Banco de México
Documentos de Investigación

Banco de México
Working Papers

N° 2007-14

**Pruebas de Eficiencia para Pronósticos con Horizontes
Múltiples**

Carlos Capistrán
Banco de México

Diciembre 2007

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively the responsibility of the authors and do not necessarily reflect those of Banco de México.

Pruebas de Eficiencia para Pronósticos con Horizontes Múltiples*

Carlos Capistrán[†]
Banco de México

Abstract

This paper develops and analyzes a series of tests to evaluate the optimality of forecasts when forecasts for more than one horizon are available. The tests are based on the property that the unconditional expected loss of optimal forecasts should not decrease with the forecast horizon (e.g., under quadratic loss the variance of optimal forecast errors should not decrease with the horizon). The tests complement existing methods of forecast evaluation, such as Mincer-Zarnowitz-type tests, by using an implication of optimality that directly concerns forecasts made at different horizons. The finite sample performance of the tests is analyzed and an illustration using the Survey of Professional Forecasters is provided.

Keywords: Forecast Evaluation, Composite Hypothesis.

JEL Classification: C12, C53, E27.

Resumen

Este documento desarrolla y analiza una serie de pruebas para evaluar la eficiencia de pronósticos cuando éstos se encuentran disponibles para más de un horizonte. Las pruebas se basan en la propiedad de que la pérdida esperada no condicional de los pronósticos óptimos no debe decrecer con el horizonte de pronóstico (e.g., para el caso de una pérdida cuadrática, la varianza de los errores de un pronóstico óptimo no debe decrecer con el horizonte). Estas pruebas complementan métodos existentes para la evaluación de pronósticos, como lo son las pruebas del tipo Mincer-Zarnowitz, al utilizar una implicación de eficiencia que concierne directamente a pronósticos hechos para distintos horizontes. Se analiza el desempeño de estas pruebas para muestras finitas y se provee un ejemplo empleando la Encuesta de Pronosticadores Profesionales (SPF, por sus siglas en inglés).

Palabras Clave: Evaluación de Pronósticos, Hipótesis Compuestas.

* Agradezco a Graham Elliot y Allan Timmermann por plantear el problema y por sus comentarios. Las conversaciones con Clive W.J. Granger fueron de gran ayuda. Agradezco a Vince Crawford, Marjorie Flavin, Antonio E. Noriega, Yixiao Sun, Elliot Williams y Carla Ysusi por sus valiosos comentarios.

[†] Dirección General de Investigación Económica. Email: ccapistran@banxico.org.mx.

1 Introducción

La forma usual de proceder en la literatura de pronóstico cuando se analiza un conjunto de pronósticos, es definiendo un pronóstico óptimo, para después, utilizar las propiedades derivadas de la optimalidad para evaluarlo. Dos propiedades, ambas derivadas bajo una pérdida de error cuadrático medio (ECM), se utilizan rutinariamente para este propósito. La primera, conocida como la propiedad de insesgamiento, indica que los errores de pronóstico deben tener media cero. La segunda, conocida como la propiedad de eficiencia, dice que no debe ser posible predecir estos errores utilizando información disponible al pronosticador al momento de hacer los pronósticos (Clements y Hendry (1998); Granger y Newbold (1996)). Las pruebas actuales de optimalidad de pronósticos están diseñadas para evaluar estas dos propiedades.¹

Sin embargo, estas propiedades se refieren a pronósticos para un horizonte fijo, lo cual contrasta con la manera en que normalmente se producen los pronósticos. Los pronosticadores no suelen predecir eventos solamente un paso hacia delante, sino que usualmente predicen dos, tres, o más pasos hacia delante al mismo tiempo. Este tipo de pronósticos se conocen como pronósticos con múltiples horizontes. Si no se incorpora esta información al momento de determinar la calidad de los pronósticos, se puede concluir incorrectamente que los pronósticos son óptimos. Lo que se necesita es una propiedad que relacione directamente pronósticos óptimos a diferentes horizontes.

Una buena candidata es una propiedad que, nuevamente bajo ECM, indica que mientras más lejos en el futuro se pronostique, menos preciso será el pronóstico óptimo, donde la precisión es medida por la varianza de los errores de pronóstico: ésta se conoce como propiedad de “precisión decreciente” (Granger y Newbold (1986)). Cuando se consideran pronósticos de eventos fijos, esto es, pronósticos del mismo evento hechos a diferentes horizontes, esta propiedad implica que mientras más cercano está el evento de interés al momento de pronosticarlo, más preciso será el pronóstico: esta se conoce como propiedad de “convergencia” (Swidler y Ketcher (1990)). Esta propiedad claramente establece una relación entre pronósticos óptimos con múltiples horizontes.

