

UN TEST CONJUNTO DE SUPERIORIDAD PREDICTIVA PARA LOS PRONÓSTICOS DE INFLACIÓN CHILENA*

Pablo M. Pincheira B. **

I. INTRODUCCIÓN

Cuando coexisten diversos pronósticos para una misma variable económica, resulta natural investigar la precisión relativa entre ellos. Tradicionalmente, esto se realiza a través de múltiples comparaciones bilaterales del Error Cuadrático Medio (ECM) de proyección. Al momento de hacer inferencia, la manera usual de proceder consiste en formular una serie de hipótesis de igualdad predictiva entre dos pares de métodos o modelos predictivos. Luego se aplica algún test como el de Diebold y Mariano (1995), el de West (1996), o el de Giacomini y White (2006). Ejemplos de artículos en los que estas estrategias han sido utilizadas son los de Pincheira y García (2009, 2012), Andersson et al. (2007), Groen et al. (2009), Ang et al. (2007), Pincheira y Álvarez (2009) y Pincheira (2010a, 2010b), por citar solo algunos.

Esta tradicional metodología para comparar la capacidad predictiva entre varios modelos presenta la natural complejidad asociada a la formulación de hipótesis múltiples, que se refiere a la dificultad de mantener bajo control el error de tipo I del test¹. Sin la formulación de un contraste conjunto que esté específicamente diseñado para tener un error de tipo I no superior a un nivel α , la aplicación de repetidas pruebas bilaterales no necesariamente garantizará un test del tamaño deseado².

La evaluación predictiva de múltiples series de pronósticos puede ser más compleja aun cuando las proyecciones se confeccionan en distintos momentos del tiempo. En ese escenario cabe esperar que las predicciones elaboradas con más información tengan un mejor desempeño que aquellas construidas con menos información. Paralelamente, también debería ser esperable que, bajo ciertos supuestos, los pronósticos a horizontes cortos tuviesen un desempeño mejor que aquellos diseñados para horizontes más largos. De hecho, Patton y Timmermann (2012) muestran que las proyecciones óptimas bajo pérdida cuadrática deberían seguir un

* Proyecto Fondecyt 11080109. Agradezco especialmente a Carlos Medel por su brillante colaboración y asistencia. Las ideas y opiniones presentadas en este trabajo no necesariamente representan la visión del Banco Central de Chile o sus autoridades. Los errores u omisiones son responsabilidad exclusiva del autor.

** Gerencia de Investigación Económica, Banco Central de Chile. E-mail: ppinchei@bcentral.cl

¹ El error de tipo I corresponde a la probabilidad de rechazar la hipótesis nula cuando ella es verdadera. A este error también se le denomina "tamaño del test".

² Una manera de paliar este problema hace uso de las llamadas cotas de Bonferroni. En términos sencillos esta metodología plantea transformar una dócima de hipótesis conjunta de tamaño α , en N pruebas de tamaño α/N que solo involucran hipótesis univariadas. No obstante la facilidad de implementación de esta estrategia, en general es criticada por poder llegar a ser en extremo conservadora, sin tener capacidad real de rechazar la hipótesis nula. Esto es mencionado en el trabajo de Patton y Timmermann (2012) y comprobado con simulaciones en la subsección IV.3 de este documento.

comportamiento de esta naturaleza³. Cuando la teoría econométrica sugiere que un cierto pronóstico debe ser mejor que otro, cobra sentido plantear una hipótesis nula de superioridad de habilidad predictiva y ya no una de igualdad de capacidad predictiva. Algunos tests de esta naturaleza han sido presentados por White (2000), Hansen (2005) y recientemente también por Patton y Timmerman (2012), quienes utilizan implícitamente tests de superioridad predictiva basados en los trabajos de Wolak (1987, 1989).

Para el caso chileno existen diversas proyecciones de inflación publicadas periódicamente en distintos momentos del tiempo. Dichas proyecciones corresponden a las de *Bloomberg*, *Consensus Forecasts* y la Encuesta de Expectativas Económicas (EEE) realizada por el Banco Central de Chile (BCCh). Adicionalmente, siempre es posible contar con proyecciones de modelos simples de series de tiempo univariados como las presentadas en Pincheira y García (2009, 2012). Al requerir solo de los rezagos de la inflación, estos simples modelos utilizan un conjunto de información muy pequeño, que supondremos está contenido en los conjuntos utilizados por los analistas que contemplan las diversas encuestas mencionadas anteriormente.

Frente a este perfil de proyecciones de inflación construidas a base de conjuntos de información que supondremos anidados, cabe naturalmente preguntarse acerca de la superioridad predictiva de aquellas proyecciones construidas sobre la base de mayor información. Paralelamente, y en el caso de proyecciones a distintos horizontes, también resulta natural preguntarse si la superioridad de las proyecciones construidas para horizontes más cercanos es tal. A objeto de responder estas preguntas, y evitar los problemas asociados al correcto control del error de tipo I en pruebas múltiples, en el presente artículo, construimos dos tests basados en los estadísticos del máximo y mínimo de un conjunto de variables aleatorias⁴. Con estos tests nos abocamos a contrastar empíricamente si los pronósticos de inflación para Chile construidos sobre la base de mayor información son superiores a los construidos a base de conjuntos de información más pequeños. También estudiamos si los pronósticos construidos para horizontes cortos son superiores a los construidos para horizontes largos. Estas evaluaciones se hacen tanto en términos del ECM de las proyecciones consideradas, como en términos del Pronóstico Cuadrático Medio (PCM), en línea con la propuesta de Patton y Timmermann (2012).

Nuestros resultados indican que en términos del ECM no es posible rechazar la hipótesis nula de optimalidad del perfil de proyecciones considerado. Esto ocurre tanto cuando evaluamos las proyecciones a evento fijo como cuando las evaluamos a evento variable⁵. Similarmente, la hipótesis nula de optimalidad tampoco puede ser rechazada cuando evaluamos los PCM a

³ De ahora en adelante usaremos simplemente el término *optimalidad* como sinónimo de la expresión *optimalidad bajo pérdida cuadrática*.

⁴ White (2000) y Pincheira (2006, 2012a) han desarrollado en forma independiente tests de esta naturaleza en contextos algo distintos al que nos ocupa en el presente documento.

⁵ La evaluación de pronósticos a evento fijo significa que se consideran proyecciones construidas en distintos momentos en el tiempo, pero todas intentando pronosticar un evento que ocurrió en una fecha inamovible. Un claro ejemplo de esto ocurre cuando a los analistas económicos se les pregunta por el pronóstico de inflación a diciembre. A lo largo del año los analistas pueden cambiar sus pronósticos a medida que se acerca el mes de diciembre y se acorta el horizonte de proyección, no obstante, a lo largo del año, el objetivo es siempre la inflación acumulada a diciembre. La evaluación a evento variable significa comparar proyecciones para distintos eventos. Un ejemplo de esto ocurre cuando se desea comparar las proyecciones de los analistas en horizontes de 12 meses. Esto significa, que a medida que pasa el tiempo, la variable a predecir también cambia. En enero será la inflación de enero del próximo año la variable a predecir, por ejemplo. En febrero será la inflación de febrero del próximo año, y así sucesivamente. Una evaluación a evento variable también puede incluir comparar proyecciones a distintos horizontes, pero realizadas en forma simultánea.

evento variable. Sin embargo, el test conjunto para los PCM rechaza la hipótesis de optimalidad cuando se consideran las proyecciones a evento fijo, sugiriendo la existencia de espacio para perfeccionar el perfil de proyecciones.

Más allá de los resultados empíricos que obtenemos en el artículo, creemos que estamos haciendo una contribución en dos dimensiones. En primer lugar, y dentro de los límites de la literatura revisada, pensamos que este es el primer artículo en testear conjuntamente para Chile el perfil no creciente de los ECM, para un conjunto de proyecciones realizadas en distintos momentos del tiempo. En segundo lugar, no conocemos de ningún otro artículo que evalúe para Chile el ordenamiento esperado que los PCM de las proyecciones de inflación deberían presentar si los pronósticos fueran óptimos.

El resto del trabajo se organiza de la siguiente manera. En la sección II se revisa brevemente la literatura relacionada. Luego, en la sección III se entregan algunas características de los datos utilizados, para después, en la sección IV resumir la derivación de las propiedades de optimalidad y presentar los métodos de inferencia que serán utilizados. El análisis empírico se realiza en la sección V. Finalmente, en la sección VI se entrega un resumen de las principales conclusiones de este trabajo.

II. REVISIÓN DE LA LITERATURA

Existen numerosos trabajos que evalúan las proyecciones públicas de diversas variables económicas tales como crecimiento del PIB e inflación. Entre ellos destacamos las recientes contribuciones de Croushore (2010), Capistrán y López-Moctezuma (2010), Groen et al. (2009), Pincheira (2010a, 2010b), así como el ya más tradicional artículo de Ang et al. (2007). En general, en casi ninguno de estos artículos se considera la dinámica temporal en la construcción de pronósticos, siendo el artículo de Pincheira y Álvarez (2009) la única excepción por nosotros conocida.

Otros trabajos evalúan pronósticos a través de sus propiedades de optimalidad bajo alguna función de pérdida, como las descritas en Elliot y Timmermann (2008). Llama fuertemente la atención que una gran mayoría de esta literatura encuentre que diversas series de proyecciones no satisfacen criterios de optimalidad bajo pérdida cuadrática. Entre estos estudios destacan los trabajos de Joutz y Stekler (2000), Capistrán y López-Moctezuma (2010), Bowles et al. (2007) y Loungani (2001).

Para el caso de Chile también hay cierta literatura relevante. Por ejemplo, Chumacero (2001) documenta una subestimación sistemática de las proyecciones privadas del PIB durante 1986-1998. Posteriormente, Bentancor y Pincheira (2008, 2010) analizan el sesgo y el exceso de autocorrelación de los pronósticos de inflación chilena provenientes de la EEE. También en esta línea de análisis se encuentra el trabajo de Pincheira y Álvarez (2009) quienes evalúan una serie de proyecciones de inflación, incluyendo aquellas provenientes de *Bloomberg*, *Consensus Forecasts* y la EEE. Uno de los objetivos de ese trabajo es evaluar la superioridad predictiva de los pronósticos contruidos sobre la base de mayor información. El análisis se basa solamente en estudiar los ECM de los distintos pronósticos mediante una secuencia de tests bilaterales de igualdad de capacidad predictiva, con lo que no queda claro que en ese trabajo se haya controlado en forma adecuada el error de tipo I. Con todo, los autores concluyen que no es posible rechazar la hipótesis nula de mayor o igual precisión mientras mayor sea el conjunto de información utilizado en la proyección.

Es importante destacar que la primera parte del presente trabajo generaliza el artículo de Pincheira y Álvarez (2009) ya que no solo se contrasta una hipótesis de superioridad predictiva similar en términos del ECM, sino que también se contrasta en términos de los PCM. Adicionalmente, en el presente artículo se desarrolla y aplica un test conjunto para tener un control adecuado del error de tipo I.

De acuerdo con esta breve y selecta revisión de la literatura, se infiere que, en un alto porcentaje, las proyecciones de variables macro están sujetas a ineficiencias, y que pocos trabajos enfocan la evaluación de proyecciones de manera conjunta y considerando la dinámica temporal en la elaboración de las proyecciones, lo que constituye un aspecto distintivo del presente trabajo. En las siguientes secciones se observará cómo nuestros resultados se ajustan a los hallazgos de la literatura relacionada.

III. DATOS

El análisis se realiza con datos chilenos de frecuencia mensual. La fuente de los datos efectivos es el Instituto Nacional de Estadísticas (INE), mientras las proyecciones provienen de *Bloomberg*, *Consensus Forecasts*, la EEE del BCCh, y una combinación de modelos de series de tiempo presentados en Pincheira y García (2009, 2012). Todas las series tienen frecuencia mensual, están denominadas en porcentaje y representan la variación anual del IPC. Para algunas pruebas y estadísticos descriptivos se considera la mayor muestra disponible sin observaciones faltantes de cada proyección, en tanto el test conjunto y otros análisis se realizan con una muestra balanceada desde diciembre del 2003 hasta diciembre del 2009 (73 observaciones). En el gráfico 1 se presenta la serie original de inflación desde enero del 2001 hasta diciembre del 2009. Se advierte un importante aumento de la inflación desde mediados del 2007 hasta fines del 2009.

Gráfico 1

Inflación efectiva en Chile, 2001-2009

(variación porcentual en doce meses)



Fuente: Instituto Nacional de Estadísticas (INE).

1. Bloomberg

Las proyecciones que denominamos bajo el nombre de “Bloomberg” corresponden a la media de las respuestas de la encuesta efectuada por *Bloomberg* a distintos analistas e instituciones financieras. Se dispone de una muestra que abarca desde enero del 2001 hasta diciembre del 2009. Estas proyecciones se hacen públicas aproximadamente una semana antes de conocerse el dato efectivo de inflación y, de hecho, entre todas las proyecciones que consideramos, son las últimas en darse a conocer antes de la publicación del dato efectivo. Haremos el supuesto de que el horizonte de proyección de esta encuesta es de un mes, aunque en términos efectivos este horizonte es de alrededor de una semana. Existe una observación faltante (noviembre 2003) la que no recibe un tratamiento especial, y es la que condiciona la selección de la muestra alternativa que se utiliza para los test de superioridad predictiva.

