

40 Jornadas Nacionales de Administración Financiera
Octubre 2020

Metodología de ajuste y proyección S.I.R. a la cantidad de infectados en pandemia

El caso del Covid-19 en Argentina

Mauro Edgardo Speranza

Martín Ezequiel Masci

Universidad de Buenos Aires

SUMARIO

1. Introducción
2. Modelos epidemiológicos de ajustamiento y proyección de casos afectados
3. Modelo S.I.R. (Kermack y McKendrick, 1927)
4. Aplicación a las observaciones de casos de Covid-19 en Argentina
5. Conclusiones

Para comentarios:
martinmasci@economicas.uba.ar

1. Introducción

A la luz del año 2020, el brote de un nuevo virus luego transformado en la pandemia internacional del Covid-19, azotó en tiempo y espacio mediante una propagación global a todo el mundo. Su evolución y aún más su impacto económico fue y todavía a la fecha de generación del presente trabajo es de extrema incertidumbre, evitando la formulación de estrategias de regulación que den respuesta equivalente a la problemática. Esta nueva variable aleatoria, incierta, sin lugar a dudas de baja frecuencia o probabilidad de ocurrencia, pero de un elevado impacto, es lo que en cualquier documento académico de aquellos que forman el estado del arte en materia de riesgos, sería definido como un *outlier*, o en este caso, el mantenimiento de uno por un tiempo indeterminado.

Por su parte, durante los últimos años se han desarrollado rápidos avances en materia de tecnología y en este aspecto particularmente se ha puesto el principal foco en nuevas metodologías de explotación de grandes volúmenes de información, disponible usualmente en formato no estructurado (comúnmente denominado *Big data* de acuerdo al idioma inglés) y que hasta el momento no se encontraban naturalmente disponibles para su aprovechamiento. Mediante las mismas, resulta posible capturar datos al respecto de la evolución de la pandemia a nivel global, permitiendo esto construir alternativas de políticas públicas que resulten acordes a las necesidades de cada sector.

El objetivo de este trabajo enmarcado en este estado de situación es, a partir de datos recabados al respecto de los casos de contagios de la pandemia en Argentina, problematizar y proponer una metodología de ajustamiento a las observaciones disponibles, que permita posteriormente proyectar un escenario de continuidad de la crisis sanitaria. Además, se evalúan los resultados obtenidos con el objetivo de que dicha proyección pueda resultar aplicada a la gestión de riesgos de organizaciones y estados vinculados con esta materia. Finalmente se proponen posibles líneas de continuación para este trabajo.

2. Modelos epidemiológicos de ajustamiento y proyección de casos afectados

La problemática que se plantea en el siguiente apartado se encuentra asociada a la proyección de la variable de casos o contagios en situación de pandemia, tomando como parámetro para su validación la situación del Covid-19 y su expansión local. Sin embargo, la discusión que prevalece en un nivel de mayor profundidad se centra en la metodología óptima de ajustamiento a partir de la información que se encuentra disponible. Dicha pregunta refiere a un marco teórico de gran amplitud, que se ha visto bastante incrementado en los últimos años a partir del desarrollo de distintos tipos de metodologías.

Inicialmente, ante conjuntos de datos de elevada frecuencia y poca estructuración general, los modelos tradicionales como los asociados a regresiones de bajo nivel suelen encontrarse con algunas deficiencias prácticas. Aun teniendo la ventaja de resultar de fácil interpretación, estos modelos tienen demasiada limitación al respecto de la predicción de las variables que buscan representar producto de la forma estructural de los mismos. Asimismo, regresiones de mayores niveles suelen traer aparejados inconvenientes de sobre estimación de parámetros,

que generan dificultades al momento de generar una predicción a partir de las mismas. Si bien existen innumerables enfoques de modelado a través de técnicas de interpolación, extrapolación y ajustamiento mediante regresiones – alternando parámetros y formas funcionales – no hay una solución única y simple a la problemática bajo estudio mediante las metodologías de ajustamiento mediante modelos de regresión. Resulta necesario, producto del postulado anterior, explorar otra tipología de modelos diferenciados de los clásicos, que permitan aumentar la capacidad predictiva y explicativa al respecto de la situación evaluada.

