# Machine Learning II Efecto del COVID en la demanda energética de Bolivia

# Nicolás Oriol Guerra - Universidad Pontificia Comillas ICAI Diciembre 2020

Este informe es un estudio de la serie histórica de demanda energética de Bolivia desde 2014 hasta el presente. En primer lugar, se estudia el efecto y la relación existente que pueden tener variables externas con el nivel de conumo energético. Se estudian GDP, temperatura, fecha (día de la semana y mes) y día especial. Con estos datos se procede a estudiar el efecto del COVID sobre la demanda de nacional energía. Para responder a esto se ha comparado la demanda hipotética que podría haber habido en 2020 sin efectos de la pandemia con los valores reales observados. Se obtiene así una recomendación de modelo ARIMA para predecir demandas futuras y una estimación del efecto de la pandemia valorada en 2-2.5 millones de Bs.

## 1 Introducción

El objetivo de este estudio es generar un modelo que permita predecir la demanda diaria de energía en Bolivia. Este modelo a corto plazo debe ser capaz de proporcionar una estimación válida del nivel de demanda para horizontes de 1 a 7 días. Como conclusión adicional, se desea también estudiar el impacto del COVID en la demanda nacional de energía observando los efectos que ha tenido y estimando posibles hipótesis a futuro.

Para obtener estas respuestas se han planteado cuatro pasos a seguir: análisis inicial de demanda y estudio de situación actual; planteamiento de modelos ARIMA sin efectos del COVID; planteamiento de modelos ARIMA con efectos del COVID y estimación de las consecuencias de la pandemia.

Con el análisis inicial se pretende estudiar cuándo se hicieron notar los efectos del COVID y si siguen afectando en el presente. Al plantear los modelos ARIMA con y sin influencia de la pandemia se desea estudiar la diferencia y validez de las respectivas predicciones. Finalemente, se podrá comparar la predicción sin efecto del COVID con los datos reales de los meses de pandemia. Esto nos permitirá medir el efecto en la demanda energética. También se compararán las predicciones más recientes tratando de obtener un modelo adecuado.

#### 2 Análisis inicial de los datos

#### 2.1 Estudio de relaciones entre variables

Para hacer el estudio se tiene una serie temporal con datos diarios desde noviembre de 2014 hasta octubre de 2020. Cada observación tiene el valor correspondiente de demanda, temperatura, GDP, día de la semana y especial (indica si es un día festivo o con algún evento extraordiario). Algunos días presentan valores vaciós de temperatura que se completan con los datos del año anterior.

Para comenzar el análisis, se estudia la relación entre los distintos datos diarios y la demanda para ver si hay alguna relación clara que permita explicar la variable objetivo. Además, si se percibe alguna relación no lineal será necesario hacer la transformación correspondiente debido al caracter lineal de los modelos ARIMA. La relación de la demanda con las variables GDP, temperatura, fecha y día especial se representa en los siguientes gráficos.

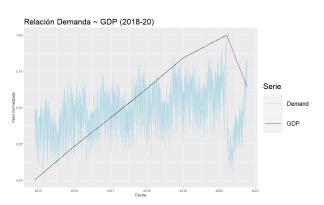


Fig. 1. Series normalizadas de demanda y GDP

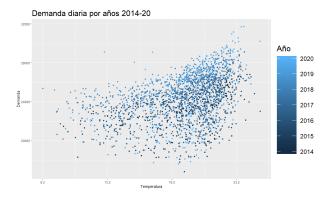


Fig. 2. Relación demanda y temperatura

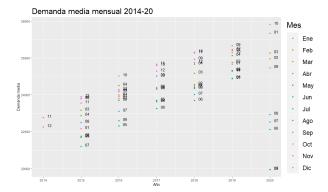


Fig. 3. Demanda según la época del año

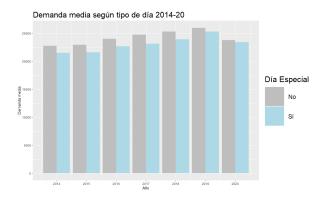


Fig. 4. Efecto del tipo de día sobre la demanda

Se pueden extraer varias conclusiones sobre las figuras 1-4. En primer lugar, la tendencia ascendente de la demanda en los últimos años se corresponde con un aumento de nivel del producto interior bruto. El carácter estacional de la serie refuerza la decisión de plantear un modelo ARIMA para explicar los datos. La temperatura presenta poca variación y no afecta a la demanda hasta superar el umbral de los 17-18 °C. Esto se tendrá en cuenta al plantear un modelo de regresión dinámica. En la figura 3, se observa que el periodo de mayor demanda se corresponde con el periodo de septiembre a diciembre.

