

# Taller 1: Análisis de Demanda y Endogeneidad

## Tópicos en Econometría

Juan Diego Moreno Sánchez

2026-02-15

### Contenido

1	EJERCICIO 1 – Responda, Defina y Explique	1
2	EJERCICIO 2 - Estructura de Costos en un Mercado Competitivo	3
3	EJERCICIO 3 – Modelo de Regresión Múltiple (Test de Chow y Variables Dummy)	9
4	EJERCICIO 4 – Regresión con Variables Instrumentales (Endogeneidad)	14

### 1 EJERCICIO 1 – Responda, Defina y Explique

#### 1.1 a) ¿Cuáles supuestos debe cumplir el modelo de regresión para obtener estimadores MELI para los parámetros estimados por MCO?

Para que el modelo de regresión esté bien especificado y los estimadores por MCO tengan buenas propiedades (y bajo normalidad coincidan con los de máxima verosimilitud), se suelen asumir varios supuestos:

1. **Linealidad:** la relación entre la variable dependiente y las explicativas debe ser lineal en los parámetros.
2. **Independencia:** las observaciones deben ser independientes entre sí.
3. **Homoscedasticidad:** la varianza del término de error debe ser constante para todas las observaciones.
4. **Normalidad:** los errores deberían distribuirse normalmente (esto es más fuerte y no siempre es estrictamente necesario).

5. **No multicolinealidad perfecta:** las variables explicativas no deben estar perfectamente correlacionadas.
6. **No autocorrelación:** los errores no deben estar correlacionados entre ellos.
7. **Exogeneidad:** las variables explicativas no deben estar correlacionadas con el error.

### 1.2 b) ¿Cuáles son las propiedades de los parámetros estimados por MCO?

Los estimadores por MCO tienen varias propiedades importantes:

1. **Insesgadez:** en promedio, el estimador coincide con el verdadero parámetro.
2. **Consistencia:** cuando el tamaño de muestra crece, el estimador se aproxima al valor real.
3. **Eficiencia:** dentro de los estimadores lineales e insesgados, MCO tiene la menor varianza.

### 1.3 c) Defina el problema de endogeneidad en un modelo de regresión. ¿Cómo se puede solucionar? ¿Con cuáles métodos?

La endogeneidad aparece cuando alguna variable explicativa está correlacionada con el término de error. Esto genera estimadores sesgados y además inconsistentes. Puede surgir por omitir variables relevantes, por simultaneidad o incluso por errores de medición.

Para corregirla se pueden usar métodos como Variables Instrumentales (VI), Mínimos Cuadrados en Dos Etapas (MC2E), modelos de efectos fijos o incluso técnicas más avanzadas dependiendo del problema.

### 1.4 d) Explique el método de estimación “Máxima Verosimilitud”.

La Máxima Verosimilitud es un método que consiste en elegir los parámetros que maximizan la probabilidad de observar los datos que realmente se tienen. Es decir, se construye una función de verosimilitud y se buscan los valores que la hacen lo más grande posible.

Este método es especialmente útil cuando se tiene una distribución específica para los errores o cuando se trabaja con modelos no lineales. Bajo ciertas condiciones, los estimadores de máxima verosimilitud son consistentes, asintóticamente normales y eficientes.

## 2 EJERCICIO 2 - Estructura de Costos en un Mercado Competitivo

Ustedes han sido contratados por una firma de consultoría nacional para estimar la función de Costos de una industria competitiva en la que participan 30 empresas. Considere la siguiente función de costos totales de la industria:

$$CT_i = \beta_0 + \beta_1 Y_i + \beta_2 Y_i^2 + \beta_3 Y_i^3 + \varepsilon_i$$

Donde  $CT_i$  representa el costo total para la empresa ( $i$ ) medido en millones de pesos y  $Y_i$  representa el nivel de producción de la empresa ( $i$ ) medido en miles de unidades.

Utilizando los datos del archivo “Función\_CostosTotales.xlsx”, usted y su pareja deben elaborar un informe que recoja el análisis econométrico.

### 2.1 a) Estime la función de costos totales de la industria y repórtela en una tabla.

Para resolver este punto vamos a estimar la función de costos totales utilizando regresión lineal. Primero, cargamos los datos desde el archivo Excel:

En este espacio correremos la regresión lineal para estimar los parametros de la función de costos totales.

Resultados de la regresión

=====	
Dependent variable:	
-----	
CT	
-----	
Y	22.548** (9.952)
I(Y2)	-3.340* (1.799)
I(Y3)	0.634*** (0.100)
Constant	185.531*** (16.340)
-----	
Observations	30
R2	0.998
Adjusted R2	0.998
Residual Std. Error	6.537 (df = 26)
F Statistic	4,849.591*** (df = 3; 26)
=====	
Note:	*p<0.1; **p<0.05; ***p<0.01

Como primer vistazo podemos observar que el modelo con la prueba  $F$  es significativo, además los coeficientes de manera individual son significativos.

## 2.2 b) Escriba la ecuación de costos estimados.

La ecuación de costos estimados es la siguiente:

$$\hat{CT}_i = 185.531 + 22.548Y_i - 3.340Y_i^2 + 0.634Y_i^3$$

## 2.3 c) Calcule y grafique las siguientes funciones a partir de la ecuación estimada.

**Pista: calcule las derivadas y luego reemplace los valores de los betas estimados en cada ecuación.**

- i. Función de costo total

la función de costo total es la siguiente:

$$\hat{CT}_i = 185.531 + 22.548Y_i - 3.340Y_i^2 + 0.634Y_i^3$$

Que la misma de la parte superior.

