Análisis de Series de Tiempo.

M. en C. César Almenara Martínez.

Universidad Nacional Autónoma de México.

10 de Noviembre de 2010.

1. Receta.

2. Análisis.

Para hacer un análisis básico de una serie de tiempo, se deben seguir los siguientes pasos.

- Graficar la serie y examinar las principales características de la misma, poniendo particular interés en:
  1. Tendencias.
  2. Estacionalidad.
  3. Cualquier cambio aparente en la conducta de la serie.
  4. "Outliers".
- Remover los componentes de tendencia y estacionalidad, para obtener una serie de residuales estacionarios.
- Elegir un modelo que se ajuste a los residuales, haciendo uso de las herramientas y técnicas vistas en clase.
- La predicción se hará primero sobre los residuales y posteriormente se invertirán las transformaciones hechas, para obtener una predicción de los datos originales.

Análisis de datos de consumo de energía eléctrica en una cierta población, en un cierto periodo de tiempo. Los datos son mensuales.

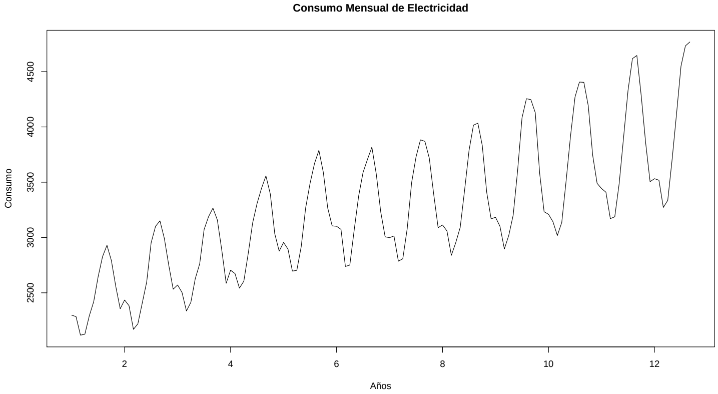


Figura: Datos Originales

El primer paso es hacer un análisis cualitativo, lo más profundo que se pueda para entender la naturaleza y comportamiento de los datos.

```
> summary(elect)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2117    2827    3159    3242    3589    4768
```

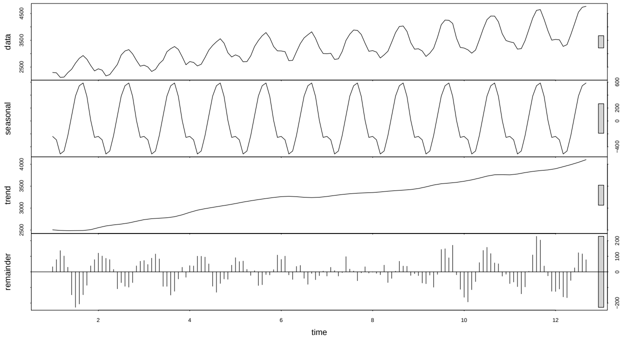


Figura: Descomposición

- Se observa un claro aumento en la variabilidad conforme el tiempo crece.
- Solución: Aplicar la transformación Box-Cox $\frac{x^\lambda - 1}{\lambda}$, para $\lambda > 0$. Cuando $\lambda \to 0$, la transformación se acerca al $\ln(x)$, que es de hecho la transformación más común.

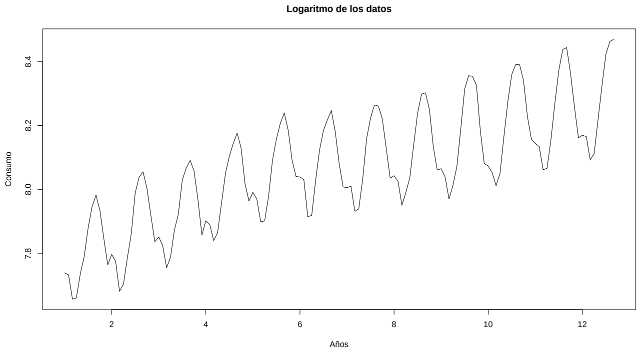


Figura: Eliminación de variabilidad.

