

ERRORES MULTIVARIANTES EN VARIABLES
EN DATOS DE PANEL:
LA DEMANDA DE EMPLEO
EN LA INDUSTRIA MANUFACTURERA ESPAÑOLA

Luis Servén
Documento 88-02

La presente es una versión totalmente revisada al Documento de trabajo FEDEA 87-11. Agradezco a Alfonso Novales y a los participantes en un Seminario en la Universidad Pontificia Católica de Rio de Janeiro por sus comentarios y sugerencias. Todos los errores que puedan quedar en el trabajo son responsabilidad de autor.

CONTENIDO:

1. Introducción
2. Cuestiones teóricas
3. Resultados
4. Conclusiones

Referencias

1. INTRODUCCION

El uso de datos de panel a menudo suscita el problema de controlar de forma adecuada los efectos omitidos específicos de los individuos y su probable correlación con las variables observadas. Mientras que transformaciones sencillas de los datos (tales como diferencias en el tiempo o con respecto a la media de cada grupo) deberían en principio eliminar esta fuente de sesgo, en la práctica a menudo conducen a estimaciones que aparecen "estar más próximas a cero" de lo que la teoría económica predeciría, incluso después de permitir un número suficientes de retardos en los efectos de las variables explicativas¹.

Un trabajo reciente por Griliches y Hausman (1986) ha mostrado que estos resultados contraintuitivos puede deberse a la presencia de errores en variables. Los intentos de eliminar los efectos específicos de los individuos utilizando la transformación "within" o similares puede incrementar el sesgo del tipo errores en variables de las estimaciones paramétricas. Más generalmente, los autores muestran que las diferentes transformaciones de los datos que tratan de eliminar los efectos específicos de los individuos tienen consecuencias diferentes para el sesgo de las estimaciones paramétricas en la presencia de errores en variables². Por lo tanto, una comparación de las estimaciones obtenidas a partir de diferentes

¹Un buen ejemplo de esta situación en el contexto de la estimación de la demanda de trabajo es Mairesse y Dormot (1985).

²Esta observación ha sido también hecho por Griliches (1983).

transformaciones de los datos puede proporcionar una buena indicación acerca de la posible existencia de errores en variables. En su modelo, con un único regresor, los autores muestran que, bajo adecuados supuestos acerca de la estructura de corrección serial del error de medida, puede disponerse de instrumentos "internos" que permitan estimación consistente.

En este artículo examinamos estos temas en el caso multivariante. Con varias variables sujetas a error de medida -una situación que no es en absoluto poco usual con datos de panel- aparecen complicaciones adicionales: Por una parte, la presencia de errores mutuamente correlacionados puede hacer todavía más difícil la identificación; por otra parte, si los errores son mutuamente incorrelacionados pero las variables subyacentes no lo son, la identificación puede ser más sencilla. En general, la identificación puede conseguirse utilizando únicamente "instrumentos internos" siempre y cuando se disponga de suficiente información acerca de la estructura de covarianza conjunta de los errores de medida multivariantes. Proporcionamos condiciones para la identificación y caracterizamos el conjunto de instrumentos válidos en el caso general.

Este tratamiento se aplica más adelante para estimar una sencilla ecuación de demanda de trabajo utilizando datos de panel para las empresas manufactureras españolas. El colapso del empleo en el sector manufacturero en España desde finales de los 70 ha contribuido a atraer una gran atención al análisis de los determinantes de la demanda de empleo. Sin embargo, la mayoría de los estudios empíricos se realizan con series temporales agre-

gadas cuya calidad ha sido a menudo criticada³. Una fuente alternativa de información viene proporcionada por la Central de Balances del Banco de España (CBBE), que desde 1981 recoge datos acerca de una gran sección cruzada de empresas manufactureras.

El artículo está organizado como sigue. Primero, discutimos las cuestiones teóricas en la Sección 2. Los resultados empíricos se presenta en la Sección 3. Las conclusiones se resumen en la Sección 4.

³Mientras que las cifras del empleo agregado carecen de homogeneidad debido a los cambios de definición, la metodología y a la frecuencia de las encuestas de empleo, los datos salariales parecen sobreestimar en gran medida los incrementos salariales de los 70.

