

Análisis de Eventos en los Casos Confirmados y Muertes por COVID-19 Utilizando el Modelo Generalizado de Bass

Fernández-Durán, Juan José y Gregorio-Domínguez, María Mercedes
ITAM

Resumen

El modelo original de difusión de Bass (Bass, 1969) es uno de los modelos más utilizados en Mercadotecnia para estudiar la difusión de nuevos productos en un mercado. Se basa en la especificación de una tasa de riesgo que es proporcional al total de individuos que ya han adquirido el producto de tal forma que es equivalente a un modelo de contagio y es un caso particular de la densidad sesgada de Gompertz. Posteriormente, Bass, Krishnan y Jain (1994) modificaron el modelo original para incluir variables explicativas tales como el precio y el gasto en publicidad. A dicho modelo se le conoce como el modelo generalizado de Bass. En este trabajo se utiliza el modelo generalizado de Bass para estudiar las características de las series de tiempo del número de casos confirmados y muertes producidos por el coronavirus COVID-19 e identificar variables explicativas relevantes en la reducción o incremento en el número de casos confirmados las cuales están relacionadas a las políticas públicas implementadas. En el caso de las muertes producidas por el COVID-19 se encuentran además efectos estacionales significativos los cuales son estudiados con el modelo desarrollado por Fernández-Durán (2014).

1. Modelo Generalizado de Bass con Efectos Estacionales

El modelo de difusión de Bass considera la variable aleatoria del tiempo que tarda un cliente en adquirir un nuevo producto (duradero), T , para la cual define su tasa de riesgo como

$$h_T(t)dt = P(t < T < t+dt \mid T \geq t) = \frac{f_T(t)dt}{F_T(t)} = p+qF_T(t) = p+\frac{q}{m}Y(t) \quad (1)$$

donde $h_T(t)dt$ es la probabilidad de que un individuo adquiriera por primera vez el producto en el intervalo $(t, t+dt)$ la cual es proporcional al porcentaje de individuos que han adquirido el producto hasta el tiempo t , $F_T(t) = \frac{Y(t)}{m}$ donde $Y(t)$ es el total de individuos que han adquirido el producto al tiempo t y m es el tamaño total del mercado, es decir, el total de individuos que acabarán adquiriendo el producto. El parámetro p representa la tasa de innovadores, es decir, los individuos que compran el producto sin haber recibido alguna recomendación de otros individuos y, q es la tasa de imitadores que son individuos que adquieren el producto sólo si les es recomendado por otros individuos. Al traducir el modelo de Bass en términos de los contagios por COVID-19, el parámetro p representa la tasa de individuos que fueron contagiados fuera de la población, por ejemplo, los extranjeros contagiados y, q , representa la tasa de contagios locales. El parámetro m representa el total de individuos en la población que serán contagiados (o al menos confirmados con el virus), $Y_T(t)$ el total de contagiados (confirmados) al tiempo t , $F_T(t)$ el porcentaje de contagiados (confirmados) al tiempo t del total m de contagiados (confirmados). La función de distribución, $F_T(t)$ satisface

$$F_T(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \quad (2)$$

Posteriormente, Bass, Krishnan y Jain (1994) modificaron el modelo de Bass para incluir el efecto de otras variables tales como el precio y el gasto en publicidad. A este modelo se le conoce como el modelo generalizado de Bass y está definido como

$$h_T(t) = (p + qF_T(t))x(t) \quad (3)$$

donde $x(t)$ es una función positiva de variables explicativas. La función de distribución del modelo de Bass generalizado, $F_T(t)$ satisface

$$F_T(t) = \frac{1 - e^{-(p+q)(X(t)-X(0))}}{1 + \frac{q}{p}e^{-(p+q)(X(t)-X(0))}} \quad (4)$$

donde $X(t) = \int_0^t x(s)ds$ es el valor acumulado de la función $x(t)$. Para el estudio de eventos en el número de confirmados y muertes por COVID-19 se pueden incluir en la función $x(t)$ variables explicativas relacionadas a las políticas públicas implementadas contra el COVID-19 tales como el inicio del confinamiento general, el inicio de etapas de regreso a la normalidad, el inicio de la movilidad generalizada para toda la población, entre otras. En particular, la existencia de efectos estacionales inducidos por el comportamiento de los individuos de la población se puede introducir en el modelo generalizado de Bass mediante el uso de sumas trigonométricas no negativas (Fernández-Durán, 2014) donde la función $x(t)$ está definida como

$$x(t) = 1 + \sum_{k=1}^n (a_k^* \cos(2\pi kt) + b_k^* \sin(2\pi kt)). \quad (5)$$