Estos son los meses más calurosos en Bolivia y concuerda con el aumento de consumo energético que se espera de los meses de verano. La figura 5 muestra esta relación entre temperatura, demanda y mes del año donde queda demostrada esta hipótesis. Respecto a la variable binaria que representa día especial se puede decir que hay una diferencia apreciable entre el valor medio de la demanda según el tipo de día (figura 4).

Algunas conclusiones finales a destacar tras este primer análisis son:

- El mes con mayor demanda fue Enero 2020.
- El mes con menor demanda fue Abril 2020.
- Los lunes presentan menos demanda con un margen considerable respecto al resto.
- El día de la semana con mayor demanda media varía entre jueves, viernes y sábado.

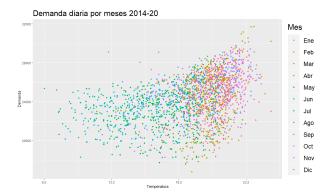


Fig. 5. Relación demanda, mes y temperatura

Las conclusiones sobre los días de la semana se pueden observar en la figura 6

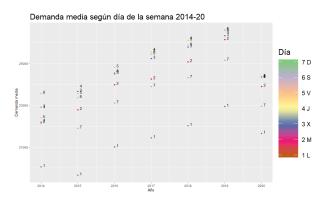


Fig. 6. Demanda media según día de la semana

## 2.2 Detección del efecto de la pandemia

Como última conclusión de este punto, se determinará si el COVID ha sido un factor importante en cuanto al consumo energético en Bolivia. En algunas de las figuras anteriores (1,3 y 4) se puede ver la bajada y ruptura de tendencia creciente en 2020, pero se tratará de mostrar más claramente en los siguientes gráficos. En primer lugar se verá más de cerca el impacto del COVID comparándolo con la tendencia de los últimos tres años (figura 7). También se mostrarán los datos únicamente de 2020 (figura 8) para observar más de cerca el cambio de tendencia.

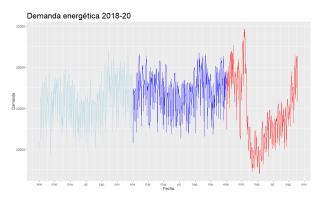


Fig. 7. Serie de demanda 2018-20

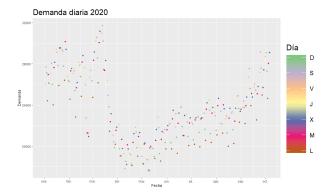


Fig. 8. Datos diarios 2020

En la figura 8 se aprecia claramente el descenso abrupto de los niveles de consumo a finales del primer trimestre de 2020. Para ver exáctamente en fechas se produce se muestra el mes de marzo de 2019 y de 2020 para ver las diferencias en la serie (figuras 9 y 10).



Fig. 9. Demanda marzo 2019

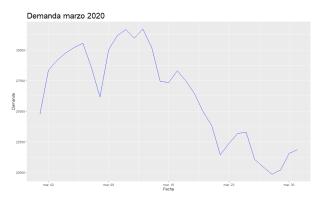


Fig. 10. Demanda marzo 2020

Se puede determinar entonces con suficiente certeza que el COVID sí que tuvo un efecto considerable en la demanda energética de Bolivia. Esto se produjo a mediados de marzo (16 marzo) de una manera extrremadamente pronunciada. Sin embargo, parece que es posible que el nivel de demanda se haya estabilizado de nuevo como indican los últimos datos de septiembre y octubre. En estos meses, la demanda media parece haberse acercado a lo esperado para niveles previos a la pandemia.

