta_1 X_{it} + \beta_2 X_{ij} + \beta_3 X_{ijt} + \beta_4 X_{ik} + \beta_5 X_{ikt} + \beta_6 X_{ijk} + \beta_7 X_{ijkt} + \varepsilon_{ijkt}$$

siendo ahora el parámetro de interés β_7 que recoge el efecto del grupo (j), y los dos tratamientos (k , t). En el modelo se incluyen los parámetros β_3 , β_5 y β_6 para recoger las interacciones de los factores (j , t), (k , t), (j , k) y los parámetros β_1 , β_2 y β_4 para los factores aislados.

b) La existencia de diferencias significativas entre individuos. Cuando, como es habitual, los individuos o los elementos que constituyen el grupo de control y el grupo objetivo son distintos, no se debe mantener la especificación inicial ya que reaccionarán de forma diferente al tratamiento en función de sus propias diferencias. En este caso es aconsejable un modelo que recoja tales diferencias en la forma

$$Y_{ijt} = \beta_0 + \beta_1 X_{it} + \beta_2 X_{ij} + \beta_3 X_{ijt} + \beta_4 Z_{ijt} + \varepsilon_{ijt} \quad [3]$$

donde la variable Z_{ijt} recoge las características de los individuos. En el caso de que los grupos se mantengan constantes antes y después del tratamiento, panel de individuos, el subíndice t de la variable Z podría ser eliminado. El parámetro β_4 debe ser igual en todos los grupos para garantizar la estimación insesgada de β_3 .

c) La utilización de períodos de tiempo más amplios. En el caso de las series económicas es muy probable que se disponga de más períodos de tiempo para la estimación que los dos descritos en la formulación más simple de “antes” y “después”. En tal caso tendremos una más importante fuente de variabilidad temporal y mayores grados de libertad para la estimación del modelo, simplemente permitiendo que $t = 1, 2, \dots, T$. Inicialmente la estimación será por MCO y tendremos un mayor número de periodos para antes del tratamiento y para después del tratamiento.

d) La utilización de un mayor número de grupos de individuos. Al igual que para el caso de la variabilidad temporal puede aumentarse la variabilidad espacial utilizando más grupos que el grupo tratamiento y el grupo de control. De hecho en muchas de las aplicaciones al mundo laboral realizadas en Estados Unidos utilizan los 50 estados existentes, cada uno un grupo, con sus consiguientes muestras de individuos. La especificación inicial del modelo se mantiene, pero considerando $j = 1, 2, \dots, K$ grupos, algunos de los cuales serán objeto del tratamiento y otros no.

e) Considerar el modelo DD como un caso especial de modelos con datos de panel. El modelo DD es un caso especial de un modelo general con datos de panel de efectos fijos de la forma

$$Y_{it} = \gamma_t + \alpha_i + \beta_1 X_{it} + u_{it}$$

donde γ_t es un término que varía con el tiempo y recoge la dinámica de la variable endógena, α_i es una variable que captura los factores no observados o no incluidos en el modelo, que afectan a la variable endógena pero no varían con el tiempo y X_{it} es la variable explicativa que varía en el tiempo y con los individuos o unidades estudiadas. A la variable α_i se le denomina efecto fijo y al modelo planteado modelo de efectos fijos. El parámetro de interés es β_1 y su estimación por MCO plantea el problema de que dado que α_i es no observable, el modelo a estimar será

$$Y_{it} = \gamma_t + \beta_1 X_{it} + V_{it}$$

con

$$V_{it} = \alpha_i + u_{it}$$

no siendo posible aceptar la incorrelación entre V_{it} y X_{it} , debido precisamente a la inclusión de α_i en el término de error. La solución al problema es sencilla ya que si suponemos la existencia de al menos dos periodos de tiempo, al tomar primeras diferencias temporales sobre el modelo inicial, tendremos

$$\Delta Y_i = \Delta \gamma + \beta_1 \Delta X_i + \Delta u_i$$

modelo en el que si podemos aceptar la incorrelación entre ΔX_i e Δu_i . Al estimador de β_1 en este modelo se le denomina estimador en primeras diferencias y plantea como uno de sus principales inconvenientes el que en muchas de las aplicaciones ΔX_i no presenta mucha variación entre las unidades consideradas y los errores estándar son muy elevados. Una solución a este problema es tomar dos periodos de tiempo alejados, en lugar de dos consecutivos, para que los incrementos presenten mayor variación entre las unidades.

El modelo es plenamente coincidente con el de DD si la variable explicativa es el tratamiento, con lo que el parámetro β_1 viene determinado por la diferencia en el tiempo y la diferencia entre el grupo de experimentación y el grupo de control.