Al mismo tiempo en que se transita el contexto de pandemia actual, comienzan a retomar importancia otra tipología de modelos matemáticos asociados a la epidemiología, a fin de intentar estimar la transmisión de casos y muertes asociados a la crisis. Estos modelos tienen un formato de series de tiempo y procesos estocásticos y también resultan de utilidad a fin de entender y predecir las variables de importancia asociadas a una pandemia.

Una de las características más representativas de los estudios de pandemias es la dificultad de encontrar factores causales que resulten adecuados para explicar la magnitud de la frecuencia de casos y contagios de una enfermedad que se expande por una población (Kermack y McKendrick, 1927). Mientras otro tipo de modelos, como los asociados a regresión, buscan encontrar una correlación en los comportamientos entre un conjunto de variables, la prioridad de este tipo de modelo epidemiológicos y matemáticos es la construcción de un marco teórico que provea la posibilidad de investigar y encontrar explicación dentro del comportamiento de un universo con el fin de alcanzar la causalidad entre los fenómenos observados.

El comportamiento de los modelos de esta naturaleza, entonces, se basa en un comportamiento poblacional donde un agente (o más) se incorpora en una comunidad siendo susceptible a una enfermedad. Por tanto, dicha enfermedad se disemina de aquella persona afectada a otras no afectadas mediante contacto entre las mismas. Al mismo tiempo, cada persona afectada transcurre dicha enfermedad y luego es removida del conjunto de agentes infectados, ya sea por su recuperación o por su fallecimiento.

Se deriva del punto anterior que al mismo tiempo que la epidemia se disemina en la comunidad, el número de agentes no afectados disminuye a tasas exponenciales. Finalmente, una vez que la epidemia avanza en el curso del tiempo, la misma llega a una finalización, ya sea porque no existen más individuos susceptibles a contraer la enfermedad dentro de la comunidad analizada –es decir, todos fueron infectados en algún momento dentro de la duración de la misma– o por una terminación anterior producto de la no diseminación incluso con individuos susceptibles aún presentes en la población que no han resultado afectados (Cooper *et al.*, 2020).

Por supuesto, una de las mayores críticas a este planteo es justamente la flexibilidad de los supuestos que el mismo propone. El primero de ellos es población, donde se propone una población sin cambios en el tiempo a excepción de los originados por la epidemia. En este sentido, la fundamentación se basa en suponer que el fenómeno tiene un curso de corto plazo en el tiempo comparado con la vida de un individuo y por tanto la población o comunidad tomada como referencia puede considerarse constante excepto por los fallecimientos generados por esta causa. Este primer argumento, si bien coherente en un contexto de corto plazo, pasa a ser más difícil de relajar al momento de aplicar el modelo a una pandemia.

No obstante, la pregunta de qué representa o cuáles son las características inequívocas de una pandemia no resulta de fácil solución. El periódico *New York Times* publicó el 8 de Junio de 2009 –en medio de la confusión producida por diseminación del virus H1N1 de influenza–

como titular: “*Is this a pandemic? Define Pandemic*”¹. Solo unos días después la Organización mundial de la Salud (OMS) anunció el primer comunicado de alerta por pandemia de dicho año. Siguiendo el analizado por Morens, Folkers y Fauci (2009), los autores recaban el estado del arte al respecto de esta materia y exponen a la definición de pandemia como una epidemia distribuida sobre un área suficientemente grande, con un movimiento geográfico de la enfermedad, con elevadas tasas de contagio de crecimiento exponencial y ante una mínima población que genera inmunidad. Por tanto, si este es el caso, los modelos de epidemia podrían resultar aplicables al fenómeno de pandemia observado.

Una segunda crítica posible a los modelos epidemiológicos es que los mismos asumen en su versión más tradicional que no existe la posibilidad de re contagio que permita perpetrar la infección. Esta característica es de importancia a la luz de eventos puntuales en el fenómeno del Covid-19 que muestra alertas tempranas de un resurgimiento de focos de contagio en zonas recuperadas. De esta manera, habrá que evaluar en cada caso la posibilidad de levantar este supuesto en la medida que el objetivo propuesto contemple un horizonte de largo plazo.

