

Aportes sobre COVID-19 *

**Modelos de series temporales adaptativos y robustos
para predecir contagios y muertes por COVID-19 en
Uruguay: predicciones resultantes de los modelos de
Castle, Doornik y Hendry**

Silvia Rodríguez-Collazo**

12 de Febrero de 2021

* El objetivo de esta serie de notas, creada a partir de abril de 2020, es contribuir a la reflexión sobre distintos aspectos de la situación de pandemia generada por el coronavirus SARS-CoV-2, con especial énfasis en Uruguay. Los documentos son elaborados por investigadores e investigadoras del Instituto de Estadística, desde las diversas perspectivas que forman parte de sus áreas de investigación, y en algunos casos en coautoría con invitados externos.

** Universidad de la República. Facultad de Ciencias Económicas y de Administración, Instituto de Estadística. Montevideo, Uruguay.
ORCID ID <https://orcid.org/0000-0002-3871-6448> Contacto: silvia@iesta.edu.uy



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Forma de citación sugerida para este documento:

Rodríguez-Collazo, Silvia. (2021). *Modelos de series temporales adaptativos y robustos para predecir contagios y muertes por COVID-19 en Uruguay: Predicciones resultantes de los modelos de Castle, Doornik y Hendry* (Serie Aportes sobre COVID-19, 12 de Febrero). Montevideo: Universidad de la República. Facultad de Ciencias Económicas y de Administración, Instituto de Estadística.

<https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/25339>

Modelos de series temporales adaptativos y robustos para predecir contagios y muertes por COVID-19 en Uruguay: Predicciones resultantes de los modelos de Castle, Doornik y Hendry.

Silvia Rodríguez-Collazo ¹ *Departamento de Métodos Cuantitativos, Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República*

Resumen

Esta nota ² pretende brindar elementos para comprender los diversos resultados que se obtienen a partir de dos tipos de metodologías para la modelización de algunas variables que informan sobre el desarrollo de la epidemia y en numerar algunos principios que deben de cumplir los procedimientos de proyección y predicción para que sean informativos y útiles.

El siguiente objetivo de la nota es dar a conocer las predicciones del acumulado de contagios y el número de fallecimientos para Uruguay provenientes de los modelos creados por Castle, Doornik y Hendry. Este grupo de científicos brindan de manera sistemática y pública tanto para Uruguay como para un amplio conjunto de países predicciones a 7 días de plazo.

En contraposición con la detallada documentación de acceso público que han generado Castle, Doornik y Hendry, en Uruguay la documentación pública generada por el equipo de Modelos y Ciencia de Datos del GACH es muy escasa. No es posible encontrar documentos de acceso público que den cuenta de las metodologías utilizadas para la estimación de sus modelos ni tampoco se dispone de información pública sobre proyecciones actualizadas que elaboran, ni el horizonte de proyección. Esta falta de información que parece ser producto de una decisión explícita, no se comprende.

El interés por la evolución de esta epidemia en nuestro país no ha decrecido, se cuenta con información diaria sobre un conjunto de variables que permiten hacer un seguimiento de la epidemia pero no se cuenta con información pública sobre proyecciones o predicciones actualizadas, por lo que se espera que esta nota pueda adicionar información sobre este tema.

¹ *email:* silvia@iesta.edu.uy, ORCID: 0000-0002-3871-6448

² Agradezco los valiosos comentarios, sugerencias recibidas, así como los fructíferos y estimulantes intercambios realizados con Fernando Lorenzo, Bibiana Lanzilotta, Laura Nalbarte, Ramón Álvarez-Vaz y Enrique Tenca. Por supuesto los errores que persistan son de mi entera responsabilidad.

Epidemia en Uruguay, su modelización y cómo estimar su evolución futura

Esta nota pretende explicitar dos tipos de aproximaciones metodológicas para la modelización de algunas variables que nos informan sobre el desarrollo de la epidemia. Se espera poder brindar elementos para comprender los diversos resultados que se obtienen cuando se aplican opciones metodológicas alternativas.

Se darán a conocer las predicciones del acumulado de contagios y el número de fallecimientos en **Uruguay** que provienen de una de las metodologías a las que haremos referencia en la nota, modelos de series temporales adaptativos y robustos. Estas predicciones se brindan de manera sistemática y pública tanto para Uruguay como para un amplio conjunto de países. Esos pronósticos son realizados por un grupo de reconocidos científicos que trabajan en Gran Bretaña y tienen una larga carrera destinada a crear conocimiento nuevo en temas de modelización y pronóstico.

Esta nota se apoya en los numerosos documentos que han elaborado los autores, así como en los datos que se ponen a disposición pública.^{3, 4.}

En la nota elaborada en junio de 2020⁵ se presentaron los principales rasgos de la estrategia de modelización de Jennifer Castle, Jurgen Doornik y David Hendry,⁶ para generar predicciones acerca de la evolución diaria, y con hasta 7 días de antelación, de los casos acumulados de infectados por el virus y del número de muertes como consecuencia del COVID-19.

