UNIDAD 4

Violaciones a los supuestos del modelo clásico

PAPIME
PE300520

Coordinador
Ing. Alejandro Augusto Pérez Pascual

Participantes
Lic. Claudia Jacqueline Bribiesca Silva,
Mtro. Ignacio Cruz López
Mtra. Hortensia Martínez Valdez
Dr. Jorge Pablo Rivas Díaz

Contenido

1. EVALUACIÓN DIAGNÓSTICA

Lee con atención y responde si los siguientes enunciados son verdaderos o falsos.

1. En el modelo clásico de regresión lineal, se establece que el valor medio de la perturbación u_i es igual a cero.

Verdadero Falso

2. En el modelo clásico de regresión lineal, se establece que la varianza del término de error, o de perturbación, cambia conforme lo hace el valor de X.

Verdadero Falso

3. En el modelo clásico de regresión lineal, se establece que no hay autocorrelación entre las perturbaciones.

Verdadero Falso

4. El problema de multicolinealidad consiste en la existencia de relaciones lineales entre dos o más variables independientes del modelo lineal uniecuacional múltiple.

Verdadero Falso

5. La heteroscedasticidad significa que la varianza condicional de Y_i, condicional a las X_i, permanece igual sin importar los valores que tome X_i.

Verdadero Falso

6. La autocorrelación se refiere a la correlación entre los elementos de series de observaciones ordenadas en el tiempo o en el espacio.

Verdadero Falso

2. PRESENTACIÓN DE LA UNIDAD

OBJETIVO PARTICULAR

Al finalizar el estudio de unidad 4, el alumno será capaz de identificar violaciones a los supuestos del modelo clásico de regresión para aplicar algún método de corrección.

TEMARIO

- 4.1 Problema de Multicolinealidad.
- 4.2 Problema de heteroscedasticidad.
    - 4.2.1 Detección, consecuencias y solución.
    - 4.2.2 Uso de deflactores.
    - 4.2.3 Pruebas de la forma funcional lineal contra log-lineal.
    - 4.2.4 Prueba de WHITE de heterocedasticidad.
- 4.3 Correlación.
    - 4.3.1 Prueba Durbin Watson.
    - 4.3.2 Prueba LM.
- 4.3.3 Modelo ARCH y correlación serial.

PRESENTACIÓN

La obtención de un modelo, por el método de MCO, no garantiza que dicho modelo sea útil a los fines de la econometría, pese a que cumpla con las hipótesis de significancia estadística de los estimadores y la significancia estadística del modelo en su conjunto. Será necesario someter a pruebas adicionales al modelo obtenido.

Existen una serie de supuestos sobre el modelo de regresión lineal general o normal (MRLN), cada autor pudiera considera un conjunto de 8 o 10 supuestos, aquí se considerarán los 10 supuestos propuestos por Gujarati y Porter (2010):

1) El modelo de regresión es lineal en los parámetros.
2) Los valores de las regresoras, las *X*, son fijos, o los valores de X son independientes del término de error. Este supuesto hace referencia a la covarianza entre u_i y cada variable *X* debe ser cero.
3) Para las *X* dadas, el valor esperado de la perturbación u_i es cero.
4) Para las *X* dadas, la varianza de u_i es constante u homoscedástica.
5) Para las *X* dadas, no hay autocorrelación, o correlación serial, entre las perturbaciones.
6) El número de observaciones n debe ser mayor al número de parámetros a estimar.
7) Debe existir variación suficiente entre los valores de las variables *X*.
8) No hay colinealidad exacta entre las variables *X*.
9) No existe sesgo de especificación del modelo, hace referencia a que el modelo está correctamente especificado.
10) El término de perturbación u_i está normalmente distribuido.

Algunos de los supuestos se establecen desde antes de la regresión, pero otros deben probarse, en la medida de lo posible. En términos generales, se presentan tres grandes problemas que generan el incumplimiento de algunos de los supuestos, a saber: la multicolinealidad, la heteroscedasticidad y la autocorrelación.

En esta unidad se analiza cada uno de estos problemas a fin de responder a las siguientes cuestiones: ¿Qué origina el problema?, ¿cómo podemos detectarlo en el modelo?, ¿cuáles son las consecuencias prácticas de la presencia del problema? y, finalmente, ¿cómo podríamos corregir el problema?

En principio, se analiza el problema de la multicolinealidad, si bien muchos autores no lo consideran un problema a resolver, otros le dedican varias cuartillas para su estudio. En esta sección se analizan las dos perspectivas para decidir lo que más convenga a la investigación.

Después se analiza el problema de la heteroscedasticidad, el cual tiene diversas implicaciones en la estimación del modelo, pues tiene que ver directamente con la varianza de los errores y, a su vez, con los errores estándar de los estimadores y del modelo. Si bien este problema suele presentarse en modelos de corte transversal, también se presentan situaciones con series de tiempo.

Finalmente, la unidad cierra con el análisis de autocorrelación, este problema suele presentarse con mayor frecuencia en los modelos de series temporales, pero no de forma exclusiva. Si bien existen otros supuestos y deben probarse, el objetivo de la unidad y del curso queda rebasado a ello.

Vale la pena hacer énfasis en que las alternativas de solución no son del todo satisfactorias, pues el "solucionar" un problema podría generar otro. Además de que pudieran presentarse más de una posible solución y no es sencillo definir cuál de ellas pudiera ser la más efectiva. Otra situación es que el modelo pudiera presentar más de un problema, por lo que nos enfrentaremos a la imposibilidad de contar con una prueba capaz de detectar todos los problemas al mismo tiempo, así como tampoco es posible darles solución a todos los problemas al unísono.

Violaciones a los supuestos del modelo clásico

En múltiples ocasiones el análisis de economía aplicada comienza con la siguiente premisa: Y y X son dos variables, el interés se centra en explicar Y en función de X, o en estudiar cómo Y varía con cambios en X.

Para explicar tal relación es necesario considerar tres aspectos:

1. Nunca hay una relación exacta entre las variables.
2. Se requiere especificar una relación funcional entre Y y X.
3. Se debe garantizar una relación *ceteris paribus* entre Y y X.

En el modelo de regresión lineal, la relación entre la variable a explicar (Y), y una variable explicativa (X) se expresa de la siguiente manera:
$Y = \beta_0 + \beta_1 X + u$

donde u representa el término de error, β_0 y β_1 los parámetros poblacionales.

Un supuesto crucial es que la esperanza del error, condicional en X, es cero, es decir, $E(u|X) = 0$. Sin embargo, existen fuentes potenciales de violaciones al supuesto de media condicional cero, por ejemplo: a) un modelo incorrectamente especificado, b) variables explicativas correlacionadas con variables omitidas, o c) errores de medida en las variables explicativas.

Es claro que la mayoría de las relaciones económicas involucran a más de dos variables. En la práctica se van a incorporar tantos controles como se crea necesario: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k K_k + u$

Un modelo de regresión lineal múltiple es más adecuado para el análisis *ceteris paribus*, ya que permite controlar explícitamente muchos de los factores que afectan simultáneamente a la variable dependiente.

Del mismo modo que en la regresión lineal simple, un supuesto crucial es que la esperanza del error, condicional en todas las X, es cero, es decir, $E(u|X_1, X_2, \ldots, X_k) = 0$

Se recomienda profundizar en las temáticas de la unidad, a través de la revisión de los materiales del curso y de la siguiente bibliografía, disponible desde la BiDi-UNAM:

- Gujarati, D y Porter, D. (2010) Capítulo 10 Multicolinealidad: ¿qué pasa si las regresoras están correlacionadas?, Capítulo 11, Heteroscedasticidad: ¿qué pasa si la varianza del error no es constante? y Capítulo 12 Autocorrelación: ¿qué pasa si los términos de error están correlacionados? En Econometría. México: McGraw-Hill.
- Pindick, R. y Rubinfeld, D. (2001) Capítulo 6 Correlación lineal y Heteroscedasticidad. En Econometría: Modelos y Pronósticos. México: McGraw-Hill Interamericana.

