



La Econometría y Aprendizaje Computacional como Herramientas de la Investigación Económica

Juan Quiñónez
jquinonez@secmca.org

Resumen

La economía es una ciencia social que ha tomado prestado recursos de otras ciencias para probar conceptos teóricos y desarrollar herramientas que han colaborado con el desarrollo de esta y otras ciencias. El incremento en el volumen de datos, conjugado con el incremento en la capacidad de análisis de éste, ha dado como resultado el desarrollo de nuevos profesionales enfocados en idear nuevos métodos de extracción de datos, manejo de bases de datos, visualización e identificación de patrones. En esta nota se recopilan diferentes trabajos de investigación donde se utilizan modelos econométricos y de aprendizaje computacional. Asimismo, se destaca el aporte de la ciencia de datos, y sus herramientas, en el desarrollo de la economía como ciencia.

Introducción

La economía es una ciencia social que ha tomado prestado recursos de otras ciencias para probar conceptos teóricos y desarrollar herramientas que han colaborado con el desarrollo de esta.

Los conceptos presentados por Adam Smith, David Ricardo y otros autores contemporáneos, basaban sus teorías en la observación y evidencia empírica. Dada la cantidad de controversias y diferencias de opiniones que surgieron con el avance de esta ciencia, Marshall (1890) y Keynes (1891) destacan la necesidad de que las opiniones de sus pares economistas deben estar respaldadas por un análisis cualitativo y cuantitativo. A partir del 1980 se produjo un fortalecimiento en los datos cuantitativos, mejorando la calidad y cantidad de la información que se publicaba, y la fundación de entidades que consolidaban los avances de la estadística como ciencia. Más detalles sobre la historia de la relación de las estadísticas y la economía son expuestos por Persons (1925).

Con el pasar de los años, la colaboración entre los profesionales de economía y estadísticas se hizo más estrecha. Durante un encuentro de la Sociedad Americana de Economía (*American Economic Society*) y la Asociación Americana de Estadística (*American Statistical Association*) se organizó una reunión de 16 profesionales quienes fundan lo que luego se convertiría la Sociedad de Econometría (*the Econometric Society*). Bjerkholt (1995) resume la historia de esta sociedad.



Notas Económicas Regionales

Secretaría Ejecutiva del Consejo Monetario Centroamericano

No. 142, mayo 2023

La primera definición formal de econometría¹ fue acuñada por Frisch (1971):

Intermedia entre las matemáticas, la estadística y la economía, encontramos una nueva disciplina que, a falta de un nombre mejor, puede llamarse econometría. El objetivo de la econometría es someter las leyes abstractas de la economía política teórica o economía "pura" a una verificación experimental y numérica, y convertir así la economía pura, en la medida de lo posible, en una ciencia en el sentido estricto de la palabra. (Frisch, 1971, p. 386)

En la actualidad la econometría combina una batería de herramientas que responden a diferentes preguntas de investigación, apoyando la toma de decisiones de política monetaria, formación de precios y estudios de mercado, entre otras áreas.

Con los avances tecnológicos, la recolección de datos ha mejorado y dado a conocer nuevos problemas e inquietudes. Se estima que para el 2025 el volumen mundial de datos alcance los 181,869 exabytes² de los cuales el 90% está compuesto por datos sin estructura (como textos, información de redes sociales, imágenes, videos, entre otros). Este incremento en el volumen de datos, conjugado con el aumento en la capacidad de análisis de esta, ha dado como resultado el desarrollo de nuevos profesionales enfocados en idear nuevos métodos de extracción de datos, manejo de bases de datos y visualización e identificación de patrones.