En esta nueva nota, luego dar un marco de análisis se intentará presentar dos visiones alternativas que actualmente se utilizan para la modelización, predicción o proyección de un conjunto de variables de interés sobre las que se sigue la evolución de la pandemia, pero el énfasis estará en una de estas aproximaciones. Constituye una continuación y profundización del tema abordado en la Nota Metodológica presentada en junio de 2020. También se presentará la aplicación de esta metodología para el acumulado de infectados por el virus y el número de fallecimientos **en Uruguay**.^{7. 8.}

Desde el inicio y como en otras partes del mundo, el gobierno decide apoyarse en el asesoramiento de equipos científicos locales, se crea el GACH (Grupo Asesor Científico

³<http://users.ox.ac.uk/~nuff0231/links.htm>

⁴<https://www.doornik.com/research.html>

⁵<https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/25413>

⁶<https://www.doornik.com/COVID-19/>

⁷https://www.doornik.com/COVID-19/index_deaths_lac.html

⁸https://www.doornik.com/COVID-19/index_lac.html

Honorario)⁹, conformado por un equipo multidisciplinario de científicos de algunas áreas de la ciencia.

En Uruguay el equipo de Modelos y Ciencia de Datos, integrantes del GACH, ha elaborado y estimado diversos modelos. Algunos de estos resultados se han dado a conocer en dos informes de carácter público, el 12 de diciembre de 2020 y el 18 de enero de 2021. Lamentablemente la documentación pública resultante de ese asesoramiento es muy escasa, por lo que no es posible encontrar informes o documentos de acceso público que den cuenta de las metodologías utilizadas para la elaboración de las proyecciones, aunque algunos de sus integrantes han brindado diversas entrevistas en los medios de comunicación en la que informan parcialmente.^{10 11} Pero estos elementos no son suficientes, no es posible dar seguimiento a sus proyecciones y diagnósticos ni evaluar sus resultados.

Disponer de predicciones actualizadas con bajos errores de pronóstico de las trayectorias de variables que dan cuenta de diversos aspectos de la evolución de la epidemia en el corto plazo, puede contribuir a reducir la incertidumbre sobre la dinámica de la misma en los próximos días y convertirse en una referencia sobre lo que probable esperar en un futuro próximo. También contribuyen a la definición de la respuesta que la autoridad sanitaria puede diseñar para el futuro próximo, permiten dar seguimiento e incluso evaluar la eficacia de las políticas sanitarias implementadas y determinar eventuales cambios de rumbo lo antes posible.

Dos aproximaciones metodológicas y la diferencia entre elaboración de escenarios y producir predicciones

Hay una enorme variedad de modelos que se han utilizado para predecir la evolución de la epidemia. En esta nota vamos a tomar dos enfoques diferentes: los modelos epidemiológicos y los modelos estocásticos basados en datos observados pasados.

Los modelos de tipo de matemático están basados en la forma en que se transmite la enfermedad, apoyados en teorías probadas de la forma en que se propagan las enfermedades. Estos modelos matemáticos dan un marco para el análisis epidemiológico y se los conoce como modelos epidemiológicos.

Se utiliza la matemáticas para tratar de explicar los procesos subyacentes de los brotes, en general adoptando una pequeña cantidad de parámetros. En ellos las proyecciones se

⁹<https://www.presidencia.gub.uy/gach/>

¹⁰<https://www.enperspectiva.net/en-perspectiva-programa/entrevistas/fernando-paganini-gach-objetivo-ahora-llegar-zona-amarilla-los-200-casos-covid-19-dia/>

¹¹<https://www.enperspectiva.net/tag/ernesto-mordecki/>

realizan en base a esos parámetros, calibrados a partir de la experiencia previa o de valores plausibles, extrapolando relaciones o parámetros.

Estos modelos pueden ser usados para analizar escenarios. Mediante la generación de escenarios alternativos ante cambios en el parámetro de base se construyen los correspondientes resultados del modelo, que constituyen resultados potenciales. A ese escenario construido a través del modelo, en ocasiones se le llama proyección. Esta denominación adiciona una dificultad a la hora de diferenciar entre el mejor escenario y una predicción. Un escenario nos puede decir lo que podría suceder, no lo que es probable que suceda, y con ello se dificulta estimar la incertidumbre asociada a ese escenario.

El enfoque de los modelos estocásticos, no pretende comprender la razón de los cambios en la evolución de la epidemia, sino que partiendo de los datos observados del pasado intenta encontrar relaciones, patrones, regularidades modelizables en esos datos, para predecir lo que ocurrirá en los próximos días. Su objetivo primordial es obtener una buena predicción y no sólo un buen ajuste a los datos, por lo que en ocasiones no son modelos de tipo explicativo sino de carácter predictivo.

El pronóstico es el resultado de una extrapolación estadística a un plazo determinado, pero estos modelos incorporan las variaciones aleatorias de la evolución de la enfermedad. Son modelos, como señalamos previamente, de tipo estocástico.

Veamos la diferencia entre el resultado de un escenario que surge a partir de los modelos epidemiológicos y un pronóstico que surge del modelo predictivo. Un pronóstico o predicción (*forecast*), es el resultado esperado de la variable en el futuro, valor esperado que resulta del modelo estimado.

Habría una variedad de aspectos a tener en cuenta a la hora de analizar los resultados de la aplicación de los modelos, pero en este breve espacio vamos a analizar el error de los pronósticos o de los escenarios construidos a partir del modelo.

