

La expansión de Covid-19 en México en 2020: un enfoque desde la econometría de series de tiempo

*The expansion of Covid-19 in Mexico in 2020:
an approach from time series econometrics*

Alejandro Rodríguez Arana
Universidad Iberoamericana, Ciudad de México
alejandro.rodriguez@ibero.mx

Resumen

Este trabajo analiza las series de infecciones y muertes acumuladas por Covid-19 en México entre fines de febrero y el 31 de diciembre de 2020. Ambas series resultan integradas de orden 1 (I(1)) y cointegradas, por lo cual se modelan con un vector de corrección de error (VEC).

Diversas estimaciones con la técnica VEC encuentran una elasticidad, de largo plazo, de muertes a contagios que es inferior a la unidad. Este resultado implica que el aumento en los contagios bajaría la tasa de letalidad de la enfermedad, aunque a un ritmo lento.

El análisis es útil para hacer extrapolaciones; si los contagios reales fueran 40 veces superiores a los registrados y las muertes fueran dos veces mayores a las oficiales, tres millones de infectados registrados implicaría un número oficial de muertos de entre 190 mil y 250 mil, pero un número real podría llegar hasta medio millón.

Palabras clave: Covid-19, contagios totales, contagios diarios, muertes totales, muertes diarias, tasa de letalidad

Abstract

This work analyzes the series of infections and deaths accumulated by Covid-19 in Mexico between the end of February and December 31, 2020. Both series are integrated of order 1 (I(1)) and cointegrated, for which they are modeled with an error correction vector (VEC).

Several VEC estimations find a long-term elasticity of deaths to infections lower than one. This result implies that the increase in infections will reduce the fatality rate of the disease, although at a slow pace.

The analysis is useful to make extrapolations, if the real infections were 40 times higher than those registered and the deaths were twice the number of the official ones, three million registered infected would imply an official death number of between 190 and 250 thousand, but a number real not very far from half a million.

Keywords: Covid-19, total infections, daily infections, total deaths, daily deaths, fatality rate.

JEL: C22, C53, I10, I12

Fecha de recepción: 19 de octubre de 2020.

Fecha de aceptación: 23 de abril de 2021.

1. Introducción

La enfermedad, que en inglés se denomina *coronavirus disease 2019* y que se conoce más comúnmente como Covid-19, surgió en China a fines de 2019 y está causando graves estragos en la salud de la mayor parte de los países del mundo. En México, los primeros casos se reportaron a fines de febrero, pero la enfermedad se expandió con gran rapidez, y por un tiempo su propagación se parecía a una curva exponencial. A fines de septiembre, se reportaron cerca de 740 mil infecciones acumuladas y más de 77 mil fallecimientos en los mismos términos. Para fines de diciembre, estos mismos números fueron de más de 1 400 000 y de más de 120 mil, en el mismo orden. La tasa de letalidad¹ de este mal en México fue cercana a 9.0% en los últimos meses de 2020, una de las más grandes del mundo.

Los efectos nocivos sobre la salud que causa esta enfermedad han propiciado que se tomen medidas de confinamiento, las cuales han sido obligatorias en muchos lugares del mundo. Las consecuencias económicas de esas medidas han sido muy graves. En México, más de diez millones de personas dejaron de percibir ingresos por su trabajo en abril. El producto interno bruto (PIB) del segundo trimestre de 2020 cayó más de 18%, aunque después ha habido una recuperación parcial. Para fin de 2020 se espera una caída anual promedio del PIB en México cercana a 10%.

Los hechos anteriores sugieren que es muy relevante conocer la dinámica de la expansión de la enfermedad y de los fallecimientos que causa, pues de esa manera pueden tomarse las medidas más adecuadas, tanto en términos de salud como económicas. En comparación con otros países, México ha llevado a cabo pocas pruebas de laboratorio para detectarla. Por tal motivo, también en términos relativos, resultan importantes otras formas de detectar la propagación de este mal.

Este trabajo utiliza técnicas de la econometría de series de tiempo para caracterizar el proceso dinámico de propagación de Covid-19 y de las muertes asociadas a esta enfermedad en México, entre los últimos días de febrero de 2020 y el 31 de diciembre de ese mismo año. En particular, el estudio estima la elasticidad de largo plazo de las muertes con respecto a los contagios. Este concepto es importante

¹ Por tasa de letalidad se entiende el número de personas fallecidas por la enfermedad divididas entre el número total de personas contagiadas por el mismo mal. En cambio, tasa de mortalidad es el número de personas fallecidas debido a la misma enfermedad, pero divididas por la población total del país en cuestión.

porque al conocer dicha variable también se pueden proyectar las muertes esperadas ante contagios crecientes y la dinámica de la tasa de letalidad de la enfermedad.

El estudio está dividido en cuatro secciones. La primera hace una descripción de los contagios y las muertes generadas por Covid-19 en México entre febrero y diciembre de 2020, como ya se comentó. El segundo apartado analiza estadísticamente las series de contagios y muertes por esta enfermedad, así como su posible relación de largo plazo. Dados los resultados de la segunda sección, la tercera estima un modelo de vector de corrección de error (VEC) bivariado para las series que se analizan. La cuarta lleva a cabo un ejercicio de bondad de ajuste de dos de los VEC estimados en la sección anterior y proyecta la relación de contagios y muertes con varios de los modelos analizados a lo largo del artículo.

Los resultados de la segunda sección de este trabajo sugieren que las series de infecciones y muertes acumuladas en México no son estacionarias, sino integradas de orden 1 (I(1)) y que muy probablemente estén cointegradas. Por tal razón se propone la estimación de un VEC bivariado en la tercera sección.

La estimación de un grupo grande de modelos VEC sugiere que la elasticidad de las muertes a los contagios es menor a la unidad. Esto indica que conforme aumentan los contagios, la tasa de letalidad disminuirá, pero de manera moderada.

Diversos modelos estimados, entre ellos dos VEC, muestran que tres millones de infecciones registradas por Covid-19 podrían ocasionar entre 190 mil y 250 mil muertes. Como es posible que haya un factor de subregistro de contagios de alrededor de 40 veces, y de muertes de entre dos y tres veces, tres millones de contagios implicaría a casi toda la población, o a una proporción tal que se estaría ya en la inmunidad de rebaño, lo que daría lugar a que al final de la pandemia podría haber hasta medio millón de fallecidos en México por Covid-19.

Hay un grupo considerable de estudios empíricos sobre las infecciones y muertes por Covid-19 utilizando técnicas de econometría de series de tiempo. Algunos de ellos usan modelos univariados tipo ARIMA, por ejemplo, Dehesh et al. (2020) para diversos países principalmente de Asia, y Benvenuto et al. (2020) para Italia. Otros, como el de Khan et al. (2020) y Kumari et al. (2021) utilizan la técnica de vectores autorregresivos (VAR) para analizar los contagios, las muertes y los enfermos recuperados en Pakistán e India, respectivamente. En el caso de México, el trabajo de Gómez Cravioto et al. (2020) aborda la expansión de contagios y muertes por Covid-19 en la primera etapa de la enfermedad con diversas técnicas, entre ellas un VAR.

El problema de utilizar la técnica VAR para analizar las muertes y los contagios de la enfermedad que nos concierne es que las series de estas variables no son estacionarias, o cuando menos no muestran los rasgos de estacionariedad tradicionales, por lo cual pueden ser mejores para el análisis otras técnicas como la de VEC. Hemos encontrado muy pocos estudios internacionales que utilicen VEC para abordar contagios y muertes por Covid-19 y no pudimos localizar ninguno para México. El trabajo de Paolo et al. (2020) lleva a cabo el análisis de un VEC para hospitalizaciones y entradas a unidades de cuidados intensivos en algunas regiones de España, Italia y Suiza.

2. La infección de Covid-19 en México

En los últimos días de 2019 se reportó en China la aparición de una nueva enfermedad denominada Covid-19, por las siglas en inglés *coronavirus disease 2019*, como ya se comentó. Esta infección es provocada por un virus identificado como SARS-Cov-2. Inicialmente, la enfermedad se caracteriza por presentar síntomas similares a los de la influenza común, pero más adelante los pulmones, y en ocasiones otros órganos vitales, pueden verse seriamente afectados. En un porcentaje todavía no del todo identificado este mal lleva a la muerte.

La enorme capacidad del virus para replicarse ha implicado que el Covid-19 se encuentre actualmente en la gran mayoría de los países del mundo. España, Estados Unidos e Italia fueron de los más afectados en los primeros meses de 2020; luego los contagios y las muertes cayeron, pero repuntaron pronto en Estados Unidos y más adelante también en España e Italia.

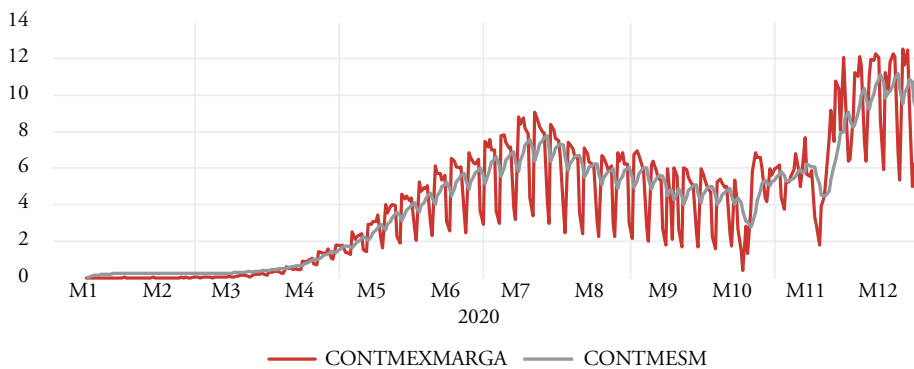
En México, los primeros casos reportados de Covid-19 se hicieron públicos a fines de febrero de 2020. Un mes después ya había más de 700. El 31 de diciembre la cifra acumulada de personas contagiadas fue superior a 1 400 000, y el número oficial de muertos de ese mismo día fue de un poco más de 123 mil.

Hasta los primeros días de octubre, las infecciones y las muertes diarias siguieron el patrón que se conoce como de curva epidémica (ver por ejemplo Guilligan, 1990; Torok, 2003; Lee y Wong, 2010; Yang et al., 2015), en el cual las variables bajo estudio, al nivel diario, suben a lo largo del tiempo, llegan a un máximo y comienzan a caer.