2. Consensus Forecasts

Las proyecciones que denominamos bajo el nombre de "Consensus" corresponden a la media de las respuestas de la encuesta efectuada por *Consensus Forecasts* a distintos analistas. La encuesta de *Consensus* se realiza alrededor de tres semanas antes de conocerse los datos efectivos de inflación. Esta encuesta provee proyecciones en horizontes de uno a seis meses⁶. La muestra existente considera datos desde julio de 1993 hasta la actualidad. Sin embargo, se dispone de datos sin observaciones faltantes solo desde abril del 2001 en adelante para los seis horizontes de proyección.

3. Encuesta de expectativas económicas

La EEE es levantada por el BCCh a comienzos de cada mes, y entrega proyecciones de inflación a uno, tres y doce meses hacia adelante⁷. Se aplica a un selecto grupo de académicos, consultores, ejecutivos y asesores del sector privado, quienes informan, además de sus proyecciones de inflación, pronósticos de otras variables económicas de interés. Dado que las proyecciones individuales de cada analista no son publicadas por el BCCh, utilizamos la mediana de las proyecciones. Este dato es de público acceso a través del sitio web del BCCh⁸. La EEE comenzó a operar el año 2000 y en varias oportunidades se modificó su contenido. Salvo cambios menores, ha permanecido con pocas alteraciones desde noviembre del 2004. Se cuenta con una muestra sin observaciones faltantes desde enero del 2001 hasta la actualidad para pronósticos un mes hacia adelante, desde noviembre del 2001 para pronósticos tres meses hacia adelante y desde septiembre del 2001 para doce meses hacia adelante. En promedio, durante el período 2000-2009, 35 personas contestaron el cuestionario cada mes⁹.

4. Proyección ESARIMA

Corresponde al promedio aritmético de los pronósticos provenientes de ocho modelos de series de tiempo de naturaleza similar. Dichos modelos se describen en Pincheira y García (2009, 2012) y corresponden a distintas especificaciones de modelos ESARIMA (*extended seasonal ARIMA*) que se pueden anidar en la siguiente expresión:

$$\pi_t - \pi_{t-1} = \delta + \rho(\pi_{t-1} - \pi_{t-2}) + \varepsilon_t - \theta_E \varepsilon_{t-1} - \theta_E \varepsilon_{t-12} + \theta_3 \varepsilon_{t-13},$$

donde π_t es la inflación en términos de variación a 12 meses en t , ε_t es un término de error, y δ , ρ , θ , θ_E y θ_3 , son parámetros a estimar (E es la frecuencia de los datos, que para el caso

⁶ También hay ciertas proyecciones a horizontes mayores, pero en número bastante limitado, razón por la cual no se utilizan en este estudio.

⁷ Los pronósticos a tres meses son deducidos por el autor a base de la información de la EEE. Básicamente, la EEE reportaba para el período de análisis el pronóstico un mes adelante y el promedio de la inflación esperada, en los dos meses siguientes. De esta información es posible construir un pronóstico tres meses adelante utilizando la aproximación logarítmica de la inflación. Este pronóstico se construye como la suma de la inflación esperada en un mes, más dos veces el promedio de la inflación esperada, en los dos meses siguientes. La aproximación logarítmica funciona adecuadamente cuando las tasas de variación anual son menores de 10%, lo que claramente ocurre para la inflación chilena en períodos cortos.

⁸ http://www.bcentral.cl/estadisticas-economicas/series-indicadores/index_ee.htm

⁹ Una descripción técnica de la EEE y una evaluación de sus resultados se encuentra en Pedersen (2010). Adicionalmente, observamos que los pronósticos de corto plazo provenientes de la EEE tienen como objetivo predecir la inflación mensual y no la acumulada en 12 meses. Ha sido necesario entonces transformar estas predicciones a inflación acumulada en 12 meses. También es importante destacar que la información reportada por la EEE contiene un solo decimal. Esto significa que podría haber discrepancias, por problemas de redondeo, con otros pronósticos contruidos directamente a partir el IPC y que consideran en su análisis dos y hasta tres decimales.

mensual corresponde a $E = 12$). Los ocho modelos se generan por la combinatoria de casos en que δ , ρ , y θ_3 son estimados o restringidos a cero. Notemos que cuando θ_3 es estimado, su estimación es restringida a satisfacer la condición $\hat{\theta}_3 = \hat{\theta}_{\theta_3}$. Estos ocho modelos siempre incluyen los términos que acompañan a los parámetros θ y ε . Con estos modelos se construyen pronósticos para siete distintos horizontes: de uno a seis meses y un año hacia adelante. La muestra para los siete horizontes considerados comienza en enero del 2001 y finaliza en diciembre del 2009. El cuadro 1 resume los períodos muestrales que cubre cada fuente de proyecciones considerada.

Cuadro 1

Períodos muestrales sin observaciones faltantes

	ESARIMA	EEE	Consensus Forecasts	Bloomberg
$h = 1$	Ene-2001 — dic-2009 (108)	Ene-2001 — dic-2009 (108)	Abr-2001 — dic-2009 (105)	Dic-2003 — dic-2009 (73)
$h = 2$	Ene-2001 — dic-2009 (108)	-	Abr-2001 — dic-2009 (105)	-
$h = 3$	Ene-2001 — dic-2009 (108)	Nov-2001 — dic-2009 (98)	Abr-2001 — dic-2009 (105)	-
$h = 4$	Ene-2001 — dic-2009 (108)	-	Abr-2001 — dic-2009 (105)	-
$h = 5$	Ene-2001 — dic-2009 (108)	-	Abr-2001 — dic-2009 (105)	-
$h = 6$	Ene-2001 — dic-2009 (108)	-	Abr-2001 — dic-2009 (105)	-
$h = 12$	Ene-2001 — dic-2009 (108)	Sep-2001 — dic-2009 (100)	-	-

Fuente: Elaboración propia.

Nota: El número entre paréntesis denota el número de observaciones.

En el cuadro 2 se presentan algunos estadísticos descriptivos de las series tanto para la muestra completa como para la submuestra utilizada en los tests de superioridad predictiva (diciembre 2003 – diciembre 2009).

Cuadro 2

Estadísticos típicos de todas las series

	Muestra completa (cuadro 1)					Muestra: dic. 2003 - dic. 2009 (73 datos)				
	Original	ESARIMA	EEE	C. Forecast	Bloomberg	Original	ESARIMA	EEE	C. Forecast	Bloomberg
Promedio	3,45	3,48	3,43	3,42	3,63	3,66	3,67	3,62	3,62	3,63
Desv. est.	2,51	2,53	2,46	2,49	2,94	2,99	3,01	2,94	2,93	2,94
Máximo	9,90	9,96	9,54	9,50	9,50	9,90	9,96	9,54	9,50	9,50
Mínimo	-2,30	-2,10	-1,81	-1,90	-1,90	-2,30	-2,10	-1,81	-1,90	-1,90
Coef. var.	0,72	0,72	0,71	0,72	0,81	0,81	0,82	0,81	0,81	0,81
RECM $h = 1$	-	0,44	0,32	0,30	0,24	-	0,45	0,34	0,32	0,25
RECM $h = 2$	-	0,81	-	0,64	-	-	0,88	-	0,71	-
RECM $h = 3$	-	1,12	0,99	0,80	-	-	1,26	1,05	0,89	-
RECM $h = 4$	-	1,43	-	0,96	-	-	1,65	-	1,10	-
RECM $h = 5$	-	1,76	-	1,14	-	-	2,05	-	1,32	-
RECM $h = 6$	-	2,08	-	1,30	-	-	2,44	-	1,53	-
RECM $h = 12$	-	3,48	2,06	-	-	-	4,15	2,34	-	-

Fuente: Elaboración propia.

Nota: Promedio, desviación estándar, máximo y mínimo de la serie un mes adelante, medida en puntos porcentuales. RECM $h = j$ significa raíz del error cuadrático medio j meses adelante. La 'Muestra completa' varía según lo reportado en el cuadro 1. Para la serie original se considera una muestra desde enero del 2001 hasta diciembre del 2009.

Las últimas filas del cuadro 2 muestran los ECM muestrales asociados a cada fuente de proyección. Recordemos que el $ECM(h)$ se define por:

$$ECM(h) = \mathbb{E}(Y_{t+h} - y_t^f(h))^2$$

en que $y_t^f(h)$ es la proyección de Y_{t+h} construida en el instante t . La raíz cuadrada del ECM se denota por $RECM$. Al ser un momento poblacional, el ECM debe ser estimado. Un estimador natural está dado por el promedio muestral de los errores de pronóstico al cuadrado:

$$\widehat{ECM}(h) = \frac{1}{T} \sum_t (Y_{t+h} - y_t^f(h))^2$$

Las últimas filas del cuadro 2 nos muestran tres hechos estilizados que nos parecen interesantes de destacar y que se verifican en las dos muestras consideradas. En primer lugar, se observa que las proyecciones presentan una estimación del ECM relativamente baja a horizontes menores de un año. En segundo lugar se observa que las proyecciones que presumiblemente se construyen con mayor información presentan un menor ECM estimado. Así, se verifica que las proyecciones de *Bloomberg* son más precisas que las de *Consensus*, las de *Consensus* más precisas que las de la *EEE*, y las proyecciones de la *EEE* más precisas que las del promedio de modelos *ESARIMA*. En tercer lugar, destacamos que para todas las proyecciones el ECM crece con el horizonte de proyección. Como mostraremos posteriormente, estos dos últimos puntos en principio son coherentes con la optimalidad del perfil de pronósticos analizado.

Así como hemos definido el ECM, podemos definir el PCM de la siguiente manera:

$$PCM(h) = \mathbb{E}(y_t^f(h))^2$$

que naturalmente puede ser estimado como el promedio al cuadrado de las proyecciones correspondientes.

El cuadro 3 presenta los estimadores punto de los PCM para todos los perfiles de proyección en las mismas dos muestras consideradas en el cuadro 2. Al observar cada columna del cuadro 3 se encuentra que el estimador puntual del PCM decrece con el horizonte de proyección para casi

Cuadro 3

Proyecciones cuadrático medias

PCM	Muestra completa (cuadro 1)				Muestra: dic. 2003 - dic. 2009 (73 datos)			
	ESARIMA	EEE	Consensus Forecast	Bloomberg	ESARIMA	EEE	Consensus Forecast	Bloomberg
$h = 1$	18,49	17,85	17,86	21,77	22,49	21,64	21,63	21,77
$h = 2$	18,70	-	17,19	-	22,59	-	20,69	-
$h = 3$	18,91	17,24	16,48	-	22,76	20,00	19,62	-
$h = 4$	19,04	-	15,66	-	22,88	-	18,45	-
$h = 5$	19,12	-	14,71	-	22,94	-	17,17	-
$h = 6$	19,21	-	13,84	-	23,10	-	16,00	-
$h = 12$	18,52	10,00	-	-	22,46	10,30	-	-

Fuente: Elaboración propia.

todos los perfiles considerados (EEE, *Consensus* y *Bloomberg*), tanto para la muestra completa como para la submuestra. La única excepción se detecta para el caso de las proyecciones ESARIMA, que presentan un perfil creciente en los seis primeros horizontes de proyección para luego decaer en el horizonte doce. Este comportamiento creciente en los primeros seis meses podría no ser compatible con un perfil de proyecciones óptimo, de acuerdo con lo mostrado por Patton y Timmermann (2012).

Si ahora observamos cada fila del cuadro 3, también encontramos algunos hechos interesantes. Por ejemplo, para el primer horizonte se observa que el estimador puntual del PCM de la proyección ESARIMA es mayor que el PCM estimado para la EEE y *Consensus* en ambos períodos muestrales considerados. Para este mismo horizonte de proyección, en el período muestral que va de diciembre del 2003 a diciembre del 2009 notamos también que el estimador puntual del PCM de los modelos ESARIMA es mayor que el de *Bloomberg*, y que este estimador para la EEE también es mayor que el de *Consensus*. El análisis del resto de los horizontes de proyección no hace más que confirmar esta tónica. De hecho, más allá del primer horizonte, el PCM siempre muestra un comportamiento decreciente en la medida en que la información para construir los pronósticos es supuestamente mayor.

Cuadro 4

Porcentaje de veces superior

	ESARIMA	EEE	<i>Consensus Forecast</i>	<i>Bloomberg</i>
$h = 1$	21,9	12,3	24,7	46,6
$h = 2$	42,5	-	57,5	-
$h = 3$	19,2	19,2	37,0	-
$h = 4$	42,5	-	57,5	-
$h = 5$	43,8	-	56,2	-
$h = 6$	37,0	-	63,0	-
$h = 12$	19,2	80,8	-	-

Fuente: Elaboración propia.

Nota: La muestra abarca de diciembre del 2003 a diciembre del 2009 (73 datos).

