Modelo econométrico aplicado a la economía mexicana

José Julián Mateo López

2022/09/12

Resumen

Después de mas de dos años de pandemia, la economía mundial volvió a sufrir otro evento catastrófico para su recuperación económica. En el 2020 la economía mundial sufrió una de las peores crisis económicas en la historia, debido a los estragos económicos y sociales generados por la pandemia. Entre las principales consecuencias de la pandemia, se encuentra el aumento significativo de la pobreza y pobreza extrema, la caída generalizada de la actividad económica, perdida masiva de empleos, perdida de ingresos per cápita, el aumento de la deuda externa, entre muchos otros impactos económicos y sociales. Sin embargo, debido al conflicto bélico entre la Federación Rusia y Ucrania la economía mundial se encuentra nuevamente en riesgo, con altos niveles de inflación, riesgos financieros, aumento de la inseguridad alimentaria y la descomunal acumulación de la deuda externa, principalmente en las Economías de Mercados Emergentes y en Desarrollo. Aquí se destaca de la importancia del crecimiento económico actualmente, es por ello, que en este trabajo nos tomamos un espacio para analizar el crecimiento económico de México a través de un tratamiento estadístico matemático. El objetivo de este trabajo es elaborar y desarrollar un modelo econométrico de regresión lineal múltiple capaz de explicar el comportamiento de la economía mexicana durante el periodo 2005-2020, ayudados por el lenguaje de programación R y su entorno de trabajo Rstudio.

ESPECIFICACIÓN DEL MODELO

La metodología econométrica plantea cuatro etapas en la elaboración de un modelo econométrico. Una primer etapa es la especificación del modelo, misma que se encargar de especificar que fenómeno se quiere explicar y en función de que variables lo queremos hacer. En este documento, se pretende explicar el Producto Interno Bruto de México en función de la Formación Bruta de Capital Fijo, la Población Económicamente Activa, el Indice Nacional de Precios al Consumidor y el Tipo de Cambio Real Bilateral a precios constantes en una frecuencia trimestral durante el periodo 2005-2020.

Obtención de la información

Como primer paso para estimar el modelo propuesto, es necesario importar la base de datos al programa, para posteriormente indicarle a R que se esta trabajando con variables de series de tiempo. Para ello, ejecutamos la siguiente linea de código. Primero, debemos activar la librería **readxl** que me permita cargar mi archivo en formato excel. La información (datos) recabada ha sido extraída del Instituto Nacional de Estadística y Geografía (INEGI) y el Banco de México (Banxico). Por lo que, al trabajar con datos económicos reales, el análisis econométrico requería un mayor esfuerzo y dedicación de nuestra parte.

```
library(readxl)
datos <- read_excel("Modelo multiple 2022.xlsx", sheet = "Hoja3")
datos_n <- datos[c("PIB","FBCF","PEA","INPC","TCRB")]
print(datos_n)
```

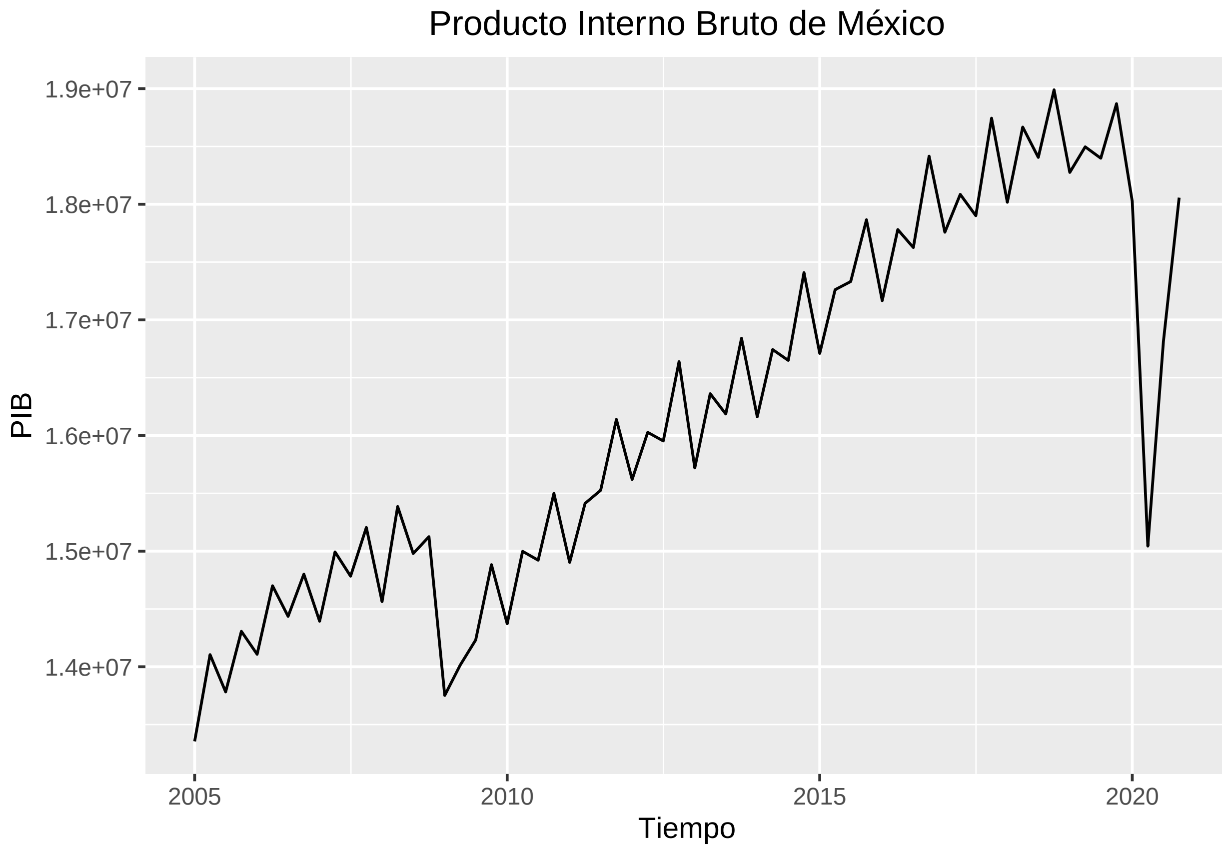
```
## # A tibble: 64 x 5
##          PIB     FBCF      PEA  INPC  TCRB
##        <dbl>    <dbl>    <dbl> <dbl> <dbl>
##  1 13354788. 2680263. 43099847  58.5  85.9
##  2 14104834. 2723325. 43180433  58.9  85.1
##  3 13782144. 2766146. 44000204  59.1  83.6
##  4 14306524. 2927996. 44245519  59.9  83.0
##  5 14107960. 2927578. 44306012  60.7  81.3
##  6 14700504. 2960384. 44611672  60.7  87.2
##  7 14435868. 3064764. 45431392  61.2  85.4
##  8 14800897. 3178540. 45580994  62.3  82.6
##  9 14393727. 3044761. 45314888  63.2  83.3
## 10 14993339. 3080037. 45569395  63.1  84.0
## # ... with 54 more rows
```

```
attach(datos)
```

Análisis Exploratorio de Datos

Una vez importada la base de datos, estamos en condiciones de llevar a cabo las estimaciones pertinentes. Sin embargo, como paso previo a la estimación del modelo, llevaremos a cabo el tratamiento previo de datos o análisis exploratorio de datos como técnicamente se le conoce. En esta sección se realizará la metodología básica para analizar el comportamiento estadístico de la variable endógena o la variable explicativa del modelo. Para ello, es necesario activar las librerías que ocuparemos.

```
library(tidyverse)
startday = as.Date("2005-1-1")
tiempo <- as.Date(seq(startday, by = "3 month", length.out = 64))
ggplot(data = datos, aes(x = tiempo, y = PIB))+geom_line()+ggtitle("Producto Interno Bruto de México")+
```



Una de las principales dificultades a la hora de trabajar un modelo de este tipo con R, es darle un formato de series de tiempo a los objetos, al menos para una persona primera en el aprendizaje del lenguaje. En este sentido, es necesario darle el formato de **time series** a las variables a utilizar en el modelo, para que mas adelante en la estimación del modelo, el lenguaje sepa que se están trabajando con series de tiempo.