El patrón descrito se rompió en octubre, cuando tanto las infecciones como las muertes diarias comenzaron a subir otra vez, en una nueva curva que todavía pudiera no alcanzar otro máximo local.

La expansión de Covid-19 en México en 2020:
un enfoque desde la econometría de series de tiempo

Gráfica 1. Curva epidémica de contagios por Covid-19 en México
Miles de infecciones diarias
Del 1 de marzo al 31 de diciembre de 2020

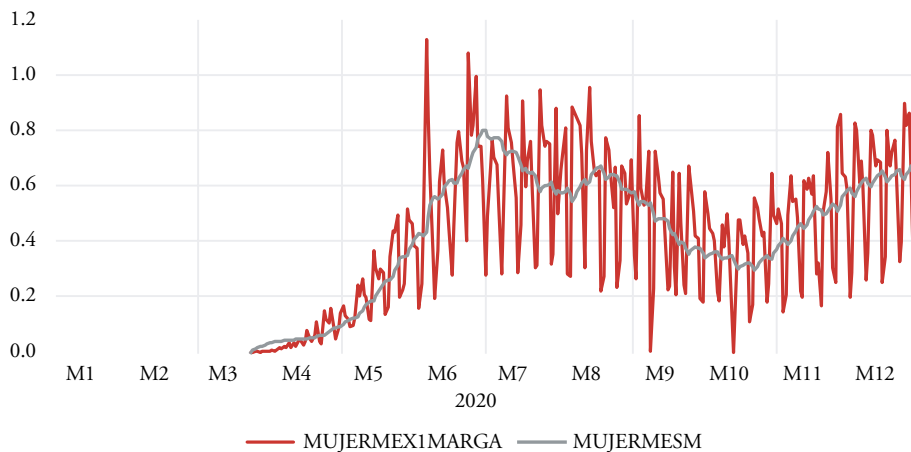


Rojo: Datos observados.

Gris: Suavización exponencial.

Fuente:Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinklr.com/>

Gráfica 2. Fallecimientos diarios por Covid-19 en México
Miles de fallecimientos diarios
Del 1 de marzo al 31 de diciembre de 2020



Rojo: Datos observados.

Gris: Suavización exponencial.

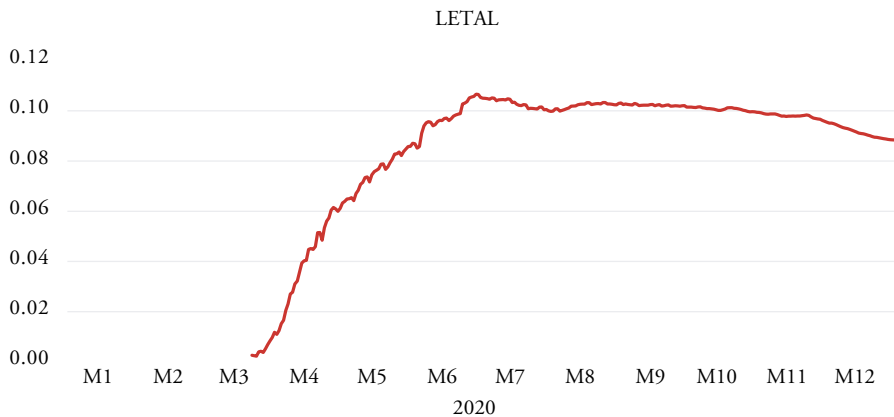
Fuente:Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinklr.com/>

Las Gráficas 1 y 2 muestran los contagios y muertes por Covid-19 en México al nivel diario. Las figuras en naranja son una suavización exponencial de los datos observados, los cuales se muestran en azul y, como claramente puede verse, tienen una enorme varianza, especialmente en el caso de los decesos. Un máximo local de los contagios diarios ocurrió en los primeros días de agosto, mientras que el de las muertes tuvo lugar a fines de junio.

Las gráficas muestran claramente que hubo un mínimo local de contagios y muertes diarias en octubre de 2020. A partir de entonces ambas variables han vuelto a aumentar. A fines de 2020, las muertes diarias apenas alcanzaron el máximo local que se observó en junio, pero los contagios superaron con creces el umbral al que llegaron en agosto.

La tasa de letalidad, que se define como la razón de muertes a contagios, puede observarse en la Gráfica 3.

Gráfica 3. Tasa de letalidad de Covid-19 en México
(Muertes totales/contagios totales)
Entre marzo y diciembre de 2020



Fuente:Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinkl.com/>

En la Gráfica 3 se observa que la mencionada tasa sube a gran velocidad entre marzo y junio de 2020. Después baja un poco y luego se estabiliza alrededor del 10% entre julio y octubre. A partir de ahí se reduce porque, como se observa en las gráficas anteriores, los contagios de la segunda ola de Covid-19 en México

suben con mayor rapidez que las muertes. Aun así, el último día de 2020 la tasa fue de 8.8%

Cabe señalar que el 10 de octubre las autoridades sanitarias de México hicieron un ajuste brusco hacia arriba tanto en la serie de contagios como en la de muertes. Esto propicia un problema cuando se utilizan en su forma pura, pues se rompe la continuidad de una serie de manera artificial.

Para paliar el problema descrito, con los datos oficiales hasta el 9 de octubre se hizo un pronóstico del día 10 con un modelo de VAR de 53 rezagos, el cual tiene una bondad de ajuste muy elevada. Luego se consideraron los datos oficiales entre el 10 de octubre y el 31 de diciembre y se obtuvieron los datos anteriores aplicando hacia atrás las tasas de crecimiento diarias de los datos oficiales hasta el 9 de octubre y el pronóstico del 10 de ese mismo mes. Esto le da una continuidad a la serie que en realidad es muy probable que exista.

Los datos oficiales muestran que en los primeros días de octubre había entre cuatro mil y cinco mil contagios diarios y entre 300 y 400 muertes diarias por Covid-19. El 10 de octubre se reportaron poco más de 30 mil contagios y tres mil muertes, pero en los días siguientes volvió a haber un número de contagios y muertes muy similares a los que ocurrieron antes del 10 de octubre.

Es claro que en el 10 de octubre hubo un ajuste por contagios y muertes que ocurrieron en días y tal vez en meses anteriores. Mantener exactamente los datos oficiales falsea la realidad y genera problemas de estimación; el primero de ellos la no normalidad de los residuales que surjan en las estimaciones.

3. Análisis de las series de infecciones y muertes acumuladas por Covid-19 en México

El objetivo de esta sección es analizar las series de contagios y muertes por Covid-19, ambas en términos acumulados, para determinar sus principales características y la probable relación que existe entre ellas.

3.1. Análisis de raíces unitarias de las series en cuestión

El primer ejercicio consiste en averiguar si las series en cuestión tienen o no raíces unitarias. Para ello, se llevan a cabo diversas pruebas, entre las cuales tal vez las más conocidas sean las de Dickey-Fuller y la de Phillips-Perron (Dickey y Fuller, 1979; Phillips y Perron, 1988).

Al aplicar estas pruebas, si resulta que las series son estacionarias, de modo que los contagios y las muertes acumuladas llegan a un máximo y ahí se estabilizan, es posible analizar su dinámica conjunta directamente sobre los logaritmos de las series en cuestión mediante la técnica de VAR. En caso de que las series resulten no estacionarias, pero estén cointegradas, la dinámica conjunta de ambas series se pueda analizar por medio de un VEC. Si las series son no estacionarias, pero no están cointegradas, el recurso para caracterizar su dinámica conjunta puede ser un VAR de sus cambios logarítmicos, siempre y cuando dichos cambios sean estacionarios.

La Tabla 1 muestra las pruebas de raíces unitarias para los logaritmos de las infecciones y las muertes acumuladas por Covid-19 en México y sus cambios logarítmicos. El análisis se lleva a cabo con datos diarios desde el 29 de febrero de 2020 al 31 de diciembre de ese mismo año.

En total se aplicaron siete pruebas.² En seis de ellas la hipótesis nula es que hay raíces unitarias. En la prueba KPSS (Kwiatkowski et al., 1992) la hipótesis nula es que la serie en cuestión es estacionaria. Solo tres rechazan la hipótesis nula de raíces unitarias tanto para el logaritmo de los contagios como el correspondiente a las muertes.³ En cambio, las seis pruebas cuya hipótesis nula es la de raíces unitarias rechazan dicha hipótesis en los cambios logarítmicos de ambas variables bajo estudio. La prueba KPSS rechaza la hipótesis de estacionariedad en todos los casos.

Aunque el análisis no es concluyente, pues la prueba KPSS rechaza incluso que los cambios logarítmicos de las variables en cuestión sean estacionarios, la mayoría sugiere que los logaritmos de las infecciones y las muertes son series no estacionarias integradas de orden uno (I(1)) y sus cambios logarítmicos constituyen series estacionarias (integradas de orden cero (I(0))).

² Para la prueba de Dickey-Fuller tradicional ver Dickey y Fuller (1979), para la prueba de Dickey-Fuller con cambio estructural ver Perron y Vogelsang (1992). Para las pruebas DFGLS y Elliot-Rothenberg-Stock de punto óptimo, ver Elliot et al. (1996). Para la prueba de Phillips-Perron, ver Phillips y Perron (1988), para la prueba KPSS ver Kwiatkowski et al. (1991), para la prueba de Ng-Perron ver Ng y Perron (2001).

³ Con más de 95% de confianza.

Tabla 1. Pruebas de raíces unitarias sobre el logaritmo y el cambio logarítmico de los contagios y las muertes por Covid-19 en México

(En términos acumulados)
Información diaria entre los primeros días de marzo y el 31 de diciembre de 2020

	<i>Logaritmo de los contagios acumulados</i>	<i>Logaritmo de las muertes acumuladas</i>	<i>Cambio logarítmico de los contagios acumulados</i>	<i>Cambio logarítmico de las muertes acumuladas</i>
ADF(x)	-3.6***	-4.0***	-3.1**	-11.5***
ADF(z) con cambio estructural	-10.3***	-23.9***	-12.6***	-12.2***
DFGLS	0.7	1.9	-3.4***	-11.5***
PP	-3.7***	-14.6***	-25.1***	-10.8***
KPSS	1.8 ^{vvv}	1.1 ^{vvv}	1.5 ^{vvv}	1.2 ^{vvv}
ERS	1507.9	6816.1	0.3***	0.5***
NgP	npru	npru	pru***	pru***

ADF(x): Prueba de Dickey-Fuller tradicional. $x = 3$ para el logaritmo de los contagios; $x = 9$ para el logaritmo de los fallecimientos; $x = 4$ para el cambio logarítmico de los contagios; $x = 8$ para el cambio logarítmico de los fallecimientos.

