SOLUCIONES EXAMEN FINAL EXTRAORDINARIO DE ECONOMETRÍA

1. (a) Este es un contraste del modelo completo con esta hipótesis nula

$$H_0 : \beta_1 = \cdots = \beta_5 = 0$$

en contra de la alternativa de que algún coeficiente es diferente de cero. Para ello se debe realizar un contraste conjunto de todos los coeficientes mediante un estadístico F. Dada la información sólo se puede calcular un F no robusto a la heterocedasticidad, con modelo restringido igual a la constante con $R_r^2 = 0$ y R_{nr}^2 dado por la regresión no restringida,

$$\begin{aligned} F &= \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2} \frac{n-k-1}{q} = \frac{R_{nr}^2}{1 - R_{nr}^2} \frac{n-k-1}{q} \\ &= \frac{0.1085580 - 0}{1 - 0.1085580} \frac{27326 - 5 - 1}{5} = 665.39, \end{aligned}$$

que es muy significativo comparado con el valor crítico de una $\chi_5^2/5$, $F = 665.39 > 11.07/5$ al 5%, y por tanto podemos rechazar la hipótesis nula, es decir el modelo es significativo. Este contraste sólo es válido bajo homocedasticidad (junto con el resto de hipótesis del modelo clásico), pero no en presencia de heterocedasticidad.

(b) En este caso la hipótesis nula es

$$H_0 : \beta_4 = 0$$

y la alternativa

$$H_1 : \beta_4 \neq 0.$$

Usando el ajuste del primer modelo se puede calcular directamente un estadístico t robusto

$$t_4 = \frac{\hat{\beta}_4}{e.e. \left(\hat{\beta}_4\right)} = \frac{0.00618235}{0.00207600} = 2.978$$

que con p-valor = 0.0029 es muy significativo a los niveles habituales al compararlo con una normal estándar, rechazando la hipótesis de igual determinación de la salud entre hombres y mujeres.

Alternativamente se puede calcular un estadístico F no robusto donde la segunda regresión estima la versión restringida del modelo original:

$$\begin{aligned} F &= \frac{SCR_r - SCR_{nr}}{SCR_{nr}} \frac{n-k-1}{q} \\ &= \frac{762.5888 - 762.3413}{762.3413} \frac{27326 - 5 - 1}{1} = 8.869\,6, \end{aligned}$$

o alternativamente,

$$\begin{aligned} F &= \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2} \frac{n-k-1}{q} \\ &= \frac{0.1085580 - 0.1082687}{1 - 0.1085580} \frac{27326 - 5 - 1}{1} = 8.866\,2, \end{aligned}$$

que es significativo comparado con el valor crítico de una $\chi_1^2/1$, $F = 8.8 > 3.84$, al 5%, y por tanto rechazamos la hipótesis nula.

(c) El efecto marginal requerido es

$$\gamma = \frac{\partial \mathbb{E}\left(SAH|\, X\right)}{\partial Age} = \beta_1 + 2\beta_2 Age = \beta_1 + 2\beta_2 \cdot 30$$

que se estima como

$$\begin{aligned}\hat{\gamma} &= \hat{\beta}_1 + 2\hat{\beta}_2 \cdot 30 \\ &= 0.04424800 + 60 \cdot (-0.00085563) = -7.0898 \times 10^{-3}.\end{aligned}$$

A su vez tenemos que el error estándar de este efecto se calcula como

$$\begin{aligned} e.e.\left(\hat{\gamma}\right) &= \left(e.e.\left(\hat{\beta}_1\right)^2 + 60^2 \cdot e.e.\left(\hat{\beta}_2\right)^2 + 2 \cdot 60 \cdot \widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)\right)^{1/2} \\ &= \left(0.00391106^2 + 60^2 \cdot 0.00014667^2 + 2 \cdot 60 \cdot \left(-2.8 \times 10^{-7}\right)\right)^{1/2} \\ &= 7.6902 \times 10^{-3}\end{aligned}$$

y el correspondiente intervalo de confianza al 95% para el efecto marginal γ es

$$-7.0898 \times 10^{-3} \pm 1.96 \cdot 7.6902 \times 10^{-3} = \left(-2.2163 \times 10^{-2}, 7.9830 \times 10^{-3}\right).$$