El uso de variables explicativas relacionadas a políticas públicas y de efectos estacionales se pueden combinar en la función $x(t)$ para separar cada uno de los efectos. El ajuste de los modelos se lleva a cabo en el software estadístico *R* mediante el uso de los paquetes *CircNNTSR* (Fernández-Durán y Gregorio-Domínguez, 2012 y 2016) para la inclusión de efectos estacionales y *minpack.lm* (Elzhov et al., 2016) para la estimación de los parámetros mediante el algoritmo de Levenberg-Marquardt para mínimos cuadrados no lineales.

2. Estudio de Eventos en España

Se utilizaron los datos sobre casos confirmados en España que se obtuvieron del sitio de internet <https://ourworldindata.org/> el 11 de junio de 2020. La base contiene el número de casos confirmados y muertes del 1 de febrero donde se tuvo el primer caso confirmado al 11 de junio. Se seleccionaron los datos de España ya que se trata de un país donde prácticamente para el 11 de junio se está dando el fin de esta ola de contagios además de tratarse de un país con cultura similar a la de México. Se ajustó el modelo de Bass

generalizado con y sin efectos estacionales del día de la semana considerando las siguientes variables explicativas relacionadas a los siguientes eventos:

1. 26 de febrero: Primer caso local, no importado
2. 4 de marzo: Primera muerte por COVID-19
3. 10 de marzo: Suspensión de las Fallas de Valencia
4. 12 de marzo: Suspensión de la Liga de fútbol
5. 13 de marzo: Inicio del estado de alarma
6. 14 de marzo: Confinamiento total
7. 3 de abril: Día con el máximo número de muertes en un día
8. 10 de mayo: Inicio de la fase 1 de regreso a la normalidad en varias Comunidades Autónomas
9. 25 de mayo: Inicio de la fase 2 de regreso a la normalidad en varias Comunidades Autónomas (Madrid y Barcelona pasan a la fase 1)

Para estos eventos se generaron variables indicadoras con valores posibles de cero y uno. Por ejemplo, para la variable indicadora asociada al inicio del confinamiento total, toma el valor de cero para los días del 1 de febrero al 13 de marzo, el valor 1 del 14 de marzo al 9 de mayo y, el valor cero del 10 de mayo (inicio de la fase 1) al 11 de junio. Para la variable indicadora del máximo de muertes, toma el valor de 0 para los días del 1 de febrero al 2 de abril, el valor 1 para los días del 4 de abril al 11 de junio.

La Figura 1 presenta el ajuste del modelo de Bass sin efectos estacionales y sin variables explicativas (modelo original de Bass) a los casos confirmados acumulados diarios. Como se puede observar, este modelo presenta un mal ajuste a los datos con periodos en donde el modelo pronostica consistentemente un mayor (menor) número de casos confirmados acumulados.

Al comparar diferentes modelos generalizados de Bass con y sin efectos estacionales y con diferentes combinaciones de variables explicativas, el modelo que presenta el mejor ajuste a los datos es un modelo generalizado de Bass sin efectos estacionales con dos variables explicativas que corresponden a la variable indicadora del inicio del confinamiento con un efecto rezagado