En tercera instancia, estos modelos matemáticos se nutren de la información observada al respecto de la evolución de las variables de interés. Sin embargo, dicha información puede resultar sesgada a características de la población –por ejemplo, una permeabilidad diferenciada de diferentes grupos poblacionales o *clusters* al contagio de la enfermedad– y también a medidas del ambiente –por ejemplo, medidas que restringen la circulación tomadas por autoridades a fin de aminorar el avance de la enfermedad. Por tanto, resultará necesario hacer un estudio profundo del funcionamiento del modelo propuesto y de las respuestas obtenidas del mismo, a fin de no generar un mal juicio sobre los resultados que se obtengan.

En el marco de los puntos anteriores, es que se propone estudiar el modelo S.I.R. originalmente propuesto por Kermack y McKendrick (1927) que resultó un modelo epidemiológico fundacional por más de un siglo. El mismo da un marco conceptual simple pero completo, sobre el que se han generado múltiples extensiones posteriores producto de su flexibilidad.

3. Modelo S.I.R. (Kermack y McKendrick, 1927)

El modelo epidemiológico denominado S.I.R. –sigla que deviene del anacronismo susceptible, infectado y removido– es una alternativa de modelo matemático epidemiológico que permite generar un ajustamiento asociado a una pandemia e incorporando la capacidad de interpretación al respecto de las variables y su relación con el contexto en que se desarrolla.

Siguiendo lo que plantea Ellison (2020), el análisis de las consecuencias de la pandemia del Covid-19 se pueden fundar en las bases de un modelo SIR homogéneo. Explicando resumidamente el mismo, se plantea una población que solamente puede estar en uno de los tres estados que dan origen a su nombre y cada uno con valores dependientes del tiempo. Estas dinámicas se resumen en las ecuaciones siguientes:

$$\frac{dI(t)}{dt} = S(t)I(t)\beta - \gamma I(t)$$

¹ De acuerdo al idioma inglés: “¿Es esto una pandemia? Defina pandemia”

$$\frac{dR(t)}{dt} = \gamma I(t)$$

$$\frac{dS(t)}{dt} = -S(t)I(t)R_0\beta$$

En este conjunto de ecuaciones, la población $S(t)$ se encuentra presentada por los individuos que no están infectados pero podrían estarlo en un futuro. Por su parte, la población $I(t)$ representa a aquellos individuos que se encuentran infectados y son capaces de transmitirlo a otros individuos susceptibles. Por último, la población $R(t)$ de removidos son aquellos individuos que se recuperaron de la enfermedad –se los asume inmunes– o murieron por esta causa.

A partir de las definiciones, se construye el ratio de reproducción de la pandemia mediante de la ecuación $R_0 = \frac{\beta}{\alpha}N$. Para ello, este modelo asume que la población (N) permanece constante en su total por cualquier otra razón poblacional.

De esta manera, en la aplicación de un esquema como el considerado, las ecuaciones diferenciales describen las leyes de la naturaleza e interacciones entre variables que buscan representar un marco de contexto. Esta es una diferencia de gran importancia al comparar estos modelos con otros asociados al ajustamiento a datos (independientemente de su grado de precisión), dado que estos últimos solo buscan encontrar los parámetros que logren minimizar errores en la estimación, pero no generar una interpretación específica de un fenómeno concreto.

Este modelo simplificado se construye sobre el supuesto de que cada individuo en la población analizada es igualmente susceptible a la enfermedad y que no existe la posibilidad de múltiples contagios posterior a la salida del conjunto de afectados –todos los cuales fueron profundizados en el apartado anterior. Desde ya, existe evidencia empírica para poner un manto de duda a algunos de los supuestos mencionados.

La finalización de la pandemia, en este caso, no necesariamente se alcanza cuando toda la población alcanza un estadio de contagio, sino que, por el contrario, producto de una combinación de algunos factores como la inefectividad, recuperación, mortalidad y políticas exógenas tomadas para todo el entorno, es factible la existencia de un umbral poblacional crítico que extingue la enfermedad. Sin dudas, es normal esperar que este punto solamente resulte alcanzando por una porción suficientemente pequeña del total poblacional inicial.

Asimismo, este modelo también tiene dentro de sus supuestos de mayor importancia la consideración de una población homogénea sobre agentes infectados y susceptibles a serlo. En ambas, la población susceptible decrece de manera monótona con límite en cero. Sin embargo, estos supuestos no son necesariamente correctos para cualquier pandemia, por ejemplo, ante la aparición de nuevos focos de infección o epicentros de la enfermedad.

