CONTRASTES DE HIPÓTESIS EN DATOS DE PANEL

Cobacho Tornel, Mª Belén
Universidad Politécnica de Cartagena

Bosch Mossi, Mariano
London School of Economics

RESUMEN

Una de las principales aportaciones del análisis de datos de panel es que permite medir efectos y comportamientos que no es posible identificar en datos puros de series temporales o de corte transversal. Cuando se plantea la cuestión de cuál es la especificación más adecuada a nuestra situación de estudio con datos de panel, o cuál el método de estimación más apropiado, los contrastes de hipótesis suponen una herramienta útil que permite dar soporte o rechazar determinados supuestos. La doble dimensionalidad de los conjuntos de datos de panel permite en algunos casos adaptar los contrastes de hipótesis ya planteados para alguna de sus dos dimensiones, y en otros, ha requerido el planteamiento de tests propios de los datos de panel. En este trabajo se recogen algunos de los contrastes más frecuentes que pueden plantearse cuando se trabaja con datos de panel, y se presenta una aplicación al estudio de los efectos de la inversión pública federal mexicana, con datos de los estados mexicanos durante el periodo 1970-2000.

1. INTRODUCCIÓN

Un panel de datos es un conjunto de datos que combinan series temporales con unidades de sección cruzada o de corte transversal (países, regiones, empresas, hogares, etc.), de forma que un estudio de los datos considerando estas dos dimensiones por separado (tiempo y sección cruzada) deja cuestiones sin resolver. En Baltagi (2001) se enumeran algunas de las ventajas e inconvenientes del uso de los datos de panel. Entre las ventajas se menciona el control sobre la heterogeneidad individual; más variabilidad, menos colinearidad entre las variables, más grados de libertad y mayor eficiencia; mejor adecuación al estudio de las dinámicas de ajuste; mejor capacidad de identificar y medir efectos que no son detectables en datos puros de sección cruzada o de series temporales y también mejor capacidad de análisis en comportamientos más complicados. Como desventajas, los datos de panel presentan el problema de recolección de datos, distorsiones por errores de medida y la corta dimensión temporal que se tiene generalmente en los conjuntos de datos.

Teniendo en cuenta sus limitaciones, y a pesar de las ventajas que presentan, cuando nos enfrentamos al análisis de un panel de datos, existen gran cantidad de cuestiones que cabe plantearse a la hora de mantener determinados supuestos y de elegir un método de estimación, para poder así dar un mayor soporte al estudio que se está realizando. Estas cuestiones se nos plantearon como parte de un estudio de un panel de datos de los estados mexicanos durante el periodo 1970-2000, con el que se pretende analizar si la inversión pública federal en México ha tenido algún efecto sobre determinados indicadores de desarrollo. En este trabajo se muestran algunas de las cuestiones que cabe plantearse a la hora de enfrentarse a un estudio aplicado de datos de panel y se recogen contrastes de hipótesis existentes en la literatura que pueden ayudar a dar respuesta a tales cuestiones. Un estudio en mayor profundidad de estos y otros contrastes, así como referencias bibliográficas a estudios teóricos y empíricos puede encontrarse en Baltagi (2001).

El trabajo se estructura de la siguiente manera: en las secciones 2 y 3 se comentan los contrastes de hipótesis de los modelos estáticos y dinámicos respectivamente. En la sección 4 se presenta una aplicación al estudio del efecto de la inversión pública federal en México. En la sección 5 se exponen las conclusiones.

2. CONTRASTES EN MODELOS ESTÁTICOS

Los modelos de datos de panel se clasifican en dinámicos o estáticos según incluyan o no en sus ecuaciones variables pertenecientes a diferentes periodos temporales. En la especificación estática, el *modelo de regresión de un solo factor* supone que el error aleatorio se descompone en $\varepsilon_{it} = \alpha_i + v_{it}$, donde cada α_i es el efecto individual (inobservado) de cada unidad de sección cruzada, invariante en el tiempo[1]. El modelo a estimar es el siguiente:

$$y_{it} = \alpha_i + X_{it}^{'}\beta + v_{it} \qquad (2.1)$$

La presencia del efecto fijo en esta ecuación hace que una estimación de β por MCO no sea consistente. Los métodos que se utilizan para solventar ese problema son la estimación de efectos fijos (EF) o *entre grupos* (*within groups*) y la estimación de efectos aleatorios (RE) por Mínimos Cuadrados Generalizados (MCG). La estimación de EF puede llevarse a cabo transformando el modelo (2.1) en otro en términos de las medias de grupo:

$$y_{it} - y_{i.} = (X_{it} - X_{i.})'\beta + (v_{it} - v_{i.}) \qquad (2.2)$$

y estimando por MCO, o también mediante una estimación por mínimos cuadrados de variables ficticias, α_i. En este contexto, pueden plantearse los siguientes contrastes.

