
Estimación del PIB de Estados Unidos, a partir de la demanda agregada

García Rodríguez L.A.

INTRODUCCIÓN

La economía es un sistema complejo en el que diversas variables interactúan entre sí, y comprender las relaciones subyacentes es fundamental para analizar su funcionamiento. En este sentido, el estudio de la demanda agregada y su impacto en el Producto Interno Bruto (PIB) se ha convertido en un tema de gran relevancia en el análisis económico.

La demanda agregada, definida como el gasto total realizado por los consumidores, las empresas, el gobierno y las exportaciones netas, es considerada una medida clave de la actividad económica de un país. Por otro lado, el PIB es ampliamente aceptado como un indicador fundamental del nivel de producción y crecimiento económico. Comprender la relación entre estas dos variables nos permitirá obtener una visión más clara de cómo los cambios en la demanda agregada afectan el nivel de actividad económica en Estados Unidos.

En este trabajo, se aplicará el método de mínimos cuadrados ordinarios (MCO) para estimar los coeficientes de la ecuación de regresión lineal, lo que nos permitirá cuantificar la relación entre los elementos de la demanda agregada y el PIB de Estados Unidos. Además, se realizará un análisis de los supuestos básicos del modelo, como la linealidad, la ausencia de correlación entre los errores y la homocedasticidad, con el fin de validar la robustez de los resultados obtenidos.

A través de este análisis, se espera obtener una mayor comprensión de cómo los elementos de la demanda agregada influye en el nivel de producción económica de Estados Unidos, lo cual puede resultar valioso para la formulación de políticas económicas y la toma de decisiones tanto a nivel gubernamental como empresarial."

Planteamiento del problema:

Objetivo

- A. El objetivo principal analizar la relación entre la demanda agregada y el PIB de Estados Unidos, utilizando un enfoque econométrico basado en el modelo de regresión lineal con mínimos cuadrados ordinarios (MCO), que es una herramienta estadística que permite cuantificar y evaluar la influencia de una variable independiente sobre una variable dependiente.
- B. Identificar si son o no significativas las variables empleadas, que son:
 - 1. Consumo
 - 2. Inversión
 - 3. Exportaciones
 - 4. Importaciones
- C. Obtener los coeficientes de regresión de las variables anteriores, así como su signo y compararlo con el esperado
- D. Aplicar las pruebas de normalidad, heterocedasticidad, multicolinealidad y autocorrelación

Hipótesis:

Existe una relación significativa y positiva entre la demanda agregada y el Producto Interno Bruto (PIB) de Estados Unidos.

Basándonos en la teoría económica y considerando la importancia de la demanda agregada en el funcionamiento de la economía, se plantea la hipótesis de que un aumento en la demanda agregada estará asociado con un aumento en el PIB de Estados Unidos.

Para probar esta hipótesis, se utilizará el modelo de regresión lineal con mínimos cuadrados ordinarios (MCO) para estimar los coeficientes de la ecuación. Si el coeficiente estimado de las variables Consumo, Inversión, Exportaciones e Importaciones resultan ser significativos, se confirmaría la hipótesis planteada y se respaldaría la idea de que un incremento en la demanda agregada se traduce en un mayor nivel de PIB en Estados Unidos.

Es importante destacar que, en el análisis econométrico, también se evaluarán otros supuestos y se realizarán pruebas de significancia y robustez de los resultados obtenidos. Además, se considerarán posibles variables de control y se examinarán aspectos como la multicolinealidad y la heterocedasticidad para garantizar la validez de los resultados.

Marco Teorico

La regresión lineal es una técnica estadística utilizada para modelar la relación entre una variable dependiente y una o más variables independientes. Se basa en la suposición de que la variable dependiente puede ser aproximada mediante una combinación lineal de las variables independientes, con un término de error.

En su forma más simple, un modelo de regresión lineal se expresa como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Donde:

- Y es la variable dependiente que se busca explicar.
- X1, X2, ..., Xn son las variables independientes que se consideran.
- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ son los coeficientes que representan el efecto o impacto de las variables independientes sobre la variable dependiente.
- ε es el término de error aleatorio, que captura las influencias no capturadas por las variables independientes y otros factores no controlados.

El objetivo de la regresión lineal es estimar los coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ de manera que se minimice la suma de los cuadrados de los errores, utilizando el método de mínimos cuadrados ordinarios (MCO).

Esto implica encontrar la línea recta (en el caso de regresión simple) o el hiperplano (en el caso de regresión múltiple) que mejor se ajuste a los datos observados.

En el contexto de la regresión lineal existen algunos supuestos clave:

1. Linealidad: Se asume que la relación entre las variables independientes y la variable dependiente es lineal.
2. Independencia: Se asume que los errores ϵ son independientes entre sí.
3. Homocedasticidad: Se asume que la varianza de los errores es constante en todos los niveles de las variables independientes.
4. Ausencia de multicolinealidad: Se asume que las variables independientes no están altamente correlacionadas entre sí.
5. Normalidad: Se asume que los errores ϵ siguen una distribución normal.

Estos supuestos son importantes para garantizar la validez de los resultados obtenidos a través del modelo de regresión lineal.

La regresión lineal es una herramienta ampliamente utilizada en diversas disciplinas, incluyendo la economía, la sociología, la psicología y muchas otras. Permite analizar y cuantificar la relación entre variables, identificar la magnitud y dirección de los efectos, y realizar predicciones y proyecciones.

Revisión de la literatura

Autores	Año	Investigación	Anotaciones
Darrell Duffie y Kenneth J. Singleton	2003	The Aggregate Demand for Treasury Debt"	Este estudio examina la relación entre la demanda agregada de bonos del Tesoro de Estados Unidos y el ciclo económico. Analiza cómo la demanda de bonos está relacionada con variables macroeconómicas, como el PIB, la inflación y las tasas de interés.
Robert J. Gordon	1984	Aggregate Demand and the Productivity Slowdown	En este trabajo, el autor investiga la relación entre la demanda agregada y la desaceleración de la productividad en Estados Unidos durante la década de 1970. Examina cómo los cambios en la demanda agregada afectaron el crecimiento económico y la productividad.
Olivier Blanchard	2010	<i>Aggregate Demand and Supply Factors in OECD Unemployment: An Update</i>	Este estudio investiga la relación entre la demanda agregada y el desempleo en los países de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), incluido Estados Unidos. Analiza cómo los cambios en la demanda agregada afectan el nivel de desempleo.