Posteriormente, realiza las actividades 1 a 6 siguiendo las instrucciones al pie de la letra.

3. UNIDAD 4 - ACTIVIDAD 1

En esta actividad el alumno describirá el problema de la multicolinealidad.

Objetivo: Explicar qué es y cuáles son las fuentes de la multicolinealidad para establecer un método que solucione el problema.

Instrucciones: con base en los siguientes textos:

- Gujarati, D. & Porter, D. (2010) Capítulo 10 Multicolinealidad: ¿qué pasa si las regresoras están correlacionadas? En Econometría. México: McGraw-Hill.
- Wooldridge, Jeffrey M. (2015) Capítulo 3 Análisis de regresión múltiple. En Introducción a la Econometría. Quinta Edición. Cengage.

1. Elabora un ***texto descriptivo*** con la siguiente estructura:

   A. Introducción: redactar en un párrafo el tema que será descrito.
   B. Desarrollo: en esta sección se deberá responder las siguientes preguntas:

      i. ¿Qué es la multicolinealidad y qué la origina?
      ii. ¿Cuáles son las consecuencias prácticas de la multicolinealidad?
      iii. ¿Cómo se detecta la multicolinealidad?
      iv. ¿De qué forma se puede corregir la multicolinealidad?

   C. Conclusión: con tus propias palabras explica qué es la multicolinealidad y cómo se soluciona dicho problema.

2. El texto debe tener la estructura de un trabajo académico: carátula, Introducción, desarrollo, conclusiones, citas, notas, referencias y fuentes de consulta en formato APA.

3. La extensión máxima del texto es de 12 cuartillas.

4. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Describe qué es la multicolinealidad y cuáles son sus fuentes	2		
2	Explica las consecuencias de la multicolinealidad	2		
3	Explica las pruebas para detectar la multicolinealidad	3		
4	Explica cómo se corrige el problema de multicolinealidad	2		
5	Recupera información de fuentes adicionales	0.35		
6	La actividad cumple el formato solicitado (máximo 12 diapositivas)	0.35		
7	Se envía la actividad en tiempo	0.30		
	Puntaje total	10		

4. UNIDAD 4 - ACTIVIDAD 2

En esta actividad se identificará si los datos presentan multicolinealidad.

Objetivo: utilizar alguno de los métodos para detectar multicolinealidad para decidir qué hacer con el problema.

Instrucciones:

1. Instala el programa de procesamiento econométrico E-Views, R, R Studio o en su defecto utiliza Excel.
2. Descargue el archivo titulado "***Ejercicio 4.1***", considere la tabla que proporciona datos sobre las importaciones de los Estados Unidos, el PIB y el IPC de 1975 a 2005.
   a) Considere el siguiente modelo: $lnImportaciones_t = \beta_1 + \beta_2 lnPIB_t + \beta_3 lnIPC_t + u_t$
   Utilice Excel, EViews o R para estimar los parámetros de este modelo.
   b) ¿Hay sospecha de multicolinealidad en los datos?
   c) Efectúe las siguientes regresiones:
      i. $lnImportaciones_t = A_1 + A_2 lnPIB_t$
      ii. $lnImportaciones_t = B_1 + B_2 lnIPC_t$
      iii. $lnPIB_t = C_1 + C_2 lnIPC_t$

   Con base en estas regresiones, ¿qué puede decir sobre la naturaleza de la multicolinealidad en los datos?
3. Suponga que existe multicolinealidad en los datos, pero que $\hat{\beta}_2$ y $\hat{\beta}_3$ son significativos individualmente en el nivel de 5%, y que la prueba global F es también significativa. En este caso, ¿debe preocupar el problema de colinealidad?
4. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Efectúa de manera correcta la regresión: $lnImportaciones_t = \beta_1 + \beta_2 lnPIB_t + \beta_3 lnIPC_t + u_t$ Y determina si hay multicolinealidad en los datos.	3		
2	Efectúe de manera correcta las regresiones: 1) $lnImportaciones_t = A_1 + A_2 lnPIB_t$ 2) $lnImportaciones_t = B_1 + B_2 lnIPC_t$ 3) $lnPIB_t = C_1 + C_2 lnIPC_t$ Y comenta el problema de la multicolinealidad en los datos.	4		
3	Responde correctamente al planteamiento: Suponga que existe multicolinealidad en los datos, pero que $\hat{\beta}_2$ y $\hat{\beta}_3$ son significativos individualmente en el nivel de 5%, y que la prueba global F es también significativa. En este caso, ¿debe preocupar el problema de colinealidad?	2.5		
4	Se envía la actividad en tiempo	0.5		
	Puntaje total	10		

5. UNIDAD 4 - ACTIVIDAD 3

En esta actividad el alumno describirá el problema de la heteroscedasticidad.

Objetivo: Explicar qué es y cuáles son las fuentes de la heteroscedasticidad para establecer un método que solucione el problema.

Instrucciones: con base en los siguientes textos:

- Gujarati, D. & Porter, D. (2010) Capítulo 11 Heteroscedasticidad: ¿qué pasa si la varianza del error no es constante? En Econometría. México: McGraw-Hill.
- Wooldridge, Jeffrey M. (2015) Capítulo 8 Heteroscedasticidad. En Introducción a la Econometría. Quinta Edición. Cengage.

1. Elabora un ***texto descriptivo*** con la siguiente estructura:

   A. Introducción: redactar en un párrafo el tema que será descrito.
   B. Desarrollo: en esta sección se deberá responder las siguientes preguntas:

      i. ¿Qué es la heterocedasticidad y por qué motivos se presenta?
      ii. ¿Cuáles son las consecuencias de utilizar MCO en presencia de heterocedasticidad?
      iii. ¿Cómo se detecta la heterocedasticidad?
      iv. ¿De qué forma se puede corregir la heterocedasticidad?

   C. Conclusión: con tus propias palabras explica qué es la heteroscedasticidad y cómo se soluciona dicho problema.

2. El texto debe tener la estructura de un trabajo académico: carátula, Introducción, desarrollo, conclusiones, citas, notas, referencias y fuentes de consulta en formato APA.

3. La extensión máxima del texto es de 12 cuartillas.

4. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Describe qué es la heteroscedasticidad y cuáles son sus fuentes	2		
2	Explica las consecuencias de la estimación de MCO en presencia de la heteroscedasticidad	2		
3	Explica las pruebas para detectar la heteroscedasticidad	3		
4	Explica cómo se corrige el problema de la heteroscedasticidad	2		
5	Recupera información de fuentes adicionales	0.35		
6	La actividad cumple el formato solicitado (máximo 12 diapositivas)	0.35		
7	Se envía la actividad en tiempo	0.30		
	Puntaje total	10		

6. UNIDAD 4 - ACTIVIDAD 4

En esta actividad se identificará si los datos presentan heteroscedasticidad.

Objetivo: utilizar alguno de los métodos para detectar heteroscedasticidad para decidir qué hacer con el problema.

