Derechos Reservados © 2026 Mundo Social
Mundo Social **06**, Notas (2026)
Disponible en https://mundo-social.com/

Violaciones de supuestos: Causas, consecuencias y soluciones

Anjaly Arcos Huaman [1]

Melany Vega Lugo [2]

Curso dictado en 25-2 por William Sanchez

[1] *Undergraduate Economist, Universidad Nacional Mayor de San Marcos*
anjaly.arcos@gmail.com
[2] *Undergraduate Economist, Universidad Nacional Mayor de San Marcos*
melany.vega@unmsm.edu.pe
Disponible en: https://mundo-social.com/

A continuación, se presenta una nota académica dedicada al estudio de Econometría, inspirada en el curso dictado por el profesor Mg. William Sanchez. Este material ha sido elaborado con el mayor respeto hacia el profesor y con fines estrictamente educativos, con el propósito de contribuir a la comprensión y difusión de sus enseñanzas, así como de asegurar que su valioso legado académico perdure en el tiempo.
Recomendamos encarecidamente la lectura de los libros y materiales del profesor, los cuales constituyen una fuente rigurosa y profunda para el estudio.
Esta nota incluye el desarrollo de los siguientes subtemas fundamentales:

- Heterocedasticidad
- Autocorrelación
- Inestabilidad de parámetros
- Multicolinealidad
- No normalidad
- Endogeneidad

Ante cualquier error, comentario o sugerencia, no dude el lector en escribirnos al correo que aparece en la portada del presente documento.

1. El Modelo Clásico de Regresión Lineal: Un Enfoque General

La econometría tiene como objetivo fundamental cuantificar relaciones económicas a partir de datos observados. Para ello, se apoya en el Modelo Lineal General (MLG) como herramienta central de análisis. Este modelo representa formalmente el proceso generador de datos (PGD) bajo el cual una variable dependiente es explicada por un conjunto de variables explicativas y un término de error que recoge factores no observados.

En su forma más general, el modelo puede expresarse como:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \tag{1}$$

Donde Y_i es la variable dependiente, X_{ji} son las variables explicativas, β_j representan los parámetros estructurales del modelo y u_i es el término de perturbación aleatoria. La correcta especificación del modelo exige que la forma funcional utilizada en la estimación sea coherente con la estructura del proceso generador de datos; es decir, si el PGD es lineal, la estimación debe mantener dicha linealidad, y si incluye términos cuadráticos o transformaciones, estos deben incorporarse adecuadamente.

El método de estimación más utilizado en este contexto es el de Mínimos Cuadrados Ordinarios (MCO), el cual obtiene los estimadores minimizando la suma de los residuos al cuadrado. Bajo ciertos supuestos, el estimador MCO es el Mejor Estimador Lineal Insesgado (BLUE [1] por sus siglas en inglés), garantizando eficiencia e insesgadez en la estimación de los parámetros.

Sin embargo, dichas propiedades dependen críticamente del cumplimiento de una serie de supuestos fundamentales. Entre ellos destacan la linealidad en los parámetros, la correcta especificación del modelo, la exogeneidad de los regresores, la homocedasticidad, la ausencia de autocorrelación y la inexistencia de multicolinealidad perfecta. En particular, la condición de exogeneidad —expresada como $E(u_i|X) = 0$— resulta esencial para garantizar la insesgadez de los estimadores, ya que implica que el término de error no está correlacionado con las variables explicativas.

Cuando alguno de estos supuestos es violado, las propiedades estadísticas del estimador MCO pueden verse afectadas. Dependiendo del supuesto incumplido, los estimadores pueden perder eficiencia, volverse sesgados o incluso inconsistentes, comprometiendo la validez de la inferencia estadística y de las conclusiones económicas derivadas del modelo.

El propósito de esta nota académica es analizar, de manera sistemática, las principales violaciones a los supuestos del Modelo Lineal General, examinando sus causas, consecuencias teóricas y alternativas de corrección. Para ello, se parte del marco general del modelo clásico y posteriormente se estudia cada caso particular de incumplimiento, con el fin de proporcionar una comprensión integral tanto conceptual como aplicada.

2. Heterocedasticidad

La heterocedasticidad ocurre cuando la varianza de los errores en un modelo de regresión no es constante a lo largo de las observaciones — expresada como $Var(u_i|X_i) \neq \sigma^2$—. Este fenómeno viola uno de los supuestos clave en la regresión lineal, afectando la eficiencia de los estimadores.

[1] Las siglas BLUE corresponden a: Best (mínima varianza), Linear (función lineal de los datos), Unbiased (su valor esperado es el parámetro poblacional) y Estimator.

 © **Mundo Social 2026**

2.1. Causas

En primer lugar, una de las causas más frecuentes es la omisión de variables explicativas relevantes. Cuando el modelo no incorpora variables que forman parte del verdadero proceso generador de datos, el término de error comienza a capturar efectos sistemáticos no modelados. Si la variable omitida está correlacionada con alguna de las variables incluidas y además influye en la dispersión de la variable dependiente, la varianza del error puede variar sistemáticamente con el nivel de los regresores, generando heterocedasticidad.

En segundo lugar, la heterocedasticidad puede presentarse cuando existen diferencias significativas en las escalas de medición de las variables. En muestras que combinan unidades económicas de distinto tamaño —por ejemplo, empresas grandes y pequeñas, o países con niveles de ingreso muy disímiles— la magnitud de la variable dependiente puede variar considerablemente entre observaciones. Esta heterogeneidad en la escala suele traducirse en una mayor dispersión del término de error a medida que aumenta el nivel de la variable explicativa, produciendo una estructura de varianza no constante.

Asimismo, la presencia de valores extremos u outliers[2] puede generar heterocedasticidad. Los outliers tienden a influir de manera desproporcionada en la estimación del modelo, alterando la distribución de los residuos y aumentando su varianza en determinados rangos de la muestra. En estos casos, la variabilidad del error no responde a un patrón aleatorio, sino que está concentrada en observaciones específicas que distorsionan la estructura de dispersión.

También puede deberse a cambios estructurales en los datos. Cuando la muestra abarca distintos periodos, regiones o regímenes económicos, es posible que la relación entre las variables cambie a lo largo del tiempo o entre grupos. Estos cambios pueden modificar no solo los parámetros del modelo, sino también la variabilidad del término de error. Si tales transformaciones no son modeladas explícitamente mediante variables dummy[3] o estimaciones por submuestras, la varianza de los residuos puede diferir entre segmentos de la muestra, generando heterocedasticidad.