ADF(z): Prueba de Dickey-Fuller tradicional con cambio estructural. $z = 0$ en el logaritmo de los contagios; $z = 0$ para el logaritmo de los decesos; $z = 0$ para el cambio logarítmico de los contagios; $z = 0$ para el cambio logarítmico de los fallecimientos.

DFGLS: Prueba ajustada de Dickey-Fuller que elimina la tendencia de la serie.

PP: Prueba de Phillips-Perron.

KPSS: Prueba de Kwiatkowski-Phillips-Shmidt-Shin.

ERS: Prueba de Elliot-Rothenberg-Stock de punto óptimo.

Ng: Prueba de Ng-Perron.

* Puede rechazar la hipótesis nula de raíces unitarias con entre 90% y 95% de confianza.

**Puede rechazarse la hipótesis nula de raíces unitarias con entre 95% y 99% de confianza.

*** Puede rechazar la hipótesis nula de raíces unitarias con más de 99% de confianza.

vvv Puede rechazar la hipótesis nula de estacionariedad de la prueba KPSS con más de 99% de confianza.

npru: Puede rechazar raíces unitarias npru: no puede rechazar raíces unitarias con 90% o más de confianza.

Fuente:Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinklr.com/>

3.2. La relación de largo plazo entre los contagios y las muertes por Covid-19 en México

Si las series de los logaritmos de contagios y muertes fueran $I(1)$ habría que analizar si están cointegradas. La cointegración es un término introducido por Engle y Granger (1987). Muchas series son crecientes a lo largo del tiempo y, por lo

tanto, tienen una gran correlación entre ellas. Sin embargo, en muchas ocasiones no tienen ninguna relación causal. La cointegración implica que la combinación lineal de una o varias series no estacionarias genera un residual estacionario. En las que no están cointegradas eso no sucede.

La cointegración es una característica que se encuentra en un conjunto de series. *Per se* no significa que las series tengan una causalidad entre ellas. Sin embargo, normalmente es más probable que estén cointegradas series en las que hay alguna causalidad, a que series no estacionarias, las cuales no tienen relación alguna, generen combinaciones lineales estacionarias.

Hay diversas pruebas de cointegración disponibles. La prueba de Engle y Granger (1987) corre una regresión simple por mínimos cuadrados ordinarios (MCO) entre las variables no estacionarias en cuestión y luego analiza los residuales. Si estos son estacionarios, entonces pueden decirse que las series están cointegradas. Stock (1987) señala que hay un sesgo en este tipo de estimaciones. Sin embargo, Banerjee et al. (1986) muestran que el sesgo disminuye de manera considerable con una bondad de ajuste grande (una R^2 elevada). Cuthbertson et al. (1992 p. 135) indican que la regresión simple propuesta por Engle y Granger está libre de endogeneidad, pues los estimadores son súper consistentes.

La prueba de cointegración de Johansen (1988) es más amplia que la de Engle y Granger (1987), pues estima diversas formas en las cuales las variables están cointegradas, es decir, puede haber varias combinaciones lineales de variables no estacionarias que produzcan un residual estacionario. La estimación de esta técnica se lleva a cabo por el método de máxima verosimilitud (MV).

Otra prueba de cointegración de importancia es la que proponen Pesaran, Shin y Smith (2001) (de aquí en adelante PSS, 2001), los cuales utilizan una variante del modelo de corrección de error (MCE) (Engle y Granger, 1987) para estimar la relación de largo plazo entre dos o más variables. Cabe señalar que estos autores muestran que es posible encontrar relaciones de largo plazo no solo en variables no estacionarias, sino también en estacionarias, o en una combinación entre ambas, siempre que sean $I(0)$ o $I(1)$.

El método que proponen PSS (2001) es el de ARDL (autorregresivo de rezagos distribuidos, por sus siglas en inglés). En el caso que nos ocupa de infecciones y muertes acumuladas por Covid-19 en México, la manera de buscar una relación de largo plazo entre estas variables sería colocando las más endógena, que presumiblemente son las muertes, en función de rezagos de ella misma y el valor contemporáneo y rezagos de los contagios.

En este caso, si los contagios se definen como I y los fallecimientos como F , la regresión ARDL a estimar sería:

$$\log(F_t) = a_0 + \sum_{i=1}^{n_1} a_{1i} \log(F_{t-i}) + \sum_{i=1}^{n_2} a_{2i} \log(I_{t-i}) + e_t \quad (1)$$

La ecuación de largo plazo de los fallecimientos con respecto a los contagios es la siguiente:

$$\log(F) = \frac{a_0}{\left(1 - \sum_{i=1}^{n_1} a_{1i}\right)} + \frac{\sum_{i=1}^{n_2} a_{2i}}{\left(1 - \sum_{i=1}^{n_1} a_{1i}\right)} + \log(I) \quad (2)$$

Una condición necesaria para que el proceso sea estable, es decir que haya una relación de largo plazo entre las variables en cuestión, es: $\sum_{i=1}^{n_1} a_{1i} < 1$. pss (2001) transforman la ecuación (1) en un MCE para obtener una prueba F del proceso de largo plazo. La hipótesis nula, en este caso, es que no hay relación de largo plazo entre las variables, por lo cual si la prueba F está por arriba de los valores críticos, se rechaza dicha hipótesis, y en caso de que las variables sean $I(1)$ no puede rechazarse que haya cointegración entre ellas.

Varios programas de cómputo llevan a cabo estimaciones de ecuaciones ARDL de manera automática. Estas rutinas escogen los números de rezagos n_1 y n_2 para lograr un cierto objetivo, en muchos casos minimizar el criterio de Akaike o el de Schwartz.⁴ Sin embargo, en algunos programas econométricos, como el E-Views 11, el número máximo de rezagos está topado a 12. En estimaciones trimestrales o anuales ese es un número relativamente alto de rezagos; en estimaciones mensuales puede apenas ser suficiente, pero en una estimación diaria ese número máximo es muy bajo.⁵

⁴ Los criterios de Akaike y Schwartz hacen un balance entre la bondad de ajuste de la R^2 y el número de parámetros a estimar en una regresión. Dichos criterios premian la mayor R^2 y castigan el mayor número de parámetros a estimar. Estos criterios son mejores cuando su valor es el más pequeño posible (Akaike, 1974).

⁵ Una de las razones por las cuales el número máximo que pueden tomar n_1 y n_2 está topado es porque el programa tiene que generar una enorme cantidad de combinaciones de n_1 y n_2 para encontrar el mínimo criterio de Akaike. Si el número máximo de rezagos fuera mayor a 12, las combinaciones posibles crecen de forma exponencial y la memoria del programa no es suficiente para generarlas.

Otro problema que puede surgir con la estimación ARDL automática es el de endogeneidad. Como la o las variables independientes contemporáneas entran como regresores, aun cuando el número de rezagos sea suficiente para eliminar la autocorrelación de los errores no puede descartarse que haya correlación de esas variables independientes contemporáneas con el error, lo cual, al utilizar MCO como método de estimación, podría generar estimadores inconsistentes.

Por lo anterior, además de una regresión ARDL decidimos incluir una variante del MCE donde no entra la variable independiente contemporánea. En el caso que nos ocupa, la ecuación MCE a estimar sería:

$$d(\log(F_t)) = a_0 + \sum_{i=1}^n a_{1i} d(\log(F_{t-i})) + \sum_{i=1}^n a_{2i} d(\log(I_{t-i})) \quad (3) \\ + b_1 \log(F_{t-1}) + b_2 \log(I_{t-1}) + v_t$$

Como el objetivo es utilizar, al menos en algunos casos, muchos más rezagos que 12, el número de rezagos de los cambios logarítmicos de las muertes y los contagios es el mismo. En este caso se podría jugar con este número fijo para ambas variables a manera de minimizar el criterio de Akaike o el de Schwartz.

Si los logaritmos de las variables en cuestión son $I(0)$ o $I(1)$, la elasticidad de largo plazo de las muertes con respecto a los contagios es:

$$\frac{d(\log(F))}{d(\log(I))} = -\frac{b_2}{b_1} \quad (4)$$

La ecuación (3) se puede estimar por MCO sin generar estimadores inconsistentes cuando el número de rezagos n es tal que el error v es un ruido blanco, pues en este caso no hay variables independientes contemporáneas. Además, puede aplicarse una variante de la prueba F de PSS (2001) para la hipótesis nula en la cual $b_1 = b_2 = 0$.

El único problema que surge con esta estimación es que quizá no sea posible recuperar la constante de largo plazo de la relación entre el logaritmo de las muertes y el logaritmo de los contagios, pues esta está mezclada con la constante a_0 . No obstante, si las muertes y los contagios son variables $I(0)$ o $I(1)$, se presume que a_0 es el valor negativo del MCE, en cuyo caso la ecuación de largo plazo de las muertes en relación con los contagios sería:

$$\log(F) = -a_0 - \frac{b_2}{b_1} \log(I) \quad (5)$$

Tabla 2. Estimación de las ecuaciones de largo plazo entre el logaritmo de las muertes acumuladas y el logaritmo de los contagios acumulados

Variable dependiente: logaritmo de las muertes acumuladas
Estadístico t en paréntesis

<i>Métodos</i>	<i>Constante</i>	<i>Elasticidad de las muertes a los contagios</i>	<i>Prueba F de Pesaran, Shin y Smith en pruebas tipo ARDL o MCE</i>
Engle y Granger	-6.3 (-43.4 ^{***})	1.31 (111.8 ^{***})	-
Johansen	0.8 (1.3)	0.77 (17.9 ^{***})	-
ARDL	0.21 (0.3)	0.82 (18.8 ^{***})	13.7 ^{***}
MCE1	2.0 (1.8 [*])	0.69 (8.5 ^{***})	85.8 ^{***}
MCE2	0.51 (0.8)	0.79 (14.5 ^{***})	10.8 ^{***}

Los estadísticos t de la prueba de Engle y Granger pueden no ser fiables.

La prueba de Johansen se estimó con 37 rezagos de las variables correspondientes.

