



Munich Personal RePEc Archive

Classical Probability of Overfitting with Information Criteria: Estimations with Chilean Macroeconomic Series

Medel, Carlos A.

18 July 2014

Online at <https://mpra.ub.uni-muenchen.de/57401/>
MPRA Paper No. 57401, posted 23 Jul 2014 01:23 UTC

Probabilidad Clásica de Sobreajuste con Criterios de Información: Estimaciones con Series Macroeconómicas Chilenas*

Carlos A. Medel[†]

July 18, 2014

Abstract

En este trabajo se estima mediante simulaciones de Monte Carlo la probabilidad clásica de sobreajuste, en un ambiente autoregresivo (AR), con los criterios de información (CI) de Akaike, Schwarz y Hannan-Quinn (AIC, BIC y HQ), calibradas con los datos chilenos de inflación total, inflación subyacente, Imacec y el retorno mensual del tipo de cambio nominal peso-dólar estadounidense. Esta probabilidad corresponde al número de veces en que un modelo candidato posee estrictamente un mayor número de coeficientes que el modelo verdadero, dividido por el número total de búsquedas. Los resultados indican que el mayor riesgo de sobreajuste se obtiene con el AIC, seguido por el HQ y finalmente el BIC. Las probabilidades más altas de sobreajuste se alcanzan con el AIC, llegando a un 32 y un 30% con el tipo de cambio y el Imacec, respectivamente, seguidas por un 25 y 22% para la inflación total y subyacente. Considerando los tres CI, la mayor probabilidad siempre se obtiene solamente con un coeficiente de sobreajuste. Asimismo, es más probable que el sobreajuste no exceda a los 10 coeficientes. Estos resultados son importantes en la medida que cuantifican el riesgo al cual se está sometido al modelar dichas variables mediante representaciones AR. Dentro de estos potenciales problemas están: (i) el de regresión espuria, (ii) distorsionar la estimación de la función de impulso-respuesta, y (iii) perjudicar la precisión predictiva de la variable de interés. Este último problema es analizado en detalle.

Códigos JEL: C22; C51; E23; E27; E31; E37; F31.

Palabras clave: *Modelación no paramétrica; criterios de información; sobreajuste; análisis fuera de muestra.*

*Se agradecen los comentarios de Rodrigo Alfaro, Rolando Campusano, Mario Giarda, Ercio Muñoz, Pablo Pincheira y Damián Romero. Todas las estimaciones han sido realizadas con el programa *Eviews 8*. Este trabajo ha sido realizado mientras el autor se encontraba afiliado al Banco Central de Chile. Cualquier error u omisión es responsabilidad del autor.

[†]Email: carlos_medel@yahoo.com.

Emperador José II de Habsburgo: *"My dear young man, don't take it too hard. Your work is ingenious. It's quality work. And there are simply too many notes, that's all. Just cut a few and it will be perfect."*

Mozart: *"Which few did you have in mind, Majesty?"*

—*Amadeus* (película), 1984, Warner Bros.

1 Introducción

El ajuste a un modelo simple de una serie de tiempo estacionaria finita $\mathbf{y}_t = (y_1, \dots, y_T)'$, donde T es el número de observaciones disponibles, proveniente del modelo verdadero $L(\mathbf{y}_t|\theta_0)$ (con un conjunto de coeficientes θ_0), es un problema recurrente en muchas disciplinas. Dentro del análisis económico, el problema ha sido tradicionalmente resuelto mediante el uso de los llamados criterios de información (CI). Los CI proveen una medida de la distancia entre un modelo estimado con un conjunto de coeficientes θ_k y el modelo verdadero. La popularidad asociada a los CI se basa, en gran medida, en su versatilidad, facilidad de manejo, y sus resultados para la modelación con fines predictivos.¹ Típicamente, su uso se ha circunscrito al ajuste del modelo $L(\mathbf{y}_t|\theta_k)$ utilizando el menor conjunto de información—*i.e.* rezagos de la misma serie \mathbf{y}_t . Así, suponiendo una familia de modelos lineales autoregresivos (AR) $\mathcal{F}_k = \{L(\mathbf{y}_t|\theta_k)|\theta_k \in \Theta_k\}$, de dimensión finita $k + 1$ (k coeficientes AR más la varianza del componente estocástico), el problema consiste en la búsqueda del vector $\hat{\theta}_k$ que maximiza la función de verosimilitud $L(\mathbf{y}_t|\theta_k)$, para obtener $L(\mathbf{y}_t|\hat{\theta}_k)$. Dada la forma AR del conjunto de información, se tiene que $k \subset \{1, \dots, k^{\max}\}$, por lo que $\Theta_1 \subset \dots \subset \Theta_{k^{\max}}$ (desde 2 hasta $k^{\max} + 1$ elementos.)

La medida de distancia que los CI proponen entre el *modelo candidato* y el verdadero es la siguiente:²

$$d_t(\theta_k, \theta_0) = \mathbb{E}_0 [-2 \log L(\theta_k|\mathbf{y}_t)], \quad (1)$$

donde \mathbb{E}_0 es la esperanza condicional en el modelo verdadero, y $L(\theta_k|\mathbf{y}_t)$ es la función de verosimilitud del modelo candidato de coeficientes θ_k , o también denominado "modelo k ". Así, la evaluación de esta medida utilizando la función de verosimilitud del modelo k , $L(\hat{\theta}_k|\mathbf{y}_t)$, genera la distancia $d_t(\hat{\theta}_k, \theta_0)$. Sin embargo, esta distancia no es computable dado que se estima de manera condicional en el modelo verdadero θ_0 . Una solución, propuesta por Akaike (1973), consiste en estimar $d_t(\hat{\theta}_k, \theta_0)$ con su estimador sesgado $\mathbb{E}_0 [-2 \log L(\hat{\theta}_{k,T}|\mathbf{y}_t)]$, es decir, $\hat{\theta}_k$ con los T datos disponibles, $\hat{\theta}_{k,T}$, y que el sesgo mencionado, equivalente a:

$$\text{Sesgo} = \mathbb{E}_0 \left[\mathbb{E}_0 [-2 \log L(\theta|\mathbf{y}_t)]_{|\theta=\hat{\theta}_k} \right] - \mathbb{E}_0 [-2 \log L(\theta_k|\mathbf{y}_t)], \quad (2)$$

sea aproximado por la expresión $2k$ (dos veces el número de elementos del conjunto de elementos del modelo k). Así, esta última expresión constituye una penalización por el número de coeficientes utilizados en la aproximación al modelo verdadero.

Varios CI basados en la función $L(\hat{\theta}_k|\mathbf{y}_t)$ utilizan esta derivación como punto de partida, y básicamente difieren en la forma de aproximación al término $d_t(\hat{\theta}_k, \theta_0)$.³ Cuando el sesgo de aproximación

¹Ver, por ejemplo, Elliot y Timmermann (2006), Stock y Watson (2007), y más recientemente, una exhaustiva comparación entre varias metodologías se encuentra en Faust y Wright (2013).