2. CUESTIONES TEORICAS

Sea la verdadera ecuación

$$(1) \quad Y_{it} = x_{it}^* \beta + \alpha_i + v_{it} \quad ; \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

donde x_{it}^* y β son vectores $1 \times K$ y $K \times 1$ respectivamente, α_i es un efecto no observado específico de cada individuo y v_{it} es una perturbación aleatoria con media cero. Suponemos que x_{it}^* es ortogonal a v_{it} pero potencialmente correlacionada con α_i . Además los verdaderos x_{it}^* no son directamente observables; sólo su medida observada con ruido, x_{it} está disponible al económetra:

$$(2) \quad x_{it} = x_{it}^* + \epsilon_{it} \quad ; \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

donde ϵ_{it} es un vector $1 \times K$ de errores de medidas. En términos de variables observables tenemos:

$$\begin{aligned} (3) \quad Y_{it} &= x_{it} \beta + (\alpha_i + v_{it} - \epsilon_{it} \beta) \\ &= x_{it} \beta + u_{it} \quad ; \quad i = 1, \dots, N; \quad t = 1, \dots, T \end{aligned}$$

Es claro que MCO o MCG aplicados a la ecuación (3) producirán estimaciones inconsistentes de β debido a la correlación de x_{it} con el término de error compuesto. Transformar los datos para eliminar el error específico de la unidad eliminará una fuente de sesgo pero no corregirá la presencia de errores de medida. Sin embargo, β puede ser identificable y estimable si pueden imponerse a priori suficientes restricciones sobre la estructura de covarianzas de las variables u_{it} . Para ver esto, ordenemos como vector columna las observaciones para un valor dado de i y reescribamos el modelo

(abandonando el subíndice i por conveniencia)

$$(4) \quad y = x^* \beta + 1_T \alpha + v$$

$$(5) \quad x = x^* + \epsilon$$

donde y y v son vectores $T \times 1$, 1_T es un vector $T \times 1$ de unos y x , x^* y ϵ son matrices $T \times K$. Para identificar β necesitamos una matriz de $T \times L$ (donde $L \geq K$) variables instrumentales -que denotaremos por Z - incorrelacionada con la perturbación compuesta pero correlacionada con los verdaderos regresores:

$$(6) \quad E(Z'u) = 0$$

$$(7) \quad \text{rank } E(Z'x^*) = K$$

En ausencia de cualquier información externa, puede construirse instrumentos válidos utilizando combinaciones lineales de los valores corrientes, retardados y futuros de las variables en el miembro derecho de la ecuación sujeta a error. Esto es, denotando por z una columna cualquiera de Z , un conjunto de instrumentos viene dado por

$$(8) \quad z = P \text{vec}(x)$$

donde P es una matriz $T \times TK$. Para que z sea un instrumento válido, (6) requiere que

$$(6') \quad E[\text{vec}(x)'P'(1_T \alpha + v - \epsilon \beta)] = 0$$

En ausencia de información más específica acerca de la estructura de correlaciones entre los verdaderos regresores y las perturbaciones específicas de cada individuo, la ecuación (6') generalmente equivale a dos requisitos diferentes:

$$(9a) \quad P'1_T = 0$$

$$(9b) \quad \text{tr} \left([(\beta' \otimes I_T) \Sigma + R] P' \right) = 0$$

donde 0 denota el producto de Kronecker, y hemos definido la matriz simétrica $TK \times TK$:

$$\Sigma = E [\text{vec}(\epsilon) \text{vec}(\epsilon)']$$

que suponemos positiva semidefinida, y la matriz $T \times TK$

$$R = E [v \text{vec}(\epsilon)']$$

La condición (9a) asegura la eliminación del efecto α específico de los individuos haciendo que las columnas de P sumen cero, imponiendo de este modo KT restricciones sobre sus elementos. La condición (9b) es una representación de la ausencia de correlación entre los instrumentos y la perturbación compuesta. De este modo, el vector de parámetros β puede identificarse si se encuentran al menos K P -matrices linealmente independientes que satisfagan (9a), (9b) y (7).