IV.- PROBLEMAS ECONOMÉTRICOS DEL MODELO DD

Como ya se ha expuesto, los modelos DD han sido ampliamente utilizados en diferentes ciencias sociales y de manera muy especial en economía. El hecho de disponer de un instrumento capaz de medir el impacto de un cambio sobre un sistema y que además el instrumento sea fácil de entender y de utilizar, ha generado una gran cantidad de aplicaciones por parte de multitud de economistas, muchos de ellos de reconocido prestigio. Pero desgraciadamente la mayoría de los análisis se han fijado en los resultados y sus consecuencias, pasando por alto, salvo honrosísimas excepciones, ciertos problemas que afectan directamente a los resultados econométricos de la estimación. De estos problemas cabe destacar tres por su importancia y generalidad: el problema de la endogeneidad del tratamiento, el problema de la correlación intragrupo y el problema de la autocorrelación en la perturbación aleatoria.

a) Endogeneidad del tratamiento. El problema de la endogeneidad en el tratamiento, cambio o ley, en los modelos DD ha recibido la atención de la econometría ya que tal situación no es infrecuente. Pongamos por ejemplo el caso en que analizamos el impacto que tiene sobre el salario de los individuos conocer un idioma. En el modelo DD la especificación podría ser del tipo

$$Y_{iLt} = A_L + B_t + CX_{iLt} + \beta T_L + \varepsilon_{iLt}$$

donde Y_{iLt} recoge el salario para el individuo i perteneciente al grupo de idioma L en el momento t , A_L sería el efecto fijo del lenguaje, B_t el efecto fijo del tiempo, X_{iLt} sería el vector de otras variables explicativas y β el parámetro de interés que recoge el impacto sobre el salario después de haber aprendido inglés(*).

Sin embargo, podría ocurrir que el hecho de que un individuo decida aprender inglés viene motivado por el salario que percibiría en el caso de no saber inglés, con lo que tendríamos una correlación entre el tratamiento T y la perturbación ε . Es decir, el tratamiento tendría una endogeneidad explicada precisamente por la variable dependiente.

(*) Nótese que esta notación es similar a la [3] pero más simple que la anterior, siendo más utilizada por la literatura especializada. Los efectos A y B serán cero para el grupo de control y el momento “antes” respectivamente, con lo que están representando solamente los parámetros de interés sin las variables ficticias. La variable T_L es la equivalente a la variable X_{ijt} de [3]

Si eliminamos la influencia del resto de variables y dejamos solamente el tratamiento, el estimador por MCO de β será

$$\hat{\beta} = (T'T)^{-1}T'Y \quad [5]$$

y sustituyendo el valor de Y

$$\hat{\beta} = \beta + (T'T)^{-1}T'\epsilon \quad [6]$$

Con lo que dada la correlación entre T y ϵ el estimador será sesgado y también inconsistente pues su límite en probabilidad no coincidirá con el parámetro poblacional. Como se recordará, entre las hipótesis básicas del modelo de regresión lineal se encuentra la hipótesis de que las variables explicativas son fijas y no aleatorias y en consecuencia se acepta su incorrelación con el término de perturbación aleatoria. Ello permite obtener estimadores insesgados por MCO, ya que bajo este supuesto en [6] $E(\hat{\beta}) = \beta$. Sin embargo la endogeneidad de T nos impide considerar al tratamiento como no aleatorio y además estará correlacionado con e .

Esta situación en la que existe correlación entre la variable dependiente y la perturbación es similar a otros dos problemas descritos en la econometría clásica, errores de medida en la variable explicativa y variables omitidas.

En el planteamiento de variables con error de medida y en el contexto del modelo lineal general (MLG) se da la situación en la que existe un modelo real

$$Y = \tilde{X}\beta + u \quad [7]$$

pero en lugar de observar \tilde{X} observamos o medimos X

$$X = \tilde{X} + V \quad [8]$$

siendo V un error de medida aleatorio desconocido. En lugar de [7] nuestro modelo será

$$Y = X\beta + (u - VB) \quad [7']$$

y dada la relación existente entre X y V , es posible que no podamos admitir la incorrelación entre estas variables, con lo que la variable explicativa

observada estará correlacionada con la perturbación aleatoria, y al igual que en [6] el estimador será sesgado

Fijémonos que aun cuando el error de medida V no esté correlacionado con la variable explicativa no observada, los problemas permanecen. Así, con

$$\text{cov}(\tilde{X}, V) = 0$$

tendremos que desde [8]

$$\text{Cov}(X, V) = E(X, V) = E(\tilde{X}, V) + E(V^2) = 0 + \sigma_v^2 \quad [9]$$

Y la covarianza entre la variable observada que aparece en el modelo estimado y el error de medida es igual a la varianza de este. Ahora la covarianza entre la variable explicativa observada y el término de error será

$$\text{Cov}(X, u - V\beta) = -\sigma_v^2\beta$$

y la estimación será inconsistente. En general la inconsistencia para un estimador $\hat{\beta}_i$, viene definida por

$$\text{plim} \hat{\beta}_i = \beta_i + \frac{\text{Cov}(X_i, \varepsilon_i)}{\text{Var}(X_i)} \quad [10]$$

y en nuestro caso

$$\text{plim} \hat{\beta}_i = \beta_i + \frac{\text{Cov}(X_i, u - \beta_i V)}{\text{Var}(X_i)}$$

y desde [8] y [9]

$$\text{plim} \hat{\beta}_i = \beta_i + \frac{(0 - \beta_i \sigma_v^2)}{\sigma_{xi}^2 + \sigma_v^2}$$

$$= \hat{\beta}_i \left(1 - \frac{\sigma_v^2}{\sigma_{xi}^2 + \sigma_v^2} \right)$$

$$= \hat{\beta}_i \left(\frac{\sigma_{xi}^2}{\sigma_{xi}^2 + \sigma_v^2} \right)$$

con lo que el término entre paréntesis siempre será menos que uno y la estimación por MCO siempre tenderá a subestimar el parámetro poblacional en grandes muestras y ante errores de medida. Es fácil observar que cuanto mayor sea la varianza de X , el sesgo será más pequeño.