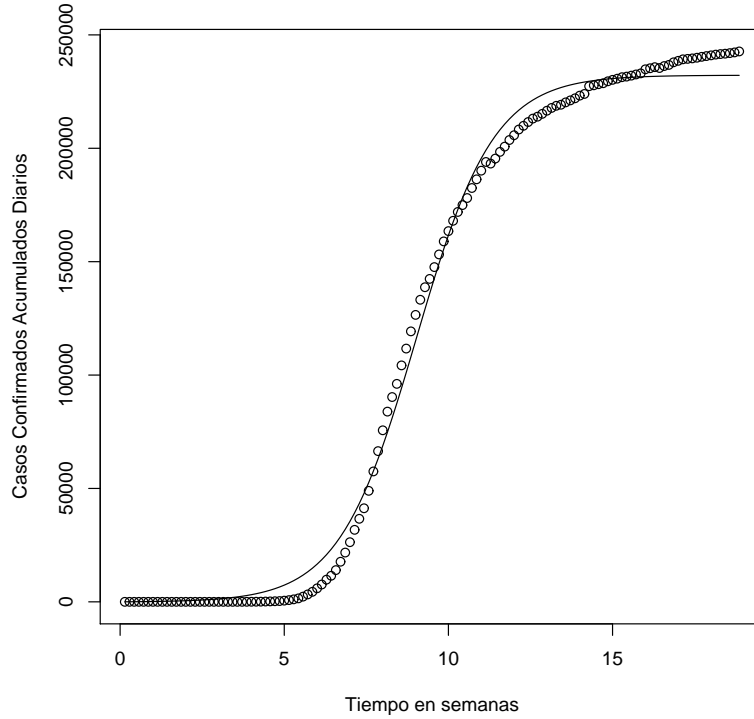


Figura 1: Modelo de Bass sin efectos estacionales y sin variables explicativas ajustado a los casos confirmados diarios en España (1-Feb-2020 a 11-Jun-2020). Los puntos representan los datos observados y la línea el modelo ajustado.

de 13 días y la indicadora del día con el máximo número de muertes. La Figura 2 presenta el ajuste del modelo.

Los parámetros estimados del mejor modelo así como sus errores estándar se incluyen en el Cuadro 1. A partir de estas estimaciones se puede obtener el tiempo al pico de confirmados diarios que es estimado en 8.27 semanas. El valor observado del tiempo al pico fue igual a 8.00 semanas. Las estimaciones de los parámetros nos enseñan que el parámetro estimado p es muy pequeño en relación al parámetro estimado q por lo que el modelo se aproxima a un modelo logístico, el número total de casos estimado es igual a 238400 (el valor observado al 11 de junio es de 242707), los signos de los coeficientes

estimados tanto para la variable indicadora del confinamiento con un rezago de 13 días como la variable indicadora del día con el máximo número de muertes implican una reducción en el valor de la estimación del parámetro q por lo que implican una reducción en la tasa de contagios, es decir, después de 13 días de iniciado el confinamientos se obtiene la mayor reducción en la tasa de contagios y, después del día con el máximo número de muertes se vuelve a reducir la tasa de contagios. Nótese que la noticia del día con el máximo número de contagios implica que los días subsecuentes presentaran una tendencia a la baja en el número de muertes.

Parámetro	Estimación	Error Estándar	Valor t	valor p
p	1.052e-05	2.546e-06	4.131	6.51e-05 ***
q	1.379e+00	3.530e-02	39.071	<2e-16 ***
m	2.384e+05	4.136e+02	576.346	<2e-16 ***
confinamiento con rezago de 13 días	-2.702e-01	2.817e-02	-9.593	<2e-16 ***
día con el máximo de muertes	-3.400e-01	2.296e-02	-14.806	<2e-16 ***

Cuadro 1: Mejor modelo para España: Estimaciones de los parámetros, errores estándar, estadísticos t y valores p.

3. Estudio de Eventos en México

Al considerar el modelo final para España con los datos de casos confirmados diarios en México, encontramos que sólo cuando se hizo el anuncio de la tercera fase de confinamiento se tienen resultados significativos. La tercera fase de confinamiento empezó el 21 de abril. La Tabla 2 presenta los resultados de este modelo que corresponde a un modelo de Bass generalizado sin efectos estacionales e incluyendo la variable indicadora de la tercera fase del confinamiento que es igual a cero para los días anteriores al 21 de abril e igual a uno para los días del 21 de abril en adelante.

A partir del modelo final para México se obtiene que la fase 3 del confinamiento redujo la tasa de contagios q de 0.7698 a 0.3433. El número proyectado del total de contagios m es estimado con el valor de 236,200. El 12 de junio se tenían un total acumulado de 133,974 casos confirmados.