Aún cuando las propiedades de los pronósticos óptimos se han considerado típicamente bajo ECM, la literatura de pronóstico reciente ha explorado el papel de funciones de pérdida más generales, por ejemplo, asimétricas (Diebold y Lopez (1996); Granger (1999)). Bajo funciones de pérdida generales, las propiedades de los pronósticos óptimos son algo diferentes. Por ejemplo, no se mantiene el hecho de que los errores de pronóstico deben de tener media cero, sino que puede existir un sesgo óptimo (Zellner (1986); Elliott, et al. (próxima

¹Veáse la revisión por Stekler (2002).

publicación)). En el caso de la propiedad que relaciona diferentes horizontes, la generalización muestra que no es la varianza de los errores el objeto que debe ser no-decreciente con el horizonte, sino su pérdida esperada (incondicional), la cual, en algunos casos (como en ECM) coincide con la varianza (Patton y Timmermann (próxima publicación)).

Es el objetivo de este documento investigar a detalle la propiedad de la precisión decreciente, o convergencia, bajo funciones de pérdida generales, así como desarrollar pruebas para ésta.

2 Pronósticos Óptimos

La manera tradicional para determinar la calidad de los pronósticos para diferentes horizontes hechos por el mismo pronosticador es comparando alguna medida estadística del tamaño de los errores, por ejemplo, el error medio absoluto (EMA) o el error cuadrático medio (ECM). El Cuadro 1 presenta un ejemplo de lo que se reporta.usualmente² ¿Cómo se puede interpretar el Cuadro 1? Si los pronosticadores estuvieran pronosticando los mismos trimestres (lo cual no es cierto) uno podría querer comparar el tamaño de los ECM entre los pronosticadores para un horizonte fijo. Para eso, uno podría utilizar pruebas de habilidad de predicción (e.g., Diebold y Mariano (1995)). ¿Pero, qué se puede hacer con la información entre diferentes horizontes? ¿Cómo se puede interpretar? Lo primero que se puede observar, es que para la mayoría de los pronosticadores, la raíz del ECM (RECM) aumenta monotónicamente con h , el horizonte de pronóstico, lo que no ocurre con los pronosticadores 1, 3, 96 y 452.³ ¿Cuáles son las implicaciones de este comportamiento? ¿Se puede evaluar la calidad de los pronósticos con base en esta diferencia?

Una revisión de la literatura sobre evaluación de pronósticos óptimos proporciona una idea sobre como contestar estas preguntas. Sin embargo, la literatura ha tomado dos caminos diferentes para probar la optimalidad de los pronósticos. Ambos caminos están relacionados porque se basan en las mismas propiedades. Pero son diferentes debido a la estructura de los datos a las cuales se han aplicado estas propiedades. La primer estructura se conoce como pronósticos de eventos-móviles, donde el pronosticador pronostica diferentes eventos fijando el horizonte de pronóstico y variando la fecha inicial (Mincer y Zarnowitz (1969)). La segunda estructura se conoce como pronósticos de eventos-fijos porque, como se describió anteriormente, se fija el evento pero se varía el horizonte, por lo que el pronosticador se

²La tabla contiene información utilizada en la aplicación al final del documento, donde se pueden encontrar los detalles acerca de su construcción. Los pronosticadores son una muestra aleatoria tomada de la Encuesta de Pronosticadores Profesionales (SPF). $RECM_h$ es la raíz del ECM para el horizonte h .

³El ECM es igual a la varianza de los errores cuando los pronósticos son insesgados.

acerca naturalmente al evento (Nordhaus (1987)).⁴ Dada esta división de la literatura, la revisión en esta sección esta dividida en tres partes, en la primera subsección se revisan las características en común y las propiedades de los pronósticos óptimos. En las siguientes dos, se revisan los aspectos específicos a cada estructura.

2.1 Pronósticos Óptimos en el Caso Lineal-Cuadrático Tradicional

La estructura tradicional para analizar pronósticos óptimos es la estructura lineal-cuadrática bajo parámetros conocidos. La parte cuadrática se refiere a la función de pérdida empleada, de tipo ECM, y la parte lineal a la forma que, se asume, tiene el modelo del pronosticador, la cual es lineal en los parámetros.

