

Ciencia de Datos para Economía y Negocios

Trabajo Práctico Final

Efectos del ingreso de China a la OMC sobre la apertura comercial global y el crecimiento

Camila Germain - 908610

M. Florencia Blanco - 883743

Noviembre 2025

Facultad de Ciencias Económicas - UBA

Hipótesis y diseño de investigación

Análisis exploratorio de los datos

Detección de outliers

Estadísticas descriptivas

Análisis de distribuciones y supuestos

Análisis de las Regresiones

Conclusiones

Hipótesis y diseño de investigación

Hipótesis principal: El ingreso de China a la OMC en 2001 impactó positivamente en la apertura comercial mundial y en el crecimiento económico.

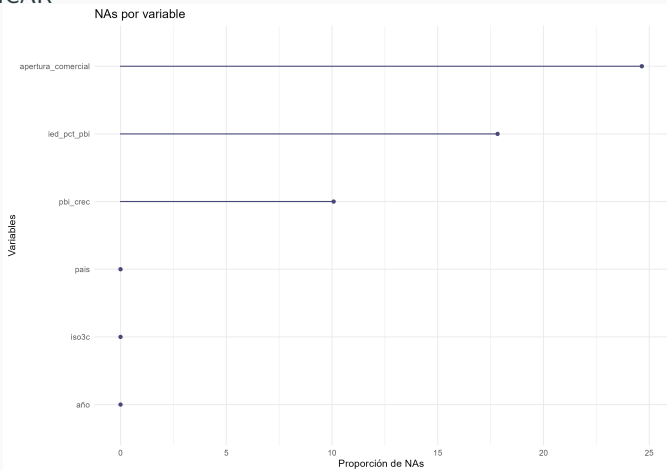
- Obtuvimos los datos del repositorio de R WDI.
- Variables utilizadas:
 - Apertura comercial
 - Tasa de crecimiento del PIB
 - IED como % del PBI
 - País y año
- Años a analizar:
 - 1990 (pre adhesión de China a la OMC)
 - 2001 (adhesión de China a la OMC)
 - 2010 (post ingreso de China a la OMC)

Análisis exploratorio de los datos

Análisis de datos faltantes

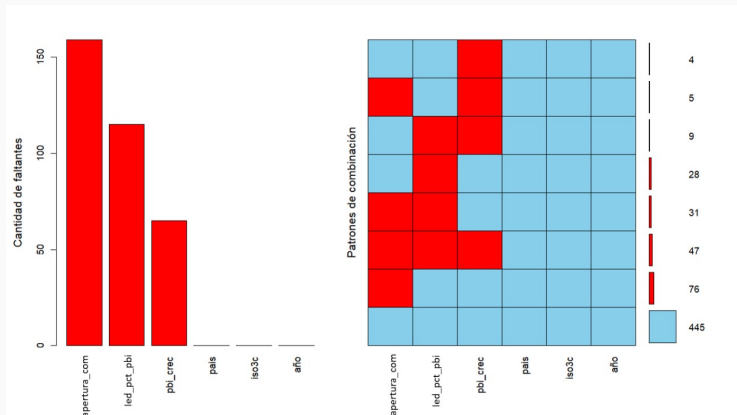
Diagnóstico inicial:

- Alta proporción de NAs para apertura comercial e IED
- Los datos faltantes no son completamente aleatorios, no siguen un patrón MCAR



Patrones de NAs

Patrones de datos faltantes para las variables analizadas



Gran parte de los datos faltantes se concentran en países donde los valores de estas variables no fueron informados, lo que sugiere un problema de cobertura

Analizamos la distribución de los datos faltantes

Prueba t: Testeamos si el PBI promedio difiere entre los grupos de datos faltantes y no faltantes de la Apertura Comercial

- Con un p-value de $0.4645 > \alpha 0.05 \rightarrow$ No rechazamos H_0
- Por lo tanto, no podemos afirmar que los datos faltantes son MAR. La variable observada (crecimiento del PBI) NO es un buen predictor de si la variable clave (apertura comercial) faltará.