- ii. Función de costo marginal (como cambia el costo total cuando cambia la producción)

la función de costo marginal se obtiene al derivar la función de costo total con respecto a la producción  $Y_i$ :

$$CM_i = \frac{d\hat{CT}_i}{dY_i} = 22.548 - 2 \cdot 3.340Y_i + 3 \cdot 0.634Y_i^2 = 22.548 - 6.680Y_i + 1.902Y_i^2$$

Así obtenemos la función de costo marginal, que nos indica cómo cambia el costo total cuando cambia la producción.

- iii. Función de costo medio (el costo total dividido la producción)

La función de costo medio se obtiene dividiendo la función de costo total por la producción  $Y_i$ :

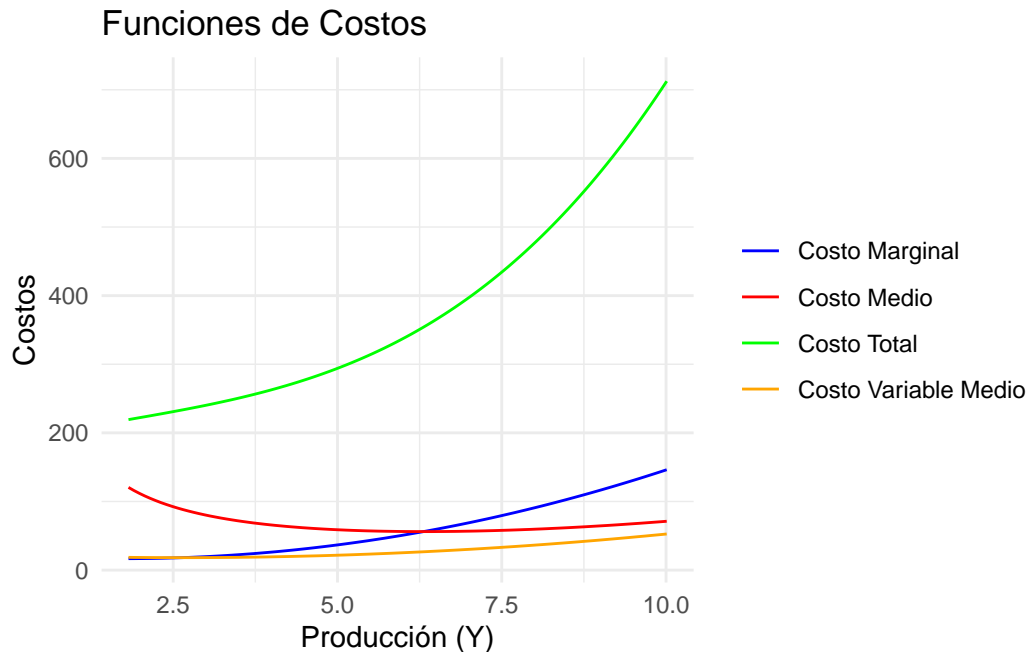
$$CM_i = \frac{\hat{CT}_i}{Y_i} = \frac{185.531 + 22.548Y_i - 3.340Y_i^2 + 0.634Y_i^3}{Y_i} = \frac{185.531}{Y_i} + 22.548 - 3.340Y_i + 0.634Y_i^2$$

- iv. Función de costo variable medio (el costo variable dividido la producción)

El costo variable se obtiene restando el costo fijo (intercepto) del costo total:

$$CV_i = \hat{CT}_i - 185.531 = 22.548Y_i - 3.340Y_i^2 + 0.634Y_i^3$$

Ahora graficaremos las funciones de costo total, costo marginal, costo medio y costo variable medio utilizando ggplot2:



**2.4 d) Interprete los coeficientes estimados (tenga en cuenta que es una función de costo total, es decir, solo hay dos costos que la componen, el costo fijo y el costo variable). Pista: Utilice la derivada de la función de producción con respecto a la producción.**

Los coeficientes estimados en la función de costos totales representan lo siguiente:

- El intercepto (185.531) representa el costo fijo de la empresa, es decir, el costo que incurre la empresa incluso cuando no produce nada.
- El coeficiente de  $Y$  (22.548) representa el costo variable por unidad de producción, es decir, el costo adicional que incurre la empresa por cada unidad adicional producida.
- El coeficiente de  $Y^2$  (-3.340) indica que el costo variable disminuye a medida que la producción aumenta, lo que sugiere economías de escala.
- El coeficiente de  $Y^3$  (0.634) indica que el costo variable aumenta a medida que la producción aumenta, lo que sugiere deseconomías de escala a niveles de producción más altos.

**2.5 e) Realice un análisis de los resultados del modelo (significancia individual, significancia conjunta,  $R^2$ ).**

El análisis de los resultados del modelo muestra que:

El resultado de todos los coeficientes son significativos al 0.01, lo que indica que cada uno de los términos ( $Y$ ,  $Y^2$ ,  $Y^3$ ) tiene un impacto significativo en el costo total.

El  $f$  calculado del modelo nos da 4,849.5 con un p-valor de 0.01 lo que indica que el modelo es significativo en su conjunto.

**2.6 f) Aplique pruebas de Autocorrelación (Correlograma (Box-Pierce-Ljung-Box), Durbin-Watson y Breusch-Godfrey), Heteroscedasticidad (Breusch-Pagan y White (sin términos cruzados)) y Normalidad (Jarque-Bera y Shapiro-Wilk). Presente el resultado de cada prueba de manera ordenada, comente sus resultados y conclusión.**

Para realizar las pruebas de autocorrelación, heteroscedasticidad y normalidad, utilizaremos las siguientes funciones en R:

Table 1: Pruebas de diagnóstico: Comparativa de niveles de significancia

Prueba	Estadístico	p-valor	Decisión (5%)
Durbin-Watson	1.6240	0.0457	Rechaza $H_0$
Breusch-Godfrey	0.4326	0.5107	No rechaza $H_0$
Ljung-Box	0.4666	0.4946	No rechaza $H_0$
Breusch-Pagan	6.0985	0.1069	No rechaza $H_0$
White	6.0985	0.1069	No rechaza $H_0$
Jarque-Bera	1.4428	0.4861	No rechaza $H_0$
Shapiro-Wilk	0.9405	0.0940	No rechaza $H_0$

Aca podemos observar como la prueba de Durbin-Watson nos indica que hay correlación con los residuos en el modelo, pero en contra parte la prueba Beusch-Godfrey no encuentra evidencia de autocorrelación, lo que puede indicar que la autocorrelación detectada por Durbin-Watson es de primer orden.

