

Forecasting the growth rate of new weekly deaths attributed by COVID-19 in Mexico using a hybrid ARIMA-ARCH model

(Pronóstico tasa de crecimiento de nuevos fallecimientos semanales atribuidos por COVID-19 en México usando un modelo híbrido ARIMA-ARCH)

Enrique Hernández-Laredo
Facultad de Medicina Universidad Autónoma del Estado de México
Toluca de Lerdo, México
ehernandezl190@alumno.uaemex.mx

ABSTRACT

This work presents the design of an ARMA(1,4)-ARCH(1) model to be used to generate useful information for the study of epidemiology and public policies for the control of viral spread. This model showed stability (residuals without serial correlation and homoscedastic), and also when comparing its forecasts with historical data, it presented an acceptable fit (60% of these forecasts remained within the 95% confidence intervals).

These results demonstrate that the ARMA(1,4)-ARCH(1) model generated in the present work is a stable, freely available and computationally inexpensive tool for making volatility forecasts and point estimates of the weekly death rate due to COVID-19 in Mexico. It also solves the limitations of ARMA models.

Keywords: *Forecasting-for-COVID-19, ARMA-ARCH, econometrics-in-health care.*

I. INTRODUCCION.

El COVID-19 es la enfermedad causada por el virus coronavirus SARS-CoV-2. La Organización Mundial de la Salud (OMS) tuvo noticia por primera vez de la existencia de este nuevo virus en el 2019, pero dado el aumento en los números de nuevos casos positivos por COVID-19 alrededor del mundo, en el mes de marzo del 2020 la OMS declaró la pandemia global por COVID-19, la cual ha generado en todo el mundo, una severa crisis económica, social y de salud (1,2). En este sentido, han surgido propuestas basadas en modelos estadísticos, matemáticos y/o de inteligencia artificial para el pronóstico en el número de fallecimientos a causa del COVID-19 (la generación de pronósticos también suelen ser llamada *forecasting*), esto con el fin de generar información útil en el estudio de la epidemiología y de políticas públicas para el control de la propagación viral (3,4).

El modelo de media móvil integrada autorregresiva (ARIMA) es el método de series temporales más empleado para pronosticar el número de nuevos casos positivos y

fallecimientos por COVID-19 (5), la mayoría de los investigadores de diferentes disciplinas utilizaron modelos ARIMA debido a su simplicidad, estructura sistemática y rendimiento aceptable en pronósticos (6), como el trabajo presentado anteriormente (7), donde se utilizó un modelo ARIMA (4,1,1) para pronosticar el número de fallecimientos semanales por COVID-19 en México en un periodo contenido de la 1er semana de junio a la 1er semana del julio del 2022, basado en los datos históricos del 1 de enero del 2020 hasta el 29 de mayo de 2022, y a pesar de que dicho modelo es capaz de estimar de manera correcta la tendencia de los fallecimientos, Aykut Ekinci (6) reconoce que existen limitantes en los modelos ARIMAs, pues son modelos lineales bajo el supuesto de homocedasticidad, en este sentido sugiere que el modelo de heterocedasticidad condicional autorregresiva (ARCH) puede tener un mejor rendimiento en comparación con los métodos estadísticos tradicionales como ARMA, dado que permiten capturar las no linealidades y la heterogeneidad de las series temporales (6,8).

El propósito del presente trabajo fue realizar un modelo híbrido ARMA-ARCH a partir de los hallazgos en (7), con el fin de eliminar las limitaciones del modelo ARIMA(4,1,1), y generar un modelo mejor ajustado que permita pronosticar la volatilidad y estimaciones puntuales de la tasa de crecimiento del número de fallecimientos semanales ocasionados por COVID-19 en México.

II. METODOLOGÍA.

A. Base de datos:

Se utilizó la base de datos "*owid-covid-data*" del repositorio público de *Our World in Data* (9), que contiene información sobre COVID-19 referente a vacunas, pruebas y positividad, pacientes en hospitales, casos confirmados, muertes confirmadas, tasa de reproducción del COVID-19 y respuesta de políticas de al menos 47 países, incluyendo México.