- Observamos un periodo igual a 12, por lo que aplicar una diferencia de rezago 12 ($\nabla_{12} X_t = X_t - X_{t-12}$), eliminará una tendencia lineal y la componente estacional (hay que recordar guardar los datos suficientes para poder regresar cada una de las transformaciones aplicadas).

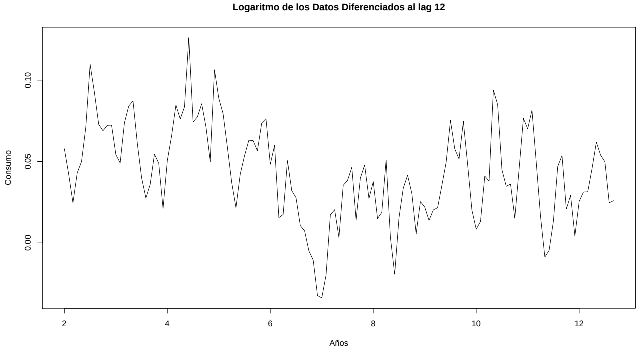


Figura: Eliminación de componentes estacional y de tendencia.

- Chequemos la estacionaridad de la serie resultante.

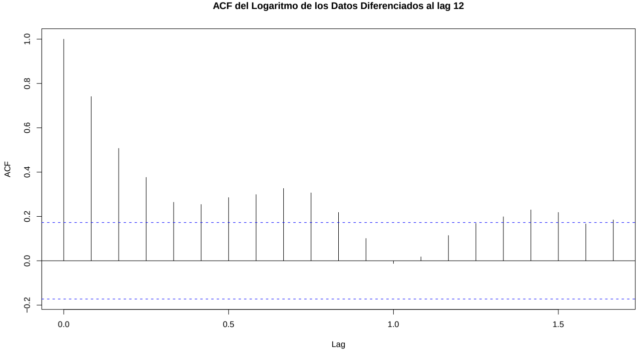


Figura: ACF de la serie resultante.

- Éste análisis gráfico, nos sugiere que la serie resultante aún no es estacionaria.

Aplicamos una diferencia más a los datos pasados, para obtener una serie de residuales estacionaria. ($\nabla X_t = \nabla_1 X_t = X_t - X_{t-1}$).

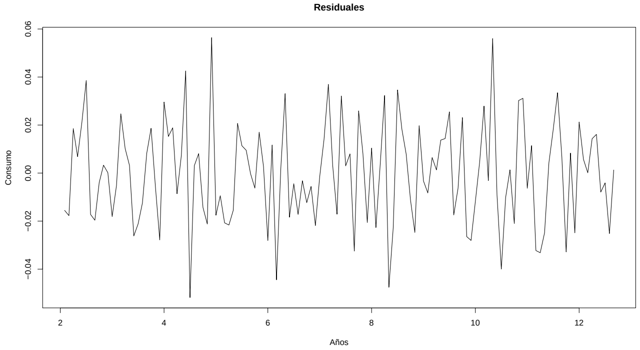
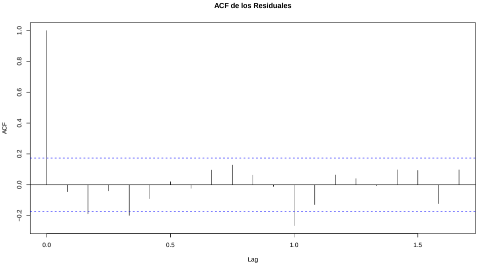
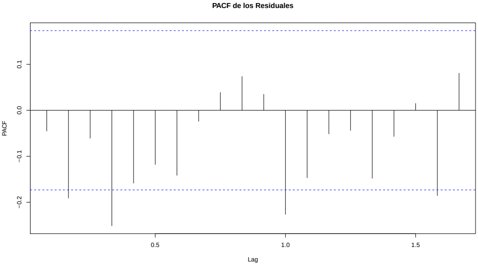


Figura: Residuales estacionarios.