Ahora la prueba de White nos indica que existe problemas de heteroscedasticidad en el modelo, lo que puede afectar la eficiencia de los estimadores.

Ademas la prueba de Jarque-Bera y Shapiro-Wilk nos indican que los residuos siguen una distribución normal.

**2.7 g) Calcule el nivel de producción y el nivel de precio al que las firmas salen de la industria. Muéstrelo matemática y gráficamente.**

Para calcular el nivel de producción al que las firmas salen de la industria, debemos encontrar el punto en el que el costo marginal es igual al precio de mercado. En un mercado competitivo, el precio de mercado es igual al costo marginal en el punto de salida.

Matemáticamente, esto se expresa como:

$$CM_i = P$$

Donde  $P$  es el precio de mercado. Para encontrar el nivel de producción, debemos resolver la ecuación del costo marginal igual al precio de mercado, que se expresa en la siguiente igualdad:

$$22.548 - 6.680Y_i + 1.902Y_i^2 = P$$

Ahora vamos a dejar el precio de mercado como una variable y resolveremos la ecuación para encontrar el nivel de producción  $Y_i$ :

$$Y = \frac{6.680 - \sqrt{(-6.680)^2 - 4 \cdot 1.902 \cdot (22.548 - P)}}{2 \cdot 1.902}$$

Y así encontramos el óptimo de producción al que las firmas salen de la industria, dependiendo del precio de mercado.

Ahora vamos a calcular el punto de cierre, el cual se expresa de la siguiente forma:

$$CT(Y) = 185.531 + 22.548Y - 3.340Y^2 + 0.634Y^3$$

Dividimos el costo total entre la producción para obtener el costo variable medio:

$$CV(Y) = 22.548Y - 3.340Y^2 + 0.634Y^3, \quad CF = 185.531$$

$$AVC(Y) = \frac{CV(Y)}{Y} = 22.548 - 3.340Y + 0.634Y^2$$

$$\text{Punto de cierre: } P_{\text{cierre}} = Cm_i(Y_{\text{cierre}}) = AVC(Y_{\text{cierre}})$$

Reemplazando la ecuación del costo marginal y el costo variable medio en la ecuación del punto de cierre, tenemos:

$$22.548 - 3.340Y_i + 0.634Y_i^2 = 22.548 - 6.680Y_i + 1.902Y_i^2 \text{ lo que es igual a } 3.340Y_i - 1.268Y_i^2 = 0$$

Despejamos  $Y_i$  para encontrar el nivel de producción al que las firmas salen de la industria:

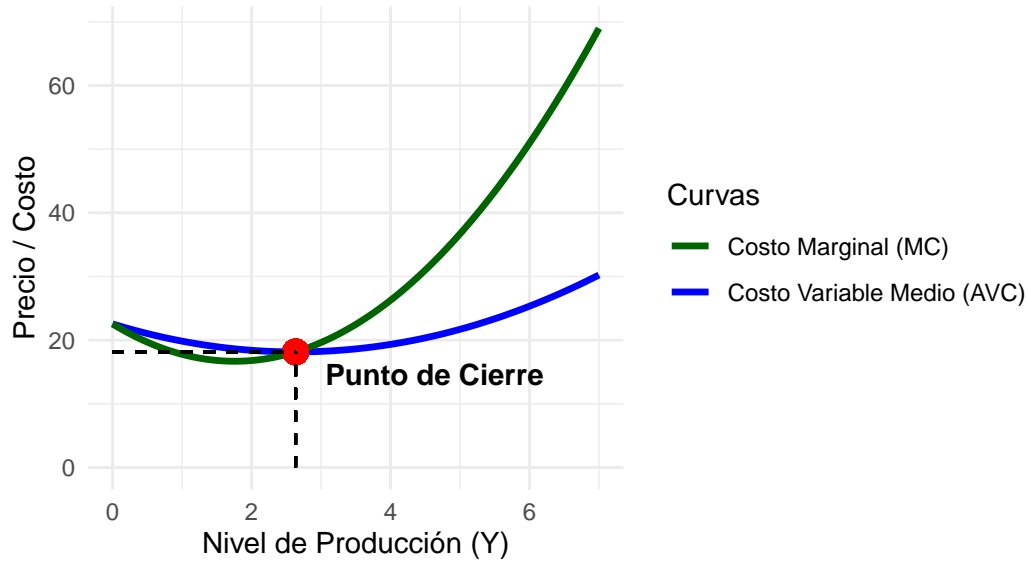
$$Y_i = \frac{3.340}{1.268} \approx 2.634$$

Ahora que tenemos el nivel de producción al que las firmas salen de la industria, podemos calcular el nivel de precio al que las firmas salen de la industria utilizando la función de costo variable medio:

$$P_{\text{cierre}} = AVC(Y_{\text{cierre}}) = 22.548 - 3.340(2.634) + 0.634(2.634)^2 \approx 18.150$$

### Punto de Cierre de la Firma

Punto de salida (corto plazo) en  $Y = 2.634$  y  $P = 18.15$



**2.8 h) Pruebe la siguiente hipótesis con un nivel de significancia del 5%. “Una de las firmas que participan en la industria le manifiestan que según sus cálculos el nivel de producción al que una firma tiene que cerrar es de 4.000 unidades”.**