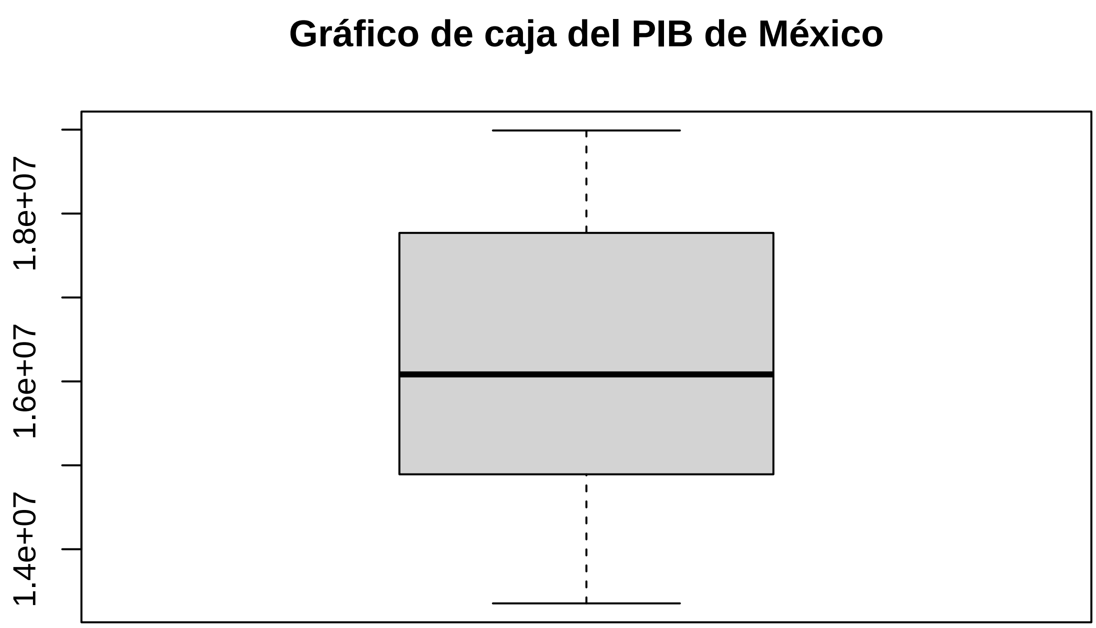
```
library(tseries)
library(psych)
PIB_ts <- ts(PIB, start = 2005, frequency = 4)
FBCF_ts <- ts(FBCF, start = 2005, frequency = 4)
INPC_ts <- ts(INPC, start = 2005, frequency = 4)
TCRB_ts <- ts(TCRB, start = 2005, frequency = 4)
PEA_ts <- ts(PEA, start = 2005, frequency = 4)
```

En el resultado siguiente se pueden apreciar las principales estadísticas de la serie a explicar. Para ello, se ha seleccionado la serie Producto Interno Bruto de México en una frecuencia trimestral. Para el periodo en el que se analiza la serie presentó un valor medio de (mean) $16,198,346 (millones de pesos a precios de 2013) y un valor de la mediana (median) muy similar $16,083,502 (millones de pesos a precios de 2013). El valor máximo (max) alcanzado en la serie es de $18,990,033 (millones de pesos a precios de 2013) en el cuarto trimestre de 2018. Por su parte, el valor minimo (min) es de $13,354,788 (millones de pesos a precios de 2013) que corresponde al segundo trimestre de 2002. La desviación típica alcanza un valor de $1,592,940 (millones de pesos a precios de 2013). Otro de los estadísticos básicos para comprender el comportamiento estadístico de una serie, son los estadísticos de asimetría (skew) y curtosis (Kurtosis) que recogen los valores respectivos de 0.13 y -1.32.

```r
describe(PIB_ts)
```

```
##    vars  n     mean      sd   median  trimmed     mad      min      max   range
## X1    1 64 16198346 1592940 16083502 16180658 1946278 13354788 18990033 5635245
##    skew kurtosis       se
## X1 0.13    -1.32 199117.5
```

```r
boxplot(PIB_ts, main = "Gráfico de caja del PIB de México")
```



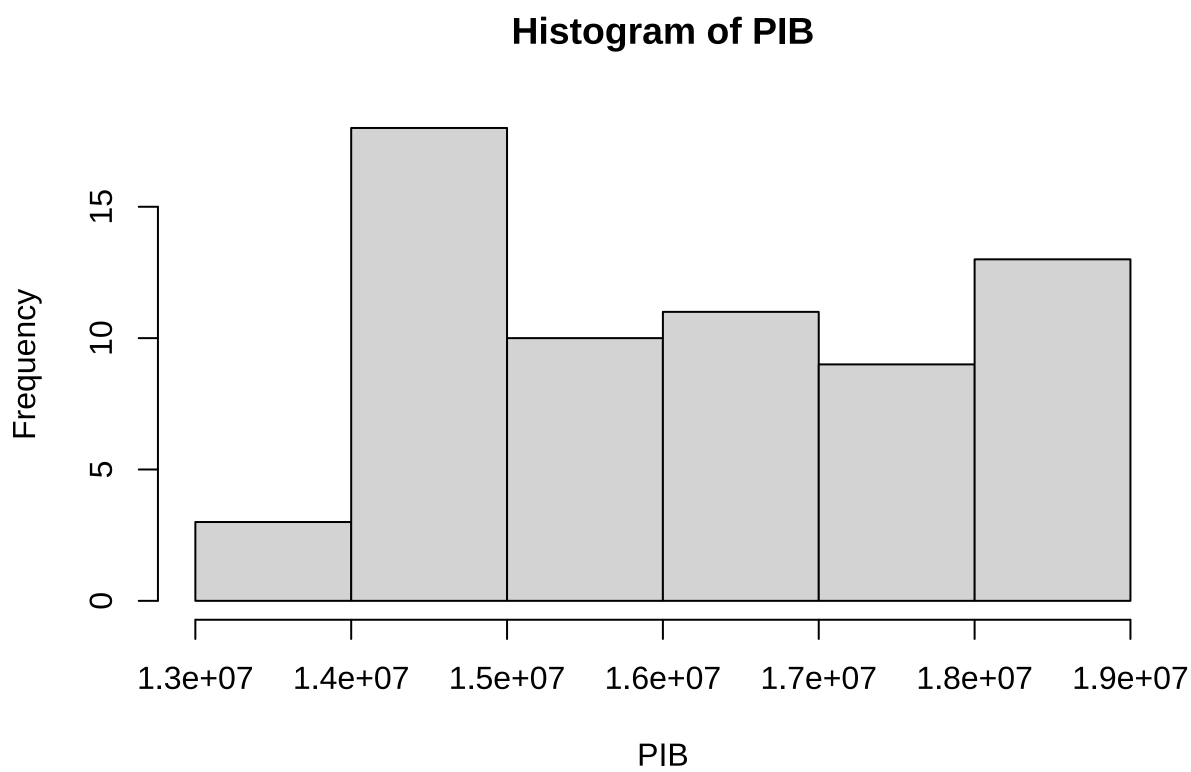
El estadístico Jarque-Bera es una prueba de bondad de ajuste para comprobar si una muestra de datos tiene la asimetria y la curtosis de una distribución normal. Bajo la hipótesis nula de una distribución normal, el estadístico jarque-bera se distribuye como una ji-cuadra con 2 g.de l. que en términos aproximados toma el valor de 6. En este sentido, se plantea como hipótesis nula que Ho: La serie se distribuye como una normal, en contraste, como hipótesis alternativa: Ha: La serie no se distribuye como una normal. La regla de decisión consiste en que si la probabilidad asociada al estadístico Jarque-Bera es mayor al 5% (0.05), no recharemos la hipótesis nula y afirmaremos que la serie PIB de México se distribuye como una normal. Como se puede apreciar en la prueba estadística, la probabilidad es mayor al 5%, al alcanzar un valor del 10%, por lo tanto, no se rechaza hipótesis nula.

```r
jarque.bera.test(PIB_ts)
```

```
## 
##  Jarque Bera Test
## 
```

```
## data:  PIB_ts
## X-squared = 4.5073, df = 2, p-value = 0.105
```