Más formalmente, una condición de orden para la identificación puede establecerse del siguiente modo: Primero, notemos que (9b) puede reescribirse

$$(10) \quad \text{vec}[\Sigma(\beta \otimes I_T) + R']' \text{vec}(P') = 0$$

El primer término en corchetes en (10) es una suma de matrices $TK \times T$ de la forma $\beta_k \otimes \Sigma_k$, donde $\Sigma_k = E[\text{vec}(\epsilon) \epsilon_k']$ y ϵ_k denota la k -ésima columna de ϵ , para $k=1, \dots, K$. Para β , Σ y R arbitrarios es fácil ver que cada elemento diferente sin restringir de esta suma impondrá en general una restricción lineal sobre la matriz P , forzando a que uno de sus elementos sea cero. Definamos una matriz S cuya j -ésima fila corresponda a la j -

ésima restricción; S tendrá entonces s filas donde s es el número de elementos diferentes sin restringir en la matriz $\Sigma(\beta \otimes I_T) + R'$. Entonces las ecuaciones (9) pueden escribirse de modo compacto:

$$(11) \quad \begin{bmatrix} S \\ \dots\dots\dots \\ I_T' \otimes I_{KT} \end{bmatrix} \text{vec}(P') = 0$$

La condición de orden requiere que existan al menos K P-matrices linealmente independiente cuyos elementos satisfagan las (s + KT) ecuaciones lineales (11). En otras palabras, el espacio de soluciones de (11) debe tener al menos dimensión K. Puesto que la dimensión del espacio de soluciones es igual a KT^2 menos el rango de la matriz que aparece en (11), se sigue que la condición de orden se cumplirá si y sólo si la matriz tenga rango menor o igual que $K(T^2 - 1)$. Una condición suficiente para que ello ocurra es:

$$(12) \quad K(T^2 - T - 1) \geq s$$

Notemos que la condición no es necesaria porque en algunos casos si (9b) o (10) se cumple entonces algunas de las restricciones implicadas por (9a) será redundante, en cuyo caso el número de ecuaciones linealmente independientes en (11) será estrictamente inferior a $s + KT^4$.

Por otra parte, la condición de rango (7) puede escribirse

⁴Esto ocurre cuando las primeras s ecuaciones en (11) fuerzan a que una columna de P sea igual a vector de ceros de dimensión T. La restricción de que sumen cero es entonces redundante.

$$(13) \quad \text{rank } E \left[\frac{\text{vec}(x)' P_1' x}{\text{vec}(x)' P_L' x} \right] = K$$

donde $P_1, \dots, P_L$ ($L \geq K$) son un conjunto de matrices linealmente independientes que satisfacen (11). En esencia, (13) requiere que cada instrumento esté correlacionado al menos con uno de los regresores, y que cada regresor esté correlacionado con al menos un instrumento.

La identificabilidad del modelo depende crucialmente de la estructura de las matrices de covarianzas Σ y R . Situaciones diferentes surgen dependiendo de la correlación mutua y de la estacionalidad de los errores de medida y de las variables subyacentes. Por razones de espacio limitaremos nuestros comentarios a cuatro casos generales.

(I) Errores de medida mutuamente incorrelacionados, y las variables verdaderas x^* 's mutuamente incorrelacionadas. En este caso sencillo, Σ es una matriz diagonal a bloques. Si los errores de medida son también incorrelacionados con el término original de perturbaciones, R es una matriz de ceros. La situación entonces similar al caso de un único regresor discutido por Griliches y Hausman, puesto que cada instrumento puede asociarse de modo único con uno de los regresores. En tal caso, las P -matrices válidas pueden expresarse como $P = (0, \dots, 0, Q, 0, \dots, 0)$, donde Q y 0 son matrices $T \times T$. Cada instrumento hace uso únicamente de una variable del miembro de la derecha de la ecuación. Las propiedades de cada submatrix $\Sigma_{kk} = E[\epsilon_k \epsilon_k']$ determinan el número de instrumentos

válidos disponibles para el k-ésimo regresor⁵. Si R no es trivial de modo que, por ejemplo, $E[v\epsilon_j']$ no es cero para algún j , el número disponible para el j-ésimo regresor puede ser pequeña. (véase pie de página número 8).