Modelo a utilizar

Recordando la hipótesis del trabajo, donde un aumento en la demanda agregada estará asociado con un aumento en el PIB de Estados Unidos, debido a que el incremento en el gasto total realizado por los consumidores, las empresas, el gobierno y las exportaciones netas tiende a estimular la actividad económica, generando un mayor nivel de producción y, por ende, un aumento en el PIB:

Se estimará la ecuación de regresión a partir de las siguientes variables que conforman la demanda agregada: $DA = Consumo(Consumo + Gasto) + Inversión + Exportaciones - Importaciones$.

La forma funcional del modelo será log - log, por lo que entonces:

$$\log(PIB_{EU}) = b_0 + b_1 \log(Consumo) + b_2 \log(Inversión) + b_3 \log(Exportaciones) - b_4 \log(Importaciones)$$

Donde:

$\log(PIB_{EU})$ = Es el logaritmo del PIB de Estados Unidos;

$\log(Consumo)$ = Es el logaritmo del consumo;

$\log(Inversión)$ = Es el logaritmo de la inversión;

$\log(Exportaciones)$ = Es el logaritmo de las exportaciones;

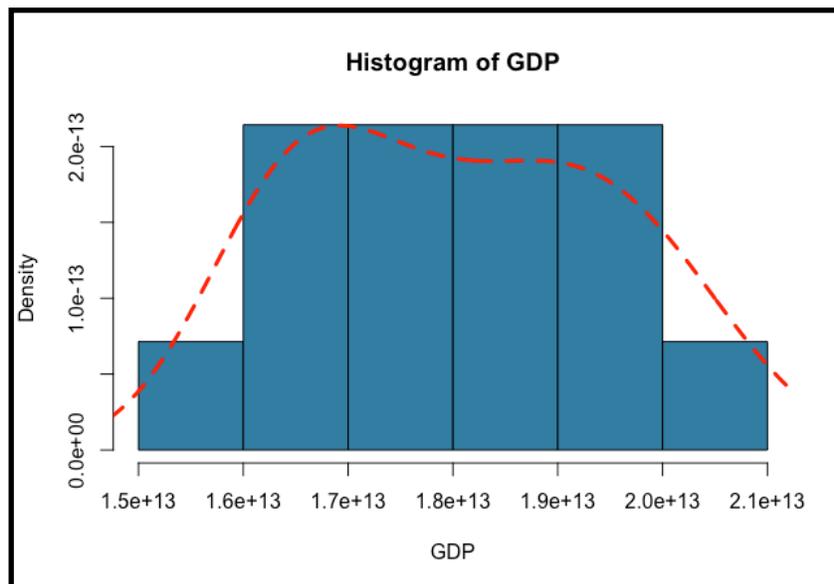
$\log(Importaciones)$ = Es logaritmo de las importaciones.

Descripción de las variables

PIB Estados Unidos

Indicador de referencia: GDP (constant 2015 US\$)

El PIB a precios de comprador es la suma del valor agregado bruto de todos los productores residentes en la economía más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. Se calcula sin hacer deducciones por depreciación de bienes fabricados o por agotamiento y degradación de los recursos naturales. Los datos están a precios constantes de 2015, expresados en dólares estadounidenses. Las cifras en dólares para el PIB se convierten de las monedas nacionales utilizando los tipos de cambio oficiales de 2015.



Consumo

Indicador de referencia: Final consumption expenditure (constant 2015 US\$).

Este indicador representa el gasto de consumo final (antes consumo total) y es la suma del gasto de consumo final de los hogares (antes consumo privado) y el gasto de consumo final del gobierno general (antes consumo del gobierno general). Los datos están a precios constantes de 2015, expresados en dólares estadounidenses.

HISTOGRAMA DEL CONSUMO DE ESTADOS UNIDOS

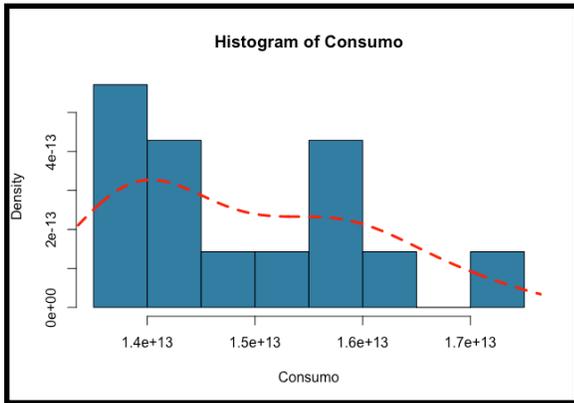
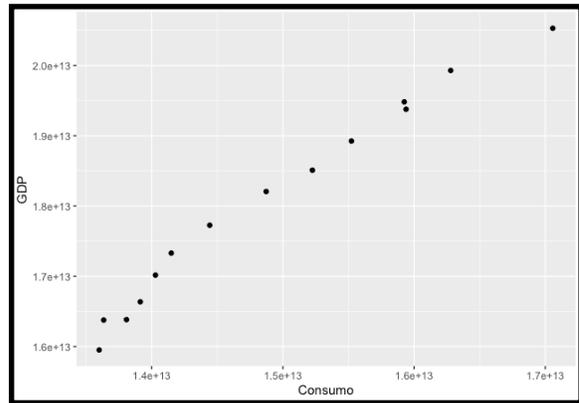