2.1. Contraste de efectos individuales en modelos de componentes de error de un solo factor

La significatividad conjunta de las variables ficticias en un modelo de componentes de error de un factor puede ser contrastada mediante un test F con la hipótesis nula $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_{N-1} = 0$. El estadístico de contraste y su distribución bajo la hipótesis nula son los siguientes:

$$F_0 = \frac{(RSS_R - RSS_U)/(N-1)}{RSS_U/(NT - N - K)} \ \overset{H_0}{:} \ F_{N-1, N(T-1)-K} \qquad (2.3)$$

que es un test de Chow en el que RSS_R es la suma de cuadrados de residuos que se obtiene de la estimación MCO en el modelo agrupado y RSS_U es la suma de cuadrados de los residuos de la estimación por mínimos cuadrados de variables dummy.

[1] α_i y v_{it} verifican los supuestos: α_i : $i.i.d.(0,\sigma_\alpha^2)$ sobre i, v_{it} : $i.i.d.(0,\sigma_v^2)$ sobre i y t, y α_i independiente de $v_{jt}\ \forall i, j, t$.

2.2 Contraste de efectos individuales y de tiempo en modelos de componentes de error de dos factores

Un *modelo de regresión de componentes de error de dos factores* incluye, además de un efecto individual invariante en el tiempo, α_i, un efecto común a todos los individuos, μ_t, el cual captura efectos temporales o macroeconómicos inobservables. El modelo de regresión es por tanto[2]:

$$y_{it} = \alpha_i + X_{it}'\beta + \mu_t + \nu_{it} \qquad (2.5)$$

El modelo de estimación *within groups* conlleva tomar medias en (2.5), tanto en los individuos como en el tiempo, y transformar el modelo en:

$$y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..} = (X_{it} - \bar{X}_{i.} - \bar{X}_{.t} + \bar{X}_{..})'\beta + (\nu_{it} - \bar{\nu}_{i.} - \bar{\nu}_{.t} + \bar{\nu}_{..}) \qquad (2.6)$$

En este modelo de componentes de error de dos factores se puede también contrastar la significatividad conjunta de las dummies de forma similar a como se indica para los modelos de un solo factor, pero para los dos grupos de variables ficticias:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_{N-1} = 0 \text{ y } \mu_1 = \mu_2 = \ldots = \mu_{T-1} = 0$$

El estadístico para este caso es:

$$F_0 = \frac{(RSS_R - RSS_U)/(N+T-2)}{RSS_U/[(N-1)(T-1)-K]} \overset{H_0}{:} F_{N+T-2,(N-1)(T-1)-K}$$

donde ahora RSS_R es la suma de cuadrados de los residuos de la estimación por MCO en el modelo agrupado y RSS_U es la suma de los cuadrados de los residuos de la regresión *within* en (2.6).

Es posible además contrastar la existencia de efectos individuales de grupo dada la existencia de efectos temporales:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_{N-1} = 0 \text{ dado que } \mu_1, \mu_2, \ldots, \mu_{T-1} \neq 0$$

En este caso, RSS_R procede de la regresión que incluye únicamente dummies temporales, o bien de la regresión en $y_{it} - \bar{y}_{.t} = (X_{it} - \bar{X}_{.t})'\beta + (\nu_{it} - \bar{\nu}_{.t})$, mientras que RSS_U sigue siendo la suma de cuadrados residuales de la regresión *within* en (2.6). La distribución del estadístico es en este caso:

---

[2] α_i y ν_{it} verificando las mismos supuestos que en el modelo de un solo factor y además μ_t *i.i.d.*: $(0, \sigma_\mu^2)$ para todo t, y α_i, μ_t y ν_{jt} mutuamente independientes para todo i, j, t.

$$F_0 \overset{H_0}{\sim} F_{(N-1),(N-1)(T-1)-K}$$

Obsérvese que la diferencia entre este test y el contrastes de efectos individuales en los modelos de un solo factor (estadístico (2.3)), es que allí se está contrastando la hipótesis $H_0 : \alpha_i = 0,\ i = 1,2,\ldots N-1$ *suponiendo* que $\mu_t = 0$ para todo $t = 1,2,\ldots,T-1$, mientras que aquí la hipótesis es $H_0 : \alpha_i = 0,\ i = 1,2,\ldots N-1$ *sabiendo* que $\mu_t = 0$ para todo $t = 1,2,\ldots,T-1$.