- Inspección de la ACF y PACF de los Residuales





Concluimos que la serie es estacionaria, ya que tanto la ACF como la PACF decaen rapidamente. Adicionalmente podemos aplicar una de las pruebas de aleatoriedad vistas en clase para verificar que existe dependencia entre los datos.

Selección de modelo.

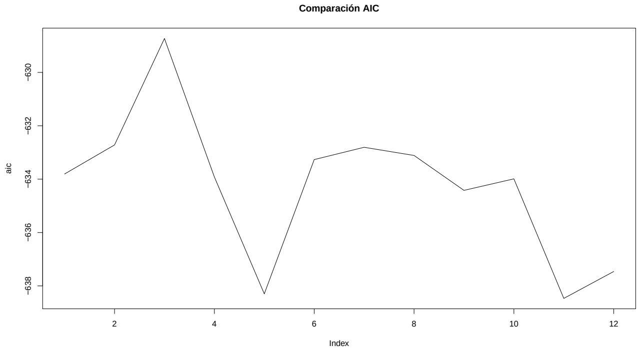
- Cualquier modelo $ARMA(p, q)$ contiene a todos los $ARMA(p', q')$, con $p' \leq p$ y $q' \leq q$.
- Recuerda que al implementar el método de máxima verosimilitud para la estimación de los parámetros del modelo, al aumentarse el número de parámetros la estimación máximo verosimil, NO se ve afectada.
- No es deseable tener más parámetros de los necesarios dentro del modelo.
- Para discriminar entre modelos, podemos hacer un gráfico de varianzas de residuales o hacer una penalización de la *verosimilitud*.

Concluimos que la serie es estacionaria, ya que tanto la ACF como la PACF decaen rapidamente. Adicionalmente podemos aplicar una de las pruebas de aleatoriedad vistas en clase para verificar que existe dependencia entre los datos.

Selección de modelo.

- Cualquier modelo $ARMA(p, q)$ contiene a todos los $ARMA(p', q')$, con $p' \leq p$ y $q' \leq q$.
- Recuerda que al implementar el método de máxima verosimilitud para la estimación de los parámetros del modelo, al aumentarse el número de parámetros la estimación máximo verosimil, NO se ve afectada.
- No es deseable tener más parámetros de los necesarios dentro del modelo.
- Para discriminar entre modelos, podemos hacer un gráfico de varianzas de residuales o hacer una penalización de la *verosimilitud*.

- $AIC = -2 \times \ln(\mathit{verosimilitud}) + 2 \times (p + q + 1)$. Cuando aumenta la verosimilitud, baja $-2 \times \ln(\mathit{verosimilitud})$. Si se agrega un parámetro adicional al modelo, bajará $-2 \times \ln(\mathit{verosimilitud})$, pero se verá compenzado con la penalización $2 \times (p + q + 1)$
- Un método sistemático para elegir el mejor modelo, es estimar modelos $ARMA(p, p)$ con $p = 1, 2, \ldots$ y de éstos, elegir el orden del modelo con el menor AIC.
- En éste caso comparamos los modelos $ARMA(p, p)$, con $p = 1, 2, \ldots, 12$, por ser el 12, el último rezago en el que se identifica alta dependencia entre los datos (según nuestro análisis gráfico de la ACF y PACF).



- Observamos que los modelos con menor AIC son los de orden 5 y 11. Elegimos el modelo $ARMA(5,5)$ por ser el de menor número de parámetros a estimar.