Si el pronóstico es para una variable observada, como el número acumulado de contagios, es sencillo verificar si el pronóstico es compatible con los datos observados, pues la comparación entre valor esperado o pronóstico y valor observado surge de forma directa. Pero si el pronóstico no es una variable observada, verificar si el pronóstico fue compatible con los datos no es tan sencillo. Por ejemplo si el modelo se apoya en el número estimado de personas que contagia cada infectado, el conocido número de reproducciones, R_t , es un parámetro estimado, y a partir de un modelo para este parámetro se estima una trayectoria para la cantidad de contagios. Entonces la comparación con los datos observados de contagios no es directa como en la situación anterior y esto dificulta la evaluación del modelo.

Un enfoque metodológico no es mejor que otro a priori, sino que ambos trabajan sobre la base de supuestos diferentes y los resultados de los modelos son conceptualmente diferentes y tienen utilidades diferentes.

Al elaborar un modelo con el objetivo de proyectar hacia adelante en el tiempo la dinámica que se viene dando, sólo se puede obtener información sobre lo que ocurriría en el futuro bajo el supuesto de que no se registran cambios en los parámetros sobre los que se construye el modelo o suponiendo que no se registren cambios en la evolución futura respecto al pasado en los datos, ni en su distribución. Cuando esos cambios ocurren, ya sea por cambios en los parámetros base o en la dinámica de la enfermedad, las predicciones fallan y los escenarios construidos no se concretan en la realidad. Bajo esa situación el grado de adaptabilidad y flexibilidad del modelo permite o no corregir el rumbo con mayor o menor velocidad y así poder re estimar y proyectar nuevamente incluyendo esos “quiebres” o cambios que se registraron.

La evolución de esta epidemia tiene mucho que ver con el comportamiento individual y colectivo de las personas, por lo que cambios en el comportamiento, ya sea que se impongan políticas de restricción de movilidad, aforos, etc. por parte del gobierno o que las personas decidan atender en mayor medida las recomendaciones sobre distanciamiento y cuidado personal, todo ello tiene la capacidad de modificar la dinámica futura de los contagios. Los modelos que no tengan la flexibilidad necesaria, producirán por más tiempo proyecciones o escenarios alejadas de la realidad.

¿Cuál es la base de modelización para Castle, Doornik y Hendry ?

La modelización de las series temporales del acumulado de casos positivos, así como del número de muertes por COVID-19 enfrenta el desafío de representar adecuadamente la forma de la no estacionariedad que presentan los datos. En este caso existen dos fuentes fundamentales de no estacionariedad: la que proviene de la propia evolución de la serie en condiciones normales que definen el carácter estocástico del componente tendencia y la que se genera como consecuencia de cambios bruscos y quiebres de diversa magnitud que afectan a la evolución de las variables.

La presencia de ambas formas de no estacionariedad implica que la estructura paramétrica de la función de distribución evoluciona a lo largo del tiempo y esto debe ser contemplado en la modelización empírica.

Para abordar esta cuestión, los modelos estadísticos utilizados por los autores combinan el uso de tendencias estocásticas con la introducción de cambios estructurales en la función de distribución de las variables analizadas.

Tanto las políticas sanitarias llevadas adelante, como las modificaciones en las metodologías de registro tanto de contagios como de muertes, influyen sobre las características de los datos que se utilizan en la construcción de los modelos. Cabe precisar, por otra parte, que los conteos de casos están sujetos a quiebres impredecibles, debido a diversas razones como el sub-reporte, cambios en los criterios de definición, variaciones en la cantidad de tests que se realizan diariamente, retrasos en el registro de los datos y hasta a errores involuntarios en las metodologías de construcción de las series, entre otros.

Castle, Doornik y Hendry adoptan un enfoque metodológico para obtener predicciones robustas de las cifras de contagios y fallecimientos, basados en el uso de modelos de tendencias estocásticas y en la consideración de la existencia de cambios en la función de distribución de los datos. La propuesta consiste en la aplicación conjunta de operaciones de diferenciación, como las habitualmente aplicadas en el tratamiento de la no estacionariedad, y técnicas de saturación, basadas en la aplicación de variables indicatrices que recogen los efectos de los quiebres que se van produciendo en los datos a lo largo del tiempo.

De este modo, atienden las dos fuentes de no estacionariedad mencionadas, construyendo modelos predictivos robustos, que se adaptan al comportamiento de los datos.

Estos modelos parten de dos supuestos referidos a las predicciones que provienen de ellos: suponen que la tendencia subyacente es informativa de la evolución de corto plazo de la epidemia y no se requiere incorporar en el modelo supuestos sobre la expansión del virus ni sobre las políticas preventivas. Esto los diferencia de los modelos epidemiológicos.

Las series de frecuencia diaria, se descomponen separando el componente tendencial (flexible) de las mismas. Hay que tener presente que la estimación de la tendencia está sujeta a varios desafíos como las políticas de intervención que se realizan con el fin de reducir o suprimir la transmisión del virus, las distintas estrategias y tecnologías de testeo que se implementan y los conteos de casos que están sujetos a quiebres estructurales, sub-reportes, cambios de definición, retrasos y errores entre otros factores.

Para resolver esos desafíos, la tendencia es estimada a partir de ventanas móviles de los datos, y se aplica el método de saturación para conseguir robustez y flexibilidad en la estimación del componente tendencial a lo largo del tiempo. Se satura con tendencias determinísticas lineales, con el objeto de captar si hay un cambio de nivel en la tasa de

crecimiento de la tendencia. El proceso de selección de estas variables se realiza con un algoritmo econométrico de *machine learning* denominado Autometrics. Posteriormente, las tendencias lineales seleccionadas son promediadas.

