



Blaconá, María Teresa^(*); Méndez, Fernanda

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística

() Consejo de Investigaciones de la UNR*

DIFICULTADES POR INFORMACIÓN CON CAMBIOS DE ESTRUCTURA DE CORRELACIÓN EN SERIES MACROECONÓMICAS ARGENTINAS

I. Introducción

En el trabajo de Blaconá, M.T y Méndez, F. (2009) se llegó a la conclusión, que problemas en los datos sugerían que no fuera conveniente aplicar los métodos del NBER (Moore y Shiskin, 1967) y el modelo de Stock-Watson (1988, 1989) para construir un índice compuesto coincidente (ICC), para Argentina. Para realizar dicha construcción se deben seleccionar distintos indicadores económicos de los que se dispone información histórica y clasificarlos de acuerdo a que estas sean series coincidentes o series líderes.

Las series coincidentes son aquellas que se mueven en aproximada sincronización con la serie de PIB y van a constituir el nivel de actividad económica mensual, expresado a través del ICC. Las series a incluir en el ICC están limitadas a las variables disponibles en una frecuencia elevada y que tengan un período muestral suficiente amplio para analizar el ciclo.

Con el fin de corroborar si se cumplían estos requisitos se realizó una validación para analizar si brindaban información sobre el ciclo económico. Las herramientas que se utilizaron para determinar cuáles eran las series coincidentes, fueron las funciones de correlación cruzada entre la serie del Producto Interno Bruto (PIB) con cada una de otras series macroeconómicas, que en el estudio mencionado fueron: importaciones (IMP), recaudaciones de impuestos nacionales (RT) el Índice de producción Industrial (IPI), ventas de supermercados (SUPER) y consumo de gas (GAS).

Al realizar esta primera etapa del estudio con datos de la Argentina, se encontró el inconveniente de no contar con series confiables durante un período suficientemente amplio de tiempo (1997-2008), esto se puso de manifiesto al estimar las funciones de correlaciones cruzadas (FCC) antes mencionadas con series que comprenden distintos períodos, uno incluyendo el año 2008 y otro sólo hasta el 2007. Estas FCC mostraron correlaciones más débiles cuando se consideró el período completo que cuando se eliminó el año 2008, incluso presentaron un patrón distinto de relación entre las variables. Estos resultados dificultaron continuar con la construcción ICC porque eliminar el año 2008 no era conveniente dado que este índice es útil para determinar los ciclos económicos los cuales deben estar actualizados.

En ese momento se consideró que la próxima etapa del trabajo era poder encontrar una alternativa a este fundamental inconveniente.



Teniendo en cuenta las observaciones correspondientes al año 2009, los inconvenientes marcados en el trabajo mencionado persistían. Por tal motivo se decidió investigar sobre las recomendaciones dadas en la bibliografía ante estas dificultades.

En la sección II se describe lo mencionado por varios autores cuando los datos no son confiables, en la sección III se realiza un breve comentario de ética en Estadística y en la sección IV se presentan las consideraciones finales.

II. La información en los métodos cuantitativos de pronóstico

En los últimos tiempos algunos autores se expresaron sobre los límites de los métodos de pronósticos cuantitativos. Por ejemplo Gordon, A. (2010), dice: "Los métodos cuantitativos (MC), incluyendo el análisis estadístico, se han vuelto una autoridad del conocimiento en las modernas sociedades occidentales".

Los MC en las ciencias son recientes, sólo a partir del siglo XIX el análisis numérico consiguió un estatus y especialmente en la investigación tiene un reconocimiento de las distintas ciencias. Una de estas áreas es la Economía que es una ciencia social y por lo tanto, en algunas ocasiones su relación con las disciplinas del área matemática puede ser conflictiva. Este patrón se repite también en la psicología o cualquier otra ciencia social.

Los MC han tomado ímpetu con el crecimiento del acceso a la modelización y los adelantos en computación. Todos los modelos predictivos operan procesando algoritmos que matemáticamente definen relaciones entre variables y puntualizan un conjunto de decisiones acerca de cuáles variables afectan a cuáles otras, bajo qué condiciones y en cuánto. Con esto determinado el modelador puede proyectar hacia el futuro apretando un botón (Gordon, A., 2010).