(d) El máximo de la función de regresión en función de la edad se alcanza en el punto crítico Age^* que satisface

$$\frac{\partial \mathbb{E}\left(SAH|\, Age^*\right)}{\partial Age} = \beta_1 + 2\beta_2 Age^* = 0,$$

es decir

$$Age^* = -\frac{\beta_1}{2\beta_2}$$

que se estima como

$$\widehat{Age}^* = -\frac{\hat{\beta}_1}{2\hat{\beta}_2} = -\frac{0.04424800}{2 \cdot (-0.00085563)} = 25.857.$$

El contraste de las hipótesis

$$\begin{aligned} H_0 &: Age^* = 20 \\ H_1 &: Age^* \geq 20\end{aligned}$$

es equivalente a

$$\begin{aligned} H_0 &: \beta_1 + 40\beta_2 = 0 \\ H_1 &: \beta_1 + 40\beta_2 \leq 0\end{aligned}$$

suponiendo que $\beta_2 < 0$, el cuál se puede realizar con un contraste de la t basado en el estadístico

$$t = \frac{\hat{\beta}_1 + 40\hat{\beta}_2}{e.e.\left(\hat{\beta}_1 + 40\hat{\beta}_2\right)}$$

que requiere conocer la covarianza estimada entre $\hat{\beta}_1$ y $\hat{\beta}_2$ o la estimación de un modelo reparametrizado:

$$\begin{aligned} e.e.\left(\hat{\beta}_1 + 40\hat{\beta}_2\right) &= \left(e.e.\left(\hat{\beta}_1\right)^2 + 40^2 \cdot e.e.\left(\hat{\beta}_2\right)^2 + 2 \cdot 40 \cdot \widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)\right)^{1/2} \\ &= \left(0.00391106^2 + 40^2 \cdot 0.00014667^2 + 2 \cdot 40 \cdot \left(-2.8 \times 10^{-7}\right)\right)^{1/2} \\ &= 5.2264 \times 10^{-3}\end{aligned}$$

por lo que con $\hat{\beta}_1 + 40\hat{\beta}_2 = 0.04424800 + 40 \cdot (-0.00085563) = 1.002\,3 \times 10^{-2}$

$$t = \frac{1.002\,3 \times 10^{-2}}{5.226\,4 \times 10^{-3}} = 1.917\,7.$$

La conclusión del contraste unilateral se basa en la comparación con el valor crítico correspondiente de la normal estándard de tal forma que se rechaza H_0 en favor de H_1 si $t < -1.645 = -z_\alpha$ al 5% de significación: en este caso $t > 0$, por lo que no se rechaza la hipótesis nula de *que* el punto crítico es 20 en favor de que es más pequeño (la evidencia empírica indica que es mayor).

(e) El modelo conjunto para las dos regresiones es

$$\begin{aligned} SAH &= \beta_0 + \beta_1 Age + \beta_2 Age^2 + \beta_3 Married + \beta_4 Female + \beta_5 Hhkids \\ &\quad + \delta_1 Age * Female + \delta_2 Age^2 * Female + \delta_3 Married * Female + \delta_5 Hhkids * Female + v \end{aligned}$$

donde $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_5)$ son los coeficientes de la regresión para los hombres y $(\beta_0 + \beta_4, \beta_1 + \delta_1, \beta_2 + \delta_2, \beta_3 + \delta_3, \beta_5 + \delta_5)$ corresponden a los coeficientes para las mujeres, es decir

$\hat{\beta}_0 =$	0,01085465
$\hat{\beta}_1 =$	0,03082032
$\hat{\beta}_2 =$	-0,00033408
$\hat{\beta}_3 =$	0,05491566
$\hat{\beta}_5 =$	-0,01782500

