Taller 1: Análisis Índice de Productividad - Canadá

Juan Pablo Calle Quintero

27 de agosto de 2015

Punto 1

En la Figura 1 se muestra la serie mensual del índice de productividad de Canadá entre enero de 1950 y diciembre de 1973. A primera vista se puede observar una clara tendencia creciente del índice de productividad, que además parece ser global , esto lo vemos confirmado en la componente de la tendencia (T_t) que se muentra en la Figura 2, donde vemos que se podría ajustar una curva suave donde los parámetros no dependan del tiempo.

También parece que la tendencia no es lineal, quizá un modelo cuadrático o exponencial ajuste mejor.

Viendo el gráfico de la serie no es muy claro si su varianza es constante o no. No es fácil de percibir cuando la variabilidad es pequeña. Como no se aprecian grandes diferencias en la varianza a través del tiempo se podría pensar en un modelo aditivo en principio.

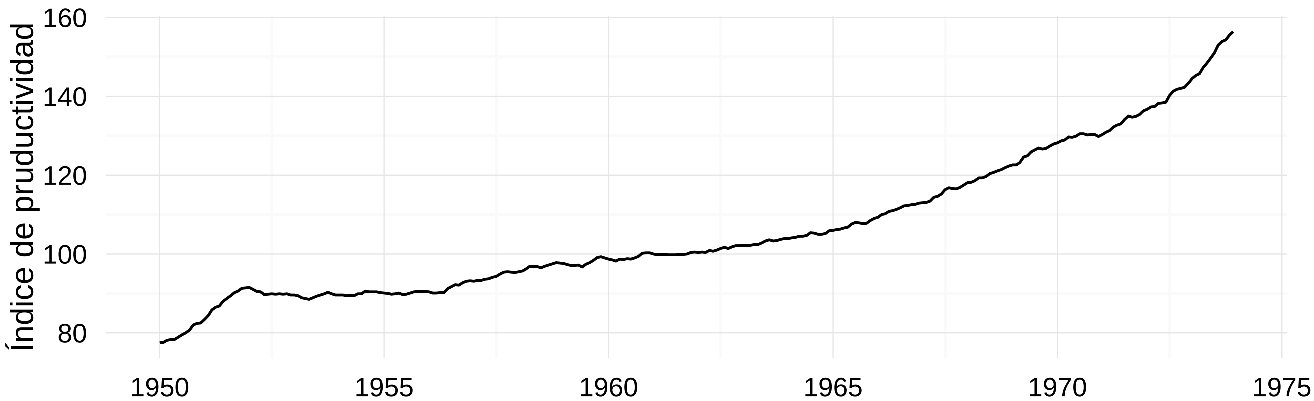

Figura 1: Serie de tiempo del índice de productividad de Canadá desde enero de 1950 hasta diciembre de 1973. Son 288 observaciones en 24 años.

Los primeros dos años de la serie (1950 y 1951) parecen tener un comportamiento diferente al resto, con un crecimiento considerablemente mayor comparado con los años siguientes, por lo que se podría pensar en un posible cambio estuctural de la serie a partir de 1952. Quizá una recesión económica o el fin de la participación de Candá en la guerra de Corea tengan que ver con este cambio. [1] También es posible que se deba a un cambio en la medición del ídice.

[1] Historica Canada. 2015. [En línea]. Tomado de http://www.thecanadianencyclopedia.ca/en/article/kore war/

La serie sí parece tener ciclos, aunque no muy perceptibles, por ejemplo entre 1950 y 1956, o entre 1956 y 1961. La componente del error (E_t) de la Figura 2 muestra algunos tramos que duran más de una año y que podrían ser de crecimiento o de recesión.

No parece haber indicios claros de ciclos, la componente del error (E_t) de la Figura 2 no tiene patrones observables mayores a un año que nos haga pensar en la presencia de ciclos.

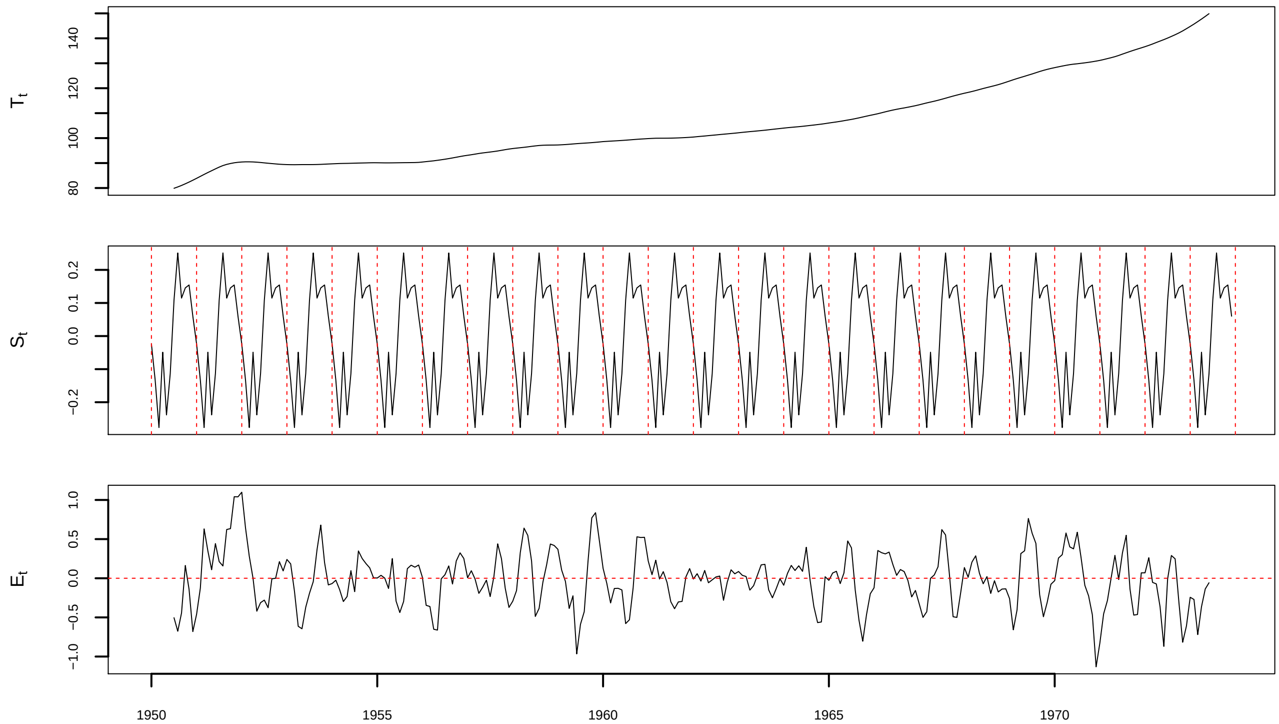

Figura 2: Descomposición de la serie de tiempo en sus componentes; tendencia (T_t), estacional (S_t) y error (E_t)

Aunque a simple vista no se aprecia un patrón estacional en el índice de productividad, la descomposición de la serie sí muestra claramente que existe un comportamiento repetitivo cada año (ver Figura 2). Debido a que la la variabilidad de la series es pequeña, la estacionalidad no fácil de apreciar a simple vista en la Figura 1. Incluso en el box plot que compara el comportamiento de los meses (ver Figura 3) es muy difícil ver la estacionalidad. En este taller vamos a omitir la componente estacional por que no es de interés por ahora.