²Esta medida corresponde a la *discrepancia* propuesta en Kullback y Leibler (1951) y Kullback (1968); también denominada *entropía cruzada*.

³Notar que los criterios basados en la función de verosimilitud $L(\hat{\theta}_k|\mathbf{y}_t)$ son típicamente computados dentro de

es igual a $2k$ se obtiene el criterio de Akaike (1974; AIC), cuando es igual a $k \times \log(T)$, el de Schwarz (1978; o bayesiano, BIC), y cuando es igual a $2k \times \log(T)$, el de Hannan y Quinn (1979; HQ):⁴

$$\text{AIC} : -2 \log L(\hat{\theta}_{k,t} | \mathbf{y}_t) + 2k, \quad (3)$$

$$\text{BIC} : -2 \log L(\hat{\theta}_{k,t} | \mathbf{y}_t) + k \times \log(T), \quad (4)$$

$$\text{HQ} : -2 \log L(\hat{\theta}_{k,t} | \mathbf{y}_t) + 2k \times \log(T). \quad (5)$$

Un modelo con menor valor en el CI representa un mejor candidato: está más cercano al modelo verdadero. Existen diversas comparaciones entre estos CI, tanto para su rendimiento dentro como fuera de muestra, con datos simulados y variables económicas, para modelos lineales y no lineales, entre otras comparaciones.⁵

Notar que a pesar de incluir y diferir en el término de penalización, una de las principales críticas al proceso de selección de modelos con CI es el llamado problema de *sobreajuste* (del inglés, *overfitting*). Dentro del esquema AR presentado, se entiende por sobreajuste cuando el modelo k presenta un mejor ajuste de acuerdo al CI que el modelo verdadero, aunque con un mayor número de coeficientes.^{6 7} Formalmente, el sobreajuste existe cuando se cumplen las siguientes dos condiciones:

$$(i) : \text{CI}(\hat{\theta}_{k,T}, k | \mathbf{y}_t) < \text{CI}(\theta_0, k_0 | \mathbf{y}_t), \text{ y} \quad (6)$$

$$(ii) : k > k_0, \quad (7)$$

donde k_0 es el número de coeficientes del modelo verdadero, el cual es sobreajustado en s coeficientes ($s = k - k_0$).

La motivación de este trabajo es estimar la probabilidad clásica de sobreajuste, es decir, reportar el cociente entre el número de veces que un modelo basado en un CI presenta un número de coeficientes estrictamente mayor al verdadero y el número de veces totales en que se realiza esa búsqueda, para cada CI utilizando series macroeconómicas específicas. Estas constituyen cuatro series que: (i) son representativas de la economía chilena—*i.e.* concitan una alta atención del público

muestra con la información disponible hasta el período T . Dado que el AIC no es consistente, generalmente es preferible utilizar toda la muestra disponible para su estimación (Canova, 2007). Si el objetivo perseguido con la modelación es predecir, se hace el supuesto implícito que fuera de muestra el modelo presentará un ajuste similar, lo que no necesariamente ocurre en la práctica. Esta relación puede perturbarse, por ejemplo, en una serie que presenta valores atípicos y/o quiebres. Mayores detalles sobre la relación y extrapolabilidad dentro/fuera de muestra a nivel teórico se presentan en Hansen (2009).

⁴Dada la similitud de estos CI, la derivación de sus propiedades asintóticas para una varianza fija del componente estocástico, puede obtenerse de manera conjunta para el *Criterio de Información Generalizado* (GIC, por sus siglas en inglés), de acuerdo a, por ejemplo, Xu y McLeod (2012). Sin embargo, el trabajo que ha guiado la obtención de las propiedades asintóticas de los CI—y del AIC, en particular—es Shibata (1976).

⁵Ver, por ejemplo, Koehler y Murphree (1988) para una comparación dentro de muestra del ajuste del AIC y BIC, Lütkepohl (1985) para una comparación con VARs, Stock y Watson (2007) para una comparación de modelos lineales versus no lineales, y Medel (2011; 2013) para una comparación fuera de muestra entre el AIC, BIC y HQ con los datos trimestrales del PIB chileno.

⁶Notar que el mejor ajuste se asocia únicamente a un menor valor del CI, pudiendo utilizarse otras medidas de bondad de ajuste, como los coeficientes R^2 y \bar{R}^2 . En el caso AR analizado en este trabajo, sin embargo, un elemento distintivo de los CI considerados es la forma de penalización de los coeficientes adicionales utilizados para el cálculo de la función de máxima verosimilitud. Con las medidas R^2 y \bar{R}^2 la penalización es ampliamente más permisiva, fomentando el sobreajuste.

⁷Es con este espíritu que Nishii (1984) deriva las propiedades asintóticas del AIC y BIC suponiendo dos funciones de riesgo: (i) para la probabilidad de estimar el modelo verdadero, y (ii) para comparar el ajuste del modelo candidato con el del modelo verdadero.

no especializado–, y (ii) poseen características estadísticas distintas. Así, el objetivo de este trabajo consiste en evaluar el riesgo de sobreajuste para variables macroeconómicas chilenas asociado a la modelación con CI comunes utilizando una función de riesgo–distribución de probabilidades– que anida las dos funciones tradicionalmente utilizadas. Las variables utilizadas son la inflación total, la inflación subyacente (inflación total excluyendo la inflación del IPC de alimentos y de energía), el Indicador Mensual de la Actividad Económica (Imacec), y el retorno mensual del tipo de cambio nominal peso–dólar estadounidense (tipo de cambio). Una revisión de la literatura relacionada indica que no se dispone de estimaciones como las presentadas en este trabajo, ni con series macroeconómicas de naturaleza similar.

Se discuten, además, algunas razones por las cuales el sobreajuste podría ser un eventual problema para la inferencia estadística. Dentro de estos potenciales problemas están: (i) en el caso con variables exógenas–no AR–, errar el efecto de una variable independiente x sobre la dependiente y (encontrar $\partial y/\partial x \neq 0$ cuando en realidad es inexistente; problema denominado en la literatura como el de *regresión espuria*; Granger y Newbold, 1974), (ii) distorsionar la estimación, especialmente de la varianza, de la función de impulso–respuesta de la variable de interés debido a la introducción de sesgo (Kilian, 2001), y (iii) perjudicar la precisión predictiva de la variable de interés (Clark, 2004). En este trabajo, en particular, se analiza y provee evidencia relativa al tercer problema mencionado.