- El resultado de la estimación es:

```
Call:
arima(x = res, order = c(5, 0, 5))

Coefficients:
         ar1     ar2      ar3     ar4     ar5      ma1      ma2     ma3     ma4
      0.1106  0.1005  -0.3235  -0.361  0.4171  -0.2837  -0.4059  0.4059  0.2837
s.e.  0.0920  0.0897   0.0942   0.098  0.0898   0.0606   0.0627  0.0613  0.0575
          ma5  intercept
      -1.0000     -3e-04
s.e.   0.0592      2e-04

sigma^2 estimated as 0.0003024:  log likelihood = 331.15,  aic = -638.3
```

- Describir cuadro.
- Los parámetros que No son significativos, son la "intersección", el $ar1$ y el $ar2$.
- Recordemos que un parámetro es significativo a un nivel del 95 %, si $1.96 \times s.e$, es menor que el valor absoluto del estimador.
- El parámetro más No-significativo es el $ar2$. Lo eliminamos (i.e. $ar2 = 0$) y volvemos a estimar.

- El resultado de la estimación es:

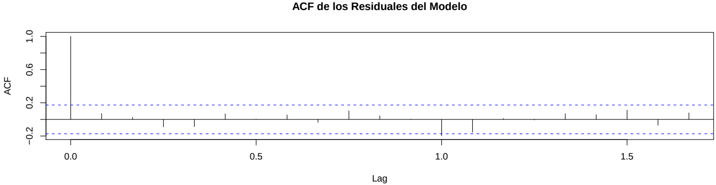
```
Call:
arima(x = res, order = c(5, 0, 5), fixed = c(NA, 0, NA, NA, NA, NA, NA, NA,
    NA, NA, NA))

Coefficients:
          ar1  ar2     ar3      ar4     ar5     ma1      ma2      ma3     ma4
      -0.6707    0  0.1719  -0.2096  0.0384  0.5870  -0.4106  -0.5664  0.0079
s.e.   0.3918    0  0.3392   0.2292  0.2396  0.3958   0.1024   0.4779  0.2388
          ma5  intercept
      -0.2822     -2e-04
s.e.   0.3265      4e-04

sigma^2 estimated as 0.0003352:  log likelihood = 327.67,  aic = -633.35
```

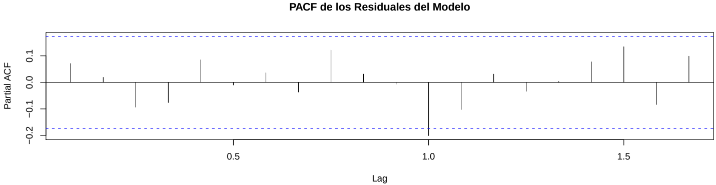
- Explicar cuadro.
- Eliminamos el parámetro *ar*1 y volvemos a estimar. Obteniendo:

```
Call:
arima(x = res, order = c(5, 0, 5), fixed = c(0, 0, NA, NA, NA, NA, NA, NA, NA,
    NA, NA))

Coefficients:
      ar1  ar2      ar3      ar4     ar5      ma1      ma2     ma3     ma4
        0    0  -0.3438  -0.4180  0.3205  -0.1992  -0.3405  0.4246  0.3381
s.e.    0    0   0.0870   0.0992  0.0931   0.0631   0.0558  0.0581  0.0682
          ma5  intercept
      -0.9193     -2e-04
s.e.   0.0689      4e-04

sigma^2 estimated as 0.0003109:  log likelihood = 330.51,  aic = -641.01
```

- Explicar cuadro.

- En éste punto encontramos que todos los parámetros son significativos y el AIC encontrado es el más bajo hasta el momento.
- Nuestro siguiente paso es checar los residuales del modelo vía la ACF y PACF.





- Notamos que en el rezago 12, tenemos una pequeña salida. Ésto nos sugiere incluir al modelo los parámetros *ar*12 y *ma*12. Con lo que obtenemos la sigueinte estimación:

```
Call:
arima(x = res, order = c(12, 0, 12), fixed = c(0, 0, NA, NA, NA, 0, 0, 0, 0,
    0, 0, NA, NA, NA, NA, NA, NA, 0, 0, 0, 0, 0, 0, NA, NA))

Coefficients:
      ar1  ar2    ar3      ar4      ar5  ar6  ar7  ar8  ar9  ar10  ar11    ar12
        0    0  0.246  -0.3652  -0.1034    0    0    0    0     0     0  0.1292
s.e.    0    0  0.158   0.1808   0.1247    0    0    0    0     0     0  0.1704
          ma1      ma2      ma3    ma4      ma5  ma6  ma7  ma8  ma9  ma10  ma11
      -0.1148  -0.1756  -0.3290  0.145  -0.0067    0    0    0    0     0     0
s.e.   0.1097   0.1097   0.1706  0.160   0.1697    0    0    0    0     0     0
         ma12  intercept
      -0.6238     -3e-04
s.e.   0.1935      3e-04

sigma^2 estimated as 0.0002971:  log likelihood = 333.41,  aic = -642.81
```