Instrucciones:

1. Descargue el archivo titulado "***Ejercicio 4.2***", considere la tabla que proporciona datos sobre 81 automóviles respecto de su MPG (millas promedio por galón), CF (caballos de fuerza de su motor), VOL (pies cúbicos de su cabina), VM (velocidad máxima en millas por hora) y su PS (peso del vehículo en cientos de lb).

   a) Considere el siguiente modelo: $MPG_i = \beta_1 + \beta_2 VM_i + \beta_3 CF_i + \beta_4 PS_i + u_i$

      Utilice Excel, EViews o R para estimar los parámetros de este modelo. Desde el punto de vista económico, ¿tiene sentido?

   b) ¿Esperaría que la varianza del error en el modelo anterior sea heteroscedástica? ¿Por qué?
   c) Con la prueba de White determine si la varianza de error es heteroscedástica.

2. Obtenga los errores estándar de White consistentes con la heteroscedasticidad, así como los valores t, y compare los resultados con los obtenidos mediante MCO.

3. Si se establece heteroscedasticidad, ¿cómo puede transformar los datos de manera que en los datos transformados la varianza del error sea homoscedástica?

5. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Efectúa de manera correcta la regresión: $MPG_i = \beta_1 + \beta_2 VM_i + \beta_3 CF_i + \beta_4 PS_i + u_i$ Responde si los parámetros tienen sentido desde el punto de vista económico.	2		
2	Aplica el método de White y determina si hay heteroscedasticidad.	2		
3	Obtiene los errores estándar de White consistentes con la heteroscedasticidad, así como los valores t, y los compara con los resultados obtenidos mediante MCO.	2.5		
4	Responde cómo puede transformar los datos de manera que en los datos transformados la varianza del error sea homoscedástica.	3		
5	Se envía la actividad en tiempo.	0.5		
	Puntaje total	10		

7. UNIDAD 4 - ACTIVIDAD 5

En esta actividad el alumno describirá el problema de la autocorrelación.

Objetivo: Explicar qué es la autocorrelación para establecer un método que solucione el problema.

Instrucciones: con base en los siguientes textos:

- Gujarati, D. & Porter, D. (2010) Capítulo 12 Autocorrelación: ¿qué pasa si los términos de error están correlacionados? En Econometría. México: McGraw-Hill.
- Wooldridge, Jeffrey M. (2015) Capítulo 12 Correlación serial y heteroscedasticidad en regresiones de series de tiempo. En Introducción a la Econometría. Quinta Edición. Cengage.

1. Elabora un ***texto descriptivo*** con la siguiente estructura:

A. Introducción: redactar en un párrafo el tema que será descrito.
B. Desarrollo: en esta sección se deberá responder las siguientes preguntas:

I. ¿Qué es la autocorrelación y qué la ocasiona?
II. ¿Cuáles son las consecuencias de la estimación de MCO en presencia de la autocorrelación?
III. ¿Cómo se detecta la autocorrelación? Pruebas de Durbin-Watson y Breusch-Godfrey
IV. ¿De qué forma se puede corregir la autocorrelación?

C. Conclusión: con tus propias palabras explica qué es la heteroscedasticidad y cómo se soluciona dicho problema.

2. El texto debe tener la estructura de un trabajo académico: carátula, Introducción, desarrollo, conclusiones, citas, notas, referencias y fuentes de consulta en formato APA.

3. La extensión máxima del texto es de 12 cuartillas.

4. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Describe qué es la correlación y por qué se presenta	2		
2	Explica las consecuencias de la estimación de MCO en presencia de correlación	2		
3	Explica las pruebas para detectar la autocorrelación	3		
4	Explica cómo se corrige el problema de la autocorrelación	2		
5	Recupera información de fuentes adicionales	0.35		
6	La actividad cumple el formato solicitado (máximo 12 diapositivas)	0.35		
7	Se envía la actividad en tiempo	0.30		
	Puntaje total	10		

8. UNIDAD 3 - ACTIVIDAD 6

En esta actividad se identificará si los datos presentan autocorrelación.

Objetivo: utilizar alguno de los métodos para detectar autocorrelación para decidir qué hacer con el problema.

Instrucciones:
1. Descargue el archivo titulado "***Ejercicio 4.3***", considere la tabla que proporciona datos sobre los inventarios y ventas en la industria manufacturera de Estados Unidos, 1950-1991 (millones de dólares).

a) Considere el siguiente modelo: $Y_t = \beta_1 + \beta_2 X_t + u_t$

Donde Y = inventarios y X = ventas, ambas medidas en miles de millones de dólares.

Utilice Excel, EViews o R para estimar los parámetros de este modelo.

b) Con los residuos estimados, investigue si hay autocorrelación positiva mediante:

i. La prueba de Durbin-Watson.
ii. La prueba de Breusch-Godfrey o ML.

2. Con base en los resultados de esta prueba, ¿cómo transformaría los datos para eliminar la autocorrelación? Muestre todos sus cálculos.

3. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Efectúa de manera correcta la regresión: $Y_t = \beta_1 + \beta_2 X_t + u_t$	2.5		
2	Aplica la prueba Durbin-Watson para determinar si hay autocorrelación.	2		
3	Aplica la prueba de Breusch-Godfrey o LM para determinar si hay autocorrelación.	2		
4	Propone una transformación para corregir el problema de autocorrelación.	3		
5	Se envía la actividad en tiempo.	0.5		
	Puntaje total	10		

9. FORMULARIO

MULTICOLINEALIDAD

Factor inflacionario de la varianza (FIV): $FIV: \frac{1}{(1-r_{23}^2)}$

Estimadores de MCO con varianzas y covarianzas grandes:

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2} FIV$$

$$var(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2} FIV$$

$$cov(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}} FIV$$

HETEROSCEDASTICIDAD

Estimadores de Mínimos Cuadrados Generalizados (MCG)

Supone que se conoce la varianza heteroscedástica (σ_i^2)

Modelo transformado: $Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^*$

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}$$

$$var(\hat{\beta}_2^*) = \frac{\sum w_i}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}$$

Donde: $w_i = \frac{1}{\sigma_i^2}$

Prueba de Park

Forma funcional: $ln\sigma_i^2 = ln\sigma^2 + \beta lnX_i + v_i$

Donde v_i es el término de perturbación estocástico.

Como σ_i^2 generalmente no se conoce, Park sugiere utilizar $\hat{u}_i^2$ como aproximación para correr la regresión: $ln\hat{u}_i^2 = ln\sigma^2 + \beta lnX_i + v_i$

Prueba de White

Pasos que seguir para desarrollar la prueba:

1. Estimar el modelo de regresión original mediante MCO, para obtener los residuos $\hat{u}_i$, por ejemplo, $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$
2. Realizar la siguiente regresión auxiliar, para obtener el valor de R^2:
   $\hat{u}_i^2 = \propto_1 + \propto_2 X_{2i} + \propto_3 X_{3i} + \propto_4 X_{2i}^2 + \propto_5 X_{3i}^2 + \propto_6 X_{2i}X_{3i} + v_i$
3. Se establecen las hipótesis a contrastar:
   $H_0: Homoscedasticidad$ $\qquad$ $H_1: Heteroscedasticidad$

Donde: $nR^2 \sim \chi^2_{gl}$

4. Si el valor de ji cuadrada obtenido en el paso 3 excede el valor de ji cuadrada crítico, en el nivel de significancia seleccionado, la conclusión es que hay heteroscedasticidad.

CORRELACIÓN

Esquema autorregresivo de primer orden de Markov, AR(1):

$$u_t = \rho u_{t-1} + \varepsilon_y; \qquad -1 < \rho < 1$$

Donde ρ es el coeficiente de autocorrelación del rezago 1 y ε_t es la perturbación estocástica que satisface los supuestos habituales de MCO.