El otro componente restante, que no es la tendencia, se predice de forma separada mediante un método denominado Cardt. Este método es el resultado de combinar, el método Delta, que especifica y estima una tendencia “amortiguada” para las tasas de crecimiento de la serie y el método Rho, que estima un modelo autorregresivo. Este último proceso puede incorporar variables indicatrices estacionales o raíces unitarias estacionales, si corresponde. Y se crean nuevas predicciones a partir de un promedio de las predicciones provenientes de estos dos métodos.

Finalmente las predicciones elaboradas en el paso anterior se toman como si fueran datos observados y se estima un modelo autorregresivo más rico que el proveniente del método Rho. Los valores estimados que se obtienen como resultado de esta última etapa se convierten en las predicciones finales.

Todos los procedimientos y metodologías utilizadas, están documentados y son accesibles al público.

Algunos elementos a considerar en el proceso de predicción

La literatura sobre la teoría de la predicción es muy amplia y constituye en sí misma un campo de estudio. El error es inherente a la tarea de predicción, pero de su cuantificación y análisis se puede extraer información muy valiosa para el proceso continuo de reelaboración y actualización de las predicciones.

Hay algunos elementos básicos que debemos preguntarnos si están presentes a la hora de elaborar predicciones o proyecciones como ¿qué variable se va a predecir (o proyectar)?, ¿cómo se producen esas predicciones?.

En Diebold (2007) se advierte sobre un conjunto de aspectos a tener presente cuando se encara la tarea de realizar predicciones (o proyecciones):

- ¿Cómo cuantificamos lo que es un buen pronóstico?
- ¿Cómo se calculan los pronósticos óptimos?, ¿Cuál es el costo o la pérdida asociada con los errores de pronóstico?, ¿es igual si el signo del error es positivo o negativo?, ¿tiene igual costo equivocarse por encima que por debajo del valor observado?
- ¿Cómo se desean formular las predicciones? como la mejor conjetura, como un rango razonable de posibles valores futuros que reflejan la incertidumbre asociada al pronóstico, como una distribución de probabilidad de los posibles valores futuros

- ¿Cuál es el horizonte de predicción?
- ¿Cuál es la información sobre la que se elaboran las predicciones?
- ¿Cuál es el método que se utiliza? ¿Es el más adecuado para el propósito trazado?.

La consideración sobre lo que es un buen pronóstico está asociada a la cuestión de cuán cerca está el pronóstico de los datos finalmente observados y por tanto qué métrica se selecciona para cuantificar esa distancia. La definición de la función de pérdida o costo del error de predicción está vinculada con el predictor que se utilizará para realizar la predicción y por supuesto esto está asociado a la metodología de modelización y predicción seleccionadas. Está demás advertir que si no se conoce la metodología, no se puede inferir sobre cuál es la mejor forma de evaluar esa predicción.

La forma como se decide formular las predicciones determinará la comunicación que se realice de las mismas y también la posibilidad de que quienes usan esas predicciones para tomar decisiones o informarse, puedan comprender los resultados del trabajo de modelización y predicción y convertirlos en información útil.

Otro aspecto refiere a dejar en claro el horizonte de predicción. No es lo mismo predecir a 7 días en adelante que a 30 o 40 días en adelante. Si la comunicación es clara, alcanza un simple cuadro, no habrá dudas al respecto.

Los aspectos relacionados a la predicción como los errores cometidos y la incertidumbre asociada a la predicción es diferente si la variable que se modeliza y predice es la misma o si la predicción es un resultado indirecto de la modelización.

Finalmente, explicitar el método que se utiliza para elaborar las predicciones y el fundamento de su elección contribuye a clarificar el entorno de la predicción.

Poder ubicar estos elementos en las metodologías presentadas en esta nota nos ayudarán a evaluar el proceso y a calidad del trabajo realizado.

Predicciones para el acumulados de personas contagiadas y fallecimientos por COVID-19 en Uruguay

En este bloque se presentan las predicciones provenientes de los modelos CDH, parte del funcionamiento del sistema de predicción y un análisis breve sobre las magnitudes de los errores de predicción. Todo ello es posible por la transparencia con que los autores manejan la información y por la actualización continua y cada pocas unidades de tiempo de las proyecciones. Este proceso se representa en los gráficos y los datos se incluyen en los cuadros. A través de este procedimiento, se incorporan las novedades que traen los nuevos datos y por tanto se fortalece la capacidad de interpretación temprana de los potenciales cambios de rumbo en la evolución de la epidemia.

En los cuadros y gráficos se presentan las predicciones que corresponden al número de contagios acumulados a través de los días y el número acumulado de fallecimientos en Uruguay. Se reportan predicciones 7 días en adelante, y las actualizaciones que se realizan días antes de culminar los 7 días predichos en la instancia previa.

En los cuadros 1 y 2 se presentan las predicciones en el punto ¹², resultantes de los modelos aplicados a los datos de Uruguay y en las figuras 1 y 2 se presentan esos resultados en forma gráfica.