La prueba ARDL se llevó a cabo con la subrutina de E-Views 11 con un máximo de 12 rezagos y resultó un ARDL (12,12), 12 rezagos para el logaritmo de las muertes y 12 rezagos para el logaritmo de los contagios, lo que minimizó el criterio de Akaike.

La prueba MCE1 considera 1 rezago de las diferencias de los logaritmos de las variables involucradas.

La prueba MCE2 considera 40 rezagos de las diferencias de los logaritmos de las variables involucradas.

* Significativo al 90% de confianza.

** Significativo al 95% de confianza.

*** Significativo al 99% de confianza.

Fuente: Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinklr.com/>

La Tabla 2 muestra diversas estimaciones de la relación de largo plazo entre el logaritmo de las muertes y el correspondiente a los contagios en México.

Todas las pruebas muestran la existencia de una relación de largo plazo positiva y significativa entre el logaritmo de las muertes y el de los contagios. El resultado de las pruebas de Engle y Granger (1987) y de Johansen (1988) implicaría que estas series están cointegradas. No obstante, cuando se analizan los residuales de la ecuación de Engle y Granger (1987), solo la prueba ADF(7) y la PP pueden rechazar raíces unitarias. En todas las demás no se puede rechazar esa hipótesis.

La prueba de Johansen (1988) encuentra un solo vector de cointegración, que es el que se muestra en la Tabla 2.

La prueba MCE1 se corrió solo con un rezago de las diferencias de las variables dependiente e independiente. En contraste, la prueba MCE2 buscó minimizar el criterio de Akaike, lo cual se logró con 40 rezagos de las diferencias de cada logaritmo. En las pruebas ARDL y MCE1 y 2 el estadístico F de PSS (2001) resultó altamente significativo, rechazando en forma contundente la hipótesis nula de no existencia de relación de largo plazo entre las variables, lo que implica la presunción de existencia de una relación positiva entre ambas variables.

Excepto por la prueba de Engle y Granger (1987), las demás muestran una elasticidad estimada de muertes a contagios menor pero no muy lejana a la unidad. Esto querría decir que conforme aumenta la cantidad de contagios la tasa de letalidad se reduce, pero poco.

4. Estimación de un modelo VEC para el análisis de la dinámica de contagios y muertes por Covid-19 en México

Los resultados de la sección anterior sugieren la conveniencia de llevar a cabo un análisis más detallado de relación entre muertes y contagios por Covid-19 en México, que compatibilice la dinámica de corto plazo con la relación existente de largo plazo entre las variables en cuestión. También conviene hacer un análisis bivariado más que univariado, como es el caso de las relaciones vistas en la sección anterior, porque, aunque parece haber una clara relación positiva de los contagios a las muertes, no puede descartarse a priori que haya una doble causalidad donde las muertes también impacten a los contagios.⁶

Por lo anterior, proponemos utilizar la metodología de un VEC, el cual analiza la dinámica de un sistema conformado por las muertes y los contagios de Covid-19, pero tomando en cuenta la relación de largo plazo entre estas variables.

En el caso que nos ocupa, la mayoría de las pruebas de raíces unitarias mostraron que las series de los logaritmos de las muertes y los contagios no son estacionarias. Las pruebas de cointegración en general también muestran que hay una relación de largo plazo entre las variables que nos conciernen. Utilizar un VEC

⁶ Aunque después se explicará esto de nuevo, las muertes podrían incrementar los contagios futuros cuando los familiares y amigos de los difuntos se juntan para hacer duelo por la persona fallecida o salen a arreglar trámites de pagos de primas de seguros y otros asuntos jurídicos. Por otra parte, al menos en teoría las muertes podrían reducir los contagios futuros cuando el incremento de dichas infecciones propicia que otras personas se cuiden más.

tiene la ventaja de que también esta metodología hace una prueba de cointegración de Johansen. Si se llegara a rechazar en esta prueba la cointegración, el VEC se transforma en un VAR. Esto quiere decir que los modelos VEC contienen a los VAR, pero no al revés. Si existe alguna duda sobre si hay o no cointegración entre las variables en cuestión, o si estas son o no estacionarias, lo mejor es utilizar un VEC.

Los modelos VEC surgen con posterioridad a los modelos de vectores autorregresivos VAR (Sims, 1980). Un modelo VAR muestra un sistema de dos o más variables, donde los valores contemporáneos de cada una de estas variables dependen de rezagos de ellas mismas y de las demás variables. Estos modelos son idóneos para variables estacionarias, pero pueden ser inviables para analizar variables no estacionarias (ver por ejemplo Sims et al., 1990).

Un modelo VEC (Johansen, 1995; Hendry y Juselius, 2000; PSS, 2000; Lutkepohl, 2005a, 2005b; Kilian y Lutkepohl, 2017) es un sistema de ecuaciones en el cual las diferencias de distintas variables se relacionan con los rezagos de ellas mismas y de las demás variables, al tiempo que en cada ecuación se impone una relación de cointegración entre las variables, la cual se estima por el método de Johansen (1988, 1995).

En el caso que nos ocupa de muertes y contagios por Covid-19 en México, el VEC correspondiente se formula como:

$$d(\log(F_t)) = a_0(\log(F_{t-1}) + \gamma_1 \log(I_{t-1}) + \gamma_0) + \sum_{i=1}^n a_{1i} d(\log(F_{t-i})) + \sum_{i=1}^n a_{2i} d(\log(I_{t-i})) + \epsilon_t \quad (6)$$

$$d(\log(I_t)) = b_0(\log(F_{t-1}) + \gamma_1 \log(I_{t-1}) + \gamma_0) + \sum_{i=1}^n b_{1i} d(\log(F_{t-i})) + \sum_{i=1}^n b_{2i} d(\log(I_{t-i})) + v_t \quad (7)$$

Donde el término $(\log(F_{t-1}) + \gamma_1 \log(I_{t-1}) + \gamma_0)$ muestra la relación de cointegración en el largo plazo entre el logaritmo de los contagios acumulados (I) y el logaritmo de las muertes (F) en los mismos términos de Covid-19 en México. Los términos ϵ y v son variables aleatorias que idealmente se distribuyen de manera normal con media cero.

Para comprender cómo funciona un VEC conviene expresar las ecuaciones (6) y (7) en notación matricial:

$$\begin{pmatrix} d(\log(F_t)) \\ d(\log(I_t)) \end{pmatrix} = \begin{pmatrix} a_0 & a_0\gamma_1 \\ b_0 & b_0\gamma_1 \end{pmatrix} \begin{pmatrix} \log(F_{t-1}) \\ \log(I_{t-1}) \end{pmatrix} \quad (8)$$

$$+ \begin{pmatrix} a_0\gamma_0 + \sum_{i=1}^n a_{1i}d(\log(F_{t-i})) + \sum_{i=1}^n a_{2i}d(\log(I_{t-i})) + \epsilon_t \\ b_0\gamma_0 + \sum_{i=1}^n b_{1i}d(\log(F_{t-i})) + \sum_{i=1}^n b_{2i}d(\log(I_{t-i})) + \nu_t \end{pmatrix}$$

El determinante de la matriz característica $\begin{pmatrix} a_0 & a_0\gamma_1 \\ b_0 & b_0\gamma_1 \end{pmatrix}$ es: $a_0b_0\gamma_1 - a_0b_0\gamma_1 = 0$.

Si las series I y F son integradas de orden 1, las series $d(\log(I))$ y $d(\log(F))$ serán integradas de orden cero y tendrán una media constante. Por lo anterior, en el plano donde los fallecimientos están en el eje vertical y los contagios en el eje horizontal, las funciones $d(\log(I)) = 0$ y $d(\log(F)) = 0$ serán la misma función, o serán dos funciones que tienen exactamente la misma pendiente.

En el caso que nos ocupa, si partimos del punto donde las dos funciones de largo plazo constituyen la misma función, cuando las muertes y los contagios están sobre dicha función $d(\log(I)) = 0 = d(\log(F)) = 0$, el sistema estará en reposo sobre algún punto de tal relación de largo plazo. Si hay un choque permanente en alguna de las funciones, entonces el sistema entrará en un movimiento donde los contagios y las muertes podrían moverse sin alcanzar reposo, pero los cambios logarítmicos sí alcanzarán un valor estable.

En general, en este trabajo nos centraremos en los casos donde el valor de ajuste a_0 es negativo y el valor b_0 es positivo o cero en las ecuaciones (6) y (7), pues es cuando, en muchos casos, el sistema es globalmente estable y, si está fuera de la relación de largo plazo $\log(F_{t-1}) + \gamma_1 \log(I_{t-1}) + \gamma_0$, tiende a dicha relación.

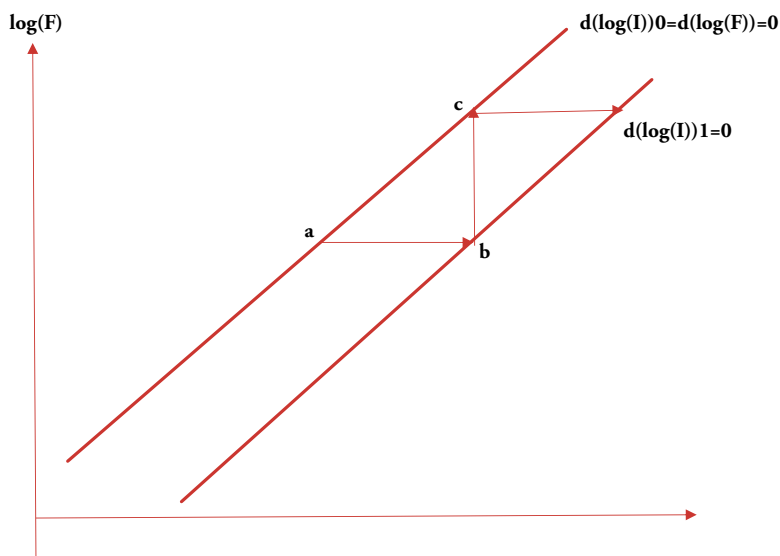
Para mostrar lo anterior, analizaremos de manera breve un sistema donde $\gamma_1 < 0$ y $a_{1i} = a_{2i} = b_{1i} = b_{2i} = 0$ para toda $i \geq 1$. Esto quiere decir que el modelo VEC no tiene rezagos de las diferencias de los logaritmos de las variables en cuestión.