Los resultados son directamente comparables, y están en línea, con los obtenidos por Seghouane (2006). Sin embargo, su obtención proviene de un ejercicio de simulaciones de una expresión analítica derivada asintóticamente. En este trabajo se pretende realizar una estimación asintótica desde un punto de vista empírico, enfrentando complejidades similares a las de la práctica habitual. Los resultados de este trabajo indican que el mayor riesgo de sobreajuste se obtiene con el AIC, seguido por el HQ, y finalmente el BIC. Las probabilidades más altas de sobreajuste se alcanzan con el AIC, llegando a un 32 y un 30% con el tipo de cambio y el Imacec, respectivamente, seguidas por un 25 y 22% para la inflación total y subyacente. El BIC, por su parte, exhibe la menor probabilidad de sobreajuste: alcanza a un 3 y 2% con el tipo de cambio y el Imacec, respectivamente, y un 1% tanto para la inflación total como subyacente. Considerando los tres CI, la mayor probabilidad siempre se obtiene solamente con un coeficiente de sobreajuste. Además, es más probable que el sobreajuste no exceda a los 10 coeficientes.

Estos resultados son importantes en la medida que representan el riesgo de sobreajuste–*i.e.* que tan alejado se está del modelo verdadero–al cual se está sometido al modelar dichas variables mediante representaciones AR, u otras con características estadísticas similares. Esto es potencialmente un problema para fines predictivos de acuerdo a los resultados de, por ejemplo, Clark (2004).⁸ En esta misma línea, este trabajo comparte algunos de estos resultados, encontrando que una vez estimado un modelo sobreajustado, es más probable que presente un peor rendimiento predictivo. Un resultado importante es que un sobreajuste promedio mayor a los 12 coeficientes–*i.e.* de un orden mayor a la frecuencia anual de los datos–incrementa el error de proyección en alrededor de un 20% respecto al error promedio de sobreajuste de hasta 12 coeficientes, prediciendo 24 meses adelante.

El trabajo continúa de la siguiente manera. En la sección 2 se describe en detalle el ejercicio de simulaciones y su parametrización. En la sección 3 se presentan los resultados, de manera

⁸El trabajo de Clark (2004), calibrado con variables de la economía estadounidense, encuentra que el sobreajuste genera predicciones menos precisas que el modelo verdadero, y que incluso un modelo más parsimonioso que el verdadero entrega un mejor rendimiento predictivo fuera de muestra.

gráfica, para la probabilidad de sobreajuste de cada CI considerado a través de los s coeficientes de sobreajuste del modelo k . Además, se analiza el impacto tanto del subajuste como del sobreajuste en la capacidad predictiva, como medida del *costo económico* de los desvíos respecto al modelo verdadero. Finalmente, se resumen los resultados en la sección 4.

2 Diseño de simulaciones de Monte Carlo

La estimación de la probabilidad clásica de sobreajuste se obtiene de la siguiente manera. Primero, las variables son transformadas para alcanzar la estacionariedad requerida por la metodología. Dado que en este trabajo se considera en el conjunto Θ_i solo coeficientes AR, la modelación de series de tiempo, de acuerdo a Box y Jenkins (1970), se obtiene con series estacionarias para la obtención de coeficientes consistentes. Las transformaciones se indican en la fila correspondiente de la tabla 1, así como también sus estadísticos descriptivos dentro de la muestra indicada. En segundo lugar, se estiman los coeficientes $\{\alpha; \rho_k; \sigma_\varepsilon^2\}$ de los veinticuatro modelos AR que surgen de la expresión $y_t = \alpha + \sum_k^{k^{\max}} \rho_k y_{t-k} + \varepsilon_t$, con $\varepsilon_t \sim iid\mathcal{N}(0, \sigma_\varepsilon^2)$, variando k^{\max} desde 1 hasta 24—*i.e.* desde el AR(1) hasta el AR(24). Así, mediante una estrategia *general a específica* (Hoover y Pérez, 1999), se obtiene un modelo AR cuyos rezagos consecutivos (*non-skipped*) sean significativos al 1%. El modelo elegido de acuerdo a esta metodología se considera como el modelo verdadero de y_t . La estimación de los modelos verdaderos se presenta en la tabla 2. Notar que esta representación supone que existen las raíces AR de las series consideradas, lo que no necesariamente puede ocurrir si la serie proviene de un proceso generador de datos constituido por medias móviles (MA). Además, convenientemente se supone que a lo largo de la muestra el modelo AR estimado no cambia. Los valores alcanzados por el estadístico Durbin-Watson sugieren que los modelos están bien especificados (independientemente de los valores de R^2 ; Granger y Newbold, 1974).

Tabla 1: Estadísticos típicos de las series analizadas

	Inflación total	Inflación subyacente	Imacec	Tipo de cambio
Transformación	Var. anual del índice	Var. anual del índice	Var. anual del índice	Var. mensual del nivel
Promedio	3.91	3.45	3.96	0.18
Desviación estándar	2.54	2.33	3.33	2.54
Máximo	9.82	9.79	13.20	16.64
Mínimo	-3.45	-1.63	-4.40	-6.97
Muestra	1995.1-2013.7	1995.1-2013.7	2009.1-2014.2	1995.1-2014.3
No. de observaciones (T)	223	223	62	231
Frecuencia	Mensual	Mensual	Mensual	Mensual
Fuente	BCCh	BCCh	BCCh	BCCh

Fuente: Elaboración propia a partir de los datos del Banco Central de Chile (BCCh).

Dado que para cada uno de los modelos de cada variable se obtienen sus coeficientes estimados $\{\hat{\alpha}; \hat{\rho}_k\}$, más la distribución del componente estocástico, $\hat{\sigma}_\varepsilon^2$, es posible realizar n replicaciones de la misma variable. En este caso, se realizan $n = 5000$ replicaciones para cada variable. Para cada una de las 5000 replicaciones se estima, posteriormente, el orden k con los tres CI considerados. En caso de encontrar algún CI el orden del modelo verdadero, el modelo estimado no necesariamente posee los mismos coeficientes que el modelo verdadero, debido a que es estimado con una serie simulada, en vez de la serie efectiva. Además, con el orden elegido por cada criterio en cada replicación (que bien puede como también no ser el orden del modelo verdadero), se reestima el modelo con $T - 24$ observaciones y se realiza una proyección a uno, doce y veinticuatro meses

adelante, $h = \{1; 12; 24\}$. Así, se obtiene un total de 5000 proyecciones, con sus correspondientes errores de proyección, provenientes de modelos con (eventuales) diferentes órdenes AR. Finalmente, una vez obtenidos los órdenes AR para el total de replicaciones con cada CI, se calcula y presenta el histograma para las tres series $s^{CI} = k^{CI} - \hat{k}_0$, $CI = \{AIC; BIC; HQ\}$, con k^{CI} variando desde 1 hasta 24; *i.e.* desde 1 hasta 24 coeficientes de sobreajuste.