La otra situación en la que nos podemos encontrar un problema de correlación entre una variable predeterminada y la perturbación aleatoria es cuando se omite alguna variable explicativa. Será el caso en el que en lugar de estimar el modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad [11]$$

estimamos el modelo

$$Y = \beta_0 + \beta_1 X_1 + (\varepsilon + \beta_2 X_2) \quad [12]$$

expresión similar a [7'] y en la que se omite la variable X2 que queda incluida en el término de error u.

$$Y = \beta_0 + \beta_1 X_1 + u \quad [13]$$

Si las variables X1 y X2 están correlacionadas, esta se transmitirá como correlación entre la variable explicativa y la perturbación y obtendremos estimadores sesgados e inconsistentes. El sesgo del parámetro estimado, desde [10], será

$$\text{plim} \hat{\beta}_1 - \beta_1 = \frac{\text{cov}(x_1, u)}{\text{Var}(x_1)}$$

y en el que como hemos dicho el numerador no es nulo.

La solución más habitual al problema general de correlación entre la variable explicativa y la perturbación es la utilización del método de variables instrumentales, procedimiento que se describe a continuación

En el contexto del modelo lineal general (MLG)

$$Y = X\beta + u$$

donde el término de perturbación aleatoria cumple las hipótesis habituales pero existe correlación entre X y u

$$\text{plim}(X'u/n) \neq 0$$

supondremos que podemos hallar P instrumentos o variables recogidas en una matriz Z, tal que cumplan las siguientes dos condiciones

a/ Existe correlación entre las Z y las X

$$\text{plim}\left(\frac{Z'X}{n}\right) = \Sigma_{zx}$$

b/ Existe incorrelación entre las Z y las u

$$\text{plim}\left(\frac{Z'u}{n}\right) = 0$$

Con lo que premultiplicando el MLG por las variables instrumentales

$$Z'Y = Z'X\beta + Z'u$$

tendremos

$$\text{COV}(Z'u) = \sigma^2(Z'Z) = \sigma^2V$$

y aplicando mínimos cuadrados generalizados al modelo transformado, estimaríamos los parámetros por el método de variables instrumentales. En el modelo transformado y estimado por MCG si se cumplirá que $\text{plim}(\hat{\beta}) = \beta$

Debe resaltarse que de las dos condiciones impuestas a las variables instrumentales, la condición (b) no se puede contrastar dada la imposibilidad de observar u y en general será una hipótesis del investigador en base a sus conocimientos. Por el contrario y para una variable explicativa cualquiera (x) si podemos contrastar la condición (a), por lo que es habitual hacer una regresión del tipo

$$X = \beta_0 + \beta_1 Z + \varepsilon$$

y hacer un contraste sobre el parámetro β_1 que nos asegure su relación con la variable X.

Evidentemente la selección de los instrumentos es la parte más compleja del proceso, pues debe tratarse de variables no aleatorias con cierta relación con las variables explicativas. Como se ha mencionado, esta selección será una decisión del investigador, pero normalmente dispondrá de un conjunto de variables exógenas, distintas a las ya incluidas en el modelo.

Si suponemos el caso de dos variables explicativas, una no correlacionada con la perturbación aleatoria (Z_1) y otra si (Y_2), podemos formular el modelo por

$$Y_1 = \beta_0 + \beta_1 Z_1 + \beta_2 Y_2 + u_1 \quad [14]$$

donde las variables correlacionadas con u_1 pueden ser consideradas como endógenas (Y_1 e Y_2). La ecuación [14] sería una ecuación estructural y el problema es encontrar una variable instrumental de forma tal que esté correlacionada con Y_2 pero no lo esté con u_1 y a la que llamaremos Z_2 . En general para analizar la correlación entre Z_2 e Y_2 , se realiza la regresión

$$Y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + V_2 \quad [15]$$

donde se contrasta que π_2 sea distinto de cero. Como se verá se incluye en la regresión la variable exógena Z_1 dada su presencia en la ecuación estructural, ya que se requiere que el instrumento Z_2 tenga una correlación de tipo parcial con Y_2 , independientemente de Z_1 . En cierto sentido la ecuación [15] es una ecuación en forma reducida al estar la endógena en función de todas las exógenas.

El procedimiento puede parecer más complejo si se tiene más de una variable instrumental, sin embargo es igualmente simple pues no hay que elegir entre ellas. Si partimos del modelo [14] y suponemos la existencia de dos posibles variables instrumentales (Z_2 y Z_3) la estimación puede describirse como un proceso en dos etapas. En la primera etapa efectuamos una regresión por MCO como en [15] para estimar la variable \hat{Y}_2

$$\hat{Y}_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3$$

que será el instrumento obtenido por combinación lineal de las posibles dos alternativas. Esta combinación lineal será más eficiente que tomar cada instrumento por separado, si bien habrá que contrastar que π_2 y π_3 no sean cero, o no lo sea alguno de los dos, pues si fuera así el instrumento estaría perfectamente correlacionado con Z_1 . A continuación y en una segunda etapa realizamos la regresión

$$Y_1 = \beta_0 + \beta_1 Z_1 + \beta_2 \hat{Y}_2 + u_1$$

siendo la estimación por MCO de esta regresión el estimador por VI.