(II) Errores de medida mutuamente incorrelacionados, pero variables x^* 's mutuamente correlacionadas. Este caso puede proporcionar posibilidades adicionales de identificación: incluso si para algún k la estructura de la matriz Σ_{kk} hace que no todas las combinaciones lineales de valores pasados y futuros del k-ésimo regresor no sea instrumentos válidos porque, por ejemplo, ϵ_k no es estacionario y está serialmente correlacionado a todos los retardos temporales-, puede utilizarse combinaciones lineales de valores pasados y futuros de otras variables del miembro derecho de la ecuación cuyos correspondientes errores de medida muestren correlación serial de orden bajo para el mismo propósito⁶.

(III) Variables x^* 's mutuamente incorrelacionadas pero errores de medida mutuamente correlacionados. En este caso, bastante improbable, los instrumentos válidos pueden asociarse de nuevo de modo único con el regresor correspondiente al menos que se disponga de información específica

⁵En este caso (9b) requiere que

$$\text{tr}[(\beta_1 \otimes \Sigma_{11}, \dots, \beta_k \otimes \Sigma_{kk})P'] = 0$$

Sin ninguna información que ligue los elementos de diferentes submatrices Σ_{kk} , cada Σ_{kk} determinará la estructura de la correspondiente submatriz TxT de orden P .

⁶Nótese la similaridad entre este caso y la situación considerada por Griliches y Hausman en la que los errores de medida son estacionarios pero las verdaderas variables no lo son, proporcionando por tanto instrumentos adicionales válidos.

acerca de la estructura de las submatrices Σ_{jk} (para j diferente de k)⁷. Para una estructura dada de correlación serial de cada error de medida individual, cada correlación cruzada no contemporánea restringe el número de instrumentos disponibles⁸. En general, la identificación será más difícil según aumente el número de correlaciones (cruzadas o seriales) en retardos temporales no nulos.

(IV) Variables x^* 's mutuamente correlacionadas, y errores de medida mutuamente correlacionados. De nuevo cada correlación cruzada no contemporánea de los errores de medida imponen restricciones adicionales sobre la estructura de las P-matrices válidas. En este caso, sin embargo, la estructura diferenciada de las correlaciones cruzadas entre los errores de medida y las variables x^* 's puede todavía permitir el uso de algunas combinaciones lineales de columnas diferentes de x como instrumentos, además de las consideradas en (III).

⁷Por supuesto, toda información específica acerca de la estructura de Σ y R puede utilizarse en este así como en los otros casos para proporcionar instrumentos válidos adicionales. Tal información contribuye a restringir el número de elementos diferentes en la matriz de coeficientes del sistema de ecuaciones (11).

⁸Nótese que únicamente las correlaciones cruzadas no contemporáneas introducen restricciones adicionales. Las correlaciones contemporáneas no comparten esta propiedad porque sin información adicional, las varianzas y covarianzas no pueden identificarse de modo separado: Las K covarianzas contemporáneas (para un dado t) entre la variable dependiente y las x 's no puede identificar K varianzas de error más $K(K-1)/2$ más los correspondientes K elementos de r . Esto refleja el hecho general de que únicamente los elementos diferentes de la matriz suma $\Sigma (\beta \otimes I) + R'$ (y no los elementos individuales de Σ y R) imponen restricciones sobre la estructura de las P-matrices.

Por supuesto, en el caso general de matrices Σ y/o R sin restringir la identificación no es posible sin la utilización de información externa. En muchos casos, sin embargo, los supuestos habituales de la estructura de estas matrices dejarán el modelo sobreidentificado. Las restricciones de sobreidentificación pueden entonces contrastarse, proporcionando información acerca de la validez de las especificaciones alternativas del modelo y de los procesos de error multivariantes.

Si el modelo está identificado, de tal modo que pueden encontrarse M instrumentos válidos con $M \geq K$, el estimador asintóticamente eficiente del vector de parámetros β puede calcularse utilizando el método generalizado de Hansen (1982). Definiendo la matriz $X = (x_1', \dots, x_N')'$ de dimensión $NT \times K$, donde x_i ($i = 1, \dots, N$) es una matriz $T \times K$; el vector $Y = (Y_1', \dots, Y_N')'$ $NT \times 1$, donde y_i ($i = 1, \dots, N$) es un vector $T \times 1$; y la matriz de instrumentos $W = (Z_1', \dots, Z_N')'$ de dimensión $NT \times L$, donde cada Z_i ($i = 1, \dots, N$) es una matriz $T \times L$, el estimador viene dado por

$$b = (X'W \Omega^{-1} W'X)^{-1} X'W \Omega^{-1} W'Y$$

donde $\Omega = E[Z'uu'Z]$ es la matriz de ponderaciones óptimas (ver Hansen (1982) o White (1982)). En la práctica, Ω se sustituye por una estimación consistente utilizando los residuos de una estimación preliminar de β .