GRAFICO DE DISPERSIÓN

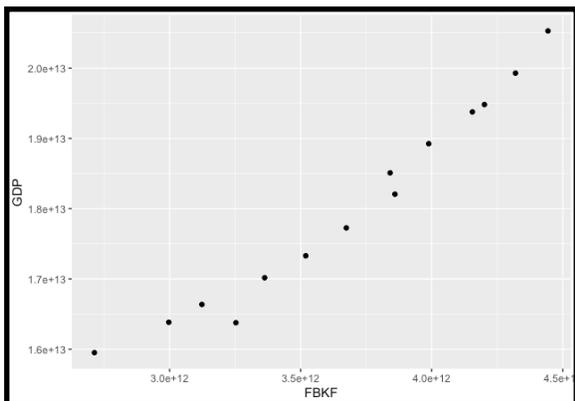


Inversión

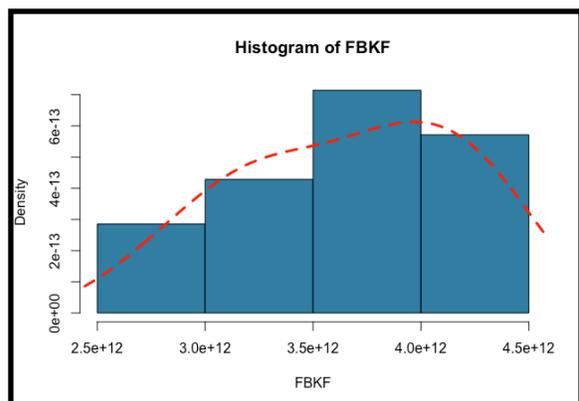
Indicador de referencia: Gross capital formation (constant 2015 US\$)

La formación bruta de capital (anteriormente, inversión interna bruta) consiste en los desembolsos por adiciones a los activos fijos de la economía más los cambios netos en el nivel de inventarios. Los activos fijos incluyen mejoras de terrenos (cercas, zanjas, desagües, etc.); compras de plantas, maquinaria y equipo; y la construcción de carreteras, vías férreas y similares, incluidas escuelas, oficinas, hospitales, viviendas residenciales privadas y edificios comerciales e industriales. Los inventarios son existencias de bienes que tienen las empresas para hacer frente a fluctuaciones temporales o inesperadas en la producción o las ventas, y el "trabajo en curso". Los datos están a precios constantes de 2015, expresados en dólares estadounidenses.

GRAFICO DE DISPERSIÓN



HISTOGRAMA DE LA INVERSIÓN DE ESTADOS UNIDOS



Exportaciones

Indicador de referencia: Exports of goods and services (constant 2015 US\$)

Las exportaciones de bienes y servicios representan el valor de todos los bienes y otros servicios de mercado proporcionados al resto del mundo. Incluyen el valor de las mercancías, fletes, seguros, transporte, viajes, regalías, derechos de licencia y otros servicios, como servicios de comunicación, construcción, financieros, de información, comerciales, personales y gubernamentales. Excluyen la remuneración de los empleados y los ingresos por inversiones (anteriormente llamados servicios de factores) y los pagos de transferencia. Los datos están a precios constantes de 2015, expresados en dólares estadounidenses.

HISTOGRAMA DE LAS EXPORTACIONES DE ESTADOS UNIDOS

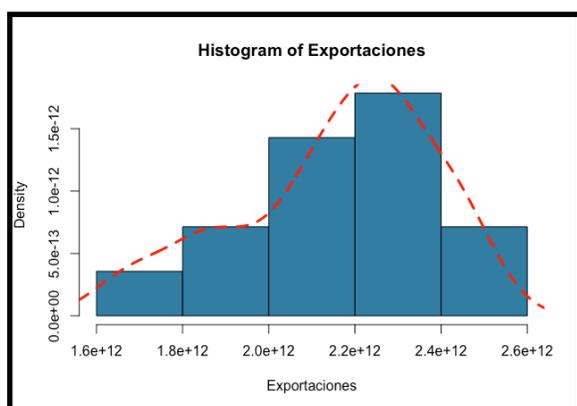
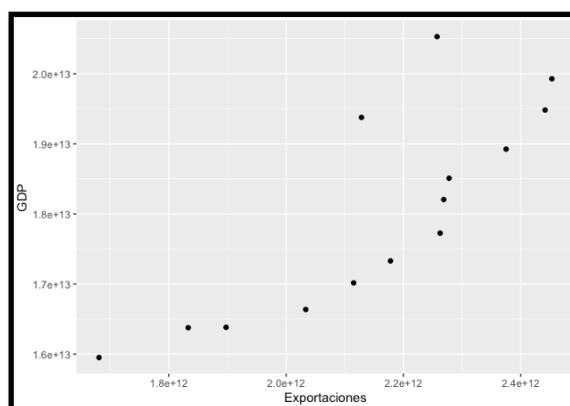


GRAFICO DE DISPERSION

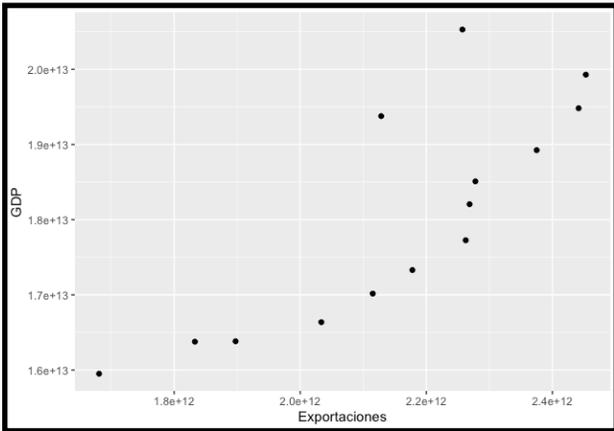


Importaciones:

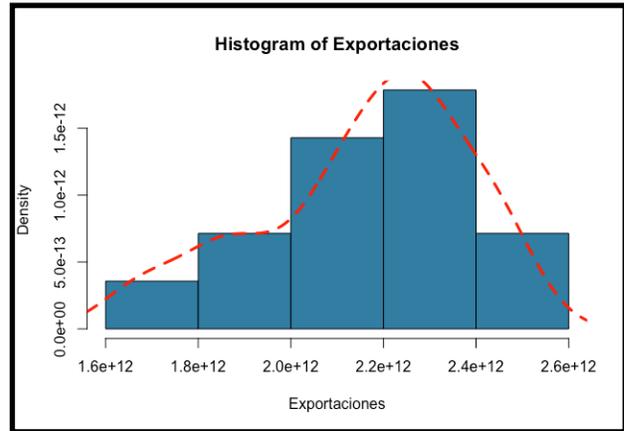
Indicador de referencia: Imports of goods and services (constant 2015 US\$)