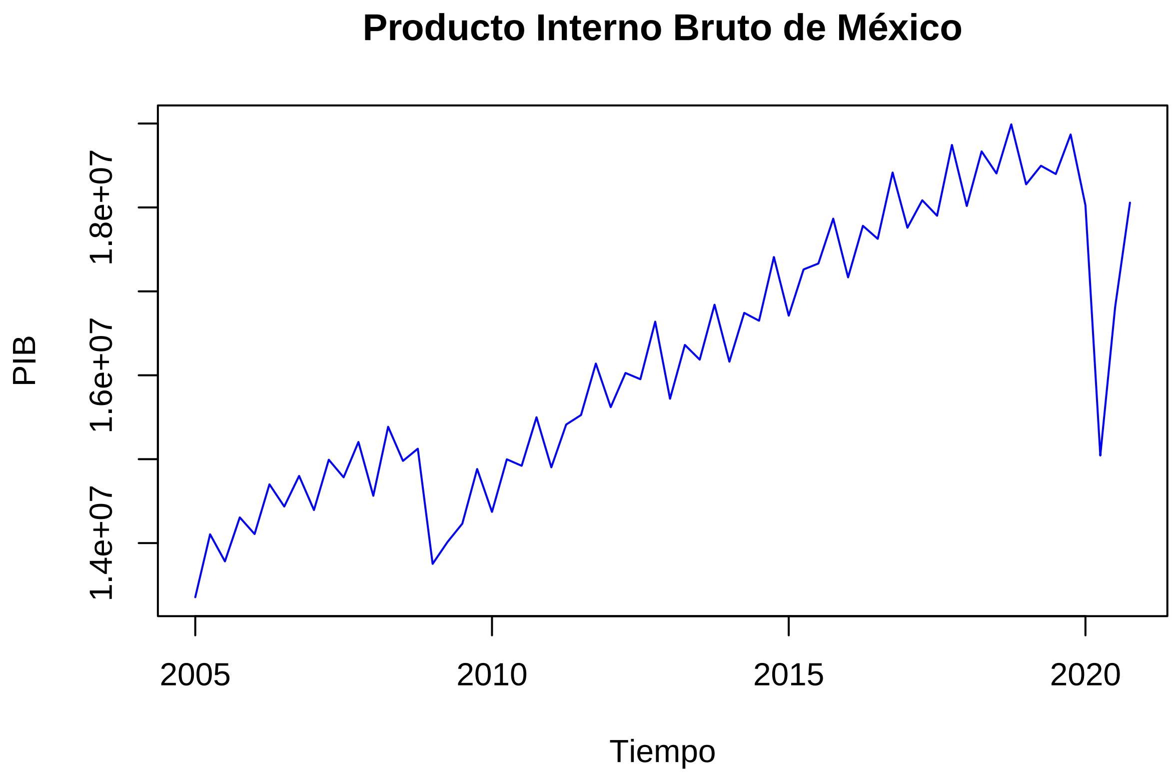
```
hist(PIB_ts, main = "Histogram of PIB", xlab = "PIB")
```



Determinacion de la estacionariedad de la serie

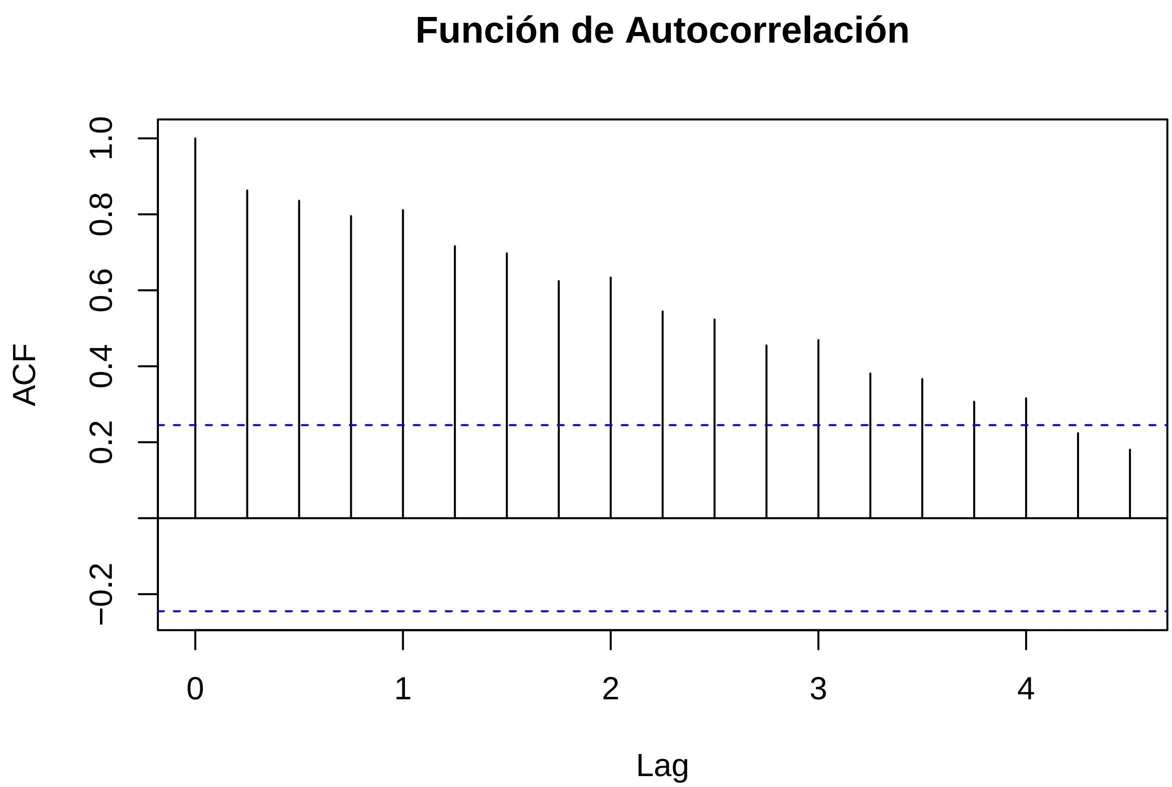
Una serie es estacionaria cuando su media y varianza son constantes a través del tiempo, es decir, son invariantes al tiempo. Si bien es cierto que existe una amplia conceptualización matemática detrás de este concepto econométrico, en este trabajo nos limitaremos únicamente a la aplicación de las pruebas estadísticas necesarias para determinar la estacionariedad de la serie. Una de las principales recomendaciones en el análisis de series de tiempo, es el análisis gráfico. El análisis gráfico representa la metodología básica para comprender el comportamiento estadístico de una serie. En la siguiente linea de código se puede apreciar la evolución temporal del Producto Interno Bruto de México.

```
plot(PIB_ts, main = "Producto Interno Bruto de México", xlab = "Tiempo", ylab = "PIB", col = "Blue")
```



En la gráfica anterior se puede observar la amplia volatilidad presentada en la serie. Se puede apreciar que la serie no presenta una media y varianza constantes, principalmente en los años 2008-2009 y en el 2020. Esto se le puede atribuir a las crisis económicas de dichos años. Recordemos que en el periodo de análisis, la economía mexicana ha pasado por fuertes e inestables ciclos económicos. Aun así, no esta demás realizar pruebas estadísticas mas formales, por poner de ejemplo, la función de autocorrelación parcial. Para obtener la función de autocorrelación ejecutamos la siguiente linea de código.

```
library(stats)
acf(PIB_ts, main = "Función de Autocorrelación")
```



Como se aprecia en la función de autoccorelación, la serie seleccionada parece presentar una raíz unitaria. El gráfico anterior es la representación típica de una serie no estacionaria, es decir, una serie de caminata aleatoria, como se le conoce en econometría de series de tiempo. Como podemos observar los coeficientes de correlación parcial disminuyen paulatinamente, por lo que, podemos asumir que la serie es no estacionaria. Es decir, su media y varianza no se distribuyen alrededor de una constante. Entonces, el Producto Interno Bruto de México presenta una tendencia estocástica.