Sin embargo, a pesar de sus limitaciones, el modelo propuesto puede ser aplicado a fin de obtener un mejor entendimiento sobre la diseminación del virus dentro de una población o comunidad. Asimismo, el modelo resulta de utilidad con el objetivo de realizar predicciones sobre el número de infectados y muertes futuras, permitiendo proveer una escala temporal para la duración de la pandemia y la construcción de escenarios de riesgo. Por tanto, el mismo provee un marco teórico conceptual y de inferencia que puede aplicarse a una situación de

pandemia y que por consiguiente en este trabajo su aplicación se realiza sobre las observaciones de casos de Covid-19 en Argentina.

4. Aplicación a las observaciones de casos de Covid-19 en Argentina

La aparición de los conceptos de grandes volúmenes de datos o *Big Data*, ha resultado un cambio paradigmático en la forma de ver y entender las ventajas competitivas de los participantes del mercado que logren adecuarse a este flujo de información y capturarlo. En este sentido, su implementación en la organización resulta un cambio necesario para explotar la oportunidad de negocios que esta revolución trae aparejada. Entre otras cosas, no solamente modelos deben ser actualizados, sino que también se requiere la presencia de nuevos roles dentro del contexto de una institución renovada y el proceso mismo de co-construcción del conocimiento ha manifestado alteraciones (Schmarzo, 2013).

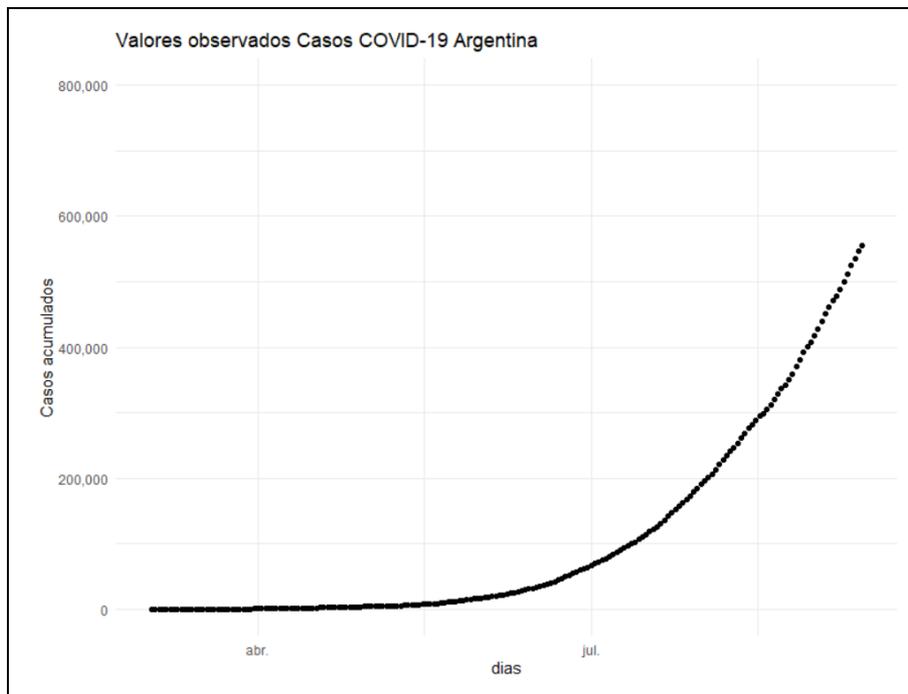
A partir del punto anterior, se manifiesta el desafío principal que trae aparejado este cambio estructural al incorporarse en medio de una de las crisis sanitarias más importantes de la historia. La crisis, inestabilidad macroeconómica global y sobre todo situación de incertidumbre masiva, se produce con un rezago en los países de América Latina, dentro de los cuales se pondrá particular enfoque en Argentina para la construcción de este trabajo. A partir del mencionado rezago, grandes volúmenes de información incipiente, poco estructurada y sobre todo no necesariamente confiable se vuelven disponibles, en el marco de una necesidad urgente de decisiones cotidianas con elevados costos y potencialmente incluso perjudiciales ante interpretaciones incorrectas al respecto de la evolución del sistema.