La Figura 3 incluye los datos acumulados así como las predicciones del model final.

Parámetro	Estimación	Error Estándar	Valor t	valor p
p	8.262e-05	1.555e-05	5.313	6.26e-07 ***
q	7.698e-01	2.792e-02	27.576	<2e-16 ***
Mmax	2.362e+05	3.954e+03	59.750	<2e-16 ***
fase 3 del confinamiento	-4.265e-01	2.231e-02	-19.120	<2e-16 ***

Cuadro 2: Mejor modelo para México: Estimaciones de los parámetros, errores estándar, estadísticos t y valores p.

4. Estacionalidad en el número de muertes en CDMX

Los mejores modelos para el número acumulado de casos confirmados para México y España resultaron ser modelos generalizados de Bass sin efectos estacionales. El siguiente análisis presenta como en los datos sobre el acumulado de muertes se presenta estacionalidad severa que requiere ser modelada mediante un modelo de Bass con efectos estacionales. Se analiza la serie de muertes acumuladas para CDMX del 18 de marzo (primera muerte reportada) al 12 de junio reportadas en la página web sobre el COVID-19 de la Dirección General de Epidemiología de la Secretaría de Salud de México (www.coronavirus.gob.mx).

La Figura 4 presenta el ajuste de un modelo de Bass sin efectos estacionales donde claramente se observan los efectos estacionales que no son capturados por el modelo.

Cuando se considera un modelo de Bass con efectos estacionales (Fernández-Durán, 2014) se obtiene un mejor ajuste a los datos (ver la Figura 5).

5. Conclusiones

El presente trabajo demuestra la necesidad de incluir el efecto de variables explicativas relacionadas a los eventos sobre políticas públicas llevadas a cabo para combatir la pandemia de COVID-19. Para el caso de casos confirmados en España, el confinamiento presenta un efecto rezagado de 13 días. La otra variable importante es el día en el cual ocurre el número máximo de muertes. En el caso de México, en el cual todavía no se ha desarrollado por completo la pandemia, se encuentra que sólo hasta la tercera fase del confinamiento se presenta una reducción en la tasa de contagios. En ninguno de los

modelos para México y España fue necesaria la inclusión de efectos estacionales, sin embargo, para el caso del número acumulado de muertes en CDMX, sí se presentan efectos estacionales importantes que deben ser considerados en los modelos predictivos a ajustar. Como investigación futura se considera la inclusión de variables explicativas que puedan modelar la reacción en el comportamiento de los habitantes ante diferentes políticas públicas implementadas por los gobiernos durante una pandemia o inclusive ante noticias dadas en los principales medios de comunicación y redes sociales.

Agradecimientos

Los autores agradecen a la Asociación Mexicana de Cultura A.C. por su apoyo para la elaboración de este trabajo.

Referencias

- [1] Bass, F.M. (1969) A New Product Growth for Model Consumer Durables. *Management Science*, 15-5, pp. 215-227.
- [2] Bass, F.M., Krishnan, V. and Jain, D.C. (1994) Why the Bass Model Fits without Decision Variables. *Marketing Science*, 13-3, pp. 203-223.
- [3] Elzhov, T.V., Mullen, K.M., Spiess, A-N. y Bolker, B. (2016) minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds, R package version 1.2-1, <https://CRAN.R-project.org/package=minpack.lm>
- [4] Fernández-Durán, J.J. and Gregorio-Domínguez, M.M. (2012). CircNNTSR: An R Package for the Statistical Analysis of Circular Data Using Nonnegative Trigonometric Sums (NNTS) Models. R package version 2.0. <http://CRAN.R-project.org/package=CircNNTSR>
- [5] Fernández-Durán, J.J. (2014) Modeling Seasonal Effects in the Bass Forecasting Diffusion Model. *Technological Forecasting and Social Change*, 88, pp. 251-264.

- [6] Fernández-Durán, J.J. and Gregorio-Domínguez, M.M. (2016) CircNN-TSR: An R Package for the Statistical Analysis of Circular, Multivariate Circular, and Spherical Data Using Nonnegative Trigonometric Sums, *Journal of Statistical Software*, 70-6.
- [7] R Development Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [8] [https : //ourworldindata.org/](https://ourworldindata.org/) Consultado el 11 de junio de 2020.
- [9] www.coronavirus.gob.mx Consultado el 12 de junio de 2020.