Se representa la evolución prevista para los siguientes 7 días para diferentes puntos de arranque, los que comienzan en las siguientes fechas: 08/01, 12/01, 14/01, 18/01, 21/01, 26/01 y 29/01.

A través de los gráficos y cuadros se muestra el proceso de corrección y actualización de los pronósticos a medida que se incorpora nueva información. En especial, en los gráficos 1 y 2 se representa la trayectoria definida por cada nuevo pronóstico y los datos observados. Esta representación toma la forma de un “cabello florecido”, donde cada florecimiento es la nueva predicción con datos hasta un cierto período y “cada nuevo florecimiento” muestra cómo se corrige la trayectoria esperada de la variable al incorporar los nuevos datos observados.

Esa corrección nos muestra la importancia de la actualización continua de las predicciones, pues la corrección incorpora la nueva información que trae consigo cada dato observado al ser incorporado al modelo. La información nueva desencadena esa corrección y con ello permite detectar tempranamente los cambios que se están procesando y que podrán consolidarse en el correr de los días.

En el caso de Uruguay, para el mes de enero, esa corrección es relevante ya que se modifica la tendencia de crecimiento de los casos acumulados y/o muertes (ver el gráfico 2). Podemos observar, en cada caso como se modifica la pendiente de la curva, ya sea pronunciando la pendiente o reduciéndola como al final del período.

Este sistema de predicción incorpora otras de las ventajas que trae la actualización cada pocos días. Se reducen los errores y se corrigen rápidamente las trayectorias pronosticadas. En momentos donde el cambio se está procesando, esto constituye una señal a atender e interpretar.

En los cuadros además de las predicciones en el punto y los valores observados se agrega un indicador para medir el nivel de error de la predicción. En este caso se selecciona el error relativo $(\hat{Y}_t - Y_t)/Y_t$, por su sencillez en la interpretación. Da cuenta de la magnitud

¹²En la web se reportan los valores de las predicciones puntuales y en el gráfico se representan también los intervalos de confianza

relativa y el signo del error.

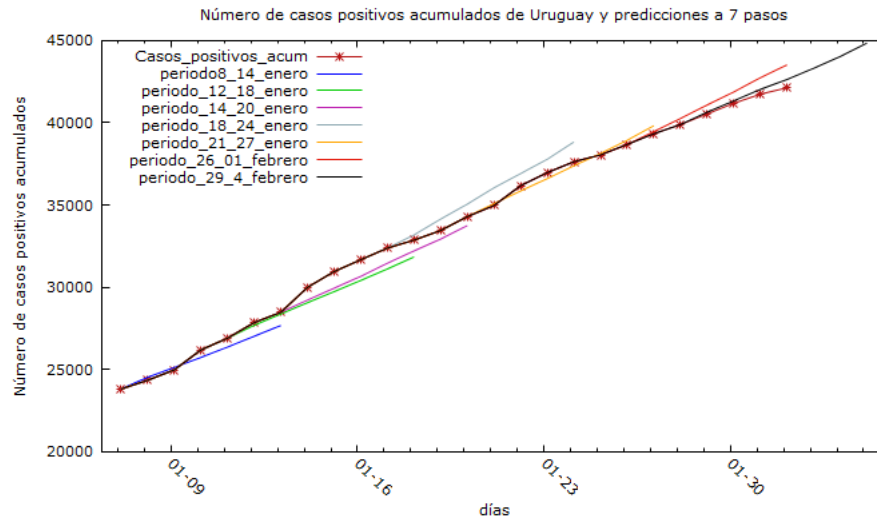


Figura 1: Fuente: SINAIE y www.doornik.com/COVID-19. Datos observados y Confirmed count forecast Latin America

Cuadro 1 - Número de Casos confirmados acumulados en Uruguay, predicciones a 7 días y errores relativos (en el cuadro ER)

2021	Dato	08-14/01	ER	12-18/01	ER	14-20/01	ER	18-24/01	ER	21-27/01	ER	26-01/02	ER	29-04/02	ER
07 enero	23807	23807													
8 enero	24339	24520	0.7												
9 enero	24974	25120	0.6												
10 enero	26186	25730	-1.7												
11 enero	26901	26360	-2.0												
12 enero	27846	27020	-3.0	27680	-0.6										
13 enero	28475	27670	-2.8	28370	-0.4										
14 enero	29989			29060	-3.1	29220	-2.6								
15 enero	30946			29730	-3.9	29940	-3.3								
16 enero	31669			30410	-4.0	30660	-3.2								
17 enero	32378			31110	-3.9	31460	-2.8								
18 enero	32863					32210	-2.0	33180	1.0						
19 enero	33446					32930	-1.5	34140	2.1						
20 enero	34294					33750	-1.6	35050	2.2						
21 enero	34965							36040	3.1	35080	0.3				
22 enero	36170							36910	2.0	35850	-0.9				
23 enero	36967							37780	2.2	36600	-1.0				
24 enero	37633									37380	-0.7				
25 enero	38041									38130	0.2				
26 enero	38680									38930	0.6	38670	-0.03		
27 enero	39328									39820		39460	0.34		
28 enero	39887											40260	0.94		
29 enero	40529											41080	1.36	40640	0.27
30 enero	41181											41870	1.67	41330	0.36
31 enero	41738											42740	2.40	42030	0.70
1 febrero	42128											43520	3.30	42630	1.19
2 febrero														43310	
3 febrero														44030	
4 febrero														44840	