En este caso, el sistema de ecuaciones en diferencias (8) se expresa como:

$$\begin{pmatrix} d(\log(F_t)) \\ d(\log(I_t)) \end{pmatrix} = \begin{pmatrix} a_0 & a_0\gamma_1 \\ b_0 & b_0\gamma_1 \end{pmatrix} \begin{pmatrix} \log(F_{t-1}) \\ \log(I_{t-1}) \end{pmatrix} + \begin{pmatrix} a_0\gamma_0 + \epsilon_t \\ b_0\gamma_0 + \nu_t \end{pmatrix} \quad (9)$$

Suponiendo γ_1 menor a cero, el diagrama de fase de este sistema puede observarse en la gráfica 4.

Gráfica 4. Diagrama de fase entre los logaritmos de las muertes y contagios por Covid-19



Fuente: Elaboración del autor.

En una situación hipotética inicial de equilibrio, las ecuaciones $d(\log(I))_0 = 0$ y $d(\log(F))_0 = 0$ son la misma ecuación y el sistema está en un punto de reposo, por ejemplo, el punto “a” en la Gráfica 3. En ese punto los contagios y las muertes no se modifican y la relación de largo plazo entre ellas es:

$$\log(F) = -\gamma_0 - \gamma_1 \log(I) \quad (10)$$

Como γ_1 es negativo, esta relación es positiva.

Supongamos que hay un choque permanente en la ecuación de contagios, de modo que se mueve a la derecha (por ejemplo, porque el parámetro v en el sistema (8) se convierte en un factor determinístico que aumenta en forma permanente). En ese caso, los contagios aumentan en el período actual, lo que se observa gráficamente como un desplazamiento de la función $d(\log(I))_0$ a $d(\log(I))_1$ y la flecha que va del punto a al punto b. Sin embargo, ese aumento actual de contagios propicia un aumento de las muertes el siguiente período, pues el parámetro $a_0 \gamma_1$

en el sistema (8) es positivo, ya que ambos a_0 y γ_1 son negativos. Este aumento de muertes al siguiente período se representa con la flecha que va del punto b al punto c en la gráfica 4. A su vez, el incremento de las muertes en el siguiente período aumenta los contagios del que sigue, pues el parámetro b_0 es positivo, lo que se observa en la gráfica con la flecha horizontal que parte del punto c. Este proceso teórico no tiene fin. Las dos variables continuarían aumentando en forma indefinida.

Si el choque mencionado fuera temporal, la ecuación $d(\log(F))_1 = 0$ regresa a donde estaba antes y el sistema converge a un nuevo punto de reposo sobre la recta original, pero con más muertes e infecciones acumuladas (un punto al noreste del punto a sobre esa misma recta).

En el caso en el cual a_0 es negativo y b_0 es cero, un choque positivo permanente en los contagios genera un movimiento similar al descrito en el caso anterior, pero en este no habría un incremento sin fin de las variables en cuestión, sino que el sistema volvería a un punto de reposo sobre la función $d(\log(F)) = 0$ aun cuando la función $d(\log(I)) = 0$ no regresara a su lugar original. En este caso la trayectoria sería la que va solo del punto a al b y de ahí al c. En el punto c terminaría el proceso dinámico, con más muertes y contagios que en el punto original, pero nuevamente en reposo.

Claramente, aun cuando sea posible caracterizar la relación de cointegración de largo plazo (10), los valores de a_0 y b_0 son cruciales porque puede ser que, aunque exista esa relación, el sistema no converja a la misma.

Es necesario enfatizar que la inclusión de rezagos de las diferencias de los logaritmos de las variables en cuestión puede modificar las dinámicas que hemos ejemplificado. Pueden darse casos en que se generen oscilaciones convergentes o divergentes o dinámicas diferentes a las que mostramos en el ejemplo más sencillo donde no hay rezagos.⁷

Se llevaron a cabo estimaciones del VEC conformado por las ecuaciones (6) y (7). Normalmente la estimación que se reporta es aquella en la cual el número de rezagos de las diferencias de las variables en cuestión minimiza el criterio de Akaike o de Schwartz. Sin embargo, en este caso el criterio de Akaike baja

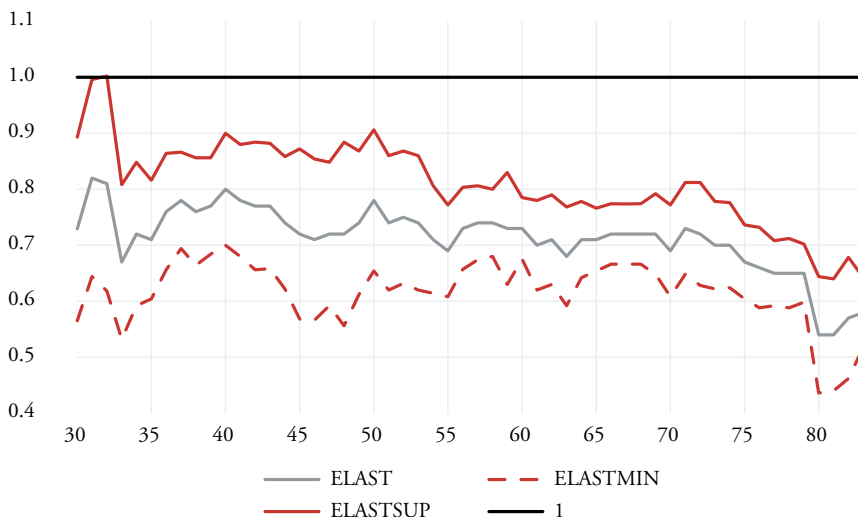
⁷ Un sistema de ecuaciones diferenciales similar a (8) sería estable si λ_1 y a_0 son menores a cero y b_0 es positivo, algo muy similar a lo que sucede en el sistema de ecuaciones en diferencias cuando no hay rezagos de las diferencias de los logaritmos de las variables en cuestión (Chiang, 1984; Takayama, 1985).

de manera monótonica al incluir más rezagos. El criterio de Schwartz se minimiza con 37 rezagos de las diferencias de las variables bajo análisis.

El concepto de la elasticidad de muertes a contagios por Covid-19 es relevante, porque indica de qué forma se altera la tasa de letalidad de la enfermedad cuando aumenta el número de contagios. Eso es muy útil para hacer cálculos aproximados de cuántas personas van a morir si se infecta un cierto porcentaje de la población. Por otra parte, el modelo VEC con 37 rezagos da resultados muy sensatos en términos de las funciones de impulso-respuesta, pero no tiene una bondad de ajuste dinámica demasiado elevada, pues termina sobreestimando, de manera considerable, tanto los contagios como las muertes en los últimos meses del año en curso.

Por las razones anteriores, se llevan a cabo varios ejercicios. El primero consiste en describir en una gráfica un resumen de estimaciones del modelo VEC (sistema 8) con entre 30 y 83 rezagos, siendo este último donde los ejercicios de causalidad de Granger (1969) muestran que cuando menos los contagios causan las muertes. Con más rezagos los ejercicios de causalidad de Granger no son significativos. No se muestran las estimaciones del modelo VEC con rezagos menores a 30, porque en general presentan alta autocorrelación de los residuales.

Gráfica 5. Elasticidad de largo plazo de muertes a contagios por Covid-19 en México en relación con los retrasos de diferentes estimaciones VEC



Fuente:Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinklr.com/>

La Gráfica 5 muestra en azul la elasticidad puntual estimada de largo plazo de muertes a contagios por Covid-19 en México conforme los retrasos de la estimación del VEC suben de 30 a 83. El promedio de las estimaciones es 0.71. La línea en verde es el valor máximo del intervalo de confianza de dicha elasticidad al 95%; la línea naranja es el valor mínimo de dicho intervalo.

En todos los casos los valores dentro de los intervalos de confianza están arriba de cero. Asimismo, excepto en un caso (cuando los retrasos son 32), el valor máximo del intervalo de confianza está debajo de la unidad. De modo que, de acuerdo con estas estimaciones, es posible afirmar con bastante confianza que la elasticidad de muertes a contagios por Covid-19 en México es inferior a uno. Un aumento en los contagios reduce la tasa de letalidad, aunque no en una gran proporción.

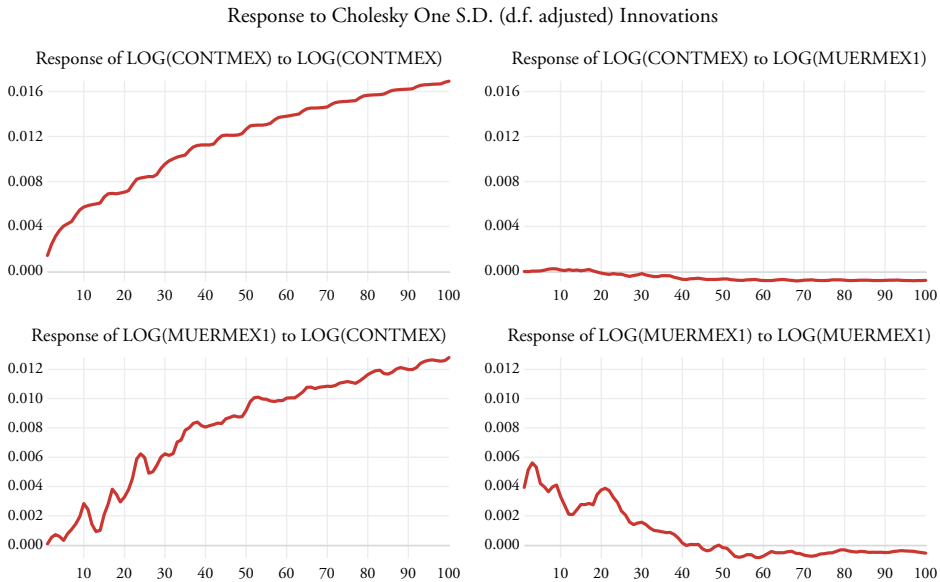
Un segundo ejercicio consiste en mostrar las funciones de impulso-respuesta del VEC con 37 rezagos, que es el que minimiza el criterio de Schwartz.

La ecuación de largo plazo del VEC con 37 rezagos es:

$$\log(F_t) = 0.48 + 0.78 \log(I_t) \tag{11}$$

Las funciones de impulso-respuesta en este caso son las siguientes.

