Índice de Precios Industriales
Bienes Intermedios

Diego Del Río Rodríguez
3.01 GIT

CEU | Universidad San Pablo

En este informe se analizará el componente del índice de precios industriales correspondiente con los precios de los bienes intermedios para tratar de realizar predicciones de dicho índice. Los bienes intermedios son aquellos que bienes que son agotados en medio de su proceso productivo, teniendo la posibilidad de ser o no adquiridos por algún otro agente económico que le de uso como mercancía de consumo o mercancía de inversión, por ejemplo, la harina, pan, metales…

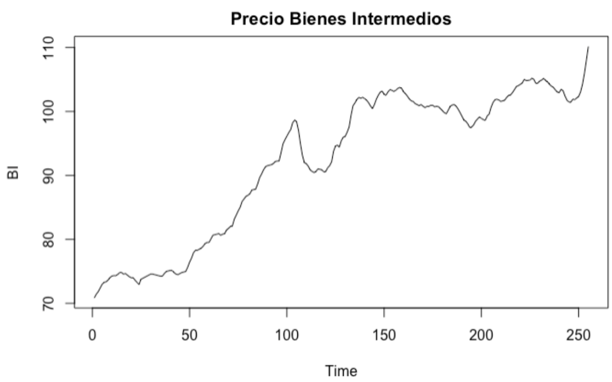
Los datos han sido extraídos de la base de datos del INE y se corresponde con una serie temporal mensual de dichos precios desde el año 2000 hasta marzo de 2021.

La base de datos consta de 256 observaciones, correspondientes a 21 años, desde el año 2000 hasta marzo de 2021, y de una única variable de tipo numérico que se corresponde con el valor que toma los bienes de equipo en dicho mes.

En un primer lugar lo que nos interesa es comprobar la estacionalidad de la serie que vamos a utilizar y si esta es estacionaria o no en media y varianza. En este caso observamos un claro ascenso de los datos durante todo el periodo estudiado excepto en algún momento puntual que estos bajan debido a factores externos, lo que nos indica que la serie no es estacionaria ni en media ni en varianza. Su tendencia se podría decir que es errática, es decir, no parece muy asentada alrededor de la media y por lo tanto no parece ser un buen resumen de la serie. Esto nos indica que deberemos introducir al menos una diferencia en la serie. Por lo que tenemos que extraer el componente estacional de la serie porque sino, la serie no podría ser estacionaria y lo que nos interesa es que lo sea para la modelización.

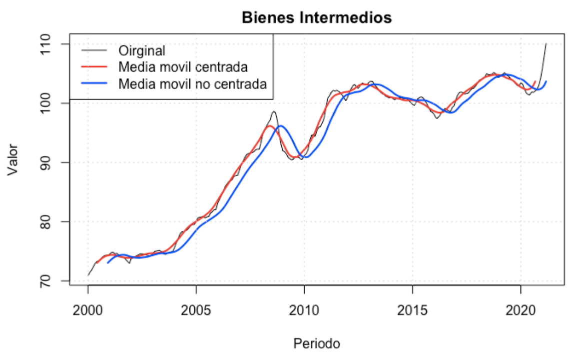
Observamos que la serie presenta un comportamiento no estacional mensual. *(Véase Anexo I)*

Anexo I: Serie Precio de los Bienes Intermedios



Tras esto pensé en realizar un grafico similar al hecho arriba solo dibujando tambien su media móvil tanto centrada como no centrada. Hice esto ya que este método puede ser útil si queremos calcular la tendencia de una serie temporal sin tener que ajustarnos a una función previa, ofreciendo así una visión suavizada o alisada de la serie, ya que promediando varios valores se elimina parte de los movimientos irregulares de la serie; también puede servirnos para realizar predicciones cuando la tendencia de la serie tiene una media constante. *(Véase Anexo II)*

Anexo II: Serie precio de los bienes intermedios con media móvil.