3. RESULTADOS EMPIRICOS

Consideremos una especificación muy sencilla de la demanda de empleo que relaciona al empleo con los salarios reales y el producto, así como la productividad del trabajo -cuya evolución viene descrita por una tendencia temporal- y otros efectos no observables específicos de las empresas. Más específicamente,

$$(14) \quad l_{it} = \alpha_i + \beta_1 t + \beta_2 q_{it} + \beta_3 w_{it} + v_{it}$$

donde l , q y w representan los logaritmos del empleo, el producto y el salario real, respectivamente, t representa el tiempo, α_i es una perturbación específica de cada empresa y v_{it} es una perturbación de media cero incorrelacionada a través de los valores del índice i . Para evitar problemas de simultaneidad, las empresas se supone que operan en un régimen keynesiano, con el producto determinado por el lado de la demanda.

Utilizamos datos del panel formado por la Central de Balances del Banco de España (CBBE) acerca de 820 empresas manufactureras para los años 1981-1985. El empleo total en estas empresas equivalía en 1985 a 360.000 trabajadores, aproximadamente el 15% del empleo manufacturero agregado en España.

La muestra proporciona únicamente información acerca de las cantidades nominales. Por tanto, para llegar a una medida del producto real y de los salarios reales utilizamos los deflatores y de los precios industriales proporcionados

por el INE, que considera una desagregación de 17 sectores del sector manufacturero. Ello puede introducir errores de medida (mutuamente correlacionados) en ambas variables. El producto se definió igual a las ventas más los cambios de existencias más trabajos realizados internamente para el stock de capital de las empresas. Sin embargo, la CBBE proporciona información únicamente acerca de los costes salariales totales de la empresa. Por lo tanto, las cifras de salarios tuvieron que construirse dividiendo la factura salarial por el número promedio de empleados, introduciendo otra posible fuente de error de medida muy probablemente correlacionada con la perturbación original v.

Para resumir, las variables en el miembro derecho de la ecuación (14) están sujetas a errores de medida debido al uso de deflatores de precios posiblemente no adecuados, a la definición endógena del salario real -y de modo más importante- al uso de salarios reales y producto verdaderos en vez de sus valores anticipados "permanente" no observables, que son las variables relevantes teóricamente, en tanto en cuanto las decisiones de contratación y restricciones de contrato encierra alguna clase de costes fijos.

Despreciando estos hechos por el momento, estimamos la ecuación de empleo utilizando métodos convencionales. La Tabla 1 muestra los resultados de la estimación por MCG así como para la especificación de "within"⁹. Mientras que las estimaciones de los coeficientes del salario real y el

⁹La especificación MCG inicialmente incluía 16 variables ficticias sectoriales, que fueron más tarde excluidas debido a su falta de significación conjunta.

TABLA 1
ESTIMACIONES DE LA ECUACION DE EMPLEO

Coeficiente de la variable

<u>Método</u>	<u>Producto</u>	<u>Salario Real</u>	<u>Tiempo</u>	<u>Número de observaciones</u>
MCG	.587 (.003)	-.645 (.014)	-.020 (.002)	4.100
Within	.461 (.011)	-.607 (.014)	-.017 (.002)	4,100

Nota: desviaciones típicas en paréntesis.

tiempo son aproximadamente similares en ambos casos, la estimación MCG del producto parece sustancialmente más alta (más de 25%) que su contrapartida en la especificación "within", con estimaciones puntuales en ambos casos claramente por debajo de la unidad.