Es importante notar que en realidad es como si la variable Y_2 , tuviera dos partes, \hat{Y}_2 que no está correlacionada con u_1 y V_2 que si lo está. Así

$$Y_2 = \hat{Y}_2 + V_2$$

con lo que el modelo final es desde [14]

$$Y_1 = \beta_0 + \beta_1 Z_1 + \beta_2 \hat{Y}_2 + (\beta_2 V_2 + u_1) \quad [17]$$

Ello significa que si bien el término de error entre paréntesis está incorrelacionado con las explicativas y tiene media cero, no podemos aplicar las fórmulas habituales de MCO para calcular los errores estándar de los parámetros ya que la varianza de la perturbación no es σ_u^2 . Adicionalmente el estimador por VI puede plantear problemas en el cálculo del R^2 y en la significación conjunta del modelo. En primer lugar ocurre que el estimador por VI puede proporcionar un peor ajuste o menor R^2 que si aplicamos MCO directamente a la ecuación estructural, pero ello no debe preocuparnos ya que el objetivo del método no es mejorar el ajuste sino proporcionar una mejor estimación del parámetro o efecto cuando existe correlación con la perturbación.

Otro aspecto importante a tomar en consideración es que debido a que el estimador VI puede tener unos errores estándar muy grandes será un estimador menos eficiente que MCO cuando las variables explicativas sean exógenas, por lo que se deberá contrastar adecuadamente la posible endogeneidad de alguna de ellas antes de su utilización. Para analizar la endogeneidad existen varias alternativas, entre las cuales la más habitual sería la de realizar la estimación por MCO y VI y ver si ambas estimaciones difieren significativamente o no. Por su parte Wooldrige (2006) propone estimar la forma reducida [15], donde el término de error V_2 estará incorrelacionado con el error u_1 , de la forma estructural [14], solamente si Y_2 está incorrelacionada con u_1 lo que permitiría rechazar la endogeneidad. Para ello estimaría \hat{V}_2 por MCO y en la ecuación

$$Y_1 = \beta_0 + \beta_1 Z_1 + \beta_2 Y_2 + \delta_1 \hat{V}_2 + u_1$$

contrastaría si $\delta_1=0$ con el estadístico t, para en el caso de aceptar la hipótesis rechazar la endogeneidad de Y_2

En el contexto que estamos analizando de los modelos DD debe señalarse el importante trabajo de Besley y Case (2000) ya que se centran en el problema de la endogeneidad en los modelos con políticas y tratamiento. Para ello utilizan la especificación

$$Y_{st} = X_{st}a + b_t + c_s + \gamma P_{st} + [u_{st} + Z_{st} \delta] \quad [16]$$

donde

X_{st} = vector de características de los individuos promediadas para el estado s en el momento t

B_t = efecto del año

c_s = efecto del estado

P_{st} = efecto del tratamiento

Z_{st} = variables económicas (Q_{st}) y variables políticas (W_{st}) que influyen sobre P_{st}

U_{st} = error que recoge otras influencias no observables sobre Y_{st} y ortogonales con z_{st}

Nótese que z_{st} son variables observables que no están recogidas en la ecuación pero que al influir sobre P_{st} inciden sobre la variable endógena y por tanto nos encontramos con la situación en que las variables omitidas generan la endogeneidad del tratamiento. Por otro lado el interés del análisis se centrará fundamentalmente en la estimación del parámetro γ .

A efectos de solucionar el sesgo por la endogeneidad de la política o tratamiento Besley y Case (2000) eliminan o filtran la influencia de las variables X , año y estado sobre la variable endógena, el tratamiento y las variables z , con lo que las nuevas variables \tilde{Y} , \tilde{P} , \tilde{Z} serán ortogonales a X , efecto año y efecto estado. A continuación reescriben [16] por

$$\tilde{Y}_{st} = \gamma \tilde{P}_{st} + [u_{st} + \tilde{Z}_{st} \delta] \quad [17]$$

y dado que en nuestra hipótesis el tratamiento es endógeno y dependiente de las variables Z ,

$$\tilde{P} = \tilde{Z}\rho + V \quad [18]$$

La estimación por MCO de γ será

$$\hat{\gamma} = (\tilde{P}\tilde{P})^{-1} \tilde{P}'\tilde{Y}$$

y sustituyendo el valor de \tilde{Y}_{st} , en forma matricial sin considerar subíndices,

$$\hat{\gamma} = \gamma + (\tilde{P}\tilde{P})^{-1} \tilde{P}'u + (\tilde{P}\tilde{P})^{-1} \tilde{P}'\tilde{Z}\delta$$

Con lo que si sustituimos \tilde{P} por su expresión [18] y hallamos el límite de probabilidad eliminando los términos iguales a cero, quedará

$$\text{Plim } \gamma = \gamma + \text{plim} \left[\left(\frac{\tilde{P}\tilde{P}}{n} \right)^{-1} \left(\frac{v'u}{n} \right) \right] + \text{plim} \left[\left(\frac{\tilde{P}\tilde{P}}{n} \right)^{-1} \rho \left(\frac{z'z}{n} \right) \right] \quad [19]$$

La anterior expresión nos indica que existen por tanto dos fuentes de sesgo, la correspondiente al primer corchete y que se debe a variables no observables que influyen sobre la endógena y el tratamiento y la del segundo corchete que se debe a variables observables pero omitidas que influyen sobre el tratamiento. Evidentemente el primer caso es más difícil de solucionar, mientras que el segundo puede abordarse mediante una adecuada especificación del modelo. Sin embargo, y para este segundo caso, Besley y Case demuestran en su trabajo el fuerte sesgo que se origina al eliminar algunas de las variables que influye sobre el tratamiento.