B. Preparación de los datos:

Se filtró la base de datos para incluir únicamente la serie de tiempo con información del número de nuevos fallecimientos diarios atribuidos por COVID-19 en México

durante el periodo contenido del 1 de marzo del 2020 hasta el 29 de mayo de 2022. Se reemplazaron valores nulos y se calcularon los promedios de muertes semanales de acuerdo con lo propuesto en (7), a partir de estos datos se construyó la serie de tiempo de la tasa de crecimiento de fallecimientos semanales atribuidos por COVID-19 en México ($TC_{(t)}$) dada la Ec.1 (6) y posteriormente se excluyeron los valores nulos, Inf y -Inf.

$$TC_{(t)} = Ln\big(Y_{(t)} \big) - Ln(Y_{(t-1)}), \tag{Ec.1}$$

para un tiempo t = 0,1, 2, ...n. donde n es el número total de muestras.

$Y_{(t)}$ = la serie de tiempo de los fallecimientos semanales atribuidos por COVID-19 en México.

C. Modelado:

A partir de $TC_{(t)}$ se desarrolló un modelo hibrido ARMA-ARCH de acuerdo con las siguientes etapas:

C.1) Estacionariedad

La estacionariedad de las series de tiempo $TC_{(t)}$, se estudió mediante los estadísticos Dickey-Fuller aumentada (DF), Phillips-Perron (PP), y KPSS, los cuales fueron ejecutadas mediante un nivel de significancia (α) del 0.05.

C.2) Identificación de la volatilidad

Se verificó la existencia de volatilidad condicional mediante la prueba ARCH (heterocedasticidad condicional). Asimismo, se calculó el cuadrado de $TC_{(t)}$, y se ejecutaron los estadístico Ljung-Box y la función de autocorrelación simple (ACF) para verificar la presencia de correlación serial, estas pruebas fueron ejecutadas mediante un $\alpha=0.05$.

C.3) Estimación del modelo ARMA-ARCH

Los coeficientes autorregresivo (p) y de medias móviles (q) del del modelo ARMA(p,q) se estimaron mediante la función auto.arima (10), y posteriormente se estimó un el modelo hibrido ARMA(p,q)-ARCH(1).

D. Evaluación del modelo:

Se verificó la estabilidad del modelo hibrido ARMA(p,q)-ARCH(1) mediante la normalidad y no autocorrelación de los residuos (r) a partir de los estadísticos Jarque-Bera y Ljung-Box respectivamente. La presencia heteroscedasticidad de los r y r^2 se verificó mediante las pruebas Ljung-Box y LM ARCH respectivamente. En caso de no presentar estabilidad en el modelo ARMA(p,q)-ARCH(1) se estimarán órdenes superiores el modelo ARCH.

E. Pronostico:

Se graficaron los datos de la serie de tiempo $TC_{(t)}$ y los valores estimados por el mejor modelo.

Por último, se realizaron las estimaciones de la tasa de crecimiento del número de fallecimientos semanales ocasionados por COVID-19 en México y su volatilidad del periodo contenido de la 1er semana de junio a la 1er semana del julio del 2022.

III. RESULTADOS

En la **¡Error! No se encuentra el origen de la referencia.** se muestra la serie de tiempo de la tasa de crecimiento de los fallecimientos semanales atribuidos por COVID-19 en México ($TC_{(t)}$) en un periodo contenido del 1 de marzo del 2020 hasta el 29 de mayo de 2022, con 113 muestras posterior a la eliminación de valores faltantes e incongruentes. Se asume que dicha serie de tiempo presenta un comportamiento estacionario pues el valor de probabilidad (p-valor) del estadístico DF y PP se encuentran por debajo del $\alpha = 0.05$, por lo tanto, al encontrarse en un error tipo I, se rechaza H0: no estacionariedad, y se concluye existe estacionariedad en la serie $TC_{(t)}$ (Ver Tabla 1).