# 3 ARIMA previo a la pandemia

Estos modelos se ajustan con datos de demanda hasta el 16 de marzo de 2020. Esta fecha ha sido determinadad en los apartados anteriores. Antes de comenzar a ajustar el modelo se guardan los 7 últimos datos de demanda como conjunto de validación. Con estos datos se harán comparaciones de los errores de predicción para los horizontes 1-7.

### 3.1 Ajuste del modelo ARIMA simple

El primero de los modelos consiste en realizar un ajuste ARIMA contando úncamente con los datos de demanda. De esta manera no se hará uso de ninguna de las variables adicionales. Inicialmente, se determina que es necesario diferenciar tanto regular como estacionalmente. La figura 11 muestra las funciones de autocorrelación tras haber realizado diferenciación de grado uno regular y estacional.

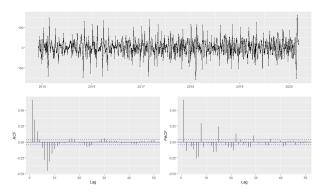


Fig. 11. Funciones de autocorrelación tras diferenciar regular y estacionalmente

En cuanto a posibles transformaciones, el aumento de nivel medio apenas explica un 1% de la variabilidad. Aun así, se construiran dos modelos paralelos con y sin transformación de Box-Cox y sus resultados serán comparados. Para el modelo que sí que hace uso de esta transformación, el valor es:  $\lambda=0.61$ . Las funciones de autocorrelación de la figura 11 no varía significativamente con la aplicación de la transformación.

Tras realizar los ajustes pertinentes a los coeficientes regulares y estacionales se obtienen unos parámetros que reflejan resultados válidos del ajuste: todos los coeficientes son significativos y los residuos no presentan correlación entre sí.

```
z test of coefficients:

z test of coefficient
```

Fig. 12. Nivel de significación de los coeficientes para ambos modelos (superior sin transformación)

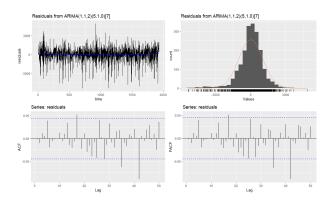


Fig. 13. Residuos de ARIMA sin transformación

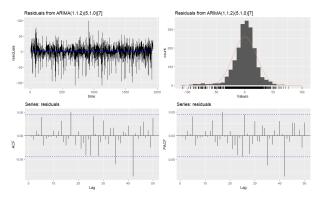


Fig. 14. Residuos de ARIMA con transformación Box-Cox ( $\lambda=0.61$ )

# 3.2 Ajuste del modelo ARIMA con regresión dinámica

Las variables que se utilizarán para el modelo de regresión dinámica son: temperatura y día especial. Esto se ha decidido observando las gráficas iniciales y comparando varias combinaciones distintas para escoger la que proporcionaba mejores resultados. Es importante remarcar que la semana y el mes se tienen en cuenta en la componente estacional del modelo ARIMA por lo que no es necesario incluir estas variables adicionales.

La temperatura no tiene relación lineal con la demanda (ver figura 2), por lo que se dividirá en dos variables distintas: TCOLD y THOT partiendo los

valores reales en 17.5. De esta manera se consigue una relación lineal que peude ser capturada por un modelo ARIMA. La figura 15 muestra la distribución de los valores de temperatura.

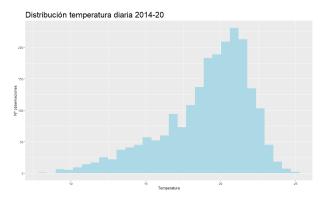


Fig. 15. Distribución de temperaturas diarias

Las distintas variables se escalan para que tengan la misma magnitud y no haya problemas de importancia de variables en el ajuste lineal del modelo de regresión dinámica. Las series resultantes se muestran en la figura 16.