Para solucionar la endogeneidad del tratamiento utilizan variables instrumentales, con la variante de que el procedimiento sería encontrar variables que se relacionan o tienen efecto sobre la política pero no sobre la variable dependiente o resultado, proponiendo los autores una nueva aproximación mediante el uso de variables políticas como instrumentales.

Entre las conclusiones básicas del trabajo hay que destacar la necesidad pero también la dificultad de que las variables económicas influyan de la misma manera a los elementos, grupos o estados, por lo que se convierte en trascendental la elección del grupo de control. Adicionalmente los autores recomiendan analizar previamente la influencia de variables económicas o políticas sobre la variable tratamiento, e incluso estimar una función de comportamiento del tipo [18], ya que ello ayudará a la elección del grupo de control y a la selección de los instrumentos finales.

b) Correlación intragrupo. Cuando en el modelo de regresión las perturbaciones son no esféricas se incumple la hipótesis de que $E(uu') = \sigma^2 I$ y en su lugar tenemos $E(uu') = \sigma^2 V$. Esta situación invalida los test de inferencia habitualmente usados cuando se estima el modelo por MCO, ya que si bien los estimadores mínimo cuadráticos son insesgados la matriz de varianzas de los parámetros no tiene la habitual fórmula

$$\text{Var}(\beta) = \sigma^2 (x'x)^{-1} \quad [20]$$

sino que la correcta expresión es

$$\text{Var}(\beta) = \sigma^2 (x'x)^{-1} x' V x (x'x)^{-1} \quad [21]$$

con lo que la aplicación de los contrastes t es incorrecta. Pero aún en el supuesto de que podamos estimar la varianza de $\hat{\beta}$ con la última expresión, si conocemos V, no podríamos utilizar su valor en el cálculo de la t ya que no se cumplen las hipótesis en base a las que se construyó el contraste. Adicionalmente, no se cumple tampoco el que los estimadores MCO tengan varianza mínima y el supuesto de independencia entre las x y la u debe ser confirmado.

La econometría conoce bien el problema y ha estudiado múltiples soluciones para la no esfericidad y la correlación de la perturbación aleatoria en el contexto de las series temporales, no así en el caso de los datos de corte transversal o ante la agregación de observaciones. Sin embargo ya en 1990 Brent R. Moulton publica un artículo en el que avisa y pone en guardia sobre los errores en que se incurre cuando en modelos que utilizan datos de microunidades, tanto en la variable dependiente como en las variables explicativas, se incluyen datos agregados entre estas últimas. Es por lo menos curioso notar como, quince años después del trabajo de Moulton, el mismo error se sigue cometiendo y son pocos los analistas que tan siquiera lo mencionan. Incluso podría decirse que este error es ahora más habitual que antes pues es ahora cuando se dispone de mayor cantidad de datos procedentes de encuestas y son mayores sus intentos de modelización.

De forma general el problema se plantea en modelos de corte transversal o de datos de panel, cuando sobre las unidades que se están estudiando, (personas, trabajadores, familias, etc.), se incluyen, además de variables

específicas de las anteriores unidades, variables agregadas sobre las características de sectores, provincias, países etc. Un ejemplo de ello lo tenemos en los modelos sobre inmigración y mercado laboral, cuando al tratar de explicar la situación de empleo o los salarios de los individuos, con datos procedentes de las encuestas se incluyen, además de las características de los individuos, variables con agregación geográfica que tratan de recoger las características del mercado laboral o de la economía de la zona. Y es que si bien desde un punto de vista teórico es acertada la inclusión de variables de esta naturaleza, no lo es la estimación resultante del modelo por mínimos cuadrados ordinarios, ya que la perturbación aleatoria incumplirá la hipótesis básica de incorrelación.

A nivel intuitivo es fácil admitir que las unidades que pertenezcan a una unidad de agregación o grupo, por ejemplo provincia, estén relacionadas con la variable incluida en el modelo, supongamos la tasa de paro, pero también con otras variables omitidas o no observables que harán que la perturbación esté correlacionada dentro de cada grupo. Moulton señala que si los investigadores omiten este hecho será porque piensan que el modelo está bien especificado y la correlación será pequeña, pero demuestra que aún con pequeñas correlaciones la desviación típica de los parámetros está fuertemente sesgado a la baja y que ello conduce a conclusiones erróneas sobre la significatividad de los parámetros.

Resulta evidente que en el contexto de los modelos DD el problema de la incorrelación intragrupo debe ser estudiado, ya que en su especificación inicial así lo plantea. En la formulación básica del modelo DD se incluirá siempre y por lo menos la variable tratamiento, que no es más una variable agregada por grupo, y no será difícil incurrir en la omisión de otras variables a nivel de grupo que estando relacionada con la variable endógena generen correlación intragrupo en la perturbación.