La pregunta que nos podemos hacer son: ¿cuán buenos son los modelos para predecir?, ¿son buenos como para extender hacia el futuro dichos modelos predictivos de las variables?. Los MC funcionan bien cuando se cumplen las condiciones expresadas anteriormente, pero son extremadamente pobres cuando suceden eventos inesperados. En el caso que nos ocupa es el cambio de comportamiento visible de la estructura de correlación de cada variable y la relación entre las mismas.

Para mejorar el desempeño de los modelos predictivos, se pueden realizar preguntas conceptuales y ver el sesgo relacionado con la base de datos o cuestiones técnicas de cómo esos datos se obtuvieron, como así también, cómo se deducen los algoritmos o se construyen modelos para tratar de mejorar los procedimientos.

No obstante, existen límites de la validación del modo de predecir. La pregunta no es ¿cómo realizar el mejor modelo de predicción para esta situación?, sino, ¿es esta la situación correcta para trabajar con un modelo de predicción? El problema lo podemos expresar, como que los modelos predictivos muy a menudo son más complejos y tiene mayor incertidumbre que lo que cualquier computador puede considerar.



Según Duncan, N. (1992) para resolver el problema de realizar pronósticos existen cuatro elementos básicos, para que eligiendo un buen algoritmo, se obtengan buenos pronósticos, a saber: forma, saltos, interacción y retardo.

- Forma: se refiere a la forma matemática de la relación entre variables de entrada y salida (futuro predecible). La forma puede variar, de una simple recta a una relación que necesita media página de álgebra para describirla.
- Saltos: son discontinuidades en las relaciones, es el punto donde el efecto del "input" cambia repentinamente los factores. Para tener en cuenta este comportamiento, existen modelos especiales. Estos modelos no serían aplicables al caso que nos ocupa, de los modelos para el ICC, porque los cambios de la relación entre las variables, parecería que se deben a manipulación en los datos.
- Interacción: está presente cuando el efecto de un factor depende de los valores de uno o más de otros factores. En general se determina cómo todos afectan a todos. Los inconvenientes hallados en el trabajo que se intenta realizar son los mismos que los enunciados en el elemento anterior.
- Retardo: tiene en cuenta cuando un resultado se ve afectado no sólo por valores actuales de una variable, sino por el anterior o anteriores, surgiendo el problema cuánto atrás se tiene que considerar que sea pertinente tomarlo como "input". En el caso del análisis que se pretende realizar el principal problema existente es que cambia la estructura de correlación de algunas series de tiempo (Blaconá, M.T. y Méndez, F., 2009).

Rittel and Webber (1973), describen la situación de tener variables incompletas, mal definidas o interdependientemente contradictorias, como un problema "perverso". Un problema "perverso", contrasta con un problema "dócil", en este último la solución es clara, mientras en el primero no hay solución. El problema encontrado en las series estudiadas, podría clasificarse como un problema "perverso".

Gladwell, M., (2007), distingue entre los problemas "rompecabezas y los "misteriosos. Un problema rompecabezas es aquel al que le falta información y que si esta fuera provista, se podría resolver. Sin embargo, un problema misterioso, aunque se tenga más información no necesariamente ayuda. El autor dice, "A veces la información que nos han dado es inadecuada y algunas veces no podemos ser tan inteligentes acerca de cómo darle sentido a lo que tenemos y algunas veces el problema en sí mismo no puede ser resuelto". En el análisis de las series que intervendrían en el ICC, como los datos a partir del 2007 cambiaron de estructura, sin alguna explicación inherente al problema, sino más bien porque parecería que los datos son incorrectos, el tener más años de información, mientras persista la situación actual, no resolverá el problema. Por lo tanto, se estaría ante un problema "misterioso".

En resumen, se debe tener conocimiento adecuado de las condiciones iniciales, para postu-



lar los modelos de pronóstico. Las proyecciones cuantitativas fallan, no porque ellas estén mal realizadas, sino porque ellas tratan de hacer lo imposible. En ciertos casos, como el que nos ocupa, no somos capaces de capturar adecuadamente la complejidad o las situaciones inciertas del presente, en consecuencia realizar pronósticos y estudiar su evolución, no se puede resolver por algoritmos, modelos o computadoras.