En ese contexto, mediante las herramientas de explotación de datos se implementarán las metodologías de interpolación y ajustamiento sobre las que se ha discutido bastamente en el apartado anterior, tomando las observaciones disponibles al respecto de la pandemia del Covid-19 en el contexto de Argentina. En esta primera instancia, el objetivo final será explicar con los modelos que mayores ventajas han mostrado –esto es, el modelo estocástico y epidemiológico de transmisión S.I.R. – las observaciones que hayan sido capturadas al respecto de la evolución de casos a nivel local. Asimismo, a continuación se explora y profundiza sobre la capacidad de explicar o pronosticar su evolución en el tiempo. Esto es dado que, la búsqueda de una causalidad o correlación entre diferentes caminos de acción, de manera de comprender dichas relaciones lleva a tomar mejores decisiones en un ámbito de restricciones en la información disponible, tal como mencionan McAfee *et al* (2012).

A partir de la información pública disponible, el primer paso es construir una serie de tiempo con la información de la variable respuesta a analizar. En el caso de este trabajo, dicha variable estará representada por la evolución de los casos acumulados de contagiados por el virus en Argentina. Para ello, tomando la información de fuentes públicas mediante metodologías de *web scrapping*,² se construye y presenta a continuación la evolución temporal de la variable en forma gráfica.

² Extracción desde sitios web de la base de datos analizada a partir del software R.

Figura 1: Evolución de casos confirmados acumulados de Covid-19 en Argentina



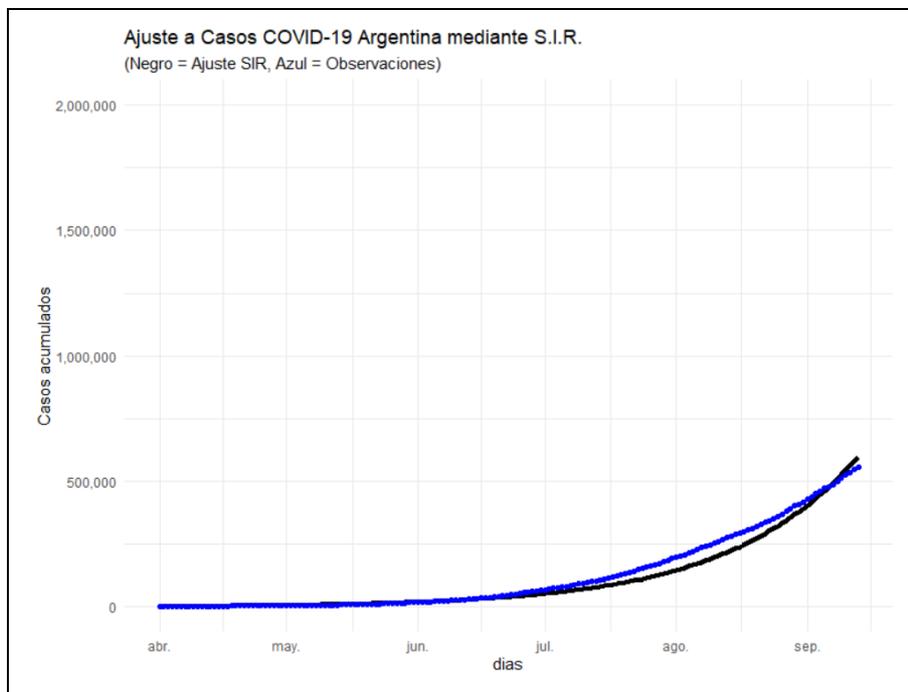
Fuente: Basado en datos de <https://www.worldometers.info/coronavirus/>

Las observaciones representadas en la Figura 1 muestran la evolución de los casos acumulados conocidos para Argentina desde el inicio de la pandemia en Marzo y hasta la última fecha registrada y publicada hasta el momento de realización de este trabajo que es el 13 de Septiembre de 2020. A partir de la recopilación de los datos y de las conclusiones arribadas en el apartado anterior, se plantea el ajustamiento mediante el modelo matemático S.I.R. que resultó el más destacados en términos de precisión y de capacidad de explicación del fenómeno planteado. Como referencia, la figura 2 muestra los resultados obtenidos a partir de la calibración de dicha forma funcional a las observaciones realizadas, a partir de l.