Un contraste de Hausman de la especificación MCG frente a la especificación "within" proporciona un estadístico que bajo la hipótesis nula se distribuye como una chi-cuadrado con tres grados de libertad. El valor calculado fue 428. Por tanto, la hipótesis nula de que la especificación MCG sea correcta se rechaza de modo abrumador, un resultado que apenas debe sorprender dada la más que probable correlación entre el producto y los salarios de las empresas y características no observables específicas de las empresas tales como la capacidad de sus directivos o sus diferencias tecnológicas.

En este punto, una posible interpretación de los resultados sería aceptar la presencia de importantes efectos específicos y concluir que la elasticidad producto de la demanda del trabajo está por debajo de .5, lo que sugiere la existencia de sustanciales economías de escala¹⁰.

Una alternativa es contrastar la existencia de errores en variables que pudieran aparecer por

¹⁰ Especificaciones alternativas que incluyen el empleo retardado y/o el producto en las ecuaciones (siguiendo la línea del modelo de coste de ajuste) condujeron a resultados similares. Más aún, cuando estos modelos fueron estimados utilizando métodos de variables instrumentales (teniendo cuidado con las variables endógenas retardadas; ver p.e. Anderson y Hsiao (1982)) el coeficiente del empleo retardado resultó siempre no significativamente diferente de cero.

las causas antes descritas. Ello puede conseguirse calculando estimadores alternativos a partir de los datos diferenciados. Como ha sido indicado por Griliches y Hausman (1986) si todas las variables se miden sin error entonces MCO aplicado a los datos diferenciados debería proporcionar estimaciones similares con independencia de la longitud del intervalo temporal de acuerdo con el cual se diferencia.

Las estimaciones resultantes se muestran en la Tabla 2. Nótese que los coeficientes del producto y del progreso tecnológico (el último dado por la constante de la regresión tras aplicar un factor de escala de forma apropiada) crece de modo monótono en valor absoluto cuando crece la longitud del periodo de diferenciación, mientras que el coeficiente del salario real se comporta de modo mas ambigüo. Un contraste de Wald de igualdad de las estimaciones correspondientes a las diferencias primera y más larga condujo a una variable chi-cuadrado con 12 grados de libertad. Su valor calculado fue 42.1, rechazando claramente la hipótesis nula.

A la vista de estos resultados, parece apropiado respecificar el modelo permitiendo tanto errores de medida como correlación no nula entre los regresores y los efectos omitidos específicos de las empresas. En ausencia de cualquier otra información exterior, tratamos la identificación y estimación del modelo utilizando tan solo instrumentos internos del tipo descrito en la sección anterior. Para cada especificación alternativa de las propiedades estocásticas de los errores de medida el conjunto de instrumentos disponibles puede encontrarse facilmente utilizando las ecuaciones (11).

TABLA 2

ESTIMACIONES DE LA ECUACION DE EMPLEO
PARA PERIODOS DE DIFERENCIACION ALTERNATIVOS

Coeficiente de				
<u>Método</u>	<u>Producto</u>	<u>Salario Real</u>	<u>Const^a</u>	<u>Número de observaciones</u>
Primeras diferencias	.378 (.011)	-.615 (.011)	-.014 (.002)	3.280
Segundas diferencias	.420 (.013)	-.653 (.015)	-.017 (.004)	2.460
Terceras diferencias	.510 (.016)	-.624 (.022)	-.019 (.006)	1.640
Cuartas diferencias	.532 (.023)	-.424 (.035)	-.019 (.010)	820

Nota: desviaciones típicas en paréntesis.

^a dividida por la longitud del intervalo de diferenciación

Comenzamos suponiendo que los errores de medida no son estacionarios pero que están incorrelacionados serialmente. Permitimos también que estén correlacionados mutuamente en un mismo período y correlacionados con la perturbación original¹¹. Dado nuestro tamaño muestral ($T=5$) esto impone diez restricciones sobre los elementos de las P-matrices definidas mas arriba, mas diez restricciones adicionales precisan para eliminar los efectos específicos de las empresas¹². Por tanto, tenemos 30 instrumentos (31 con la tendencia temporal) y el modelo está claramente sobreidentificado. Si permitimos la no estacionariedad conjunta del proceso de medias móviles de primer orden de los errores de medida reduciríamos adicionalmente números disponibles a 14 (22 si las correlaciones cruzadas no contemporáneas se suponen no nulas. Un proceso conjunto de segundo orden proporcionaría únicamente cuatro instrumentos válidos (16 si todas las correlaciones cruzadas no contemporáneas son nulas)¹³. Un proceso de orden superior para los errores de medida dejarían el modelo identificado sin información adicional (p.e. estacionariedad) sobre sus propiedades estocásticas.