Con el esquema AR(1) se demuestra que:

$$var(u_t) = E(u_t^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

$$cov(u_t, u_{t+s}) = E(u_t u_{t-s}) = \rho^s \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

$$cor(u_t, u_{t+s}) = \rho^s$$

$$var(\hat{\beta}_2)_{AR1} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \cdots + 2\rho^{n-1}\frac{x_1 x_n}{\sum x_t^2}\right]$$

Estimador MELI en presencia de autocorrelación:

$$\hat{\beta}_2^{MCG} = \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + C$$

$$var\hat{\beta}_2^{MCG} = \frac{\sigma^2}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + D$$

Prueba *d* de Durbin-Watson

$$d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2}$$

Regla de decisión:

Hipótesis nula	Decisión	Si
No hay autocorrelación positiva	Rechazar	$0 < d < d_L$
No hay autocorrelación positiva	Sin decisión	$d_L \leq d \leq d_U$
No hay correlación negativa	Rechazar	$4 - d_L < d < 4$
No hay correlación negativa	Sin decisión	$4 - d_U \leq d \leq 4 - d_L$
No hay autocorrelación positiva o negativa	No rechazar	$d_U < d < 4 - d_U$

Prueba Breusch-Godfrey BG (Prueba LM)

Partiendo del modelo de regresión de 2 variables, donde se pueden incluir valores rezagados: $Y_t = \beta_1 + \beta_2 X_t + u_t$

El término de error $\hat{u}_t$ sigue un esquema autorregresivo de orden ρ, $AR(\rho)$:

$$\hat{u}_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_\rho u_{t-\rho} + \varepsilon_t$$

La hipótesis nula, H_0, por demostrar es: H_0: $\rho_1 = \rho_2 = \cdots = \rho_\rho = 0$

Pasos para desarrollar la prueba BG o LM:

1. Estime el modelo de regresión de dos variables mediante MCO y obtenga los residuos $\hat{u}_t$
2. Haga la regresión $\hat{u}_t$ sobre la X_t original y $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-p}$. En resumen, realice la regresión: $\hat{u}_t = \propto_1 + \propto_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_\rho \hat{u}_{t-\rho} + \varepsilon_t$; y obtenga R^2 de esta regresión (auxiliar).
3. Si el tamaño de la muestra es grande, Breusch y Godfrey demostraron que: $(n-\rho)R^2 \sim \chi^2_\rho$

Es decir, asintóticamente, $(n-\rho)$ veces el valor de R^2 obtenido en la regresión auxiliar sigue la distribución ji cuadrada con ρ gl. Si en una aplicación $(n-\rho)R^2$ excede el valor crítico ji cuadrada, en el nivel de significancia seleccionado, se puede rechazar la hipótesis nula, en cuyo caso, por lo menos una ρ es significativamente diferente de cero.

10. NOTAS DE APOYO

Violaciones a los supuestos del Modelo clásico de regresión lineal: Multicolinealidad, heteroscedasticidad y autocorrelación

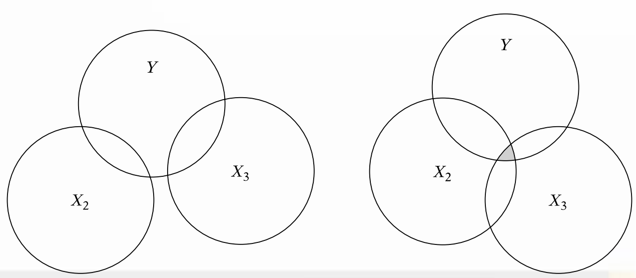
Una vez que se realiza la regresión de un modelo con varias variables explicativas, surgen diversos problemas más allá de una incorrecta especificación. Las limitaciones que surgen por el origen y calidad de las muestras pueden generar una serie de problemas, algunos relacionados con las variables explicativas, otros relacionados con la calidad y comportamiento de las perturbaciones y otros más al modelo.

MULTICOLINEALIDAD
Recuerde el supuesto 8, el cual tiene que ver con la naturaleza de las variables explicativas (*Xi*). El supuesto expresa que no debería existir colinealidad exacta entre las variables, lo que se entiende es que cada una de las variables debe ser independiente del resto, deben medir el fenómeno al que se refieren de forma que no esté contaminado por el resto de las variables. Es decir, el problema de multicolinealidad aparece cuando las variables explicativas de nuestro modelo están correlacionadas entre sí.

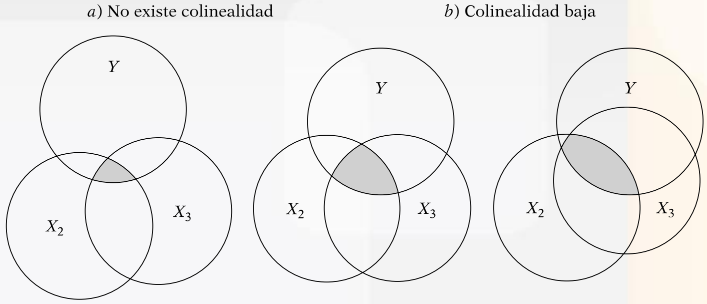
En la realidad, este problema se debe más al grado de asociación que presentan las variables explicativas, pues muchas de las variables sostienen una correlación en su medición; es decir, no se contabilizan de forma "pura", sino en relación de otras. En este sentido, el origen de este problema tiene que ver con la calidad de la muestra, más que a un problema de la población. Aunque veremos más adelante las posibles causas que generan este problema.

Para conocer los problemas o las consecuencias que genera se debe tener en cuenta el grado en que se presenta la colinealidad.

Tipos de multicolinealidad
Observe el gráfico de Ballentine sobre multicolinealidad y sus grados:



a) No existe colinealidad *b*) Colinealidad baja



c) Colinealidad moderada *d*) Colinealidad alta *e*) Colinealidad muy alta

Multicolinealidad perfecta. La situación se observa fácilmente en la matriz de variables explicativas, pues en el caso de multicolinealidad perfecta el determinante de dicha matriz será cero y no se podrá llevar a cabo la regresión.

Ejemplo: considere las siguientes variables explicativas X_1, X_2 y X_3:

X_1	X_2	X_3*
4	6	10
6	7	13
5	7	12
8	8	16
9	6	15

Como puede apreciarse X_3 resulta de X_1+X_2, por lo que, al trabajar la matriz X'X se tiene lo siguiente:

$$X'X = \begin{bmatrix} 4 & 6 & 5 & 8 & 9 \\ 6 & 7 & 7 & 8 & 6 \\ 10 & 13 & 12 & 16 & 15 \end{bmatrix} * \begin{bmatrix} 4 & 6 & 10 \\ 6 & 7 & 13 \\ 5 & 7 & 12 \\ 8 & 8 & 16 \\ 9 & 6 & 15 \end{bmatrix} = \begin{bmatrix} 222 & 219 & 441 \\ 219 & 234 & 453 \\ 441 & 453 & 894 \end{bmatrix}$$

Donde:

$|X'X|$=0 Esto imposibilita la obtención de $\hat{\boldsymbol{\beta}} = (X'X)^{-1}(X'Y)$

Por ello, en presencia de multicolinealidad perfecta no será posible determinar los coeficientes de la regresión para las variables explicativas y los errores estándar se vuelven infinitos.

Multicolinealidad menos que perfecta. En este caso, los coeficientes de regresión sí podrán ser estimados, pero poseen grandes errores estándar, por lo que los coeficientes no pueden ser estimados de forma precisa.