De forma análoga, se puede contrastar la existencia de efectos temporales conociendo la existencia de efectos individuales:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_{T-1} = 0 \text{ dado que } \alpha_1, \alpha_2, \ldots, \alpha_{N-1} \neq 0 ,$$

donde ahora RSS_R viene dado por la regresión en (2.2) y RSS_U procede de la regresión en (2.6). El estadístico de contraste en ese caso es:

$$F_0 \overset{H_0}{\sim} F_{(T-1),(N-1)(T-1)-K}$$

Breusch y Pagan (1980) plantean un test LM (Multiplicadores de Lagrange) para contrastar la hipótesis: $H_0 : \sigma_\alpha^2 = \sigma_\mu^2 = 0$.

Trabajando a partir de la función de verosimilitud que se emplea para la estimación por máxima verosimilitud en (2.5) , se construye el estadístico:

$$LM = LM_1 + LM_2 \overset{H_0}{\sim} \chi_2^2$$

donde $LM_1 = \left[1 - \frac{\hat{\varepsilon}'(I_N \otimes J_T)\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}}\right]^2 \frac{NT}{2(T-1)}$, $LM_2 = \left[1 - \frac{\hat{\varepsilon}'(J_N \otimes I_T)\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}}\right]^2 \frac{NT}{2(N-1)}$,

siendo $\hat{\varepsilon}$ los residuos de la estimación MCO en (2.5), I_N e I_T las matrices identidad de tamaño N y T respectivamente, y J_T, J_N matrices de unos de tamaños T y N respectivamente. Si lo que se quiere es contrastar $H_1 : \sigma_\alpha^2 = 0$, el estadístico de contraste es precisamente LM_1, que, bajo H_1, se distribuye asintóticamente según una χ_1^2. Análogamente, bajo la hipótesis $H_2 : \sigma_\mu^2 = 0$ se obtiene el estadístico LM_2, cuya distribución asintótica bajo H_2 es también una χ_1^2.

2.3 Test de especificación de Hausman

A la hora de elegir el método de estimación de un modelo de componentes de error de un solo factor (modelo (2.1)), juega un papel importante la existencia de correlación entre los regresores y los términos de error, y resulta arriesgado suponer que tal correlación no existe, es decir, que $E[\varepsilon_{it} \mid X_{it}] = 0$, puesto que $\varepsilon_{it} = \alpha_i + \nu_{it}$ contiene el efecto fijo inobservado, que puede estar correlacionado con los regresores X_{it} (y de hecho suele estarlo), y por tanto conducir a estimadores inconsistentes. Cuando $E[\alpha_i \mid X_{it}] \neq 0$, sólo el estimador por EF es consistente, mientras que, bajo $H_0 : E[\alpha_i \mid X_{it}] = 0$, tanto MCO como EF y MCG son consistentes, siendo MCG el estimador lineal insesgado óptimo.

Un *contraste de Hausman* se utiliza para analizar la posible correlación entre los α_i y los regresores y poder así decidir entre una estimación por EF o por RE. Bajo $H_0 : E[\alpha_i \mid x_{it}] = 0$, el estadístico de Hausman, converge en distribución a una χ^2_{NT}:

$$Q_{FE,RE} = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'(\hat{\sigma}^2_{\hat{\beta}_{FE}} - \hat{\sigma}^2_{\hat{\beta}_{RE}})^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \overset{H_0}{:} \chi^2_{NT} \qquad (2.7)$$