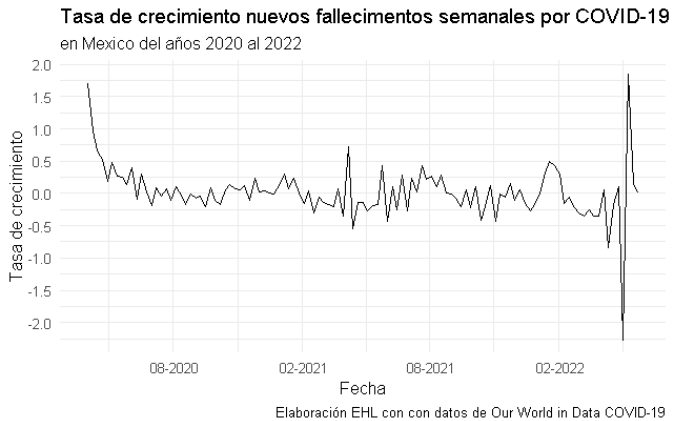


Fig. 1 Serie de tiempo tasa de crecimiento de los fallecimientos semanales atribuidos por COVID-19 en México.

Tabla 1 Pruebas estadísticas de estacionariedad para $TC_{(t)}$.

Prueba	p-valor
Dickey-Fuller aumentada	0.0168
Phillips-Perron	<0.010
KPSS	0.0229

De la prueba ARCH se obtuvo un p-valor <0.001 menor al $\alpha = 0.05$, por lo que, mediante un error tipo I, se rechaza H0: no heterocedasticidad condicional, y se concluye existen efectos ARCH en la serie $TC_{(t)}$.

En la Fig. 2 y Fig. 3 se muestran la ventana de volatilidad ($TC_{(t)}^2$) y su ACF respectivamente, a pesar de que dicha ACF no muestra indicios de correlación serial para rezagos mayores a 3, se asume que $TC_{(t)}^2$ presenta correlación serial pues la prueba Ljung-Box presento un p-valor = 0.001 menor al $\alpha = 0.05$, por lo que, mediante un error tipo I, se rechaza H0: los datos se distribuyen independientemente, por lo que concluye que los datos no se distribuyen de manera independiente y presentan correlación serial.

Tasa de crecimiento al cuadrado de nuevos fallecimientos diarios por COVID-19
en México del años 2020 al 2022



Fig. 2 Tasa de crecimiento al cuadrado de los fallecimientos semanales atribuidos por COVID-19 en México (ventana de volatilidad).

Función de autocorrelación simple

Tasa de crecimiento nuevos fallecimientos semanales por COVID-19

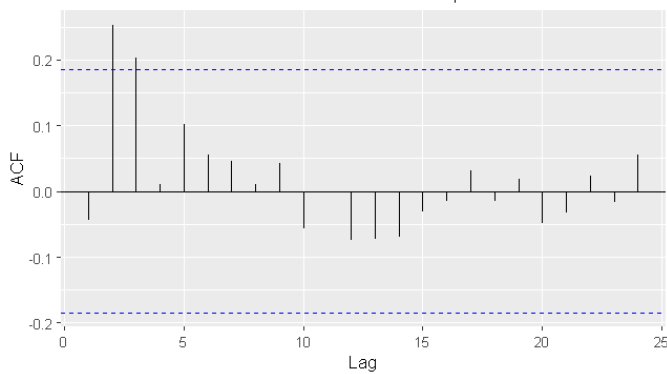


Fig. 3 Función de autocorrelación de la tasa de crecimiento al cuadrado de los fallecimientos semanales atribuidos por COVID-19 en México.

La función `auto.arima` estimó que el modelo ARMA con mejor ajuste a la serie $TC_{(t)}$ es el ARMA(1,4), el cual permitió generar el modelo híbrido ARMA(1,4)-ARCH(1) (Ver Tabla 2).

Tabla 2 Modelo ARIMA(1,4)-ARCH(1).