Para probar la hipótesis de que el nivel de producción al que una firma tiene que cerrar es de 4.000 unidades, podemos plantear las siguientes hipótesis:

$$H_0 : Y_{\text{cierre}} = 4 \text{ la firma cierra a 4.000 unidades}$$

$$H_a : Y_{\text{cierre}} \neq 4 \text{ (La firma no cierra a 4.000 unidades)}$$

otra forma de plantear la hipótesis es:  $H_0 : b_2 + 8b_3 = 0$  (La firma cierra a 4.000 unidades)

Linear hypothesis test:

$$I(Y^2) + 8 I(Y^3) = 0$$

Model 1: restricted model

Model 2:  $CT \sim Y + I(Y^2) + I(Y^3)$

Res.Df	RSS	Df	Sum of Sq	Chisq	Pr(>Chisq)
--------	-----	----	-----------	-------	------------



```

1      27 1235.5
2      26 1111.1  1      124.31 2.9088      0.0881 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Podemos observar que el valor p de la prueba es igual a 0.0851, lo que nos indica que no podemos rechazar la hipótesis nula al nivel de significancia del 5%, lo que sugiere que no hay suficiente evidencia para afirmar que el nivel de producción al que una firma tiene que cerrar es diferente de 4.000 unidades.

### 2.9 i) Aplique una prueba de Normalidad, Homoscedasticidad y no Autocorrelación, para determinar si el modelo cumple con los supuestos clásicos. ¿Existe algún problema econométrico?

Ya fue realizado en la parte superior del ejercicio, donde se aplicaron las pruebas de Jarque-Bera y Shapiro-Wilk para normalidad, Breusch-Pagan y White para heteroscedasticidad, y Durbin-Watson y Breusch-Godfrey para autocorrelación.

## 3 EJERCICIO 3 – Modelo de Regresión Múltiple (Test de Chow y Variables Dummy)

(Utilice R para las estimaciones)

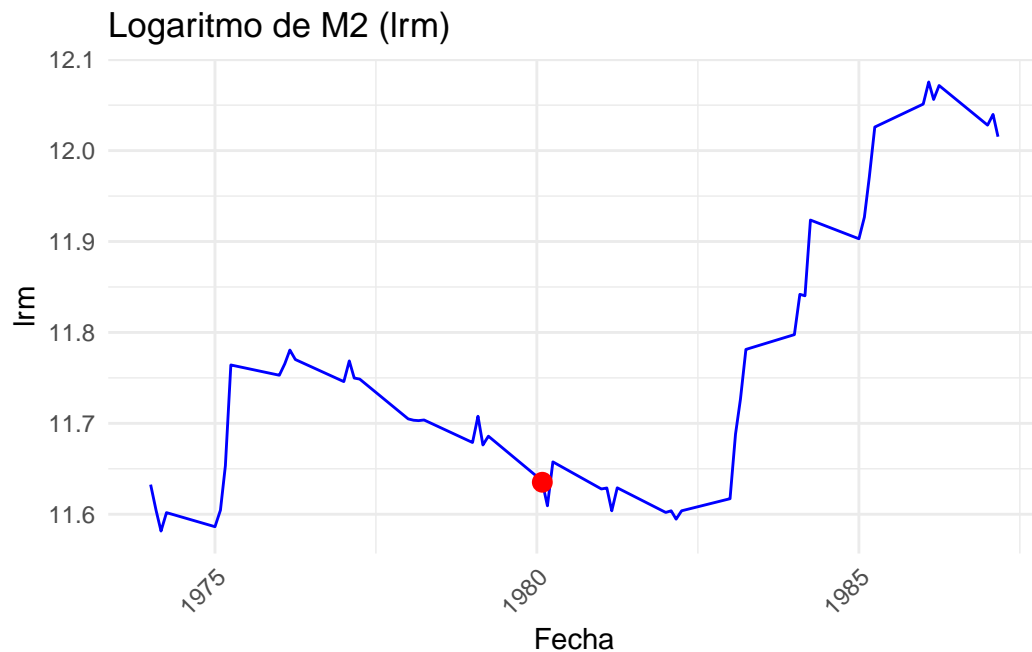
El archivo “demandadinero.xls” contiene la información para estimar una función de demanda por dinero para Dinamarca. La información es trimestral y cubre el periodo 1974:1 a 1987:3.

**Modelo a estimar:**

$$lrm_t = \beta_0 + \beta_1 lry_t + \beta_2 ib_t + \beta_3 id_t + \varepsilon_t$$

Donde:

- $lrm_t$  es el logaritmo de M2 (en términos reales).
- $lry_t$  es el logaritmo del ingreso real.
- $ib_t$  es una tasa de interés sobre bonos.
- $id_t$  es una tasa de interés sobre depósitos.



### 3.1 a) Test de Chow

Utilizando el *test de Chow*, compruebe la existencia (o no) de un cambio estructural en la función de demanda de dinero para Dinamarca en 1980:2. Comente el resultado.

- Escriba las Hipótesis (nula y alterna).
- Escriba los modelos a estimar (R y U).
- Presenta una tabla con la estimación de cada modelo.
- Muestre el cálculo del Estadístico de Prueba (F).
- Concluya.

Para resolver este punto, primero debemos definir las hipótesis para el test de Chow: **Hipótesis Nula** ( $H_0$ ): No existe cambio estructural en 1980:2. Todos los coeficientes asociados a la variable dummy y sus interacciones son iguales a cero.

$$H_0 : \alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 0$$

**Hipótesis Alterna** ( $H_1$ ): Existe cambio estructural en 1980:2. Al menos uno de los coeficientes asociados a la variable dummy o sus interacciones es distinto de cero.

$$H_1 : \text{Al menos un } \alpha_i \neq 0 \text{ para } i = 0, 1, 2, 3$$

Ahora vamos a definir los modelos restringido (R) y no restringido (U):

Modelo Restringido (R):

$$lrm_t = \beta_0 + \beta_1 lry_t + \beta_2 ib_t + \beta_3 id_t + \varepsilon_t$$

Modelo No Restringido (U):

$$lrm_t = \beta_0 + \beta_1 lry_t + \beta_2 ib_t + \beta_3 id_t + \alpha_0 D_t + \alpha_1 (D_t \cdot lry_t) + \alpha_2 (D_t \cdot ib_t) + \alpha_3 (D_t \cdot id_t) + \varepsilon_t$$

Donde  $D_t$  es una variable dummy que toma el valor de 0 para los periodos anteriores a 1980:2 y 1 para los periodos posteriores a 1980:2.

Ahora vamos a estimar ambos modelos de series de tiempo utilizando regresión lineal y presentar los resultados en una tabla:

#### Resultados de las Regresiones

Dependent variable:		
	lrm	
	(1)	(2)
D		-2.608 (1.936)
lry	1.296*** (0.094)	1.144*** (0.210)
ib	-2.616*** (0.328)	-2.506*** (0.836)
id	0.619 (0.691)	0.693 (1.071)
D:lry		0.436 (0.321)
D:ib		0.926 (1.082)
D:id		-1.341 (1.910)
Constant	4.394*** (0.581)	5.266*** (1.190)
Observations	55	55
R2	0.926	0.932
Adjusted R2	0.922	0.922
Residual Std. Error	0.043 (df = 51)	0.043 (df = 47)
F Statistic	213.279*** (df = 3; 51)	91.627*** (df = 7; 47)
Note: *p<0.1; **p<0.05; ***p<0.01		

Ahora vamos a realizar la prueba del estadístico F para comparar ambos modelos:

### Analysis of Variance Table

Model 1:  $lrm \sim lry + ib + id$

Model 2:  $lrm \sim D * (lry + ib + id)$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	51	0.092543				
2	47	0.085589	4	0.0069547	0.9548	0.4411

El p-value que obtenemos con la prueba F es de 0.4411 lo que significa que no podemos rechazar la hipótesis nula a ningún valor de significancia, por lo tanto esto nos demuestra que no hay evidencia suficiente para afirmar un choque estructural en la función de demanda por dinero para Dinamarca en 1980-2.

Por esto nos inclinamos a tomar el modelo restringido como el modelo preferido.

### 3.2 b) Variables Dummy

Continuando con el ejercicio anterior, construya una *variable dummy* que toma el valor de cero entre 1974:1 y 1980:1, y uno entre 1980:2 y 1987:3. Utilizando esta *variable dummy* pruebe si hay evidencia estadística para demostrar la existencia (o no) de un cambio estructural en la función de demanda por dinero para Dinamarca en 1980:2. Específicamente:

- Escriba las Hipótesis (nula y alterna).
- Escriba los modelos a estimar (R y U).
- Presenta una tabla con la estimación de cada modelo.
- Muestre el cálculo del Estadístico de Prueba (F).
- Concluya.

En este punto vamos a establecer el cambio en el intercepto.

**Hipótesis Nula ( $H_0$ ):** No existe cambio estructural en 1980:2. El coeficiente asociado a la variable dummy es igual a cero.

$$H_0 : \alpha_0 = 0$$