2.2. Consecuencias

Aunque los coeficientes son insesgados, los estimadores no son los más precisos. Un punto clave es que los estimadores obtenidos por MCO siguen siendo insesgados siempre que se mantenga la exogeneidad. Esto significa que, en promedio, los coeficientes estimados coinciden con los verdaderos parámetros poblacionales. Sin embargo, el problema aparece en la precisión de esas estimaciones.

La heterocedasticidad provoca que los estimadores sean ineficientes. En otras palabras, aunque los coeficientes estén "bien" en promedio, ya no son los más precisos posibles. Se pierde la propiedad BLUE, porque existe otra forma de estimar que podría producir varianzas más pequeñas. Esto implica que los coeficientes pueden presentar una mayor dispersión de la necesaria.

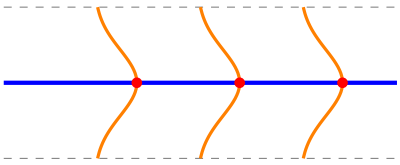
Además, uno de los efectos más importantes se observa en la inferencia estadística. Las pruebas de hipótesis, como la prueba t y la prueba F, dependen del supuesto de varianza constante de los errores. Cuando este supuesto se viola, los errores estándar calculados por MCO dejan de ser

[2] Un outlier es una observación que se encuentra a una distancia anormal de otros valores en una muestra aleatoria. En econometría, estos pueden distorsionar la estimación MCO al aumentar desproporcionadamente la suma de los residuos al cuadrado.

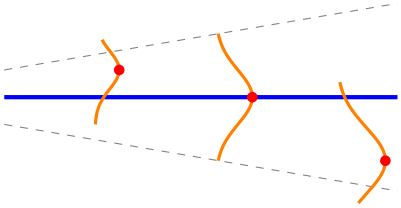
[3] Una variable dummy es un indicador binario que toma los valores 0 o 1 para representar categorías cualitativas o cambios en el comportamiento de los datos. Se profundiza en su aplicación en la sección 4.3.

correctos. Como consecuencia, los estadísticos t y F pueden estar mal estimados, lo que lleva a conclusiones equivocadas sobre la significancia de las variables.

Relacionado con lo anterior, los intervalos de confianza también se ven afectados. Si los errores estándar están mal calculados, los intervalos pueden resultar demasiado amplios o demasiado estrechos, generando una percepción incorrecta sobre la precisión de los coeficientes estimados.



σ^2 cte (Homocedasticidad)



$\sigma^2 \neq$ cte (Heterocedasticidad)

2.3. Detección y solución

- Detección visual: Gráficos de residuos vs. valores ajustados.
- Técnicas para estimar mediante Test: Breusch - Pagan, White
- Soluciones: Errores robustos, transformaciones logarítmicas, GLS.

a. Test Breusch - Pagan - Godfrey

Es una prueba estadística que se utiliza en econometría para detectar la presencia de heterocedasticidad en un modelo de regresión lineal.

- Etapa 1: Estimar el modelo de regresión y obtener los residuos.

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k + \hat{\varepsilon} \tag{2}$$

- Etapa 2: Obtener el cuadrado de los residuos

$$\hat{\varepsilon}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2 + \cdots + \hat{\alpha}_k X_k + \hat{v} \tag{3}$$

$\hat{\varepsilon}^2$: representa el proxy de la varianza

- Prueba de Hipótesis:

H_0 : Homocedasticidad p $> 5\,\%$ No se rechaza la hipótesis nula

H_1 : Heterocedasticidad p $< 5\,\%$ Se rechaza la hipótesis nula

- Calcular el estadístico LM:

$$\underbrace{LM}_{\text{Multiplicador Lagrange}} = \frac{SEC}{2(\sigma^2)^2} \quad \rightarrow \text{ de la Etapa II}$$

- Comparar con la distribución Chi - cuadrada

$$LM \sim \chi^2_{(k)}$$

b. Test de Harvey

Es una prueba estadística utilizada para evaluar si la varianza de los errores depende de alguna variable explicativa del modelo, pero con una formulación más flexible que ya no requiere que la relación sea estrictamente lineal.

- Etapa 1: Estimar el modelo de regresión y obtener los residuos.

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k + \hat{\varepsilon} \tag{4}$$

- Etapa 2: Obtener el cuadrado de los residuos

$$\hat{\varepsilon}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2 + \cdots + \hat{\alpha}_k X_k + \hat{v} \tag{5}$$

- Prueba de Hipótesis:

$$\begin{aligned} H_0 &: \text{Homocedasticidad p} > 5\,\% \quad \text{No se rechaza la hipótesis nula} \\ H_1 &: \text{Heterocedasticidad p} < 5\,\% \quad \text{Se rechaza la hipótesis nula} \end{aligned}$$

- Calcular el estadístico LM:

$$LM = \frac{SEC}{\Psi_{0,5}} \quad \rightarrow (\text{Ajuste para distribución } \chi^2)$$

Función Gamma:

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$$

- Comparar con la distribución Chi - cuadrada

$$LM \sim \chi^2_{(k)}$$

c. Test de Glejser

Este test se basa en examinar la relación del valor absoluto de los residuos y las variables explicativas del modelo.

- Etapa 1: Estimar el modelo de regresión y obtener los residuos.

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k + \hat{\varepsilon} \tag{6}$$

- Etapa 2: Obtener el cuadrado de los residuos

$$|\hat{\varepsilon}^2| = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2 + \cdots + \hat{\alpha}_k X_k + \hat{v} \tag{7}$$

- Prueba de Hipótesis:

$$H_0 : \text{Homocedasticidad p} > 5\,\% \quad \text{No se rechaza la hipótesis nula}$$
$$H_1 : \text{Heterocedasticidad p} < 5\,\% \quad \text{Se rechaza la hipótesis nula}$$

- Calcular el estadístico LM:

$$LM = \frac{SEC}{\underbrace{\left(1 - \frac{2}{\pi}\right)}_{\text{constante de corrección}} \quad (\hat{\sigma}^2)^2} \quad \rightarrow \quad \in \text{ Etapa II}$$

- Comparar con la distribución Chi - cuadrada

$$LM \sim \chi^2_{(k)}, \quad k : \#\text{ regresores}$$

d. Test de White

Es una prueba estadística general para detectar heterocedasticidad sin necesidad de especificar una forma funcional de cómo varía la varianza del término error.