Las fuentes de la multicolinealidad, según Montgomery y Peck, son:

1. El método de recolección de la información.
2. Restricciones en el modelo o en la población que se está estudiando.
3. Especificación del modelo.
4. En el caso de un modelo sobre determinado. Es decir, el modelo cuenta con más variables explicativas que el número de observaciones.
5. La tendencia común que pudieran presentar las variables regresoras en un modelo de series de tiempo.

Como consecuencias de la multicolinealidad (no perfecta), se tiene el problema de la precisión de los coeficientes para las variables explicativas, derivado de errores estándar muy grandes.
En resumen, entre las consecuencias prácticas de la multicolinealidad están las siguientes:

1. En el caso de colinealidad menos que perfecta, los estimadores de Mínimos Cuadrados Ordinarios (MCO) aún son los mejores estimadores lineales e insesgados (MELI), pero al tener varianzas y covarianzas muy grandes, hacen poco precisos a los estimadores.
2. Por la misma razón del punto anterior, los intervalos de confianza tienden a ser muy amplios y esto dificulta verificar con certeza la hipótesis de comprobación individual para los estimadores.
3. Una explicación adicional al punto 1, es que el estadístico *t* para uno o varios coeficientes tiende a ser estadísticamente no significativa. Es decir, que la posibilidad de rechazar un estimador, estadísticamente, aumenta.

4. R^2 muy alta, pese a que uno o varios estimadores de los coeficientes sea no significativo. Es decir, la bondad de ajuste suele mostrarse muy alta, aunque varios de los estimadores resulten no significativos en lo individual.
5. Debido a que el origen se suele presentar en el origen de los datos muestrales, los estimadores MCO y sus respectivos errores estándar suelen ser sensibles a pequeños cambios en los datos.

Para detectar la multicolinealidad
Un primer indicio de la presencia de multicolinealidad se presenta cuando se tiene una R^2 muy alta y varios estimadores no significativos en lo individual. Por lo que podemos suponer que existe, en algún grado, multicolinealidad entre nuestras variables.

El problema de la multicolinealidad no es un problema de existencia o no existencia sino del grado en que se presenta el problema, por ello, el investigador se debe apoyar en correlaciones entre las variables regresoras y diagramas de dispersión.

Al detectar alguna(s) correlaciones se sugiere realizar regresiones auxiliares con y sin las variables responsables de la multicolinealidad y comparar las $R^{2.}$

Algunos autores utilizan el Factor de inflación de la varianza (FIV) como un indicador de la presencia de multicolinealidad. Como regla práctica, si el valor FIV de una variable es mayor a 10, se dice que la variable es muy colineal.

Para verificar de forma más completa, se utiliza el factor de tolerancia (TOL), comparando la cercanía de este factor respecto de cero, mayor grado de colinealidad; si es más cercano a 1, menor evidencia de que la variable regresora sea colineal con las demás.

En los casos donde se detecte multicolinealidad entre las variables del modelo se tienen las siguientes alternativas:

1. No hacer nada. Esta idea, planteada por Blanchard, refiere al hecho de que, si se ha elegido el modelo y las respectivas variables para analizar un determinado fenómeno, no se resolverá el problema de ninguna forma, pues ese es el modelo y las variables elegidas y se recomienda no hacer nada. Por el contrario, si estamos abiertos a modificar nuestro modelo, podemos utilizar algunas de las siguientes sugerencias:
2. Búsqueda de información *a priori*, derivada de estudios previos que vinculan las variables que pretendemos incluir.
3. En los casos que sea posible, incluir información de corte transversal y de series de tiempo.
4. Eliminación de una o varias variables, más colineales, de nuestro modelo. Si bien esta solución puede generar un sesgo de especificación del modelo.
5. Una última sugerencia será la transformación de variables. Si identificamos algún patrón temporal, se pueden incluir las primeras diferencias de las variables colineales. También se sugieren otro tipo de transformaciones.
6. En caso de que sea posible, dado que el problema se genera por la calidad y tamaño de la muestra, incluir nuevos datos o datos adicionales para la regresión.
7. En casos donde las variables regresoras sean expresadas en formas polinomiales sean las que generan la colinealidad, se recomienda que se expresen como desviaciones respecto a su valor medio, esto tiende a disminuir el problema sustancialmente; en caso de que no suceda, se recomienda el uso de otras técnicas que rebasan al curso.
8. Finalmente, otros métodos utilizados para disminuir el problema de la multicolinealidad son: el análisis de factores y de componentes principales o la regresión en cadena, temas que salen del objetivo del curso.

Conclusión:
Cuando el objetivo de nuestro modelo es la predicción, el problema de multicolinealidad no se considera grave y el modelo puede utilizarse para tal fin.

Ejercicios:
1.- Considere las siguientes matrices X'X y señale en cuáles de ellas se podría detectar un problema de multicolinealidad y por qué:

$a)\begin{pmatrix}7 & 8 & 9\\ 4 & 5 & 6\\ 1 & 2 & 3\end{pmatrix}$ $b)\begin{pmatrix}1 & 3 & 15\\ 5 & 9 & 12\\ 8 & 9 & 13\end{pmatrix}$ $c)\begin{pmatrix}4 & 7 & 9\\ 5 & 2 & 3\\ 1 & -5 & -6\end{pmatrix}$

SOLUCIÓN:
En las matrices a) y c) el determinante de dichas matrices es igual a cero, por lo que se verifica la presencia de colinealidad perfecta. Y no pueden obtenerse los estimadores de la regresión.

2.- Para la siguiente salida de una regresión, ¿considera que puede existir un problema de multicolinealidad? ¿por qué?

Variable dependiente: *Y*
Muestra: 1947-1962

Variable	Coeficiente	Error estándar	Estadístico *t*	Probabilidad
C	65720.37	10624.81	6.185558	0.0000
PIBR	9.736496	1.791552	5.434671	0.0002
X_4	-0.687966	0.322238	-2.134965	0.0541
X_5	-0.299537	0.141761	-2.112965	0.0562

R cuadrada	0.981404	Media de la variable dependiente	65317.00
R cuadrada ajustada	0.976755	Desviación estándar de la variable dependiente	3511.968
Error estándar de la regresión	534.4492		
Suma de cuadrados residual	3440470.	Criterio de información de Akaike	15.61641
Log verosimilitud	-120.9313	Criterio de Schwarz	15.80955
Estadístico de Durbin-Watson	1.654069	Estadístico *F*	211.0972
		Probabilidad (estadístico *F*)	0.000000

SOLUCIÓN: Se observan valores altos para la bondad de ajuste (R^2), pese a que dos de las tres variables explicativas resultan no significativas en lo individual y el modelo sí resulta estadísticamente significativo, tenemos evidencia de la existencia de multicolinealidad. Sabemos que no es perfecta, porque sí fue posible realizar la estimación de los coeficientes.

3.- Para el modelo de regresión múltiple dado por: $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$, se verifica que $X_{1t} = 4X_{2t}$. Considerando esta información, indique cuáles de los parámetros se pueden estimar, bajo las siguientes condiciones:

a) No se tiene información a priori sobre los coeficientes, y
b) Cuando se sabe que $\beta_2 = 5$

SOLUCIÓN:
a) Cuando no se tiene información sobre los coeficientes, salvo que $X_{1t} = 4X_{2t}$, entonces procedemos a sustituir en el modelo original la restricción conocida.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$
$$Y_t = \beta_0 + 4\beta_1 X_{2t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$
$$Y_t = \beta_0 + (4\beta_1 + \beta_2) X_{2t} + \beta_3 X_{3t} + u_t$$

De esta forma sabemos que los parámetros $\beta_1 \; y \; \beta_2$ y la combinación lineal de ambos, dada por $4\beta_1 + \beta_2$, no se podrán estimar porque no se cuenta con información adicional.

b) Si se conoce que $\beta_2 = 5$, entonces la ecuación a estimar quedaría de la siguiente forma:

$$Y_t = \beta_0 + (4\beta_1 + 5)X_{2t} + \beta_3 X_{3t} + u_t$$

Lo que haría posible la obtención de los estimadores par los parámetros $\beta_0, \; \beta_1 \; y \; \beta_3$

HETEROSCEDASTICIDAD

Retome ahora el supuesto 4 del MRLN que hace referencia a la varianza de las perturbaciones.