En el cuadro 4 se muestra el porcentaje de veces en que una proyección arroja el menor error. La suma puede ser mayor de 100% toda vez que algunas proyecciones sean iguales. Se observa que ninguna lo hace mejor el 100% de las veces. De hecho, el mayor valor alcanzado es cercano al 80% para el caso de la EEE proyectando a doce meses. En definitiva, no existe un predominio uniforme de un pronóstico sobre otro, sino diferencias en el promedio de sus precisiones predictivas.

Otra característica importante del conjunto de pronósticos es la alta correlación entre ellos, al menos en los horizontes más cortos. En el cuadro 5 se presenta la matriz de correlaciones de los pronósticos para los dos horizontes predictivos en que existen más de dos proyecciones:

Cuadro 5

Matriz de correlaciones

	Proyecciones							
	$h = 1$				$h = 3$			
	ESARIMA	EEE	<i>Consensus Forecast</i>	<i>Bloomberg</i>	ESARIMA	EEE	<i>Consensus Forecast</i>	<i>Bloomberg</i>
ESARIMA	1	-	-	-	1	-	-	-
EEE	0,9963	1	-	-	0,9730	1	-	-
<i>Consensus Forecast</i>	0,9957	0,9995	1	-	0,8444	0,8934	1	-
<i>Bloomberg</i>	0,9942	0,9985	0,9991	1	-	-	-	-

Fuente: Elaboración propia.

$h = 1$ y $h = 3$. Para los casos en que el horizonte predictivo es igual a $h = 2, 4, 5, 6$ y 12 , las correlaciones caen mientras mayor es el horizonte, pasando de $0,94$ ($h = 2$) a $0,01$ ($h = 12$).

Luego de revisar los datos y sus estadísticas descriptivas, en la próxima sección pasamos a describir las propiedades de optimalidad que nos interesan, así como la metodología de evaluación que utilizaremos.

IV. OPTIMALIDAD, METODOLOGÍA E INFERENCIA

Un resultado tradicional en la literatura predictiva es aquel que muestra que el pronóstico óptimo bajo pérdida cuadrática corresponde a la esperanza condicional de la variable a predecir, con respecto al conjunto de información disponible¹⁰. Este resultado lleva a que el error de predicción óptimo cumpla las propiedades de insesgamiento y de ortogonalidad con respecto a cualquier variable presente en el conjunto de información. Una derivación de algunas de estas propiedades se puede encontrar en Elliot y Timmermann (2008), Bentancor y Pincheira (2008, 2010) y en Capistrán y López-Moctezuma (2010).

Más allá de las propiedades de sesgo y de ortogonalidad que caracterizan el error de predicción óptimo, en este artículo estamos interesados en dos propiedades que restringen de una manera particular a un perfil de pronósticos construidos en distintos momentos del tiempo. En las próximas líneas seguiremos de cerca a Patton y Timmermann (2012) para derivar estas propiedades.

1. Un perfil de proyecciones óptimo

Consideremos el siguiente proceso:

$$Y = \{\Upsilon_t\}_{t=0}^{\infty}.$$

que supondremos adaptado a una filtración denominada $\Upsilon = \{\Upsilon_t\}_{t=0}^{\infty}$. Recordemos que, al ser Υ una filtración, se cumple que:

$$\Upsilon_t \subseteq \Upsilon_{t'}, t' > t.$$

Los elementos de la filtración constituyen los conjuntos de información disponibles en cada instante t . Supondremos además, que la siguiente esperanza condicional existe:

$$\mathbb{E}[Y_t | \Upsilon_{t-h}] \text{ para todo } t, h = 0, 1, 2, \dots \text{ tal que } t - h \geq 0$$

Tal como hemos señalado, el pronóstico óptimo bajo pérdida cuadrática corresponde a la esperanza condicional de la variable a predecir con respecto al conjunto de información utilizado en la predicción:

$$y_{t-h}^f(h)^* \equiv \mathbb{E}[Y_t | \Upsilon_{t-h}]$$

¹⁰ Ver, por ejemplo, Elliot y Timmermann (2008).

El error de pronóstico óptimo $e_{t-h}(h)^*$ se define como:

$$e_{t-h}(h)^* \equiv Y_t - y_{t-h}^f(h)^*$$

y su ECM se define consecuentemente como:

$$ECM \equiv \mathbb{E}(e_{t-h}(h)^*)^2$$

A su vez, el PCM queda simplemente definido como:

$$PCM \equiv \mathbb{E}(y_{t-h}^f(h)^*)^2$$

A continuación mostramos las propiedades que debe cumplir un perfil óptimo de proyecciones en términos de ECM y PCM.

Perfil óptimo de errores cuadráticos medios

Si nuestro objetivo es predecir Y_t y nuestro conjunto de información está dado por Υ_{t-h_1} , $h_1 > 0$, $t - h_1 > 0$, entonces se cumple que

$$\mathbb{E}[(Y_t - y_{t-h_1}^f(h_1)^*)^2] \leq \mathbb{E}[(Y_t - u)^2] \text{ para todo } u \in \Upsilon_{t-h_1}$$

en particular, si tomamos

$$\bar{u} = \mathbb{E}[Y_t | \Upsilon_{t-h_2}]; h_1 < h_2 \text{ tal que } 0 \leq t - h_2$$

se tiene

$$\bar{u} \in \Upsilon_{t-h_1}$$

porque

$$0 \leq t - h_2 < t - h_1$$

y por lo tanto

$$\bar{u} = \mathbb{E}[Y_t | \Upsilon_{t-h_2}] \in \Upsilon_{t-h_2} \subseteq \Upsilon_{t-h_1}$$

luego

$$\mathbb{E}[(Y_t - y_{t-h_1}^f(h_1)^*)^2] \leq \mathbb{E}[(Y_t - y_{t-h_2}^f(h_2)^*)^2] \text{ para todo } h_1 < h_2 \text{ tal que } 0 \leq t - h_2$$

lo que dice simplemente que el ECM del pronóstico óptimo decrece débilmente (no crece) en la medida en que el pronóstico se construye con mayor información; esto es, cada vez más cerca del instante t en el que se observa la variable a predecir.

Si suponemos adicionalmente que el proceso de predecir cómo sus pronósticos óptimos sean conjuntamente estacionarios y nos planteamos como objetivo predecir Y_{t+h} con la información en t , entonces se cumple que

$$\mathbb{E}[(Y_{t+h} - \mathbb{E}[Y_{t+h} | \Upsilon_t])^2] \leq \mathbb{E}[(Y_{t+h} - \mathbb{E}[Y_{t+h} | \Upsilon_{t-k}])^2]; k > 0$$

porque

$$\mathbb{E}[Y_{t+h} | \Upsilon_{t-k}] \in \Upsilon_{t-k} \subseteq \Upsilon_t$$

pero nuestro supuesto de estacionariedad implica que

$$\mathbb{E}[(Y_{t+h} - \mathbb{E}[Y_{t+h} | \Upsilon_{t-k}])^2] = \mathbb{E}[(Y_{t+h+k} - \mathbb{E}[Y_{t+h+k} | \Upsilon_t])^2]$$

por lo que se concluye que

$$\mathbb{E}[(Y_{t+h} - y_t^f(h)^*)^2] \leq \mathbb{E}[(Y_{t+h+k} - y_t^f(h+k)^*)^2] \text{ para todo } k > 0$$

lo que dice simplemente que el ECM del pronóstico óptimo crece débilmente (no decrece) con el horizonte de proyección.

Perfil óptimo de pronósticos cuadráticos medios

Tal como se mencionó al inicio de esta sección, el error de predicción óptimo bajo pérdida cuadrática cumple las propiedades de insesgamiento y de ortogonalidad con respecto a cualquier variable presente en el conjunto de información utilizado para predecir. Esto significa que

$$\mathbb{E}[e_{t-h}(h)^*] = 0 \quad (\text{insesgamiento})$$

$$\text{Cov}[y_{t-h}^f(h)^*, e_{t-h}(h)^*] = \mathbb{E}[y_{t-h}^f(h)^* e_{t-h}(h)^*] = 0 \quad (\text{ortogonalidad})$$

entonces

$$\begin{aligned} \mathbb{E}[Y_t]^2 &= \mathbb{E}[y_{t-h}^f(h)^* + e_{t-h}(h)^*]^2 \\ &= \mathbb{E}[y_{t-h}^f(h)^*]^2 + \mathbb{E}[e_{t-h}(h)^*]^2 + 2\mathbb{E}[y_{t-h}^f(h)^* e_{t-h}(h)^*] \\ &= \mathbb{E}[y_{t-h}^f(h)^*]^2 + \mathbb{E}[e_{t-h}(h)^*]^2 \end{aligned}$$

lo que se puede reformular como:

$$\mathbb{E}[y_{t-h}^f(h)^*]^2 = \mathbb{E}[Y_t]^2 - \mathbb{E}[e_{t-h}(h)^*]^2$$

Esta expresión indica que el PCM óptimo presenta un comportamiento opuesto al del ECM, esto es, crece débilmente (no decrece) en la medida en que el pronóstico se construye con mayor información.

Si ahora nuestro objetivo es predecir Y_{t+h} y nuestro conjunto de información está dado por Υ_t podemos realizar un razonamiento análogo al anterior para mostrar que

$$\mathbb{E}[y_t^f(h)^*]^2 = \mathbb{E}[Y_{t+h}]^2 - \mathbb{E}[e_t(h)^*]^2 \quad (1)$$

No obstante, si suponemos adicionalmente que tanto el proceso a predecir como sus pronósticos óptimos son conjuntamente estacionarios, tenemos que

$$\mathbb{E}[Y_{t+h}]^2 = \mathbb{E}[Y_t]^2 \quad \text{para todo } h$$

$$\mathbb{E}[e_t(h_1)^*]^2 \leq \mathbb{E}[e_t(h_2)^*]^2 \quad \text{si } h_1 \leq h_2$$

por lo que podemos reescribir (1) como

$$\mathbb{E}[y_t^f(h)^*]^2 = \mathbb{E}[Y_t]^2 - \mathbb{E}[e_t(h)^*]^2 \quad \text{para todo } h$$

y concluir que el PCM del pronóstico óptimo decrece débilmente (no crece) con el horizonte de proyección.

2. Nuestras hipótesis nulas

En esta sección planteamos formalmente las diversas hipótesis nulas que evaluaremos y que se derivan de las propiedades de optimalidad demostradas en la subsección anterior.

La primera hipótesis nula de interés, que denotamos como H_{01}^h , plantea, para cada horizonte predictivo h , que el ECM del pronóstico óptimo decrece débilmente (no crece) en la medida en que el pronóstico se construye con mayor información. Así, bajo el supuesto de optimalidad, deberíamos esperar para cualquier horizonte predictivo h :

$$H_{01}^h: ECM_{Bloomberg}^h \leq ECM_{Consensus}^h \leq ECM_{EEE}^h \leq ECM_{ESARIMA}^h$$

Nuestra segunda hipótesis nula de interés, que denotamos como H_{02}^h , plantea, para cada horizonte predictivo, que el pronóstico que se construye en un instante más cercano a la publicación de la variable de interés debería presentar un PCM mayor o igual al de aquellos pronósticos que se construyen en instantes más alejados de la publicación del dato de inflación relevante. Así, tenemos que para cualquier horizonte predictivo h deberíamos esperar:

$$H_{02}^h: PCM_{Bloomberg}^h \geq PCM_{Consensus}^h \geq PCM_{EEE}^h \geq PCM_{ESARIMA}^h$$

La tercera hipótesis nula que nos interesa estudiar se refiere a la consistencia temporal de cada perfil de proyecciones en términos del ECM. Esto significa que, tanto para las proyecciones de *Bloomberg* como para las de *Consensus*, así como para las de la *EEE* y las *ESARIMA*, se debería verificar que las proyecciones a horizontes cortos tengan un ECM menor o igual al de las proyecciones a más largo plazo. Así, para cada fuente de proyección $j \in \{Bloomberg, Consensus, EEE, ESARIMA\}$ debería cumplirse la siguiente hipótesis nula H_{03}^j :

$$H_{03}^j: ECM(h=1) \leq ECM(h=2) \leq \dots \leq ECM(h=6) \leq \dots \leq ECM(h=12)$$

en que $h = 1, \dots, 12$ son los horizontes de proyección.

Finalmente, la cuarta hipótesis a evaluar es similar a la anterior, pero ahora se refiere a la consistencia temporal de cada perfil de proyecciones en términos del PCM. Esto significa que las proyecciones a horizontes cortos deberían tener un PCM mayor o igual al de las proyecciones a más largo plazo. Así, para cada fuente de proyección $j \in \{Bloomberg, Consensus, EEE, ESARIMA\}$, debería cumplirse la siguiente hipótesis nula H_{04}^j :

$$H_{04}^j: PCM(h = 1) \leq PCM(h = 2) \leq \dots \leq PCM(h = 6) \leq \dots \leq PCM(h = 12)$$

en que $h = 1, \dots, 12$ son los horizontes de proyección.

Las cuatro hipótesis nulas formuladas anteriormente son similares. En particular, todas ellas corresponden a un conjunto de desigualdades entre momentos de diversas variables aleatorias. En la siguiente subsección mostraremos la estrategia de inferencia que utilizaremos para evaluar todas las hipótesis de nuestro interés.