A partir del análisis de los mismos no se observan diferencias significativas en términos de ajuste a la serie temporal de datos recopilada. Se puede afirmar que el error cuadrático se minimiza al aplicar los siguientes parámetros mediante un sistema de optimización de las ecuaciones diferenciales que plantea el modelo: $\alpha = 0.1105$ y $\gamma = 0.07734$. De esta manera se alcanza un umbral semejante a las observaciones de casos totales mediante el modelo de proyección, que marcaron los 555.537 contagiados acumulados en Argentina al 13 de Septiembre de este año.

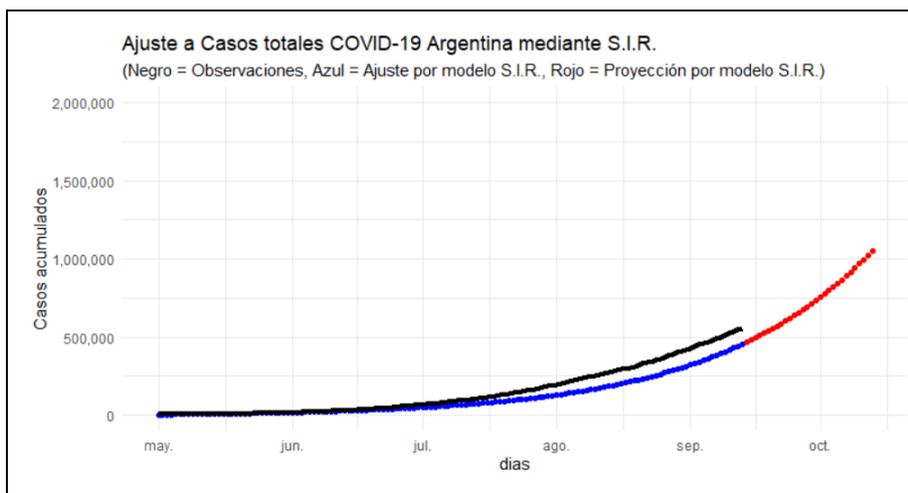
Posteriormente, dado que el modelo matemático S.I.R. tiene la ventaja de estar constituido como una serie temporal, permite la proyección a partir de la calibración inicial con los parámetros que explican la evolución de la enfermedad. En este caso, dada la variabilidad que trae aparejado el entorno, se realizó una proyección al respecto de los primeros 30 días posteriores al último dato observado. De esta manera, se obtuvo el resultado que se presenta en la Figura 3.

Figura 2: Evolución y ajuste mediante el modelo SIR, de casos totales confirmados acumulados de Covid-19 en Argentina



Fuente: Basado en datos de <https://www.worldometers.info/coronavirus/>

Figura 3: Ajuste y proyección mediante el modelo SIR, de casos totales confirmados acumulados de Covid-19 en Argentina



Fuente: Basado en datos de <https://www.worldometers.info/coronavirus/>

De esta manera, el modelo S.I.R. permite proveer detalles y predicciones al respecto de la diseminación del virus en una comunidad –o como en este caso, en un país–, mientras que los datos en su estado natural no permiten proveer esta información. Por tanto, este proceso

muestra la importancia del modelado de la pandemia con un instrumento como el propuesto, ayudando a evaluar el impacto futuro de la misma.

La proyección, entendida como un intento cuantitativo de predecir los eventos del horizonte analizado, no considera ningún tipo de cambio en las condiciones iniciales. No obstante, durante el curso de una pandemia puede resultar de interés entender los futuros cursos de acción y su efecto en tiempo real (Nishiura, 2011). Desde ya, la aplicación de este modelo tiene como limitaciones la incapacidad del mismo, en su versión más tradicional, de capturar cambios de conductas de los agentes de la comunidad evaluada. Esto resulta de gran importancia al incorporar restricciones a la movilidad de los individuos y otras medidas que los gobiernos están implementando como método de combate contra este fenómeno. No obstante, algunas alteraciones al modelo permitirían su correcta adecuación.

5. Conclusiones

En este ensayo se discute la problemática de la adecuación, modelización y proyección de los casos de contagios durante el transcurso de una pandemia, realizando una aplicación al fenómeno del Covid-19 en Argentina. Apartándose de los modelos de regresión comúnmente utilizados por su capacidad de adaptación, se planteó una reducción metodológica al modelo matemático y epidemiológico S.I.R. de Kermack y McKendrick (1927) dado que resulta en mayores ventajas al momento de generar un ajuste de una variable compleja como lo es los casos de infectados en una pandemia.