Las importaciones de bienes y servicios representan el valor de todos los bienes y otros servicios de mercado recibidos del resto del mundo. Incluyen el valor de las mercancías, fletes, seguros, transporte, viajes, regalías, derechos de licencia y otros servicios, como servicios de comunicación, construcción, financieros, de información, comerciales, personales y gubernamentales. Excluyen la remuneración de los empleados y los ingresos por inversiones (anteriormente llamados servicios de factores) y los pagos de transferencia. Los datos están a precios constantes de 2015, expresados en dólares estadounidenses

GRAFICO DE DISPERSION



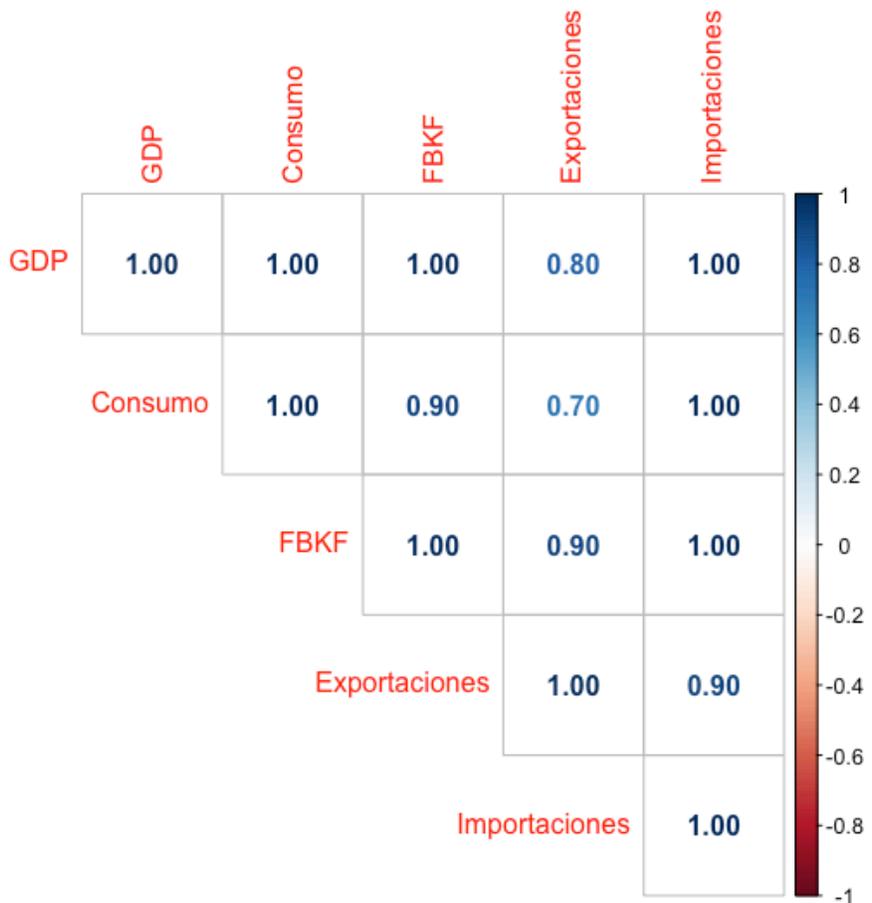
HISTOGRAMA DE LAS IMPORTACIONES DE ESTADOS UNIDOS



Correlaciones:

Puede observarse en la matriz de correlaciones, que hay correlación positiva e intensa entre todas las variables explicativas respecto a la explicada. Lo que da indicios de un buen ajuste del modelo, y una buena regresión.

Sin embargo, también se aprecia alta correlación entre las variables independientes, lo que advierte de problemas de dependencia lineal entre las variables exógenas



Resultados Obtenidos:

Los resultados de la estimación de la ecuación de regresión con la forma funcional log - log establecida son los siguientes:

Regresión:

$$\log(\text{PIB_EU}) = 0.14 + 0.85 \log(\text{Consumo}) + 0.19(\text{Inversión}) + 0.12 \log(\text{Exportaciones}) - 0.16 \log(\text{Importaciones})$$

Variable	Estimadores	Std.Error	Valor t	P-value
Intercepto	0.141109	0.109166	1.29260942051555	0.228
log(Consumo)	0.852807	0.005935	143.691154170177	2×10^{-16}
log(Inversion)	0.197206	0.003409	57.8486359636257	6.94×10^{-13}
log(Exportaciones)	0.121159	0.003163	38.3050901043313	2.80×10^{-11}
log(Importaciones)	-0.161799	0.006392	-25.3127346683354	1.13×10^{-9}

$R^2 = 0.99$, $R^2_{ajustada} = 0.99$, P-value del estadístico F = 2.2×10^{-16}

El resumen del modelo de regresión nos indica que el intercepto b_0 , resulta ser no significativo ya que su p-value es mayor al nivel de significancia, por lo que no hay evidencia suficiente para rechazar la $H_0: b_0 = 0$. El resto de las variables resultan ser significativas a un nivel de confianza del 95%, ya que el p-value asociado a la probabilidad de su estadístico t es menor al nivel de significancia de 0.05, por lo que se rechaza la hipótesis nula $H_0: b_i = 0$, y se acepta la $H_1: b_i \neq 0$, por lo que las variables son significativas individualmente. En el caso del estadístico F, su respectivo p-value es menor al nivel de significancia elegido, de 0.05, por lo que se rechaza la H_0 , y se acepta la alternativa; $H_1: b_0 + b_1 + b_2 + b_3 + b_4 \neq 0$.

Interpretación de los parámetros:

Primero tenemos a B_0 , el valor de la constante, que no está explicado por las variables independientes, y que cuyo valor será de 0.14 cuando el crecimiento de las variables explicativas sea 0. Cabe resaltar que aunque esta constante no es significativa, no se puede excluir dado que estaríamos sesgando la estimación del modelo.