Supuesto 4: Para las *X* dadas, la varianza de u_i es constante u homoscedástica.

La violación a este supuesto genera un problema de heteroscedasticidad y provoca que nuestro modelo pierda eficiencia para la predicción, entre otros problemas. Simbólicamente $E(u_i^2) = \sigma^2 \qquad i = 1, 2, 3, \ldots, n$

En principio este problema se origina por la violación de este supuesto, lo que significa que el comportamiento de los errores no es igual ente ellos, lo que genera diferentes varianzas entre ellos: $E(u_i^2) = \sigma_i^2$

En este modelo de regresión por MCO, se tiene que la matriz de varianzas y covarianzas del vector de variables aleatorias de perturbación es: $E(uu') = \sigma_u^2\Omega$, donde Ω es una matriz definida positiva y se considera normalizada, es decir que la traza de Ω es igual al número de observaciones. Esta hipótesis contempla una doble suposición:

$$E(u_i, u_j) = \begin{cases} \sigma_i^2 & \forall \; i = j \\ \neq 0 & \forall \; i \neq j \end{cases}$$

En otras palabras, los componentes del vector *u* no cuentan con varianza igual, provocando la heteroscedasticidad, además de que las covarianzas son diferentes de cero provocando que los elementos del vector *u* presenten autocorrelación.

Las perturbaciones siguen ahora una distribución normal de la forma $u \sim N(0, \sigma_i^2\Omega)$.

Con estas nuevas especificaciones el estimador por MCO sigue siendo insesgado pero deja de tener varianza mínima, por lo que deja de ser eficiente. Esto provoca que al utilizar la matriz de varianzas y covarianzas que está mal calculada, las pruebas de hipótesis quedan invalidadas, provocando que se requieran otros métodos de estimación como el método de Mínimos Cuadrados Generalizados (MCG), el cual consiste en minimizar la suma de los cuadrados de los residuos ponderada por la matriz Ω^{-1}. Con la utilización de este método se obtienen estimadores lineales, insesgados y óptimos. En términos matriciales la estimación se lleva a cabo mediante:

$$b_G = \left(X'^{\Omega^{-1}}X\right)^{-1} X'^{\Omega^{-1}}Y$$

Las causas posibles de la heteroscedasticidad:

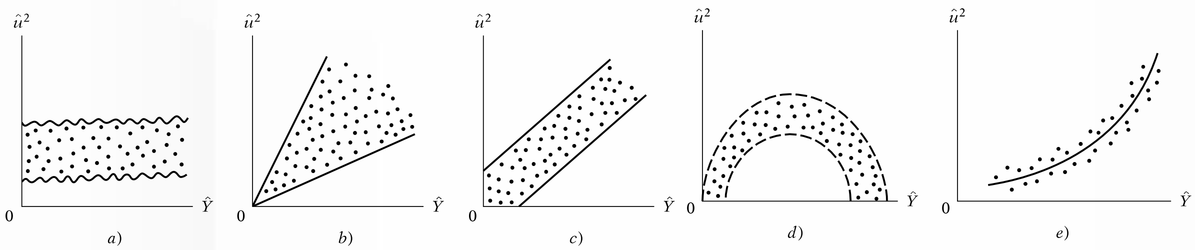
a) Causas estructurales, se producen generalmente en los modelos de corte transversal, cuando las observaciones muestran un comportamiento muy heterogéneo.
b) Causas muestrales, cuando no es posible disponer de todos los datos de una muestra se suele trabajar con valores medios o agregados correspondientes a submuestras, de la muestra total.

c) Causas espurias: es posible que la presencia de heteroscedasticidad sea consecuencia de la existencia de otros problemas en el modelo.

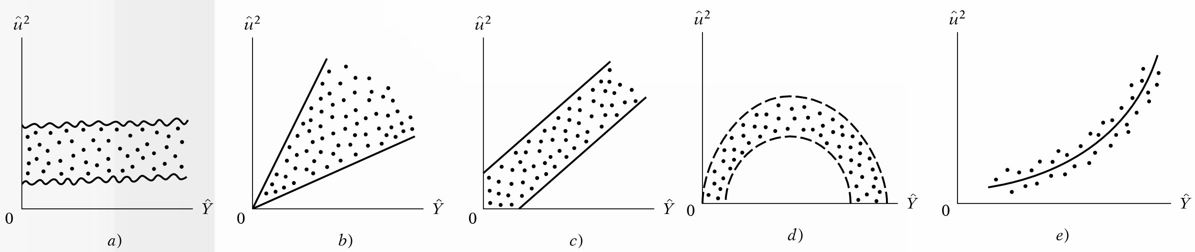
Para identificar la existencia de heteroscedasticidad en el modelo se sugieren los siguientes procedimientos:

Métodos gráficos
Lo primero que se debe hacer es realizar la regresión mediante MCO y guardar los residuos como una serie sobre la que se irá trabajando.

El primer gráfico será la línea de regresión de los residuos contra los valores estimados de la variable dependiente, para observar la existencia de algún patrón sistemático entre ambas variables, en cuyo caso es posible la existencia de heteroscedasticidad. Considere los siguientes gráficos, definidos en Gujarati y Porter (2010):



El Segundo método gráfico es el de línea o líneas de regresión de los residuos (o su cuadrado) frente al regresor causante de la heteroscedasticidad (o su cuadrado). Si la representación gráfica presenta un comportamiento similar es indicativo de que puede existir heteroscedasticidad. Nuevamente, observemos las propuestas de Gujarati y Porter:



Dentro de las pruebas formales se tienen las siguientes:

a) La prueba de Breusch-Pagan
b) La prueba de Goldfeld y Quandt
c) Test de White

La **prueba de Breusch-Pagan** supone heteroscedasticidad lineal, es decir, la varianza del término de error en cada observación dependerá de un vector de variables z_i de dimensión p (esto es, p variables explicativas en las que se incluye una constante). La función que define esta relación está dada por:

$$\sigma_i^2 = f(z_i, \alpha) = \alpha_1 + \alpha_2 z_{2i} + \alpha_3 z_{3i} + \cdots + \alpha_p z_{pi}$$

Pasos de la prueba:

1. Se realiza la estimación por MCO y se obtienen los residuos e_i
2. Se calcula el estimador de la varianza de la siguiente forma: $\hat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n} e_i^2$
3. Se genera una nueva variable, g_i, de los residuos al cuadrado normalizados, mediante: $g_i = \frac{e_i^2}{\hat{\sigma}^2_{MV}}$
4. Se realiza una regresión auxiliar de los cuadrados de los residuos normalizados, g_i , sobre las variables z_i y se obtiene la suma de cuadrados explicada en dicha regresión: $g_i = z_i'\alpha + u$
5. Recordemos que z_i' incluye término independiente.
6. Se establecen las hipótesis a contrastar:

$$H_0: Homoscedasticidad \qquad H_1: Heteroscedasticidad$$

Con $\chi^2_{exp} = \frac{SCE}{2}$ y $\chi^2_{tco} = \chi^2_{p-1}$, donde p es el número de regresores de la regresión auxiliar del punto 4.