Cuadro 2 - Número de fallecidos por COVID-19 en Uruguay, predicciones a 7 días y errores relativos (en el cuadro ER)

2021	Dato	08-14/01	ER	12-18/01	ER	14-20/01	ER	18-24/01	ER	21-27/01	ER	26-01/02	ER	29-04/02	ER
07 enero	231														
8 enero	240	239	-0.4												
9 enero	248	248	0.0												
10 enero	256	257	0.4												
11 enero	262	266	1.5												
12 enero	269	275	2.2	269											
13 enero	275	285	3.6	277	0.7										
14 enero	280	295	5.4	284	1.4	282	0.7								
15 enero	291			292	0.3	289	-0.7								
16 enero	298			300	0.7	297	-0.3								
17 enero	311			308	-1.0	304	-2.3								
18 enero	319			316	-0.9	312	-2.2	320	0.3						
19 enero	330					319	-3.3	327	-0.9						
20 enero	336					327	-2.7	335	-0.3						
21 enero	347							342	-1.4	343	-1.2				
22 enero	364							352	-3.3	353	-3.0				
23 enero	367							361	-1.6	361	-1.6				
24 enero	376							370	-1.6	371	-1.3				
25 enero	390									378	-3.1				
26 enero	401									387	-3.5	400	-0.25		
27 enero	407									394		409	0.49		
28 enero	415											418	0.72		
29 enero	425											428	0.71	424	-0.24
30 enero	431											436	1.16	434	0.69
31 enero	436											446	2.29	443	1.58
1 febrero	444											455	2.48	453	1.99
2 febrero														463	
3 febrero														473	
4 febrero														484	

Fuente: SINAIE y www.doornik.com/COVID-19

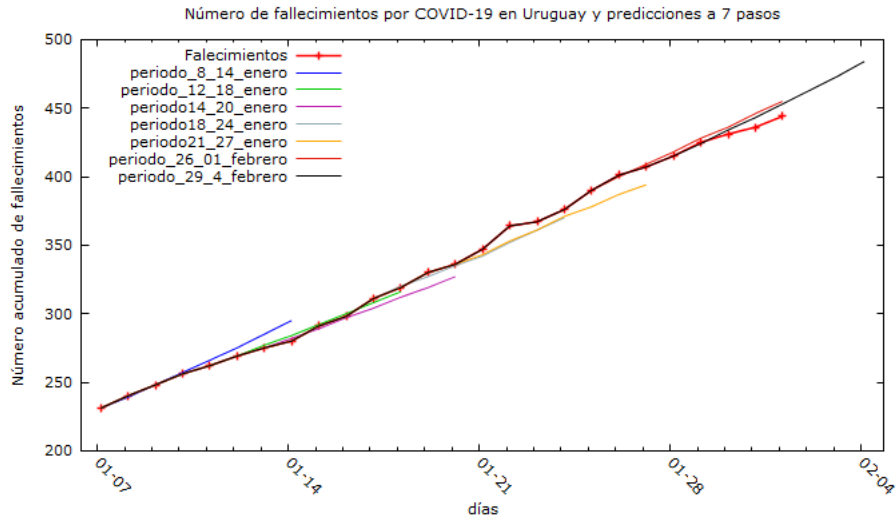


Figura 2: Fuente: SINAE y www.doornik.com/COVID-19. Datos observados y Deaths count average forecast Latin America

Como sabemos, los errores se incrementan a medida que nos alejamos del último dato observado. Es por ello que es interesante separar los errores a diferentes pasos para evaluarlos, y por ello en el Cuadro 3 se presentan los errores relativos para cada conjunto de predicciones agrupados por horizonte de predicción.

El error relativo a un paso, de mayor magnitud, para el caso del acumulado de contagios fue de 2,56 % para el período del 14 al 20 de enero y en el caso de total de fallecimientos, el máximo error relativo alcanza el 1,15 % para el período del 21 al 27 de enero.

En el horizonte a 6 días, donde se pueden esperar errores más altos que los presentados previamente, el error de mayor magnitud en la variable contagios acumulados fue de 3,92 % registrado en las predicciones correspondientes al período que va del 12 al 18 de enero. En el total de fallecimientos fue de 3,64 % registrado entre el 8 y el 14 de enero.

Los niveles de error a diferentes horizontes de predicción son bajos, por lo que con estos niveles de error, más la actualización permanente de las predicciones, este sistema de predicción se convierte en un instrumento muy útil para mirar unos días en adelante la evolución de la epidemia.