3. Estrategia de inferencia

Algunas de las hipótesis de nuestro interés corresponden a hipótesis univariadas, que involucran solo una desigualdad entre dos momentos poblacionales. Esto ocurre para las hipótesis $H_{01}^2, H_{02}^2, H_{01}^4, H_{02}^4, H_{01}^5, H_{02}^5, H_{01}^6, H_{02}^6, H_{01}^{12}, H_{02}^{12}$. En otras ocasiones, las hipótesis de nuestro interés son conjuntas. En este contexto no es directo construir un test que controle adecuadamente por el error de tipo I, como lo veremos en las próximas líneas.

Necesidad de un test conjunto

Cuando el interés del investigador se centra en testear una hipótesis conjunta que se compone de múltiples hipótesis univariadas, se puede adoptar básicamente dos estrategias. La primera consiste en desarrollar un test conjunto, y la segunda en construir múltiples tests para cada una de las hipótesis univariadas y luego, mediante alguna regla, obtener una conclusión de aceptación o rechazo de la hipótesis conjunta bajo evaluación. Algunas reglas simples son las siguientes:

1. Rechazar la hipótesis nula conjunta a un nivel de significancia α si cualquiera de las hipótesis univariadas que componen la hipótesis conjunta es rechazada a un nivel de significancia α .
2. Rechazar la hipótesis nula conjunta a un nivel de significancia α si cualquiera de las N hipótesis univariadas que componen la hipótesis conjunta es rechazada a un nivel de significancia α/N . A esta regla se la suele denominar *regla de Bonferroni*.

No obstante la simpleza que ofrecen estas reglas, ambas pueden presentar serias dificultades en el control del error de tipo I. Para analizar un caso particular en el que esto puede ocurrir, pensemos en la siguiente hipótesis nula multivariada:

$$H_0: \vec{\mu} = \mathbf{0}_{N \times N} \quad (2)$$
$$\vec{\mu} = \mathbb{E}X_t$$

En otras palabras, deseamos testear que los N componentes de la esperanza del vector aleatorio X_t sean cero. La hipótesis alternativa plantea que al menos un componente de $\vec{\mu}$ sea mayor que cero:

$$H_A: \exists i \in \{1, \dots, N\} \text{ tal que } \vec{\mu} > 0 \quad (3)$$

Supongamos que disponemos de una colección de tests $T_i, i = 1, \dots, N$ que son una función de la muestra y que están diseñados para testear la siguiente hipótesis nula

$$H_0^{(i)} : \mu_i = 0$$

con respecto a la alternativa

$$H_A^{(i)} : \mu_i > 0$$

de manera que cada test T_i rechaza la nula $H_0^{(i)}$ a un determinado nivel α cuando, evaluados en la muestra disponible de datos D ,

$$T_i(D) > c$$

Evaluemos el tamaño del error de tipo I de la regla 1 bajo el supuesto de independencia del vector

$$\vec{T} = (T_1, T_2, \dots, T_N)'$$

En este caso tenemos que:

$$\Pr(\exists i \in \{1, \dots, N\} \text{ tal que } T_i(D) > c \mid H_0) = \Pr\left(\sum_{i=1}^N W_i > 0 \mid H_0\right)$$

en que

$$W_i = \begin{cases} 1 & \text{si } T_i(D) > c \\ 0 & \text{si no} \end{cases}$$

Como W_i es una variable aleatoria Bernoulli(p) de parámetro $p < \alpha$, tenemos que $\sum_{i=1}^N W_i$ es una binomial de parámetros (N, p) . De esta forma

$$\begin{aligned} \Pr\left(\sum_{i=1}^N W_i > 0 \mid H_0\right) &= 1 - \Pr\left(\sum_{i=1}^N W_i = 0 \mid H_0\right) \\ &= 1 - \Pr(T_i(D) \leq c, \forall i \in \{1, \dots, N\} \mid H_0) \\ &= 1 - (1 - p)^N \rightarrow 1 \text{ cuando } N \rightarrow \infty \end{aligned}$$

Es decir, la regla 1 pierde completamente el control del error de tipo I a medida que la dimensión del problema tiende a infinito, bajo el supuesto de independencia del vector de tests individuales.

En el caso de la regla de Bonferroni, su problema usual es plantear un test a veces demasiado conservador, puesto que se cumple la siguiente desigualdad

$$\Pr(\exists i \in \{1, \dots, N\} \text{ tal que } T_i(D) > \tilde{c} \mid H_0) \leq \sum_{i=1}^N \Pr(T_i(D) > \tilde{c} \mid H_0).$$

En el caso de la regla 2 se escoge $\tilde{c}$ tal que

$$\Pr(T_i(D) > \tilde{c} \mid H_0) = \frac{\alpha}{N}, \forall i \in \{1, \dots, N\}$$

luego, estas cotas garantizan que

$$\Pr(\exists i \in \{1, \dots, N\} \text{ tal que } T_i(D) > \tilde{c} \mid H_0) \leq \sum_{i=1}^N \frac{\alpha}{N} = \alpha$$

pero nada asegura que esta relación no se cumpla con desigualdad estricta y que en realidad el tamaño del test sea bastante inferior al deseado. El cuadro 6, a continuación, muestra el error de tipo I cometido por el test conjunto que desarrollaremos posteriormente y por la aplicación de las reglas 1 y 2 cuando las hipótesis nula y alternativa corresponden a (2) y (3) respectivamente con $N = 2$. Los valores de cada celda corresponden a los tamaños empíricos calculados en 100 mil simulaciones de Monte Carlo en las que los parámetros verdaderos están fijos en cero y corresponden a la esperanza de una normal bivariada con varianza V de diagonal unitaria y donde se prueban varias configuraciones de correlaciones. El tamaño nominal de la prueba conjunta se fija en 0.10, por lo que se espera que el tamaño empírico se aproxime tanto como sea posible a 0.10. Los correspondientes tests individuales son tests t tradicionales.

Se observa que el test conjunto obtiene un error de tipo I perfecto, mientras que las reglas de tipo 1 y 2 tienen errores de tipo I excesivamente altos y bajos, respectivamente, lo que justifica la construcción de una prueba conjunta como la que veremos próximamente¹¹.

Tests univariados

Para el caso de aquellos horizontes predictivos en los que solo se dispone de un par de proyecciones distintas, una forma fácil de proceder consiste en utilizar algún test de igualdad de capacidad predictiva pero a una sola cola. Alternativas naturales de estos tests

Cuadro 6

Error de tipo I con tamaño nominal de 10%

Evaluación de un test conjunto y reglas 1 y 2 cuando $N = 2$			
Correlación	Test conjunto	Regla 1	Regla 2
$r = -0,90$	0,1	0,200	0,099
$r = -0,50$	0,1	0,201	0,098
$r = 0,00$	0,1	0,188	0,098
$r = 0,50$	0,1	0,168	0,091
$r = 0,90$	0,1	0,133	0,069

Fuente: Elaboración propia.

¹¹ En simulaciones, no reportadas aquí simplemente para ahorrar espacio, exploramos las distorsiones de tamaño asociadas a la regla 2 para $N=20$ y una matriz de varianzas con diagonal unitaria y en la que fuera de la diagonal aparece un número $r < 1$. Cuando $r=0$, la regla 2 se comporta bastante bien, presentando un tamaño empírico del 9,7% cuando el tamaño nominal es de 10%. No obstante, cuando la correlación entre las variables crece y consideramos $r=0,5$, el tamaño empírico decae a 5,6%. Con $r=0,95$ el tamaño empírico decae a 1,4%. Finalmente, cuando $r=0,99$, tenemos un tamaño empírico ínfimo de solo 0,76%. La intuición es bien sencilla. La regla 2 rechaza al 10% cuando cualquiera de los estadísticos t individuales rechaza al 10%/N. Cuando la regla 2 rechaza si algún estadístico t rechaza al 5%, pero cuando $N=20$, la regla 2 rechaza si al menos uno de los 20 t -statistics rechaza al 0,5%. Cuando la correlación entre las variables es muy alta, los estadísticos se comportan básicamente en forma similar. Así, si uno de los 20 estadísticos t indica rechazo, probablemente todos los demás también indiquen rechazo y vice versa. Intuitivamente, una correlación alta implica que la aplicación de la regla 2 sea equivalente a considerar un test en el que se evalúe un "t-statistic representativo" al 0,5% cuando $N=20$, o al 5% cuando $N=2$. Es por eso que se transforman las cotas de Bonferroni en tests muy conservadores cuando los estadísticos individuales están correlacionados, y de ahí la necesidad de contar con un buen test conjunto que no sea tan conservador como esta regla 2. Dado que a priori no está claro si las diferencias de ECM o PCM van a tener correlaciones altas o bajas, es importante contar con una prueba estadística no sujeta a este tipo de críticas.

son los atribuidos a Diebold y Mariano (1995) y West (1996), o bien la versión incondicional del test de Giacomini y White (2006), que denotaremos por GW. Formalmente, para estos casos utilizaremos el siguiente estadístico de GW:

$$t_{n(h)} = \frac{\overline{\Delta L_h}}{\sigma_{n(h)} / \sqrt{n(h)}},$$

con:

$$\overline{\Delta L_h} = \frac{1}{n(h)} \sum_{t=1}^{n(h)} \Delta L_{th},$$

donde ΔL_{th} es el diferencial de la función de pérdida entre dos series de pronósticos, $n(h)$ es el número de observaciones (proyecciones) a un horizonte h , y $\hat{\sigma}_{n(h)}$ es un estimador HAC de la desviación estándar de $\overline{\Delta L_h} / \sqrt{n(h)}$. Giacomini y White (2006) muestran que este estadístico tiene una distribución asintótica normal estándar. Cuando estamos interesados en comparaciones del ECM, el diferencial de funciones de pérdida entre la serie de pronósticos A y B luce como sigue:

$$\Delta L_{th} = e_{A,t}^2(h) - e_{B,t}^2(h)$$

Análogamente, cuando nos centramos en comparaciones del PCM, el diferencial de funciones de pérdida entre la serie de pronósticos A y B luce como sigue:

$$\Delta L_{th} = (y_{A,t}^f(h))^2 - (y_{B,t}^f(h))^2$$

Con el test de GW a una sola cola abordaremos la inferencia de las siguientes hipótesis univariadas: H_{01}^2 , H_{02}^2 , H_{01}^4 , H_{02}^4 , H_{01}^5 , H_{02}^5 , H_{01}^6 , H_{02}^6 y H_{01}^{12} , H_{02}^{12} .

Un test conjunto de superioridad predictiva

Los cuadros 2 y 3 nos muestran que para el primer y el tercer horizonte de proyección las hipótesis nulas H_{01}^1 , H_{01}^3 , H_{02}^1 y H_{02}^3 son hipótesis que relacionan tres o más series de pronósticos, lo que conlleva la necesidad de trabajar con un test conjunto. En esta sección intentaremos diseñar un test de superioridad predictiva conjunto que controle adecuadamente por el error de tipo I. Consideremos la siguiente hipótesis nula genérica:

$$H_0^G : \mathbb{E}f(X_{1t}) \leq 0; \mathbb{E}f(X_{2t}) \leq 0; \dots; \mathbb{E}f(X_{lt}) \leq 0$$

contra la hipótesis alternativa genérica

$$H_A^G : \exists j \in \{1, 2, \dots, l\} \text{ tal que } \mathbb{E}f(X_{jt}) > 0$$

en que

$$X_t = (X_{1t}, \dots, X_{lt})'$$

representa a un vector de l variables aleatorias, y $f : \mathbb{R} \rightarrow \mathbb{R}$ es una función continua. Denotaremos por Z_t el vector

$$Z_t = f(X_t)$$

Evidentemente, nuestras hipótesis de interés son casos particulares de H_0^G . Por ejemplo, cuando queremos testear

$$H_{01}^3: ECM_{Consensus}^3 \leq ECM_{EEE}^3 \leq ECM_{ESARIMA}^3$$

basta tomar

$$X_{1t} = e_{Consensus,t}^2(3) - e_{EEE,t}^2(3)$$

$$X_{2t} = e_{EEE,t}^2(3) - e_{ESARIMA,t}^2(3)$$

y $f(x) = x$ para expresar H_{01}^3 en términos de H_0^G . Cuando, en cambio, consideramos

$$H_{02}^3: PCM_{Consensus}^3 \geq PCM_{EEE}^3 \geq PCM_{ESARIMA}^3$$

basta tomar

$$X_{1t} = (Y_{Consensus,t}^f(3))^2 - (Y_{EEE,t}^f(3))^2$$

$$X_{2t} = (Y_{EEE,t}^f(3))^2 - (Y_{ESARIMA,t}^f(3))^2$$

y la misma función f definida anteriormente para expresar H_{02}^3 en términos de H_0^G .