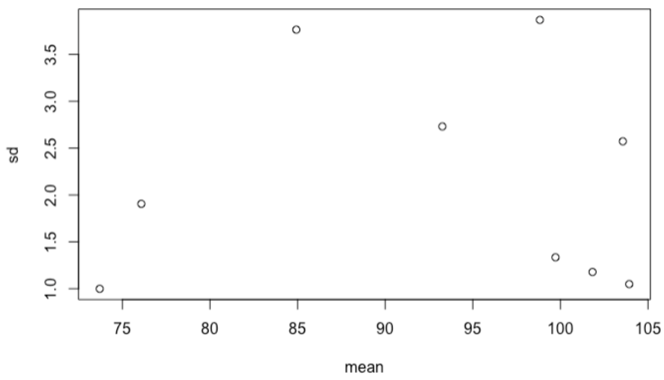


Como queremos obtener unas predicciones que sean coherentes hemos decidido no prescindir de ningún dato de la serie temporal, ya que conocer como afecto el COVID a la evolución de los bienes intermedios es importante tenerlo en cuenta ya que sino estaríamos realizando predicciones con valores pre-covid lo cual seria algo incoherente.

Para analizar el comportamiento de la varianza lo que haremos será un grafico media ~ desviación típica. De este modo sabremos si podremos linealizar la serie introduciendo logaritmos.

(Véase Anexo III)

Anexo III: Grafico media – desviación típica

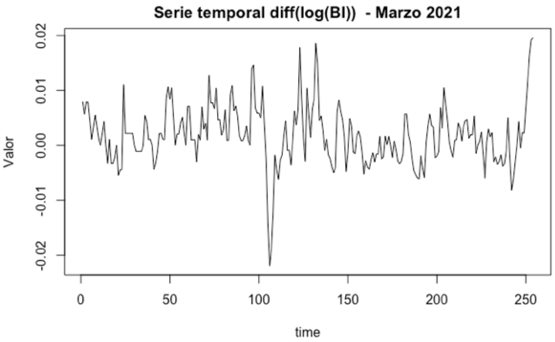


Con este grafico podemos observar que no parece clara la necesidad de logaritmo ya que no hay una relación directa entre la media y la desviación típica (localmente) en la serie.

A continuación, lo que hicimos fue tratar de solucionar la no estacionariedad en varianza mediante la aplicación de un logaritmo a la variable. Una vez realizado esto corregiremos los datos con una primera diferencia para hacer que nuestra serie sea estacionaria en media y de este modo que presente un comportamiento mas estable.

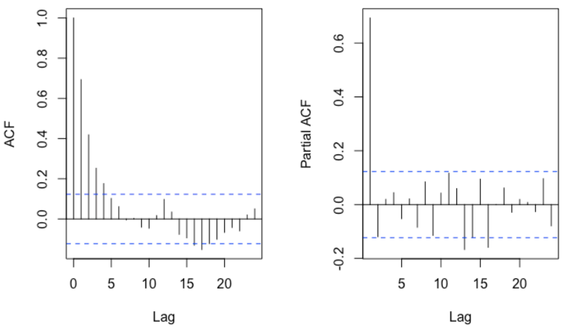
(Véase Anexo IV)

Anexo IV: Nueva serie estacionaria en media.



Como podemos observar en el grafico de encima la tendencia creciente de la serie a desaparecido, pero permanece el comportamiento estacional. Una vez tenemos los datos corregidos en media y varianza, creamos un modelo con los datos corregidos y la variable dummy, posteriormente observamos los correlogramas resultantes. En los correlogramas se pueden observar, además, la parte regular y la parte estacional. *(Véase Anexo V)*

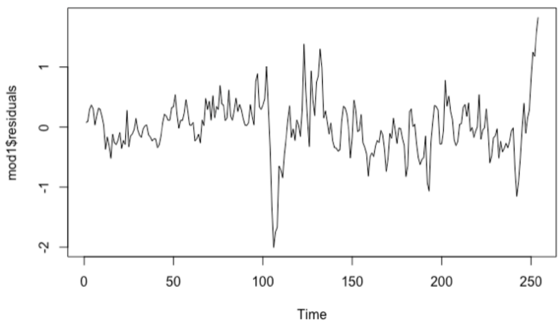
Anexo V: Correlogramas Acf y Pacf.



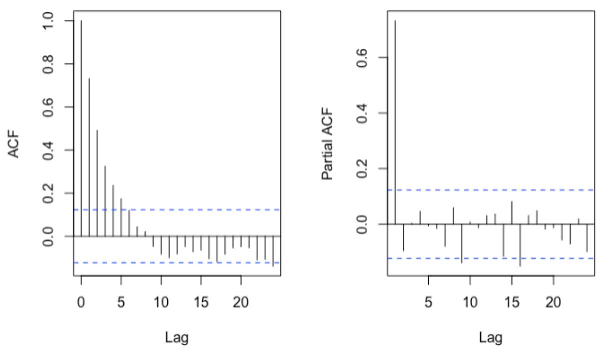
Fijándonos en el componente estacional de nuestro correlograma observamos que persiste estacionalidad en la serie ya que aumenta de un “lag” a otro y representan valores significativos. En este correlograma podemos observar un decaimiento exponencial cada 12 meses, lo que quiere decir que en la parte estacional encontramos un autorregresivo de orden 1, dado que en el PACF solo vemos un palo significativo. Por otro lado, en la parte regular observamos varios palos en la ACF y un decaimiento exponencial en los primeros meses, por lo que nos encontramos ante una media móvil de posible orden 1,2,3 e incluso 4.