Como se puede observar, $Q_{FE,RE}$ es el cociente del cuadrado de la diferencia entre los dos estimadores y la diferencia entre las varianzas de éstos. Así, bajo H_0, tanto EF como MCG son consistentes y por tanto deben tender al mismo valor cuando $NT \to \infty$, de modo que la diferencia entre los estimadores debe ser pequeña. Puesto que el estimador $\hat{\beta}_{RE}$ es más eficiente que $\hat{\beta}_{FE}$, la varianza de aquél es pequeña en comparación con la de éste y por tanto la diferencia entre las varianzas es grande. La combinación de ambas cosas dará como resultado un valor del estadístico $Q_{FE,RE}$ cercano a 0 y que por tanto haya que rechazar la hipótesis nula. Si, por el contrario, H_0 no es cierta, entonces $\hat{\beta}_{FE}$ es consistente pero $\hat{\beta}_{RE}$ no lo es, con lo que debe haber diferencia notable entre los valores de estos estimadores. Esto implicará que el valor del estadístico $Q_{FE,RE}$ será alto, pudiendo así rechazar la hipótesis nula (Greene (1998)).

Hausman y Taylor (1981) demostraron que la misma hipótesis puede ser contrastada utilizando cualquier par de diferencias $\hat{\beta}_{MCG} - \hat{\beta}_{EF}$, $\hat{\beta}_{MCG} - \hat{\beta}_{EF}$ o $\hat{\beta}_{EF} - \hat{\beta}_{BE}$ (siendo $\hat{\beta}_{BE}$ el estimador de β mediante una estimación *entre grupos o between groups*) intercambiando estas diferencias y sus varianzas en (2.7), dado que los estadísticos que se obtienen difieren unos de otros en una matriz no singular.

Para el modelo de dos factores, el test de Hausman se basa igualmente en la diferencia entre el estimador de efectos aleatorios por MCG y el estimador EF (con variables dummy individuales y de tiempo), sólo que la equivalencia de los contrastes intercambiando los estimadores $\hat{\beta}_{MCG}, \hat{\beta}_{EF}, \hat{\beta}_{BE}$ no se mantiene en este caso, aunque otro tipo de equivalencias han sido establecidas (véase Baltagi (2001)).

2.5 Contrastes de agrupabilidad de los datos

Otra cuestión que se plantea en el análisis de datos de panel es si agrupar o no los datos, es decir, si plantear un modelo como en (2.1), con todos los β iguales para todos los grupos y para todos los periodos (modelo restringido), o si incluir un parámetro diferente para cada individuo o para cada periodo de tiempo (modelo no restringido), de modo que se tendría una ecuación de regresión para cada i, $y_{it} = X'_{it}\beta_i + \varepsilon_{it}$, o una para cada t, $y_{it} = X'_{it}\beta_t + \varepsilon_{it}$.

En el primero de los casos (sobre si agrupar o no los datos de corte transversal), se trata de contrastar la hipótesis: $H_0 : \beta_i = \beta \ \forall i = 1, 2, \ldots, N$.

Si $\varepsilon : N(0, \sigma^2 I)$, se puede construir un estadístico:

$$F = \frac{(e'e - e_1'e_1 - e_2'e_2 - \ldots - e_N'e_N)/(N-1)(K+1)}{(e_1'e_1 + e_2'e_2 + \ldots + e_N'e_N)/N(T-K-1)} :^{H_0} F_{(N-1)(K+1),\, N(T-K-1)},$$

donde $e = (I_{NT} - X(X'X)^{-1}X')y$ y $e_i = (I_T - X_i(X_i'X_i)^{-1}X_i')y_i$ para cada i, que es precisamente un test de Chow extendido al caso de N regresiones lineales.

Cuando $\varepsilon : N(0, \Omega)$, el estadístico de Chow no sigue una distribución F y no es correcto utilizar este test. En este caso, si se puede escribir $\Omega = \sigma^2\Sigma$, basta con premultiplicar por $\Sigma^{-1/2}$ las variables en el modelo (2.1) y aplicar al modelo transformado el test de Chow anterior[3]. El estadístico F para el test de Chow en el modelo transformado contiene en su expresión la matriz Σ, de modo que cuando ésta no es observable se deberá utilizar un estimador consistente de Σ. El estadístico de contraste sigue también una distribución $F_{(N-1)(K+1),\, N(T-K-1)}$.[4]

[3] Esto puede hacerse porque las perturbaciones transformadas, $\Sigma^{-1/2}\varepsilon$, tienen varianzas homoscedásticas.

[4] Véase Baltagi (2001) para un estudio detallado de estos y referencias a otros contrastes de agrupación.