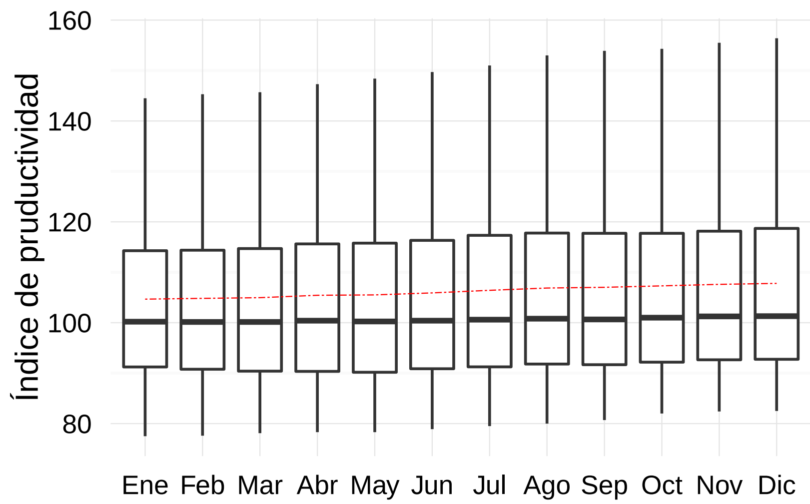

Figura 3: Box-plot de la serie por mes. La línea punteada roja corresponde a los promedios.

Punto 2

De acuerdo con lo visto gráficamente, se proponen tres modelos aditivos, el primero de tendencia cuadrática, el segundo de tendencia cúbica y un último modelo no lineal de tendencia exponencial cúbica.

Modelo aditivo de tendencia cuadrática

El modelo teórico con tendencia cuadrática está dado por:

$$\text{Índice Producción} = \beta_0 + \beta_1 t + \beta_2 t^2 + E_t \qquad E_t \overset{iid}{\sim} N(0, \sigma^2) \quad (1)$$

En la Tabla 1 se muestran los resultados de la estimación del modelo 1. Los parámetros del modelos son significatavos con un nivel de confianza alto, mayor incluso al 99 %. Sin embargo esto no significa que el modelo sea el más adecuado. **Los residuales no parecen cumplir con el supuesto de independencia, precisamente por la incapacidad del modelo para ajustarse a los ciclos, por lo que no tendría mucho sentido realizar la prueba de normalidad.** sin embargo se muestra el gráfico para hacerce una idea de la posible distribución de los residuales. La varianza de los errores sí parece costante ya que no se aprecian dendencias o patrones a través del tiempo.

La ecuación ajustada es:

$$\text{Índice Producción} = 87.7218 - 0.0278t + 0.0008t^2 \quad (2)$$

	Estimado	Error Estd.	Valor t	Pr(>|t|)	
β_0	87.7218	0.5160	170.000	$< 2E^{-16}$	***
β_1	-0.0278	0.0082	-3.368	0.0009	***
β_2	0.0008	$2.7630E^{-5}$	29.285	$< 2E^{-16}$	***
	$\sqrt{MSE} = 2.899$	C^*_{AIC}=2.140	C^*_{BIC}=2.157		

Tabla 1: Ajuste modelo tendencia cuadrática. C^*_{AIC} y C^*_{AIC} son los criterios de información por mínimos cuadrados equivalentes a los criterios AIC y BIC respectivamente.

En ambos gráficos de residuales, vs ajustados y el tiempo, de la Figura 4 vemos que los errores del modelo no parecen seguir un patrón aleatorio. En gran parte debido a la carencia de ajuste, ya que el modelo no logra capturar los posibles ciclos presentes en la serie. Ademá se ve que no es capaz de ajustarse adecuadamente a los primeros años de la serie, debido al comportamiento un poco diferente al resto de los años. Esto se ve reflejado en los grandes valores de los residuales en estos primeros periodos.

Modelo aditivo de tendencia cúbica

El modelo con tendencia cúbica está dado por:

$$\text{Índice Producción} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + E_t \qquad E_t \overset{iid}{\sim} N(0, \sigma^2) \quad (3)$$

En la Tabla 2 se muestran los resultados de la estimación del modelo 3. Este modelo parece ajustar un poco mejor que el de tendencia cuadrática ya que el el término cúbico le da más flexibilidad para adaptarse a las curvas (ciclos). Todos los parámetros del modelo son significativos y las medidad de comparación de modelos nos confirman que este modelo es mejor, tanto la raiz del error cuadrático medio, como el AIC y el BIC son menores que el modelo anterior. Ver Tabla 2.

	Estimado	Error Estd.	Valor t	Pr(>|t|)	
β_0	81.3298	0.3964	205.18	$< 2E^{-16}$	***
β_1	0.2354	0.0119	19.85	$< 2E^{-16}$	***
β_2	-0.0015	0.0001	-15.36	$< 2E^{-16}$	***
β_3	$5.242E^{-6}$	$2.167E^{-7}$	24.19	$< 2E^{-16}$	***
	$\sqrt{MSE} = 1.660$	C^*_{AIC}=1.027	C^*_{BIC}=1.039		

Tabla 2: Ajuste modelo tendencia cúbica

La ecuación ajustada es:

$$\text{Índice Producción} = 81.3298 + 0.2354t - 0.0015t^2 + 5.242E^{-6}t^3 \quad (4)$$

A pesar del mejor ajuste, los residuales siguen mostrando la debilidad del modelo para capturar los cilcos de la serie (ver Figura 5, pues estos no muestran un comportamiento eleatorio al rededor de cero. En este caso tampoco tendría mucho sentido interpretar las pruebas de normalidad, pero se muestran como ejercicio de verificación.

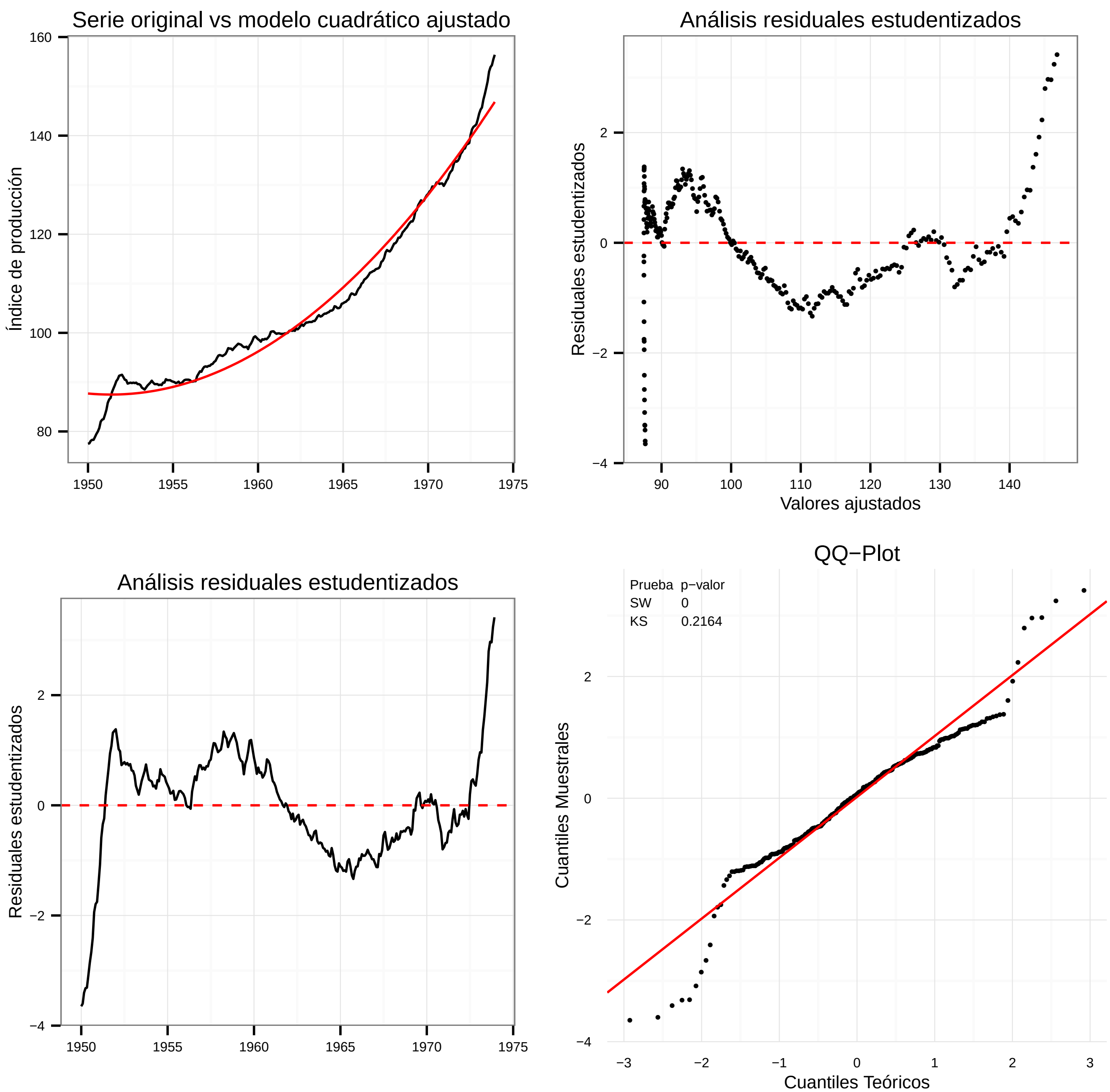

Figura 4: Gráficos de diagnóstico modelo tendencia cuadrática

Modelo aditivo exponencial cúbico

El modelo con tendencia exponencial cúbica está dado por:

$$\text{Índice Producción} = e^{\beta_0+\beta_1 t+\beta_2 t^2+\beta_3 t^3+E_t} \qquad E_t \overset{iid}{\sim} N(0,\sigma^2) \qquad (5)$$

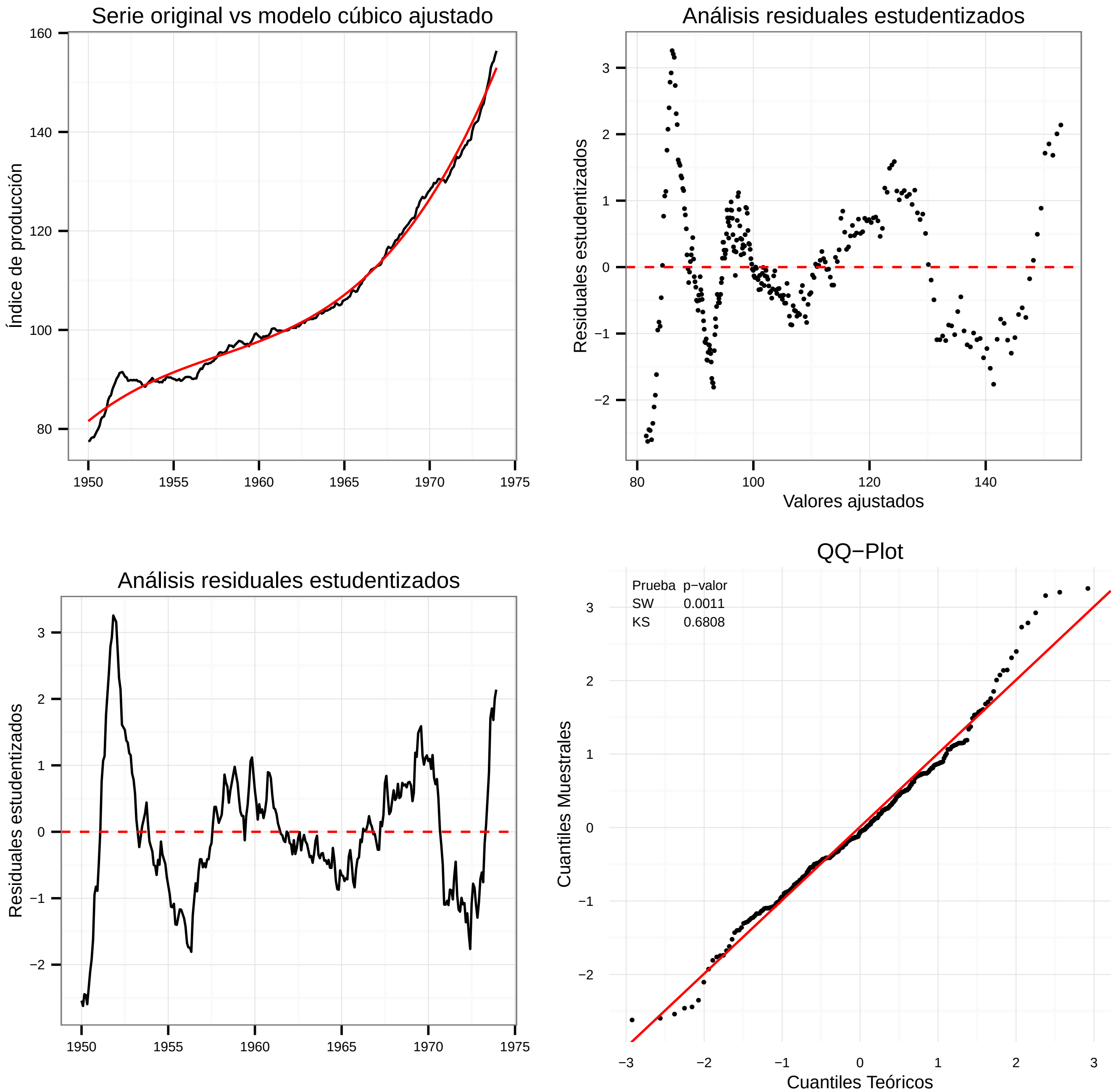

Figura 5: Gráficos de diagnóstico modelo tendencia cúbica

En la Tabla 3 se muestran los resultados de la estimación del modelo 5. El modelo de tendencia cúbica exponencial muestra un comportamiento muy parecido al modelo cúbico, con mejor ajuste en llas curvas de la serie. Sin embargo, aqunque todos los parámetros son significativos con un nivel de confianza del 99 %, todos los criterios de comparación de modelos, $\sqrt{MSE}$, $C^*_($$AIC)$ y C^*_{BIC} muestran que es mejor el modelo cúbico (3) para representar los datos.

	Estimado	Error Estd.	Valor t	Pr(>|t|)	
β_0	4.411	$4.655E^{-3}$	947.60	$< 2E^{-16}$	***
β_1	0.0021	$1.280E^{-4}$	16.54	$< 2E^{-16}$	***
β_2	$-1.003E^{-5}$	$9.673E^{-7}$	-10.37	$< 2E^{-16}$	***
β_3	$3.553E^{-8}$	$2.099E^{-9}$	16.93	$< 2E^{-16}$	***
	$\sqrt{MSE} = 1.718$	C^*_{AIC}=1.096	C^*_{BIC}=1.108		

Tabla 3: Ajuste modelo tendencia cúbica exponencial

La ecuación ajustada es:

$$\text{Índice Producción} = e^{4.411+0.0021t-1.003E^{-5}t^2+3.553E^{-8}t^3} \tag{6}$$

Comparación de los tres modelos

De acuerdo con los criterios de selección de modelos mostrados en la Tabla 4, el de tendencia cúbica tiene todos los valores menosres que los otros dos modelos de tendencia cuadrática y cúbica exponencial.

Modelo	$\sqrt{MSE}$	C^*_{AIC}	C^*_{BIC}
Tendencia cuadrática (1)	2.899	2.139	2.157
Tendencia cúbica (3)	1.660	1.027	1.039
Tendencia cúbica exponencial (5)	1.718	1.096	1.108

Tabla 4: Comparación de los modelos. Entre menor el criterio, mejor el modelo.

Punto 3

Ahora sin tener en cuenta los tres primeros años de la serie se ajustan los modelos nuevamente.

Modelo aditivo de tendencia cuadrática

El modelo teórico con tendencia cuadrática está dado por:

$$\text{Índice Producción} = \beta_0 + \beta_1 t + \beta_2 t^2 + E_t \qquad E_t \overset{iid}{\sim} N(0, \sigma^2) \tag{7}$$

En la Tabla 5 se muestran los resultados de la estimación del modelo 7. Los parámetros del modelo son ignificativos todas con un nivel de confianza del 99 %.

La ecuación ajustada es:

$$\text{Índice Producción} = 100.3677 + 0.1025t + 0.0018t^2 \tag{8}$$

Si bien en este caso el comportamiento de los residuales mejoró un poco, en el sentido de que en los primeros años ya no se ven los error tan pronunciados, el modelo sigue incapaz de capturar lo ciclos.