- Continuamos excluyendo parámetros de acuerdo a su nivel de significancia.

- Después de eliminar todos los parámetros No-significativos, nos quedamos con el modelo

```
Call:
arima(x = res, order = c(12, 0, 12), fixed = c(0, 0, 0, NA, 0, 0, 0, 0, 0, 0,
    0, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA, NA))

Coefficients:
      ar1  ar2  ar3      ar4  ar5  ar6  ar7  ar8  ar9  ar10  ar11  ar12
        0    0    0  -0.2104    0    0    0    0    0     0     0     0
s.e.    0    0    0   0.0876    0    0    0    0    0     0     0     0
          ma1      ma2  ma3  ma4  ma5  ma6  ma7  ma8  ma9  ma10  ma11     ma12
      -0.2601  -0.2601    0    0    0    0    0    0    0     0     0  -0.4016
s.e.   0.0838   0.0884    0    0    0    0    0    0    0     0     0   0.0749
      intercept
         -2e-04
s.e.      2e-04

sigma^2 estimated as 0.0003278:  log likelihood = 329.99,  aic = -647.98
```

- Explicar cuadro, por que el modelo es adecuado y escribirlo de manera análitica (teniendo cuidado con la definición de R para la intercepción).
- Análisis la ACF y PACF del modelo.

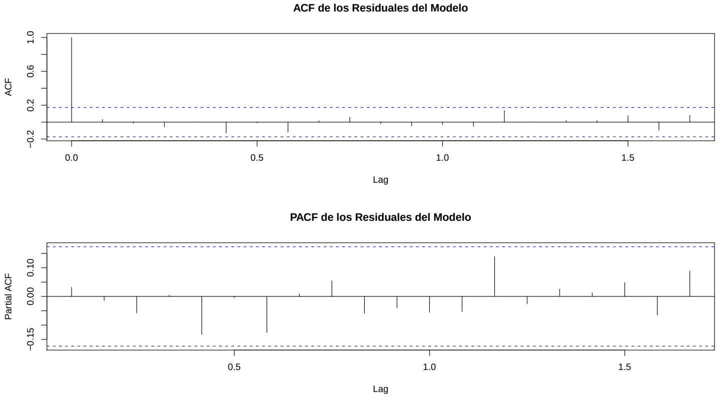


Figura: Residuales del modelo estimado.

El paso inmediato es realiazar el análisis de residuales, de acuerdo a los métodos vistos en clase.

Usamos éste modelo para realizar una predicción de la serie (inversión de transformaciones hechas).

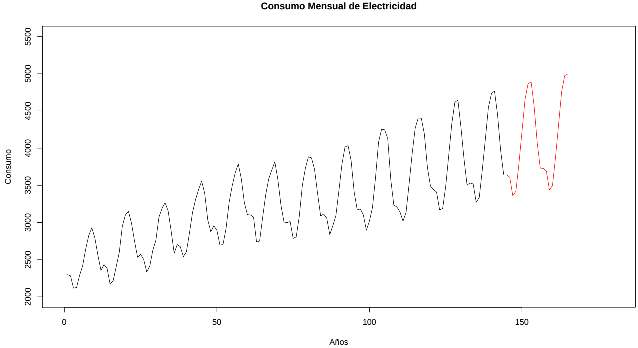


Figura: Residuales del modelo estimado.

Explicar los datos obtenidos en la predicción.