Al tener un modelo log - log, debemos interpretar en tasas de crecimiento. Por lo que pasaremos a presentar las elasticidades correspondientes.

Elasticidad Consumo del PIB:

Por cada punto porcentual de crecimiento del Consumo (que es la suma del gasto de consumo final de los hogares y el gasto de consumo final del gobierno general), el PIB de EU, crecerá en 0.85%, siendo esta la variable con mayor influencia dentro de la demanda agregada, y por ende del PIB.

Elasticidad Inversión del PIB:

Por cada punto porcentual de crecimiento de la inversión (FBKF) (que son los activos fijos como mejoras de terrenos, compras de plantas, maquinaria y equipo; y la construcción de carreteras, vías férreas y similares, incluidas escuelas, oficinas, hospitales, viviendas residenciales privadas y edificios comerciales e industriales.), el PIB de EU, crecerá en 0.19%

Elasticidad Exportaciones del PIB:

Por cada punto porcentual de crecimiento en las exportaciones (que incluyen el valor de las mercancías, fletes, seguros, transporte, viajes, regalías, derechos de licencia y otros servicios, como servicios de comunicación, construcción, financieros, de información, comerciales, personales y gubernamentales que son ofrecidos al resto del mundo), el PIB de EU, crecerá en 0.12%

Elasticidad Importaciones del PIB:

Por cada punto porcentual de crecimiento en las importaciones (que incluyen el valor de las mercancías, fletes, seguros, transporte, viajes, regalías, derechos de licencia y otros servicios, como servicios de comunicación, construcción, financieros, de información, comerciales, personales y gubernamentales que son recibidos del resto del mundo), el PIB de EU, decrecerá en - 0.16%

Pruebas de 5/6 supuestos del teorema de Gauss Markov:

- 1 - La derivada de nuestros coeficientes es 0, se cumple con la linealidad en los parámetros.
- 2- Los errores se distribuyen como una normal con media 0 y varianza constante

Shapiro-Wilk Normality Test
W = 0.95861
P-Value = 0.7001
Media de los errores = 2.902222×10^{-21}

Se observa que se pasa con éxito la prueba de normalidad Shapiro-Wilk, ya que el p-value asociado es mayor a el nivel de significancia 0.05, por lo que no existen elementos suficientes para rechazar la H0: Los errores tiene distribución normal.

- 3- No multicolinealidad, es decir, no dependencia lineal entre las variables explicativas

Se observa que no se pasa con éxito la prueba del Factor de Inflación de la varianza, debido a que el valor de los factores en promedio e individualmente es mayor a 10.

Factor de Inflación de la Varianza - VIF
log(Consumo) = 19.22
log(FBKF) = 25.86
log(Exportaciones) = 12.29
log(Importaciones) = 19.22

- 4 - No heterocedasticidad, es decir la varianza es constante en la regresión

Breusch Pagan Test for Heteroskedasticity
DF = 1
Chi2 = 0.99
P-value = 0.319

Se pasa con éxito la prueba Breusch Pagan para heterocedasticidad, ya que el p-value es mayor a 0.05, así que no hay evidencia suficientes para rechazar la H0: la varianza es constante

5 - No autocorrelacion, es decir la correlación entre los errores y sus valores pasados

Se pasa con éxito la prueba Durbin Watson, dado que el p-value es mayor al nivel de significancia, y el DW cae en la zona de aceptación entre el dU y $4 - dU$. Por lo que no existe autocorrelacion de primer orden

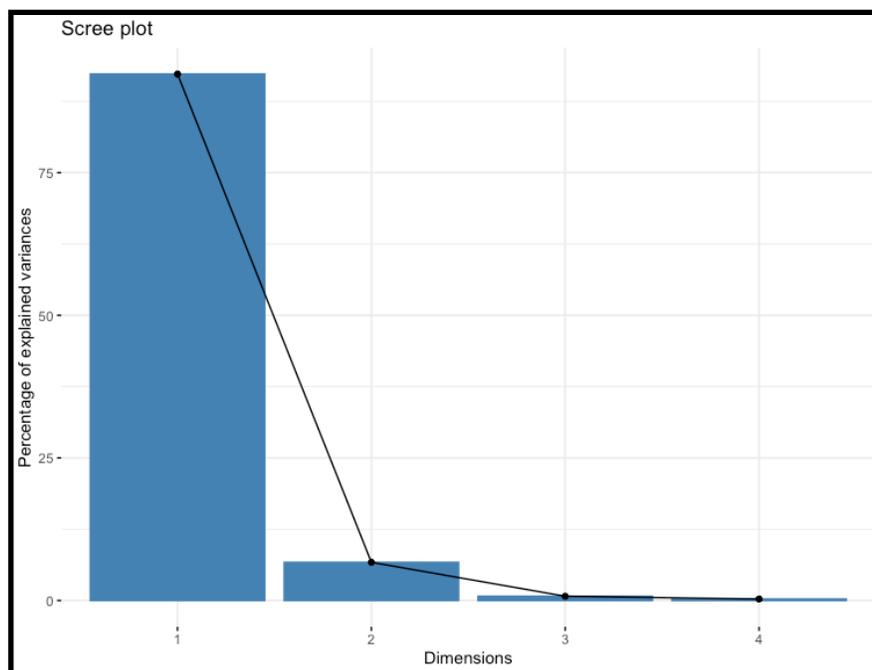
Durbin Watson Test
DW = 1.9798
P - value = 0.2095

Se lograron cumplir los supuestos de linealidad, normalidad, homocedasticidad, no autocorrelacion, pero desafortunadamente no se logró el supuesto de no multicolinealidad. Para tratar este asunto, vamos a recurrir al método PCA, componentes principales, el cual de forma resumida, creará nuevas variables explicativas que recolecten la varianza explicada de nuestras variables que conforman la demanda agregada, para crear así componentes principales de la demanda agregada, las cuales tomaremos como nuevas variables explicativas, aprovechando una de sus características principales, que es la ortogonalidad. Or lo que quedará solucionado el problema de multicolinealidad.