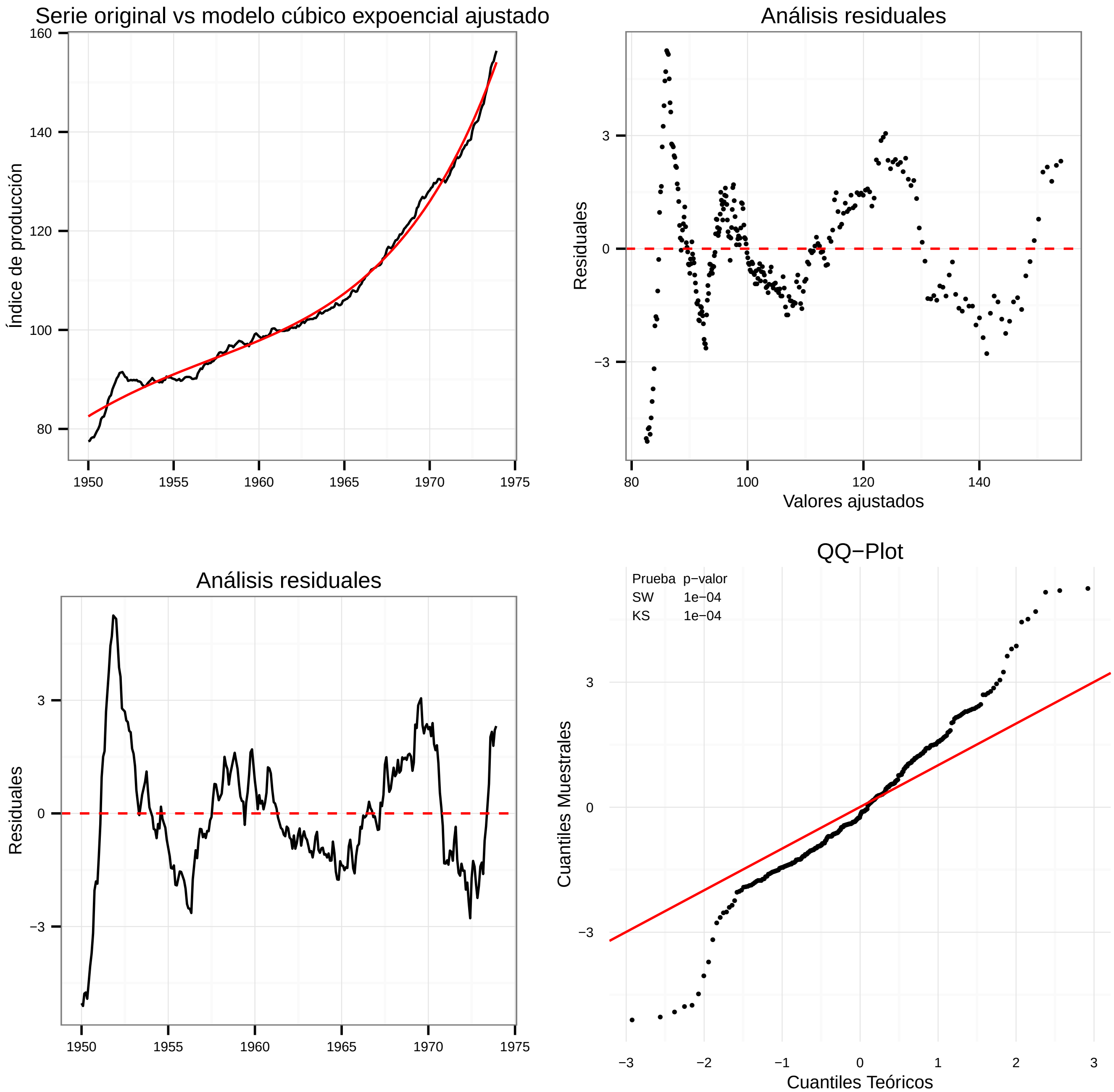

Figura 6: Gráficos de diagnóstico modelo tendencia cúbica exponencial

Modelo aditivo de tendencia cúbica

El modelo con tendencia cúbica está dado por:

$$\text{Índice Producción} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + E_t \qquad E_t \overset{iid}{\sim} N(0, \sigma^2) \quad (9)$$

	Estimado	Error Estd.	Valor t	Pr(>|t|)	
β_0	100.3677	0.3334	301.05	$< 2E^{-16}$	***
β_1	0.1025	0.0106	9.65	$< 2E^{-16}$	***
β_2	0.0018	0.0001	24.97	$< 2E^{-16}$	***
	$\sqrt{MSE} = 2.899$	C^*_{AIC}=1.404	C^*_{BIC}=1.424		

Tabla 5: Ajuste modelo tendencia cuadrática sin tres primeros años

En la Tabla 6 se muestran los resultados de la estimación del modelo 9. En este modelo todos los parámetros son significativos con un nivel de confianza del 99 %. Además mejora considerablemente los criterios de comparación, por ejemplo el C^*_{BIC} pasa de 1.424 a 0.447, casi un tercio del anterior.

	Estimado	Error Estd.	Valor t	Pr(>|t|)	
β_0	87.3384	0.3154	276.945	$< 2E^{-16}$	***
β_1	0.1608	0.0108	14.927	$< 2E^{-16}$	***
β_2	-0.0009	0.0001	-9.308	$< 2E^{-16}$	***
β_3	$5.207E^{-6}$	$2.569E^{-7}$	20.274	$< 2E^{-16}$	***
	$\sqrt{MSE} = 1.233$	C^*_{AIC}=0.435	C^*_{BIC}=0.447		

Tabla 6: Ajuste modelo tendencia cúbica sin tres primeros años

La ecuación ajustada es:

$$\text{Índice Producción} = 87.3384 + 0.1608t - 0.0009t^2 + 5.207E^{-6}t^3 \quad (10)$$

Los residuales siguen mostrando el patrón de correlación debido a los ciclos (ver Figura 8), por lo que no es apropiado realizar los test de normalidad. Sin embargo, suponiendo que cumpliera, los residuales sí parecen tener una distribución noremal, el gráfico y los test de Shapiro-Wilk (p-valor=0.0873) y Kolmogorov-Smirnov (p-valor=0.6639) no muestran evidencia suficiente para rechazar H_0: Los errores se distribuyen normal vs. H_a: Los errores no se distribuyen normal. Sin embargo esta conclusión no es válida por violar el supuesto de incorrelación de los errores.

Modelo aditivo exponencial cúbico

El modelo con tendencia exponencial cúbica está dado por:

$$\text{Índice Producción} = e^{\beta_0+\beta_1 t+\beta_2 t^2+\beta_3 t^3+E_t} \qquad E_t \overset{iid}{\sim} N(0,\sigma^2) \quad (11)$$

En la Tabla 7 se muestran los resultados de la estimación del modelo 11. Este modelo, además de que todos los parámetros son significativos co un nivel de confianza del 99 %, los criterios de información no son mejores que en los dos modelos anteriores, por lo que se podría pensar que este es un mejor modelo que los dos anteriores.