Tabla 2: Especificación de los modelos verdaderos (*)

Var. dep.:	Inflación total	Inflación subyacente	Imacec	Tipo de cambio
Primer rezago	1.341	1.289	0.483	0.379
	[20.168]	[17.749]	[4.301]	[7.519]
Segundo rezago	-0.442	-0.382	0.316	×
	[-4.272]	[-2.939]	[3.035]	×
Tercer rezago	0.290	0.233	×	×
	[2.751]	[1.846]	×	×
Cuarto rezago	-0.330	-0.168	×	×
	[-2.889]	[-2.395]	×	×
Quinto rezago	0.275	×	×	×
	[2.427]	×	×	×
Sexto rezago	-0.186	×	×	×
	[-2.487]	×	×	×
Constante	0.192	0.079	0.979	0.094
	[2.931]	[2.147]	[2.343]	[0.628]
R^2	0.964	0.973	0.674	0.143
No. de observaciones	201	207	58	227
Durbin-Watson stat.	1.980	2.008	2.039	1.937

(*) Entre [·] se presenta el t -Statistic. Fuente: Elaboración propia.

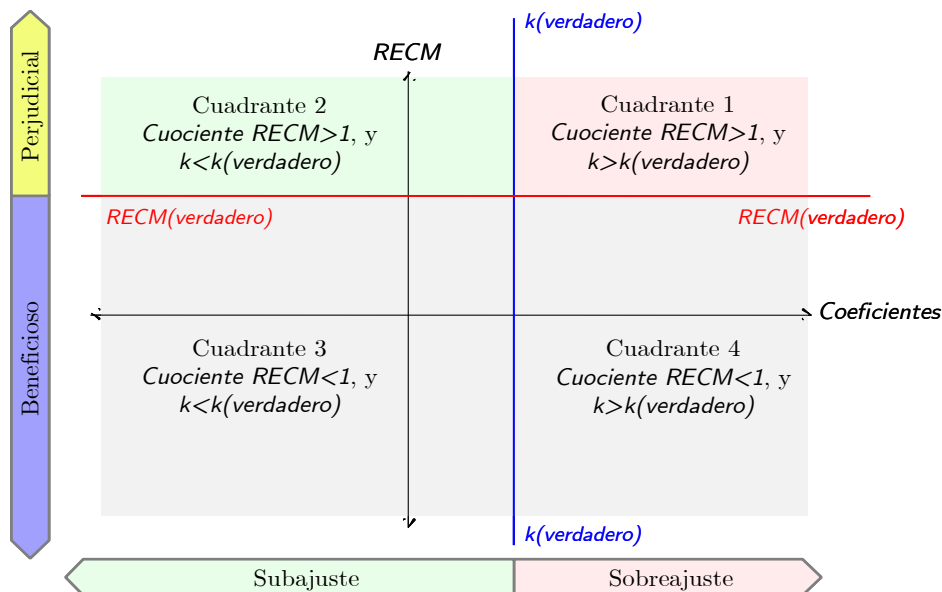
Para el análisis fuera de muestra se calcula, para cada orden desde 1 hasta 24 coeficientes, el cociente entre el promedio de la raíz cuadrada del error cuadrático medio (RECM) del modelo de orden k , $k \in \{1, \dots, 24\}$, y la RECM del modelo verdadero:

$$\text{Cuociente RECM} = \frac{\text{RECM Promedio modelo } k}{\text{RECM Modelo verdadero}} = \frac{\left[\frac{1}{T} \sum_{t=1}^T (y_{t|t} - y_{t|t-h}^{k,CI})^2 \right]^{\frac{1}{2}}}{\left[\frac{1}{T} \sum_{t=1}^T (y_{t|t} - y_{t|t-h}^{\text{verdadero}})^2 \right]^{\frac{1}{2}}}, \quad (8)$$

donde $y_{t|t-h}^{k,CI}$ es la proyección de y_t , h pasos adelante, con el orden k , escogido con el $CI = \{AIC; BIC; HQ\}$. $y_{t|t-h}^{\text{verdadero}}$ es la proyección del modelo verdadero, e $y_{t|t}$ es la serie original en el momento t . De esta forma, valores mayores (menores) a 1 indican que el modelo k presenta un peor (mejor) rendimiento predictivo que el modelo verdadero. Notar que para algunos valores de k no necesariamente existe una RECM (en ninguna de las 5000 replicaciones fue escogido, generando valores omitidos que no reciben tratamiento alguno), mientras que para otros valores de k puede existir más de una RECM, en cuyo caso se considera el promedio. Finalmente, se grafican los valores del *Cuociente RECM* a través de k , para los tres CI considerados. El gráfico resultante queda dividido en cuatro cuadrantes, definidos de acuerdo a: (i) *Cuociente RECM* mayor o menor a uno, y (ii) valores de k mayores y menores que el valor verdadero, tal como se presenta en la figura 1. El Cuadrante 1 (cuando *Cuociente RECM* > 1 y $k > k^{\text{verdadero}}$) señala los casos en los cuales

el sobreajuste es perjudicial para la capacidad predictiva.⁹ Para efectos predictivos, cualquier par ordenado (*Coficiente, Cuociente RECM*) que se ubique en los cuadrantes 1 y 2 es perjudicial. Por otro lado, cabe mencionar que si una proporción importante de casos de ubica en el Cuadrante 2, tampoco existirían ganancias de precisión predictivas asociadas al *subajuste*.

Figura 1: Clasificación de los resultados predictivos



Un punto en el Cuadrante 1 (4) indica que el *sobreajuste* es perjudicial (beneficioso) para la capacidad predictiva. En el Cuadrante 2 (3), el *subajuste* es perjudicial (beneficioso) para la capacidad predictiva. Fuente: Elaboración propia.

El tamaño de muestra disponible para las cuatro series analizadas puede considerarse *largo*. Por lo anterior, no se consideran versiones corregidas de los CI como los propuestos en Hurvich y Tsai (1993), Cavanaugh (1997), o McQuarrie (1999). Asimismo, en Medel y Salgado (2013) se realiza un ejercicio de comparación conjunta entre el rendimiento del AIC y BIC tanto dentro como fuera de muestra, encontrando que el BIC es superior tanto dentro como fuera de muestra sólo en un contexto asintótico. Dentro de muestra, sin embargo, el sobreajuste del AIC es estadísticamente significativo, pero no lo suficientemente despreciable respecto al BIC. Así, la elección de uno u otro CI en un contexto de muestra larga y finita es una respuesta dependiente, en último término, de los datos.