Un test para H_0^G es el desarrollado por White (2000) bajo los supuestos que guían la teoría asintótica fuera de muestra desarrollada por West (1996). Si en vez de considerar esta teoría asintótica, usamos la propuesta por Giacomini y White (2006), podemos desarrollar un test similar al de White (2000). En efecto, tenemos que, cuando el número de predicciones involucradas, P , tiende a infinito, se cumple que

$$\sqrt{P}(\bar{Z} - \mathbb{E}Z_t) \overset{A}{\rightsquigarrow} N(0, \Omega)$$

$$\bar{Z} = \frac{1}{P} \sum_{t=1}^P Z_t; Z_t = f(X_t)$$

en que Ω representa una matriz definida positiva que algunas veces es denominada como “matriz de varianzas-covarianzas de largo plazo”. Como resultado de la aplicación del teorema de la función continua para convergencia en distribución, tenemos que en la medida en que el número de pronósticos P vaya a infinito, entonces

$$\text{Max}_{j \in \{1, 2, \dots, I\}} \left[\sqrt{P} \frac{1}{P} \sum_{t=1}^P [Z_{jt} - \mathbb{E}Z_{jt}] \right]_{I \times 1} \xrightarrow{D} \text{Max}_{k \in \{1, \dots, I\}} \{u_k\}$$

donde $\{u_1, u_2, \dots, u_I\}$ es un vector de dimensión I que sigue una distribución normal multivariada $N(0, \Omega)$. A partir de este resultado es posible construir un test basado en el estadístico del máximo:

$$\text{Max}_{j \in \{1, 2, \dots, I\}} \left[\sqrt{P} \frac{1}{P} \sum_{t=1}^P [Z_{jt} - \mathbb{E}Z_{jt}] \right]_{I \times 1}$$

No obstante, y como señalan Hansen (2005) y Romano y Wolf (2005), es intuitivamente preferible construir un test sobre un vector de variables Z_t estandarizado. Para ello procedemos de la siguiente manera:

Consideremos la siguiente matriz diagonal $\mathbf{D}$, definida como

$$\mathbf{D}_{ii} = \sigma_i^{-1}; i = 1, \dots, I$$

en que

$$\sigma_i^2 = \Omega_{ii}; i = 1, \dots, I.$$

Entonces, se cumple que

$$\sqrt{P} \mathbf{D}(\bar{Z} - \mathbb{E}Z_t) \overset{A}{\rightsquigarrow} N(0, \mathbf{D}\Omega\mathbf{D})$$

con el beneficio de que ahora

$$[\mathbf{D}\Omega\mathbf{D}] = 1; \text{ para todo } i = 1, \dots, I$$

Desafortunadamente, tenemos tres términos desconocidos: $\mathbb{E}Z_t$, $\mathbf{D}$ y Ω . Notemos que la hipótesis nula H_0^G se puede expresar como

$$H_0^G : \mathbb{E}Z_t \leq 0$$

Esto muestra que hay una infinidad de vectores $\mathbb{E}Z_t$ que son coherentes con la hipótesis nula. En particular,

$$\mathbb{E}Z_t = 0 \tag{4}$$

es un valor coherente y es el menos favorable a la alternativa, lo que asegura la construcción de un test con un tamaño acotado por el tamaño nominal. Por esta razón, y en coherencia con la propuesta de White (2000) construimos nuestro test imponiendo (4). También tenemos que $\mathbf{D}$ y Ω son desconocidos, pero mediante el método de Newey y West (1987, 1994) por ejemplo, es posible construir un estimador consistente y semidefinido positivo para W que denotaremos $\hat{\Omega}$.

Luego, podemos definir un estimador consistente de $\mathbf{D}$ como sigue¹²:

$$\hat{\mathbf{D}} = [\hat{\Omega}_{ii}]^{-1/2}; i = 1, \dots, I$$

¹² Esto es posible siempre que los elementos de la diagonal sean no nulos.

Así se cumple que bajo la hipótesis nula

$$\sqrt{P} \hat{\mathbf{D}}(\bar{\mathbf{Z}}) \overset{A}{\rightsquigarrow} N(0, \tilde{\Omega})$$

donde

$$\tilde{\Omega} \equiv \mathbf{D} \Omega \mathbf{D}$$

Con esto, nuestro estadístico de interés se puede escribir como

$$\text{Max}_{j \in \{1, 2, \dots, l\}} \sqrt{P} \hat{\mathbf{D}} \bar{\mathbf{Z}}$$

en que $j \in \{1, 2, \dots, l\}$ representan los distintos componentes del siguiente vector

$$\hat{\mathbf{D}} \bar{\mathbf{Z}}.$$

Los valores críticos de nuestro estadístico pueden ser calculados a través de simulaciones de Monte Carlo, como sugiere White (2000), de acuerdo con los siguientes pasos

1. Calcular la siguiente descomposición de Cholesky sobre $\hat{\mathbf{D}} \hat{\Omega} \hat{\mathbf{D}} = \mathbf{G}' \mathbf{G}$
2. Definir un número de replicaciones N , que en el presente artículo corresponderá a $N = 1,000,000$.
3. Para cada una del millón de replicaciones se obtiene una realización independiente v de una normal multivariada $N(0, I_{l \times l})$
4. Se define w como $w = \mathbf{G}' v$, luego w es una realización independiente de $N(0, \hat{\mathbf{D}} \hat{\Omega} \hat{\mathbf{D}})$
5. Se define m como $m = \max_{j \in \{1, \dots, l\}} \{w_j\}$
6. Luego se ordena el millón de términos m obtenidos y se definen los valores críticos como los correspondientes cuantiles.

Es importante recalcar que bajo la hipótesis nula genérica,

$$\text{Max}_{j \in \{1, 2, \dots, l\}} \mathbb{E} Z_t \leq 0$$

mientras que, bajo la hipótesis alternativa genérica, se tendrá que

$$\text{Max}_{j \in \{1, 2, \dots, l\}} \mathbb{E} Z_t > 0,$$

lo que implica que nos concentraremos en un test de una sola cola, cuya región de rechazo estará definida por una relación del tipo

$$\text{Max}_{j \in \{1, 2, \dots, l\}} \sqrt{P} \hat{\mathbf{D}} \bar{\mathbf{Z}} > c.$$

Naturalmente, también es posible construir de manera análoga un test para el estadístico del mínimo, lo cual es muy útil para el caso en el que la hipótesis nula se exprese como

$$H_0 : \mathbb{E} Z_t \geq 0$$

V. ANÁLISIS EMPÍRICO

En esta sección veremos qué nos dicen estos tests univariados y conjuntos basados en los estadísticos del máximo y del mínimo, cuando los aplicamos a las hipótesis de nuestro interés. Comenzaremos con los tests univariados para luego pasar a los tests conjuntos.

1. Tests univariados

En los cuadros 7 y 8 se muestran los resultados del test de GW a una cola para los ECM y PCM, respectivamente. Estos cuadros contienen en cada celda el numerador del estadístico t de GW y su denominador entre paréntesis redondos. A partir de estos cuadros es posible deducir el valor de los tests para todas las posibles comparaciones bilaterales entre proyecciones. El cuadro 7 nos indica que las hipótesis nulas H_{01}^2 , H_{01}^4 , H_{01}^5 , H_{01}^6 y H_{01}^{12} no pueden ser rechazadas a ningún nivel de significancia usual (10%, 5% o 1%). De hecho, este cuadro nos indica que cualquier hipótesis nula bilateral que plantee superioridad en términos del ECM por parte de los pronósticos construidos en un instante más cercano a la publicación de la variable de interés, no podría ser rechazada a los niveles de significancia usuales. Aún más, todos los signos de los estadísticos apuntan en la dirección de establecer un perfil de estimaciones puntuales del ECM coherentes con un perfil de proyecciones óptimas a cualquier horizonte predictivo.

El cuadro 8 nos muestra un panorama algo distinto. Si bien este cuadro indica que las hipótesis nulas H_{02}^2 , H_{02}^4 , H_{02}^5 , H_{02}^6 tampoco pueden ser rechazadas a ningún nivel de significancia usual (10%, 5%, o 1%), la hipótesis H_{02}^{12} efectivamente es rechazada a un nivel de significancia del 5%. Más aún, este cuadro nos indica que en varias ocasiones una hipótesis nula bilateral, que plantee que los pronósticos que se construyen en un instante más cercano a la publicación

Cuadro 7

Test de GW para ECM

Muestra completa (cuadro 1)

	Bloomberg y Consensus Forecasts	Bloomberg y EEE	Bloomberg y ESARIMA	Consensus Forecasts y EEE	Consensus Forecasts y ESARIMA	EEE y ESARIMA
$h = 1$	-0,038 (0,010)	-0,051 (0,010)	-0,143 (0,044)	-0,013 (0,006)	-0,104 (0,039)	-0,091 (0,037)
$h = 2$	-	-	-	-	-0,269 (0,310)	-
$h = 3$	-	-	-	-0,303 (0,328)	-0,786 (0,488)	-0,482 (0,416)
$h = 4$	-	-	-	-	-1,519 (0,776)	-
$h = 5$	-	-	-	-	-2,465 (1,145)	-
$h = 6$	-	-	-	-	-3,602 (1,694)	-
$h = 12$	-	-	-	-	-	-11,759 (6,684)

Fuente: Elaboración propia.

Nota: Entre paréntesis, se presenta la desviación estándar.

Cuadro 8
Test de GW para PCM

Muestra completa (cuadro 1)

	Bloomberg y Consensus Forecasts	Bloomberg y EEE	Bloomberg y ESARIMA	Consensus Forecasts y EEE	Consensus Forecasts y ESARIMA	EEE y ESARIMA
$h = 1$	0,135 (0,222)	0,128 (0,292)	-0,716* (0,539)	-0,006 (0,135)	-0,851** (0,451)	-0,845** (0,378)
$h = 2$	- -	- -	- -	- -	-1,89 (1,576)	- -
$h = 3$	- -	- -	- -	-0,382 (2,101)	-3,131 (2,79)	-2,748** (1,214)
$h = 4$	- -	- -	- -	- -	-4,431 (3,966)	- -
$h = 5$	- -	- -	- -	- -	-5,775 (5,044)	- -
$h = 6$	- -	- -	- -	- -	-7,103 (5,859)	- -
$h = 12$	- -	- -	- -	- -	- -	-12,161** (6,666)

Fuente: Elaboración propia.

Nota: Entre paréntesis, se presenta la desviación estándar. (*) Significativo al 10%, (**) Significativo al 5%.

de la variable de interés debería presentar un PCM mayor o igual al de aquellos pronósticos que se construyen en instantes más alejados de la publicación oficial, sería rechazada a los niveles de significancia usuales. De hecho, varios signos de los estadísticos apuntan en la dirección de establecer un perfil de estimaciones puntuales del PCM que no son coherentes con un perfil de proyecciones óptimas bajo pérdida cuadrática.

Hasta ahora hemos dado una respuesta satisfactoria a la evaluación de propiedades de optimalidad para el perfil de proyecciones en aquellos horizontes en los que solo tenemos dos proyecciones. En la siguiente subsección analizaremos lo que ocurre en aquellos horizontes en los que hay más de un par de proyecciones disponibles.

2. Evaluaciones conjuntas: una primera aproximación

En esta sección deseamos evaluar las siguientes hipótesis nulas utilizando las reglas 1 y 2 presentadas en la sección IV.3:

$$H_{01}^1 : ECM_{Bloomberg}^1 \leq ECM_{Consensus}^1 ; ECM_{Consensus}^1 \leq ECM_{EEE}^1 \text{ y } ECM_{EEE}^1 \leq ECM_{ESARIMA}^1$$

$$H_{01}^3 : ECM_{Consensus}^3 \leq ECM_{EEE}^3 \text{ y } ECM_{EEE}^3 \leq ECM_{ESARIMA}^3$$

$$H_{02}^1 : PCM_{Bloomberg}^1 \geq PCM_{Consensus}^1 ; PCM_{Consensus}^1 \geq PCM_{EEE}^1 \text{ y } PCM_{EEE}^1 \geq PCM_{ESARIMA}^1$$

$$H_{02}^3 : PCM_{Consensus}^3 \geq PCM_{EEE}^3 \text{ y } PCM_{EEE}^3 \geq PCM_{ESARIMA}^3$$

Definamos un nivel de significancia del 10% y evaluemos estas cuatro hipótesis de acuerdo con las reglas 1 y 2. Observamos que las hipótesis H_{01}^1 y H_{01}^3 no son rechazadas por ninguna de las dos reglas, puesto que ninguna de las relaciones bilaterales es rechazada ni siquiera al 10%, como puede apreciarse en el cuadro 7. Encontramos un escenario opuesto cuando evaluamos las hipótesis H_{02}^1 y H_{02}^3 . No es difícil calcular los valores p del test de GW a una sola cola para evaluar las siguientes dos hipótesis bilaterales:

$$\begin{aligned} PCM_{EEE}^1 - PCM_{ESARIMA}^1 &\geq 0 \\ PCM_{EEE}^3 - PCM_{ESARIMA}^3 &\geq 0 \end{aligned}$$

Estas hipótesis bilaterales se rechazan a un nivel de significancia del 5%. De hecho, los respectivos valores p corresponden a 0,013 y 0,012 respectivamente. Como consecuencia, tanto la regla 1 como la 2 rechazan las hipótesis nulas H_{02}^1 y H_{02}^3 a un nivel de significancia del 10%, e incluso del 5% también.

Los cuadros 9 y 10 nos muestran los resultados de los test de GW, pero esta vez practicado sobre cada conjunto de proyecciones. Así evaluamos si el ECM de las proyecciones ESARIMA, por ejemplo, presenta un comportamiento creciente con el horizonte de proyección. El cuadro 9 en particular se centra en evaluar el perfil creciente de los ECM para las proyecciones ESARIMA, de la EEE y de *Consensus*. El cuadro 10 evalúa el supuesto perfil decreciente de los PCM para los respectivos conjuntos de proyecciones.