Prueba de Dickey-Fuller Aumentada

Para la determinación de la estacionaridad de la serie llevaremos a cabo la aplicación de la prueba de Dickey-Fuller. Para ello, importamos la librería **urca**. Se plantea como hipótesis nula que existe una raíz unitaria, en contraste, como hipótesis alternativa Ha: No existe raíz unitaria. Como se puede observar, el estadístico calculado del test ADF recoge el valor de 0.91, por lo que, no rechazamos la hipótesis nula en los 3 niveles de significancia estadística (1%,5% y 10%) y afirmamos que la serie presenta al menos una raíz unitaria. Es decir, su media y varianza no son constante, para posteriormente ser una serie integrada de orden 1 o I(1).

```
library(urca)
summary(ur.df(PIB_ts, type = "none", selectlags = "AIC"))
```

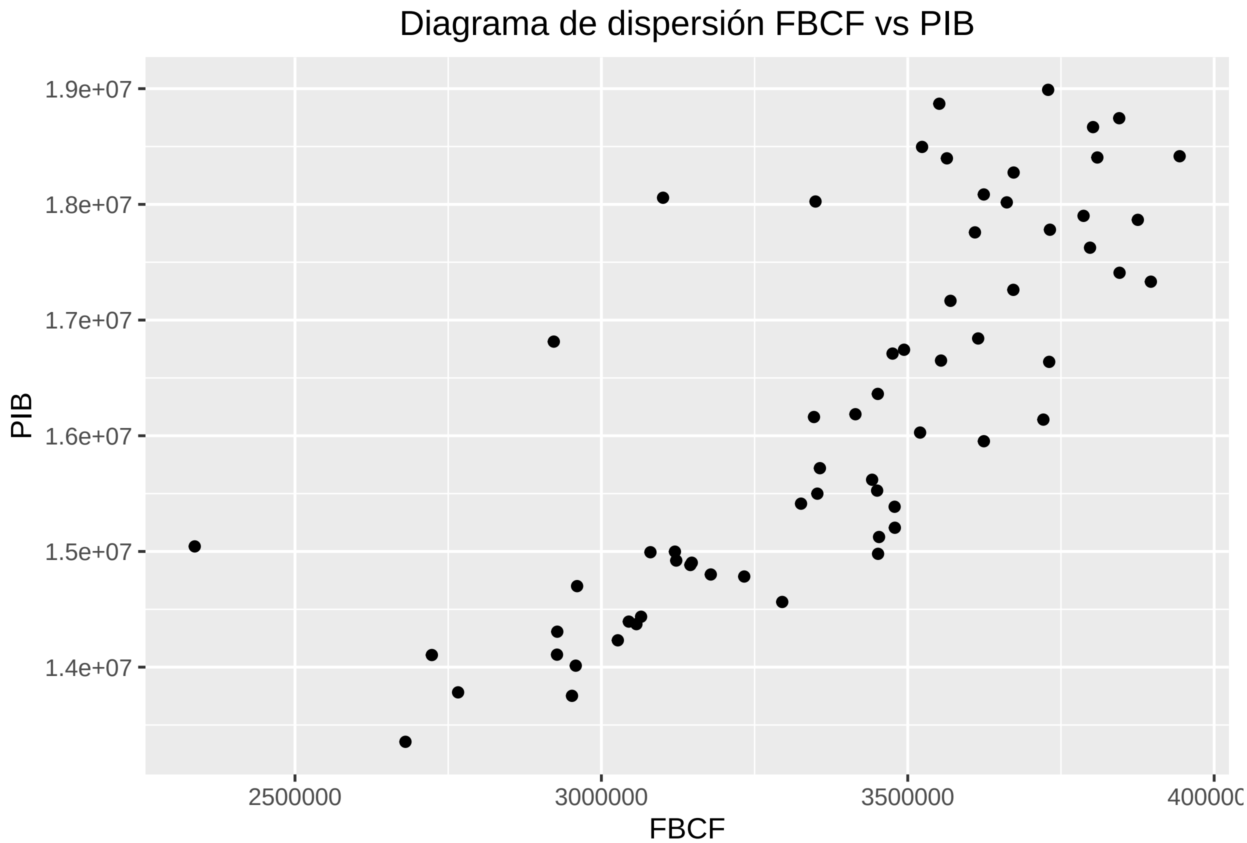
```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
```

```
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3436104  -266351    94546   340361  1940051
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## z.lag.1     0.004698   0.005133   0.915 0.363661