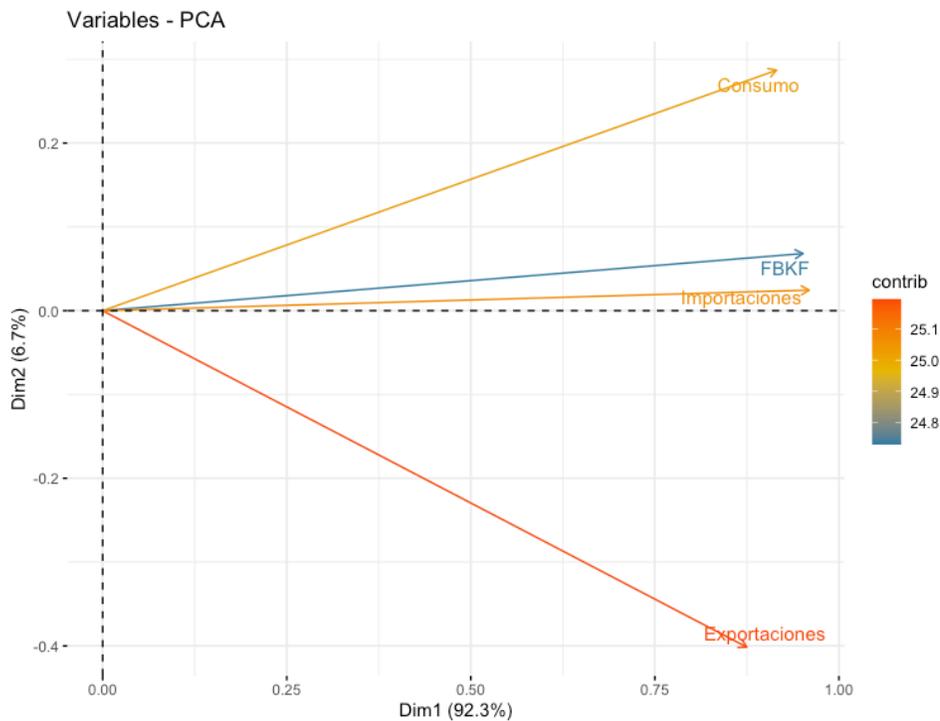
El proceso es el siguiente:

1. Normalización de los datos: Se deben normalizar las variables para que tengan media cero y varianza unitaria. Esto es importante para evitar que las variables con mayor escala dominen el análisis.
2. Cálculo de la matriz de covarianza o de correlaciones: Se calcula la matriz de covarianza o la matriz de correlación a partir de los datos normalizados. Esta matriz muestra las relaciones de covarianza o correlación entre las variables originales.

3. Cálculo de los autovalores y autovectores: Se obtienen los autovalores y autovectores de la matriz de covarianza. Los autovectores representan las direcciones en las que los datos muestran mayor variabilidad, mientras que los autovalores indican la cantidad de varianza explicada por cada componente principal.
4. Selección de componentes principales: Se seleccionan las primeras k componentes principales que expliquen la mayor cantidad de varianza acumulada. Por lo general, se elige un umbral, como explicar al menos el 80% o 90% de la varianza total.
5. Transformación de los datos: Los datos originales se transforman en un nuevo espacio de menor dimensión utilizando los autovectores correspondientes a las componentes principales seleccionadas.



En este scree plot, podemos ver como el porcentaje de variación explicada de todas las variables de la demanda agregada se puede explicar a partir de dos componentes principales



Este es un loading plot, donde podemos observar el peso de las variables explicativas originales sobre los componentes principales, en el eje X, se muestra el primer componente que explica el 92.3% de la variabilidad. Mientras que en el segundo se explica el 6.7%

El peso e influencia exacto de cada variables sobre los componentes es el siguiente:

Los resultados del promedio de los loading del método de PCA, nos dejan ver las variables más importantes del componente. Por lo que este PCA, sugiere que la demanda agregada es mas sensible a el Consumo, la Inversión, las Importaciones, y por ultimo a las exportaciones.

VARIABLE	PCA_1_ DA	PCA_2_ DA	Promedi o
Consumo	0.494	0.575	0.5345
Inversión	0.514	0.137	0.3255
Exportaciones	0.472	-0.805	-0.1665
Importaciones	0.518	0	0.259

Se propone una nueva ecuación de regresión y forma funcional de la siguiente forma
 Regresión:

$$\log(\text{PIB}_{EU}) = B_0 + B_1 (\text{CP}_{DA_1}) + B_2(\text{CP}_{DA_2})$$

Donde:

$\log(\text{PIB}_{EU})$ = Es el logaritmo del PIB de Estados Unidos;

CP_{DA_1} = Es el componente principal 1 de la demanda agregada; y

CP_{DA_2} = Es el componente principal de la demanda agregada.

Se obtiene:

$$\log(\text{PIB}_{EU}) = 30.52 + 0.0008 (\text{CP}_{DA_1}) + 0.003(\text{CP}_{DA_2})$$

Variable	Estimadores	Std.Error	Valor t	P-value
Intercepto	30.52	0.001647	18530.6618093503	2×10^{-16}
$\log(\text{Consumo})$	0.04175	0.0008897	46.925930088794	5.04×10^{-14}
$\log(\text{Inversion})$	0.0245	0.0033	7.42424242424242	1.33×10^{-5}

$R^2 = 0.99$, $R^2_{ajustada} = 0.99$, P-value del estadístico F = 1.86×10^{-13}

El resumen del modelo de regresión nos indica que el intercepto b_0 , resulta ser significativo ya que su p-value es menor al nivel de significancia, por lo que se rechaza la H_0 , y se acepta H_1 : $b_0 \neq 0$. El resto de las variables resultan ser significativas a un nivel de confianza del 95%, ya que el p-value asociado a la probabilidad de su estadístico t es menor al nivel de significancia de 0.05, por lo que se rechaza la hipótesis nula H_0 : $b_i = 0$, y se acepta la H_1 : $b_i \neq 0$, por lo que las variables son significativas individualmente. En el caso del estadístico F, su respectivo p-value es menor al nivel de significancia elegido, de 0.05, por lo que se rechaza la H_0 , y se acepta la alternativa; H_1 : $b_0 + b_1 + b_2 + b_3 + b_4 \neq 0$, significancia conjunta

Pruebas de 5/6 supuestos del teorema de Gauss Markov:

- 1 - La derivada de nuestros coeficientes es 0, se cumple con la linealidad en los parámetros.
- 2- Los errores se distribuyen como una normal con media 0 y varianza constante

Shapiro-Wilk Normality Test
W = 0.98578
P-Value = 0.9957
Media de los errores =-5.421011e-19

Se observa que se pasa con éxito la prueba de normalidad Shapiro-Wilk, ya que el p-value asociado es mayor a el nivel de significancia 0.05, por lo que no existen elementos suficientes para rechazar la H0: Los errores tiene distribución normal.