Por esta línea se ha desarrollado la rama de aprendizaje computacional (*machine learning*) la cual puede ser descrita como un conjunto de herramientas y métodos que buscan inferir y extraer hallazgos de datos (Gutiérrez, 2015). Las herramientas de esta rama tienen la particularidad de ser capaces de "aprender" y mejorar sus propios algoritmos sin ser explícitamente programados (Pagliaran, 2019). Esta definición fue formalizada por Mitchell (1997):

Se dice que un programa informático aprende de la experiencia E con respecto a alguna clase de tarea T y rendimiento P , si su rendimiento en la tarea en T , medido por P mejora con la experiencia E . (Mitchell, 1997, p. 2)

El ejemplo clásico para aterrizar esta definición es un programa de clasificación de e-mails como "correo basura" (*spam*). Si se desarrolla un programa para esta tarea y el mismo logra mejorar la clasificación de estos correos mientras aumenta la cantidad de e-mails que reciba, entonces el mismo cumple con la definición de aprendizaje computacional. Esto descubre uno de los grandes atractivos del aprendizaje computacional: logra descubrir patrones generalizables los cuales no necesariamente deben ser especificados por el investigador (Mullainathan & Spiess, 2017).

Ambas ciencias utilizan datos como insumos para su desarrollo de modelos. Entre los diferentes tipos de datos se destacan los de corte transversal, los cuales capturan distintas características de

¹ El primer uso de la palabra "econometría" fue realizado por Ciompa (1919), como referencia al estudio de la estadística y a la contabilidad.

² La información es almacenada en bytes y las unidades de medida incrementan en 1,000. Un exabyte equivale a 1,000 millones de gigabytes o $1e^{18}$ bytes



Notas Económicas Regionales

Secretaría Ejecutiva del Consejo Monetario Centroamericano

No. 142, mayo 2023

diferentes observaciones para un mismo momento en el tiempo. Para este tipo de datos es común los estudios de correlaciones y relaciones causales y ejercicios de clasificación. Otro tipo de datos son los llamados series de tiempo, los cuales capturan una característica de una observación a lo largo del tiempo. Este tipo de datos responde a cambios de tendencia, ciclicidad y estacionalidad del tiempo, por lo que requiere un tratamiento especial, de acuerdo con el estudio que se realice.

En esta nota se recopilan diferentes trabajos de investigación donde se utilizan modelos econométricos y de aprendizaje computacional. Estos trabajos muestran que el uso conjunto de ambos enfoques (tanto de modelos econométricos como de modelos de aprendizaje computacional) pueden proveer más información a la hora de realizar un estudio. Asimismo, se destaca el aporte de la ciencia de datos, y sus herramientas, en el desarrollo de la economía como ciencia.

Revisión de literatura reciente

En esta sección se revisa una muestra de la literatura publicada por los Bancos Centrales de la región de Centroamérica y República Dominicana y se resalta el uso de modelos de ambos enfoques.

Los resultados de los modelos de aprendizaje computacional se pueden dividir en dos grupos: clasificación (donde el resultado es una categoría) y regresión (donde los resultados pueden ser variables continuas). Aunque las técnicas de esta rama se pueden aplicar para ambos grupos, es conveniente dividir los resultados de esta forma para su comparación con los modelos econométricos.

Sobre los modelos de clasificación se destaca el documento donde de Óleo & Ramírez (2021) estudian los determinantes de la desocupación en la región Norte de la República Dominicana utilizando modelos logísticos. En esta categoría de resultados Bournigal & García (2022) comparan el desempeño de diferentes técnicas de clasificación (Logit, XGBoost, Random Forest, Support Vector Machine, entre otros) para determinar la probabilidad de incumplimiento de una cartera de crédito. Este estudio compara el rendimiento de modelos econométricos contra algunas técnicas de clasificación de aprendizaje computacional.

Para los Bancos Centrales que han adoptado el Esquema de Metas de Inflación (EMI) es de vital importancia contar con un sistema robusto de proyecciones de variaciones de precios. En este rubro se destaca el ejercicio realizado por Ramírez & Ramírez (2019) donde utilizan diferentes modelos econométricos para proyectar variaciones de precios de los artículos que componen el Índice de Precios al Consumidor de la República Dominicana y luego agregan las mismas utilizando sus respectivas ponderaciones para obtener una proyección de la inflación total. Rodríguez (2020) realiza un ejercicio de pronóstico de inflación de corto plazo de Costa Rica comparando modelos univariados autorregresivos contra una batería de modelos de aprendizaje computacional. Sus resultados muestran un desempeño superior (medido por la raíz del error



Notas Económicas Regionales

Secretaría Ejecutiva del Consejo Monetario Centroamericano

No. 142, mayo 2023

cuadrático medio de las proyecciones) de los modelos de aprendizaje computacional frente a los modelos univariados tradicionales. De manera similar Fuentes (2022) compara el desempeño de diferentes modelos de econometría y aprendizaje computacional para estimar el comportamiento de la inflación total de Guatemala en diferentes horizontes.