3. CONTRASTES EN MODELOS DINÁMICOS

Las situaciones en las que las variables presentan un carácter autorregresivo se estudian a través de modelos dinámicos, caracterizados por presentar entre los regresores la variable dependiente retardada:

$$y_{it} = \delta y_{i,t-1} + X'_{it}\beta + \varepsilon_{it} \qquad (3.1)$$

En estos modelos, las estimaciones por MCO, EF o MCG no son consistentes debido a la correlación de los regresores con el término de error. Para solventar este problema, Anderson y Hsiao (1981) sugieren tomar primeras diferencias en el modelo dinámico:

$$\Delta y_{it} = \delta\, \Delta y_{i,t-1} + \Delta X'_{it}\beta + \Delta\varepsilon_{it} \qquad (3.2)$$

y realizar una estimación por variables instrumentales (IV) con $y_{i,t-2}$ como instrumento, dado que $y_{i,t-2}$ está correlacionada con $\Delta y_{i,t-1}$, pero no lo está con el error $\Delta\varepsilon_{it} = \Delta\nu_{it}$, siempre que no exista correlación serial entre los errores. Para mejorar la eficiencia del estimador, Arellano y Bond (1991) sugieren utilizar todos los instrumentos disponibles: $\Delta y_{i,t-k}$, con $k > 1$, y una estimación por GMM, también bajo la hipótesis de no correlación serial de las perturbaciones. Este contexto da lugar a los siguientes contrastes.

3.1 Contrastes de correlación serial de las perturbaciones

Es importante estudiar la existencia de correlación serial en las perturbaciones, dado que ésta provoca que un estimador por IV con variables retardadas como instrumentos no sea consistente. Arellano y Bond (1991) sugieren tres contrastes con la ausencia de correlación serial en los errores como hipótesis nula.

Un primer test trabaja bajo la hipótesis de no correlación serial de segundo orden en las perturbaciones del modelo en diferencias: $E[\Delta\nu_{it}\Delta\nu_{i,t-2}] = 0$. Bajo H_0, el estadístico m_2 construido por Arellano y Bond (1991) se distribuye asintóticamente según una $N(0,1)$.[5] Este estadístico es flexible en el sentido de que puede estar basado en cualquier estimador consistente GMM no necesariamente eficiente, aunque la potencia del test m_2 sí depende de la eficiencia del estimador utilizado.

[5] Puede consultarse la expresión del estadístico en Arellano y Bond (1991).

Existen situaciones en las que el estadístico m_2 no está definido. Para ese caso Arellano y Bond (1991) sugieren un test de Sargan de sobre-identificación de restricciones cuyo estadístico de contraste es:

$$s = \Delta\hat{v}' Z (\sum_{i=1}^{N} Z_i' \Delta\hat{v}_i \Delta\hat{v}_i')^{-1} Z' \Delta\hat{v} \;:\; \chi^2_{p-K-1} \;,$$

donde $\Delta\hat{v}$ denota los residuos de la estimación GMM en dos pasos propuesta por Arellano y Bond (1991), $Z = [Z_1', Z_2', \ldots, Z_N']'$ es la matriz de instrumentos, es decir, Z_i es una matriz diagonal por bloques cuyo bloque en la posición s es $(y_{i1}, \ldots, y_{is})$, y p es el número de columnas de Z.[6]

Una tercera opción propuesta por los mismos autores consiste en construir un test de Hausman basado en la diferencia de los estimadores de $\hat{\delta}_1 - \hat{\delta}$, donde $\hat{\delta}_1$ es el estimador GMM en dos etapas suponiendo que las perturbaciones en niveles siguen un proceso e medias móviles MA(1) y $\hat{\delta}$ suponiendo que son MA(0). El estadístico de contraste es:

$$h = (\hat{\delta}_1 - \hat{\delta})'(a\,\mathrm{v\hat{a}r}_{\hat{\delta}_1} - a\,\mathrm{v\hat{a}r}_{\hat{\delta}})^{-}(\hat{\delta}_1 - \hat{\delta}) \;:\; \chi^2_r \;,$$

donde $a\,\mathrm{v\hat{a}r}_{\hat{\delta}_1}$ y $a\,\mathrm{v\hat{a}r}_{\hat{\delta}}$ son estimadores de las varianzas asintóticas de $\hat{\delta}_1$ y de $\hat{\delta}$ respectivamente, r es el rango de la varianza asintótica de $\hat{\delta}_1 - \hat{\delta}$ y $()^{-}$ indica la inversa generalizada.