¹¹Nótese que estas correlaciones cruzadas no imponen ninguna restricción adicional sobre las matrices de instrumentos válidos (ver pie de página 8).

¹²Suponemos que la tendencia temporal es un instrumento válido. Por tanto, nos referiremos únicamente al subsistema (q,w) cuando se calcula el número de instrumentos disponibles.

¹³De nuevo, nótese que los instrumentos seleccionados para los casos de estructuras conjuntas MA(0), MA(1) y MA(2) permanecen válidos si los errores de medida están correlacionados con la perturbación original y hasta los retardos temporales de orden cero, uno o dos, respectivamente.

Para cada una de estas especificaciones, el estimador eficiente asintóticamente fue calculado utilizando el Método Generalizado de Momentos (MGM). La matriz de ponderaciones óptima se construyó utilizando residuos de una estimación preliminar por mínimos cuadrados en 2 etapas.

Las estimaciones MGM resultantes se muestran en la Tabla 3. La elasticidad estimada del producto crece de .5 bajo el supuesto de errores de medida incorrelacionados serialmente a casi .9 bajo el supuesto MA(1) o MA(2) sin correlación. Las últimas estimaciones no son estadísticamente diferentes de la unidad. El valor absoluto de la elasticidad con respecto al salario real muestra una caída paralela¹⁴. Nótese también que según permitimos más correlaciones seriales y cruzadas no nulas entre los errores de medida las desviaciones típicas estimadas (que permiten heteroscedasticidad condicional) crecen gradualmente, reflejando la disminución en el número de instrumentos válidos.

La última columna de la Tabla 3 muestra el estadístico para el contraste de las restricciones de sobreidentificación (Hansen (1982)), que proporciona un contraste formal acerca de la validez de los supuestos que subyacen a las especificaciones alternativas. Bajo la hipótesis nula de una especificación correcta, dicho estadístico se distribuye como una chi-cuadrado con un número de grados de libertad igual al número de instrumentos, menos el número de parámetros estimados. Para

¹⁴ Nótese que el error de medida que surge de la definición de los salarios tendería a sesgar el coeficiente estimado hacia -1.

TABLA 3
ESTIMACIONES MGM DE LA ECUACION DEL EMPLEO

Coeficiente en				
Proceso de error <u>Supuesto</u>	<u>Producto</u>	<u>Salario Real</u>	<u>Tiempo</u>	Contraste de las restricciones de de sobreidentifi- cación (g.d.e)
MA(0) ^a	.505 (.030)	-.736 (.063)	.018 (.002)	59.7 (28)
MA(1) conjunto ^b	.590 (.163)	-.560 (.097)	-.022 (.004)	12.5 (12)
MA(1) ^a	.826 (.137)	-.636 (.107)	-.027 -.028	17.2 (20)
MA(2) ^a	.891 (.168)	-.589 (.140)	-.028 (.005)	14.2 (14)

Notas: ^a Permitiendo unicamente correlación cruzada contemporánea de los errores
 ^b Permitiendo también correlación cruzada en el primer retardo temporal.

el caso de errores de medida serialmente incorrelacionados el estadístico es 59.7 con 28 grados de libertad, lo que proporciona clara evidencia de error de especificación. En contraste, el estadístico no puede rechazar la validez de las restricciones de sobreidentificación para las otras especificaciones¹⁵.

Concluimos esta sección con un breve resumen de nuestros resultados empíricos. Hemos encontrado evidencia acerca de que las estimaciones de datos de panel y de la sección cruzada convencionales de la demanda de empleo en el sector manufacturero español pueden estar recogiendo sesgos de errores en las variables debido tanto a la probable importancia de los efectos emitidos específicos de las empresas como a la existencia de errores de medida en las variables. En particular, la elasticidad producto de la demanda de trabajo obtenida a partir de las estimaciones convencionales "within" o con primeras diferencias aparece estar sesgada claramente hacia el origen.