Continuando con los estudios relacionados con la inflación, la literatura reciente ha destacado la importancia de analizar y considerar la evolución de la volatilidad de esta variación de precios. Este tema, enfocado en la economía dominicana, es estudiado por Quiñónez & Ramírez (2022) y Díaz, Ramírez & Sánchez (2022), quienes analizan los cambios en la volatilidad de la serie de inflación e inflación subyacente en el contexto del COVID-19. Otros autores analizan los efectos de este choque sobre la precisión de los modelos de pronósticos de inflación (Mejía & Ramírez, 2022). Este bloque de documentos utiliza técnicas de econometría (como modelos ARIMA y GARCH) para responder diferentes preguntas de investigación.

En cuanto a la actividad económica, los modelos de aprendizaje computacional se han destacado por incorporar una gran cantidad de variables sin incurrir en altos costos en términos de grados de libertad. Esto ha permitido el diseño de modelos de proyección en tiempo real (*nowcast*), los cuales utilizan variables independientes de alta frecuencia para estimar variables dependientes de una frecuencia mayor. Sobre este tema, Amaya (2022) realiza una revisión de literatura donde resalta el rol de estos modelos en la toma de decisiones de corto plazo. León & Ortega (2018) realizan proyecciones de la actividad económica utilizando información de pagos electrónicos y modelos de redes neuronales. Una aproximación similar es presentada por Santana (2017), quien utiliza cerca de 150 términos de búsquedas en Google (información provista por Google Trends) para proyectar la evolución de la actividad económica mensual de la economía dominicana, y también realizar pronósticos de la inversión y consumo privado. Asimismo, se destaca el trabajo de Santana & Giraldi (2019) quienes estudian la dinámica de la actividad del sector turismo utilizando datos de alta frecuencia y modelos estructurales bayesianos de series de tiempo (BSTS).

El desarrollo de los algoritmos utilizados en el aprendizaje computacional ha sido nutrido por el aumento en la disponibilidad de información de alta frecuencia y así como mejoras en técnicas de extracción de información. Sobre este punto se destaca el trabajo de Rosa (2021) quien utiliza diferentes métodos de extracción de datos de páginas web (*webscraping*) para desarrollar un índice de precios de viviendas para la República Dominicana.

Sobre la extracción de información se destacan los avances en la minería de texto (*text mining*), donde se recopila información de fuentes no estructuradas, en su mayoría textos (como libros, artículos de periódicos y de opinión, transcripciones de audio y videos, entre otros). Este tipo de datos son utilizados por Santana (2018) donde se desarrolla un índice de confianza del consumidor para la República Dominicana. Este indicador logra mantener una correlación de 0.6 con el consumo privado (que representa cerca de un 70% del PIB dominicano) y logra anticipar cambios de dirección o puntos de giro en el comportamiento de esta variable.



Notas Económicas Regionales

Secretaría Ejecutiva del Consejo Monetario Centroamericano

No. 142, mayo 2023

El desarrollo de índices basados en datos no estructurados (textos en su mayoría) ha tomado un significativo incremento en su consideración para los tomadores de decisiones de política monetaria. Este tipo de indicadores son construidos con una alta periodicidad (desde semanal hasta por horas), permitiendo analizar la reacción del público general ante el anuncio de eventos de alto impacto para la economía. Tal es el caso de Ahir, Bloom & Furceri (2018) quienes desarrollan un índice de incertidumbre internacional, recolectando datos de artículos de periódicos de distintos países. Un ejercicio similar es desarrollado por Santana, Rosa & Quiñónez (2020) quienes toman noticias de determinados portales de información para construir un indicador de incertidumbre, logrando desagregar el mismo por bloques, lo que permite identificar la fuente (o incidencias) de cada bloque sobre el índice.