## z.diff.lag -0.438103   0.118666  -3.692 0.000482 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 655300 on 60 degrees of freedom
## Multiple R-squared:  0.1887, Adjusted R-squared:  0.1616
## F-statistic: 6.976 on 2 and 60 DF,  p-value: 0.001888
##
##
## Value of test-statistic is: 0.9154
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

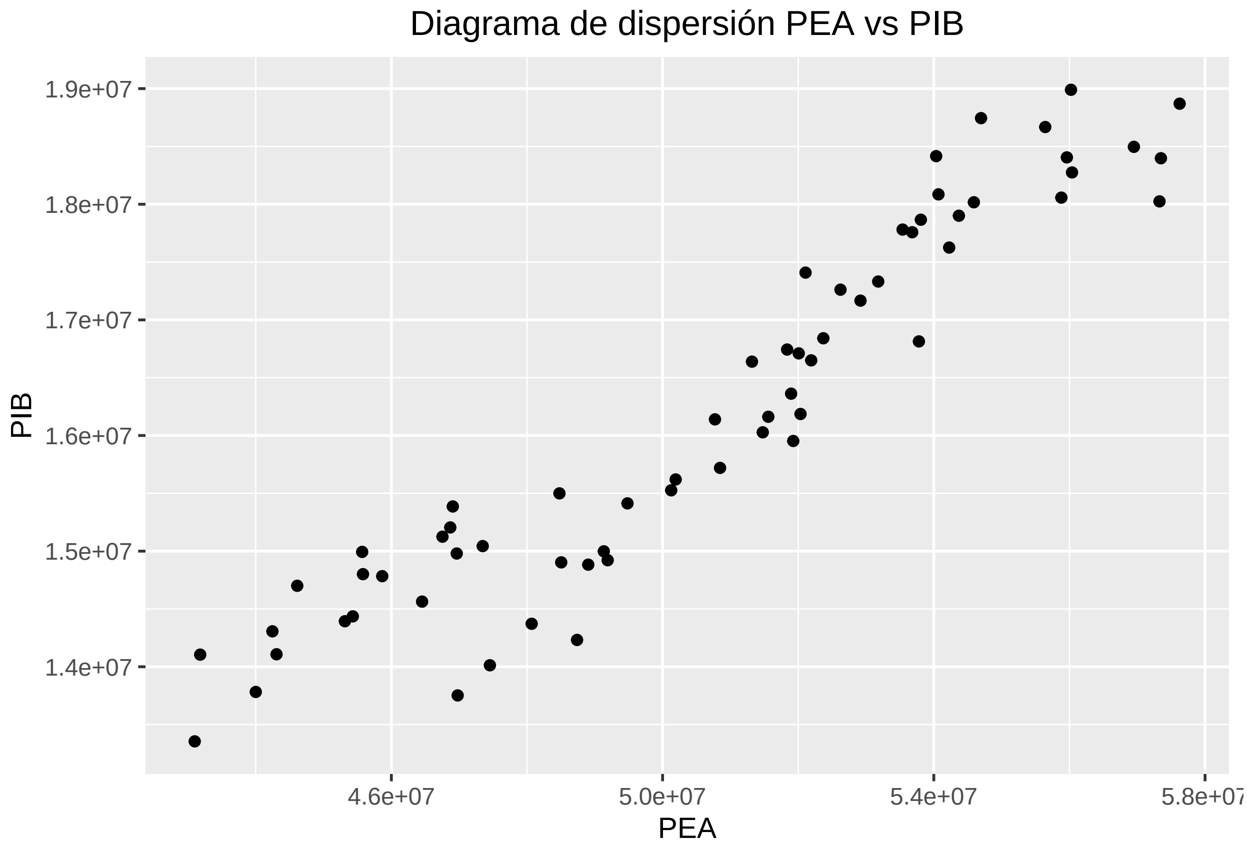
Diagramas de dispersión

Con el objetivo de analizar la relación que existe entre la variable endógena y las variables explicativas del modelo, procedemos a realizar los diagramas de dispersión, mismo que son de gran utilidad para comprender las relaciones inexactas de las variables del modelo.

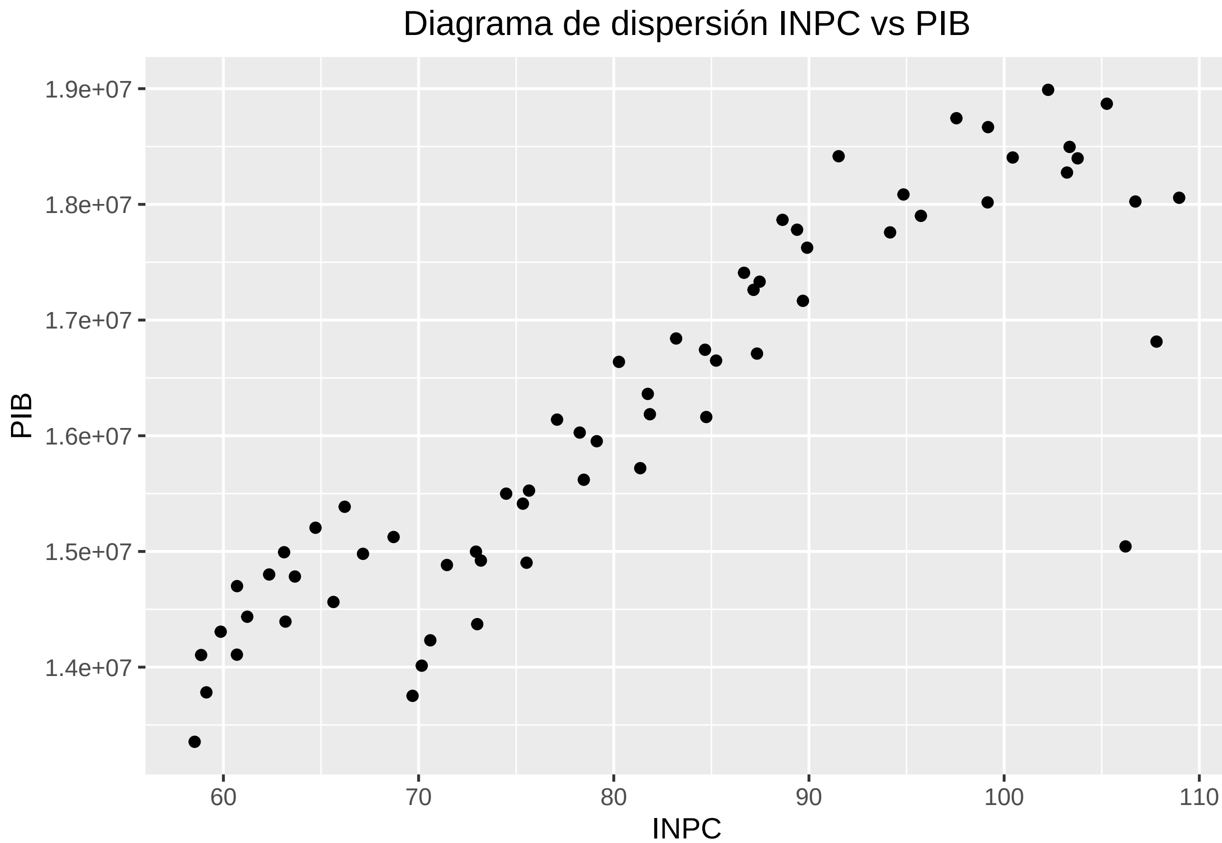
```
library(graphics)
par(mfrow=c(2,2), mar = c(4,4,4,1) + .1)
ggplot(data = datos, aes(x = FBCF, y = PIB))+
  geom_point()+ggtitle("Diagrama de dispersión FBCF vs PIB")+
  theme(plot.title = element_text(hjust = 0.5))
```



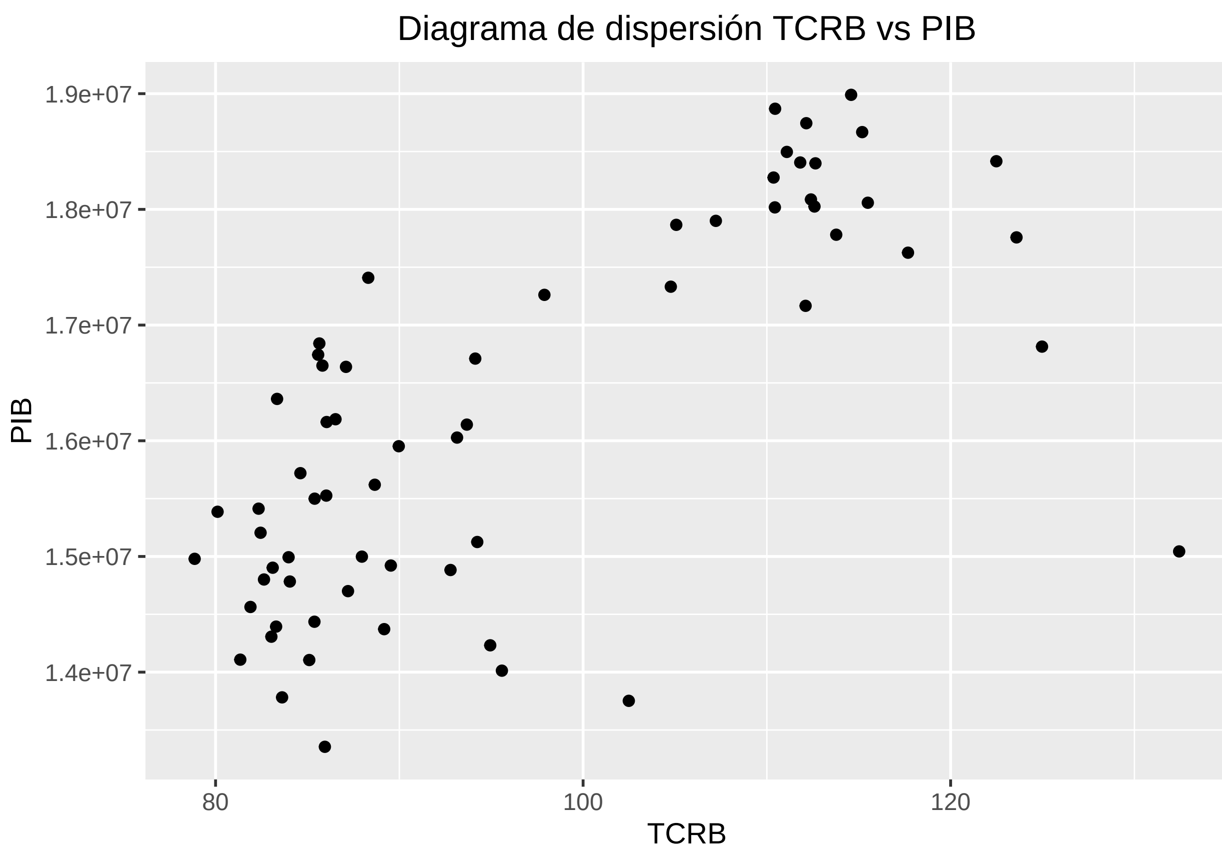