- Etapa 1: Estimar el modelo de regresión y obtener los residuos.

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k + \hat{\varepsilon} \tag{8}$$

- Etapa 2: Obtener el cuadrado de los residuos

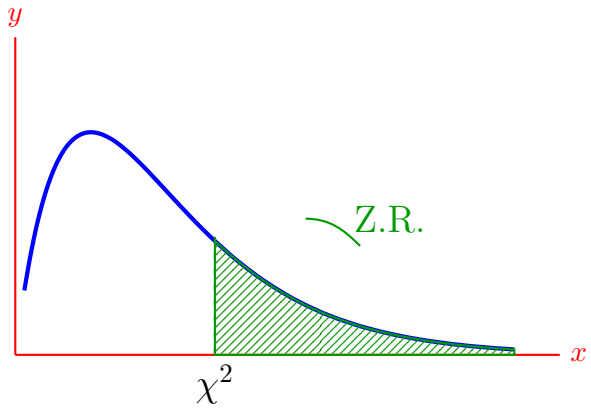
$$\hat{\varepsilon}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2 + \cdots + \hat{\alpha}_k X_k + \hat{v} \tag{9}$$

- Prueba de Hipótesis:

$$H_0 : \text{Homocedasticidad}$$
$$H_1 : \text{Heterocedasticidad}$$

- **Estadístico de prueba:**

$$(\#\text{ observaciones}) \quad R^2 \sim \chi^2_{(k-1)}$$
$$nR^2 \sim \chi^2_{(k-1)}$$



Deficiencia: Este test tiene mayor probabilidad de rechazar la hipótesis nula (Homocedasticidad). Dado que este test considera una mayor cantidad de regresores en la Etapa 2. Si aumenta el número de regresores k entonces aumenta el R^2.

2.4. Aplicación en STATA

Primero, realizamos la detección mediante el test de Breusch-Pagan y visualmente con los residuos.
Utilizar la base de datos *Data1.dta* y el script *Dofileclase2025.do*.

```
* 1. Estimar el modelo original por MCO
regress lnsalario educ edad edad2 urban casado femenino

* 2. Obtener residuos para diagnostico visual
predict uhat, resid
kdensity uhat
```

Listing 1: Detección de Heterocedasticidad

A continuación, se presenta el resultado del test formal de Breusch-Pagan obtenido en Stata:

```
. hettest educ edad edad2 urban casado femenino, iid

   Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: educ edad edad2 urban casado femenino

         chi2(6)      =   34.52
         Prob > chi2  =   0.0000
```

Listing 2: Resultado - Test de Breusch-Pagan

Como el p-valor es 0.0000, rechazamos la hipótesis nula de varianza constante. Por lo tanto, procedemos a corregir el modelo utilizando errores estándar robustos:

```
* 3. Estimacion con correccion de White (errores robustos)
reg lnsalario educ edad edad2 urban casado femenino, robust
```

Listing 3: Solución - Errores Estándar Robustos

A continuación, se presentan los resultados de la estimación del modelo utilizando errores estándar robustos para corregir la pérdida de eficiencia provocada por la heterocedasticidad.

```
------------------------------------------------------------------------------

             |               Robust
   lnsalario | Coefficient  std. err.      t    P>|t|     [95% conf.
   interval]
-------------+----------------------------------------------------------------

        educ |   .0820252   .0018096    45.33   0.000       .0784778
   .0855726
        edad |   .0620389   .0048979    12.67   0.000       .0524376
   .0716402
       edad2 |  -.0006103   .0000647    -9.43   0.000      -.0007372
   -.0004835
       urban |   .1201113   .0142284     8.44   0.000       .0922193
   .1480034
      casado |   .1107829   .0222779     4.97   0.000       .0671115
   .1544543
    femenino |  -.2554387   .0274603    -9.30   0.000      -.3092691
   -.2016083
       _cons |   .6242313   .0766603     8.14   0.000       .4739538
   .7745088
------------------------------------------------------------------------------
```

Listing 4: Resultado - Regresión con Errores Estándar Robustos

Como se observa en el Cuadro 4, aunque los coeficientes se mantienen, los errores estándar han sido ajustados, lo que permite realizar inferencias y pruebas de hipótesis válidas.

3. Autocorrelación

Es la correlación entre los términos de error de diferentes observaciones. Común en series temporales donde el error presente depende de errores pasados.

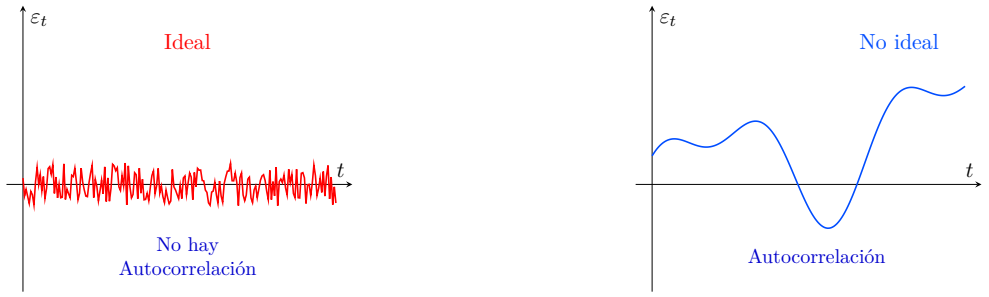


Figura 1: Comparación de residuos en el tiempo.

3.1. Causas

Una de las causas más frecuentes es la omisión de variables relevantes. Cuando el modelo no incluye factores que influyen en la variable dependiente y que además evolucionan en el tiempo, esos efectos terminan acumulándose en el término de error. Si dichos factores presentan cierta persistencia temporal, los errores también comenzarán a mostrarla, generando correlación entre periodos consecutivos.

Otra causa importante es la mala especificación del modelo, especialmente cuando se utiliza una forma funcional incorrecta. Por ejemplo, si el verdadero proceso generador de datos tiene una estructura dinámica —como una relación con rezagos o un comportamiento no lineal— y el modelo estimado no incorpora esos elementos, los residuos pueden reflejar patrones sistemáticos. En estos casos, la autocorrelación no es más que una señal de que el modelo no está capturando adecuadamente la dinámica de la variable.