Cuadro 3 - Errores relativos de predicción^a. Modelo para el número acumulado de casos positivos

Errores a diferentes pasos

Período		1 paso	2 pa- sos	3 pa- sos	4 pa- sos	5 pa- sos	6 pa- sos
08/01 14/01	a	0.74	0.58	-1.74	-2.01	-2.97	-2.83
12/01 18/01	a	-0.60	-0.37	-3.10	-0.37	-3.98	-3.92
14/01 20/01	a	-2.56	-3.25	-3.19	-2.84	-1.99	-1.54
18/01 24/01	a	0.96	2.07	2.20	3.07	2.05	2.20
21/01 27/01	a	0.33	-0.88	-0.99	-0.67	0.23	0.65
26/01 01/02	a	-0.03	0.34	0.94	1.36	1.67	2.40

^a Elaboración propia en base a datos del SINAE y www.doornik.com/COVID-19

Cuadro 4 - Errores relativos de predicción^b. Modelo para el número de fallecidos

Errores a diferentes pasos

Período		1 paso	2 pa- sos	3 pa- sos	4 pa- sos	5 pa- sos	6 pa- sos
08-14/01		-0.42	0.00	0.39	1.53	2.23	3.64
12-18/01		0.73	1.43	0.34	0.67	-0.96	-0.94
14-20/01		0.71	-0.69	-0.34	-2.25	-2.19	-3.33
18-24/01		0.31	-0.91	-0.30	-1.44	-3.30	-1.63
21-27/01		-1.15	-3.02	-1.63	-1.33	-3.08	-3.49
26-01/02		-0.25	0.49	0.72	0.71	1.16	2.29

^bElaboración propia en base a datos del SINAE y www.doornik.com/COVID-19

Si los modelos generan resultados muy distanciados de la realidad de forma sistemática dejan de ser útiles, razón por la cual contar con la información actualizada de las predicciones o de los escenarios y las correcciones que se generan por la incorporación de nueva información es sumamente relevante tanto para poder evaluar las predicciones como para tener mayor información sobre el rumbo esperado de la epidemia, información importante en los momentos donde se procesan cambios.

Se puede pensar que estos modelos aunque son muy útiles para reducir la incertidumbre en el corto plazo y pueden colaborar a la hora de pensar nuevas medidas o mantener las vigentes, no son adecuados para evaluar los pasos a seguir en un horizonte de mediano plazo, lo cual es cierto. Pero en momentos donde la evolución de la epidemia sufre cambios, los pronósticos de corto plazo son fundamentales, porque los pueden detectar rápidamente. Cuando los contagios y muertes se estabilizan los pronósticos a mediano plazo son útiles. En esta epidemia los contagios se dan producto de un virus desconocido, del que se van conociendo sus características a medida que transcurre la epidemia. Reducir la incertidumbre a corto plazo es un insumo sumamente relevante.

¿Cuáles son las predicciones más recientes provenientes de los modelos CDH?

La última actualización al momento de escribir esta nota, comprende el período 9 a 15 de febrero. Los modelos pronostican que al 15 de febrero el acumulado de casos positivos alcanzaría los 48980 casos. Recordemos que al día 9 de febrero el dato observado es de 46143 casos. Estas cifras dan lugar a una pequeña corrección a la baja de la trayectoria de los casos acumulados respecto a la proyección anterior y una tasa de crecimiento promedio de los 7 días dos décimas por debajo de la proyección anterior. El error relativo promedio para el día 9 fue de $-0,18\%$

En cuanto a cantidad de defunciones, al 9 de febrero se acumulan 506 defunciones. Los modelos CDH pronostican para el 15 de febrero que el acumulado de fallecimientos podría llegar a 550. Estas nuevas predicciones corrigen a la baja las anteriores en términos de número de fallecimientos y corrigen tres décimas a la baja la tasa de crecimiento promedio para los 7 días de pronóstico. El error relativo del día 9 fue de $-0,4\%$.

Cuadro 5 - Pronóstico para el número acumulado de casos positivos del 9 al 15 de febrero

Fecha	Pronóstico de casos positivos
9 de febrero	46060
10 de febrero	46540
11 de febrero	47040
12 de febrero	47530
13 de febrero	48010
14 de febrero	48510
15 de febrero	48980

Fuente: www.doornik.com/COVID-19

Cuadro 6 - Pronóstico del número acumulado de fallecimientos del 9 al 15 de febrero

Pronóstico

Fecha	Pronóstico de fallecimientos
9 de febrero	504
10 de febrero	511
11 de febrero	519
12 de febrero	526
13 de febrero	534
14 de febrero	542
15 de febrero	550

Fuente: www.doornik.com/COVID-19

Modelos de predicción a mediano plazo

Castle, Doornik y Hendry exploran la posibilidad de crear modelos de predicción para el mediano plazo. Para ello utilizan la información que se generó en China, donde la epidemia ha recorrido las fases de inicio, desarrollo y control de la epidemia. Exploran si esta experiencia recorrida por China es utilizable para otros países. Pero esta aproximación metodológica quedará para otra nota metodológica futura.

Comentarios finales

Disponer de predicciones lo más precisas posibles sobre la trayectoria de algunas variables que informan sobre la evolución de la epidemia en el corto plazo, pueden contribuir a reducir la incertidumbre y tener una referencia sobre lo que es probable esperar en un futuro próximo. Esas predicciones pueden ser utilizadas como insumos para la definición de la respuesta que la autoridad sanitaria puede dar frente a la evolución que se espera tome la epidemia en un futuro próximo y con ello eventualmente, determinar cambios de rumbo en la estrategia a seguir para el control de la expansión del virus.

Los equipos internacionales encargados de hacer seguimiento de la epidemia a través de los datos y de la construcción de modelos, han desarrollado múltiples estrategias cuyos resultados aportan información diferente, con grados de precisión y de error diferentes. De las dos aproximaciones que se mencionan en esta nota, un tipo de estrategia de modelización genera información que fundamentalmente permite construir escenarios futuros posibles bajo ciertos supuestos y otros producen un conjunto de valores esperados sobre la trayectoria futura, también apoyados en supuestos, pero otros supuestos.