Para eliminar la pauta estacional vamos a introducir variables de estacionalidad determinista para quedarnos solo con la parte libre de estacionalidad (Dummies). *(Véase Anexo VII)*

Anexo VI: Residuos de la serie.

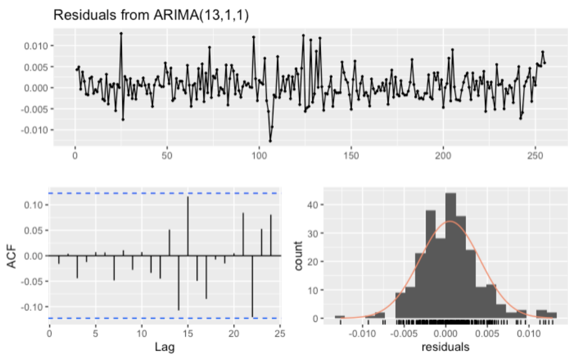


Anexo VII: Correlograma de los residuos.



Tras esto pasamos a modelar todo en su conjunto. Para esto tuvimos en cuenta la totalidad de los datos disponibles y la posible presencia de una media móvil en la estacionalidad. Para decidir cual escoger nos fijaremos en el ACF de sus residuos, escogeremos el que menores correlaciones presente, es decir lo que trataremos de buscar es que se de un ruido blanco. Realizamos un total de 13 modelos con presencia de media móvil pero no de manera estacional sino en la parte regular de los modelos hasta obtener un ruido blanco en el ACF. Que obtuviésemos un ruido blanco en un Arima(13,1,1) nos pareció una barbaridad ya que eso indicaba que teníamos que mirar 13 momentos hacia el pasado para poder predecir el momento futuro. *(Véase Anexo VIII)*

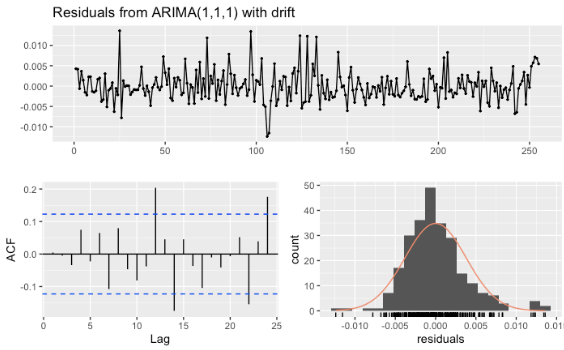
Anexo VIII: Modelo ARMA(13,1,1)



Por lo que decidimos hacer un auto.arima para que nos devolviese el modelo arima que mejor se ajusta a nuestros datos modificados.

Podemos hacer también un auto.arima para que nos devuelva el modelo ARIMA que mejor se ajusta a nuestros datos modificados. Utilizando este método nos salió que el modelo que se ajusta mejor a nuestros datos seria un Arima(1,1,1) por lo que decidimos tener este modelo en cuenta pese a que apareciesen valores significativos en el ACF de los residuos.*(Véase Anexo IX)*

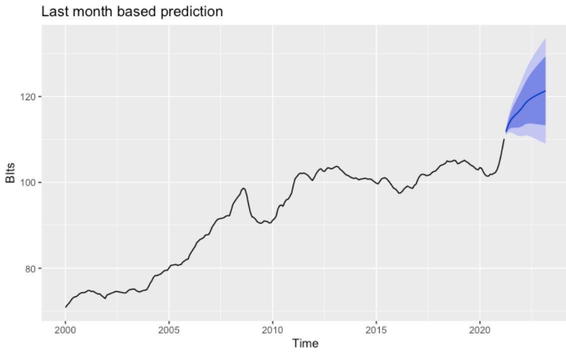
Anexo IX: Modelo ARMA(1,1,1)



En este caso en el grafico ACF podemos observar varios valores significativos por lo que en este caso no nos encontramos ante un ruido blanco. Observamos correlaciones muy elevadas en el mes 12, 14, 22 y 24, esto lo que quiere decir es que el valor de los bienes intermedios de hace 12, 14, 22 y 24 meses influyen en los posibles valores de los bienes intermedios de hoy.