La **prueba de Golfeld-Quandt** supone que la varianza de la perturbación depende de una variable z_i. Generalmente se trata de una de las variables explicativas, pero no necesario que así sea.

Pasos de la prueba:

1. Se ordenan en forma ascendente los valores de z_i, variable causante de la heteroscedasticidad.
2. Se omiten las c observaciones centrales. Generalmente, $c \simeq \frac{n}{3}$.
3. Se ajustan dos regresiones por separado para las primeras observaciones y otra sobre las últimas observaciones.
4. Se calculan las sumas de cuadrados de los residuos de las regresiones con las primeras y las últimas observaciones.
5. Se establecen las hipótesis a contrastar:

$$H_0: Homoscedasticidad \qquad H_1: Heteroscedasticidad$$

Con $F_{exp} = \frac{SCR_1}{SCR_2}$ y $F_{tco} = F^{(n-c)/2-k}_{(n-c)/2-k}$

La **prueba de White** no requiere especificar las variables causantes de la heteroscedasticidad y se considera un contraste robusto y asintótico.

Fases de la prueba

1. Se estima el modelo original mediante MCO y se obtienen los residuos de la regresión.
2. Se realiza una regresión auxiliar del cuadrado de los residuos de la regresión original sobre una constante, los regresores del modelo original, sus cuadrados y sus productos cruzados de segundo orden. De esta regresión se obtiene el valor del coeficiente de determinación.
3. Se establecen las hipótesis a contrastar:

$$H_0: Homoscedasticidad \qquad H_1: Heteroscedasticidad$$

Con $\chi^2_{exp} = n \cdot R^2$ y $\chi^2_{tco} = \chi^2_{p-1}$. Donde p es el número de regresores de la regresión auxiliar del punto 2.

AUTOCORRELACIÓN

Retome ahora el supuesto 5 del MRLN que hace referencia a la autocorrelación entre las perturbaciones:

Supuesto 5: Para las X dadas, no hay autocorrelación, o correlación serial, entre las perturbaciones.

El modelo clásico de regresión lineal supone que no existe autocorrelación entre las perturbaciones u_i, esto significa que el término de perturbación relacionado con una observación cualquiera no recibe influencia del término de perturbación relacionado con cualquier otra observación. No obstante, si se observa tal dependencia, entonces existe autocorrelación.

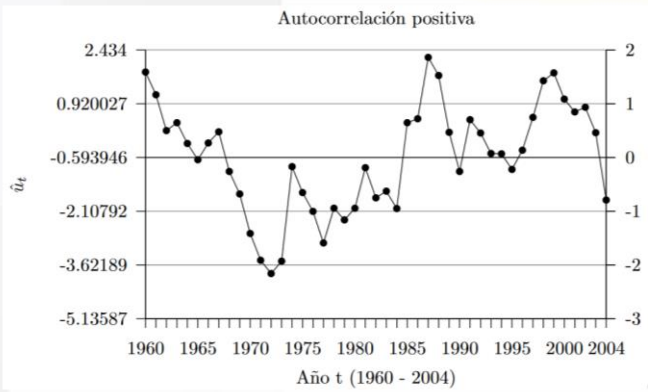
Las causas posibles de la correlación son diversas, a continuación, se listan las principales:

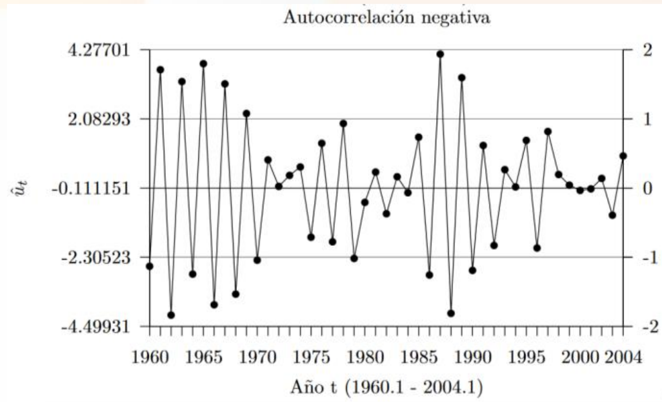
1. Inercia: que es una característica de la mayoría de las series de tiempo económicas.
2. Sesgo de especificación (caso de variables excluidas): este se presenta cuando son excluidas variables del modelo de regresión; con frecuencia, el problema se corrige con la inclusión de dichas variables.
3. Sesgo de especificación (forma funcional incorrecta): este se presenta cuando se trabaja con modelo ajustado de forma incorrecta, el resultado puede ser muy distinto del modelo verdadero.
4. Fenómeno de la telaraña: según el teorema de la telaraña, el mercado se adapta a una determinada modificación o proceso de reajuste, donde la producción es cambiante o discontinua. Un caso donde aplica el teorema de la telaraña es, por ejemplo, el mercado agrícola donde existe un periodo de espera entre la producción y la decisión de consumo de los demandantes. El precio del trigo no es el mismo en el momento que se cultiva que cuando se recolecta.
5. Rezagos: en una regresión de series de tiempo es relativamente común que el valor en el periodo actual de una variable dependa, entre otras cosas, del valor de la misma variable en el periodo anterior.
6. Por la manipulación de los datos.

Para identificar la existencia de autocorrelación en el modelo se sugieren los siguientes procedimientos:

Método gráfico
Hay diversas formas de examinar los residuos, se pueden graficar respecto del tiempo, con una gráfica secuencial de tiempo. Por otro lado, se pueden graficar los residuos estandarizados respecto del tiempo, los residuos estandarizados son tan sólo los residuos $(\hat{u}_t)$ divididos entre el error estándar de la regresión $\sqrt{\hat{\sigma}^2}$, es decir, ${(\hat{u}_t)}/{\hat{\sigma}}$

El grafico temporal de los residuos es una herramienta muy útil para detectar la presencia de autocorrelación. Si los residuos están correlacionados, entonces se deben distribuir aleatoriamente por encima y por debajo de su media, que será cero si el modelo incluye un término constante. No obstante, el conocimiento de los residuos previos a un instante t no predice si el residuo $\hat{u}_t$ se encuentra por encima o por debajo de la media. Por el contrario, si los residuos están positivamente correlacionados, entonces se observa en el grafico temporal rachas de residuos por debajo y por encima de la media. Cuando los residuos estén negativamente correlacionados, el gráfico temporal revelara una alternancia en el signo de los residuos.