3.2 Contraste para efectos individuales en modelos autorregresivos

Es posible contrastar la existencia de efectos fijos haciendo uso de restricciones de sobre-identificación del modelo dinámico. Considerando un modelo dinámico sin regresores, $y_{it} = \delta y_{i,t-1} + \varepsilon_{it}$, y suponiendo T=3 (para facilitar la explicación), esta ecuación puede ser estimada utilizando los dos últimos periodos. Bajo la hipótesis $H_0 : \alpha_i = 0$, se tienen las condiciones: $E[y_{i1}\varepsilon_{i2}] = 0$, $E[y_{i1}\varepsilon_{i3}] = 0$, $E[y_{i2}\varepsilon_{i3}] = 0$, es decir, tres condiciones para determinar un único parámetro. Las dos condiciones "sobrantes" pueden ser utilizadas para contrastar la existencia de efectos fijos. Si reescribimos las condiciones anteriores de la forma:

[6] Ver Arellano y Bond (1991) para una construcción más detallada del estadístico de Sargan.

$$E[y_{i1}(\varepsilon_{i3}-\varepsilon_{i3})]=0 \qquad (3.3)$$
$$E[y_{i1}\varepsilon_{i2}]=0 \qquad (3.4)$$
$$E[y_{i2}\varepsilon_{i3}]=0 \qquad (3.5)$$

la primera de éstas puede ser utilizada para identificar δ incluso bajo la presencia de efectos fijos, mientras que la hipótesis de no existencia de efectos fijos impone sólo las dos condiciones adicionales (3.4) y (3.5).

Se puede entonces reestablecer el sistema inicial en los términos siguientes:

$$y_{i3}-y_{i2}=\delta(y_{i2}-y_{i1})+(\varepsilon_{i3}-\varepsilon_{i2})$$
$$y_{i3}=\delta y_{i2}+\varepsilon_{i3}$$
$$y_{i2}=\delta y_{i1}+\varepsilon_{i2}$$

y en forma vectorial:

$$y^{*}=Y^{*}\delta+\varepsilon^{*} \qquad (3.6)$$

donde $y^{*}=(y_3'-y_2',y_3',y_2')'$, $Y^{*}=(y_2'-y_1',y_2',y_1')'$ y $\varepsilon^{*}=(\varepsilon_3'-\varepsilon_2',\varepsilon_3',\varepsilon_2')'$. Generalizando al caso de cualquier T, estas expresiones serían: $y^{*}=(\Delta y_T',y_T',y_{T-1}',\ldots,y_2')'$, $Y^{*}=(\Delta y_{T-1}',y_{T-1}',\ldots,y_1')'$ y $\varepsilon^{*}=(\Delta\varepsilon_T',\varepsilon_T',\varepsilon_{T-1}',\ldots,\varepsilon_2')'$.

Hotlz y Eakin (1988) estiman este sistema de ecuaciones simultáneas con variables instrumentales (diferentes instrumentos para cada ecuación). Llamando $Z=diag(Z_i)$ a la matriz de instrumentos y premultiplicando (3.6) por Z se obtiene por MCG un estimador $\hat{\delta}$ de δ, y con él, el estadístico que permite contrastar la existencia de efectos fijos en el modelo dinámico:

$$SSQ=(y^{*}-Y^{*}\hat{\delta})'Z\hat{\Omega}^{-1}Z'(y^{*}-Y^{*}\hat{\delta})/N \sim \chi_r^2,$$

donde r es el número de restricciones de sobre-identificación.

4. APLICACIÓN EMPÍRICA

Algunos de los tests anteriores han sido aplicados en el estudio del impacto sobre la tasa de mortalidad infantil de la inversión pública federal en México en actividades productivas y en el sector social. Partiendo del modelo $y_{it}=X_{it}'\beta+\varepsilon_{it}$, donde las variables explicativas son la inversión per cápita en actividades productivas y la inversión social, y la variable dependiente es la tasa de mortalidad infantil. La tabla 1 muestra los cálculos realizados para el test de Chow contrastando la hipótesis nula de que los coeficientes β son los mismos para las regiones de la zona centro-norte que para las regiones de la zona centro-sur.