En el modelo con perturbaciones correlacionadas,

$$y = X\beta + u$$

$$E(u) = 0$$

$$E(uu') = \sigma^2 v = \sigma^2 [(1 - \rho)I_n + \rho ZZ']$$

siendo

y = vector $n \times 1$ correspondiente a la variable endógena

X = matriz $n \times k$ de variables explicativas

Z = matriz $n \times p$ de 0-1, indicando la pertenencia del individuo al grupo

u = vector $n \times 1$ de perturbaciones aleatorias

ρ = correlación intragrupo de la perturbación

admitimos que la matriz de varianzas covarianzas de la perturbación aleatoria no es diagonal, que tiene valores no nulos (ρ) en aquellos elementos que recogen la covarianza de las perturbaciones de un mismo grupo, pero que la varianza si es constante y con valor s_2 .

La matriz de varianzas covarianzas de las estimaciones mínimo cuadráticos viene dada por

$$\begin{aligned} E(uu') &= \sigma^2 (X'X)^{-1} X' V X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} [I + \rho(N-I)] \end{aligned}$$

siendo

$$N = X' Z Z' X (X'X)^{-1}$$

Con lo que si utilizamos la habitual forma de calcular la varianza, $\sigma^2 (X'X)^{-1}$ el sesgo, que cometemos se aproximará a $[I + \rho(N-I)]$.

Este sesgo depende del tamaño de los grupos, resultado del producto zz' y su multiplicación por x , la correlación intragrupo ρ , pero también de la correlación intragrupo de las variables explicativas. Moulton (1990) propone para el cálculo de las covarianzas la fórmula propuesta por Klock (1981) en el caso en que todas las explicativas sean agregadas por grupos

$$C = \sigma^2 (X^1 X^1)'^{-1} [1 + (m-1)\rho]$$

siendo m el tamaño de los grupos. Si bien la expresión no es exacta proporciona una simple y aceptable aproximación. Mediante un modelo de determinación de salarios, con una baja correlación de la perturbación intragrupo pero con un

elevado tamaño muestral, demuestra el fuerte sesgo que se produce en la varianza de los parámetros y en consecuencia la escasa fiabilidad de los contrastes t de los parámetros. Mas recientemente Pepper (2001) plantea que si la población está dividida en cluster o grupos y estos son suficientemente elevados en número, la estimación puede realizarse tomando como unidad el cluster en lugar de los individuos y obtener así los parámetros por mínimos cuadrados ordinarios, estimando su matriz de varianzas covarianzas mediante

$$V_c = (X'X)^{-1} \sum_{c=1}^C \{ (X'_c \hat{u}_t)(u'_t X_c) \} (X'X)^{-1}$$

siendo C el número de grupos y \hat{u}_c el vector de residuos mínimo cuadráticos del cluster o grupo c. Con un ejemplo Pepper demuestra como el tamaño del grupo o cluster tiene fuertes efectos sobre la correlación de la perturbación dentro del grupo y como esta afecta a la desviación típica de los parámetros.

c) Autocorrelación en la perturbación aleatoria El trabajo más importante en esta línea es el desarrollado por Bertrand, Duflo y Sendhil (2004), donde parten de la siguiente formulación del modelo DD

$$Y_{ist} = A_s + B_t + cX_{ist} + \beta t_{st} + \varepsilon_{ist} \quad [19]$$

en el que A y B son efectos fijos para estados y tiempo, X_{ist} es la variable de características de los individuos, T es una variable ficticia con valor 1 cuando el tratamiento afecto al grupo s en el momento t y el parámetro de impacto del tratamiento es β , normalmente estimado por MCO. El problema se centra en que el modelo planteado tenga correlación serial, una correlación que puede producirse debido a tres causas. Primero, porque los modelos DD plantean muchas veces largas series de tiempo, tal y como demuestran Bertrand y otros (2004) en su revisión de trabajos previos y en el que estiman un promedio de 16,5 períodos para todos los trabajos analizados. Segundo, porque las variables dependientes, usadas normalmente en los modelos DD, presentan correlación serial positiva. Tercero, y más importante, porque la variable tratamiento T_{st} cambia muy poco para un mismo grupo a lo largo del tiempo.

El problema surge, como es bien conocido en econometría, cuando existiendo autocorrelación se estima por mínimos cuadrados ordinarios, ya que si bien los estimadores serán insesgados y consistentes(*), también serán ineficientes y en consecuencia los procedimientos de inferencia habituales no

(*) La consistencia se mantendrán si entre las variables explicativas no figura la variable endógena desplazada.

serán correctos. Así y al igual que se exponía en el apartado anterior, cuando $E(uu') = \sigma^2 v$, la verdadera matriz de varianzas covarianzas de los estimadores no será [20] sino la expresión recogida en [21].