**Hipótesis Alterna ( $H_1$ ):** Existe cambio estructural en 1980:2. El coeficiente asociado a la variable dummy es distinto de cero.

$$H_1 : \alpha_0 \neq 0$$

Modelo Restringido (R):

$$lrm_t = \beta_0 + \beta_1 lry_t + \beta_2 ib_t + \beta_3 id_t + \varepsilon_t$$

Modelo No Restringido (U):

$$lrm_t = \beta_0 + \beta_1 lry_t + \beta_2 ib_t + \beta_3 id_t + \alpha_0 D_t + \varepsilon_t$$

Ahora vamos a estimar ambos modelos utilizando regresión lineal y presentar los resultados en una tabla:

## Resultados de las Regresiones

Dependent variable:		
	lrn	
	(1)	(2)
lry	1.296*** (0.094)	1.271*** (0.110)
ib	-2.616*** (0.328)	-2.541*** (0.371)
id	0.619 (0.691)	0.385 (0.869)
D		0.007 (0.016)
Constant	4.394*** (0.581)	4.549*** (0.680)
Observations	55	55
R2	0.926	0.926
Adjusted R2	0.922	0.921
Residual Std. Error	0.043 (df = 51)	0.043 (df = 50)
F Statistic	213.279*** (df = 3; 51)	157.505*** (df = 4; 50)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Ahora continuaremos con la prueba del estadístico F para comparar ambos modelos:

## Analysis of Variance Table

Model 1:  $lrn \sim lry + ib + id$

Model 2:  $lrn \sim lry + ib + id + D$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	51	0.092543				
2	50	0.092172	1	0.00037146	0.2015	0.6555

El p-value obtenido es de 0.6555, lo que nos demuestra que ni siquiera hubo un cambio en el intercepto, osea que no hubo cambio en el nivel base de la demanda por dinero después de 1980:2. Por lo tanto, no hay evidencia suficiente para afirmar que existe un cambio estructural en la función de demanda por dinero para Dinamarca en 1980:2.

### 3.3 c) Supuestos Clásicos

Aplique una prueba de Normalidad, Homoscedasticidad y no Autocorrelación, para determinar si el modelo cumple con los supuestos clásicos. ¿Existe algún problema econométrico? ¿Como podría corregirse? (No debe corregirlo, solo explicar cómo).

Table 2: Pruebas de diagnóstico: Comparativa de niveles de significancia

Prueba	Estadístico	p-valor	Decisión (5%)
Durbin–Watson	0.7449	0.0000	Rechaza H0
Breusch–Godfrey	19.9943	0.0000	Rechaza H0
Ljung–Box	20.4649	0.0000	Rechaza H0
Breusch–Pagan	0.8777	0.8308	No rechaza H0
White	6.8468	0.0326	Rechaza H0
Jarque–Bera	0.1778	0.9149	No rechaza H0
Shapiro–Wilk	0.9846	0.7030	No rechaza H0

Primero con el test de Durbin-Watson, Breusch-Godfrey y Ljung-Box nos indican que hay una alitisima evidencia de autocorrelación en el modelo, lo que puede afectar la validez de los resultados. Para corregir este problema, se podrían utilizar métodos como la regresión con errores estándar robustos o modelos de series de tiempo que tengan en cuenta la autocorrelación, como modelos con errores autoregresivos.

Segundo, en el test de Breusch-Pagan nos indica que no hay evidencia de heteroscedasticidad, pero en contraste el test de White nos indica que si hay evidencia de heteroscedasticidad, lo que sugiere que la heteroscedasticidad no es lineal. Para corregir este problema, se podrían utilizar errores estándar robustos o transformar la variable dependiente para estabilizar la varianza.

Tercero, en el test de Jarque-Bera y Shapiro-Wilk nos indican que los residuos si siguen una distribución normal, lo que es un buen resultado para el modelo y no necesita corrección en este aspecto. Pero si lo necesitará seria utilizar transformaciones de la variable dependiente o utilizar métodos de estimación que no requieran normalidad, como la regresión robusta o modelos de regresión no paramétrica.

## 4 EJERCICIO 4 – Regresión con Variables Instrumentales (Endogeneidad)

Una herramienta en la búsqueda de la reducción de las enfermedades y las muertes causadas por el tabaco —y en los costos o externalidades, en los que se incurre debido a esas enfermedades y que son soportados por el resto de la sociedad— consiste en gravar los cigarrillos tan duramente como para conseguir que el número de fumadores actuales se reduzca y se disuada a los potenciales nuevos fumadores de adquirir el hábito. Durante los últimos 40 años, diferentes países han aumentado los impuestos sobre el cigarrillo con el objetivo de reducir su consumo, basándose en estimaciones de la elasticidad de la demanda de cigarrillos.

El archivo “**Cigarrillos.xlsx**” contiene información sobre el consumo de cigarrillos en Estados Unidos para 48 estados y para los años 1985 y 1995.

A continuación, una descripción de las variables.

Variable	Descripción
<i>state</i>	48 estados
<i>year</i>	1985, 1995
<i>cpi</i>	índice de precios al consumidor
<i>population</i>	población en cada estado
<i>packs</i>	número de paquetes de cigarrillos vendidos per cápita en cada estado
<i>income</i>	ingreso nominal en cada estado
<i>tax</i>	impuesto especial (al cigarrillo) estatal promedio para el año fiscal
<i>price</i>	es el precio nominal promedio por paquete de cigarrillos (incluidos todos los impuestos)
<i>taxs</i>	impuesto especial (al cigarrillo) estatal promedio para el año fiscal, incluido el impuesto sobre las ventas

Teniendo en cuenta que las variables de precios, ingresos e impuestos están en términos nominales, debemos transformarlas para tenerlas en términos reales, utilizando el índice de precios al consumidor (*cpi*); adicionalmente, el ingreso debe estar medido en términos per cápita. Lo primero que deben hacer, luego de cargar los datos, es crear las siguientes variables:

$$rprice = \frac{price}{cpi} \quad rincome = \frac{\left(\frac{income}{population}\right)}{cpi} \quad rtdiff = \frac{(taxs - tax)}{cpi}$$

El modelo que nos interesa inicialmente es:

$$\log(Q_{state}^{cigarrillos}) = \beta_0 + \beta_1 \log(P_{state}^{cigarrillos}) + \varepsilon_t$$

**4.1 a) Estime la ecuación (4) por MCO para 1985 y 1995. Presente los resultados en una misma Tabla. ¿Cuál es el valor de la elasticidad precio de la demanda en cada año? Interprete los valores.**

Resultados de las Regresiones MCO