Nuestro objetivo es evaluar, de acuerdo con las reglas 1 y 2, las siguientes hipótesis:

$$\begin{aligned} H_{03}^{Consensus} : & ECM_{Consensus}^{h=1} \leq ECM_{Consensus}^{h=2} \leq \dots \leq ECM_{Consensus}^{h=6} \\ H_{03}^{EEE} : & ECM_{EEE}^{h=1} \leq ECM_{EEE}^{h=3} \leq ECM_{EEE}^{h=12} \\ H_{03}^{ESARIMA} : & ECM_{ESARIMA}^{h=1} \leq ECM_{ESARIMA}^{h=2} \leq \dots \leq ECM_{ESARIMA}^{h=6} \leq ECM_{ESARIMA}^{h=12} \\ \\ H_{04}^{Consensus} : & PCM_{Consensus}^{h=1} \geq PCM_{Consensus}^{h=2} \geq \dots \geq PCM_{Consensus}^{h=6} \\ H_{04}^{EEE} : & PCM_{EEE}^{h=1} \geq PCM_{EEE}^{h=3} \geq PCM_{EEE}^{h=12} \\ H_{04}^{ESARIMA} : & PCM_{ESARIMA}^{h=1} \geq PCM_{ESARIMA}^{h=2} \geq \dots \geq PCM_{ESARIMA}^{h=6} \geq PCM_{ESARIMA}^{h=12} \end{aligned}$$

en que $h = 1, \dots, 12$ denotan los horizontes de proyección.

Al igual que más arriba, aplicaremos las reglas 1 y 2 usando los test de GW bilaterales entre horizontes adyacentes. Así, por ejemplo, rechazaremos la hipótesis nula $H_{03}^{Consensus}$ con un nivel de significancia del 10% si al menos uno de los tests de GW para las siguientes hipótesis

$$\begin{aligned} ECM_{Consensus}^{h=1} - ECM_{Consensus}^{h=2} &\leq 0 \\ ECM_{Consensus}^{h=2} - ECM_{Consensus}^{h=3} &\leq 0 \\ ECM_{Consensus}^{h=3} - ECM_{Consensus}^{h=4} &\leq 0 \\ ECM_{Consensus}^{h=4} - ECM_{Consensus}^{h=5} &\leq 0 \\ ECM_{Consensus}^{h=5} - ECM_{Consensus}^{h=6} &\leq 0 \end{aligned}$$

se rechaza a un nivel del 10% usando la regla 1, o bien a un nivel del 2% aplicando la regla 2.

Cuadro 9
**Resultados del test de GW
entre horizontes para ECM**

Muestra: dic. 2003 - dic. 2009 (73 datos)

	ESARIMA	EEE	Consensus Forecasts
$h = 1$ menos $h = 2$	-0,574 (0,181)	-0,986 (0,275)	-0,408 (0,203)
$h = 2$ menos $h = 3$	-0,804 (0,305)	- -	-0,287 (0,195)
$h = 3$ menos $h = 4$	-1,146 (0,447)	-4,381 (1,476)	-0,413 (0,195)
$h = 4$ menos $h = 5$	-1,483 (0,621)	- -	-0,536 (0,207)
$h = 5$ menos $h = 6$	-1,716 (0,758)	- -	-0,579 (0,234)
$h = 6$ menos $h = 12$	-11,311 (6,138)	- -	- -

Fuente: Elaboración propia.

Nota: Para los casos de EEE, las comparaciones son: 1 - 3 y 6 - 12. Entre paréntesis, se presenta la desviación estándar.

Cuadro 10
**Resultados del test de GW
entre horizontes para PCM**

Muestra: dic. 2003 - dic. 2009 (73 datos)

	ESARIMA	EEE	Consensus Forecasts
$h = 1$ menos $h = 2$	-0,099 (1,017)	1,636 (1,923)	0,94 (0,627)
$h = 2$ menos $h = 3$	-0,167 (1,114)	- -	1,072 (0,59)
$h = 3$ menos $h = 4$	-0,13 (1,179)	9,708 (4,398)	1,169 (0,559)
$h = 4$ menos $h = 5$	-0,06 (1,248)	- -	1,283 (0,582)
$h = 5$ menos $h = 6$	-0,156 (1,257)	- -	1,171 (0,636)
$h = 6$ menos $h = 12$	0,643 (6,489)	- -	- -

Fuente: Elaboración propia.

Nota: Para los casos de EEE las comparaciones son: 1 - 3 y 6 - 12. Entre paréntesis, se presenta la desviación estándar.

Una rápida inspección al cuadro 9 nos indica que no existe ningún rechazo bilateral. De hecho, todos los estadísticos de GW son negativos, lo que implica que no hay evidencia para rechazar la hipótesis nula de un perfil creciente del ECM para todos los conjuntos de proyecciones que poseen pronóstico a más de un horizonte.

Similar resultado se obtiene cuando se analiza el cuadro 10. Al no haber siquiera un rechazo al 10% usando el test de GW, no se puede rechazar la hipótesis nula de un perfil de PCM decreciente con el horizonte de predicción para ninguno de los conjuntos de pronósticos considerados: ESARIMA, EEE y *Consensus*.

En la siguiente sección veremos si todos estos resultados que hemos obtenido aplicando las cuestionables reglas 1 y 2 sobreviven a la aplicación de un test conjunto más apropiado.

3. Aplicación del test de superioridad de habilidad predictiva

Errores cuadráticos: fuente variable y evento fijo

Bajo el criterio del ECM se desea probar el cumplimiento de los siguientes perfiles:

$$H_{01}^1 : ECM_{Bloomberg}^1 \leq ECM_{Consensus}^1 \leq ECM_{EEE}^1 \leq ECM_{ESARIMA}^1$$

$$H_{01}^3 : ECM_{Consensus}^3 \leq ECM_{EEE}^3 \leq ECM_{ESARIMA}^3$$

lo que implica el cumplimiento simultáneo de seis desigualdades para el caso en que $h = 1$ y de tres desigualdades para el caso $h = 3$. Los resultados se presentan en el cuadro 11, desde donde se aprecia que no es posible rechazar la hipótesis nula, ya que los estadísticos calculados se encuentran en el extremo izquierdo de la distribución, corroborando los resultados obtenidos usando las reglas 1 y 2.

Cuadro 11

Test de superioridad de habilidad predictiva

Errores cuadrático medios: Fuente variable y horizonte fijo

Estadístico efectivo y distintos percentiles de su distribución bajo la hipótesis nula							
$h = 1$				$h = 3$			
Percentil		Percentil		Percentil		Percentil	
Calculado:	-1,953	50%	0,695	Calculado:	-0,817	50%	0,603
1%	-1,184	60%	0,909	1%	-1,141	60%	0,817
5%	-0,646	70%	1,141	5%	-0,652	70%	1,049
10%	-0,355	80%	1,415	10%	-0,388	80%	1,326
20%	0,000	90%	1,803	20%	-0,063	90%	1,716
30%	0,259	95%	2,122	30%	0,177	95%	2,045
40%	0,483	99%	2,735	40%	0,394	99%	2,661

Fuente: Elaboración propia.

*Proyecciones cuadráticas: fuente variable y evento fijo*Bajo el criterio del PCM se desea probar el cumplimiento de los siguientes perfiles:¹³

$$H_{02}^1 : PCM_{Bloomberg}^1 \geq PCM_{Consensus}^1 \geq PCM_{EEE}^1 \geq PCM_{ESARIMA}^1$$

$$H_{02}^3 : PCM_{Consensus}^3 \geq PCM_{EEE}^3 \geq PCM_{ESARIMA}^3$$

Entonces, para el caso en que $h = 1$ deben satisfacerse seis condiciones para el cumplimiento conjunto de la desigualdad anterior. Para el caso en que $h = 3$, la dimensión del problema se reduce a tres desigualdades.

La forma de estimar la fdp es similar a la del caso anterior, con la única diferencia de que el estadístico se construye sobre el mínimo de los componentes y ya no sobre el máximo. Con esto, bajo la nula se espera que el mínimo de las diferencias del PCM sea positivo. Bajo la alternativa, en cambio, el mínimo debiera ser negativo, lo que implica que nos concentraremos en un test de una sola cola, cuya región de rechazo estará definida por una relación del tipo

$$\min_{j \in \{1, 2, \dots, I\}} \sqrt{P} \hat{D} \bar{Z} > c.$$

Los resultados de la aplicación de este test se encuentran en el cuadro 12. Estos indican que al 5% de significancia se rechaza la hipótesis nula de optimalidad desde la perspectiva de los PCM para los dos horizontes analizados, ya que los estadísticos calculados son menores que el cuantil correspondiente al 5% de probabilidad acumulada sobre la distribución bajo la hipótesis nula, confirmando lo encontrado en las secciones anteriores con las reglas 1 y 2. En consecuencia, el supuesto de optimalidad bajo pérdida cuadrática es rechazado para el perfil de proyecciones.

¹³ Como ya se ha mencionado en el artículo, la propiedad de optimalidad que plantean las hipótesis nulas y que no requiere el supuesto de estacionariedad. Consecuentemente, el rechazo a la condición de eficiencia de los pronósticos no depende de este tipo de supuestos.

Cuadro 12
Test de superioridad de habilidad predictiva

Proyecciones cuadráticas: Fuente variable y horizonte fijo

Estadístico efectivo y distintos percentiles de su distribución bajo la hipótesis nula							
$h = 1$				$h = 3$			
Percentil		Percentil		Percentil		Percentil	
Calculado:	-2,337	50%	-0,436	Calculado:	-2,255	50%	-0,716
1%	-2,608	60%	-0,206	1%	-2,784	60%	-0,504
5%	-1,963	70%	0,033	5%	-2,162	70%	-0,281
10%	-1,621	80%	0,313	10%	-1,835	80%	-0,025
20%	-1,211	90%	0,703	20%	-1,446	90%	-0,323
30%	-0,915	95%	1,020	30%	-1,167	95%	0,609
40%	-0,665	99%	1,611	40%	-0,932	99%	1,142

Fuente: Elaboración propia.

Errores cuadráticos: fuente fija, evento y horizonte variable

Bajo el criterio del ECM, se desea probar el cumplimiento de los siguientes perfiles:

$$\begin{aligned}
H_{03}^{\text{Consensus}} : ECM_{\text{Consensus}}^{h=1} &\leq ECM_{\text{Consensus}}^{h=2} \leq \dots \leq ECM_{\text{Consensus}}^{h=6} \\
H_{03}^{\text{EEE}} : ECM_{\text{EEE}}^{h=1} &\leq ECM_{\text{EEE}}^{h=3} \leq ECM_{\text{EEE}}^{h=12} \\
H_{03}^{\text{ESARIMA}} : ECM_{\text{ESARIMA}}^{h=1} &\leq ECM_{\text{ESARIMA}}^{h=2} \leq \dots \leq ECM_{\text{ESARIMA}}^{h=6} \leq ECM_{\text{ESARIMA}}^{h=12}
\end{aligned}$$

Es decir, se desea evaluar si cada fuente de información entrega una estructura temporal de pronósticos a distintos horizontes compatible con optimalidad bajo pérdida cuadrática. Los distintos percentiles correspondientes a la distribución del estadístico bajo la hipótesis nula, así como el cálculo del estadístico propiamente tal, se presentan en los cuadros 13, 14 y 15. Los estadísticos son todos negativos y se encuentran muy alejados hacia la izquierda en la distribución, indicando que no es posible rechazar la hipótesis nula de optimalidad del perfil del ECM.

Pronósticos cuadráticos: fuente fija, evento y horizonte variable

Bajo el criterio del PCM, se desea probar el cumplimiento de las siguientes hipótesis conjuntas:

$$\begin{aligned}
H_{04}^{\text{Consensus}} : PCM_{\text{Consensus}}^{h=1} &\geq PCM_{\text{Consensus}}^{h=2} \geq \dots \geq PCM_{\text{Consensus}}^{h=6} \\
H_{04}^{\text{EEE}} : PCM_{\text{EEE}}^{h=1} &\geq PCM_{\text{EEE}}^{h=3} \geq PCM_{\text{EEE}}^{h=12} \\
H_{04}^{\text{ESARIMA}} : PCM_{\text{ESARIMA}}^{h=1} &\geq PCM_{\text{ESARIMA}}^{h=2} \geq \dots \geq PCM_{\text{ESARIMA}}^{h=6} \geq PCM_{\text{ESARIMA}}^{h=12}
\end{aligned}$$

en que $h = 1, \dots, 12$ denotan los horizontes de proyección. Es decir, se desea evaluar si cada fuente de información entrega una estructura temporal de pronósticos a distintos horizontes compatible con optimalidad, pero en términos de un ordenamiento sobre el pronóstico cuadrático medio. Observamos tanto en el cuadro 16 como en el 17 y el 18

Cuadro 13**Test de superioridad de habilidad predictiva para *Consensus Forecasts***

Errores cuadrático medios: Fuente fija y evento y horizonte variables

Estadístico efectivo y distintos percentiles de su distribución bajo la hipótesis nula			
	Percentil		Percentil
Calculado:	-1,453	50%	0,842
1%	-0,671	60%	1,035
5%	-0,268	70%	1,246
10%	-0,042	80%	1,503
20%	0,245	90%	1,869
30%	0,463	95%	2,181
40%	0,655	99%	2,786

Fuente: Elaboración propia.