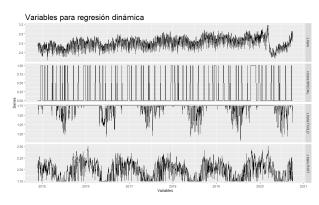


Fig. 16. Series de las variables para la regresión dinámica

Para ajustar el modelo, en primer lugar se hace el análisis de sensibilidad para cada una de las variables explicativas. Se determina que es necesario un nivel de diferenciación estacional y que ninguna de las variables precisa de retardo (lag). Una vez se ha comprobado esto y se han obtenido los coeficientes ARMA de las variables explicativas, se procede a ajustar el modelo. Los resultados se muestran en las figuras 17 y 18. Es importante indicar que se ha comprobado si existe correlación cruzada entre los residuos y las variables explicativas y se ha descartado la posibilidad tras hacer los ajustes pertinentes (ver código para más detalle).

Para el ajuste de los tres modelos ARIMA representados se ha tenido en cuenta también la estabilidad de las raíces, el test de Ljung-Box y la correlación cruzada entre residuos y variables explicativas. Estos gráficos y resultados no han sido representados por simplificación del proceso de explicación. Para ver los valores correspondientes se puede consultar el código adjunto.

#### z test of coefficients:

```
Estimate
                        Std. Error
                                       value
             0.4174195
                         0.0229918
                                                2.2e-16
ma1
                                     18.1552
             0.0887601
                         0.0224145
                                      3.9599
                                                497e-05
sar1
             0.1265474
                         0.0306318
                                      4.1312
                                             3.608e-05
sma1
             0.8014572
                         0.0189301
                                     -42.3378
                                                2.2e-16
SPECIAL-MAO
                         0.0072044
            -0.0914125
                                     -12.6884
                                                        222
SPECTAL -MA1
            -0.0837310
                         0.0075392
                                    -11.1061
                                                2.2e-16
SPECIAL-MA2
             0.0154375
                         0.0071257
                                      2.1665
                                                0.03028
TCOLD-MA0
             0.1179421
                         0.0171376
                                      6.8821
                                             5.900e-12
THOT-AR1
             0.3098405
                         0.0260130
                                     11.9110
                                                2.2e-16
THOT-MAO
             0.4439109
                         0.0151694
                                     29.2637
                                                2.2e-16
                0 '***
Signif, codes:
                         0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Fig. 17. Nivel de significación de los coeficientes

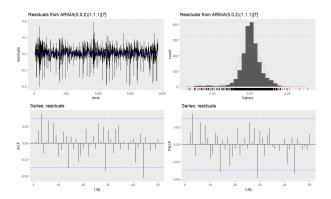


Fig. 18. Residuos de ARIMA con reg. din.

# 3.3 Coeficientes de los modelos ajustados y error de estimación

Los distintos modelos ajustados en el apartado 3 se muestran en la tabla 1 junto a los errores de estimación. Estos errores representan el RMSE de predicción a un horizonte de 1 día calculado con las 7 muestras del conjunto de validación (Rolling Error).

Modelo ARIMA	DReg	RMSE
$(1,1,2)(5,1,0)[7] (\lambda = 0.61)$	0	2029.00
(1,1,2)(5,1,0)[7]	0	2016.76
(0,0,2)(1,1,1)[7](0,2)(0,0)(1,0)	1	1326.72

Tabla 1. Comparación modelos I

### 4 ARIMA considerando la pandemia

Para ajustar este modelo se utilizan todos los datos de demanda. Se utilizan las mismas variables que en la regresión dinámica anterior pero se le añade una variable de intervanción que indica si se está en época de COVID o no. Se utiliza un conjunto de validación de 7 datos para que los resultados de error sean comparables.

En este caso, no se realizará un modelo ARIMA sin regresión dinámica porque se ha visto en la Tabla 1 que los resultados de la regresión dinámica son considerable mejores que un modelo ARIMA simple. Además, sin regresión dinámica no habría manera de indicar que los datos pertenecen al periodo COVID.

# 4.1 Ajuste del modelo ARIMA con regresión dinámica

De manera similar al caso anterior, se muestran las series temporales de las variables explicativas junto a la de demanda. Estan han sido escaladas para tener magnitudes similares.

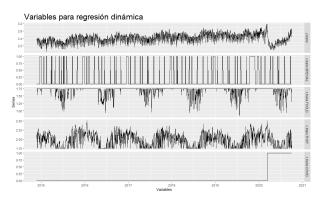


Fig. 19. Residuos de ARIMA con reg. din.

Se ajusta el modelo tras hacer el análisis de sensibilidad y realizar una diferenciación estacional de grado uno. Se obtienen de esta manera los coeficientes ARMA de las variables explicativas. Esto se introduce en el ajuste final del modelo ARIMA con regresión dinámica.

En este caso, sí que es necesario introducir un retardo en una de las variables. La variable COVID resulta no ser significativa en las primeras muestras del análisis de sensibilidad y ha de retrasarse antes del ajuste. Las figuras 20 y 21 muestran los resultados del modelo.

### z test of coefficients:

```
Estimate Std. Error
                                     z value
                                               Pr(>|z|)
              0.5015891
                         0.0231490
                                     21.6678
ma1
                                              < 2.2e-16
              0.2212277
                         0.0229974
ma3
              0.1135807
                         0.0209730
                                       5.4156
                                              6
                                                110e-08
                                     29.1679
sma1
             0.6596223
                         0.0226146
                                                2.2e-16
sma2
             -0.0859620
                         0.0229571
                                      3.7445
SPECTAL -MAO
             -0.0810889
                         0.0071289
                                     -11.3747
                                                2.2e-16
SPECIAL-MA1
              0.0716112
                         0.0070378
                                     -10.1752
                                                2.2e-16
TCOLD-MA0
              0.1069393
                         0.0179089
                                        .9713
TCOLD-MA1
              0.0461973
                         0.0174224
                                       2,6516 0,0080112
THOT-MAO
              0.4717988
                         0.0155171
                                     30.4051
                                                2.2e-16
COVID-MA0
             0.5493079
                         0.0370368
                                     -14.8314
                                                2.2e-16
                0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1
```

Fig. 20. Nivel de significación de los coeficientes

De nuevo, al ajustar este modelo se han hecho las comprobaciones correspondientes sobre la correlación cruzada de residuos y variable explicativa. Además se ha hecho el test de Ljung-Box y se ha asegurado que las raíces del modelo sean estables. Este proceso está detallado en el código adjunto.

En este caso, la variable COVID parece necesitar una diferenciación regular pero este parámetro no está aceptado en la función utilizada. En caso de querer mejorar la predicción de este modelo, se propone como posible mejora ajustar esta función. Los resultados de este modelo se ven junto al resto de modelos en la tabla 2

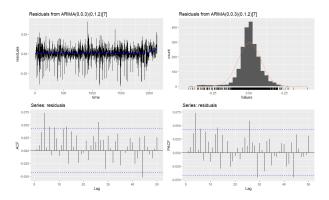


Fig. 21. Nivel de significación de los coeficientes

Modelo ARIMA	RMSE
$(1,1,2)(5,1,0)[7] (\lambda = 0.61)$	2029.00
(1,1,2)(5,1,0)[7]	2016.76
(0,0,2)(1,1,1)[7](0,2)(0,0)(1,0)	1326.72
(0,0,3)(0,1,2)[7](0,1)(0,1)(0,0)(0,0)	4932.976

Tabla 2. Comparación modelos II

#### 5 Efecto del COVID sobre la demanda

Para proporcionar una estimación válida del efecto de la pandemia sobre la demanda se compararán los valores esperados con los valores observados durante los meses afectados por el COVID. Para poner en contexto este error, se hará el mismo procedimiento para el periodo de tiempo correspondiente de 2019 y se compararán los errores. Finalmente, se determinará el impacto económico causado por el brusco descenso de demanda.

### 5.1 Comparación de estimación y serie real

Para hacer esta comparación se hace uso del modelo ARIMA con regresión dinámica ajustado con sin los datos de la pandemia. Con este modelo se realiza una predicción sobre el periodo de tiempo de la pandemia y se compara con el valor real de estos datos. Estos resultados se ven en la figura 22.



Fig. 22. Demanda real y esperada 2020