Asimismo, muchos procesos económicos presentan una inercia natural. Variables como el crecimiento económico, la inflación o el consumo suelen mostrar tendencias, ciclos o persistencia en el tiempo. Esta característica propia de los fenómenos económicos puede generar que los errores de un periodo estén relacionados con los del periodo anterior. Si el modelo no incorpora mecanismos que recojan esa dinámica (por ejemplo, términos autorregresivos o variables rezagadas), la autocorrelación aparece como consecuencia.

3.2. Consecuencias

Lo primero que conviene tener claro es que, si se mantiene la exogeneidad, los estimadores MCO siguen siendo insesgados. Es decir, en promedio, los coeficientes estimados continúan representando correctamente los verdaderos parámetros poblacionales. Sin embargo, dejan de ser eficientes. Esto significa que ya no son los estimadores con menor varianza posible dentro de la clase de estimadores lineales insesgados. En términos más simples, el modelo no está

aprovechando toda la información disponible en la estructura temporal de los datos.

Un efecto especialmente relevante se observa en los errores estándar. Cuando existe autocorrelación, los errores estándar calculados bajo el supuesto clásico suelen estar mal estimados y, en muchos casos, resultan subestimados. Esto genera una falsa sensación de precisión en los coeficientes estimados, como si el modelo fuera más confiable de lo que realmente es.

Como consecuencia directa, las pruebas de hipótesis pueden volverse engañosas. Tanto las pruebas t individuales como la prueba F conjunta dependen de errores estándar correctamente calculados. Si estos están distorsionados, el modelo puede mostrar significancia estadística cuando en realidad no existe. Es decir, podríamos concluir que una variable es relevante cuando en verdad no lo es, lo que afecta la interpretación económica y las decisiones basadas en el modelo.

3.3. Detección y solución

- Detección: Durbin - Watson, Breusch-Godfrey, ACF/PACF.
- Modelado: ARMA, variables rezagadas, diferenciación.
- Corrección: GLS, Cochrane-Orcutt, Prais-Winsten, HAC.

a. Durbin - Watson

Es una prueba estadística que permite detectar autocorrelación de primer orden.

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + u_t \tag{10}$$

- Paso 1: Regresión original
- Paso 2: Calculamos los residuos
- Paso 3: Evaluamos si los residuos siguen un proceso autorregresivo AR(1)

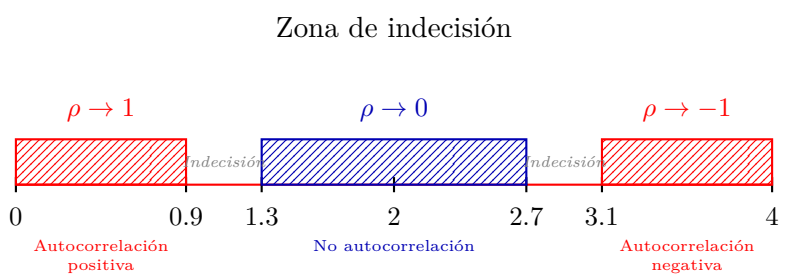
$$u_t = \rho u_{t-1} + \varepsilon_t \tag{11}$$

- Paso 4: Calculamos el estadístico DW

$$DW = \frac{\sum_{t=2}^{n}(u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2} \approx 2(1-\rho) \tag{12}$$

Valor de DW	**Valor de** ρ	**Interpretación**
Próximo a 2	$\rho \approx 0$	**Ideal:** No hay autocorrelación (Residuos aleatorios).
Próximo a 0	$\rho \approx 1$	**No ideal:** Autocorrelación positiva fuerte.
Próximo a 4	$\rho \approx -1$	**No ideal:** Autocorrelación negativa fuerte.

Crítica del Test Durwin - Watson: Solo analiza la correlación de primer orden.



b. Breusch-Godfrey

Es una prubea estadística que sirve para detectar si los errores del modelo están correlacionados entre sí en varios rezagos (autocorrelación serial).

- **Etapa 1: Regresión original**

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + u_t \tag{13}$$

- **Etapa 2: Estimar la regresión auxiliar**

$$\hat{u}_t = \rho_0 + \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_{pt} \tag{14}$$

- **Prueba de Hipótesis:**

  $H_0 : \rho_1 = \rho_2 = \cdots = \rho_p = 0$ (No hay autocorrelación)

  $H_1 : \exists \rho_j \neq 0$ (Existe autocorrelación)

- **Calcular el estadístico LM:**

$$LM = (n-p)R^2_{aux} \sim \chi^2_p$$

- **Regla de Decisión**

  $LM > \chi^2_{1-\alpha,p} \to$ Rechazar H_0 : hay autocorrelación

  $LM \leq \chi^2_{1-\alpha,p} \to$ No rechazar H_0

3.4. Aplicación de Stata

```
* Nota: Para series de tiempo o datos ordenados
* 1. Definir estructura temporal
tsset tiempo

* 2. Ejecutar test de Durbin-Watson tras la regresion
regress lnsalario educ edad edad2 urban casado femenino
. estat dwatson

  Durbin-Watson d-statistic(  7,  3953) = .4523108
```

Listing 5: Test de Autocorrelación (Durbin-Watson)

Fuente: Elaboración propia en base a las sesiones del curso dictado por el profesor William Sanchez (2025-2).

4. Inestabilidad de parámetros

La inestabilidad de parámetros se presenta cuando los coeficientes del modelo no permanecen constantes a lo largo de la muestra, sino que cambian en el tiempo o entre distintos grupos de observaciones. Este supuesto es importante porque el modelo de regresión asume que existe una única relación estructural que describe el proceso generador de datos.

4.1. Causas

Una de las causas más comunes son los cambios estructurales en la economía. Por ejemplo, procesos de apertura comercial, cambios en el modelo productivo o transformaciones en la estructura del mercado pueden alterar la forma en que las variables interactúan entre sí. En estos casos, los coeficientes estimados antes y después del cambio pueden diferir significativamente.

También pueden generar inestabilidad las crisis financieras o políticas, ya que estos eventos suelen modificar el comportamiento de los agentes económicos. Durante una crisis, variables como el consumo, la inversión o el empleo pueden reaccionar de manera distinta a como lo hacían en periodos de estabilidad, lo que implica un cambio en los parámetros del modelo.

Otra causa importante son las reformas regulatorias significativas, como modificaciones tributarias, laborales o monetarias. Estas reformas pueden alterar los incentivos y, por tanto, la relación entre las variables explicativas y la variable dependiente.