Prueba *d* de Durbin Watson

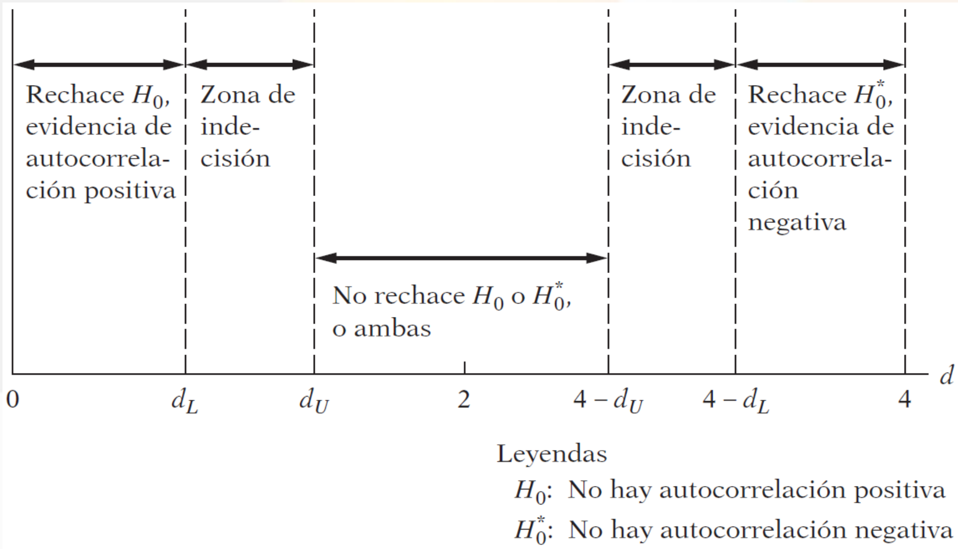
La prueba más conocida para detectar correlación serial es la de los estadísticos Durbin y Watson, que se define como la razón de la suma de las diferencias al cuadrado de residuos sucesivos sobre la SCR, simbólicamente se expresa de la siguiente manera:

$$d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2}$$

Los supuestos en los cuales se basa el estadístico son:

4. El modelo de regresión incluye el término del intercepto.
5. Las variables explicativas, X, son no estocásticas.
6. Las perturbaciones u_t se generan mediante el esquema autorregresivo de primer orden: $u_t = \rho u_{t-1} + \varepsilon_t$. Por tanto, no se pueden utilizar para detectar esquemas autorregresivos de orden superior.
7. Se supone que el término de error ut está normalmente distribuido.
8. El modelo de regresión no incluye valor(es) rezagado(s) de la variable dependiente como una variable explicativa.
9. No hay observaciones faltantes en los datos.

Durbin y Watson encontraron un límite inferior d_L y un límite superior d_U tales que, si el valor *d* calculado cae por fuera de estos valores críticos, puede tomarse una decisión respecto de la presencia de correlación serial positiva o negativa. Gráficamente se tienen las siguientes zonas de decisión.



El mecanismo de la prueba de Durbin-Watson es el siguiente, si suponemos que se cumplen los supuestos de la prueba:

1. Efectuar la regresión por MCO y obtener los residuos.
2. Calcular *d* a partir de la fórmula: $d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2}$
3. Para un tamaño de muestra dado y un número de variables explicativas dado, determinar los valores críticos d_L y d_U.
4. Ahora se siguen las reglas de decisión de la siguiente tabla: Regla de decisión:

Hipótesis nula	**Decisión**	**Si**
No hay autocorrelación positiva	Rechazar	$0 < d < d_L$
No hay autocorrelación positiva	Sin decisión	$d_L \leq d \leq d_U$
No hay correlación negativa	Rechazar	$4 - d_L < d < 4$
No hay correlación negativa	Sin decisión	$4 - d_U \leq d \leq 4 - d_L$
No hay autocorrelación positiva o negativa	No rechazar	$d_U < d < 4 - d_U$

Prueba de Breusch-Godfrey o LM
Breusch y Godfrey elaboraron una prueba para la autocorrelación que es general porque permite regresoras no estocásticas, como los valores rezagados de la regresada; esquemas autorregresivos de orden mayor, como el AR(1), AR(2), etcétera; y promedios móviles simples o de orden superior de los términos de error de ruido blanco.

Pasos para desarrollar la prueba BG o LM:
1. Estime el modelo de regresión de dos variables mediante MCO y obtenga los residuos $\hat{u}_t$
2. Realice la regresión: $\hat{u}_t = \propto_1 + \propto_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_\rho \hat{u}_{t-\rho} + \varepsilon_t$; y obtenga R^2 de esta regresión auxiliar.
3. Si el tamaño de la muestra es grande, Breusch y Godfrey demostraron que: $(n-\rho)R^2 \sim \chi_\rho^2$
Es decir, asintóticamente, $(n-\rho)$ veces el valor de R^2 obtenido en la regresión auxiliar sigue la distribución ji cuadrada con ρ gl. Si en una aplicación $(n-\rho)R^2$ excede el valor crítico ji cuadrada, en el nivel de significancia seleccionado, se puede rechazar la hipótesis nula, en cuyo caso, por lo menos una ρ es significativamente diferente de cero.

11. AUTOEVALUACIÓN

Lee con atención y responde si los siguientes enunciados son verdaderos o falsos.

1. Las principales causas que producen multicolinealidad en un modelo son:

   a) relación causal entre variables explicativas del modelo;
   b) escasa variabilidad en las observaciones de las variables independientes; y
   c) reducido tamaño de la muestra.

   Verdadero Falso

2. Una de las posibles soluciones al problema de multicolinealidad es NO eliminar las variables que se sospechan son causantes de la multicolinealidad.

   Verdadero Falso

3. Si todos los factores de inflación de la varianza FIV son igual a 1, significa que no hay multicolinealidad.

   Verdadero Falso

4. La prueba de White sirve para detectar la presencia de multicolinealidad.

   Verdadero Falso

5. La autocorrelación es un fenómeno que se presenta en muestras que contengan datos asociados al tiempo, aunque también puede presentarse cuando se trabaja con datos de corte transversal. Las causas posibles son:

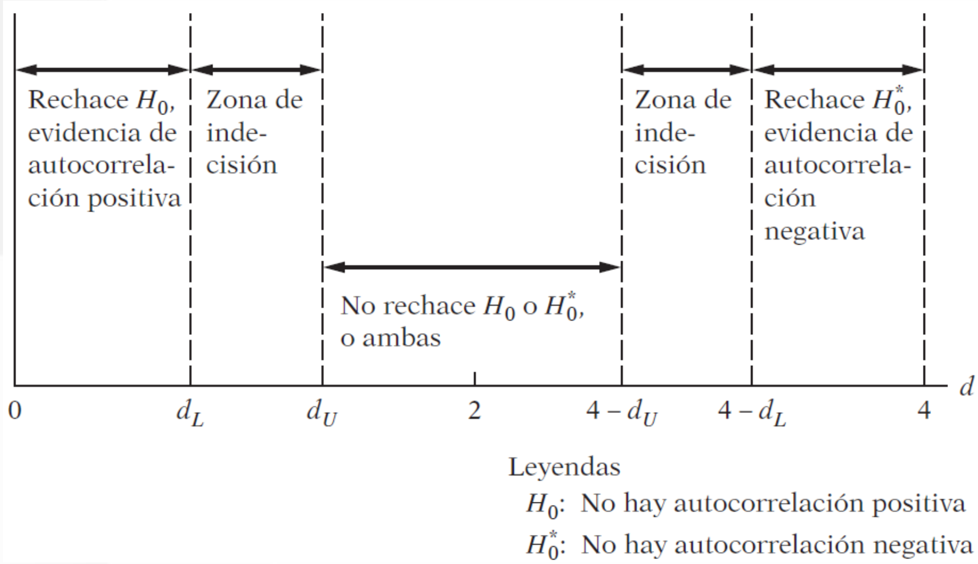
   a) la existencia de ciclos y tendencias en los datos;
   b) cuando hay errores de especificación inicial en el modelo; y
   c) cuando se comete un error de especificación en la forma funcional del modelo.

   Verdadero Falso

6. El modelo autorregresivo más común es el esquema autorregresivo de primer orden de Markov, que supone que la perturbación en el tiempo actual está linealmente relacionada con el término de perturbación en el tiempo anterior, el coeficiente de autocorrelación ρ que da el grado de interdependencia. Este mecanismo se conoce como esquema AR(1).

   Verdadero Falso

7. La siguiente figura muestra las zonas de rechazo del estadístico Durbin-Watson en sospecha de multicolinealidad.



Verdadero

Falso