Cuadro 14**Test de superioridad de habilidad predictiva para la EEE**

Errores cuadrático medios: Fuente fija y evento y horizonte variables

Estadístico efectivo y distintos percentiles de su distribución bajo la hipótesis nula			
	Percentil		Percentil
Calculado:	-3,001	50%	0,482
1%	-1,478	60%	0,702
5%	-0,914	70%	0,940
10%	-0,611	80%	1,222
20%	-0,239	90%	1,618
30%	0,031	95%	1,946
40%	0,264	99%	2,574

Fuente: Elaboración propia.

Cuadro 15**Test de superioridad de habilidad predictiva para modelos de series de tiempo**

Errores cuadrático medios: Fuente fija y evento y horizonte variables

Estadístico efectivo y distintos percentiles de su distribución bajo la hipótesis nula			
	Percentil		Percentil
Calculado:	-1,862	50%	0,802
1%	-1,066	60%	1,015
5%	-0,529	70%	1,243
10%	-0,242	80%	1,514
20%	0,111	90%	1,896
30%	0,368	95%	2,212
40%	0,592	99%	2,818

Fuente: Elaboración propia.

**Cuadro 16****Test de superioridad de habilidad predictiva para *Consensus Forecasts***

Pronósticos cuadráticos: Fuente fija y evento y horizonte variables

Estadístico efectivo y distintos percentiles de su distribución bajo la hipótesis nula			
Cuadráticas <i>Consensus Forecasts</i>			
Percentil		Percentil	
Calculado	1,505	50%	-0,605
1%	-2,741	60%	-0,377
5%	-2,104	70%	-0,135
10%	-1,770	80%	0,146
20%	-1,366	90%	0,537
30%	-1,078	95%	0,857
40%	-0,832	99%	1,443

Fuente: Elaboración propia.

Cuadro 17**Test de superioridad de habilidad predictiva para la EEE**

Pronósticos cuadráticos: Fuente fija y evento y horizonte variables

Estadístico efectivo y distintos percentiles de su distribución bajo la hipótesis nula			
Proyecciones Cuadráticas EEE			
Percentil		Percentil	
Calculado:	0,902	50%	-0,543
1%	-2,568	60%	-0,336
5%	-1,953	70%	-0,116
10%	-1,632	80%	0,136
20%	-1,251	90%	0,484
30%	-0,981	95%	0,767
40%	-0,753	99%	1,286

Fuente: Elaboración propia.

Cuadro 18**Test de superioridad de habilidad predictiva para modelos de series de tiempo**

Pronósticos cuadráticos: Fuente fija y evento y horizonte variables

Estadístico efectivo y distintos percentiles de su distribución bajo la hipótesis nula			
Cuadráticas ESARIMA			
Percentil		Percentil	
Calculado:	-0,129	50%	-0,868
1%	-2,838	60%	-0,668
5%	-2,235	70%	-0,456
10%	-1,922	80%	-0,211
20%	-1,552	90%	0,119
30%	-1,291	95%	0,389
40%	-1,070	99%	0,884

Fuente: Elaboración propia.

que el estadístico calculado cae completamente fuera de la región de rechazo para las tres fuentes de pronósticos consideradas. Para *Consensus* el estadístico calculado entrega un valor de 1,505, mientras que la región de rechazo está dada por valores menores que -1,770 al 10% de significancia. Para la EEE y las proyecciones ESARIMA, los estadísticos calculados son 0,902 y -0,129, respectivamente, en tanto las zonas de rechazo están dadas por valores inferiores a -1,632 y -1,922, respectivamente.

Estos resultados confirman los obtenidos en secciones anteriores con la aplicación de las reglas 1 y 2 y no pueden rechazar una estructura temporal, de cada una de las fuentes de proyección, que es coherente con optimalidad bajo pérdida cuadrática.

4. Implicancias de los resultados

La ineficiencia encontrada en este artículo plantea que, si tenemos dos pronósticos distintos e insesgados $y_{t-h}^{(1)}(h)$ y $y_{t-h}^{(2)}(h)$, entonces

$$Y_t = y_{t-h}^{(1)}(h) + e_{t-h}^{(1)}(h) = y_{t-h}^{(2)}(h) + e_{t-h}^{(2)}(h)$$

Luego

$$\mathbb{E}(Y_t^2) = \mathbb{E}(y_{t-h}^{(1)}(h))^2 + \mathbb{E}(e_{t-h}^{(1)}(h))^2 + 2\mathbb{E}(y_{t-h}^{(1)}(h)e_{t-h}^{(1)}(h))$$

$$\mathbb{E}(Y_t^2) = \mathbb{E}(y_{t-h}^{(2)}(h))^2 + \mathbb{E}(e_{t-h}^{(2)}(h))^2 + 2\mathbb{E}(y_{t-h}^{(2)}(h)e_{t-h}^{(2)}(h))$$

Si se cumple simultáneamente

$$\mathbb{E}(e_{t-h}^{(1)}(h))^2 < \mathbb{E}(e_{t-h}^{(2)}(h))^2 \text{ y } \mathbb{E}(y_{t-h}^{(1)}(h))^2 \leq \mathbb{E}(y_{t-h}^{(2)}(h))^2$$

entonces debe tenerse que

$$\mathbb{E}(y_{t-h}^{(1)}(h))^2 + \mathbb{E}(e_{t-h}^{(1)}(h))^2 < \mathbb{E}(y_{t-h}^{(2)}(h))^2 + \mathbb{E}(e_{t-h}^{(2)}(h))^2$$

o equivalentemente

$$\mathbb{E}(Y_t^2) - 2\mathbb{E}(y_{t-h}^{(1)}(h)e_{t-h}^{(1)}(h)) < \mathbb{E}(Y_t^2) - 2\mathbb{E}(y_{t-h}^{(2)}(h)e_{t-h}^{(2)}(h))$$

que es lo mismo que

$$\mathbb{E}(y_{t-h}^{(1)}(h)e_{t-h}^{(1)}(h)) - \mathbb{E}(y_{t-h}^{(2)}(h)e_{t-h}^{(2)}(h)) > 0$$

lo que implica que al menos uno de los términos de la izquierda en la última expresión es no nulo, lo que claramente viola una condición de ortogonalidad implicada por la optimalidad bajo pérdida cuadrática.

La exploración de los cuadros 2 y 3 nos indica que una situación como la descrita anteriormente, es decir

$$\mathbb{E}(e_{t-h}^{(1)}(h))^2 < \mathbb{E}(e_{t-h}^{(2)}(h))^2 \text{ y } \mathbb{E}(y_{t-h}^{(1)}(h))^2 \leq \mathbb{E}(y_{t-h}^{(2)}(h))^2$$

sucede para los siguientes casos:

1. Comparaciones un mes hacia adelante entre *Bloomberg* y ESARIMA, *Consensus* y ESARIMA y *EEE* y ESARIMA.
2. Comparaciones tres meses adelante entre *EEE* y ESARIMA.
3. Comparaciones 12 meses adelante entre *EEE* y ESARIMA.

Los cuadros 19 y 20 nos muestran los resultados de estas cinco comparaciones y de los ocho términos individuales que forman parte de estas cinco comparaciones.

En estos cuadros puede apreciarse que efectivamente la hipótesis

$$H_o : \mathbb{E}(y_{t-h}^{(1)}(h)e_{t-h}^{(1)}(h)) - \mathbb{E}(y_{t-h}^{(2)}(h)e_{t-h}^{(2)}(h)) = 0$$

es rechazada en todos los casos destacados en los puntos 1-3 en favor de la alternativa

$$H_A : \mathbb{E}(y_{t-h}^{(1)}(h)e_{t-h}^{(1)}(h)) - \mathbb{E}(y_{t-h}^{(2)}(h)e_{t-h}^{(2)}(h)) > 0 \quad (5)$$

Cuadro 19

Resultados del test de eficiencia débil. Muestra restringida

	Pruebas diferenciales			Muestra: dic. 2003 - dic. 2009 (73 datos)			
	EEE y ESARIMA	CF y ESARIMA	Bloomberg y ESARIMA	ESARIMA	EEE	CF	Bloomberg
$h = 1$	0,4683 *** (0,1974)	0,4779 ** (0,2368)	0,4298 * (0,2874)	-0,2157 (0,4017)	0,2525 (0,3253)	0,2622 (0,3137)	0,2141 (0,2171)
$h = 3$	1,6157 ** (0,6822)			-1,0384 (1,532)	0,5774 (1,2401)		-
$h = 12$	11,9606 * (6,3629)		-	-8,7199 (7,2220)	3,2407 * (1,9069)		-

Fuente: Elaboración propia.

Nota: Entre paréntesis, se presenta la desviación estándar. * Significativo al 10%. ** Significativo al 5%; *** Significativo al 1%.

Cuadro 20

Resultados del test de eficiencia débil. Muestra completa

	Pruebas diferenciales			Muestra completa			
	EEE y ESARIMA	CF y ESARIMA	Bloomberg y ESARIMA	ESARIMA	EEE	CF	Bloomberg
$h = 1$	0,4683 *** (0,1473)	0,4779 ** (0,1770)	0,4298 ** (0,1549)	-0,2372 (0,2893)	0,1224 (0,2361)	0,1309 (0,2337)	0,1617 (0,1500)
$h = 3$	1,6157 ** (0,5291)			-0,9822 (1,1086)	0,2021 (0,9443)		-
$h = 12$	11,9606 ** (5,0953)		-	-6,2391 (5,3484)	2,1608 * (1,5897)		-

Fuente: Elaboración propia.

Nota: Entre paréntesis, se presenta la desviación estándar. * Significativo al 10%. ** Significativo al 5%; *** Significativo al 1%.

De esta forma, los paneles del lado izquierdo en los cuadros 19 y 20 indican que hay información no eficientemente utilizada por alguno de los pronósticos considerados en la evaluación. Curiosamente, los paneles del lado derecho de los mismos cuadros solo dan cuenta de una ineficiencia estadísticamente significativa para la EEE en proyecciones a 12 meses. No obstante, también se debe apreciar que, al usar niveles de significancia algo más holgados, los mismos paneles del lado derecho del cuadro 20 muestran un rechazo de la hipótesis nula de eficiencia para las proyecciones ESARIMA a 12 meses a un nivel de significancia del 13% en tests de una sola cola. Similarmente, el cuadro 20 también muestra un rechazo de la eficiencia de las proyecciones de *Bloomberg* a una sola cola al 15%. Adicionalmente, y en un resultado que no incluimos simplemente por abreviar espacio, es posible mostrar que la eficiencia de las proyecciones de *Consensus* a horizontes de tres, cuatro, cinco y seis meses es rechazada en la muestra completa con un nivel de significancia del 5%.

La existencia de ineficiencias en los pronósticos plantea el desafío de realizar ajustes que permitan aminorarlas y disminuir el ECM de proyección. En particular, la evidencia en favor de (5) sugiere que al menos uno de los coeficientes β de las siguientes regresiones es no nulo:

$$\begin{aligned} e_{t-h}^{(1)}(h) &= \alpha_h^{(1)} + \beta_h^{(1)} y_{t-h}^{(1)}(h) + \varpi \\ e_{t-h}^{(2)}(h) &= \alpha_h^{(2)} + \beta_h^{(2)} y_{t-h}^{(2)}(h) + \varepsilon \\ \mathbb{E}\varpi &= \mathbb{E}y_{t-h}^{(1)}(h)e_{t-h}^{(1)}(h) = \mathbb{E}y_{t-h}^{(2)}(h)e_{t-h}^{(2)}(h) = \mathbb{E}\varepsilon = 0 \end{aligned}$$

esto es así porque los coeficientes β satisfacen

$$\beta_h^{(1)} \equiv \cos^{-1} \theta \frac{\mathbb{E}y_{t-h}^{(1)}(h)e_{t-h}^{(1)}(h)}{\mathbb{V}(y_{t-h}^{(1)}(h))}; \beta_h^{(2)} \equiv \frac{\mathbb{E}y_{t-h}^{(2)}(h)e_{t-h}^{(2)}(h)}{\mathbb{V}(y_{t-h}^{(2)}(h))}$$

y de acuerdo con (5), al menos uno de los términos

$$\mathbb{E}(y_{t-h}^{(1)}(h)e_{t-h}^{(1)}(h)); \mathbb{E}(y_{t-h}^{(2)}(h)e_{t-h}^{(2)}(h))$$

es no nulo. De esta manera, es posible construir los siguientes “pronósticos ajustados”:

$$y_{t-h}^{(1A)} = \alpha_h^{(1)} + (1 + \beta_h^{(1)})y_{t-h}^{(1)}(h) \text{ y } y_{t-h}^{(2A)} = \alpha_h^{(2)} + (1 + \beta_h^{(2)})y_{t-h}^{(2)}(h)$$

Que, tal como se muestra en Pincheira (2012b), tendrán un menor ECM que las predicciones originales.