Los cambios tecnológicos disruptivos pueden transformar profundamente la dinámica económica. Innovaciones que modifican los procesos productivos o los patrones de consumo pueden hacer que los coeficientes estimados en periodos anteriores ya no representen adecuadamente la nueva realidad.

4.2. Consecuencias

Cuando existe inestabilidad de parámetros, el principal problema es que los coeficientes estimados dejan de representar una relación estructural única y estable. En otras palabras, el modelo puede estar promediando distintas etapas o regímenes en una sola estimación, por lo que los parámetros obtenidos no reflejan con precisión ninguna de ellas. Esto afecta directamente la validez del análisis econométrico.

Una consecuencia importante es que los pronósticos se vuelven poco confiables. Si la relación entre las variables cambia a lo largo del tiempo y el modelo no captura esos cambios, las proyecciones basadas en parámetros inestables pueden resultar imprecisas o incluso engañosas.

Además, la inestabilidad puede conducir a interpretaciones económicas erróneas. Si asumimos que los coeficientes son constantes cuando en realidad no lo son, podríamos extraer conclusiones equivocadas sobre la magnitud o dirección del efecto de una variable sobre otra. En ese sentido, el problema no solo afecta la estimación técnica, sino también el análisis económico que se deriva de ella.

4.3. Detección y solución

Detección:

- Prueba de Chow
- Residuos Recursivos
- Test CUSUM
- Test de Quandt-Andrews

Corrección:

- Segmentación: Dividir la muestra y estimar modelos separados.

- Variables Dummy: Introducir indicadores para períodos diferentes.
- Modelos Tiempo-Variables: Parámetros que cambian gradualmente.
- Ventanas Rodantes: Estimaciones con muestras móviles de tamaño fijo.

a. Pruebas de Tipo Chow

Consideramos el siguiente modelo:

$$y = x\beta + \epsilon$$

Se divide la muestra en subconjuntos independientes y se hace una estimación en cada subconjunto con las mismas variables independientes.

$$n = n_1 + n_2$$
$$n_1 : y = x\beta^1 + \epsilon$$
$$n_2 : y = x\beta^2 + \epsilon$$

Nuestra Hipótesis inicial es:

$$H_0 : \beta^1 = \beta^2$$
$$H_1 : \beta^1 \neq \beta^2$$

- Análisis de Corte Transversal

  El modelo base de corte transversal se define como:
  log w_i[4]

$$\log w_i = \beta_0 + \beta_1 Educ_i + \beta_2 gen_i + u_i \tag{15}$$

Donde la variable *gen* se divide en:

$$gen = \begin{cases} 1 & \text{F (Femenino)} \\ 0 & \text{M (Masculino)} \end{cases}$$

Comparación por grupos:

- **Caso femenino (n^F):**

$$\log w_i = \beta_0^F + \beta_1^F Educ_i + \beta_2^F gen_i + u_i \tag{16}$$

- **Caso masculino (n^M):**

$$\log w_i = \beta_0^M + \beta_1^M Educ_i + \beta_2^M gen_i + u_i \tag{17}$$

Prueba de Hipótesis:

Planteamos las hipótesis para evaluar la estabilidad de los parámetros entre grupos:

$$H_0 : \beta_1^F = \beta_1^M \quad \text{(Estabilidad)}$$
$$H_1 : \beta_1^F \neq \beta_1^M \quad \text{(Inestabilidad)}$$

[4]El uso de logaritmos en la variable dependiente (salario) permite interpretar los coeficientes como semielasticidades, es decir, el cambio porcentual en el salario ante un cambio unitario en la variable explicativa[cite: 395, 397].

Si rechazamos H_0, procedemos con la corrección incorporando un término de interacción:

$$\log w_i = \beta_0 + \beta_1 Educ_i + \beta_2 gen_i + \beta_3 (Educ_i \times gen_i) + u_i \tag{18}$$

El efecto marginal de la educación sobre el logaritmo del salario es:

$$\frac{\partial \log w_i}{\partial Educ} = \beta_1 + \beta_3 gen_i \tag{19}$$

- Series Temporales

  Análisis de Quiebre Estructural: Términos de Intercambio

  El modelo original para la inversión (inv_t) basado en los términos de intercambio (TI_t) es:

$$inv_t = \alpha_0 + \alpha_1 TI_t + \epsilon_t \tag{20}$$

  Análisis por Subperiodos:

  Definimos una variable **Dummy = D1** para identificar los dos periodos de análisis:

  - **Periodo 1 (n_1):** 1994 T1 – 2008 T4

$$inv_t = \alpha_0^1 + \alpha_1^1 TI_t + \epsilon_t$$

  - **Periodo 2 (n_2):** 2008 T4 – 2019 T4

$$inv_t = \alpha_0^2 + \alpha_1^2 TI_t + \epsilon_t$$

  Prueba de Hipótesis:

  Planteamos la estabilidad de los parámetros entre ambos periodos:

$$\begin{aligned} H_0 &: \alpha_1^1 = \alpha_1^2 \quad \text{(Estabilidad)} \\ H_1 &: \alpha_1^1 \neq \alpha_1^2 \quad \text{(Inestabilidad)} \end{aligned}$$

  Si se rechaza la H_0, procedemos con la **corrección**.

  Modelos de Corrección:

  a. Corrección por Variable Dummy (Intercepto):

  Si el quiebre solo afecta al intercepto:

$$inv_t = \alpha_0 + \alpha_1 TI_t + \gamma_1 D_1 \tag{21}$$

  b. Corrección en ambas (Intercepto y Tendencia):

  Para capturar tanto el **Quiebre de Intercepto** como el **Quiebre de Tendencia**, el modelo completo es:

$$inv_t = \alpha_0 + \underbrace{\gamma_1 D_1}_{\text{Quiebre Intercepto}} + \alpha_1 TI_t + \underbrace{\gamma_2 (TI_t \times D_1)}_{\text{Quiebre Tendencia}} + \epsilon_t$$

  Donde el efecto total sobre la tendencia en el segundo periodo está dado por:

$$(\alpha_1 + \gamma_2 D_1) TI_t$$

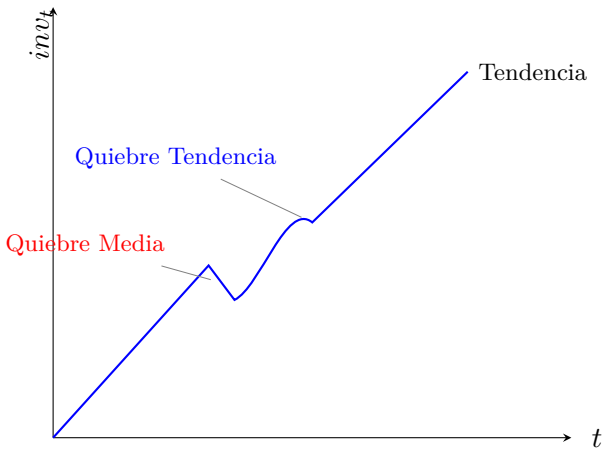


Figura 2: Representación gráfica de Quiebre Estructural en Inversión.