```
ggplot(data = datos, aes(x = PEA, y = PIB))+
  geom_point()+ggtitle("Diagrama de dispersión PEA vs PIB")+
  theme(plot.title = element_text(hjust = 0.5))
```




```r
ggplot(data = datos, aes(x = INPC, y = PIB))+
  geom_point()+ggtitle("Diagrama de dispersión INPC vs PIB")+
  theme(plot.title = element_text(hjust = 0.5))
```




```
ggplot(data = datos, aes(x = TCRB, y = PIB))+
  geom_point()+ggtitle("Diagrama de dispersión TCRB vs PIB")+
  theme(plot.title = element_text(hjust = 0.5))
```



MATRIZ DE CORRELACIÓN

```
datos_n <- datos[c("PIB","FBCF","PEA","INPC","TCRB")]
cor(datos_n)
```

```
##            PIB      FBCF       PEA      INPC      TCRB
## PIB  1.0000000 0.7778353 0.9453630 0.8862566 0.7012421
## FBCF 0.7778353 1.0000000 0.7141709 0.4859116 0.2834881
## PEA  0.9453630 0.7141709 1.0000000 0.9246693 0.6917029
## INPC 0.8862566 0.4859116 0.9246693 1.0000000 0.8303773
## TCRB 0.7012421 0.2834881 0.6917029 0.8303773 1.0000000
```

En la matriz anterior, se puede observar las correlaciones lineales entre las variables del modelo. Una matriz de correlación es de gran ayuda para determinar a través del coeficiente de correlación de pearson el nivel de asociación que existe entre las variables. Los resultados indican una relación positiva entre el PIB y la FBCF, recogiendo un coeficiente de 0.77. Por su parte, la relación entre el PIB y la PEA es muy buena, excelente al parecer, arrojando un coeficiente de correlación de 0.94. También se puede observar que existe una asociación lineal fuerte con el resto de variables, recogiendo un nivel de asociación 0.88 respecto a la inflación y 0.70 con el Tipo de Cambio. Los resultados aquí presentados, parecen corroborar lo planteado por los diagramas de dispersión. Por lo que, podemos concluir que las regresoras del modelo parecen aportar un amplio poder explicativo al modelo.

Estimación del modelo

```
library(lmtest)
library(dynlm)
modelo <- lm(PIB_ts ~ FBCF_ts + PEA_ts + INPC_ts + TCRB_ts)
summary(modelo)
```

```
##
## Call:
## lm(formula = PIB_ts ~ FBCF_ts + PEA_ts + INPC_ts + TCRB_ts)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -885966 -254050   -5312  196864  847427
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.619e+06  1.335e+06   1.213   0.2301
## FBCF_ts     1.889e+00  2.626e-01   7.194 1.27e-09 ***
## PEA_ts      6.349e-02  5.437e-02   1.168   0.2476
## INPC_ts     4.956e+04  1.417e+04   3.497   0.0009 ***
## TCRB_ts     9.820e+03  6.591e+03   1.490   0.1416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 383000 on 59 degrees of freedom
## Multiple R-squared:  0.9459, Adjusted R-squared:  0.9422
## F-statistic: 257.7 on 4 and 59 DF,  p-value: < 2.2e-16
```

En la ecuación anterior se puede apreciar la primer regresión con datos originales. Uno de los principales estadísticos a la hora de seleccionar un modelo econométrico es el R2, el cual me determina la bondad de ajuste del modelo o su capacidad explicativa. Sin embargo, no es el único elemento a tomar en cuenta, por su parte, podemos evaluar un modelo de este tipo por su parte econométrica y económica. La parte econométrica consiste en la aplicación de múltiples pruebas estadísticas que me permitan verificar el correcto funcionamiento del modelo, por otro lado, podemos evaluar la parte económica por las magnitudes y signos presentados por los coeficientes parciales de regresión esperados por la teoría económica. Respecto a la significancia estadística individual, podemos decir que únicamente la Formación Bruta de Capital y el Indice Nacional de Precios al Consumidor son estadisticamente significativos al numero de referencia 2 (5%), por su parte, tanto el intercepto como la PEA y el TCRB no estadisticamente significativos. La bondad de ajuste es muy buena, recogiendo un R2 corregido de 0.94, es decir, la variación del PIB de México queda explicada en un 94% por las variables regresoras del modelo. La prueba F, me indica que el modelo es estadíscamente significativo a nivel global.

Ecuación 2

En el trabajo aplicado se recomienda ampliamente trabajar con logaritmos neperianos con el objetivo de obtener mejores resultados. Para ello, transformamos las variables originales en términos logarítmicos.

```
logPIB <- log(PIB_ts)
logFBCF <- log(FBCF_ts)
logPEA <- log(PEA_ts)
```

```r
logINPC <- log(INPC_ts)
logTCRB <- log(TCRB_ts)
modelo2 <- lm(logPIB ~ logFBCF + logPEA + logINPC + logTCRB)
summary(modelo2)
```

```
##
## Call:
## lm(formula = logPIB ~ logFBCF + logPEA + logINPC + logTCRB)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.070200 -0.017823  0.000063  0.013513  0.055915
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.89541    2.55478   2.699  0.00906 **
## logFBCF      0.38009    0.05740   6.622 1.18e-08 ***
## logPEA       0.14087    0.19451   0.724  0.47179
## logINPC      0.25386    0.07796   3.256  0.00187 **
## logTCRB      0.08274    0.04124   2.006  0.04944 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02521 on 59 degrees of freedom
## Multiple R-squared:  0.9384, Adjusted R-squared:  0.9342
## F-statistic: 224.8 on 4 and 59 DF,  p-value: < 2.2e-16
```

En la ecuación de predicción 2 se puede apreciar que únicamente el logaritmo de la Población Económicamente activa no es estadisticamente significativa a los tres niveles habituales de significancia. Por su parte, respecto a la teoría económica, tanto el Indice Nacional de Precios al Consumidor como el Tipo de Cambio Real no cumplen con lo formulado por la teoría, puesto que, ante un incremento de estos fenómenos económicos, se esperaría una disminución del PIB. La bondad de ajuste del modelo es muy buena, recogiendo un R2 corregido de 0.93, es decir, la variación del LOGPIB queda explicado en un 93% por las variables explicativas del modelo. La prueba F, nos indica que efectivamente el modelo es estadisticamente significativo globalmente. Ante estos deficiente resultados, nos vemos en la necesidad de seguir realizando diversas aplicaciones, de forma que nos permita encontrar la recta de mejor ajuste, que se ajuste a los datos.

Ecuación 3

```r
modelo3 <- dynlm(logPIB ~ logFBCF + logPEA + stats::lag(logPIB,-1) + stats::lag(logFBCF,-1))
summary(modelo3)
```