Prueba sobre la IED:

- Con un p-value de $0.04308 < \alpha 0.05 \rightarrow$ Rechazamos H_0
- Podemos afirmar que los datos faltantes de apertura comercial son MAR con respecto a la IED

Decisión: Aplicamos imputación múltiple (MICE) como estrategia para reemplazar los NA

Detección de outliers

Detección de outliers

- Obtenemos los datos limpios luego de ejecutar la imputación múltiple
- A partir de los datos filtrados y los datos limpios, validamos si los valores de un vector están dentro de un rango
- Detectamos y reportamos los datos atípicos utilizando el IQR agrupado por año

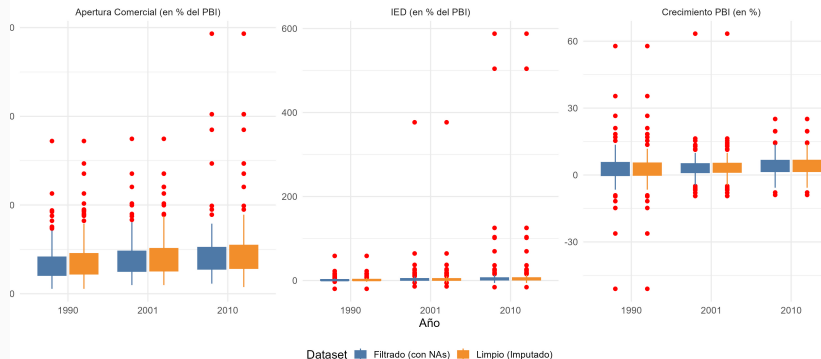
Conclusiones:

- Observamos un aumento en los outliers luego de la imputación múltiple
- Si los datos económicos son naturalmente volátiles, la imputación múltiple imita esa volatilidad, lo que explica el incremento de NAs

Outliers: comparación gráfica

Comparación de Distribución y Atípicos (IQR) antes y después de Imputación

Boxplots por Año: 1990 vs. 2001 vs. 2010



Fuente: Elaboración propia en base a datos del Banco Munc

Estadísticas descriptivas

Estadísticas descriptivas

Agrupamos por dataset, año y variable para calcular las métricas de forma independiente.

Calculamos:

Medidas de tendencia central.

- Media
- Mediana
- Moda

Medidas de dispersión.

- Desvío estándar
- IQR

Hallazgos:

- Tendencia central similar entre dataset crudo y limpio
- Diferencias en dispersión tras imputación

Evaluación del impacto de la limpieza

Variación de los estadísticos descriptivos:

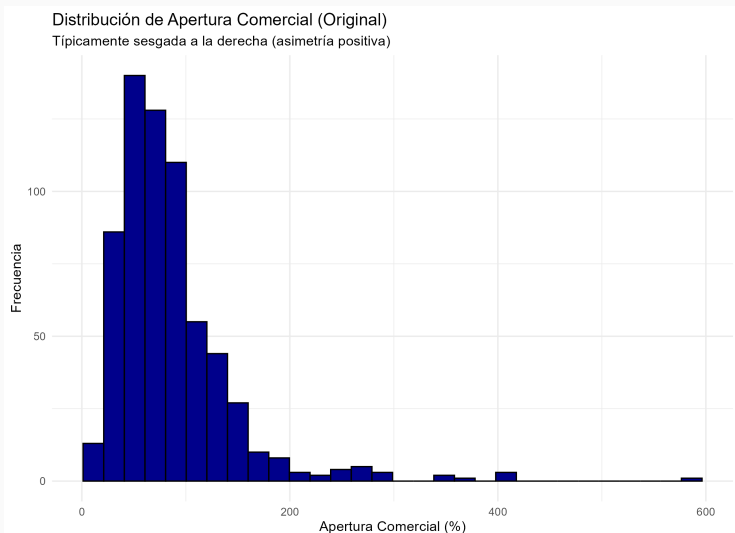
Analizamos el cambio porcentual en los estadísticos del dataset "limpio" (imputado) respecto del "filtrado"(con NAs)

Variable	Año	Media	Mediana	Moda	Desvío	IQR
Apertura comercial	1990	9,5171 %	5,2092 %	278,9578 %	9,1820 %	11,6420 %
Apertura comercial	2001	4,0316 %	2,3657 %	-61,0223 %	5,6944 %	11,2774 %
Apertura comercial	2010	6,6150 %	4,5827 %	435,8863 %	12,7924 %	6,4561 %
IED (% del PBI)	1990	6,3913 %	24,1379 %	57300,0000 %	-10,4440 %	20,5607 %
IED (% del PBI)	2001	-7,6897 %	-12,0120 %	-7833,3333 %	-6,7340 %	-2,5702 %
IED (% del PBI)	2010	17,9443 %	0,3400 %	254,0316 %	12,2156 %	1,5213 %
Crecimiento PBI	1990	-5,1834 %	-3,1530 %	-132,1011 %	-4,1035 %	-6,1049 %
Crecimiento PBI	2001	5,9317 %	3,9567 %	-138,7829 %	-1,7900 %	0,6823 %
Crecimiento PBI	2010	-0,1457 %	-0,4864 %	-84,1039 %	0,4041 %	0,5635 %