3 Resultados

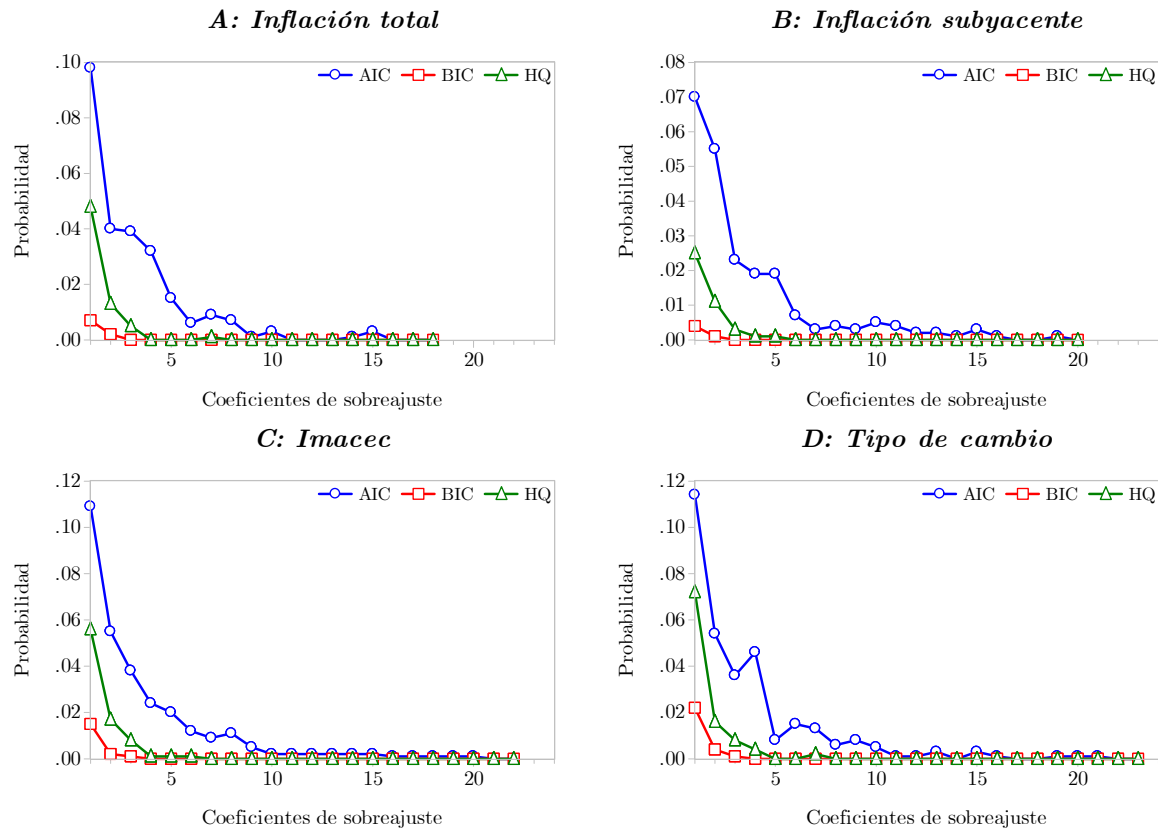
3.1 Probabilidad clásica de sobreajuste

Los resultados de la probabilidad clásica de sobreajuste se presentan en la figura 2. Como se mencionó anteriormente, estos se presentan en términos de los histogramas de las tres series s^{IC} .

⁹El Cuadrante 2 (cuando $Cuociente RECM > 1$ y $k < k^{verdadero}$) indica que el *subajuste* es perjudicial para la precisión predictiva. A su vez, el Cuadrante 3 (cuando $Cuociente RECM < 1$ y $k < k^{verdadero}$) indica que el *subajuste* es beneficioso para la precisión predictiva. Finalmente, el Cuadrante 4 (cuando $Cuociente RECM < 1$ y $k > k^{verdadero}$) indica que el *sobreajuste* es beneficioso para la precisión predictiva.

Los valores cuando $s^{IC} \leq 0$ no son presentados dado que no representan sobreajuste. Se observa que el mayor riesgo de sobreajuste se obtiene con el AIC, seguido por el HQ, y finalmente el BIC en los tres casos. Notar que este ordenamiento no es sorprendente debido a que el término de penalización del HQ es de grado intermedio comparado con el del AIC (bajo) y del BIC (alto).

Figura 2: Probabilidad clásica de sobreajuste con AIC, BIC y HQ



Fuente: Elaboración propia.

La mayor probabilidad alcanzada por el AIC es de un 32% con el Imacec, seguida de un 30% con el tipo de cambio. Las medidas de inflación presentan una probabilidad de 25 y 22% de sobreajuste con AIC, para la inflación total y la subyacente. Las menores probabilidades se alcanzan con el BIC. Estas corresponden a un 2 y 3% con el Imacec y el tipo de cambio. Con ambas medidas de inflación no sobrepasa un 1%. Considerando los tres CI, la mayor probabilidad siempre se obtiene solamente con un coeficiente de sobreajuste. Además, es más probable que el sobreajuste no exceda a los 10 coeficientes.¹⁰

Con todos los CI considerados, para las cuatro series analizadas, la probabilidad es no-creciente a partir de $s = 1$, excepto con el AIC para el tipo de cambio. Esto es indicativo de que la estrategia general a específica de términos consecutivos entrega efectivamente el modelo con menor sobreajuste—y de acuerdo al estadístico Durbin-Watson—, bien especificado. Sin embargo, el ejercicio

¹⁰Contrariamente al hallazgo de Granger y Jeon (2004), el valor k^{\max} pareciera no jugar un rol en la determinación de k , esto es, nunca se acumula masa de probabilidad en el valor k^{\max} . Esto probablemente se deba a la naturaleza estacionaria de las series y su estructura de autocovarianzas propiciada por las transformaciones.

predictivo fuera de muestra presentado más adelante indica un efecto perjudicial en la capacidad predictiva, pero de una magnitud distinta, entre el sub y el sobreajuste. Esto sugiere que las dificultades provienen en último término de errar el modelo verdadero independiente de la dirección.

3.2 ¿Por qué es importante analizar el sobreajuste?

Las consecuencias del sobreajuste no son directas ni han sido estudiadas en detalle. El problema se asocia, típicamente, a un menor rendimiento predictivo.¹¹ El argumento tradicional es que coeficientes adicionales exageran la dinámica de la variable dependiente y se es más propenso a errores de diversos tipos (Hawkins, 2004; Granger y Jeon, 2006). La idea de la modelación de una variable de interés, tanto de manera económica–basada en lo que la teoría económica indica–como estadística–basada en lo que indican pruebas estadísticas exclusivamente–, subyace en la idea de contar con una representación manejable que permita la experimentación, lo que en la práctica es prohibitivo. Tal como se señala en Hawkins (2004), el objetivo de tal tipo de experimentación son dos: (i) cuantificación, y (ii) predicción. Para ambos fines, el sobreajuste puede influir negativamente mediante los tres problemas antes mencionados.¹² En esta subsección se analiza el efecto del sub y sobreajuste sobre la predicción en particular. El problema de cuantificación se asocia primordialmente con la inferencia dentro de muestra.