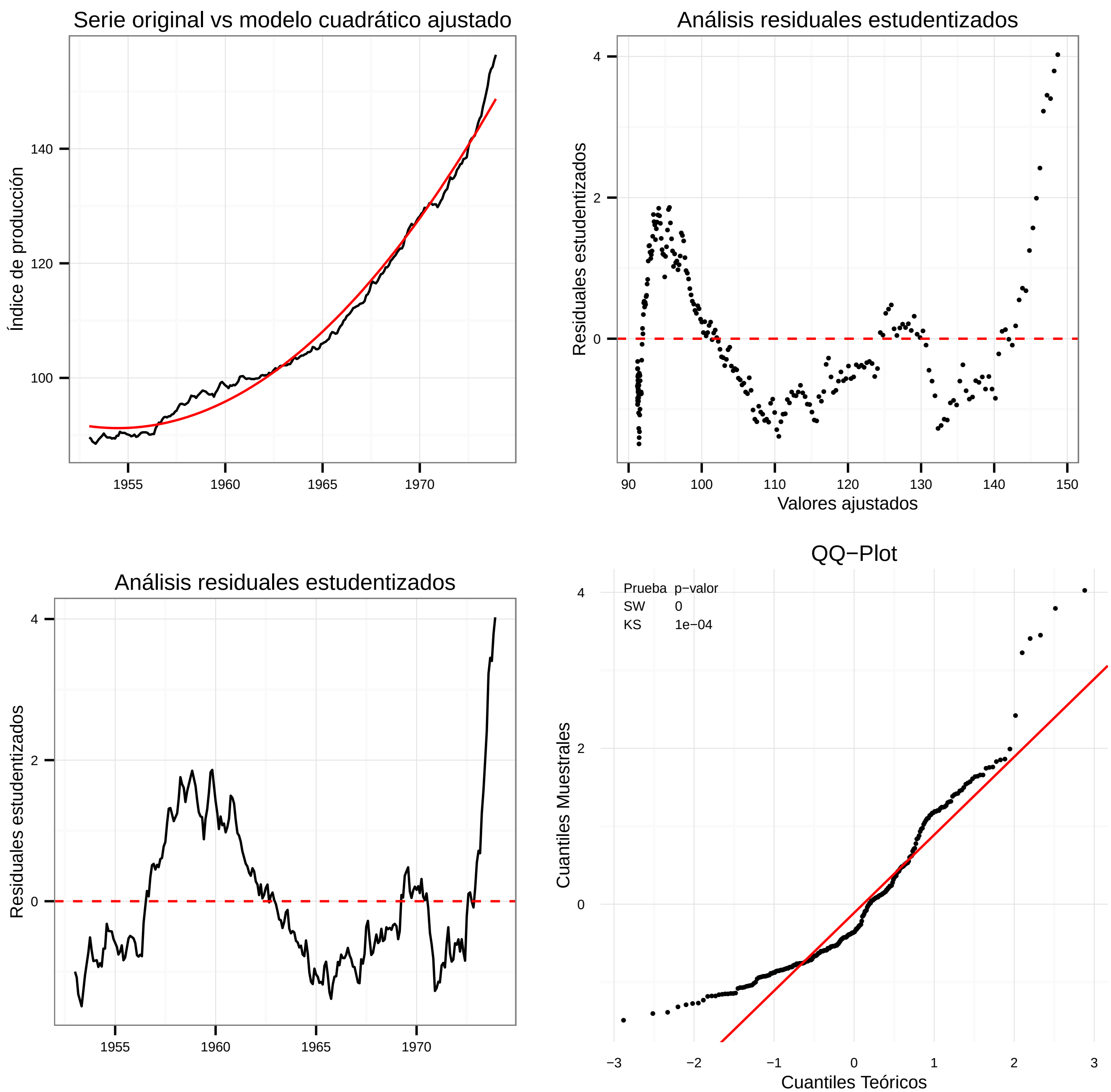

Figura 7: Gráficos de diagnóstico modelo tendencia cuadrática sin tres primeros años

La ecuación ajustada es:

$$\text{Índice Producción} = e^{4.476+0.0014t-4.246E^{-6}t^2+3.027E^{-8}t^3} \qquad (12)$$

Los residuales del modelo siguen sin poder ajustarse a los ciclos de la serie, pues presentan (ver Figura 9) muchas rachas crecientes y

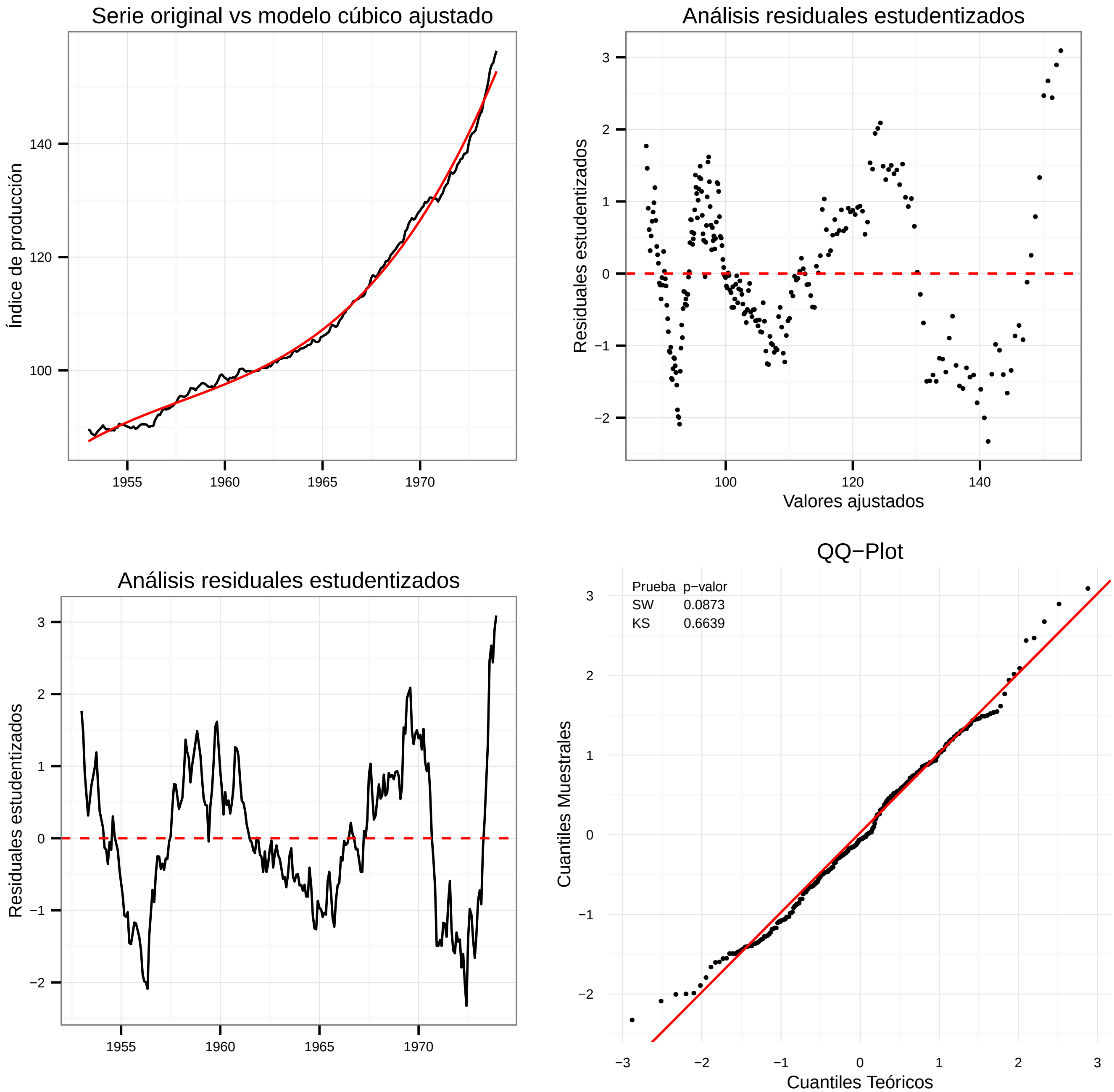

Figura 8: Gráficos de diagnóstico modelo tendencia cúbica sin tres primeros años

decreciantes. Si mejoran un poco respecto al mismo modelo con los tres años iniciales. Tambien vemos que la varianza de los residuales no parece constante, ya que va aumentando de a poco con el tiempo.