Tras esto nos lanzamos a realizar una serie de predicciones hasta el mes de Junio de 2021. En el siguiente grafico veremos una estela de los posibles valores que pueden tomar los valores de los bienes intermedios en los próximos cuatro meses, podemos ver la predicción de los puntos y luego una serie de valores por encima o por debajo de dicha predicción. *(Véase Anexo X)*

Anexo X: Predicción de los valores futuros.

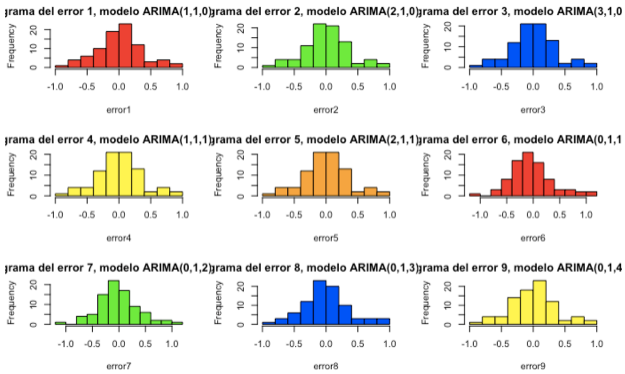


	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Apr 2021	111.7667	111.3186	112.2148	111.0813	112.4521
May 2021	112.9527	112.0132	113.8921	111.5159	114.3894
Jun 2021	113.8522	112.4175	115.2869	111.6580	116.0464

Tras esto decidimos ir almacenando los errores de predicción de los modelos con el fin de obtener información asociada a la habilidad predictiva de cada modelo. Para esto realizamos la validación cruzada.

Una vez realizada la validación cruzada lo que vamos a hacer es plotear una serie de histogramas con la distribución de los errores de cada uno de los modelos. *(Véase Anexo XI)*

Anexo XI: Distribución de los errores.



Observando el Anexo XI vemos como la distribución de todos los errores es muy similar y esta se asemeja mucho a una distribución normal en la que el error este concentrado en torno a 0, es decir en el centro, pese a que haya alguno que otro que tenga una mayor concentración del error en el lado izquierdo/derecho.

En cuanto a los errores de los modelos, estos han sido medidos de tres maneras diferentes durante la validación cruzada, de manera que hemos obtenido estos tres tipos de errores:

- El error de predicción: valor real – valor predicho
- El error porcentual: (valor real – valor predicho) / (valor real)
- El error cuadrático medio: suma(valor real- calor predicho)^2

Una vez extraídos los errores mediante la validación cruzada vemos como tanto el error de predicción, como el error porcentual como el error cuadrático medio presentan valores muy pequeños y próximos a 0. Como todos los errores se asemejan mucho y en general no hay un criterio claro de que nos permita decantarnos por un único modelo nos fijaremos en el modelo que presente el menor error posible, este modelo es el modelo 4 el cual se corresponde con un Arima(1,1,1) y que a su vez se corresponde con el modelo que realizo R a partir del auto.arima que en teoría es el modelo que mejor se ajusta a nuestros datos. Este modelo presenta un error de predicción muy bajo y cercano a 0, tiene un error porcentual también muy bajo y por último es el modelo con el error cuadrático medio mas bajo.

Tras esto lo que realizamos fue un modelo de regresión lineal en el que tratamos de explicar los datos reales a partir de las predicciones realizadas previamente y obtenemos las siguientes ponderaciones que se atribuyen a cada modelo:

- Modelo1 = 0.20
- Modelo2 = 36.74
- Modelo3 = 37.80
- Modelo4 = 23.70
- Modelo5 = -99
- Modelo6 = -0.21
- Modelo7 = -0.91
- Modelo8 = 0.54
- Modelo9 = 2.13

Con esto observamos como el modelo 4 es uno de los modelos que tiene una de las ponderaciones mas altas, eso quiere decir que sus predicciones ponderan mucho.

Y, por ultimo, estas conclusiones junto con las sacadas anteriormente con lo de la validación cruzada, es decir que este modelo presenta errores muy bajos y próximos a 0, hacen del modelo 4 el mejor candidato para realizar predicciones.