Análisis de distribuciones y supuestos

Análisis de la distribución

Histograma de la apertura comercial. En el mismo se puede observar el sesgo de la variable

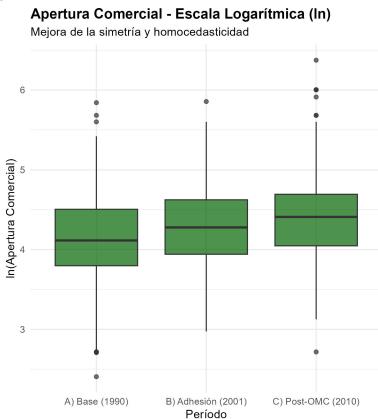
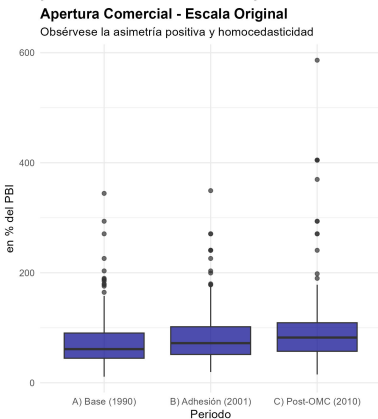


Fuente: Elaboración propia con datos del Banco Mundial

Análisis de la distribución

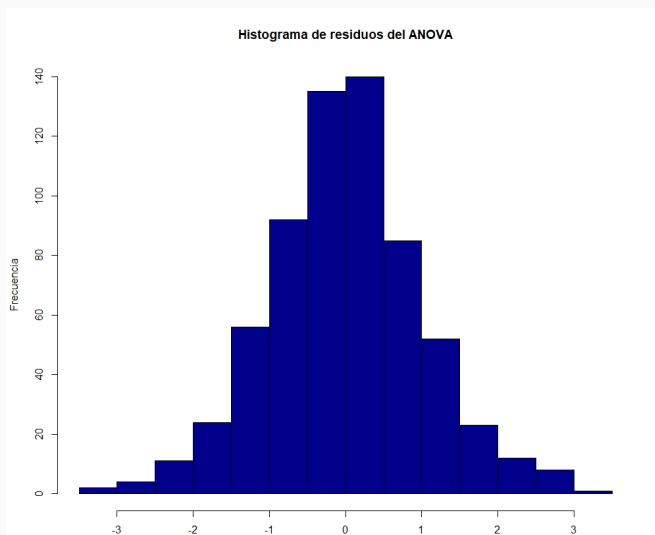
La Apertura Comercial (como % del PIB) muestra una fuerte asimetría a la derecha con valores extremos, un comportamiento típico de variables económicas positivas. Por lo que aplicamos una transformación logarítmica para reducir ésta asimetría.

Comparación de Distribución de Apertura Comercial por Periodo



ANOVA - Análisis estadístico de los supuestos

Testeamos normalidad de los residuos a través de la prueba Shapiro-Wilk y concluimos que éstos no siguen una distribución normal.



Testeamos los datos originales de Apertura Comercial

- Realizamos la prueba de Levene
- Se observa asimetría en los datos y presencia de valores extremos
- Las varianzas no difieren de forma estadísticamente significativa
- Se observa homocedasticidad

Testeamos los datos de Apertura Comercial en escala logarítmica

- Realizamos la prueba de Levene
- Se observa una reducción en la asimetría
- Las varianzas tampoco difieren de forma estadísticamente significativa
- Se mantiene la homocedasticidad

Kruskal–Wallis:

- Dada la presencia de outliers y a la ausencia de normalidad en los residuos, aplicamos este test para evaluar si existen diferencias significativas en la apertura comercial entre los distintos períodos
- El test detecta diferencias significativas globales en la apertura comercial (en escala logarítmica) entre los tres períodos, 1990, 2001 y 2010
- Esto justifica realizar comparaciones post-hoc