En general, las fórmulas habituales, si la autocorrelación es de primer orden y positiva, generará una infravaloración del error estándar, mientras que si es negativa tenderá a sobrevalorarlo. El sesgo dependerá de la correlación serial de las variables explicativas y como señalan Bertrand, Duflo y Sendhil, en los modelos DD la variable tratamiento está muy correlacionada ya que varía muy poco para un estado, 0 antes del tratamiento y 1 después. Estos autores analizan 92 trabajos de publicaciones relevantes, determinando la longitud del periodo analizado, la naturaleza de la variable dependiente y el método usado para estimar el error estándar de los parámetros, encontrando que 69 de ellos tenían más de dos periodos, que solamente 4 agrupaban o colapsaban los datos en antes y después del tratamiento y que de los 65 restantes solamente 5 utilizaban algún procedimiento para solucionar la posible autocorrelación.

Características de artículos con modelos DD (1990-2000)

Artículos modelos DD	92
Artículos con más de 2 períodos.....	69
Modelos que colapsan (antes-después).....	4
Modelos con potenciales problemas autocorrelación.....	65
Modelos con solución a la correlación.....	5
MCG	4
Estimación Ω	1
Promedio de periodos de tiempo.....	16,5

Fuente: Bertrand, Duflo, Sendhil (2004)

En el mismo trabajo plantean un modelo para el logaritmo de los salarios de las mujeres entre 25 y 50 años, con variables tales como tipo de empleo, educación, edad y estado de residencia para el periodo 1979-1999, lo que les permite definir 1050 células (50 estados por 21 años) y en cada una de ellas 300 mujeres. Los autores encontraron que los residuos obtenidos por la regresión del logaritmo de los salarios sobre ficticias de estado y tiempo, comunes para toda la muestra, estaban altamente autocorrelacionados. Adicionalmente y de forma aleatoria generaron leyes o impactos que afectaron a algunos estados en el periodo 1985-1995, de forma tal que disponían de

observaciones suficientes antes y después de la intervención. Consideraron a la mitad de los estados afectados por la ley, es decir 25 estados, aunque no hubiera sido así, mediante la variable ficticia T_{st} , con valor 1 para las mujeres que viven en el estado después de la intervención. Para esta situación simulada estimaron el modelo DD [19] por MCO, obteniendo el parámetro β y su error estándar. La simulación, con lo que denominan intervención “Placebo”, se repitió un gran número de veces, esperando que la hipótesis nula del parámetro se rechazara un 5% de las veces, ya que el parámetro realmente era nulo dada la inexistencia de tratamiento y la selección aleatoria de los estados. Adicionalmente generan otra simulación en la que a la variable endógena la sumaron $t_t * 0,02$, con lo que si generaron un efecto directo debido al tratamiento. Sin embargo en la simulación “Placebo” la hipótesis nula no se rechazó un 5% como era de esperar sino casi un 67,5% de las veces, demostrándose la tendencia a sobrerrechazar la hipótesis nula en el modelo DD. Cuando se utilizó la segunda simulación, en la que si existe efecto, la hipótesis nula se rechazó el 85,5% de las veces.

La primera explicación para que se produzca este error al rechazar la hipótesis nula reside en la crítica de Moulton (1990), ya que las estimaciones anteriores no consideran la probable correlación intragrupo, en este caso año y estado y la matriz de varianzas covarianzas de la perturbación aleatoria en lugar de ser diagonal puede ser bloque-diagonal. Para solucionarlo puede aceptarse que hay un efecto aleatorio y plantear el modelo como

$$Y_{ist} = A_s + \beta_t + cX_{it} + \beta T_{st} + U_{st} + \varepsilon_{it}$$

Estimando este modelo con la corrección de White (1984), rechazan la hipótesis nula en la simulación un 44% de las veces. Otra alternativa de solución es agregar los datos al nivel de estado-año, obteniendo un panel de estados sobre el tiempo. Para ello primero se efectúa una regresión de los salarios sobre las variables de control (educación y edad) y se obtienen los residuos. Las medias de estos residuos serán los datos a utilizar en el modelo

$$\tilde{Y}_{st} = \alpha_s + \gamma_t + \beta T_{st} + \varepsilon_{st}$$

En este caso se rechaza la hipótesis nula el 43,5% de las veces, y dado que la correlación intragrupo forzosamente se ha solucionado con la agregación, Bertrand, Duflo y Mullainathan concluyen que hay otra razón importante y que no es otra que la correlación serial del término T_{st} , es decir de la propia intervención. Para demostrarlo generan diferentes tipos de intervenciones, aleatorizando la fecha después de la cual los estados del grupo tratamiento son afectados. Seleccionan 10 fechas entre 1979 y 1989 y la intervención se define

cuando la observación es de un estado del grupo tratamiento en uno de los 10 años de intervención. Con esta simulación encuentran que la hipótesis nula solamente se rechaza un 6% de las veces.

Adicionalmente y con nuevas simulaciones en las que utilizan variables endógenas alternativas demuestran como el número de rechazos aumenta con los tamaños de los coeficientes de autocorrelación de estas variables y como manipulando los datos y la estructura del error, también el número de rechazos aumenta según aumenta el parámetro en el caso de un AR (1). En lo que se refiere al número de estados o unidades obtienen que no afecta al número de rechazos pero si observan que disminuye con el número de periodos de tiempo.

Ante el problema, que como se ha visto es bastante importante, de la correlación serial caben varias soluciones. En primer lugar estimar el verdadero error estándar de los parámetros y no mediante la fórmula habitual de MCO. Los citados autores estiman sobre los residuos minimocuadráticos la matriz Ω , pero se reduce solo hasta el 34,5% los rechazos de la hipótesis nula, consecuencia de las pocas observaciones temporales y de la dificultad de conocer la verdadera estructura del proceso.