	Estimado	Error Estd.	Valor t	Pr(>|t|)	
β_0	4.476	$3.391E^{-3}$	1329.290	$< 2E^{-16}$	***
β_1	0.0014	$1.076E^{-4}$	12.646	$< 2E^{-16}$	***
β_2	$-4.246E^{-6}$	$9.341E^{-7}$	-4.545	$< 2E^{-16}$	***
β_3	$3.027E^{-8}$	$2.325E^{-9}$	13.022	$< 2E^{-16}$	***
	$\sqrt{MSE} = 1.227$	C^*_{AIC}=0.425	C^*_{BIC}=0.438		

Tabla 7: Ajuste modelo tendencia cúbica exponencial sin tres primeros años

Comparación de los tres modelos

De los tres modelos plantedos sin los tres primeros años, el de tendencia cúbica exponencial supera a los otros dos en todos los criterios de información, aunque en muy poco al modelo cúbico. Valdría la pena estudiar si es mejor escoger el modelo cúbico en aras de una mayor simplicidad.

Modelo	MSE	AIC	BIC
Tendencia cuadrática (7)	2.006	1.404	1.424
Tendencia cúbica (9)	1.233	0.435	0.447
Tendencia cúbica exponencial (11)	1.227	0.425	0.438

Tabla 8: Comparación de los modelos sin los tres primeros años. Entre menor el criterio, mejor el modelo.

Punto 4

Los primeros tres años sí estaban teniendo un efecto importante en la estimación de la serie, pues los criteries de información mostraron mejoras considerables al eliminar estas observaciones. Con todas las observaciones el mejor modelo en términos de criterios de información parecía ser el cúbico, pero con la eliminación de los primeros tres años cambió al modelo cúbico exponencial. Sin embargo, suponiendo que el supuesto de incorrelación d elos errores se cumpliera en los modelos, cosa que no es cierta, el modelo cúbico sería mejor ya que es el que mejor se acerca a cumplir con los supuestos. Al parecer el cambio estructural que tenía la serier a partir de esos priemros trs años estaba inflando los errores, pues los modelos eran incapaces de modelar un cambio tan drástico.