El ejercicio predictivo descrito en la sección 2 permite cuantificar el costo de incurrir en sub o sobreajuste, *i.e.* de errar el modelo verdadero. Se consideran ambas desviaciones para complementar y dar robustez al análisis. En la figura 3 se presenta el *Cuociente RECM* para cada número de coeficientes k (para los cuales existe el *Cuociente*) elegidos con cada uno de los tres CI considerados, para los tres horizontes predictivos analizados. Valores mayores (menores) a 1 indican un peor (mejor) rendimiento predictivo respecto al modelo verdadero. Además se presenta, en líneas entrecortadas, el promedio de *Cuociente RECM* obtenido con cada CI. Notar que el Cuadrante 1, denominado $C(1)$, indica sobreajuste, mientras que el $C(2)$ subajuste, determinados por la línea vertical $k = k_0$ (ver figura 1). De esta manera, es posible visualizar y comparar el rendimiento predictivo promedio del sub versus sobreajuste con los tres CI simultáneamente.

La misma figura 3 además presenta, en la esquina superior derecha de cada gráfico, la probabilidad condicional del rendimiento predictivo, dado el sub o sobreajuste. Esto es, la probabilidad clásica de, dado el sub o sobreajuste, presentar un *Cuociente RECM* mayor a 1. Estas probabilidades se denominan por "Pr(1)" o "Pr(2)" para el Cuadrante 1 y 2, respectivamente, y se calcula promediando la probabilidad de los tres CI. Así, un valor "compactado" $\text{Pr}(1) = \alpha$ indica que del total de simulaciones que presentaron sobreajuste, una proporción α de ellas presentó un rendimiento predictivo inferior al del modelo verdadero (y una proporción $1 - \alpha$, un rendimiento superior). Como regla general, en especial porque considera los tres CI, una probabilidad mayor a 50% puede considerarse como un hecho regular que asienta alguna conclusión general.

En la tabla 3 se presentan dos paneles de resultados. El primero contiene el promedio de *Cuociente RECM* para los dos cuadrantes en los cuales el rendimiento predictivo es peor que el del modelo verdadero, $C(1)$ y $C(2)$. Este promedio se obtiene con las observaciones del *Cuociente RECM* para cada número de coeficientes k mayor y menor que el verdadero, graficadas en la figura 3. Así, por ejemplo, cualquier número mayor a 1 en la tabla que pertenezca a la columna $C(1)$, indica que

¹¹Este argumento es más común en un contexto AR, considerado como de coeficientes *endógenos*. En el caso con variables exógenas, el problema comúnmente analizado es el de regresión espuria.

¹²Estos son: (i) el de regresión espuria (Granger y Newbold, 1974), (ii) distorsionar la estimación de la función de impulso-respuesta de la variable de interés (Kilian, 2001), y (iii) perjudicar la precisión predictiva (Clark, 2004).

dado el sobreajuste (debido a que el $C(1)$ indica sobreajuste), el error de predicción es mayor que el del modelo verdadero. Por otro lado, si el valor es menor a 1, indica que dado el sobreajuste, la capacidad predictiva mejora respecto al modelo verdadero. De esta forma, si el sobreajuste es mayormente perjudicial para la capacidad predictiva que el subajuste, se espera observar valores mayores a 1 en la columna $C(1)$ que además sean mayores que su par $C(2)$. Por su parte, valores menores a 1 en la columna $C(2)$, indican que dado el subajuste, el rendimiento predictivo es mejor que el del modelo verdadero (representando la situación más plausible *a priori*). El segundo panel de la tabla 3, reproduce la probabilidad condicional de sub y sobreajuste de la figura 3 para facilitar la comprensión de los resultados. Finalmente, el tercer panel de la tabla 3 presenta el cociente, a través de cada CI y cada variable para un horizonte dado, entre el promedio de *Cuociente RECM* para un sobreajuste *bajo* ($k \leq 12$) y un sobreajuste *alto* ($k > 12$). Así, si este valor alcanza $(100 + \beta)\%$, entonces el sobreajuste alto provoca un empeoramiento promedio de $\beta\%$ respecto al sobreajuste bajo.

Estos resultados reflejan efectos dispares entre el sub y sobreajuste en la capacidad predictiva. En parte, este resultado es compartido con las conclusiones presentadas en Kilian (2001) relativas a la función de impulso respuesta. No es directo que un modelo con menor número de coeficientes que el modelo verdadero presente una reducción absoluta en la RECM de igual magnitud que un incremento en mismo número de coeficientes tiene aumentando la RECM. Esto es evidente en todos los casos, aunque menos marcado en la inflación total prediciendo a un mes adelante. En ese caso, el subajuste presenta (efectivamente) influencias positivas en la capacidad predictiva.