```
=====
Dependent variable:
-----
log(packs)
```

	(1)	(2)
log(rprice)	-0.985*** (0.255)	-1.213*** (0.216)
Constant	9.285*** (1.165)	10.339*** (1.035)
Observations	48	48
R2	0.245	0.406
Adjusted R2	0.228	0.393
Residual Std. Error (df = 46)	0.148	0.190
F Statistic (df = 1; 46)	14.891***	31.409***
=====		
Note:	*p<0.1; **p<0.05; ***p<0.01	

Intepretación: En 1985 la elasticidad precio de la demanda de los cigarrillos es de -0.985, es decir que ante un aumento del 1% en el precio real de los cigarrillos, se espera que la cantidad demandada de cigarrillos disminuya en un 0.985%, manteniendo todo lo demás constante. Esto mostrando una demanda inelastica.

En 1995 la elasticidad precio de la demanda de los cigarrillos es de -1.213, es decir que ante un aumento del 1% en el precio real de los cigarrillos, se espera que la cantidad demandada de cigarrillos disminuya en un 1.213%, manteniendo todo lo demás constante. Esto mostrando una demanda elastica.

#### 4.2 b) Explique por qué la regresión MCO estimada no resulta apropiada para estimar la elasticidad precio de la demanda de cigarrillos.

La regresión MCO estimada no resulta apropiada para estimar la elasticidad precio de la demanda de cigarrillos debido a un posible problema de endogeneidad. En este caso, el precio real de los cigarrillos (rprice) podría estar correlacionado con el término de error ( ) en la ecuación de demanda. Esto puede ocurrir si hay factores no observados que afectan tanto el precio como la cantidad demandada, como cambios en las preferencias de los consumidores, campañas antitabaco, o políticas gubernamentales que no se han incluido en el modelo.

Además existen un problema muy claro de simultaneidad ya que el precio de los cigarrillos puede ser determinado por la cantidad demandada, es decir que el precio y la cantidad pueden influenciarse mutuamente, lo que viola uno de los supuestos clave del modelo de regresión lineal clásico y conduce a estimaciones sesgadas e inconsistentes de la elasticidad precio de la demanda.

#### 4.3 c) Explique por qué la variable *rtdiff* es un instrumento válido para la regresión de variables instrumentales de la ecuación (4).

Para que una variable sea un instrumento válido, debe cumplir con dos condiciones principales: relevancia y exogeneidad.



En este caso, la variable *rtdiff* (diferencia en impuestos reales) podría ser un instrumento válido para la regresión de variables instrumentales de la ecuación (4) porque la diferencia entre impuestos reales puede afectar al precio del cigarrillos (*rprice*) y, por lo tanto, influir en la cantidad demandada de cigarrillos (*packs*). Esto cumple con la condición de relevancia.

Ahora, en cuanto a la exogeneidad, *rtdiff* no debería estar correlacionada con el término de error ( ) en la ecuación de demanda. Si los cambios en los impuestos reales son determinados por factores externos al mercado de cigarrillos y no están relacionados con las preferencias de los consumidores o factores no observados que afectan la demanda, entonces *rtdiff* cumpliría con la condición de exogeneidad.

Por lo tanto, si se cumplen ambas condiciones, *rtdiff* sería un instrumento válido para estimar la elasticidad precio de la demanda de cigarrillos utilizando regresión de variables instrumentales.

#### 4.4 d) ¿Es la variable *rtdiff* un instrumento débil?

Para aceptar o negar esto tendremos que realizar una prueba de instrumentos débiles, que generalmente se basa en la estadística F del primer paso de la regresión de variables instrumentales. Si el valor de F es menor a 10, se considera que el instrumento es débil.

Call:

```
ivreg(formula = log(packs) ~ log(rprice) | rtdiff, data = df95)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.64619	-0.07732	0.02981	0.11283	0.41904

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.7199	1.5141	6.420	6.79e-08 ***
log(rprice)	-1.0836	0.3166	-3.422	0.00131 **

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	1	46	40.956	7.27e-08 ***
Wu-Hausman	1	45	0.314	0.578
Sargan	0	NA	NA	NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1904 on 46 degrees of freedom

Multiple R-Squared: 0.4011, Adjusted R-squared: 0.3881

Wald test: 11.71 on 1 and 46 DF, p-value: 0.001313

Aca nos fijamos en el estadístico F de la prueba de “Weak instruments”, que nos da un  $F > 10$  exactamente de 40.95, lo que nos indica que *rtdiff* no es un instrumento débil para la regresión de variables instrumentales de la ecuación (4).

**4.5 e) Estime la regresión de la ecuación (4) utilizando la variable *rtdiff* como instrumento, para 1985 y 1995. Presente los resultados en una misma Tabla. ¿Cuál es el valor de la elasticidad precio de la demanda? Interprete los valores.**

Para esta regresión usaremos MC2E (Mínimos Cuadrados en Dos Etapas) utilizando la variable *rtdiff* como instrumento para el precio real de los cigarrillos (*rprice*).

Resultados de las Regresiones IV

Dependent variable:		
	log(packs)	
	(1)	(2)
log( <i>rprice</i> )	-1.178** (0.482)	-1.084*** (0.317)
Constant	10.169*** (2.198)	9.720*** (1.514)
Observations	48	48
R2	0.235	0.401
Adjusted R2	0.218	0.388
Residual Std. Error (df = 46)	0.149	0.190
Note: *p<0.1; **p<0.05; ***p<0.01		

Como podemos ver en 1985 la elasticidad precio de la demanda de los cigarrillos es de -1.178, es decir que ante un aumento del 1% en el precio real de los cigarrillos, se espera que la cantidad demandada de cigarrillos disminuya en un 1.178%, manteniendo todo lo demás constante. Esto mostrando una demanda elastica.

En este apartado pasamos de un lugar en el que (MCO) nos indicaba una demanda inelastica, a un lugar en el que (MC2E) nos indica una demanda elastica, lo que resalta la importancia de corregir por endogeneidad para obtener estimaciones más precisas y confiables de la elasticidad precio de la demanda de cigarrillos.

Ademas en 1995 la elasticidad precio de la demanda de los cigarrillos es de -1.084, es decir que ante un aumento del 1% en el precio real de los cigarrillos, se espera que la cantidad demandada de cigarrillos disminuya en un 1.084%, manteniendo todo lo demás constante. Esto mostrando una demanda elastica.

Ahora en este apartado si nos mantenemos en elasticidad aunque ligeramente mas reducida.

**4.6 f) Ahora, comparemos los resultados presentados en los literales a) y e), en términos de política pública. ¿Cuál debería ser el precio de venta después de impuestos de los cigarrillos necesario para lograr una reducción del 20% en el consumo de cigarrillos, en 1995? Haga los cálculos utilizando las elasticidades del literal a) sin corregir el problema de endogeneidad y después con las elasticidades del literal e) corrigiendo por endogeneidad. Compare los resultados, resaltando la importancia de corregir por endogeneidad.**

Para hacer estos cálculos, utilizaremos la fórmula de la elasticidad precio de la demanda:

$$\text{Elasticidad} = \frac{\% \Delta Q}{\% \Delta P}$$

Donde  $\% \Delta Q$  es el cambio porcentual en la cantidad demandada y  $\% \Delta P$  es el cambio porcentual en el precio.