```
##
## Time series regression with "ts" data:
## Start = 2005(2), End = 2020(4)
##
## Call:
## dynlm(formula = logPIB ~ logFBCF + logPEA + stats::lag(logPIB,
##     -1) + stats::lag(logFBCF, -1))
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.030577 -0.016706 -0.001253  0.013018  0.048249
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -1.32143    0.64912  -2.036  0.04636 *
## logFBCF                 0.42544    0.05025   8.467 1.02e-11 ***
## logPEA                  0.29027    0.09291   3.124  0.00278 **
## stats::lag(logPIB, -1)  0.74936    0.08598   8.716 3.92e-12 ***
## stats::lag(logFBCF, -1) -0.40314    0.05950  -6.775 6.99e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02038 on 58 degrees of freedom
## Multiple R-squared:  0.9579, Adjusted R-squared:  0.955
## F-statistic: 330.2 on 4 and 58 DF,  p-value: < 2.2e-16
```

En la ecuación 3 se puede apreciar la regresión con la Formación Bruta de Capital Fijo, la Población Económicamente Activa, un termino autorregresivo del logaritmo del PIB y un termino autorregresivo del logaritmo de la Formación Bruta de Capital Fijo como variables explicativas del modelo. Aquí es importante mencionar la importancia de estimar el modelo con la función **Dynamic Linear Models and Time Series Regression**, ya que R no interpretaría los términos autorregresivos agregados a la ecuación. Como podemos observar, los coeficientes de regresión parcial son estadisticamente significativos al numero de referencia 2, es decir, los estimadores son estadisticamente significativos a un nivel del 5%. Los signos de las pendientes son los esperados por la teoría económica. Ante un incremento de la inversión, se esperaría un incremento del Producto Interno Bruto. De igual forma, ante un incremento en una unidad de la PEA, se esperaría un incremento del PIB de México. La bondad de ajuste del modelo es muy buena, recogiendo un valor de 0.95, es decir, las variables explicativas explican la variación de la variable dependiente en un 95%. La prueba F nos indica que el modelo es estadisticamente significativo a nivel global.

Una vez establecidos los criterios de evaluación, y teniendo en cuenta que la ecuación se puede mejorar, los resultados de la ecuación 3 parecen ya relativamente satisfactorios al menos para un ejercicio de este tipo. Por lo que, procederemos a le evaluación del modelo.

Contrastes de hipótesis sobre la perturbación aleatoria

No normalidad

En esta sección nos encargaremos de la metodología básica para el análisis del comportamiento de la perturbación aleatoria. Como se ha comentado anteriormente, detrás de todos estos cálculos que hace R, hay una amplia conceptualización matemática, no obstante, en esta sección únicamente se llevaran a cabo las pruebas estadísticas y la interpretación de los resultados, que nos permitan determinar el correcto funcionamiento del modelo, dejando de lado (al menos por el momento) el tratamiento estadístico matemático. El modelo clásico de regresión lineal plantea el supuesto de que las perturbaciones Ui se distribuyen como una normal, con media 0 y varianza constante. Por lo que, para el contraste sobre la normalidad de los errores, procederemos a la aplicación del test de Jarque-Bera. Para ello, realizamos el siguiente comando.

```r
resid <- modelo3$residuals
resid <- ts(resid, start = 2005, frequency = 4)
jarque.bera.test(resid)
```

```
##
```

```
##  Jarque Bera Test
##
## data:  resid
## X-squared = 2.9676, df = 2, p-value = 0.2268
```

Se plantea como hipótesis nula, HO: Normalidad en el modelo, en contraste, Ha: No hay normalidad en el modelo. La regla de decisión, esta determinado, como en pruebas anteriores en el valor de la probabilidad asociada al estadístico Jarque-Bera, si esta probabilidad es menor al 5%, se rechaza la hipótesis nula, y se asume que los residuos del modelo no siguen un distribución normal. Como se puede apreciar el P-value del test es de 0.22, por lo que, no se rechaza la hipótesis nula, y asumimos que los errores se distribuyen como una normal.

Tratamiento del riesgo y la varianza

tratamiento de la varianza en modelos uniecuacionales: Heterocedasticidad

Una de las hipótesis básicas que plantea el modelo clásico de regresión lineal es que la varianza de los errores debe ser constantes a través del tiempo. Es decir, el modelo cumple la característica de ser homocedástico. La presencia de heterocedasticidad se puede deber a diversas razones, por ejemplo, a una especificación errónea del modelo (ya sea por inclusión de variables irrelevantes, una forma funcional errónea o exclusión de variables relevantes) o bien se podría dar el caso de inclusión de valores atípicos. No obstante esto no es muy común en modelos con datos de series de tiempo.

En estadística se utiliza el test de Breusch-Pagan para determinar la heterocedasticidad en un modelo de regresión lineal. Para determinar la existencia de dicha propiedad econométrica, ejecutamos el test a traves del siguiente comando.

```
bptest(modelo3)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  modelo3
## BP = 3.5484, df = 4, p-value = 0.4706
```

Ho: El modelo es homocedastico

Ha: El modelo es heterocedastico

P-value > 0.0.5, por lo tanto, no se rechaza la hipótesis nula, y afirmamos que los residuos del modelo siguen una varianza constante. Un segundo contraste realizado en este trabajo es el test de ARCH. Para aplicación de este test es necesario indicar el numero de retardos o rezagos que queremos incluir en la ecuación de contraste, se evalúa con 12 rezagos.

```
library(FinTS)
ArchTest(resid, lags = 12, demean = FALSE)
```

```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  resid
## Chi-squared = 16.431, df = 12, p-value = 0.1723
```

Ho: Los residuos no siguen un comportamiento tipo ARCH

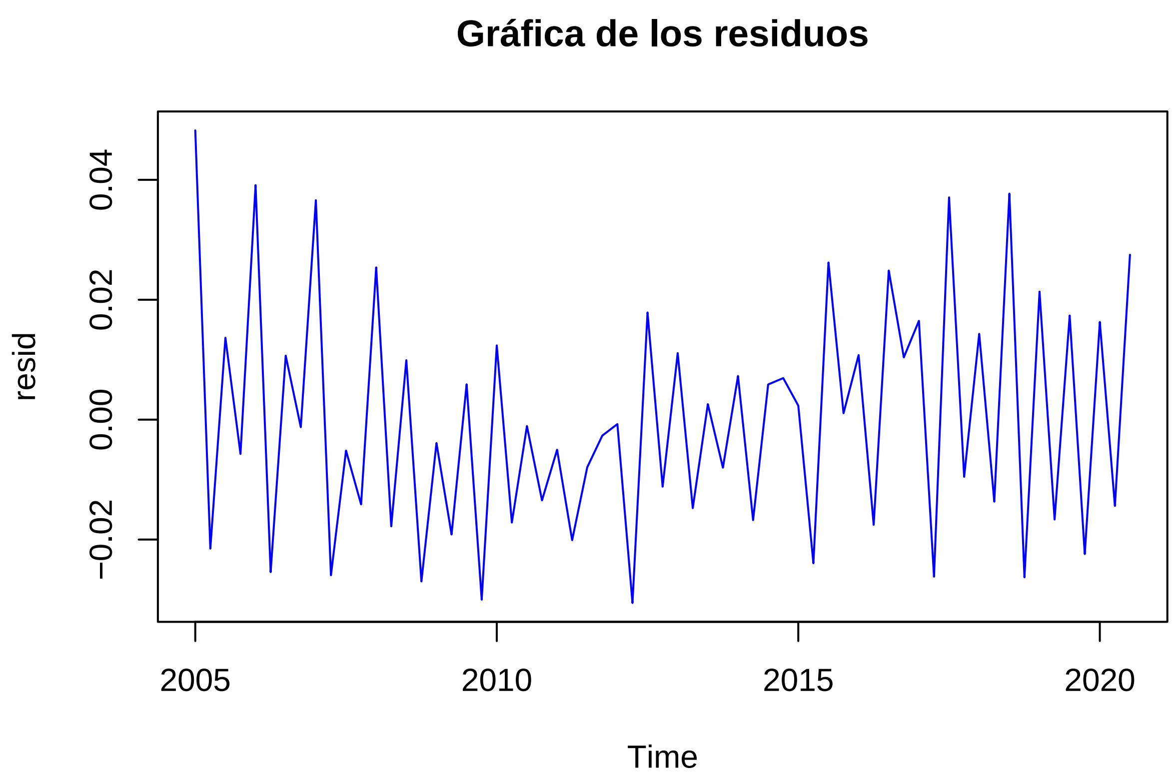
Ha: Los residuos siguen un comportamiento tipo ARCH

Regla de decisión: Si la prob.Asoc > 0.05, entonces no se rechaza la hipótesis nula. Como se puede observar, la probabilidad asociada al test de ARCH es mayor al 5%, por lo que, no se rechaza la hipótesis nula. Es decir, los residuos del modelo no siguen un comportamiento tipo ARCH. Si bien, en la literatura existe una amplia gama de pruebas estadísticas para este contraste, parece que los resultados alcanzados parecen ya satisfactorios para afirmar que el modelo es homocedastico.

Autocorrelación

Una vez verificado dos de las hipótesis básicas que plantea el modelo clásico de regresión, lo que nos queda, respecto a la evaluación perturbación estocástica es verificar que los residuos del modelo no estén correlacionados. Una de las propiedades econométricas que plantea el MCRL es que cada Ui esta independiente distribuida y no esta autocorrelacionada. Uno de los principales problemas a la hora de trabajar con un modelo de series de tiempo es la autocorrelación residual. Esto se puede deber a múltiples causas, por ejemplo, a una mala especificación del modelo, la no estacionariedad de las series o el incuplimiento de alguna del resto de las hipótesis.