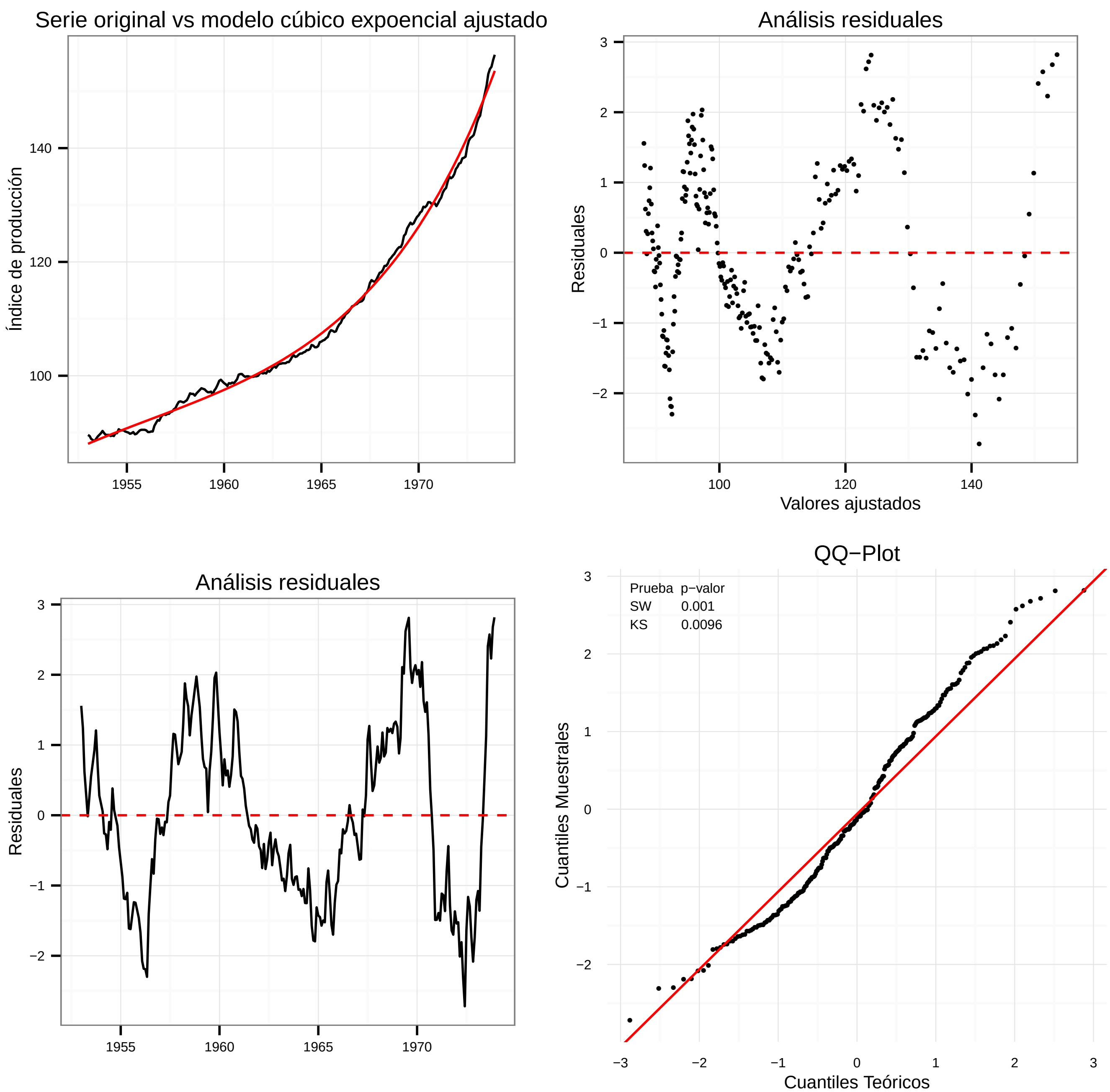

Figura 9: Gráficos de diagnóstico modelo tendencia cúbica exponencial

Código de R utilizado

```
#-----------------------------------------------------------------------------#
# Código R taller 1: Análisis índice de productividad Canadá                  #
#-----------------------------------------------------------------------------#
```

```
# Librerías requeridas
library(ggplot2) # graficos
library(grid) # para poder utilizar la función unit() en ggplot2
library(gridExtra) # funciones auxiliares graficar grid.arrange()
library(dplyr) # manipulación de datos
library(tidyr) # organizar los datos
library(xtable) # exportar tablas en formato LaTeX
theme_set(theme_bw(base_size = 7)) # definir tamaño de fuente gráficos ggplot2

## Funciones auxiliares
ggqqplot <- function(mod, resid.type = 'student'){
  # Devuelve un gráfico qqplot al estilo ggplot2

  if (resid.type == 'student'){
    res_std <- rstudent(mod)
  } else if (resid.type == 'regular'){
    # cuando el modelo ajustado es o lineal (función nls())
    res_std <- resid(mod)
  }

  val_aju <- fitted(mod)

  # Almacenar los datos en un data frame para poder usar ggplot
  dat_tmp <- data.frame(res_std, val_aju)

  qq_val <- qqnorm(res_std, plot.it = FALSE)
  dat_tmp$q_teo <- qq_val$x # añadir cuantiles teóricos al data frame
  dat_tmp$q_muestra <- qq_val$y # añadir cuantiles muestrales al data frame

  slope_qq <- diff(quantile(res_std, c(0.25, 0.75), type = 7)) /
    diff(qnorm(c(0.25, 0.75)))

  inter_qq <- quantile(res_std, c(0.25, 0.75), type = 7)[1L] -
    slope_qq * qnorm(c(0.25, 0.75))[1L]

  # pruebas de normalidad Shapiro-Wilk y Kolmogorov-Smirnov
  sw_vp <- shapiro.test(res_std)$p.value
  ks_vp <- ks.test(res_std, 'pnorm')$p.value

  pruebas_norm <- paste('Prueba  p-valor\n',
                        'SW       ', round(sw_vp, 4), '\n',
                        'KS        ', round(ks_vp, 4), sep = '')

  ub_y <- max(dat_tmp$q_muestra) - 0.2 # ubicación de los p-valores
  ub_x <- min(dat_tmp$q_teo)

  # qqplot con ggplot
  ggplot(data = dat_tmp, aes(x = q_teo, y = q_muestra)) + theme_bw(8) +
    geom_point(size = 1) +
    geom_abline(intercept = inter_qq, y = slope_qq, colour = 'red') +
    labs(x = 'Cuantiles Teóricos', y = 'Cuantiles Muestrales',
```

```
         title = 'QQ-Plot') +
    annotate('text', x = ub_x, y = ub_y, size = 2, parse = FALSE,
             label = pruebas_norm, hjust = 0) +
    theme(plot.margin = unit(c(0.1,0.1,0,0), "lines"),
          panel.border = element_blank(),
          axis.ticks = element_blank())
}

C_p <- function(residuales, type = 'AIC', p = 1){
  # calcula criterio de comparación de modelos C*(p)
  # recibe los residuales en un vector y el tipo de criterio, AIC o BIA
  # p = número de parámetros incluido el intercepto

  n <- length(residuales)

  if (type == 'AIC'){
    c_p <- log((1 / n) * sum(residuales^2)) + 2*p/n
  } else if (type == 'BIC'){
    c_p <- log((1 / n) * sum(residuales^2)) + 2*log(n)/n
  }

  return(c_p)
}

## Lectura de datos
datos_orig <- read.table(file = '../datos/IndProductividad.txt', header = FALSE)
head(datos_orig)

## Organizamos los datos y le damos el formato adecuado
ind_prod <- gather(as.data.frame(t(datos_orig)))[, 2]
fecha <- seq.Date(from = as.Date('1950-01-01', '%Y-%m-%d'),
                  to = as.Date('1973-12-01', '%Y-%m-%d'),
                  by = 'month')
mes <- reorder(months.Date(fecha), rep(1:12, 24))
datos <- data.frame(mes, fecha, ind_prod)
head(datos)
str(datos)

## Análisis exploratorio inicial
summary(datos)

## Punto 1
p1_ind_prod <- ggplot(data = datos) + theme_bw() +
  geom_line(aes(x = fecha, y = ind_prod)) +
  labs(x = '', y = 'Índice de pruductividad') +
  theme(
    axis.line = element_blank(),
    panel.border = element_blank(),
    axis.ticks = element_blank()
  )
p1_ind_prod # ver el gráfico
```

```
# Guardar gráfico
ggsave(filename = '../img/p1_ind_prod.pdf', plot = p1_ind_prod,
       width = 7, height = 2.5)

# Punto 1
datos_ts <- ts(datos$ind_prod, start = c(1950, 1), end = c(1973, 12),
               frequency = 12)
ts_desc <- decompose(datos_ts)
plot(ts_desc)
str(ts_desc)

pdf('../img/p1_descomp.pdf', width = 7, height = 4)
opar <- par()
par(mfrow = c(3, 1),
    oma = c(1, 1, 1, 1),
    mar = c(1, 4, 1, 1),
    xaxt = 'n',
    cex.axis = 0.6,
    cex.lab = 0.9,
    lwd = 0.5)
plot(ts_desc$trend,
     xlab = '',
     ylab = expression(paste(T[t])))
plot(ts_desc$seasonal,
     xlab = '',
     ylab = expression(paste(S[t])))
abline(v = 1950:1974, lty = 2, col =  'red')
par(xaxt = 's')
plot(ts_desc$random,
     xlab = 'Tiempo',
     ylab = expression(paste(E[t])))
abline(h = 0, lty = 2, col =  'red')
par(opar)
dev.off()

prom_mes <- summarise(group_by(datos, mes), prom = mean(ind_prod))

meses <- c('Ene', 'Feb', 'Mar', 'Abr', 'May', 'Jun',
           'Jul', 'Ago', 'Sep', 'Oct', 'Nov', 'Dic')

p1_boxplot_mes <- ggplot(data = datos) + theme_bw() +
  geom_boxplot(aes(x = mes, y = ind_prod)) +
  geom_line(data = prom_mes, aes(x = 1:12, y = prom),
            size = 0.2, colour = 'red', linetype = 6) +
  scale_x_discrete(labels = meses) +
  labs(x = '', y = 'Índice de pruductividad') +
  theme(
    axis.line = element_blank(),
    panel.border = element_blank(),
    axis.ticks = element_blank()
```

```
  )

p1_boxplot_mes
ggsave(filename = '../img/p1_boxplot_mes.pdf', plot = p1_boxplot_mes,
       width = 4.5, height = 3)

## Punto 2

tiempo <- 1:length(datos_ts)
mod1 <- lm(datos_ts ~ tiempo + I(tiempo^2))
summary(mod1)
xtable(summary(mod1),
       caption = 'Ajuste modelo tendencia cuadrática',
       label = 'tab:mod1_comp',
       align = 'lrrrr')

datos$aju_mod1_com <- fitted(mod1)
datos$res_mod1_com <- rstudent(mod1)

p2_mod1_g1 <- ggplot(data = datos) + theme_bw(8) +
  geom_line(aes(x = fecha, y = ind_prod)) +
  geom_line(aes(x = fecha, y = aju_mod1_com), colour = 'red') +
  labs(x = '',
       y = 'Índice de producción',
       title = 'Serie original vs modelo cuadrático ajustado')

p2_mod1_g2 <- ggplot(data = datos) + theme_bw(8) +
  geom_point(aes(x = aju_mod1_com, y = res_mod1_com), size = 1) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = 'Valores ajustados',
       y = 'Residuales estudentizados',
       title = 'Análisis residuales estudentizados')

p2_mod1_g3 <- ggplot(data = datos) + theme_bw(8) +
  geom_line(aes(x = fecha, y = res_mod1_com)) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = '',
       y = 'Residuales estudentizados',
       title = 'Análisis residuales estudentizados')

p2_mod1_g4 <- ggqqplot(mod1)

pdf('../img/p2_diag_mod1_comp.pdf', width = 7, height = 7)
grid.arrange(p2_mod1_g1, p2_mod1_g2, p2_mod1_g3, p2_mod1_g4,
             ncol = 2, nrow = 2)
dev.off()

mod2 <- lm(datos_ts ~ tiempo + I(tiempo^2) + I(tiempo^3))
summary(mod2)
xtable(summary(mod2),
       caption = 'Ajuste modelo tendencia cúbica',
```

```
       label = 'tab:mod2_comp',
       align = 'lrrrr')

datos$aju_mod2_com <- fitted(mod2)
datos$res_mod2_com <- rstudent(mod2)

p2_mod2_g1 <- ggplot(data = datos) + theme_bw(8) +
  geom_line(aes(x = fecha, y = ind_prod)) +
  geom_line(aes(x = fecha, y = aju_mod2_com), colour = 'red') +
  labs(x = '',
       y = 'Índice de producción',
       title = 'Serie original vs modelo cúbico ajustado')