$\hat{\beta}_4 = (\widehat{\beta_0 + \beta_4}) - \hat{\beta}_0 =$	-0,20116524	$-\hat{\beta}_0 =$	-0,20116524	$-$	0,01085465
$\hat{\delta}_1 = (\widehat{\beta_1 + \delta_1}) - \hat{\beta}_1 =$	0,05933135	$-\hat{\beta}_1 =$	0,05933135	$-$	0,03082032
$\hat{\delta}_2 = (\widehat{\beta_2 + \delta_2}) - \hat{\beta}_2 =$	-0,00147495	$-\hat{\beta}_2 =$	-0,00147495	$-$	(-0,00033408)
$\hat{\delta}_3 = (\widehat{\beta_3 + \delta_3}) - \hat{\beta}_3 =$	0,11519095	$-\hat{\beta}_3 =$	0,11519095	$-$	0,05491566
$\hat{\delta}_5 = (\widehat{\beta_5 + \delta_5}) - \hat{\beta}_5 =$	-0,01321821	$-\hat{\beta}_5 =$	-0,01321821	$-$	(-0,01782500)

Dado que estimar las dos regresiones de hombres y mujeres por separado es equivalente a estimar el modelo conjunto (o no restringido) con los 10 coeficientes, sus residuos deben ser los mismos y por tanto

$$SCR_{norestringido} = SCR_{male} + SCR_{female} = 384.3125 + 372.6560 = 756.97.$$

La SCR$_{restringida}$ = 762.5888 se obtiene de la regresión 2, donde no se incluye ninguna distinción entre hombres y mujeres para la muestra conjunta, y el correspondiente estadístico F para contrastar igualdad de modelos entre las dos subpoblaciones,

$$\begin{aligned} H_0 &: \beta_4 = \delta_1 = \delta_2 = \delta_3 = \delta_5 = 0 \\ H_1 &: H_0 \text{ es falsa} \end{aligned}$$

se calcula como

$$\begin{aligned} F &= \frac{SCR_r - SCR_{nr}}{SCR_{nr}} \frac{n-k-1}{q} \\ &= \frac{762.5888 - 756.97}{756.97} \frac{27326 - 5 - 1}{5} = 40.\,558, \end{aligned}$$

que es muy significativo comparado con una $\chi^2_5/5$, $F = 40.\,558 > 11.07/5$, por lo que se rechaza la hipótesis nula de que los dos modelos para hombres y mujeres son iguales.

2. .

(a) Para contrastar la relevancia de los instrumentos debemos computar el estadístico F correspondiente a la hipótesis nula

$$H_0 : \pi_1 = \pi_2 = \pi_3 = 0$$

(en contra de que algún coeficiente es diferente de cero) en la forma reducida para el regresor endógeno $educ$,

$$educ = \pi_0 + \pi_1 motheduc + \pi_2 fatheduc + \pi_3 nearc4 + \pi_4 exper + \pi_5 smsa + \pi_6 south + v.$$

La estimación del Modelo 3 nos dice que ese estadístico $F = 104{,}402$, que es muy significativo a cualquier nivel y mayor que 10, por lo que concluimos que los instrumentos son relevantes y no débiles.

Para contrastar la exogeneidad, o equivalentemente las $m - k = 3 - 1 = 2$ restricciones de sobreidentificación, debemos computar el estadístico F correspondiente a la hipótesis nula

$$H_0 : \delta_1 = \delta_2 = \delta_3 = 0$$

en la regresión de los residuos 2SLS sobre todas las variables exógenas del modelo

$$\hat{u}^{2SLS} = \delta_0 + \delta_1 motheduc + \delta_2 fatheduc + \delta_3 nearc4 + \delta_4 exper + \delta_5 smsa + \delta_6 south + \varepsilon,$$

y a partir del F de la Salida 2, el estadístico J :

$$J = mF = 3 \cdot 0.461289 = 1.383\,9$$

y compararlo con el valor crítico de una $\chi^2_{m-k} = \chi^2_2$: en este caso $J = 1.3839 < 5.99$ por lo que no se puede rechazar la hipótesis nula de exogeneidad de los instrumentos.

(b) En este caso la estrategia para contrastar la relevancia del único instrumento se basará en computar el estadístico F correspondiente a la hipótesis nula

$$H_0 : \pi_1 = 0$$

en la forma reducida para el regresor endógeno $educ$, incluyendo solo $motheduc$, pero no $fatheduc$ o $nearc4$:

$$educ = \pi_0 + \pi_1 motheduc + \pi_4 exper + \pi_5 smsa + \pi_6 south + v.$$

La estimación MCO de este modelo aparece en la estimación del Modelo 5, donde el instrumento $motheduc$ es muy significativo, $F = t^2 = 16.4907^2$, concluyendo que el instrumento es relevante y muy fuerte (como también se puede concluir a partir de los resultados en (a)).