Nosotros tenemos  $\% \Delta Q = -20\%$  (una reducción del 20% en el consumo de cigarrillos).

Además, tenemos dos elasticidades diferentes para 1995: la elasticidad sin corregir por endogeneidad (MCO) que es de -1.213, y la elasticidad corregida por endogeneidad (MC2E) que es de -1.084.

Primero, utilizando la elasticidad sin corregir por endogeneidad (MCO):

$$-1.213 = \frac{-20\%}{\% \Delta P}$$

$$\% \Delta P = \frac{-20\%}{-1.213} \approx 16.5\%$$

Esto significa que, según la elasticidad sin tratamiento por endogeneidad, el precio de venta después de impuestos de los cigarrillos debería aumentar en aproximadamente un 16.5% para lograr una reducción del 20% en el consumo de cigarrillos en 1995.

Ahora, utilizando la elasticidad corregida por endogeneidad (MC2E):

$$-1.084 = \frac{-20\%}{\% \Delta P}$$

$$\% \Delta P = \frac{-20\%}{-1.084} \approx 18.45\%$$

Esto significa que, según la elasticidad tratada por endogeneidad, el precio de venta después de impuestos de los cigarrillos debería aumentar en aproximadamente un 18.45% para lograr una reducción del 20% en el consumo de cigarrillos en 1995.

La implicación de política pública es una subestimación del precio real que debería aumentar el gobierno para lograr desincentivar el consumo de cigarrillos.

**4.7 g) Ahora estime la ecuación (4) incluyendo *rincome*, para 1995. Presente una Tabla con los resultados de la estimación MCO y la estimación MC2E utilizando como variables instrumentales *rtdiff* y *rtax*.**

Para esto, primero debemos crear la variable *rtax* que es el impuesto real, y luego estimar ambos modelos (MCO y MC2E) incluyendo *rincome* como variable explicativa adicional. Finalmente, presentaremos los resultados en una tabla.

Resultados de las Regresiones MCO y MC2E con ingreso

Dependent variable:		
	log(packs)	
	OLS	instrumental
		variable
	(1)	(2)
log(rprice)	-1.407*** (0.251)	-1.277*** (0.263)
log(rincome)	0.344 (0.235)	0.280 (0.239)
Constant	10.342*** (1.023)	9.895*** (1.059)
Observations	48	48
R2	0.433	0.429
Adjusted R2	0.408	0.404
Residual Std. Error (df = 45)	0.187	0.188
F Statistic	17.165*** (df = 2; 45)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

En esta tabla podemos observar como el metodo de OLS sobreestima la elasticidad precio de la demanda de los cigarrillos, y como el ingreso real tiene un efecto positivo sobre la cantidad demandada de cigarrillos, lo que sugiere que los cigarrillos pueden ser un bien normal.

**4.8 h) Interprete los coeficientes de la estimación MC2E.**

A partir de la estimación por Mínimos Cuadrados en Dos Etapas (MC2E) para 1995, utilizando *rtdiff* y *rtax* como instrumentos, se obtienen las siguientes elasticidades: Elasticidad Precio ( $\log(rprice)$ ): El coeficiente es de  $-1.277$ . Esto indica que, manteniendo el ingreso constante, ante un aumento del 1% en el precio real de los cigarrillos, la cantidad demandada disminuye en un 1.277%. La demanda se mantiene en el rango elástico ( $|-1.277| > 1$ ), y el coeficiente es altamente

significativo ( $p < 0.01$ ). Elasticidad Ingreso ( $\log(rincome)$ ): El coeficiente es de 0.280. Esto implica que, ante un aumento del 1% en el ingreso real per cápita, el consumo de cigarrillos aumenta en un 0.280%, manteniendo el precio constante. Dado que el signo es positivo, los resultados sugieren que los cigarrillos se comportan como un bien normal, aunque con una respuesta relativamente inelástica al ingreso.

**4.9 i) Presente los resultados una Tabla con las pruebas de diagnóstico (*Instrumentos débiles, Exogeneidad de Wu-Hausman, Sobreidentificación de Sargan*) sobre la estimación MC2E del literal g). Presente la hipótesis nula y la interpretación del resultado de cada una.**

Call:

```
ivreg(formula = log(packs) ~ log(rprice) + log(rincome) | rtdiff +
      rtax + log(rincome), data = df95)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.6006930	-0.0862223	-0.0009999	0.1164699	0.3734227

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.8950	1.0586	9.348	4.12e-12 ***
$\log(rprice)$	-1.2774	0.2632	-4.853	1.50e-05 ***
$\log(rincome)$	0.2804	0.2386	1.175	0.246

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	2	44	244.734	<2e-16 ***
Wu-Hausman	1	44	3.068	0.0868 .
Sargan	1	NA	0.333	0.5641

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1879 on 45 degrees of freedom

Multiple R-Squared: 0.4294, Adjusted R-squared: 0.4041

Wald test: 13.28 on 2 and 45 DF, p-value: 2.931e-05

El análisis muestra que para estimar la demanda de cigarrillos hay que tratar con cuidado la endogeneidad. Aunque las pruebas estadísticas sugieren una exogeneidad apenas marginal, el modelo de Variables Instrumentales (MC2E) entrega estimaciones más alineadas con la teoría económica. La diferencia en las elasticidades encontradas ( $-1.407$  en MCO frente a  $-1.277$  en MC2E) deja ver que si se ignora la simultaneidad, se podrían diseñar políticas de impuestos con una idea equivocada de qué tan sensible es el consumo al precio, y eso puede afectar los objetivos de salud pública. Por

eso, es importante usar métodos que corrijan la endogeneidad para tener estimaciones más precisas y políticas más efectivas.

Para ver el código de R dar click en el siguiente enlace. [link](#)