Se ve la caída inicial debida al súbido impacto en la demanda y una progresiva recuperación de los niveles de demanda a lo largo de los meses siguientes. Se puede además observar cómo la componente estacional no se ve afectada. La serie de valores esperados sigue de cerca las variaciones observadas en la serie real. Esto nos permite afirmar que, aunque el efecto ha sido importante, no ha alterado de manera significativa los hábitos de consumo de la población. Es importante indicar que desde la tercera semana de septiembre, los niveles de demanda han alcanzado el valor de la predicción. Podría considerarse una recuperación aunque serían necesarios más datos temporales para poder tener más confianza en esta afirmación.

Para tener una base de comparación sólida para los errores de estimación, se ha hecho el mismo proceso con los datos de 2019. Con el mismo modelo ARIMA, se han predicho los valores de 2019 correspondientes al mismo periodo de tiempo del COVID. Los resultados de los errores de 2019 y 2020 se muestran en las imágenes 23 y 24.

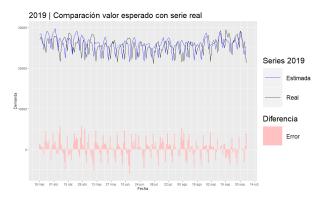


Fig. 23. Errores predicción 2019

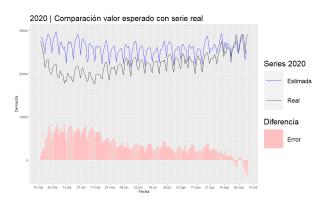


Fig. 24. Errores predicción 2020

Los resultados son los esperados: el error de estimación de demanda en 2019 parece seguir la estructura de ruido blanco y el error de 2020 claramente tiene una relación que no está siendo calculada. Los valores negativos de error  $(D_{esperada} - D_{observada})$  refuerzan la idea anterior sobre la recuperación de niveles de demanda. Con estas diferencias calculadas se puede hacer una estimación del impacto económico de la pandemia en lo referente a la demanda energética.

#### 5.2 Cálculo del impacto económico

Para calcular el impacto económico se calculará cuántos KWh han dejado de consumirse por efecto de la pandemia y se hará una estimación del precio de esa energía no consumida. Esto será considerado como el impacto o pérdida económica debido a la pandemia.

$$Impacto = (\Delta Demanda_{2020}) * precio$$

#### 5.2.1 Estimación de variación de demanda

La demanda esperada y observada son las dos series representadas en la figura 22 y 23, pero antes de hacer la multiplicación se eliminarán ciertos valores de error. Los valores eliminados serán los correspondientes a fechas posteriores al 23 de septiembre de 2020. Esta fecha es la primera que tiene un valor negativo de error, por lo que será considerada la "fecha de recuperación". Los errores posteriores a esta fecha se consideran errores esperados y no se tendrán en cuenta en el cálculo del impacto económico.

Como medida de correción adicional, a todas las variaciones se les restará el error medio de 2019. Este error medio es pequeño comparado con el valor absoluto de los errores (error en 2019 se aproxima a ruido blanco). Se tendrá en cuenta como factor de correción del modelo al tener un cierto nivel de significación (6% del error máximo de 2019). En el cálculo se utilizará un rango de valores de demanda que tendrá como límites la diferecia con y sin corrección de 2019.

$$\Delta D_{2020} = D_{2020}^{esperada} - D_{2020}^{observada} (-avg(\Delta D_{2019}))^*$$

\* Solo se aplica en el límite inferior de la variación de demanda.

#### 5.2.2 Estimación de precio

Para estimar el precio por KWh se ha tomado como referecia datos del diario Economy (Bolivia). Este diario publicó en  $2019^1$  los datos de la tarifa de electricidad tras una reducción de precios importante acometida por el gobierno de Jeanine Áñez. Estos precios están entre los más bajos de América del Sur $^2$  con una media de 9cts $^4$ /KWh. Los bajos precios se deben a que un 70% de energía es generada en plantas termoeléctricas subvencionadas por el gobierno.

