¿Cómo citar?: Montero. R (2005): *Test de Hausman.* Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España

Test de Hausman

Roberto Montero Granados
Universidad de Granada

22 de septiembre de 2005

1. Introducción

El test propuesto por Hausman (1978) es un test chi cuadrado que determina si las diferencias son sistemáticas y significativas entre dos estimaciones.

Se emplea fundamentalmente para dos cosas:

a) saber si un estimador es consistente.
b) saber si una variables es o no relevante.

2. Test de Consistencia

Supongamos que disponemos de dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ y sabemos además que uno de ellos, $\hat{\theta}_2$ es el más eficiente (tiene menor varianza). El test calcula con una formulación especial (más abajo), que sigue una chi cuadrado, las diferencias en las estimaciones comunes a ambos modelos. Si las diferencias, aunque sean altas, no son sistemáticas (no tienen un sesgo definido), entonces ambos estimadores son consistentes (la estimación muestral tiende al parámetro poblacional) y nos quedaremos con el más eficiente $\hat{\theta}_2$. Si las diferencias son sistemáticas entonces nuestra hipótesis no se cumple, ambos no son consistentes y ahora tenemos un dilema: pensar que el modelo está mal especificado en ambos casos o quedarnos con el estimador consistente, que es $\hat{\theta}_1$

Si el valor de la prueba es alto (p.e. p-valor menor de 0.05) la hipótesis de diferencias no sistemáticas se rechaza, por lo que: o se reelabora el modelo o se elige al que se considera consistente en cualquier caso $\hat{\theta}_1$

Si el valor de la prueba es bajo (p.e. p-valor mayor de 0.05) la hipótesis nula, de diferencias no sistemáticas, se cumple y podemos elegir cualquiera de los dos estimadores, normalmente el que suponemos más eficiente, $\hat{\theta}_2$

Esta prueba se puede realizar con cualesquiera dos modelos de regresión que queramos comparar. $\hat{\theta}_1$ será el modelo del que estemos más seguros, que suponemos consistente en cualquier caso y $\hat{\theta}_2$ será el modelo que queremos testar, que es más eficiente pero no estamos seguros de que sea consistente. Si los coeficientes de ambos modelos no tienen errores sistemáticos podremos quedarnos con $\hat{\theta}_2$, si, por el contrario, aparecen errores sistemáticos entonces $\hat{\theta}_2$ no es consistente y debemos quedarnos con $\hat{\theta}_1$. Por ejemplo esta prueba se puede realizar para saber si es mejor el estimador de efectos fijos o variables en una base de datos de panel. Para ello se estima el modelo de efectos fijos ($\hat{\theta}_1$) y el de efectos variables ($\hat{\theta}_2$) si no existen diferencias o sesgo significativo (p-valor alto) nos quedamos con el de efectos variables, más eficiente, pero si se detectan diferencias sistemáticas (p-valor bajo) debemos quedarnos con el de efectos fijos, que hemos supuesto siembre consistente.

Es importante hacer notar que estamos suponiendo que un modelo es siempre consistente ($\hat{\theta}_1$) y que, en caso de igualdad en las estimaciones, otro es el más eficiente ($\hat{\theta}_2$), estas suposiciones son difíciles de contrastar y, a menudo, se incumplen. Pero esa es otra historia.

3. Test de independencia o irrelevancia (IIA)

Igual que en el caso anterior, el test compara las estimaciones de dos modelos de regresión. En uno de los cuales se ha omitido una variable. Si la diferencia entre el resto de parámetros es sistemáticamente significativa, podemos suponer que el parámetro omitido es relevante. Es decir si el p-valor que resulta del test es alto podemos asumir que las diferencias entre ambos modelos no son sistemáticas y que, por tanto la variable omitida es irrelevante. Por el contrario, si el p-valor es bajo entonces la hipótesis de igualdad se rechaza y, por tanto, la variable o variables omitidas sí que eran relevantes.

4. Fórmula

$$H = (\beta_c - \beta_e)'(V_c - V_e)^{-1}(\beta_c - \beta_e), \quad H \sim X_n^2$$

donde

β_c es el vector de estimaciones del estimador consistente $\hat{\theta}_2$.

β_e es el vector de estimaciones del estimador eficiente $\hat{\theta}_1$.

V_c es la matriz de covarianzas del estimador consistente.

V_e es la matriz de covarianzas del estimador eficiente.

n son los grados de libertad de la χ_n^2 (número de variables incluida la constante, en su caso)

El test de Hausman está implementado en Stata como ado_file

rutina:

xtreg *vardep var_indep*, fe
estimates store *name_consistent*

xtreg *var_dep var_indep*, re

hausman *name_consistent* .

Si p valor < 0.05 se rechaza la hipótesis nula de igualdad al 95% de confianza y se deben asumir las estimaciones de efectos fijos. Por el mismo criterio, si p valor <0.05 se rechaza la hipótesis nula de igualdad al 95% de confianza y se debe rechazar la hipótesis de independencia o irrelevancia de las variables.

Por el contrario, si p-valor > 0.05 se debe admitir la hipótesis nula de igualdad de estimaciones y entonces el estimador mas eficiente, el del efectos variables, debe ser seleccionado. Igualmente, si el p-valor >0.05 debe asumirse con el 95% de confianza, que la variable introducida en el modelo de contraste no es irrelevante.

En ocasiones, cuando en la muestra hay pocos individuos (menos de 50 o 60) el resultado se la prueba, es decir el valor de la Chi2, puede arrojar un número negativo (lo cual es imposible) pero que a los efectos de la prueba se debe interpretar como una fuerte evidencia de que no puede rechazarse la hipótesis nula (stata reference manual AJ, 447)

5. Bibliografía

Hausman, J.A. (1978): "Specification test in econometrics". *Econometrica*. 46: 1251-1271.
Hausman, J. and McFadden, C. (1984): "Specification test in econometrics", *Econometrica*, 52, 1219-1240.
Stata (2005) *Reference manual A-J.* Stata Pres. Texas, 441-448.