Gráfica 6. Funciones de impulso-respuesta de infecciones y muertes acumuladas por Covid-19 en México con un modelo vec de 37 rezagos



Fuente:Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinklr.com/>

Los valores calculados para los parámetros de estabilidad del VEC a_0 y b_0 de las ecuaciones (6) y (7), o del sistema (8), son negativos, pero solo el estimador del parámetro a_0 es significativo. El parámetro b_0 , aunque puntualmente negativo, es muy pequeño.

Las funciones de impulso-respuesta a 100 días muestran resultados muy sensatos: un choque temporal positivo de contagios (contmex en la gráfica) afecta positivamente tanto a los contagios como a las muertes (muermex1) por algún tiempo y luego se estabiliza. Un choque de muertes no tiene un efecto significativo en los contagios (porque el parámetro b_0 es muy cercano a cero) y el efecto sobre las mismas muertes se va diluyendo.

El ejercicio de impulso-respuesta que se presenta aquí sugiere que los contagios afectan positivamente a las muertes, pero las muertes no afectan a los contagios. Llevamos a cabo pruebas de causalidad de Granger en las cuales utilizamos los mismos rezagos que se muestran en las estimaciones tipo VEC (entre 30 y 83 rezagos). En todos los casos los contagios causan las muertes con más de 95% de confianza, pero las muertes causan los contagios con ese porcentaje de confiabilidad solo cuando se utilizan entre 30 y 47 rezagos.

Siguiendo con lo anterior, los contagios causan las muertes con más de 99% de confianza cuando se usan entre 30 y 80 rezagos, pero las muertes causan los contagios con ese nivel de confiabilidad solo cuando se utilizan entre 30 y 43 rezagos.

La causalidad de los contagios a las muertes es muy lógica; menos claro es por qué las muertes podrían causar más o menos contagios. Una posibilidad de que los fallecimientos pudieran generar más contagios es que cuando alguien fallece de Covid-19, los familiares y las personas cercanas van a las agencias funerarias o a las compañías de seguros para arreglar los asuntos jurídicos que surgen de la muerte de esa persona, lo cual hace que se expongan a contagios que eventualmente causan más fallecimientos.

En cuanto a por qué las muertes podrían, tal vez en algunos casos, causar menos contagios, está el hecho de que conforme las estadísticas de fallecimientos por la enfermedad en cuestión aumentan, muchas personas se cuidan más y dejan de salir de casa, lo que podría redundar en menos contagios y eventualmente en menos muertes.

De todas las estimaciones tipo VEC que se llevaron a cabo, el parámetro b_0 es a veces positivo, a veces negativo, pero siempre es muy pequeño, y en todos los casos no es significativo estadísticamente, mientras que el parámetro a_0 es negativo y significativo en todas las estimaciones. Si el modelo VEC no tuviera rezagos, eso haría que un choque positivo en las muertes no tuviera un efecto en los contagios, mientras un choque positivo en los contagios sí tendría un efecto positivo significativo en las muertes, como ya se describió con anterioridad. También como ya se mencionó, las dinámicas

descritas pueden alterarse por la presencia de los rezagos de las diferencias de los logaritmos de las variables en cuestión.

En las gráficas de impulso-respuesta de todos los modelos VEC estimados (de las cuales solo se está mostrando la del VEC con 37 rezagos), se observa que en todos los casos un choque positivo de los contagios tiene un efecto considerable sobre las muertes, mientras que los choques de las muertes tienen efectos generalmente pequeños en los contagios, excepto cuando se utilizan entre 65 y 76 rezagos; en esos casos las muertes tienen un efecto positivo relativamente elevado en los contagios futuros.

5. Simulación dinámica de dos modelos VEC y proyecciones de contagios y muertes por Covid-19 en los diferentes modelos estimados en este trabajo

Esta sección muestra, primero, una simulación dinámica del modelo VEC con 37 rezagos (VEC37) y de un VEC con 68 rezagos (VEC68) el cual, aun teniendo más de 150 grados de libertad, es uno de los modelos estimados que muestran un alto nivel de bondad de ajuste dinámico.

5.1. Simulación dinámica de dos modelos VEC

Una simulación dinámica consiste en estimar los parámetros de un modelo con toda la información disponible y después simularlo hasta la fecha donde se llevó a cabo la estimación, pero para realizar dicha simulación solo se utiliza la mínima información necesaria. En una simulación estática se estiman los parámetros de igual forma, y cada período (en este caso cada día) se simula con los datos realmente observados en él o en anteriores (días), por lo cual al ajustar los datos de cada período (día), normalmente, las simulaciones estáticas tienen mucho mejor bondad de ajuste que las dinámicas.⁸

Las Gráficas 7 y 8 muestran los datos oficiales de las infecciones y muertes acumuladas por Covid-19, así como las simulaciones dinámicas de los modelos VEC con 37 y 68 rezagos, respectivamente. Por su parte, las Gráficas 9 y 10 muestran estas mismas figuras, pero en el nivel diario.

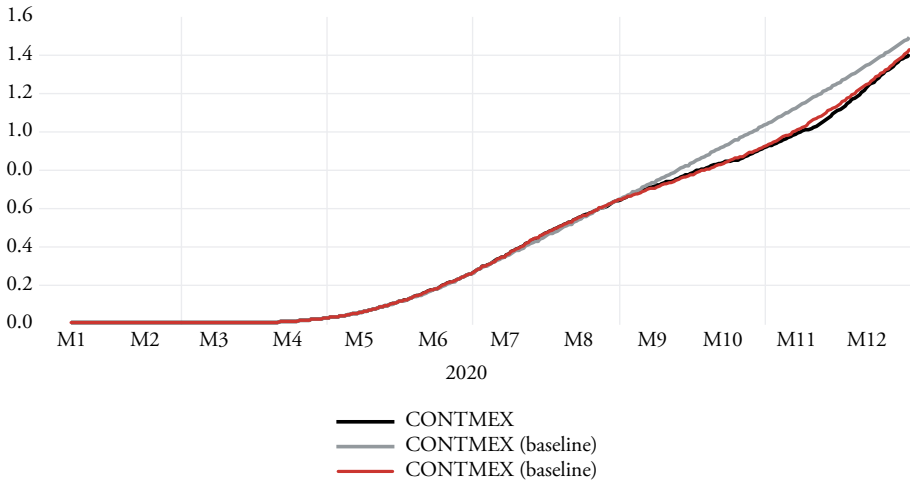
⁸ Por ejemplo, supongamos el modelo $Y_t = 0.5 + 0.2Y_{t-1}$. Si Y es por ejemplo el PIB, entonces si los parámetros 0.5 y 0.2 se obtuvieron con una regresión entre 1990 y 2020, la simulación dinámica consiste en simular la ecuación utilizando únicamente el dato de Y_{t-1} para 1989 y todos los demás de 1990 a 2020 se generan dentro de la simulación. En cambio, una simulación estática consiste en encontrar el PIB cada año poniendo en Y_{t-1} el que realmente se observó el año anterior.

La expansión de Covid-19 en México en 2020: un enfoque desde la econometría de series de tiempo

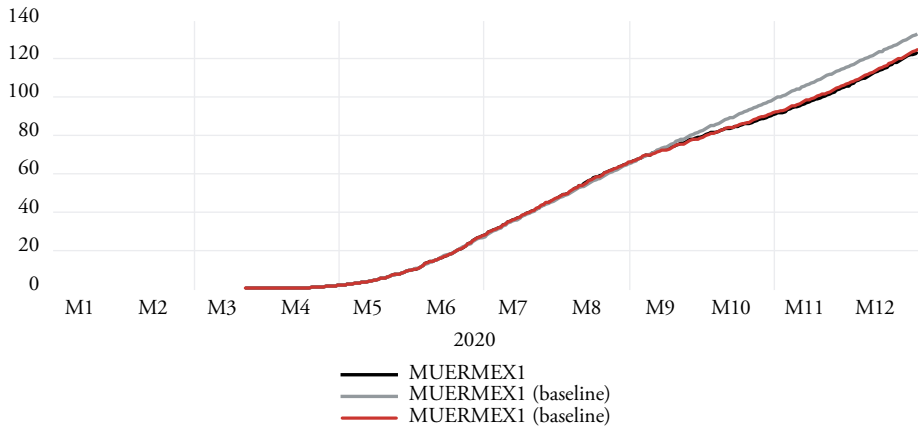
Gráficas 7 y 8. Simulación dinámica de las infecciones y muertes acumuladas por Covid-19 en México

Entre el 4 de mayo y el 31 de diciembre para el VEC de 37 rezagos
y entre el 29 de mayo y el 31 de diciembre para el VEC de 68 rezagos

Millones de contagios



Miles de muertes



Negro: Datos observados.

Gris: Simulación con el vec de 37 rezagos. Correlación de los contagios con los datos observados 0.9998. Correlación de las muertes con los datos observados: 0.9998.

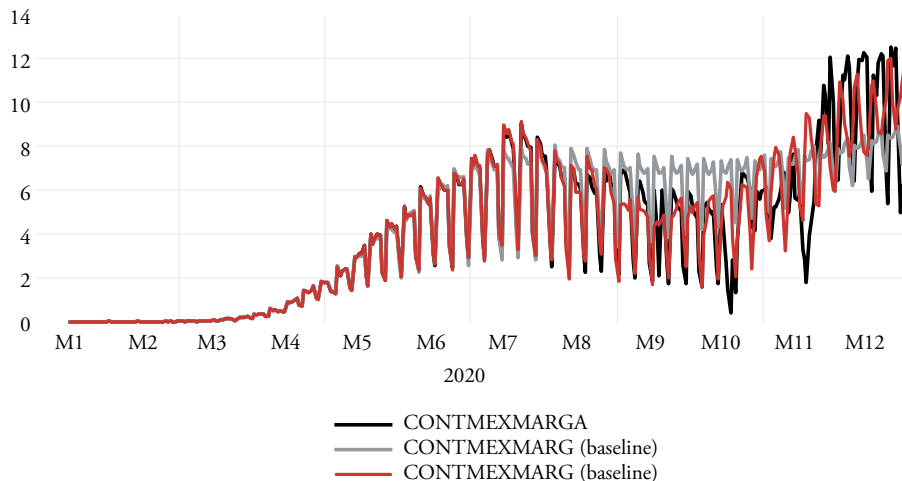
Rojo: Simulación con el vec de 68 rezagos. Correlación de los contagios con los datos observados 0.9971. Correlación de las muertes con los datos observados: 0.9999.