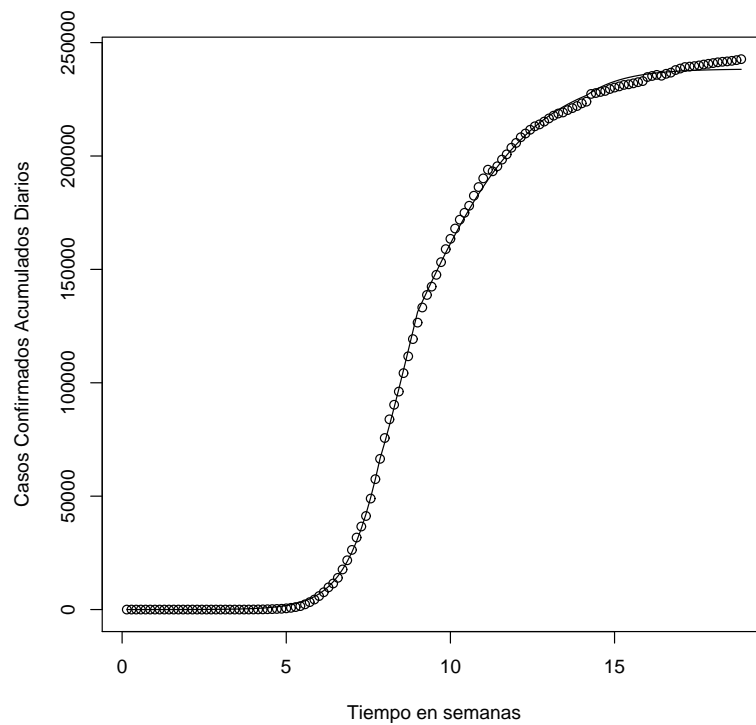


Figura 2: Mejor modelo para los datos de España: Modelo de Bass sin efectos estacionales y con dos variables explicativas que corresponden a la indicadora del inicio del confinamiento con un efecto rezagado de 13 días y la indicadora del día con el máximo número de muertes (1-Feb-2020 a 11-Jun-2020). Los puntos representan los datos observados y la línea el modelo ajustado.

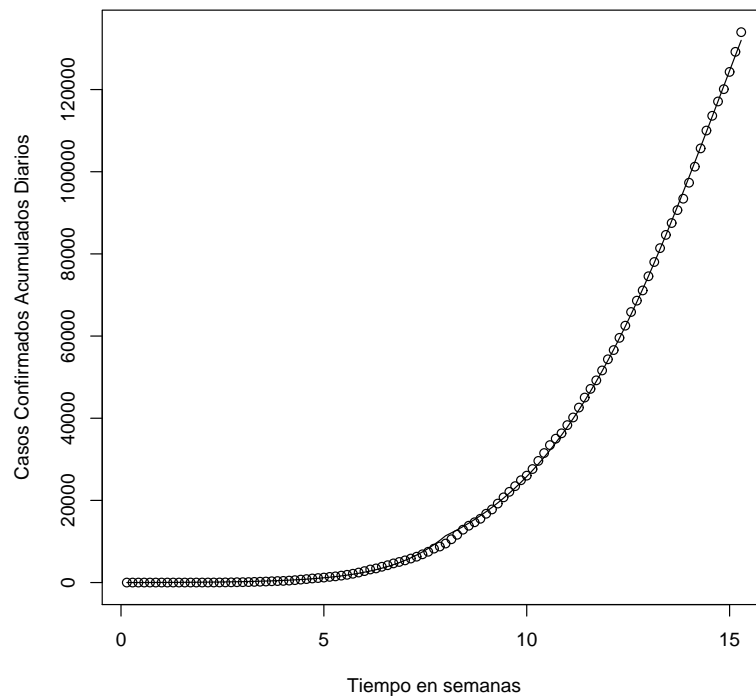


Figura 3: Mejor modelo para los datos de México: Modelo de Bass sin efectos estacionales y con una variable explicativa que corresponde a la indicadora del inicio de la tercera fase del confinamiento el 21 de abril (1-Feb-2020 a 11-Jun-2020). Los puntos representan los datos observados y la línea el modelo ajustado.