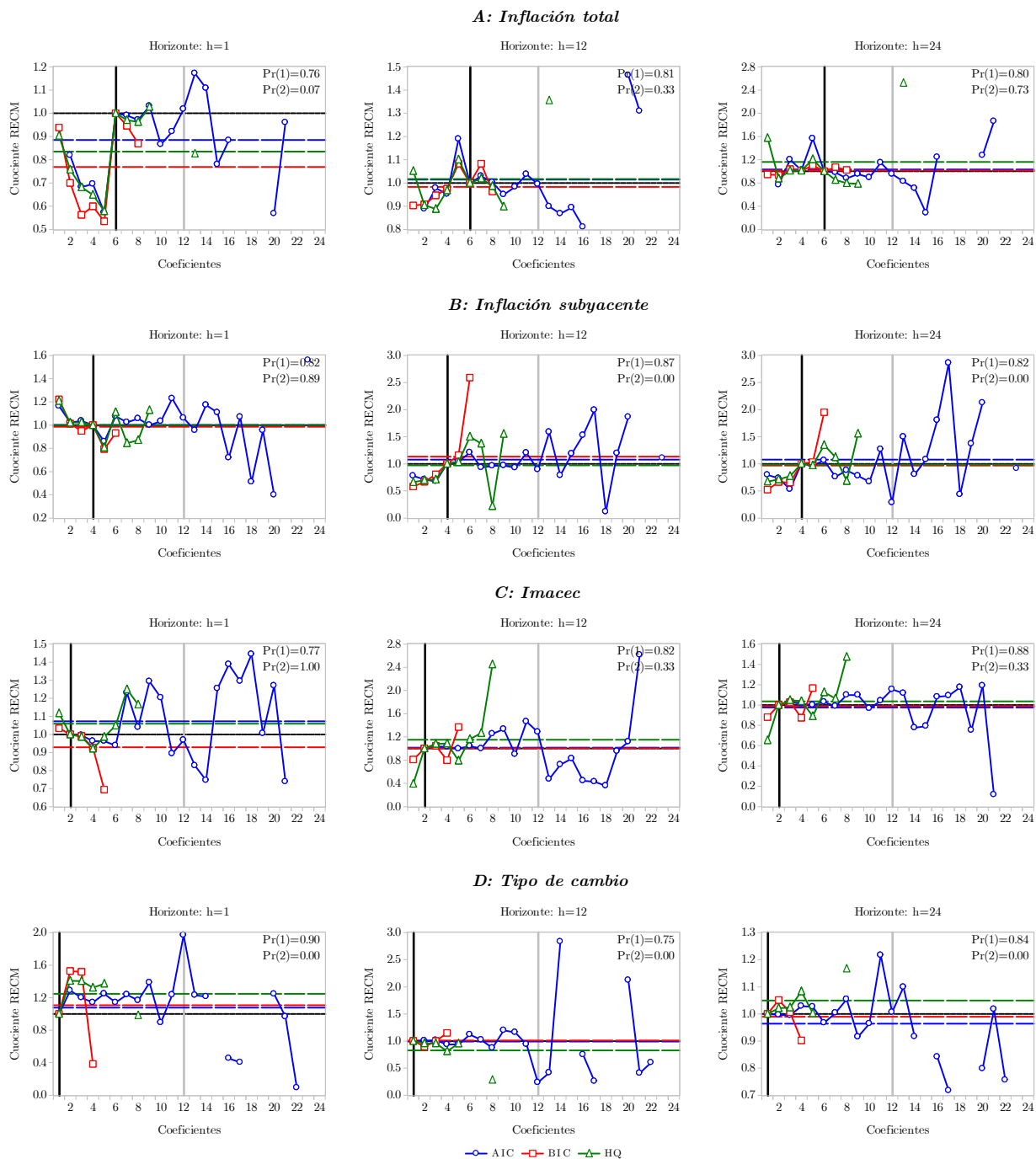
Se observa también que la dinámica exagerada de los modelos (número de coeficientes mayor a su frecuencia estacional, *i.e.* 12), con un número de coeficientes mayor a 13, estropea la capacidad predictiva fuertemente en todos los horizontes para las cuatro series. Notar que este resultado se encuentra con el AIC, dado que permite mayor sobreajuste. Cabe destacar que con algunos CI se logra un *Cuociente RECM* promedio menor que 1. Esto se hace distinguible con el BIC para la inflación total en los horizontes $h = 1$ y 12, el Imacec para $h = 1$, y el tipo de cambio para $h = 24$. Este resultado es particularmente relevante para el tipo de cambio, dado que el modelo verdadero (de un coeficiente) no permite subajuste.

La probabilidad condicional de peor rendimiento predictivo indica que el sobreajuste es mayormente perjudicial para la capacidad predictiva que el subajuste, con excepciones para la inflación subyacente y el Imacec prediciendo a un mes adelante. Esto se debe a que en todos los casos (exceptuando los anteriormente mencionados) el valor de $\text{Pr}(1)$ es mayor a 0.50, mientras que el valor de $\text{Pr}(2)$ es conjuntamente menor a 0.50, llegando incluso a 0.00 (en casos distintos a los que por definición es 0.00).

En resumen, los resultados confirman que los efectos del sub y sobreajuste son igualmente negativos para la capacidad predictiva, aunque en distinta magnitud. La predicción al largo plazo es la que se ve menos perturbada por subajuste, y fuertemente afectada con el sobreajuste. El AIC es el CI que fomenta mayormente el sobreajuste, estropeando la capacidad predictiva en todas las series y horizontes.

Se observan también efectos heterogéneos entre las series debido al sobreajuste. Para la inflación subyacente y el tipo de cambio el riesgo se avizora como más importante, especialmente prediciendo en el plazo más corto. Para la inflación subyacente y el Imacec, el riesgo es potencialmente problemático con un alto número de coeficientes de sobreajuste. En promedio para todas las series y CI, se encuentra que un sobreajuste promedio mayor a los 12 coeficientes—*i.e.* de un orden mayor a

Figura 3: *Cuociente RECM* y probabilidad condicional (*)



(*) La línea vertical cercana al origen indica k (verdadero), mientras que la segunda indica $k=12$. Las líneas entrecortadas indican el promedio de *Cuociente RECM* a través del total de coeficientes. $Pr(c)$ indica la probabilidad empírica obtenida del total de casos ubicados en el cuadrante c . Fuente: Elaboración propia.

Tabla 3: Estadísticos para análisis predictivo multihorizonte

	Inflación total		Inflación subyacente		Imacec		Tipo de cambio	
	$C(1)$	$C(2)$	$C(1)$	$C(2)$	$C(1)$	$C(2)$	$C(1)$	$C(2)$ (*)
A. Cuociente RECM promedio (**)								
$h=1$	0.931	0.691	0.934	1.073	1.002	1.075	1.174	0.00
$h=12$	1.036	0.982	1.386	0.698	1.131	0.601	0.932	0.00
$h=24$	1.095	1.094	1.259	0.673	1.034	0.766	1.002	0.00
B. Probabilidad condicional de sub y sobreajuste (***)								
$h=1$	0.76	0.07	0.82	0.89	0.77	1.00	0.90	0.00
$h=12$	0.81	0.33	0.87	0.00	0.82	0.33	0.75	0.00
$h=24$	0.80	0.73	0.82	0.00	0.88	0.33	0.84	0.00
C. Cuociente entre RECM de sobreajuste $k \leq 12$ y RECM de sobreajuste $k > 12$								
$h=1$					91%			
$h=12$					101%			
$h=24$					121%			

(*) Valores iguales a 0.00 por definición. (**) Celdas destacadas indican valores mayores a 1. (***) También presentada en la figura 3. Fuente: Elaboración propia.

la frecuencia anual de los datos incrementa el error de proyección en alrededor de un 20% respecto al error promedio de sobreajuste de hasta 12 coeficientes, prediciendo 24 meses adelante.

4 Resumen y comentarios finales

En este trabajo se estima mediante simulaciones de Monte Carlo la probabilidad clásica de sobreajuste con los criterios de información (CI) de Akaike, Schwarz y Hannan-Quinn (AIC, BIC y HQ), calibrada con los datos chilenos de inflación total, inflación subyacente, Imacec y retorno mensual del tipo de cambio nominal peso-dólar estadounidense.