Aplicando técnicas de variables instrumentales hemos llegado a una especificación cuya validez no puede ser rechazada sobre la base de los contrastes convencionales y que coloca a la elasticidad producto de la demanda de trabajo entre

¹⁵El contraste de especificación de Hausman (1978) puede también utilizarse para comparar las estimaciones paramétricas bajo los supuestos alternativos acerca del proceso de errores. En nuestro caso, conducen a estadísticos chi-cuadrado con tres grados de libertad. Por ejemplo, el valor calculado cuando se comparan las especificaciones conjuntas MA(0) y MA(1) fue 27.3, claramente por encima de los niveles de significación habituales. Para la comparación del supuesto de ausencia de correlación frente a la especificación MA(1) conjunta, el valor fue calculado 5.0, que conduce a no rechazar la primera especificación.

.8 y .9 (y no significativamente distinta de 1) y la elasticidad salarial real alrededor de .6. Estos resultados son de hecho bastante similares a los obtenidos utilizando datos de series temporales para el sector manufacturero español¹⁶.

¹⁶Por ejemplo, las estimaciones de series temporales obtenidas por Raymond y otro (1986) produjeron la siguiente ecuación de empleo a largo plazo (desviaciones típicas en paréntesis):

$$l = q - .456w/p - .035t$$

(.157) (.010)

4. CONCLUSIONES

El uso de datos de panel en la investigación econométrica a menudo conduce a estimaciones paramétricas que parecen "muy próximas a cero" a juzgar por la teoría economía habitual. Este tema ha sido investigado recientemente por Griliches y Hausman (1986) en el caso de un único regresor. Estos autores muestran que en muchos casos, estos resultados poco satisfactorios pueden deberse a la presencia de errores de medida y puntualizan que, bajo ciertas condiciones, existen suficientes "instrumentos internos" como para permitir la estimación consistente de los parámetros incluso en ausencia de cualquier información exterior.

En este artículo hemos extendido estos resultados al modelo de regresión multivariante. La identificación sin instrumentos externos puede ser posible si existe suficiente información acerca de las propiedades del error de medida multivariante. Más específicamente, la estructura de correlaciones cruzadas entre los errores de medida se convierte en uno de los determinantes clave de la identificación del modelo. Hemos mostrado también como puede hallarse el conjunto de instrumentos disponibles, para una especificación dada de la matriz de covarianzas del error medida multivariante.

Hemos seguido este procedimiento para investigar la respuesta del empleo manufacturero al producto y a los salarios reales, una cuestión clave en el actual debate político acerca de los remedios al problema del desempleo en España. Utilizando un panel de 820 empresas manufacture-

ras, hemos estimado un sencillo modelo de demanda de empleo. Hemos encontrado que es muy probable que estimaciones convencionales de datos de panel subestimen la respuesta del empleo a cambios del producto real, debido a la presencia de errores de medida.

Estimando el modelo bajo supuestos alternativos acerca de las propiedades estocásticas de los errores de medida multivariantes, hemos podido obtener estimaciones consistentes de la ecuación de demanda de empleo, que en particular sitúan la elasticidad-producto cerca de la unidad y que parecen estar bastante de acuerdo con las obtenidas a partir de datos de series temporales.

REFERENCIAS

Anderson, T. y C. Hsiao, (1982): "Formulation and Estimation of Dynamic Models using Panel Data", Journal of Econometrics 18, p 47-82.

Griliches, Z. (1983): "Sibling Models and Data in Economics: Beginnings of a Survey", Journal of Political Economy 87, p. S37-S64.

Griliches, Z. y J. Hausman (1986): "Errors in Variables in Panel Data", Journal of Econometrics 31, p. 93-118.

Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica 50, p. 1029-1054.

Hausman, J. (1978): "Specification tests in Econometric", Econometrica, 45, p. 1251-1271.

Mairesse, J. y B. Dormot (1985): "Labor and Investment Demand at the Firm Level", NBER Working Paper 1551.

Raymond, J. et.al. (1986): "Factores explicativos de la demanda de empleo", Papeles de Economía Española, 26'

White, H. (1982): "Instrumental Variables Regression with Independent Observations", Econometrica 50, p. 483-499.