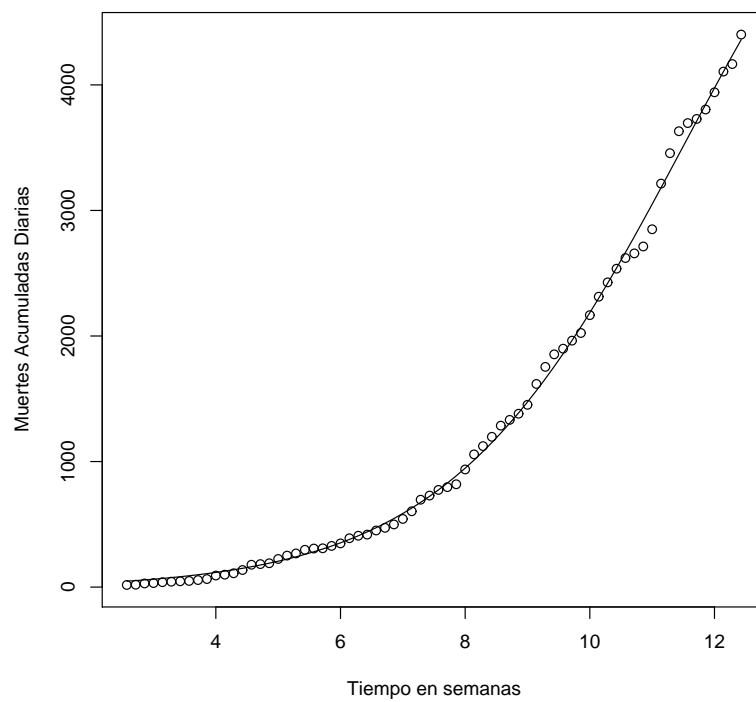


Figura 4: Modelo de Bass sin efectos estacionales para el número acumulado de muertes en CDMX (18-Mar-2020 a 12-Jun-2020). Los puntos representan los datos observados y la línea el modelo ajustado.

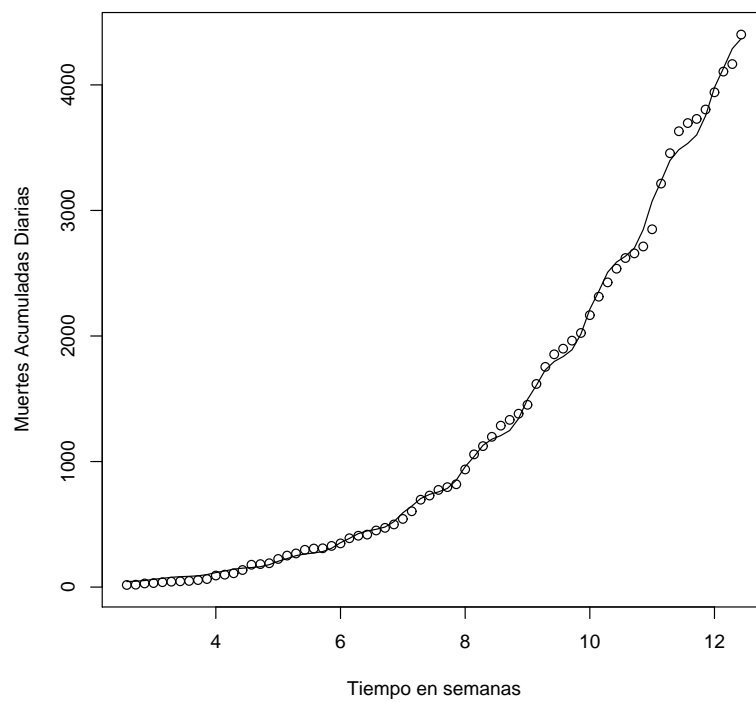


Figura 5: Modelo de Bass sin efectos estacionales para el número acumulado de muertes en CDMX (18-Mar-2020 a 12-Jun-2020). Los puntos representan los datos observados y la línea el modelo ajustado.