En resumen, la detección de ineficiencias en el conjunto de pronósticos analizado, sugiere que existe espacio para modificarlos de manera de hacerlos más precisos aun. Estos resultados son coherentes con los mostrados por Bentancor y Pincheira (2008, 2010) y Pincheira y Fernández (2011a, 2011b) para las proyecciones de la EEE en Chile y las de *Consensus* en varios países.

VI. RESUMEN Y CONCLUSIONES

En este trabajo hemos analizado la optimalidad bajo pérdida cuadrática de una serie de pronósticos de inflación para Chile. Nuestro análisis ha explotado propiedades de optimalidad que plantean restricciones de desigualdad sobre la estructura temporal de las proyecciones, lo cual, de acuerdo con nuestra revisión de la literatura, no se ha hecho anteriormente para el caso chileno. Si bien algunas de las restricciones estudiadas corresponden a propiedades vastamente conocidas que relacionan los Errores Cuadrático Medios (ECM) de proyecciones construidas en distintos momentos o para distintos horizontes, otras propiedades evaluadas se refieren a la estructura temporal de los pronósticos propiamente tales y no de sus errores. Estas propiedades han sido recientemente derivadas por Patton y Timmermann (2012).

Una segunda contribución de nuestro trabajo se refiere a la manera en la cual estas propiedades son evaluadas. Patton y Timmermann (2012) diseñan tests basados en las contribuciones de Wolak (1987, 1989) que requieren la solución de uno o varios problemas de programación cuadrática convexa. En cambio en este trabajo nosotros diseñamos tests basados en los estadísticos del mínimo y del máximo que no requieren solucionar un problema de optimización. Estos tests se inspiran en los resultados de White (2000), Pincheira (2006, 2012a), Hansen (2005) y Romano y Wolf (2005).

Desde el punto de vista empírico, en este trabajo aplicamos un test conjunto de superioridad predictiva a las proyecciones de inflación chilena disponibles a través de las siguientes fuentes: *Bloomberg*, *Consensus Forecasts*, la Encuesta de Expectativas Económicas del Banco Central de Chile (EEE) y un promedio de modelos ESARIMA. Los resultados indican que en términos de Error Cuadrático Medio no es posible rechazar la hipótesis nula de optimalidad para el perfil conjunto de proyecciones. No obstante, cuando el test es aplicado a las Proyecciones al Cuadrado, en línea con la propuesta de Patton y Timmermann (2012), el test conjunto rechaza la hipótesis de optimalidad. El análisis de estos resultados sugiere que existe una violación a una condición de ortogonalidad implícita en la optimalidad bajo pérdida cuadrática, y que esta violación es transversal al conjunto de proyecciones bajo consideración y al horizonte predictivo, lo que sugiere la existencia de espacio para perfeccionar el perfil de proyecciones.

REFERENCIAS

Andersson, M., G. Karlsson. y J. Svensson (2007). "The Riksbank's Forecasting Performance". *Economic Review* 3: 59–75.

Ang, A., G. Bekaert y M. Wei (2007). "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics* 54(4): 1163–212.

Bentancor, A. y P. Pincheira (2008). "Forecasting Inflation Forecast Errors". Documento de Trabajo N°477, Banco Central de Chile.

Bentancor, A. y P. Pincheira (2010). "Predicción de Errores de Proyección de Inflación en Chile". *El Trimestre Económico* LXXVII(305): 129–54.

Bowles, C., R. Friz, V. Genre, G. Kenny, A. Meyler y T. Rautanen (2007). "The ECB Survey of Professional Forecasters (SPF) – A Review After Eight Year's Experience". Occasional Paper Series 59, Banco Central Europeo.

Capistrán, C. y G. López-Moctezuma (2010). "Las Expectativas Macroeconómicas de los Especialistas. Una Evaluación de Pronósticos de Corto Plazo en México". *El Trimestre Económico* LXXVII(306): 275–312.

Chumacero, R. (2001). "Empirical Analysis of Systematic Errors in Chilean GDP Forecasts". *Journal of Forecasting* 20(1): 37–45.

Croushore, D. (2010). "An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data". *The B.E. Journal of Macroeconomics* 10(1).

Diebold, F.X. y R. Mariano (1995). "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics* 13(3): 253–63.

Elliot, G. y A. Timmermann (2008). "Economic Forecasting". *Journal of Economic Literature* 46(1): 3–56.

Giacomini, R. y H. White (2006). "Tests of Conditional Predictive Ability". *Econometrica* 74(6): 1545–78.

Groen, J., G. Kapetanios y S. Price (2009). "A Real Time Evaluation of Bank of England Forecasts of Inflation and Growth". *International Journal of Forecasting* 25: 74–80.

Hansen, P.R. (2005). "A Test of Superior Predictive Ability". *Journal of Business & Economic Statistics* 23: 365–80.

Joutz, F. y H.O. Stekler (2000). "An Evaluation of the Prediction of the Federal Reserve". *International Journal of Forecasting* 16: 17–38.

Loungani, P. (2001). "How Accurate are Private Sector Forecasts? Cross-Country Evidence from Consensus Forecast of Output Growth". *International Journal of Forecasting* 17(3): 419–32.



Newey, W. y K. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". *Econometrica* 55(3): 703–8.

Newey, W. y K. West (1994). "Automatic Lag Selection in Covariance Matrix Estimation". *Review of Economic Studies* 61(4): 631–53.

Patton, A. y A. Timmermann (2012). "Forecast Rationality Tests Based on Multi-Horizon Bounds". *Journal of Business & Economic Statistics* 30(1): 1–17.

Pedersen, M. (2010). "Una Nota Introductoria a la Encuesta de Expectativas Económicas". Estudio Económico Estadístico N°82, Banco Central de Chile.

Pincheira, P. (2006). "Shrinkage Based Tests of the Martingale Difference Hypothesis". Documento de Trabajo N°376, Banco Central de Chile. DOI: 10.1002/for.1270

Pincheira, P. (2010a). "A Real Time Evaluation of the Central Bank of Chile GDP Growth Forecasts". *Money Affairs* XXIII(1): 37–73.

Pincheira, P. (2010b). "A Real Time Evaluation of the Central Bank of Chile GDP Growth Forecasts". Documento de Trabajo N°556, Banco Central de Chile.

Pincheira, P. (2012a). "Shrinkage-Based Tests of Predictability". *Journal of Forecasting*, DOI: 10.1002/for.1270.

Pincheira, P. (2012b). "Are Forecast Combinations Efficient?" Documento de Trabajo N°661, Banco Central de Chile.

Pincheira, P. y R. Álvarez (2009). "Evaluation of Short-Run Inflation Forecasts and Forecasters in Chile". *Money Affairs* XXII(2): 159–80.

Pincheira, P. y N. Fernández (2011a). "Jaque Mate a las Proyecciones de Consenso". Documento de Trabajo N°630, Banco Central de Chile.

Pincheira, P. y N. Fernández (2011b). "Corrección de Algunos Errores Sistemáticos de Predicción de Inflación". *Monetaria* XXXIV(1): 37–61.

Pincheira, P. y A. García (2009). "Forecasting Inflation in Chile with an Accurate Benchmark". Documento de Trabajo N°514, Banco Central de Chile.

Pincheira, P. y A. García (2012). "En Búsqueda de un Buen Marco de Referencia Predictivo para la Inflación en Chile". *El Trimestre Económico* LXXIX(313): 85–123.

Romano, J.P. y M. Wolf (2005). "Stepwise Multiple Testing as Formalized Data Snooping". *Econometrica* 73(4): 1237–82.

West, K. (1996). "Asymptotic Inference about Predictive Ability". *Econometrica* 64(5): 1067–84.

White, H. (2000). "A Reality Check for Data Snooping". *Econometrica* 68(5): 1067–84.



Wolak, F.A. (1987). "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model". *Journal of the American Statistical Association* 82: 782–93.

Wolak, F.A. (1989). "Testing Inequality Constraints in Linear Econometric Models". *Journal of Econometrics* 31: 205–35.

APÉNDICE

En este apéndice mostramos algunas propiedades de poder y tamaño del test del máximo que hemos presentado en el presente artículo. El ejercicio básico es uno en el que consideramos cuatro pronósticos de inflación contruidos en momentos sucesivos en el tiempo, por lo que deseamos testear una hipótesis nula análoga a la siguiente:

$$H_{01}^1 : ECM_{Bloomberg}^1 \leq ECM_{Consensus}^1 ; ECM_{Consensus}^1 \leq ECM_{EEE}^1 \text{ y } ECM_{EEE}^1 \leq ECM_{ESARIMA}^1$$

Si definimos

$$\mu_1 = ECM_{Bloomberg}^1 - ECM_{Consensus}^1$$

$$\mu_2 = ECM_{Consensus}^1 - ECM_{EEE}^1$$

$$\mu_3 = ECM_{EEE}^1 - ECM_{ESARIMA}^1$$

La hipótesis nula relevante es

$$H_0 : \vec{\mu} = (\mu_1, \mu_2, \mu_3) \leq (0, 0, 0)$$

o equivalentemente

$$H_0 : \text{Max}(\mu_1, \mu_2, \mu_3) \leq 0.$$

La hipótesis alternativa se refiere a cualquier violación de la nula, es decir, a cualquier situación en la que se cumpla

$$H_A : \mu_1 > 0 \text{ ó } \mu_2 > 0 \text{ ó } \mu_3 > 0.$$

Usando 5,000 replicaciones, simulamos tres series de datos de largo T que provienen de una distribución normal multivariada con esperanza $\vec{\mu}$ y varianza $\mathbb{V}$. Usamos

$$\mathbb{V} = \begin{pmatrix} 2,500 & 1,125 & 1,250 \\ 1,125 & 2,000 & 0,250 \\ 1,250 & 0,250 & 2,250 \end{pmatrix}$$

que, tal como se muestra en Pincheira (2012b), se define como positiva.

Los valores críticos del estadístico del máximo se obtienen tal como se especifica en los pasos 1-6 de la sección IV.3, vía simulaciones de Monte Carlo, usando también 5.000 replicaciones.

Los siguientes cuadros muestran resultados de tamaño y poder para distintos valores de T , y distintas especificaciones de $\vec{\mu}$:

Cuadro A1

Tamaño empírico del test del máximo

con datos generados por una normal multivariada de esperanza $\mu = (0;0;0)$

Observaciones	Error de tipo I al 10%	Error de tipo I al 5%
50	10,56%	5,14%
100	9,94%	5,42%
200	10,16%	5,14%
500	9,98%	4,90%
1000	10,08%	4,94%

Nota: Los valores esperados de los momentos diferenciales son todos nulos.

Cuadro A2

Tamaño empírico del test del máximo

con datos generados por una normal multivariada de esperanza $\mu = (0;-0,1;0)$

Observaciones	Error de tipo I al 10%	Error de tipo I al 5%
50	8,32%	4,06%
100	7,46%	3,38%
200	7,46%	3,20%
500	6,96%	3,30%
1000	6,84%	3,34%

Nota: Los valores esperados de los momentos diferenciales son todos nulos, salvo el segundo que es -0,1.

Cuadro A3

Potencia del test del máximo

con datos generados por una normal multivariada de esperanza $\mu = (0;0,1;0)$

Observaciones	Poder del test al 10%	Poder del test al 5%
50	17,36%	9,80%
100	19,46%	11,08%
200	26,18%	16,46%
500	46,26%	33,66%
1000	69,98%	57,12%

Nota: Los valores esperados de los momentos diferenciales son todos nulos, salvo el segundo que es 0,1.

Cuadro A4

Potencia del test del máximo

con datos generados por una normal multivariada de esperanza $\mu = (0,1;0,1;0,1)$

Observaciones	Poder del Test al 10%	Poder del Test al 5%
50	24,8%	15,50%
100	30,82%	19,24%
200	42,66%	28,88%
500	69,06%	54,48%
1000	89,58%	80,94%

Nota: Los valores esperados de los momentos diferenciales son todos 0,1.

El cuadro A1 muestra el error empírico de tipo I cometido por el test del máximo. Este cuadro es generado por series que efectivamente provienen de una normal con valor esperado nulo, por lo que se espera el buen comportamiento del tamaño empírico que se ve reflejado en ese cuadro.

El cuadro A2 muestra resultados de tamaño empírico cuando $\vec{\mu} = (0;-0,1;0)$. Debemos recalcar que esta configuración de valores esperados es compatible con la hipótesis nula, por lo que tiene sentido evaluar el desempeño del test en estas condiciones. Como es de esperar, el test es más conservador que lo que plantea el tamaño nominal. Al 10% el test tiene un tamaño empírico en torno al 7%. Cuando el tamaño nominal es 5%, el test tiene un tamaño empírico de alrededor del 3%.

Los cuadros A3 y A4 muestran resultados de potencia contra dos alternativas distintas. En el cuadro A3 se tiene $\vec{\mu} = (0;0,1;0)$, mientras en el cuadro A4 se tiene $\vec{\mu} = (0,1;0,1;0,1)$. La potencia del test es baja con muestras de tamaño pequeño y moderado, mientras que sube importantemente cuando el tamaño muestral se encuentra en torno a las 500 observaciones o más.