Consideraciones finales

Con el pasar de los años la economía, como ciencia, ha logrado grandes avances, al pasar de una ciencia “pura”, con teorías basadas en observaciones, a una ciencia apoyada de las más rigurosas técnicas de estimaciones. A lo largo del tiempo se ha logrado avanzar en la validación de teorías y en la toma de decisiones, al contar con más información y mejores capacidades técnicas para el procesamiento de esta.

Con el avance de la disponibilidad de datos, y las nuevas capacidades de extracción de datos, se ha desarrollado una ciencia nueva, conocida como ciencia de datos. La misma se ha encargado de consolidar y desarrollar técnicas de extracción de información, como lo son el *webscraping* y la minería de texto, al igual que a desarrollar y optimizar algoritmos capaces de identificar relaciones no lineales entre un alto volumen de información.

Estas nuevas herramientas, disponibles para los economistas, ha logrado probar su utilidad en diferentes investigaciones, señaladas anteriormente. Las técnicas de aprendizaje computacional ayudan a los tomadores de decisiones de política monetaria a tener más información a la hora de estudiar diferentes escenarios y también mejora la visibilidad en el corto plazo para tomar en cuenta la posible reacción del público general.

Se espera que el uso de estas herramientas continúe su expansión y la adopción de estas no signifique un abandono (total o parcial) de las diferentes técnicas tradicionales de econometría. Es criterio del analista determinar la mejor herramienta para responder a la pregunta de investigación de la manera más rigurosa posible.



Bibliografía

- Ahir, H., Bloom, N., & Furceri, D. (2018). *The world uncertainty index*. Obtenido de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3275033
- Amaya, P. (2022). *Modelos nowcasting y su importancia para la toma de decisiones de corto plazo*. Obtenido de <https://www.bcr.gob.sv/documental/Inicio/descarga/bed382efc32050a7660f9e816bbafc39.pdf>
- Bjerkholt, O. (1995). *Ragnar Frisch and the Foundation of the Econometric Society and Econometrica*. Obtenido de https://www.ssb.no/a/histstat/doc/doc_199509.pdf
- Buornigal, G., & García, M. (2022). *Predicción del Default en Carteras de Crédito: Un Enfoque de Machine Learning (ML)*. Obtenido de <https://cdn.bancentral.gov.do/documents/biblioteca/documents/2022-cuarto.pdf>
- Burgener, E., & Rydning, J. (s.f.). *High Data Growth and Modern Applications Drive New Storage Requirements in Digitally Transformed Enterprises*. Obtenido de <https://www.delltechnologies.com/asset/en-us/products/storage/industry-market/h19267-wp-idc-storage-reqs-digital-enterprise.pdf>
- Cielen, D., Meysman, A., & Ali, M. (2016). *Introduction to Data Science*. Obtenido de <http://bedford-computing.co.uk/learning/wp-content/uploads/2016/09/introducing-data-science-machine-learning-python.pdf>
- Ciomba, P. (1910). *Outline of Econometrics and Bookkeeping Theory*. Obtenido de https://www.researchgate.net/publication/360697004_Pawel_Ciomba%27s_econometric_theory_of_bookkeeping
- de Óleo, C., & Ramírez, C. (2021). *Determinantes de la desocupación en la región Norte: un acercamiento por el lado de la oferta*. Obtenido de <https://www.bancentral.gov.do/a/d/2584-documentos-de-trabajo#>
- Díaz, C., Ramírez, N., & Sánchez, J. (2022). *Persistencia de la inflación subyacente en el contexto del COVID-19*. Obtenido de https://cdn.bancentral.gov.do/documents/trabajos-de-investigacion/documents/coleccion_ensayos_vol_xvi_no3.pdf?v=1684344767284
- Frisch, R. (1971). *On a Problem of Pure Economics*. Obtenido de <https://www.sv.uio.no/econ/om/tall-og-fakta/nobelprisvinnere/ragnar-frisch/published-scientific-work/rf-published-scientific-works/rf1971d.pdf>
- Fuentes, G. (2022). *Pronósticos de inflación en Guatemala: Modelos de series de tiempo o algoritmos de machine learning?* Obtenido de