Courtney at al, (1997) postulan cuatro niveles de incertidumbre, basándose sobre el grado de complejidad y el ritmo de los cambios en la situación bajo estudio. Estos niveles son:

- Nivel 1: pequeños movimientos, producen situaciones bien establecidas del resultado dependiente.
- Nivel 2: existe un conjunto limitado, predeterminado de resultados posibles que pueden ocurrir, pero no se sabe cuál.
- Nivel 3: el resultado es indeterminado, pero limitado dentro de un rango factible que puede ser definido. Los resultados no factibles o imposibles pueden ser identificados y descartados.
- Nivel 4: los resultados son desconocidos y no se pueden conocer, y los rangos de resultados son ilimitados.

Cuando los niveles de incertidumbre son altos, por ejemplo, el estudio que presentamos estaría en el nivel 4, los enfoques cuantitativos rápidamente se vuelven inadecuados, porque cuando los supuestos fallan los modelos colapsan.

III. Ética en Estadística

Según la Guía Ética de la American Statistical Association, (1999), "el propósito de la guía ética de la ASA es asegurar que el trabajo estadístico sea ético y efectivo cuando se realizan trabajos del medio ambiente y asistir a los estudiantes a aprender a desarrollar el trabajo estadístico razonablemente".

Entre los tópicos más salientes de la Guía ASA se pueden distinguir:

- Profesionalismo: se debe puntualizar la necesidad de competencia, juicio, diligencia, auto respecto y respecto a las otras personas.
- Responsabilidades para los que financian, clientes y empleadores. El estadístico debe asegurar que su trabajo sea adecuado a las necesidades y recursos de los que pagan por él.



- Responsabilidades en publicaciones y testimonios: se debe asentar la necesidad de reportear información suficiente en forma clara y entendible sobre el objetivo del trabajo, cómo fue desarrollado y cualquier limitación sobre su validación.
- Responsabilidades al investigar sujetos, se debe describir los requerimientos para proteger los intereses de los humanos, animales y medio ambiente que son investigados no sólo durante la recolección de los datos, sino también en el análisis, interpretación y publicación de los resultados encontrados.
- Responsabilidades con los colegas del grupo de investigación, definir las responsabilidades mutuas de los profesionales participando en equipos de investigación multidisciplinarios.
- Responsabilidades hacia otros Estadísticos, se debe notar la interdependencia de los profesionales que realizan trabajos similares, sean de la misma o diferentes organizaciones.
- Responsabilidades en la revisión de alegatos de malas conductas, puntualizando los que algunas veces son desagradables procesos de investigación de la potencial violación ética y tratando a los involucrados con justicia y respeto.
- Responsabilidades de empleadores, incluyendo organizaciones, individuos u otros clientes que emplean al Estadístico, fomentando a empleadores y clientes a reconocer la alta interdependencia natural de la ética estadística y la validación estadística. El empleador y los clientes no deben ejercer presión al Estadístico para producir un "resultado" particular sin mirar su validación estadística. Se debe eliminar el potencial daño social que puede resultar de la difusión de trabajos estadísticos falsos o mal realizados.

Si bien no se puede comprobar fehacientemente, los resultados encontrados en el trabajo Blaconá, M.T., Méndez, F. (2009) hacen sospechar sobre la metodología utilizada para determinar los valores publicados. Al respecto se pueden considerar lo expresado por autores reconocidos, sobre el tema.

Seltzer, W. (2005), dice: "Un importante tópico del Fundamental Principles of Official Statistics, es el mantener e incrementar la integridad del sistema estadístico nacional. ... pueden surgir amenazas a la integridad de muchas formas, incluyendo entre otras, conceptos de manipulación política arbitraria, definiciones e información de los datos muy atrasada, informando los datos reales manipulados, usando la agencia para análisis políticos y otros trabajos políticos y politizando el personal técnico de la agencia".