Fuente: Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es>. La dirección donde directamente aparecen los datos es <https://who.sprinklr.com/>

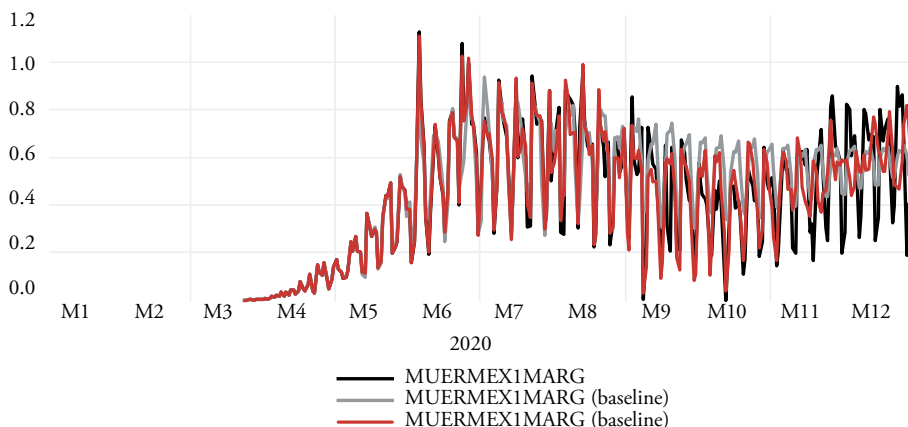
Gráficas 9 y 10. Simulación dinámica de las infecciones y muertes diarias por Covid-19 en México

Entre el 4 de mayo y el 31 de diciembre para el vec de 37 rezagos y entre el 29 de mayo y el 31 de diciembre para el vec de 68 rezagos

Millones de contagios diarios



Miles de muertes diarias



Negro: Datos observados.

Gris: Simulación con el vec de 37 rezagos. Correlación de los contagios con los datos observados 0.893. Correlación de las muertes con los datos observados: 0.868

Rojo: Simulación con el vec de 68 rezagos. Correlación de los contagios con los datos observados 0.9081. Correlación de las muertes con los datos observados: 0.896.

Fuente: Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinklr.com/>

Para generar la simulación dinámica del VEC37 se utilizaron los datos observados hasta el 4 de mayo, y todos los demás resultados de muertes y contagios entre el 5 de mayo y el 31 de diciembre surgen de la simulación. En el caso del VEC68, se emplearon los datos observados hasta el 29 de mayo, y los resultados de muertes y contagios hasta fin de 2020 también se generan exclusivamente en la simulación. En cambio, si hubieran sido simulaciones estáticas, los resultados de muertes y contagios de cada día se obtendrían utilizando todos los datos observados hasta el día anterior.

La bondad de ajuste del modelo VEC con 68 rezagos es mayor que la del VEC de 37 rezagos, que es el que minimiza el criterio de Schwartz.

Los dos modelos VEC sobrestiman los contagios y las muertes por Covid-19 en México al día 31 de diciembre, pero la sobrestimación del VEC37 es muy superior en los dos casos (1.6% para el VEC68 y 6.9% para el VEC37 en los contagios; 0.8% para el VEC68 y 7.3% para el VEC35 en las muertes).

La sobrestimación del VEC37 es mayor porque este modelo dinámico no puede reproducir la caída de las muertes y los contagios diarios que tuvo lugar aproximadamente entre julio y octubre de este año (Gráficas 9 y 10, 1 y 2), mientras que el VEC68 reproduce mucho mejor dicha caída.

En la estimación de modelos de este tipo tal vez sea conveniente llevar a cabo muchas variantes para tener resultados menos sesgados. En caso de que se hubiera escogido el VEC37 solo porque minimiza el criterio de Schwartz, la reproducción de la dinámica observada de las variables de interés de este artículo no hubiera sido la mejor que se pudiera encontrar.

5.2. Proyecciones de contagios y muertes a largo plazo con diferentes modelos estimados en el artículo

Esta misma sección muestra una proyección muy simple de muertes y contagios por Covid-19 en México con los valores observados hasta el 31 de diciembre, así como con los calculados por las ecuaciones de largo plazo de varios de los modelos aquí estimados.

La relación de largo plazo entre muertes y contagios por Covid-19 se calculó con varios modelos. En esta sección se incluyen la ecuación de Johansen, la del modelo ARDL con 12 rezagos y la que surge del modelo MCE2, estimaciones se llevaron a cabo en la segunda sección del trabajo. Se incluyen también las relaciones de largo plazo que surgen de los modelos VEC37 y VEC68 que se estimaron en la tercera sección.

En este caso se excluyen las ecuaciones de largo plazo de Engle y Granger y la que surge del modelo MCE1 de la segunda sección. La primera porque es poco contundente que el método de Engle y Granger muestre cointegración, y la segunda porque el modelo MCE2 calculado en la segunda sección engloba y tiene mejor bondad de ajuste que el MCE1.

Las ecuaciones de largo plazo de todos estos modelos que se incluyen son las siguientes:

$$\log(F) = 0.8 + 0.77 \log(I) \quad (11) \text{ Johansen}$$

$$\log(F) = 0.5 + 0.78 \log(I) \quad (12) \text{ VEC37}$$

$$\log(F) = 1.4 + 0.72 \log(I) \quad (13) \text{ VEC68}$$

$$\log(F) = 0.2 + 0.82 \log(I) \quad (14) \text{ ARDL}(12,12)^9$$

$$\log(F) = 0.5 + 0.79 \log(I) \quad (15) \text{ MCE2}$$

La Gráfica 11 muestra una relación de contagios en el eje horizontal y muertes en el eje vertical que surge de todos estos modelos.

Los valores observados al 31 de diciembre coinciden bastante con las proyecciones de la ecuación de Johansen y la del MCE2, y quedan por arriba de las proyecciones de los dos modelos VEC y por debajo de las proyecciones del modelo ARDL.

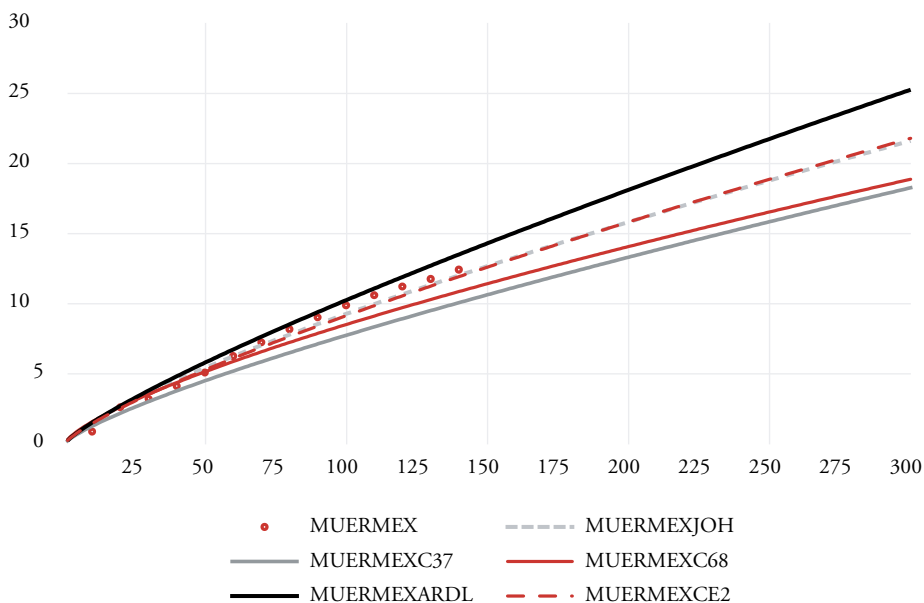
Si el número de personas infectadas llegara a tres millones, el modelo VEC37 pronostica 182 mil muertes, el VEC37, 188 mil, Johansen, 216 mil, MCE2, 217 mil y el ARDL 253 mil. Con 1.4 millones de contagios registrados ocurrieron 123 mil muertes también registradas.

El número de tres millones de infectados tiene importancia porque es posible que el subregistro de contagios sea cerca de 40 veces. La encuesta preliminar Ensanut de 2020 del Instituto Nacional de Salud Pública de México (INSP) obtuvo, entre agosto y noviembre de ese año, muestras de sangre en las cuales se observó que alrededor de 25% de los encuestados tenía anticuerpos contra Covid-19 (Instituto Nacional de Salud Pública de México, 2020). Si esta es representativa de toda la población, implicaría que en esos meses ya se hubieran infectado de esta

⁹ El programa automático de ARDL en E-Views 11 con hasta 12 rezagos escogió 12 rezagos para los logaritmos de las dos variables bajo análisis.

**Gráfica 11. Proyección de contagios y muertes por Covid-19 en México
(millones/10 000)**

Contagios en el eje horizontal, muertes en el eje vertical



Puntos rojos: Valores observados hasta el 31 de diciembre de 2020.

Punteada gris: Ecuación de Johansen.

Gris: VEC37.

Rojo: VEC68.

Negro: Modelo ARDL.

Punteada roja: MCE2.

Fuente: Elaborada por el autor con datos de la página de Internet de la Organización Mundial de la Salud <https://www.who.int/es> La dirección donde directamente aparecen los datos es <https://who.sprinkl.com/>

enfermedad más de 30 millones de personas en México (de hecho, la cifra mencionada es 32 millones). No obstante, el promedio de infecciones registradas en esos meses fue de alrededor de 770 mil, un factor de diferencia de más de 41.6 veces.

Lo anterior implicaría que si hubiera entre dos y tres millones de contagios registrados, y continuara el mismo factor de subregistro de contagios, habría realmente entre 80 millones y 120 millones de personas contagiadas. Es decir, entre dos terceras partes de la población y más de 95% de la población total del país.

Claramente, el número de muertos también tiene un subregistro, que algunos autores calculan entre 3 y 3.5 veces (Kánter, 2020; Romero y Despeghe, 2020),

pero que tal vez sea un poco menor. La Secretaría de Salud calculó para el 12 de diciembre un exceso de mortandad en 2020 de poco más de 274 mil personas (Secretaría de Salud, 2020). Ese día la cifra oficial de muertos por Covid-19 fue de entre 108 mil y 109 mil personas, lo que implica que podría haber hasta 165 mil o 166 mil personas adicionales que hayan fallecido por este mal. Esto implicaría un factor de subregistro de hasta 2.5 veces. Sin embargo, es posible que este factor sea menor porque no todas las muertes adicionales fueron necesariamente causadas por Covid-19. Un cierto número de ellas ha tenido lugar porque los espacios hospitalarios para atender pacientes con otras enfermedades se han reducido de manera considerable, especialmente en los últimos meses.