En este caso no podemos contrastar la exogeneidad individual de $motheduc$ a partir de la estimación 2SLS de este modelo, porque estaría exactamente identificado con un sólo instrumento, pero usando las conclusiones del apartado anterior, (a), no se pudo rechazar la exogeneidad de los tres instrumentos disponibles, por lo que podemos concluir que $motheduc$ es exógena.

La conclusión es que esta estimación es también válida, pero debería ser menos eficiente que la del apartado (a), como se puede comprobar en los errores estándar del coeficiente de $educ$ en las salidas del Modelo 1 y el 4, además de proporcionar menos información (sobre exogeneidad).

(c) En este caso para contrastar la relevancia de los tres instrumentos debemos computar el estadístico F correspondiente a la hipótesis nula

$$H_0 : \pi_1 = \pi_2 = \pi_3 = 0$$

en la forma reducida para el regresor endógeno *educ*, pero sin tener en cuenta los regresores exógenos *smsa* y *south*, que no aparecen en la forma estructural:

$$educ = \pi_0 + \pi_1 motheduc + \pi_2 fatheduc + \pi_3 nearc4 + \pi_4 exper + v.$$

No disponemos de la estimación de este modelo, pero a partir de las conclusiones del apartado (a), el modelo 3 nos dice los tres instrumentos eran muy significativos en un modelo con más regresores, por lo que podemos concluir que los instrumentos son relevantes y no débiles en este caso también.

(d) En el apartado anterior hemos concluido que no debería haber problemas de relevancia de los instrumentos, pero además hay que contrastar su exogeneidad. Como en (a), para contrastar la exogeneidad, o las $m-k=3-1=2$ restricciones de sobreidenticación, debemos computar el estadístico F correspondiente a la hipótesis nula

$$H_0 : \delta_1 = \delta_2 = \delta_3 = 0$$

en la regresión de los residuos 2SLS sobre todas las variables exógenas del modelo

$$\hat{u}^{2SLS} = \delta_0 + \delta_1 motheduc + \delta_2 fatheduc + \delta_3 nearc4 + \delta_4 exper + res,$$

en este caso omitiendo *south* y *smsa*, y a partir del estadístico F aportado en el Modelo 7 calcular el estadístico J :

$$J = mF = 3 \cdot 3.16551 = 9.496\,5.$$

Comparado con el valor crítico de una $\chi^2_{m-k} = \chi^2_2$ es muy significativo: $J = 9.496 > 5.99$ al 5%, por lo que se rechaza la hipótesis nula de exogeneidad de los instrumentos: en este caso omitir las variables de control *south* y *smsa* invalida la validez del conjunto de instrumentos (que era válido en el modelo considerado en (a)) porque los factores omitidos ahora pueden estar correlados con los instrumentos, por ejemplo *nearc*4 con *south* o *smsa*.

(e) Si tanto *educ* como *exper* son consideradas endógenas entonces deberemos estimar sendas formas reducidas o primeras etapas para cada uno de esos regresores endógenos a partir de los tres instrumentos disponibles y en este caso sin los regresores exógenos adicionales (ni *exper* o *educ*, respectivamente) y con los valores ajustados para *educ* y *exper* realizar una segunda etapa basada en la estimación MCO de

$$\log(wage) = \beta_0 + \beta_1 \widehat{educ} + \beta_2 \widehat{exper} + res.$$

En este caso la comprobación de la relevancia debería asegurar además que no existe multicolinealidad en este modelo en la población, mientras que el contraste de exogeneidad es posible dado que existe todavía una $(m-k=3-2=1)$ restricción de sobreidentificación (aunque dado el resultado en (d) tendríamos dudas de su exogeneidad).

La razón por la que *exper* puede ser endógena en *(2)* es la misma por la que los instrumentos pueden dejar de ser exógenos en (d): las variables omitidas pueden estar correladas con *exper*, que está incluido en V, aunque U no lo estuviese.