Los horizontes futuros sobre los que se genera información, ya sea escenarios como predicciones son variados. En esta nota nos concentramos en aquellos que generan predicciones de corto plazo.

Los modelos CDH vienen brindando predicciones en tiempo real desde mediados de marzo de 2020 para un conjunto de países. CDH han evaluado sus niveles de error y se han comparado, para el caso de Inglaterra, con los modelos epidemiológicos¹³ y han detectado menores errores de predicción en el corto plazo que los resultantes de los modelos epidemiológicos, pero en el mediano plazo las conclusiones se revierten, por lo que concluyen que ambas estrategias son útiles y complementarias.

¹³Ver Nota metodológica de junio de 2020, <https://hdl.handle.net/20.500.12008/25413>

En Uruguay no es posible hacer este tipo de comparaciones. La información pública disponible, resultante de la modelización realizada por el equipo del GACH encargado de estos temas es muy escasa, se brinda muy esporádicamente y en ninguna oportunidad se cuenta con una tabla con las proyecciones a un horizonte de tiempo claramente definido y tampoco se actualiza la información, con lo cual es imposible sistematizar los resultados y hacer un seguimiento y análisis de los mismos.

A partir de enero Uruguay se integra a la base de datos donde se reportan las predicciones para los siguientes 7 días que producen los modelos CDH. Estas predicciones son públicas y se publican en la web indicada en esta nota. Estos pronósticos se revisan con una periodicidad aproximada de 4 días.

Dado que esta información no es muy conocida pero el sistema de predicción montado tiene atributos deseables, en esta nota se pretende darles visibilidad y subrayar las características del procedimiento de predicción que utilizan.

Aspectos relevantes como la explicitación de la metodología utilizada, de la variable que se modeliza y predice, del horizonte de predicción, la periodicidad de la actualización y por supuesto la información sobre las predicciones pone a esta forma de proceder como un ejemplo a seguir y desnuda carencias en la forma como se procede en Uruguay.

Referencias bibliográficas

- Castle,J.; Doornik,J; Hendry,D. (2020) “Short-term forecasting of the Coronavirus Pandemic- 2020-05-14”. Nuffield College.2020 W06.
https://www.nuffield.ox.ac.uk/economics/Papers/2020/2020W06_COVID-19_shortterm_forecasts.pdf
- Castle,J.; Doornik,J; Hendry,D. (2020) “Short-term forecasting of the Coronavirus Pandemic”. 'IJF special section: Epidemics and forecasting with focus on COVID-19'.
<https://forecasters.org/blog/2020/04/30/short-term-forecasting-of-the-coronavirus-p>
- Castle,J.; Doornik,J; Hendry,D. (2020) “Medium-term Forecasting of the Coronavirus Pandemic”.
https://www.doornik.com/COVID-19/doc/COVID-19_mediumterm_forecasts.pdf
- Castle,J.; Hendry,D. (2019) “ Modelling our Changing World”. Palgrave Text in Econometrics. Palgrave Macmillian.
- Diebold,F. (2007) “Elements of Forecasting”. Fourth Edition. Thomson South-West.

- Doornik JA, Castle JL, Hendry DF (2020) “Statistical Short-term Forecasting of the COVID-19 Pandemic”. *J Clin Immunol Immunother* 6: 046.
- Castle,J.; Doornik,J; Hendry,D. (2019) “Some forecasting principles from the M4 competition”. Nuffield College.2019 W01.
<https://www.nuffield.ox.ac.uk/economics/Papers/index.html#2019>
- Castle,J.;Clements, M.; Hendry,D. (2014) “Robust Approaches to Forecasting. University of Oxford. Department of Economics”. Working Papers 697. January 2014.
<https://www.economics.ox.ac.uk/department-of-economics-discussion-paper-series/robust-approaches-to-forecasting>
- Doornik,J; Castle,J.; Hendry,D. (2020) “Card forecasts for M4”. *International Journal of Forecasting* 36.
- Doornik,J. (2009) *Autometrics . The Methodology and Practice of Econometrics*. Edited by Jenifer Castle y Neil Shepard. Oxford University Press
Hendry, D. F. (1999) An econometric analysis of US food expenditure, 1931?1989.
- GACH-Área Modelos y Ciencia de Datos. “Análisis de la situación de la epidemia en Uruguay. Informe del 12 de diciembre de 2020”.
https://medios.presidencia.gub.uy/11p_portal/2020/GACH/INFORMES/expertos-modelos-ciencia-de-datos-12-diciembre-2020.pdf
- GACH-Área Modelos y Ciencia de Datos.“Análisis de situación de la epidemia en Uruguay al 18 de enero”.
https://medios.presidencia.gub.uy/11p_portal/2021/GACH/INFORMES/analisis_situacion_final_18_01_21.pdf
- Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. *The Methodology and Practice of Econometrics*. Edited by Jenifer Castle y Neil Shepard. Oxford University Press.
- Rodríguez Collazo,S.(2020) . “Modelos de series temporales adaptativos y robustos para predecir contagios y muertes por COVID-19: la respuesta de Castle, Doornik y Hendry”. Notas de reflexión. Instituto de Estadística de la Facultad de Ciencias Económicas y Administración. Junio 2020. <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/25413>