4.4. Aplicación de Stata

Utilizar la base de datos *Data1.dta* y el script *Dofileclase2025.do*.
Para evaluar la estabilidad de los parámetros (por ejemplo, si la educación afecta igual al salario de hombres y mujeres), se implementa el Test de Chow y el uso de términos de interacción.

```
* 1. Estimación para el grupo: Femenino
reg lnsalario educ edad edad2 urban casado viudo femenino if femenino==1
scalar ssrum=e(rss)
scalar dfum=e(df_r)

* 2. Estimación para el grupo: Masculino
reg lnsalario educ edad edad2 urban casado viudo femenino if femenino==0
scalar ssruf=e(rss)
scalar dfuf=e(df_r)

* 3. Cálculo del estadístico de Chow
scalar ssru = ssrum + ssruf
scalar dfu = dfum + dfuf
scalar numer = ((ssrc - ssru) / g)
scalar denom = (ssru / dfu)
scalar Chow_test = numer / denom
display Chow_test
display Ftail(g, dfu, Chow_test)
```

Listing 6: Test de Chow manual

```
. test i1 i2 i3 i4 i5

( 1)  i1 = 0
( 2)  i2 = 0
( 3)  i3 = 0
( 4)  i4 = 0
( 5)  i5 = 0

       F(  5,  3941) =   12.45
            Prob > F =    0.0000
```

Listing 7: Resultado - Test de Wald

5. Multicolinealidad

La multicolinealidad es la alta correlación entre sus variables explicativas y va a dificultar y separar sus efectos sobre la variable dependiente.

5.1. Causas

- Metodológicas: Inclusión de variables redundantes o transformaciones de las mismas variables en el modelo.
- Económicos: Variables correlacionadas por relaciones económicas subyacentes.

5.2. Consecuencias

- Estimadores inestables: Pequeños cambios en los datos pueden causar grandes cambios en los coeficientes estimados.
- Varianza grande: Errores estándar inflados reducen la precisión de las estimaciones.
- Signo Incorrectos: Coeficientes con signos contrarios a la teoría económica o la intuición.
- Pérdida de Significancia: Variables importantes pueden aparecer como no significativas estadísticamente.

5.3. Detección y solución

a. VIF (Factor de inflación de Varianza)

Mide cuantas veces se infla la varianza del estimador debido a la multicolinealidad. Valores mayores indican mayor grado de correlación lineal con otras variables.

$$VIF(j) = \frac{1}{1 - R_j^2} \tag{22}$$

Regla Práctica

Si VIF supera a 10, sugiere multicolinealidad problemática. En caso supere a 5 puede ser preocupante en algunos contextos aplicados.
Aplicación en STATA Utilizar la $qlfs \quad q2 \quad 2016 \quad microdata\_edit\_son.dta$. Disponible en HarvardVerse.

```
* 1. Corres tu regresión
regress lnwage_hours educ age rural

* 2. Calculas los VIF
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
        educ |      1.21    0.829687
         age |      1.19    0.838301
       rural |      1.01    0.986215
-------------+----------------------
    Mean VIF |      1.14

end of do-file
```

Listing 8: Resultado - VIF

b. Correlación Simple

Valores cercanos a 1 (valor absoluto) entre pares de variables indican posible multicolinealidad.

Aplicación en STATA
Utilizar la *qlfs q2 2016 microdata_edit_son.dta*. Disponible en HarvardVerse.

```
* correlacion simple
. correlate lnwage_hours educ age rural
(obs=3,953)

             | lnwage~s     educ       age     rural
-------------+------------------------------------
lnwage_hours |   1.0000
        educ |   0.3891    1.0000
         age |   0.3674    0.4020    1.0000
       rural |  -0.1738   -0.1168   -0.0582    1.0000
.
end of do-file
```

Listing 9: Resultado - Correlación Simple

c. Solución

- Eliminar variables redundantes. Quitar variables altamente correlacionadas cuando sea teóricamente justificable.
- Crear índices compuestos.
- Análisis de componentes principales.
- Aumentar tamaño muestral.

6. No normalidad

6.1. Causas

- Presencia de valores atípicos (outliers).
- Especificación incorrecta del modelo.
- Variables omitidas con distribución no normal.
- Errores de medición en las variables.
- Asimetrías inherentes en los datos económicos

6.2. Consecuencias

- Pruebas t y F no válidas en muestras pequeñas
- Intervalos de confianza incorrectos
- Estimación por máxima verosimilitud ineficiente
- Pronósticos menos confiables

6.3. Detección y solución

a. Test de Jarque - Bera

El test de Jarque–Bera se utiliza para evaluar si una variable aleatoria (o los residuos de un modelo econométrico) se aproxima a una distribución Normal, a partir de sus momentos de asimetría y curtosis.

Paso 1: Planteamiento de hipótesis

Las hipótesis del test son:

$$H_0 : \text{La variable (o los residuos) siguen una distribución Normal}$$
$$H_1 : \text{La variable (o los residuos) no siguen una distribución Normal}$$

Paso 2: Estadístico de prueba

El estadístico de Jarque–Bera se define como:

$$JB = \frac{T-k}{6}\left[S^2 + \frac{(K-3)^2}{4}\right]$$

donde:

- T es el tamaño de la muestra,
- k es el número de regresores del modelo,
- S es el coeficiente de asimetría,
- K es el coeficiente de curtosis.

Bajo el supuesto de normalidad, se cumple que $S = 0$ y $K = 3$.