p2_mod2_g2 <- ggplot(data = datos) + theme_bw(8) +
  geom_point(aes(x = aju_mod2_com, y = res_mod2_com), size = 1) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = 'Valores ajustados',
       y = 'Residuales estudentizados',
       title = 'Análisis residuales estudentizados')

p2_mod2_g3 <- ggplot(data = datos) + theme_bw(8) +
  geom_line(aes(x = fecha, y = res_mod2_com)) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = '',
       y = 'Residuales estudentizados',
       title = 'Análisis residuales estudentizados')

p2_mod2_g4 <- ggqqplot(mod2)

pdf('../img/p2_diag_mod2_comp.pdf', width = 7, height = 7)
grid.arrange(p2_mod2_g1, p2_mod2_g2, p2_mod2_g3, p2_mod2_g4,
             ncol = 2, nrow = 2)
dev.off()

mod3_init <- lm(log(datos_ts) ~ tiempo + I(tiempo^2) + I(tiempo^3))
coef_init <- mod3_init$coefficients

mod3 <- nls(datos_ts ~ exp(b0 + b1*tiempo + b2*I(tiempo^2) + b3*I(tiempo^3)),
            start = list(b0 = coef_init[1],
                         b1 = coef_init[2],
                         b2 = coef_init[3],
                         b3 = coef_init[4]))
summary(mod3)

datos$aju_mod3_com <- fitted(mod3)
datos$res_mod3_com <- resid(mod3)

p2_mod3_g1 <- ggplot(data = datos) + theme_bw(8) +
  geom_line(aes(x = fecha, y = ind_prod)) +
  geom_line(aes(x = fecha, y = aju_mod3_com), colour = 'red') +
  labs(x = '',
```

```
       y = 'Índice de producción',
       title = 'Serie original vs modelo cúbico expoencial ajustado')

p2_mod3_g2 <- ggplot(data = datos) + theme_bw(8) +
  geom_point(aes(x = aju_mod3_com, y = res_mod3_com), size = 1) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = 'Valores ajustados',
       y = 'Residuales',
       title = 'Análisis residuales')

p2_mod3_g3 <- ggplot(data = datos) + theme_bw(8) +
  geom_line(aes(x = fecha, y = res_mod3_com)) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = '',
       y = 'Residuales',
       title = 'Análisis residuales')

p2_mod3_g4 <- ggqqplot(mod3, resid.type = 'regular')

pdf('../img/p2_diag_mod3_comp.pdf', width = 7, height = 7)
grid.arrange(p2_mod3_g1, p2_mod3_g2, p2_mod3_g3, p2_mod3_g4,
             ncol = 2, nrow = 2)
dev.off()

# Comparación de los modelos
C_p(residuales = resid(mod1), type = 'AIC', p = 3)
C_p(residuales = resid(mod2), type = 'AIC', p = 4)
C_p(residuales = resid(mod3), type = 'AIC', p = 4)
C_p(residuales = resid(mod1), type = 'BIC')
C_p(residuales = resid(mod2), type = 'BIC')
C_p(residuales = resid(mod3), type = 'BIC')
summary(mod1)
summary(mod2)
summary(mod3)
AIC(mod1)
AIC(mod2)
AIC(mod3)
BIC(mod1)
BIC(mod2)
BIC(mod3)

## Punto 3

tiempo2 <- 1:(length(datos_ts)-3*12)
datos_ts2 <- datos_ts[37:288]
mod12 <- lm(datos_ts2 ~ tiempo2 + I(tiempo2^2))
summary(mod1)
xtable(summary(mod1),
       caption = 'Ajuste modelo tendencia cuadrática sin tres primeros años',
       label = 'tab:mod1_sin3',
       align = 'lrrrr')
```

```
datos2 <- datos[37:288,]
datos2$aju_mod12_com <- fitted(mod12)
datos2$res_mod12_com <- rstudent(mod12)

p3_mod12_g1 <- ggplot(data = datos2) + theme_bw(8) +
  geom_line(aes(x = fecha, y = ind_prod)) +
  geom_line(aes(x = fecha, y = aju_mod12_com), colour = 'red') +
  labs(x = '',
       y = 'Índice de producción',
       title = 'Serie original vs modelo cuadrático ajustado')

p3_mod12_g2 <- ggplot(data = datos2) + theme_bw(8) +
  geom_point(aes(x = aju_mod12_com, y = res_mod12_com), size = 1) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = 'Valores ajustados',
       y = 'Residuales estudentizados',
       title = 'Análisis residuales estudentizados')

p3_mod12_g3 <- ggplot(data = datos2) + theme_bw(8) +
  geom_line(aes(x = fecha, y = res_mod12_com)) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = '',
       y = 'Residuales estudentizados',
       title = 'Análisis residuales estudentizados')

p3_mod12_g4 <- ggqqplot(mod12)

pdf('../img/p3_diag_mod1_sin3.pdf', width = 7, height = 7)
grid.arrange(p3_mod12_g1, p3_mod12_g2, p3_mod12_g3, p3_mod12_g4,
             ncol = 2, nrow = 2)
dev.off()

mod22 <- lm(datos_ts2 ~ tiempo2 + I(tiempo2^2) + I(tiempo2^3))
summary(mod22)
xtable(summary(mod22),
       caption = 'Ajuste modelo tendencia cúbica sin tres primeros años',
       label = 'tab:mod2_sin3',
       align = 'lrrrr')

datos2$aju_mod22_com <- fitted(mod22)
datos2$res_mod22_com <- rstudent(mod22)

p3_mod22_g1 <- ggplot(data = datos2) + theme_bw(8) +
  geom_line(aes(x = fecha, y = ind_prod)) +
  geom_line(aes(x = fecha, y = aju_mod22_com), colour = 'red') +
  labs(x = '',
       y = 'Índice de producción',
       title = 'Serie original vs modelo cúbico ajustado')

p3_mod22_g2 <- ggplot(data = datos2) + theme_bw(8) +
```

```
  geom_point(aes(x = aju_mod22_com, y = res_mod22_com), size = 1) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = 'Valores ajustados',
       y = 'Residuales estudentizados',
       title = 'Análisis residuales estudentizados')

p3_mod22_g3 <- ggplot(data = datos2) + theme_bw(8) +
  geom_line(aes(x = fecha, y = res_mod22_com)) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = '',
       y = 'Residuales estudentizados',
       title = 'Análisis residuales estudentizados')

p3_mod22_g4 <- ggqqplot(mod22)

pdf('../img/p3_diag_mod2_sin3.pdf', width = 7, height = 7)
grid.arrange(p3_mod22_g1, p3_mod22_g2, p3_mod22_g3, p3_mod22_g4,
             ncol = 2, nrow = 2)
dev.off()

mod32_init <- lm(log(datos_ts2) ~ tiempo2 + I(tiempo2^2) + I(tiempo2^3))
coef_init2 <- mod32_init$coefficients

mod32 <- nls(datos_ts2 ~ exp(b0+b1*tiempo2 + b2*I(tiempo2^2) + b3*I(tiempo2^3)),
            start = list(b0 = coef_init2[1],
                         b1 = coef_init2[2],
                         b2 = coef_init2[3],
                         b3 = coef_init2[4]))
summary(mod32)

datos2$aju_mod32_com <- fitted(mod32)
datos2$res_mod32_com <- resid(mod32)

p3_mod32_g1 <- ggplot(data = datos2) + theme_bw(8) +
  geom_line(aes(x = fecha, y = ind_prod)) +
  geom_line(aes(x = fecha, y = aju_mod32_com), colour = 'red') +
  labs(x = '',
       y = 'Índice de producción',
       title = 'Serie original vs modelo cúbico expoencial ajustado')

p3_mod32_g2 <- ggplot(data = datos2) + theme_bw(8) +
  geom_point(aes(x = aju_mod32_com, y = res_mod32_com), size = 1) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = 'Valores ajustados',
       y = 'Residuales',
       title = 'Análisis residuales')

p3_mod32_g3 <- ggplot(data = datos2) + theme_bw(8) +
  geom_line(aes(x = fecha, y = res_mod32_com)) +
  geom_abline(intercept = 0, slope = 0, colour = 'red', linetype = 2) +
  labs(x = '',
```

```
       y = 'Residuales',
       title = 'Análisis residuales')

p3_mod32_g4 <- ggqqplot(mod32, resid.type = 'regular')

pdf('../img/p3_diag_mod3_sin3.pdf', width = 7, height = 7)
grid.arrange(p3_mod32_g1, p3_mod32_g2, p3_mod32_g3, p3_mod32_g4,
             ncol = 2, nrow = 2)
dev.off()

# Comparación de los modelos
C_p(residuales = resid(mod12), type = 'AIC', p = 3)
C_p(residuales = resid(mod22), type = 'AIC', p = 4)
C_p(residuales = resid(mod32), type = 'AIC', p = 4)
C_p(residuales = resid(mod12), type = 'BIC')
C_p(residuales = resid(mod22), type = 'BIC')
C_p(residuales = resid(mod32), type = 'BIC')
summary(mod12)
summary(mod22)
summary(mod32)
AIC(mod12)
AIC(mod22)
AIC(mod32)
BIC(mod12)
BIC(mod22)
BIC(mod32)
```