El precio usado en la estimación es la media los precios actuales representados en el artículo del diario Economy. Esta estimación, junto con la horquilla proporcionada en demanda permite tener confianza sobre el valor final del impacto económico.

Distribuidora / Ciudad	Consumo	Coste por kWh (Bs.)
Delapaz (La Paz)	+ 1.000 kWh	0.77
	De 500 a 1.000 kWh	0.73
Sepsa (Potosí)	+ 1.000 kWh	0.90
	De 500 a 1.000 kWh	0.80
ENDE (Oruro)	+ 1.000 kWh	0.90
	De 500 a 1.000 kWh	0.85
Setar (Tarija)	+ 1.000 kWh	0.98
	De 500 a 1.000 kWh	0.98
CRE (Santa Cruz)	+ 1.000 kWh	0.98
	De 500 a 1.000 kWh	0.93
CESSA (Chuquisaca)	+ 1.000 kWh	0.99
	De 500 a 1.000 kWh	0.91
ENDE (Beni)	+ 1.000 kWh	0.99
	De 500 a 1.000 kWh	0.85
Elfec (Cochabamba)	De 500 a 1000 kWh	1.00
	+ 1.000 kWh	1.04
	MEDIA	0.91

Fig. 25. Precios energía Bolivia (Diario Economy)

#### 5.2.3 Impacto económico

$$Impacto = (\Delta Demanda_{2020}) * precio$$

Usando las aproximaciones expuestas anteriormente se obtiene que el impacto económico es superior a medio millón de bolivianos.

$$Impacto \approx 612.000 - 673.000Bs$$

Este valor parece ser más pequeño del esperado. Para explicar esto es importante tener en cuenta dos factores. En primer lugar, este impacto se corresponde únicamente con la pérdida debida a la energía no vendida. No tiene en cuenta las contributciones adicionales derivadas del beneficio perdido que hubiera generado la consumición de esa energía. En segundo lugar, como se ha remarcado antes, el precio de la energía es muy bajo debido a subvenciones estatales. Esto quiere decir que el impacto real es mayor, pero al estar gran parte cubierto por el gobierno no se aprecia con el método utilizado. Si se hiciera un ajuste para tener en cuenta el 70% de subvención estatal, el impacto sería 2-2.25 M Bs que es una cifra con más sentido.

#### 6 Conclusión

La pandemia tuvo un efecto inmediato sobre la demanda energética en Bolivia. Esto se hizo notar a partir de la segunda quincena de marzo. Los niveles de consumo energético bajaron bruscamente en tándem con el decremento del producto interior bruto. Tras estos efectos iniciales, el consumo llegó a cifras mínimas en las primeras semanas de mayo y desde entonces ha seguido una tendencia ascendente. Durante todo este periodo, a pesar de que los niveles se vieran afectados de una manera importante, las tendencias y hábitos de consumo no variaron. Esto, junto con la recuperación de niveles previos a la pandemia observada a partir de las últimas semanas de septiembre permite una perpectiva optimista de futuro.

<sup>&</sup>lt;sup>1</sup>VARELA. E., Beni recibe la mayor reducción de tarifas de electricidad, pero La Paz paga el costo más barato por kilovatio-hora. Economy, Noviembre 2019

<sup>&</sup>lt;sup>2</sup>VALDEZ. C., Nuevo gobierno de Bolivia rebaja la tarifa eléctrica. Associated Press, Diciembre 2019

Por esa razón, se propone que como modelo de predicción de demanda se use el ARIMA con regresión dinámica expuesto en el apartado 3.2. Este modelo consigue ajustarse de manera adecuada a los valores observados en el periodo de recuperación, aunque no vale para la etapa de la pandemia (no se considera etapa de pandemia en lo referente a consumo desde el 23 de septiembre de 2020). Esta elección debería ser revisada con los valores del último trimestre de 2020. En caso de presentar problemas, se podría ajustar el modelo con la variable de intervención del COVID aunque este actualmente presente más error al tener menos datos con los que entrenarse. En caso de seguir este segundo camino, se sugiere que la variable de intervención esté activa desde el 16 de marzo hasta el 23 de septiembre.

En cuanto al impacto económico de la pandemia sobre la demanda energética, se ha estimado una variación de 6.7-7.4 GWh que corresponden a aproximadamente 2-2.5 millones de Bs.