Con un factor de 40 como subregistro de contagios y un factor de 2 como subregistro de muertes, tres millones de infectados oficiales implicarían a una gran parte de la población contagiada y entre 376 mil y poco más de 500 mil muertes por Covid-19 en México. El problema es que esos números son demasiado grandes.

En la explicación de cómo funciona un modelo VEC pusimos un ejemplo en el cual si los contagios generan más muertes y las muertes más contagios, eventualmente estas variables crecen sin parar. También ahí se muestra que un cambio en los contagios puede detener completamente esta dinámica perversa. Eso lo lograron China y otros países asiáticos desde hace un buen tiempo. Es necesario implementar medidas que detengan este crecimiento entre contagios y muertes en México, el cual, hasta ahora, ya ha generado demasiadas pérdidas humanas y materiales.

Los cálculos realizados en esta subsección deben tomarse como simples extrapolaciones que tienen alguna lógica, pero que pueden fallar, entre otras cosas porque los factores de subregistro de contagios y muertes por Covid-19 no necesariamente son constantes en el tiempo.

6. Conclusiones

El Covid-19 avanza en el mundo a grandes pasos. En México, entre marzo y septiembre de 2020 los contagios y las muertes por esta enfermedad siguieron el patrón que se conoce como de curva epidémica, en el cual estas variables, medidas diariamente, primero crecen, llegan a un máximo y luego decrecen. Sin embargo, ese patrón se rompió en octubre cuando se generó una segunda curva ascendente

en las variables bajo análisis, y para fin de 2020 no era claro que hubiera llegado a un nuevo máximo local.

El análisis de los contagios y las muertes sugiere que es probable que constituyan series no estacionarias integradas de orden 1 ($I(1)$) y que con alta probabilidad estén cointegradas, por lo cual se estimó un VEC para caracterizar la dinámica de las expansión de las variables bajo análisis.

Los modelos VEC bivariados estimados en este trabajo y otros modelos también aquí analizados,¹⁰ sugieren que la elasticidad de largo plazo de las muertes acumuladas a los contagios en los mismos términos es menor a la unidad. Este resultado implica que el incremento de los contagios a lo largo del tiempo bajaría la tasa de letalidad. Sin embargo, la reducción de dicha tasa sería probablemente lenta.

La elasticidad de muertes a contagios y la relación de largo plazo entre estas variables son conceptos útiles para hacer proyecciones simples de cómo va a influir a las muertes un incremento continuo de los contagios. En este caso se observa que diversas ecuaciones estimadas donde se supone a priori una causalidad que va solo de los contagios estimadas y los modelos VEC 37 y 68 bivariados, donde puede haber doble causalidad, muestran relaciones de muertes a contagios similares a las observadas hasta ahora, y donde con un máximo de tres millones de infecciones registradas habría entre 190 mil y 250 mil muertes también registradas.

El número de tres millones es relevante porque si, como todo indica, hay un factor de subregistro de contagios cercano a 40 veces y un factor de subregistro de muertes de entre 2 y 2.5 veces, los tres millones de contagiados implicarían que casi toda la población ya se contagió. En este caso, el número de muertes reales ascendería a entre 360 mil y poco más de 500 mil, lo que sin duda representa escenarios posibles muy negativos, porque implicaría que si al 31 de diciembre ya hubo más de 123 mil muertes oficiales y probablemente el doble de muertes reales, todavía podría haber un número similar de muertes antes de que termine la pandemia.

Este análisis muestra entonces la necesidad imperiosa de ahondar en la estrategia para detener los contagios de esta enfermedad. La vacuna será sin duda de gran ayuda, pero la logística para poder aplicarla a toda la población implica todavía mucho tiempo de espera y es necesario detener ya la gran mortandad creada por este virus.

¹⁰ La técnica de Engle y Granger (1987), Johansen (1988) y el ARDL de PSS (2001).

Referencias

- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.
- Banerjee, A., Dolado, J., Hendry, D. y Smyth, G. (1986). Exploring equilibrium relationships in econometrics through state models: some Monte Carlo evidence. *Oxford Bulletin in Economics and Statistics*, 42, 1-25.
- Benvenuto, D., Giovanetti, M., Angeletti, V. y Cicozzi, M. (2020). Application of the ARIMA model of the Covid-19 epidemic dataset. <http://doi.org/10.16/j.dib2020.105340>
- Chiang, A. (1984). *Fundamental methods of mathematical economics*. McGraw Hill.
- Cuthbertson, S., Hall, S. y Taylor, M. (1992). *Applied Econometric Techniques*. Harvester Wheatsheaf.
- Dehesh, T., Mardani-Fard, H. y Dehesh, P. (2020). Forecasting of Covid19 confirmed cases in different countries with ARIMA models. <https://www.medrxiv.org/content/medrxiv/early/2020/03/18/2020.03.13.20035345.full.pdf>
- Dickey, D. A. y Fuller, W. A. (1979). Distributions of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427-431.
- Elliot, G., Rothenberg, T. y Stock, J. (1996). Efficient tests for an autorregressive unit root. *Econometrica*, 64, 813-836.
- Engle, R. y Granger, W. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55, 251-276.
- Gómez Cravioto, C., Díaz-Ramos, R., Cantú-Ortiz, F. y Ceballos, H. (2020). An exploration and forecast of Covid-19 in Mexico with Machine Learning. *Cognitive Computation*, en prensa.
- Granger, W. (1969). Investigating causal relations by econometric models and cross spectral methods. *Econometrica*, 37, 424-438.
- Guilligan, B. (1990). Comparison of disease progress curves. *New Phytologist*, 115, 223-243.
- Hendry, D. y Juselius, K. (2000). Explaining cointegration analysis. *Energy Journal*, 21, 1-42.

- Instituto Nacional de Salud Pública de México. (1 de diciembre de 2020). *Resultados preliminares de la Encuesta Nacional de Salud y Nutrición COVID-19*. <https://www.insp.mx/avisos/resultados-preliminares-de-la-encuesta-nacional-de-salud-y-nutricion-covid-19#:~:text=Los%20resultados%20preliminares%20de%20la%20ENSANUT%20COVID%2D19%20muestran%20que,de%20personas%20a%20nivel%20nacional>.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economics, Dynamics and Control*, 12, 231-254.
- Johansen, S. (1995). Identifying restrictions of linear equations with applications to simultaneous equations and co-integration. *Journal of Econometrics*, 69, 111-132.
- Kánter, I. (2020). Muertes por Covid-19 en México. *Mirada Legislativa* No. 190. Ciudad de México: Instituto Belisario Domínguez, Senado de la República,
- Khan, F., Saeed, A. y Ali, S. (2020). Modelling and forecasting of new cases, deaths and recover cases of Covid-19 by using Vector Autorregresive model in Pakistan. *Chaos Solitons Fractals*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7405884/>
- Kilian, L. y Lutkepohl, H. (2017). *Structural vector autorregresive analysis*. Cambridge University Press.
- Kumari, R., Kumar, S., Chandra Poonia, R., Singh, V., Raja, L., Bhatnagar V. y Argawal P. (2021). *Big Data Mining and Analytics*, 4, 65-75.
- Kwiatkowski, D., Phillips, P., Shmidt P. y Shin Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54, 159-178.
- Lee, S. y Wong, N. (2010). Reconstruction of epidemic curves for pandemic influenza A(H1N1) 2001 at city and sub-city levels. *Virology Journal*. <https://virologyj.biomedcentral.com/articles/10.1186/1743-422X-7-321>
- Lutkepohl, H. (2005a). Structural vector autorregresive analysis of cointegrated variables. *EUI Working Paper ECO* No. 2005/2. Florencia: European University Institute.
- Lutkepohl, H. (2005b). *New introduction to multiple time series analysis*. Springer.
- Ng, S. y Perron, P. (2001). Lag length selection and the construction of unit root test with good size and power. *Econometrica*, 69, 1519-1554.

- Organización Mundial de la Salud. (Septiembre y diciembre de 2020). *WHO Coronavirus (COVID-19) Dashboard*. <https://who.sprinklr.com/>
- Paolo, B., Paruolo, P. P., Varzillo, S. y Lovaglio, G. (2020). A bivariate prediction approach for adapting the healthcare system response to the spread of Covid-19. *Plos One*, 15. <http://doi.org/10.1371/journal.pone.0240150>
- Perron, P. y Vogelsang, T. J (1992). Testing for a unit root in a time series with a changing mean: Corrections and extensions. *Journal of Business and Economic Statistics*, 10, 467-470.
- Pesaran, M., Shin, Y. y Smith, R. (2000). Structural analysis of vector error correction models with exogenous I(1) variables. *Journal of Econometrics*, 97, 293-343.
- Pesaran, M., Shin, Y. y Smith, R. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, 16, 289-326.
- Phillips, P. C. B. y Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75, 335-346.
- Romero, M. y Despeghel, L. (2020). ¿Qué nos dicen las actas de defunción de la Ciudad de México? Actualización al 31 de mayo de 2020. Taller de Datos Nexos. <http://datos.nexos.com.mx/>
- Secretaría de Salud. (1 de diciembre de 2020). *Exceso de mortalidad por todas las causas, durante la emergencia por COVID-19, México*. <https://coronavirus.gob.mx/exceso-de-mortalidad-en-mexico/>
- Sims, C. (1980). Macroeconomics and reality. *Econometrica*, 48, 1-48.
- Sims, C., Stock, J. y Watson, M. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58, 113-144.
- Stock, J. (1987). Asymptotic properties of least squares estimators of co-integrating vectors. *Econometrica*, 55, 1035-1056.
- Takayama, A. (1985). *Mathematical Economics*. Cambridge University Press.
- Torok, M. (2003). Epidemic curves ahead. *Focus on Field Epidemiology*, 1, 1-6.
- Yang, L., Chan, K., Suen, L., Chan, P., Wang, X., Cao, P., He, D., Peiris, J. y Wong C. (2015). Impact of the 2009 H1N1 pandemic on age specific epidemic curves of other respiratory virus: A comparison of pre-pandemic and post-pandemic periods in a subtropical city. *PLoS ONE*, 10. <https://journals.plos.org/plosone/article/file?type=printable&id=10.1371/journal.pone.0125447>