Los modelos como el planteado, fundados en teorías matemáticas, son herramientas eficaces a fin de representar una evolución de un virus o en este caso una pandemia, puesto que permiten construir una predicción al respecto de su evolución, al mismo tiempo que evaluar el impacto de la intervención de diferentes medidas a tomar. En este sentido, el modelo utilizado (S.I.R.) provee un marco teórico para investigar la evolución de los casos del Covid-19 en Argentina solamente a partir de los datos observados y tomando discrecionalidades a partir de la demografía y contexto de una población determinada.

Una de las ventajas del mismo radica en que, en la medida que se incorpora nueva información al modelo, resulta simple ajustar sus parámetros, recalcular el mejor ajuste a las curvas observadas y volver a proyectar la continuación del mismo en el tiempo. En este contexto, la extensión del modelo permite además proveer estimaciones sobre los números de casos futuros y sobre la evolución de su diseminación en la sociedad.

El modelo que se ha aplicado es la versión tradicional del mismo, no obstante, diferentes restricciones o agregados al mismo podrían permitir mejorar la precisión y entender la reacción de su evolución ante medidas tomadas. No obstante, el resultado sigue siendo permeable a la precisión y veracidad de los datos observados.

Cabe resaltar que en fases iniciales de una pandemia existe una elevada incertidumbre científica sobre la enfermedad y su proceso de transmisión. Por tanto, es aceptable la hipótesis a incorporar al respecto de que aún no es conocida o revelada la dinámica final del modelo necesario para su aproximación. La estimación de parámetros en momentos tempranos de su evolución podría llevar a una seria subestimación del daño final y la extensión de la pandemia.

Finalmente, queda como primera línea de investigación futura a partir del presente trabajo el generar un análisis de modelos alternativos y complementarios, que permitan levantar alguno de los supuestos limitantes encontrados, mediante la incorporación de nuevas variables o nuevos estados de los agentes que permitan la introducción de un sistema de ecuaciones diferentes más completas, como es el caso de lo que plantean Sarkar, Khajanchiy Nieto (2020) o Hall, Gani, Hughes y Leach (2007). Asimismo, como segundo punto resulta de interés el estudio sobre la capacidad de los presentes modelos para representar otras variables de interés en este contexto – particularmente los muertos observados en una población. Por último, como tercer punto se presenta la necesidad de generar un *backtest* o pruebas de validación, para confirmar que las predicciones hechos por los mismos tienen la representatividad necesaria a fin de poder construir medidas institucionales fundadas sobre los mismos.

REFERENCIAS

- Cooper, I., Mondal, A. & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139, 110057
- Ellison, G. (2020). *Implications of heterogeneous SIR models for analyses of COVID-19* (No. w27373). National Bureau of Economic Research
- Hall, I. M., Gani, R., Hughes, H. E. & Leach, S. (2007). Real-time epidemic forecasting for pandemic influenza. *Epidemiology & Infection*, 135(3), 372-385
- Ioannidis, J. P., Cripps, S. & Tanner, M. A. (2020). Forecasting for COVID-19 has failed. *International Journal of Forecasting*
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). Springer
- Kermack, W. O. & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 115(772), 700-721
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J. & Barton, D. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60-68
- Morens, D. M., Folkers, G. K. & Fauci, A. S. (2009). *What is a pandemic?*, The New York Times, June 8, 2009
- Ng, A. Y. (2011). *Cs229 Lecture notes on machine learning*. Technical report, Stanford University, Department of Computer Science, 2011. 39, 116, 140
- Nishiura, H. (2011). Real-time forecasting of an epidemic using a discrete time stochastic model: a case study of pandemic influenza (H1N1-2009). *Biomedical Engineering online*, 10(1), 15
- Sarkar, K., Khajanchi, S. & Nieto, J. J. (2020). Modeling and forecasting the COVID-19 pandemic in India. *Chaos, Solitons & Fractals*, 139, 110049
- Schmarzo, B. (2013). *Big Data: Understanding how data powers big business*. Wiley