Una segunda solución es la denominada “block bootstrap” proceso por el que se estiman en la ecuación [19] los residuos para cada estado, y con ellos unas nuevas variables endógenas y unos nuevos parámetros, obteniéndose así una distribución de los parámetros y elaborándose las pruebas de rechazo. Con este procedimiento en la simulación la mejora es escasa.

En tercer lugar se puede utilizar la información de los estados para estimar el proceso de autocorrelación, suponiendo que este proceso es el mismo para todos los estados. De esta forma tendremos tantas estimaciones del proceso como estados, lo que equivale al caso econométrico de un “sistema de ecuaciones aparentemente correlacionadas”. Utilizando este procedimiento el número de rechazos desciende al 7,75%, pero también es bajo en la simulación en la que si existe efecto, lo que la hace poco aconsejable.

En cuarto lugar, un método que obtiene buenos resultados cuando el número de unidades es suficientemente grande, se basa en una generalización de la fórmula de White (1984) para calcular la matriz de varianzas covarianzas

$$V = (X'X)^{-1} \left(\sum_{j=1}^{n_s} u_j' u_j \right) (X'X)^{-1}$$

siendo: X = matriz de variables explicativas

n_c = total de estados

$$u_j = \sum_{t=1}^T e_{jt} \chi_{jt}$$

E_{jt} = residuo en estado j en el momento t

χ_{jt} = vector fila de variables independientes incluyendo la constante

Con este método el número de rechazos alcanza el 6% y cuando existe efecto del tratamiento llega al 78%.

En quinto lugar, otra solución alternativa para evitar la correlación serial es eliminar la variación temporal cuando se calcula el error estándar. Para ello se promediaría la información antes y la información después con lo que habría solamente dos momentos de tiempo para la estimación, manteniendo toda la variación espacial. Con este procedimiento en el experimento, el número de rechazos cae hasta el 6%, pero en el caso de impacto real solo sube hasta el 29,5%.

La aproximación más adecuada para Bertrand, Duflos y Sendhil es la que denominan “inferencia aleatorizada” en la que calculan los modelos DD para leyes aleatorias placebo, para desde estas estimaciones construir la distribución empírica de los efectos de estas leyes “falsas” y posteriormente efectuar con ellas los contrastes de significación de las verdaderas leyes

Referencias

Ashenfelter, O. & Greenstone, M. (2004): "*Using Mandated Speed Limits to Measure the Value of a Statistical Life*". Journal of Political Economy, University of Chicago Press, vol. 112(S1), pages S226-S267, February.

Bertrand, M., Duflo, E., Mullainathan, S. (2004): "*How Much should We Trust Difference-in-Differences Estimators*". The Quarterly Journal of Economics, 2004, Vol 119, 1, pages 249-275.

Besley, T., Case, A. (2000): "*Unnatural experiments? Estimating the incidence of endogenous policies*". The Economic Journal, 110. Nuremberg 2000 páginas F672-F694.

Buchmueller, T. y Valletta, R. (1999): "*The Effect of Health Insurance on Married Female Labor Supply*". The Journal of Human Resources, Vol. 34, N°.1, pp. 42-70.

Card, D. (1990): "*The Impact of the Mariel Boatlift on the Miami Labor Market*". Industrial and Labor Relations Review 43 (January 1990).

Card, D. E. & Krueger, A.B.(1995): "*Myth and Measurement: The New Economics of the Minimum Wage*". Princeton: Princeton University Press.

Grogger, J y Willis, M. (2000): "*The Emergence of Crack Cocaine and the Rise in Urban Crime Rates*". The Review of Economics and statistics, Vol. 82, N° 4, pp. 519-529.

Gruber, J. (2000): "*Disability Insurance Benefits and Labor supply*". The Journal of Political Economy, Vol. 108, N°. 6, pp. 1162-1183

Johnston, J. y Dinardo, J. (2001): "*Econometric Methods*". Vicens-Vives, 2001.

Katz, L. y Kessler, D. (2001): "*Prevailing Wage Laws and Construction Labor Markets*". Industrial and Labor Relations Review, Vol. 54, N° 2, pp. 259-274.

Maki, Dean M. (2001): "*Household Debt and the Tax Reform Act of 1986*". The American Economic Review, Vol. 91, N°.1, pp. 305-319

Moulton, B.R., (1990): “*An illustration of a Pit fall in estimating the effects of aggregate variables on micro units*”. The Review of Economics and Statistics 72, 334-338.

Neumark, D. y Wascher, W. (2001): “*Minimum Wages and Training Revisited*”. Journal of Labor Economics, Vol. 19, N° 3, pp. 563-595.

Pepper, J.V., (2002): “*Robust inferences from random clustered samples: an application using data from the panel study of income dynamics*”. Economics letters, Vol 75, N° 3, May 2002, 341-345.

Vicéns, J. (1995): “*Modelos con Variables cualitativas Dicotómicas*”. Instituto de Predicción Económica Lawrence R. Klein. Documento 95/5

White, H. (1984): “*Asymptotic Theory for Econometricians*”. San Diego, CA: Academic Press.

Wooldridge, J. (2006): “*Introducción a la econometría*”. Thomson.