Paso 3: Distribución del estadístico

Bajo la hipótesis nula, el estadístico JB sigue una distribución chi-cuadrado con dos grados de libertad:

$$JB \sim \chi^2(2)$$

Paso 4: Regla de decisión

Para un nivel de significancia del 5 %, el valor crítico es:

$$\chi^2_{(5\,\%,\,2)} = 5{,}99$$

La regla de decisión es:

- Si $JB < 5{,}99$, no se rechaza la hipótesis nula.
- Si $JB \geq 5{,}99$, se rechaza la hipótesis nula.

Paso 5: Conclusión

Si no se rechaza la hipótesis nula, existe evidencia estadística de que la variable (o los residuos del modelo) se aproximan a una distribución Normal. En caso contrario, se concluye que no se cumple el supuesto de normalidad, lo que puede afectar la validez de la inferencia estadística.

b. Aplicación STATA

Utilizar la $qlfs\quad q2\quad 2016\quad microdata\_edit\_son.dta$. Disponible en HarvardVerse.

```
* 1. Estimar el modelo
regress lnwage_hours educ age rural

* 2. Calcular los residuos
predict r, resid

* 3. Ejecutar el test de normalidad (Jarque-Bera)
. sktest r

                    Skewness/Kurtosis tests for Normality
                                                          ------ joint
   ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob
   >chi2
-------------+-------------------------------------------------------------

           r |      3,953     0.0000         0.0000           .
   0.0000

.
end of do-file
```

Listing 10: Resultado - Test Jarque Bera

7. Endogeneidad

La endogeneidad surge cuando alguna de las variables explicativas está correlacionada con el término de error, es decir,

$$\mathbb{E}(x_i\varepsilon_i) \neq 0,$$

lo que provoca que el estimador de MCO sea sesgado e inconsistente. A continuación, se presentan las principales causas de endogeneidad.

7.1. Causas

A. Sesgo por variables omitidas

El sesgo por variables omitidas ocurre cuando una variable explicativa relevante, correlacionada con alguna de las variables incluidas en el modelo, es excluida del mismo.
Supóngase el siguiente modelo poblacional:

$$y = \alpha + \beta x + \gamma w + \varepsilon,$$

donde x es una variable observada y w es una variable relevante omitida. Si w está correlacionada con x, el estimador MCO de β estará sesgado. La dirección y magnitud del sesgo dependen de:

- La correlación entre x y w.
- El signo y magnitud del coeficiente γ.

	$\text{corr}(x,w) > 0$	$\text{corr}(x,w) < 0$
$\gamma > 0$	Sesgo positivo	Sesgo negativo
$\gamma < 0$	Sesgo negativo	Sesgo positivo

Ejemplo: ecuación de Mincer

Un ejemplo clásico de este problema se presenta en la ecuación de Mincer:

$$\ln(\text{Salario}) = \alpha + \beta\,\text{Educación} + \gamma\,\text{Habilidad} + \varepsilon.$$

La habilidad afecta directamente al salario y, además, está positivamente correlacionada con los años de educación. Si la habilidad no se incluye en el modelo, el estimador de β captará parcialmente su efecto, generando sesgo por variable omitida.

B. Simultaneidad o causalidad inversa

La simultaneidad se refiere a situaciones en las que la relación causal entre dos variables es bidireccional, es decir, cuando x afecta a y y, al mismo tiempo, y afecta a x.
En este caso, la variable explicativa resulta endógena, ya que se determina simultáneamente con la variable dependiente, lo que implica:

$$\mathbb{E}(x_i\varepsilon_i) \neq 0.$$

Ejemplo

Un ejemplo típico de simultaneidad se encuentra en la relación entre el PBI per cápita y la calidad de las instituciones: economías con mejores instituciones tienden a crecer más, pero un mayor nivel de ingreso también permite mejorar la calidad institucional.

C. Error de medición y sesgo de atenuación

Otra fuente importante de endogeneidad es el error de medición en las variables explicativas. Supóngase que la variable verdadera x^* no es observable y que, en su lugar, el investigador observa:

$$x = x^* + u,$$

donde u es un error de medición clásico, caracterizado por tener media cero, ser independiente de x^* y no estar correlacionado con el término de error estructural ε.
Considérese el modelo verdadero:

$$y = \alpha + \beta x^* + \varepsilon.$$

Si el modelo se estima utilizando la variable observada x en lugar de x^*, el término de error efectivo se convierte en $\varepsilon + \beta u$, lo que induce correlación entre el regresor observado y el error del modelo, violando el supuesto de exogeneidad requerido por el método de Mínimos Cuadrados Ordinarios (MCO).
El estimador MCO del parámetro β puede expresarse como:

$$\hat{\beta} = \frac{\text{Cov}(y, x)}{\text{Var}(x)}.$$

Bajo los supuestos del error de medición clásico, se tiene que:

$$\text{Cov}(y, x) = \beta\text{Var}(x^*), \qquad \text{Var}(x) = \text{Var}(x^*) + \text{Var}(u).$$

Por lo tanto, el valor esperado del estimador resulta:

$$\mathbb{E}[\hat{\beta}] = \beta\frac{\text{Var}(x^*)}{\text{Var}(x^*) + \text{Var}(u)}.$$

El factor de atenuación
Esta medida se define mediante la siguiente expresión:

$$\lambda = \frac{\text{Var}(x^*)}{\text{Var}(x^*) + \text{Var}(u)}, \qquad 0 < \lambda < 1$$

Este parámetro resume la magnitud del sesgo inducido por el error de medición. Con esta notación, el resultado fundamental puede escribirse como:

$$\mathbb{E}[\hat{\beta}] = \lambda\beta.$$

El estimador esperado es proporcional al parámetro verdadero, pero reducido por un factor estrictamente menor que uno.

Límites del factor
El comportamiento de λ permite interpretar económicamente el problema:

- Si $\text{Var}(u) \to 0$, entonces $\lambda \to 1$: no hay error de medición y el estimador es insesgado.
- A medida que $\text{Var}(u)$ aumenta, λ disminuye: el sesgo se vuelve más severo.

Por tanto, la precisión en la medición de la variable explicativa determina directamente la magnitud del sesgo.

Interpretación del sesgo de atenuación
El rasgo fundamental del sesgo de atenuación es que **siempre empuja el estimador hacia cero**, independientemente del signo del parámetro verdadero:

- Si $\beta > 0$, entonces $\hat{\beta}$ se aproxima a cero desde valores positivos: el efecto positivo verdadero es subestimado.
- Si $\beta < 0$, entonces $\hat{\beta}$ se aproxima a cero desde valores negativos: el efecto negativo verdadero también es subestimado en magnitud.