Prueba de Dunn (no paramétrica): Observamos cambios en la significatividad al realizar esta prueba

- 1990–2010: significativo
- 1990–2001: no significativo
- 2001–2010: ya no significativo bajo Dunn

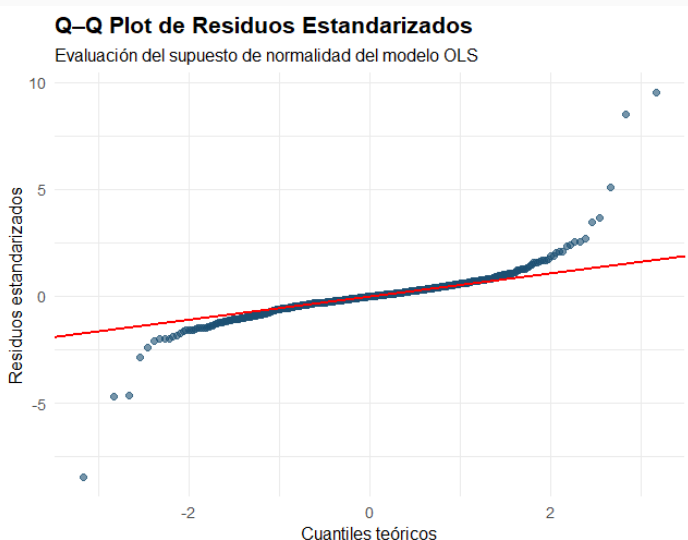
Análisis de las Regresiones

Regresión lineal simple

Variable	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.886090	0.447680	8.681	0.000
apertura_comercial	-0.005136	0.004339	-1.184	0.237
Resumen del modelo				
Residual Std. Error:		6.367 (643 df)		
Multiple R-squared:		0.002175		
Adjusted R-squared:		0.0006228		
F-statistic:		1.401 on 1 and 643 df (p-value = 0.2369)		

- No se encuentra evidencia estadísticamente significativa de que la apertura comercial afecte el crecimiento económico
- El modelo tiene muy poca capacidad explicativa, por lo que la relación entre ambas variables requiere incluir más variables o un enfoque más complejo

Q-Q Plot de los residuos



Fuente: Elaboración propia con datos del proyecto

Análisis de correlación

- En el gráfico de dispersión podemos visualizar la relacion Lin-lin
- La variable de apertura comercial presenta asimetría positiva,
- Aunque por TCL la media se acerca a una distribucion normal, vamos a aplicar logaritmo para que la distribucion de la variable se vuelva más simétrica y se acerque aún más a la normalidad

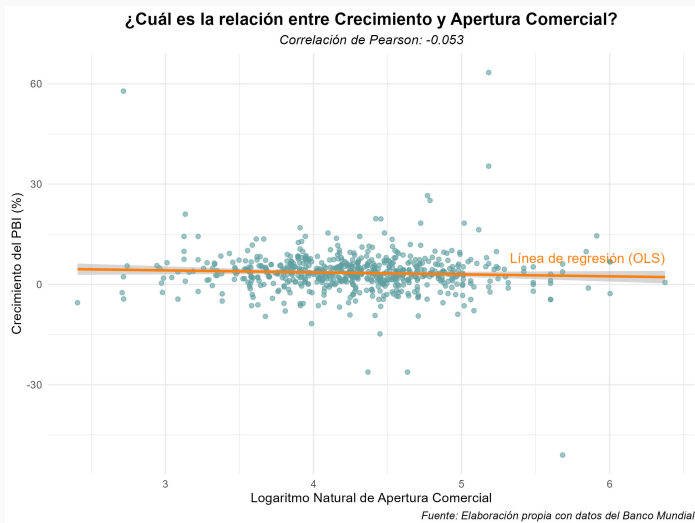
Ajustamos el modelo OLS con logaritmo y dummies según el periodo (año base 1990)

- El siguiente grafico muestra la relación bivariada entre apertura comercial y crecimiento
- Para las regresiones incluimos controles por periodo (dummies de año) pero para la visualización gráfica mantenemos la relación simple entre ambas variables

Análisis de las Regresiones

Análisis de correlación

Crecimiento del PBI (nivel) vs Apertura Comercial y año



Test de White

- Elegimos este test ya que funciona bien incluso si los residuos no son normales
- Realizamos el test y concluimos que los datos presentan heterocedasticidad
- Utilizamos entonces los errores estandar robustos (errores de White)