Notas Económicas Regionales

Secretaría Ejecutiva del Consejo Monetario Centroamericano

No. 142, mayo 2023

- <https://www.banguat.gob.gt/sites/default/files/banguat/Publica/Banca/BancaCentral84.pdf>
- Giraldi, A., & Santana, L. (2020). *Evaluación de la dinámica del sector turismo en la República Dominicana: perspectivas, vulnerabilidades y nuevas oportunidades*. Obtenido de <https://www.bancentral.gov.do/a/d/2584-documentos-de-trabajo#>
- Gutiérrez, D. (2015). *Machine Learning and Data Science: An Introduction to Statistical Learning Methods with R*. Obtenido de <https://www.statlearning.com/>
- Keynes, J. N. (1891). *The Scope and Method of Political Economy*. Obtenido de <https://historyofeconomicthought.mcmaster.ca/keynesjn/Scope.pdf>
- León, C., & Ortega, F. (2018). *Nowcasting economic activity with electronic payments data: a predictive modeling approach*. Obtenido de <https://www.redalyc.org/journal/5095/509557191006/html/>
- Marshall, A. (1890). *Principles of Economics*. Obtenido de <https://link.springer.com/book/10.1057/9781137375261>
- Mejía, M., & Ramírez, N. (2022). *Precisión del pronóstico de la inflación subyacente en el contexto del COVID-19*. Obtenido de <https://www.bancentral.gov.do/a/d/2584-documentos-de-trabajo#>
- Mullainathan, S., & Spiess, J. (2107). *Machine Learning: An Applied Econometric Approach*. Obtenido de <https://www.aeaweb.org/articles?id=10.1257/jep.31.2.87>
- Pagliarani, A. (2019). *Big Data Mining and Machine Learning Techniques Applied to Real World Escenarios*. Obtenido de http://amsdottorato.unibo.it/8904/1/Pagliarani_Andrea_tesi.pdf
- Persons, W. (1925). *Statistics and Economic Theory*. Obtenido de <https://www.jstor.org/stable/1928417>
- Quiñónez, J., & Ramírez, N. (2022). *Persistencia de la inflación en el contexto del COVID-19*. Obtenido de https://cdn.bancentral.gov.do/documents/trabajos-de-investigacion/documents/coleccion_ensayos_vol_xvi_no2.pdf?v=1684345924969
- Ramírez, F., & Ramírez, N. (2019). *Pronóstico de la inflación de corto plazo a partir de un modelo de inflación por artículos (MIPA) en la República Dominicana*. Obtenido de <https://www.bancentral.gov.do/a/d/2584-documentos-de-trabajo#>
- Rodríguez, A. (2020). *Forecasting Costa Rican Inflation with Machine Learning Methods*. Obtenido de <https://www.sciencedirect.com/science/article/pii/S2666143820300120>
- Rosa, J. (2021). *Construcción de un índice de precios de viviendas para la República Dominicana*. Obtenido de <https://www.bancentral.gov.do/a/d/2584-documentos-de-trabajo#>



Notas Económicas Regionales

Secretaría Ejecutiva del Consejo Monetario Centroamericano

No. 142, mayo 2023

Santana, L. (2017). *Nowcasting con Google Trends*. Obtenido de https://cdn.bancentral.gov.do/documents/trabajos-de-investigacion/documents/doc_trabajos/documento_2017_03.pdf?v=1684346418151

Santana, L., Rosa, J., & Quiñónez, J. (2020). *Incertidumbre, gestión de la política monetaria y entropía de las expectativas en la República Dominicana: un análisis basado en algoritmos de minería de texto y redes neuronales*. Obtenido de https://cdn.bancentral.gov.do/documents/trabajos-de-investigacion/documents/2020-02_Incertidumbre_Internacional.pdf?v=20200917?v=1684346418151