```r
resid <- modelo3$residuals
resid <- ts(resid, start = 2005, frequency = 4)
plot(resid, main = "Gráfica de los residuos", col = "blue")
```



Una de las pruebas clasicas para determinar este contraste es el estadístico Durbin-Whatson. No obstante, este prueba estadística me permite verificar la autocorrelación de primer orden, no de un orden mayor. Para la aplicación del test de Whatson utilizaremos la libreria "lmtest".

```
dwtest(modelo3)
```

```
##
##  Durbin-Watson test
##
## data:  modelo3
## DW = 2.9993, p-value = 1
## alternative hypothesis: true autocorrelation is greater than 0
```

El estadístico es de 2.99, lo que al parecer nos indica un posible problema de autocorrelación. La teoría nos indica que para que existe duda de autocorrelación el estadístico de Whatson debe ser igual a 2 o por lo menos, cercano a el. Por lo que, podemos afirmar que el modelo presenta problemas de autocorrelación. Una prueba de carácter mas general es la denominada la prueba de Breusch-Godfrey o la prueba del multiplicador de lagrange. La cual queda especificado con el siguiente comando, en la cual es necesario que introduzcamos el numero de retardos con la que la queremos contrastar. Para el caso, evaluamos el modelo3 con 12 rezagos.

```
bgtest(modelo3, order = 12)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 12
##
## data:  modelo3
## LM test = 43.482, df = 12, p-value = 1.87e-05
```

Como se habrá dado cuenta el lector, en la mayoría de las pruebas aplicadas en este trabajo, los estadísticos vienen acompañados de una probabilidad asociada. En este sentido, si el P-value del estadístico de Breusch-Godfrey es mayor al 5%, podemos afirmar que el modelo evaluado con 12 rezagos no presentar problemas de autocorrelación. Como se puede apreciar, la probabilidad asociada es mucho menor al 0.05% solicitado. Se plantea como hipótesis: Ho: Hay autocorrelación en los residuos Ha: No hay autocorrelación en los residuos De acuerdo con los resultados presentados, no se rechaza la hipótesis nula, por lo que, los errores del la ecuación de predicción 3 están correlacionados.

Contraste de hipótesis estructurales

Una vez evaluada la parte aleatoria del modelo, podemos proceder a la evaluación de su parte determinista del mismo. Si bien es cierto, no es necesario continuar con el desarrollo del método de predicción, debido al imcuplimiento de una de las hipótesis básicas del modelo clásico de regresión lineal, y lo correcto sería corregir este problema, pero únicamente lo haremos para el lector pueda ver como se aplican este tipo de contraste en el lenguaje. Por consiguiente, en esta sección se presenta la metodología básica para evaluar los supuestos estructurales.

Especificación erronea del modelo

Una de las partes mas difíciles en la elaboración de un modelo econométrico, a mi parecer, es determinar la correcta especificación del mismo. Cuando el método de predicción es validado por la teoría económica, pero existe cierta discrepancia respecto a las pruebas o supuestos, nos podría estar indicando que existe una especificación errónea del modelo. En la literatura econométrica, una especificación errónea del modelo puede deberse a múltiples causas, que por lo general, se debe a tres causas fundamentales. 1) La inclusión de variables irrelevantes, 2) la omisión de variables relevantes y por ultimo, 3) una forma funcional incorrecta en el sentido de la no linealidad de las variables. En un primer momento, la teoría económica nos permite especificar el modelo a estimar, sin embargo, en el trabajo aplicado, esto puede resultar un tanto complicado y tedioso. Otra de las posibles razones de una incorrecta especificación es el incumplimiento de algunos de los supuesto del termino de errónea (autoccorelación, normalidad y heterocedasticidad), e incluso, podríamos destacar aquí la importancia del papel que juegan los datos. En ocasiones, debido a los datos el modelo pide a gritos una transformación, y aquí esta el asunto, determinar la forma funcional que mas se adecue al modelo, tal vez no sea un regresión lineal, sino logarítmica, logística e incluso se necesito una metodología no lineal.

Recordemos, que los residuos del modelo están autocorrelacionados, por lo que, no nos debería sorprender que el modelo no esta especificado correctamente. Una de las principales pruebas para verificar la correcta forma funcional, en el sentido de la linealidad es la prueba de Ramsey Reset. Para la aplicación de este test ejecutamos la siguiente linea de código.

```
# Prueba de Ramsey Reset a la primer potencia
resettest(modelo3, power = 2:1)
```

```
##
##  RESET test
##
## data:  modelo3
## RESET = 1.1303, df1 = 2, df2 = 56, p-value = 0.3302
```

```
# Prueba de Ramsey Reset a la segunda potencia
resettest(modelo3, power = 2:2)
```

```
##
##  RESET test
##
## data:  modelo3
## RESET = 2.3009, df1 = 1, df2 = 57, p-value = 0.1348
```

```
#Prueba de Ramsey Reset a la tercer potencia
resettest(modelo3, power = 2:3)
```

```
##
##  RESET test
##
## data:  modelo3
## RESET = 5.4056, df1 = 2, df2 = 56, p-value = 0.007136
```

Con base en los resultados obtenidos, el modelo no cumple el supuesto de correcta especificación. Lo cual, era de esperarse.

Multicolinealidad

Para finalizar con este sección, aplicaremos la prueba de Factor de Inflación de Varianza para verificar la colinealidad de las variables regresoras del modelo. Para este contraste, ejecutamos el siguiente comando:

```
library(car)
vif(modelo3)
```

```
##                logFBCF                    logPEA  stats::lag(logPIB, -1) 
##               4.106605                  7.841706               10.602319 
## stats::lag(logFBCF, -1) 
##               6.146289
```

Se puede apreciar que el Factor de Inflación de Varianza no es mayor a 10, por lo que, el modelo cumple que las variables explicativas no tienen alta relacion lineal. Es decir, cumplen esa interdependencia entre si.

CONCLUSIÓN

El trabajo aplicado no se construye de manera mecánica, se necesitan habilidad, intuición y experiencia. Se puede concluir, que el modelo presentado no cumple con las hipótesis básicas del **MCRL**, por lo que, sería conveniente trabajarlo aun mas. Puesto que, no se cumplieron los supuestos básicos, no podemos proceder a la ultima etapa en la elaboración de un modelo econométrico, que es la predicción. El único problema es el supuesto de autoccorelación, que bien se puede deber a una incorrecta especificación del modelo, de hecho, es una de las principales fuentes por las cuales puede existe heterocedasticidad y autocorrelación en un modelo de este tipo. Ahora que si se ha especificado bien el modelo y este aun continua presentando problemas de esta índole, esto se podría corregir a través del método de Minimos Cuadrados Generalizados (MCG). En conclusión, es necesario trabajar aun mas el modelo, tarea que quedará pendiente, al menos, por el momento.