- 3- No multicolinealidad, es decir, no dependencia lineal entre las variables explicativas

Se observa que se logra solucionar el problema que teníamos de multicolinealidad ya que el valor de los factores ni en promedio ni individualmente es mayor a 10.

Factor de Inflación de la Varianza - VIF
CP_DA_1 = 1
CP_DA_2 =

- 4 - No heterocedasticidad, es decir la varianza es constante en la regresión

Breusch Pagan Test for Heteroskedasticity
DF = 1
Chi2 = 0.411
P-value = 0.5213

Se pasa con éxito la prueba Breusch Pagan para heterocedasticidad, ya que el p-value es mayor a 0.05, así que no hay evidencia suficientes para rechazar la H0: la varianza es constante

5 - No autocorrelación, es decir la correlación entre los errores y sus valores pasados

Se pasa con éxito la prueba Durbin Watson, dado que el p-value es mayor al nivel de significancia, y el DW cae en la zona de aceptación entre el dU y $4 - dU$. Por lo que no existe autocorrelación de primer orden

Durbin Watson Test
DW = 2.3053
P - value = 0.4951

Se lograron pasar las pruebas de normalidad, heterocedasticidad, multicolinealidad, y autocorrelación. Nuestro r^2 y r^2 ajustado es de 0.99, y tenemos significancia individual y conjunta, por lo que tenemos un buen modelo a partir de los componentes principales.

Por lo que cuando los scores de CP_DA_1 aumenten en una unidad, el PIB lo hará en 0.041, mientras que cuando los scores de CP_DA_2 aumenten en una unidad, el PIB lo habrá en 0.024%

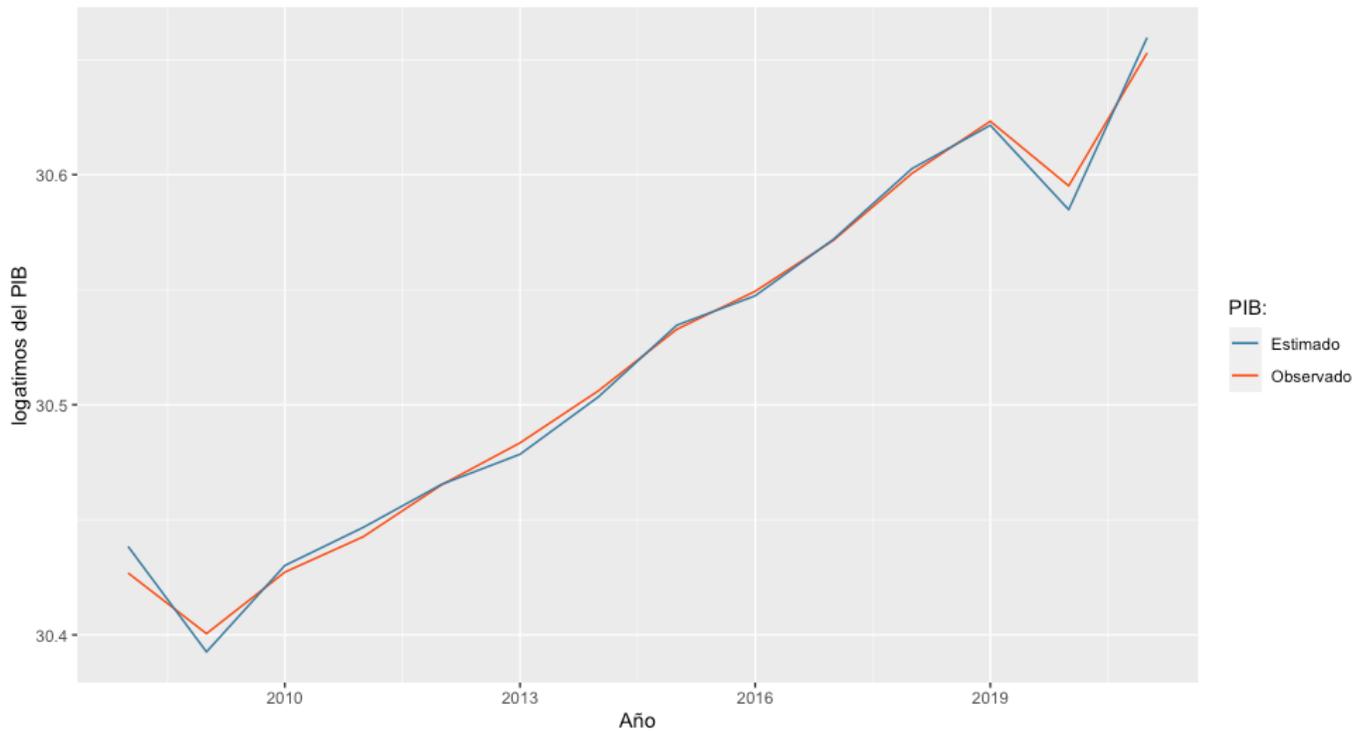
De esta forma concluimos que la demanda agregada y sus componentes tienen una relación estrictamente lineal con el PIB en Estados Unidos. Además, se pudo reconocer la influencia de las variables explicativas iniciales sobre el PIB, y se comporta de la siguiente forma. Siendo 1 la variable con más influencia y 4 la de menor influencia:

1. Consumo
2. Inversión
3. Importaciones
4. Exportaciones

ESTIMACIONES CON EL MODELO

Estimaciones del modelo vs Observaciones reales del PIB de Estados Unidos (2008-2021)

Fuente: Elaboracion propia con datos del Wod Bank



Código R:

```
# Librerias

library(readr)
library(tidyverse)
library(olsrr)
library(tseries)
library(broom)
library(lmtest)
library(car)
library(lubridate)
library(ggfortify)
library(ggthemes)
library(strucchange)
library(forecast)
library(readxl)
library(ggplot2)
library(corrplot)

Base_2 <- read_excel("~/Desktop/Base_M2_TallerVI_Epo.xlsx")
View(Base_2)

Base_3 <- Base_2[,c(2,7,4,6,8)]
Base_4 <- Base_3[,c(2,3,4,5)]

View(Base_3)
View(Base_4)

attach(Base_2)

#Graficos Descriptivos

# Histograma + densidad

hist(GDP, freq = F, col = "#367DA2")
lines(density(GDP), col = "red", lty = 2, lwd = 3)

hist(Consumo, freq = F, col = "#367DA2" )
lines(density(Consumo), col = "red", lty = 2, lwd = 3)
```

```
hist(FBKF, freq = F, col = "#367DA2")
lines(density(FBKF), col = "red", lty = 2, lwd = 3)

hist(Exportaciones, freq = F, col = "#367DA2")
lines(density(Exportaciones), col = "red", lty = 2, lwd = 3)

hist(Importaciones, freq = F, col = "#367DA2")
lines(density(Importaciones), col = "red", lty = 2, lwd = 3)
```

```
# Graficos de dispersion
```

```
## Scatterplot (Diagrama de dispersion)
```

```
ggplot(Base_2) +
  aes(x=Consumo, y=GDP) +#
  geom_point()
```

```
ggplot(Base_2) +
  aes(x=FBKF, y=GDP) +#
  geom_point()
```

```
ggplot(Base_2) +
  aes(x=Exportaciones, y=GDP) +#
  geom_point()
```

```
ggplot(Base_2) +
  aes(x=Importaciones, y=GDP) +#
  geom_point()
```

```
# Correlacion matriz
```

```
correlacion<-round(cor(Base_3), 1)
corrplot(correlacion, method="number", type="upper")
```

```
# Modelo
```

```
C <- lm(log(GDP) ~ log(Consumo) + log(FBKF) + log(Exportaciones) + log(Importaciones))
summary(C)
```

```
mean(C$residuals)
```

```
shapiro.test(C$residuals)
```

```
jarque.bera.test(C$residuals)
```

```
ols_test_breusch_pagan(C)
```

```
dwtest(C)
```

```
vif(C)
```

```
ggplot(Base_2, aes(x = Año)) +  
  geom_line(aes(y = log(GDP), color = "PIB")) +  
  geom_line(aes(y = log(FBKF), color = "FBKF")) +  
  geom_line(aes(y = log(Exportaciones), color = "Exp")) +  
  geom_line(aes(y = log(Consumo), color = "Consumo")) +  
  geom_line(aes(y = log(Importaciones), color = "Importaciones")) +  
  labs(title = "Variación anual de variables seleccionadas (2008-2021)",  
        subtitle = "Banco Mundial",  
        x = "Fecha",  
        y = "Variación porcentual",  
        colour = "Variables")
```

```
ggplot(Base_2, aes(x = Año)) +  
  geom_line(aes(y = log(GDP), color = "PIB")) +  
  geom_line(aes(y = C$fitted.values, color = "Estimado")) +  
  labs(title = "Variación anual de variables seleccionadas (2008-2021)",  
        subtitle = "Banco Mundial",  
        x = "Fecha",  
        y = "Variación porcentual",  
        colour = "Variables")
```

```
#PCA
```

```
#Dependencias
```

```
library(ggplot2)  
library(factoextra)  
library(FactoMineR)  
library(readr)
```

```
#Seleccionar una Submatriz
```

```
datos <- scale(Base_4)  
datos
```

```
#Primera forma de PCA
```

```
pca1 <- princomp(datos)  
pca1$loadings  
pca1$scores
```

```
CP_DA_1 <- pca1$scores[,c(1)]  
CP_DA_1
```

```
CP_DA_2 <- pca1$scores[,c(2)]  
CP_DA_2
```

```
CP_DA_3 <- pca1$scores[,c(3)]  
CP_DA_3
```

```
Base_5 <- cbind(Base_3$GDP, CP_DA_1, CP_DA_2, CP_DA_3)  
View(Base_5)  
Base_5 <- as.data.frame(Base_5)
```

```
#Screeplot
```

```
fviz_eig(pca1, col="#00AFBB")
```

```
#Scoreplot
```

```
fviz_pca_ind(pca1,  
             col.ind = "cos2",  
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))
```

```
#loadingplot
```

```
fviz_pca_var(pca1,  
             col.var = "contrib",
```

```
gradient.cols = c("#367DA2", "#E7B800", "#FC4E07"),  
repel = TRUE)
```

```
fviz_pca_biplot(pca1,  
  col.var = "#367DA2",  
  col.ind = "#FC4E07")
```

```
co <- cor(Base_4)  
E <- eigen(co)  
E$vector
```

```
M3 <- lm(log(V1) ~ CP_DA_1 + CP_DA_2, data = Base_5)  
summary(M3)
```

```
shapiro.test(M3$residuals)  
ols_test_breusch_pagan(M3)  
vif(M3)  
dwtest(M3)
```

```
mean(M3$residuals)
```

```
ggplot(Base_2, aes(x = Año)) +  
  geom_line(aes(y = log(GDP), color = "Observado")) +  
  geom_line(aes(y = M3$fitted.values, color = "Estimado")) +  
  labs(title = "Estimaciones del modelo vs Observaciones reales del PIB de Estados Unidos (2008-2021)",  
        subtitle = "Fuente: Elaboracion propia con datos del Wod Bank",  
        x = "Año",  
        y = "logatimos del PIB",  
        colour = "PIB:") +  
  scale_color_manual(values = c("Observado" = "#FC4E07", "Estimado" = "#367DA2"))
```