	Inv.prod.	Std.err.	**Inv. Soc.**	Std.err.
$\mathbf{b_N}$	0.072	(0.484)	-0.183*	(0.089)
$\mathbf{b_S}$	0.000	(0.046)	-0.142*	(0.070)

N	**T**	**K**	$\mathbf{RSS_U}$	$\mathbf{RSS_R}$	$\mathbf{n_1}$	$\mathbf{n_2}$
32	4	3	1.8318	1.8837	93	32
		$F_0 =$ 0.0877141		prob >F = 1		

Tabla 1. Test de Chow para la igualdad de coeficientes zonas norte y sur

Los resultados muestran que no se puede rechazar la hipótesis de que los coeficientes sean iguales. De hecho, puede observarse que los coeficientes obtenidos son muy similares para las dos zonas.

Para analizar la posible correlación de los efectos fijos con los regresores se ha utilizado un test de Hausman, cuyos resultados se muestran en la tabla 2.

	EF	Std.err.	**RE**	Std.err.
Inv.prod.	0.037	(0.031)	0.022	(0.031)
Inv.soc.	-0.161***	(0.055)	-0.149***	(0.052)
	chi2(5) = 3.46		Prob>chi2 = 0.629	

Tabla2. Test de Hausman para efectos fijos o aleatorios

Puede observarse que en este caso los coeficientes de las estimaciones por EF y por RE son muy similares. De hecho, el test no puede rechazar la hipótesis nula de no correlación entre los efectos fijos y los regresores.

Finalmente, considerando una especificación dinámica, $y_{it} = \delta y_{i,t-1} + X'_{it}\beta + \varepsilon_{it}$, los tests de Arellano y Bond (1991) permiten contrastar la hipótesis de no autocorrelación serial de segundo orden en las perturbaciones. Los resultados del test de sargan y del estadístico m_2 se muestran en la tabla 3.

Test de Sargan de restricciones de sobre-identificación
chi2(5) = 0.72 Prob > chi2 = 0.9820
Test de Arellano y Bond de no autocorrelación serial de segundo orden
z = -0.06 Prob > z = 0.9554

Tabla 3. Contrastes de no autocorrelación

Ambos contrastes muestran que no se puede rechazar la hipótesis nula de que los errores no están serialmente correlacionados, lo cual da soporte a una estimación por IV del modelo dinámico en diferencias.

5. CONCLUSIONES

En este trabajo se realiza una recopilación de diversos contrastes de hipótesis existentes en la literatura de datos de panel, los cuales constituyen una herramienta útil en la elección de la especificación de modelos de datos de panel, así como en la elección del método de estimación a emplear. Estos contrastes aportan información acerca de si la situación de estudio verifica o no los supuestos necesarios para los métodos de estimación. Varios de estos contrastes han sido planteados en el estudio del impacto de la inversión pública federal en México. Encontramos que no se pueden rechazar las hipótesis de igualdad de coeficientes para las zonas Norte y Sur, de no correlación de los efectos fijos con los regresores (especificación estática), y de no autoccorrelación serial de las perturbaciones para poder llevar a cabo una estimación por IV en el caso dinámico.

REFERENCIAS BIBLIOGRÁFICAS

- ANDERSON, T.W y C. HSIAO (1981): "Estimation of dynamic models with error components". *Journal of the American Statistical Association*, 76, 598-606.
- ARELLANO, M. y S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo. Evidence and an Application to Employment Equations". *Review of Economic Studies*, Vol. 58, pp. 277-297.
- ARELLANO, M. (2003): *Panel Data Econometrics (Advanced Texts in Econometrics).* Oxford Press.
- BALTAGI, B. (2001): *Econometric Analysis of Panel Data.* 2nd Edition. Wiley.
- BREUSCH, T. y A. PAGAN (1980): "The Lagrange Multiplier test and its applications to model specification in econometrics". *Review of Economic Studies*, 47, 239-253.
- GREENE, W.H. (1998). *Análisis Econométrico*. Tercera edición. Prentice Hall.
- HAUSMAN, J.A. y W.E. TAYLOR (1981): "Panel data and unobservable individual effects". *Econometrica*, 49, 1377-1398.
- HOLTZ-EAKIN, D. (1988): "Testing for individual effects in autoregressive models". *Journal of Econometrics*, 39, 297-307.