El sesgo no altera necesariamente el signo, pero sí reduce sistemáticamente la magnitud estimada del efecto.

Consecuencia econométrica
La consecuencia directa es que el estimador MCO **subestima sistemáticamente** el efecto verdadero de x^* sobre y. Este fenómeno se conoce como *sesgo de atenuación*, porque el impacto estimado de la variable explicativa se reduce artificialmente debido al ruido en su medición.

7.2. Consecuencias

- Estimadores inconsistentes: No convergen al valor verdadero ni con muestras infinitas.
- Inferencia No válida: Tests de hipótesis y p-valores incorrectos.
- Causalidad Distorsionada: Conclusiones erróneas sobre relaciones causales.
- Predicciones imprecisas: Modelos que generan pronósticos poco confiables.

7.3. Detección y solución de la endogeneidad

a. Test de Hausman

La prueba de Hausman compara dos estimadores del mismo parámetro:

- un estimador consistente bajo la hipótesis nula y alternativa (por ejemplo, Variables Instrumentales — VI), y
- un estimador eficiente pero consistente solo bajo la hipótesis nula (por ejemplo, MCO).

La lógica de la prueba es que, si los regresores son exógenos, ambos estimadores deben converger al mismo valor poblacional. Sin embargo, si existe endogeneidad, el estimador MCO se vuelve inconsistente, mientras que el estimador VI permanece consistente, generándose una diferencia sistemática entre ambos.
Las hipótesis se formulan como:

$$H_0 : \text{Los regresores son exógenos}$$

$$H_1 : \text{Al menos un regresor es endógeno}$$

Una diferencia estadísticamente significativa entre los estimadores lleva al rechazo de H_0, lo que constituye evidencia de endogeneidad.
Esta prueba se utiliza cuando ya se dispone de un instrumento válido, es decir, una variable correlacionada con el regresor endógeno pero no con el término de error estructural.

Aplicación en STATA Utilizar la *qlfs q2 2016 microdata_edit_son.dta*. Disponible en HarvardVerse.

```
1
2  * 1. Estimar IV
3  ivregress 2sls lnwage_hours age  rural (educ = father_educ)
4  estimates store modelo_iv
5
6  * 2. Estimar MCO
7  regress lnwage_hours educ age  rural
8  estimates store modelo_mco
9
10 * 3. Comparar (siempre se pone primero el modelo consistente/IV)
11 . hausman modelo_iv modelo_mco
12
13                 ---- Coefficients ----
14              |      (b)          (B)            (b-B)     sqrt(diag(V_b-
       V_B))
15              |   modelo_iv    modelo_mco     Difference          S.E.
16 -------------+------------------------------------------------------------
17         educ |    .0717511     .0372662        .0344849        .0050935
18          age |    .0122489     .0204729        -.008224        .0012509
19        rural |    -.106342     -.130694         .024352        .0051924
20 ------------------------------------------------------------------------------
21                       b = consistent under Ho and Ha; obtained from
       ivregress
22             B = inconsistent under Ha, efficient under Ho; obtained from
       regress
23
24     Test:  Ho:  difference in coefficients not systematic
25
26                   chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
27                           =       45.84
28                 Prob>chi2 =      0.0000
29 end of do-file
```

Listing 11: Resultado - Test Hausman

b. Test de Durbin–Wu–Hausman (DWH)

El test de Durbin–Wu–Hausman es una variante operativa y robusta de la prueba de Hausman, diseñada específicamente para contrastar endogeneidad en modelos con variables instrumentales. El procedimiento típico es el siguiente:

1. Se estima la primera etapa, donde el regresor potencialmente endógeno se explica mediante los instrumentos y las variables exógenas.
2. Se obtienen los residuos de esta regresión.
3. Se incluyen dichos residuos como regresores adicionales en la ecuación estructural original.

Si el coeficiente asociado a esos residuos es estadísticamente significativo, se concluye que el regresor está correlacionado con el error estructural, lo que confirma la presencia de endogeneidad.
Formalmente, el contraste mantiene las mismas hipótesis:

$$H_0 : \text{Los regresores son exógenos}$$

$$H_1 : \text{Existe endogeneidad}$$

En la práctica empírica, este test es ampliamente implementado en software econométrico. Por ejemplo, en Stata puede aplicarse después de una regresión de variables instrumentales mediante comandos específicos para contrastar endogeneidad.

Aplicación en STATA Utilizar la $qlfs \quad q2 \quad 2016 \quad microdata\_edit\_son.dta$. Disponible en HarvardVerse.

```
* Primero estimas tu modelo IV
ivregress 2sls lnwage_hours age rural (educ = father_educ)

* Luego aplicas el test de Durbin-Wu-Hausman
. estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Durbin (score) chi2(1)          =  49.5334  (p = 0.0000)
  Wu-Hausman F(1,3948)            =  50.0985  (p = 0.0000)
```

Listing 12: Resultado - Test Durbin Wu Hausman

c. Solución

- Variables instrumentales (VI): Utiliza variables que afectan X pero no están correlacionadas con el error.
- Mínimos Cuadrados en Dos Etapas (MC2E): Implementación práctica del método VI mediante regresiones secuenciales.
- Método de Momentos Generalizados (GMM): Generalización que permite múltiples instrumentos y heterocedasticidad.
- Ecuaciones Estructurales: Modela explícitamente las relaciones simultáneas entre variables.

Referencias

[Greene, 1997] Greene, W. H. (1997). *Análisis Econométrico*. Prentice Hall, 3ra edition.

[Greene, 2003] Greene, W. H. (2003). *Econometric Analysis*. Prentice Hall, 5th edition.

[Johnston and DiNardo, 1997] Johnston, J. and DiNardo, J. (1997). *Econometric Methods*. McGraw-Hill, 4th edition.

[Sanchez, 2025] Sanchez, W. (2025). Material práctico de econometría: Data1.dta y Dofileclase2025.do. Carpeta de recursos académicos. Disponible en: `https://drive.google.com/drive/folders/17ELTptYpTnVGqA4n9qhCuN4U8p0auvS_?usp=sharing`.

[Wooldridge, 2010] Wooldridge, J. M. (2010). *Introducción a la econometría. Un enfoque moderno*. Cengage Learning, 4 edition.