Por otro lado, teniendo en cuenta lo referido por Habermann, H. (2005): "En un sentido amplio, existe también tensión entre la agencia estadística y el uso del dato estadístico (especialmente en el contexto de la aplicación de la ley y otras actividades de defensa o intelligen-



cia) y la tensión relatada que existe con respecto a la autonomía de la agencia estadística y el contexto de la agencia dentro del gobierno. ...". "Es importante saber que esta tensión puede no resolverse. Las agencias del gobierno, como pequeñas organizaciones dentro de una estructura de gobierno mucho más amplia, no están bien equipadas para ejercitar autoridad moral."

El profesional estadístico tiene que estar capacitado para advertir con mayor premura los daños que se podrían cometer a partir de información estadística construida a partir de metodología no adecuada. Utilizar metodología apropiada no necesariamente implica comportamiento ético: Pero el uso de metodología no adecuada necesariamente va a acarrear problemas éticos.

Por ello, en el estudio que nos preocupa además de los problemas metodológicos se nos plantearon problemas éticos.

IV. Consideraciones finales

Para construir un índice compuesto coincidente ICC para Argentina, se deben seleccionar distintos indicadores económicos, de los cuales se debe disponer información histórica y clasificarlas de acuerdo a que estas sean series coincidentes o series líderes.

Las series a incluir en el ICC están limitadas a las variables disponibles en una frecuencia elevada y que tengan un período muestral suficientemente amplio para analizar el ciclo. Para corroborar si se cumplen estos requisitos se realiza una validación para analizar si brindan información sobre el ciclo económico. Las herramientas que se utiliza para determinar cuáles son las series coincidentes, son entre otras, las funciones de correlación cruzada entre la serie del Producto Interno Bruto (PIB) con cada una de otras series macroeconómicas, que en el estudio mencionado son: importaciones (IMP), recaudaciones de impuestos nacionales (RT), el Índice de producción Industrial (IPI), ventas de supermercados (SUPER) y consumo de gas (GAS).

Al realizar esta primera etapa del estudio con datos de la Argentina, se encontró el inconveniente de no contar con series confiables durante un período suficientemente largo de tiempo (1997-2009), esto se puso de manifiesto al estimar las funciones de correlaciones cruzadas (FCC) y no presentarse el comportamiento esperado según lo estipulado tanto por la teoría económica como los resultados empíricos de otros países.

Para solucionar el problema planteado se revisó bibliografía sobre la forma de resolver el problema, llegando a la conclusión de que lo que dicen la mayoría de los autores, es que cuando los datos contienen evidente falta de coherencia, no existe método cuantitativo que pueda solucionarlo. Por lo tanto, los pronósticos encontrados por MC dejarían de tener sentido.

Por último se puede expresar que en el estudio que nos preocupa, además de los problemas metodológicos, se nos plantearon problemas éticos, que nos impidieron continuar con



el análisis propuesto.

REFERENCIAS BIBLIOGRÁFICAS

- American Statistical Association, (1999), "Ethical Guidelines for Statistical Practice", <http://www.amstat.org/about/ethicalguidelines.cfm>.
- Blaconá, M.T., Méndez, F. (2009), "Cambios en la estructura de correlación de variables macroeconómicas argentinas", Decimocuarta Jornadas de "Investigaciones en la Facultad de Ciencias Económicas y Estadística".
- Courtney, H., Kirkland, J. and Viguier, P., (1997), "Strategy under uncertainty", Harvard Business Review.
- Duncan, N., (1992), "Why we can't predict", New Scientist, 136, 47.
- Gladwell, M., (2007), "Open secrets", The New Yorker, January 8.
- Gordon, A., (2010), "The boundaries of quantitative forecasting methods: respecting the limits of determinism", Foresight, Issue 19.
- Habermann, H., (2005), "Ethics, Confidentiality, and Data Dissemination", US Census Bureau.
- Rittel, H. and Webber, M. (1973), "Dilemmas in a general theory of planning", Policy Sciences, 4.
- Seltzer, W. (2001): "US Federal Statistics and Statistical Ethics: The role of the American Statistical Associations. Ethical Guidelines for Statistical Practice". Seminar organized by the Methodology Section, Washington Statistical Society.
- Seltzer, W. (2005): "Official Statistics and Statistical Ethics. Selected Issues". International Statistical Institute, 55th Session.
- United Nations Statistical Commission (1994): "Fundamental Principles of Official Statistics", <http://unstats.un.org/unsd/methods/statorg/FP-English.htm>.