Coefficientes	Error estándar	Error	p-valor
mu	0.010517	0.04088	0.796973
ar1	-0.176795	0.202115	0.381724
ma1	0.242675	0.192473	1.261
ma2	0.632522	0.079357	7.971
ma3	0.346367	0.108794	3.184
ma4	0.305786	0.12137	2.519
omega	0.021963	0.005994	3.664
alpha1	1	0.312064	3.204

Significancia estadística p valor < 0 *** , < 0.001 ** , < 0.01 *

El p-valor=0.017 de la prueba Jarque-Bera para los r del modelo ARMA(1,4)-ARCH(1) es menor al $\alpha = 0.05$, por lo que, mediante un error tipo I, se rechaza H_0 : distribución normal, y se concluye que los r no se distribuyen de forma normal. Asimismo, para r la prueba Ljung-Box estimó p-valores = 0.774, 7.825, y 11.882 para los rezagos 10, 15 y 20 respectivamente (todos mayores al $\alpha = 0.05$); por lo que, mediante un error tipo

II, no hay evidencia suficiente para rechazar H_0 : los r se distribuyen independientemente, por lo que concluye la no presencia de correlación serial en los r . Por otra parte, los p valores = 14.3246 y 13.719 de los estadístico Ljung-Box (rezago=20) y LM ARCH respectivamente, se encuentran por encima del $\alpha = 0.05$, por lo tanto, al encontrarse en un error tipo II, no hay evidencia suficiente para rechazar H_0 : no heteroscedasticidad y se concluye que existe homocedasticidad en los r y r^2 . De tal forma que se asume que el modelo ARMA(1,4)-ARCH(1) es estable.

Se calcularon los pronósticos de la tasa de crecimiento y volatilidad del número de fallecimientos semanales ocasionados por COVID-19 en México del periodo contenido de la 1er semana de junio a la 1er semana del julio del 2022 (Ver Tabla 3 y Apéndice 1).

Tabla 3 Estimación del modelo ARMA(1,4)-ARCH(1) en el periodo contenido de la 1er semana de junio a la 1er semana del julio del 2022.

TC real	Estimación ARMA(1,4)-ARCH(1)				
	TC	volatilidad	SD	Mínimo IC 95% TC	Máximo IC 95% TC
0.492	0.571	0.931	0.931	-1.253	2.397
-0.364	0.326	0.945	0.943	-1.526	2.178
0.215	0.002	1.117	0.954	-2.188	2.1936
-0.215	-0.271	1.152	0.966	-2.530	1.988
0.307	0.058	1.192	0.977	-2.278	2.394

TC= tasa de crecimiento fallecimientos semanales por COVID-19 en México
IC= intervalo de confianza

IV. DISCUSIÓN

La generación de modelos pronósticos es de importancia para la toma de decisiones en política pública, y así lograr un buen control de la propagación viral durante la pandemia por COVID-19 (3,4). Sin embargo, dicho fenómeno tiene fuertes olas de contagio (incrementos acelerados de nuevos casos positivos por COVID-19 en cortos periodos), lo que genera varianzas heterogéneas y distribuciones de cola pesada en la serie de tiempo(6). En este sentido, el propósito del presente trabajo fue dar respuesta a la limitación del incumplimiento de homocedasticidad de varianzas del modelo ARIMA(4,1,1) (7), por lo que se generó un modelo híbrido ARMA(1,4)-ARCH(1) estable.

Los presentes hallazgos muestran que el modelo ARMA(1,4)-ARCH(1) fue capaz de realizar pronósticos de la tasa de crecimiento de los fallecimientos semanales por COVID-19 en México con una precisión menor a 80 centésimas en el 40% de las veces, y además 60% de dichos pronósticos se mantuvieron dentro de los intervalos de confianza al 95%, esto para el periodo contenido de la 1er semana de junio a la 1er semana del julio del 2022

La principal limitación del modelo propuesto es la capacidad predictora, pues a comparación de los modelos ARIMAs, los modelos híbridos de la familia ARCH son incapaces de realizar estimaciones puntuales directamente en el número de

fallecimientos semanales, por tal motivo, están limitados a estimar las tasas de crecimiento de la variable de interés junto con la volatilidad de esta (11).