Los resultados de este trabajo indican que el mayor riesgo de sobreajuste se obtiene con el AIC, seguido por el HQ y finalmente el BIC, para las cuatro series consideradas. Las probabilidades más altas de sobreajuste se alcanzan con el AIC, llegando a un 32 y un 30% con el tipo de cambio y el Imacec, respectivamente, seguidas por un 25 y 22% para la inflación total y subyacente. El BIC, por su parte, exhibe la menor probabilidad de sobreajuste: alcanza a un 3 y 2% con el tipo de cambio y el Imacec, respectivamente, y un 1% tanto para la inflación total como subyacente. Considerando los tres CI, la mayor probabilidad siempre se obtiene solamente con un coeficiente de sobreajuste. Además, es más probable que el sobreajuste no exceda a los 10 coeficientes.

Estos resultados son importantes en la medida que representan el riesgo de sobreajuste—*i.e.* que tan alejado se está del modelo verdadero—al cual se está sometido al modelar dichas variables, u otras con características estadísticas similares, mediante CI de uso frecuente. A través de un análisis predictivo multihorizonte fuera de muestra, se analiza y compara el rendimiento de cada CI con sub y sobreajuste. A pesar de no ser simétricos, estos resultados indican que, en promedio, se asocia al sobreajuste una mayor probabilidad de que el error de proyección sea mayor que el del modelo verdadero, comparada con el caso con subajuste. Lo anterior representa un desafío latente para la inferencia estadística—o prueba empírica de una teoría económica—basada en modelos no paramétricos. Finalmente, no todas las series consideradas presentan el mismo deterioro en la capacidad predictiva debido al sobreajuste. Este es más perjudicial sólo cuando es de un orden, aproximadamente, mayor al de la frecuencia de los datos. Un resultado importante es que un

sobreajuste promedio mayor a los 12 coeficientes—*i.e.* de un orden mayor a la frecuencia anual de los datos—incrementa el error de proyección en alrededor de un 20% respecto al error promedio de sobreajuste de hasta 12 coeficientes, prediciendo 24 meses adelante.

Referencias

1. Akaike, H., 1973, *Information Theory and an Extension of the Maximum Likelihood Principle*, en Petrov, B.N. y F. Csaki, editores, *Proceedings of the 2nd Symposium on Information Theory*, Akademia Kiado, Budapest, Hungría.
2. Akaike, H., 1974, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control* **19**(6): 716-723.
3. Box, G.E.P. y G.M. Jenkins, 1970, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, EEUU.
4. Canova, F., 2007, *Methods for Applied Macroeconomic Research*, Princeton University Press, EEUU.
5. Cavanaugh, J.E., 1997, "Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria," *Statistics and Probability Letters* **33**(2): 201-208.
6. Clark, T.E., 2004, "Can Out-of-Sample Forecast Comparisons Help Prevent Overfitting?," *Journal of Forecasting* **23**(2): 115-139.
7. Elliott, G. y A. Timmermann, 2008, "Economic Forecasting," *Journal of Economic Literature* **46**(1): 3-56.
8. Faust, J. y J. Wright, 2013, *Forecasting Inflation*, en Elliot, G. y A. Timmermann, editores, *Handbook of Economic Forecasting, volume 2*, Elsevier, Holanda.
9. Granger, C.W.J. y Y. Jeon, 2004, "Forecasting Performance of Information Criteria with Many Macro Series," *Journal of Applied Statistics* **31**(10): 1227-1240.
10. Granger, C.W.J. y Y. Jeon, 2006, "Dynamics of Model Overfitting Measured in Terms of Autoregressive Roots," *Journal of Time Series Analysis* **27**(3): 347-365.
11. Granger, C.W.J. y P. Newbold, 1974, "Spurious Regressions in Econometrics," *Journal of Econometrics* **2**(1): 111-120.
12. Hansen, P.R., 2009, "In-Sample Fit and Out-of-Sample Fit: Their Joint Distribution and its Implications for Model Selection," versión preliminar del 23 de Abril de 2009, Departamento de Economía, Stanford University, EEUU.
13. Hannan, E.J. y B.G. Quinn, 1979, "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society B* **41**: 190-195.
14. Hawkins, D.M., 2004, "The Problem of Overfitting," *Journal of Chemical Information and Computer Sciences* **44**: 1-12.
15. Hoover, K.D. y S.J. Pérez, 1999, "Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search," *The Econometrics Journal* **2**(2): 167-191.

16. Hurvich, C.M. y C.-L. Tsai, 1993, "A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection," *Journal of Time Series Analysis* **14**: 271-279.
17. Kilian, L., 2001, "Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order," *Journal of Forecasting* **20**(3): 161-179.
18. Koehler, A.B. y E.S. Murphree, 1988, "A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order," *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **37**(2): 187-195.
19. Kullback, S. y R.A. Leibler, 1951, "On Information and Sufficiency," *Annals of Mathematical Statistics* **22**: 79-86.
20. Kullback, S., 1968, *Information Theory and Statistics*, Dover, New York, EEUU.
21. Lütkepohl, H., 1985, "Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process," *Journal of Time Series Analysis* **6**(1): 35-52.
22. McQuarrie, A.D., 1999, "A Small-Sample Correction for the Schwarz SIC Model Selection Criterion," *Statistics and Probability Letters* **44**(1): 79-86.
23. Medel, C.A., 2011, "¿Akaike o Schwarz? ¿Cuál utilizar para predecir el PIB chileno?," *Monetaria* **XXXIV**(4): 591-615.
24. Medel, C.A., 2013, "How Informative are In-sample Information Criteria to Forecasting? The Case of Chilean GDP," *Latin American Journal of Economics* **50**(1): 133-161.
25. Medel, C.A. y S. Salgado, 2013, "Does the BIC Estimate and Forecast Better than the AIC?," *Economic Analysis Review* **28**(1): 47-64.
26. Nishii, R., 1984, "Asymptotic Properties of Criteria for Selection of Variables in Multiple Regression," *Annals of Statistics* **12**(2): 758-765.
27. Schwarz, G.E., 1978, "Estimating the Dimension of a Model," *Annals of Statistics* **6**(2): 461-464.
28. Seghouane, A.-K., 2006, "A Note on Overfitting Properties of KIC and KIC_c ," *Signal Processing* **86**(10): 3055-3060.
29. Shibata, R., 1976, "Selection of the Order of an Autoregressive Model by Akaike Information Criterion," *Biometrika* **63**(1): 117-126.
30. Stock, J.H. y M.W. Watson, 2007, *A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series*, en Engle, R.E. y H. White, editores, *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press.
31. Stock, J.H. y M.W. Watson, 2009, *Phillips Curve Inflation Forecasts*, en Fuhrer, J., Y. Kordzycki, J.S. Little y G. Olivei, editores, *Understanding Inflation and the Implications for Monetary Policy, A Phillips Curve Restrospective*, MIT Press, EEUU.
32. Xu, C.J. y I. McLeod, 2012, "Further Asymptotic Properties of the Generalized Information Criteria," *Electronic Journal of Statistics* **6**: 656-663.