Modelo con errores robustos

- Observamos así que en 2010, el crecimiento promedio del PBI es aproximadamente 1.65 puntos porcentuales mayor que en el año base, controlando por apertura comercial
- El resto de los coeficientes son no significativos (salvo el intercepto que no tiene una interpretacion economica clara)

Test de Breusch-Godfrey

- Utilizamos este test ya que que los residuos no se distribuyen normalmente
- Analizamos autocorrelación
- El resultado del test indica que no existe evidencia estadística de autocorrelación en los residuos del modelo

Test de Ramsey (Reset)

- Testeamos la especificacion del modelo
- Concluimos que el modelo esta bien especificado

Test de Ramsey (Reset)

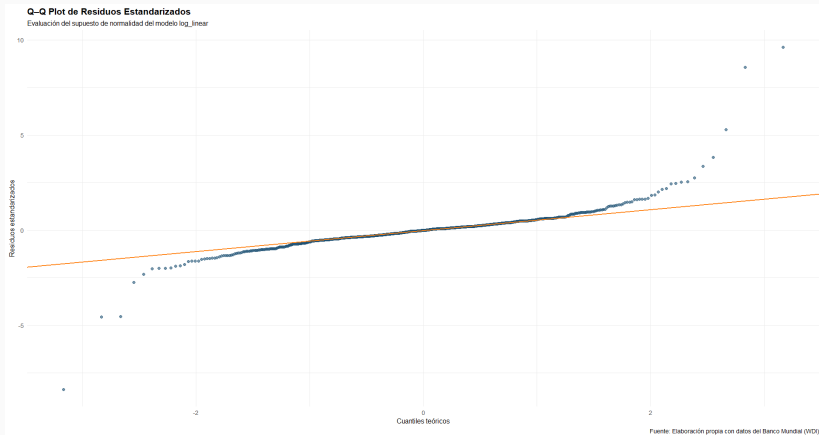
- Testeamos la especificación del modelo y concluimos que el modelo está bien especificado

Coeficiente de correlación de Pearson

- El coeficiente de correlación mide la relación simple entre las variables
- El modelo con dummies controla las diferencias sistemáticas entre periodos
- Realizamos la Prueba de Correlación y concluimos la correlación no es estadísticamente significativa

Análisis de las Regresiones

La distribución de los residuos no cambió ya que éstos dependen únicamente del modelo especificado y del comportamiento de la variable dependiente



Conclusiones

El coeficiente asociado a la apertura comercial (en logaritmo) no resulta estadísticamente significativo, incluso luego de corregir los errores estándar por heterocedasticidad mediante el estimador de White. Esto indica que, una vez controladas las diferencias sistemáticas entre años, no puede afirmarse que la apertura comercial tenga un efecto independiente y robusto sobre el crecimiento del PBI. Los resultados econométricos sugieren que no existe evidencia empírica robusta de que la apertura comercial, por sí sola, impulse el crecimiento económico en los períodos analizados.

Este hallazgo es coherente con parte de la literatura que cuestiona la robustez del vínculo entre ambas variables y subraya la importancia de considerar mecanismos más complejos o factores adicionales para comprender la dinámica del crecimiento.

Conclusiones - Otras perspectivas y líneas de análisis

El análisis puede profundizarse mediante:

- **Controles estructurales adicionales:** incorporar variables como inversión, estabilidad macroeconómica, IED entre otras permitiría captar mecanismos alternativos del crecimiento que hoy no están modelados
- **Modelos dinámicos:** el impacto de la apertura puede manifestarse con rezagos. Es útil estimar modelos con retardos o panel dinámico para identificar efectos de mediano plazo
- **Heterogeneidad entre países:** el impacto de la apertura puede diferir entre las distintas economías. Estimar por subgrupos o incluir interacciones podrían revelar efectos diferenciados
- **Medidas alternativas de apertura:** el indicadores utilizado puede no captar completamente la integración comercial. Índices de tarifas promedio, barreras no arancelarias o participación en cadenas globales de valor ofrecen aproximaciones más precisas
- **Enfoques causales:** técnicas como variables instrumentales, diferencias en diferencias o eventos de liberalización comercial podrían identificar efectos más claros, reduciendo sesgos de simultaneidad