V. CONCLUSIONES

El modelo ARMA(1,4)-ARCH(1) generado en el presente trabajo es una herramienta estable, de acceso gratuito y que requiere bajo nivel computacional para realizar pronósticos de la volatilidad y estimaciones puntuales de la tasa de fallecimientos semanales a causa del COVID-19 en México. Y que además resuelve las limitaciones de los modelos ARMAs y/o ARIMAs.

VI. ANEXO

El código para realizar el pronóstico de la tasa de fallecimientos semanales a causa del COVID-19 en México y su volatilidad basado en un modelo ARMA(1,4)-ARCH(1), se encuentra disponible en: [Github-enriquehdez98](#). Es importante aclarar que dicho código permite actualizar los datos históricos de manera automática a la fecha de consulta, en este sentido permite reajustar el modelo ARMA(1,4)-ARCH(1), y generar nuevos pronósticos correspondientes a las próximas cinco semanas a partir de la fecha de consulta.

VII. REFERENCIAS

1. Información básica sobre la COVID-19 [Internet]. [cited 2022 Jul 3]. Available from: <https://www.who.int/es/news-room/questions-and-answers/item/coronavirus-disease-covid-19>
2. Heredia Lima C, Ruiz Gallo Lambayeque. El nuevo Coronavirus y la pandemia del Covid-19. Revista Medica Herediana [Internet]. 2020 Jul 31 [cited 2022 Jul 3];31(2):125–31. Available from: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1018-130X2020000200125&lng=es&nrm=iso&tlng=es
3. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The Lancet Infectious Diseases [Internet]. 2020 May 1 [cited 2022 May 29];20(5):553. Available from: </pmc/articles/PMC7158569/>
4. Sharma A, Ahmad Farouk I, Lal SK. COVID-19: A Review on the Novel Coronavirus Disease Evolution, Transmission, Detection, Control and Prevention. Viruses [Internet]. 2021 Feb 1 [cited 2022 May 29];13(2). Available from: </pmc/articles/PMC7911532/>
5. Sotomayor D, Carlos FB. Aplicación del método autorregresivo integrado de medias móviles para el análisis de series de casos de covid-19 en Perú. Revista de la Facultad de Medicina Humana [Internet]. 2021 Jan 12 [cited 2022 May 29];21(1):65–74. Available from: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2308-05312021000100065&lng=es&nrm=iso&tlng=es
6. Ekinci A. Modelling and forecasting of growth rate of new COVID-19 cases in top nine affected countries: Considering conditional variance and asymmetric effect. Chaos, Solitons & Fractals. 2021 Oct 1;151:111227.
7. Hernandez-Laredo E. Forecasting number of new weekly deaths attributed to COVID-19 in Mexico [Internet]. 2022 [cited 2022 Jul 3]. p. 1–7. Available from: https://www.researchgate.net/publication/361151466_Forecasting_number_of_new_weekly_deaths_attributed_to_COVID-19_in_Mexico
8. Mandelbrot BB. The variation of certain speculative prices. Fractals and Scaling in Finance [Internet]. 1997 [cited 2022 Jul 3];371–418. Available from: https://link.springer.com/chapter/10.1007/978-1-4757-2763-0_14
9. Appel Cameron, Beltekian Diana. Data on COVID-19 (coronavirus) by Our World in Data [Internet]. 2022 [cited 2022 May 20]. Available from: <https://github.com/owid/covid-19-data/tree/master/public/data>
10. auto.arima function - RDocumentation [Internet]. [cited 2022 May 29]. Available from: <https://www.rdocumentation.org/packages/forecast/versions/8.16/topics/auto.arima>
11. Larre TF, Auza J. Modelización financiera mediante modelos híbridos Arima–Garch: evidencia para Argentina. Ciencias Económicas [Internet]. 2020 May 29 [cited 2022 Jul 5];2:29–45. Available from: <https://bibliotecavirtual.unl.edu.ar/publicaciones/index.php/CE/article/view/9268>

APÉNDICE 1.