La pérdida ECM implica que el pronóstico óptimo es la media condicional de las variables de interés, Y_t , el condicionamiento se hace bajo el conjunto de información $\Omega_t : Y_{t-j}, j = 0, 1, \dots$ el cual se conoce como el conjunto de información propio.⁵ La media condicional es aproximada por un modelo lineal, el cual puede ser una buena aproximación si Y_t es un proceso estacionario en covarianza.⁶

Sea $f_{t+h,t}$ el pronóstico de una variable aleatoria Y_{t+h} hecho en el tiempo t (el pronóstico h-pasos hacia delante). El pronóstico óptimo es: $f_{t+h,t}^* = \sum_{j=0}^{\infty} \psi_{h+j} \varepsilon_{t-j}$ con el error de pronóstico correspondiente: $\varepsilon_{t+h,t}^* = y_{t+h} - f_{t+h,t}^* = \sum_{j=0}^{h-1} \psi_j \varepsilon_{t+h-j}$ (Granger y Newbold (1986)). Las propiedades usuales de los pronósticos óptimos bajo este contexto son (Granger y Newbold (1986)):

1. El error de pronóstico óptimo tienen media condicional e incondicional cero. Esto es, el pronóstico óptimo es insesgado.
2. El error de pronóstico óptimo es ortogonal a cualquier función de cualquier elemento contenido en el conjunto de información Ω_t . Una implicación de esta propiedad, cuando se tiene un conjunto de información propio, es que los errores h-pasos hacia delante tienen la misma autocorrelación que un $MA(h-1)$.

Estas propiedades han sido utilizadas extensivamente en economía. Casi siempre están

⁴Con datos de pronósticos con múltiples horizontes uno siempre puede considerar que los datos tienen cualquiera de las dos estructuras.

⁵En el caso lineal-cuadrático tradicional, el conjunto de información típicamente incluye el pasado y el presente de la variable de interés. Sin embargo, el conjunto también puede incluir otras variables, siempre y cuando éstas se conozcan al momento de hacer el pronóstico.

⁶El teorema de la representación de Wold indica que cualquier proceso estacionario en covarianza puede ser representado como un promedio móvil infinito mas un término lineal determinístico $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + v_t$, donde $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, siempre que los coeficientes del MA sean absolutamente sumables (Hamilton (1994)).

ligadas a la literatura de expectativas racionales, pero también se aplican en otros contextos.⁷

Las propiedades se refieren a un solo horizonte de pronóstico. Sin embargo, Nordhaus (1987) presenta una manera de utilizar la segunda propiedad cuando se dispone de datos de pronósticos con múltiples horizontes. Nordhaus señala que si los pronósticos son del tipo de eventos-fijos, la propiedad 2 implica que las revisiones de los pronósticos no deben ser pronosticables cuando se tiene información disponible hasta antes de la revisión. Al usar información contenida en diferentes horizontes, la prueba sirve para analizar la manera en que se incorpora nueva información en los pronósticos.

Pero los pronósticos óptimos tienen otras propiedades que relacionan más directamente a los pronósticos con diferentes horizontes. En particular, es posible derivar una tercera propiedad a partir de la varianza de los errores de pronósticos óptimos, $var(e_{t+h,t}^*) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} \psi_j^2$ (Granger y Newbold (1986)):

3. La varianza de un pronóstico óptimo es una función no-decreciente de h , el horizonte de pronóstico.

Bajo este marco conceptual, tanto la varianza condicional como la incondicional del error de pronóstico es la misma, por lo que la propiedad aplica para ambas. Cuando los pronósticos son de tipo eventos-móviles, la propiedad 3 se aplica hacia delante y se interpreta como una pérdida de precisión conforme uno pronostica más lejos en el futuro. Cuando los pronósticos son de tipo eventos-fijos, la propiedad 3 se aplica hacia atrás y se interpreta como una ganancia en precisión conforme uno se aproxima al evento de interés.

2.2 Pronósticos de Eventos-Móviles

En contraste con las dos primeras propiedades, no se puede encontrar en la literatura una prueba de la propiedad de precisión decreciente, aún cuando varios investigadores han encontrado que esta propiedad es útil para determinar la optimalidad de los pronósticos.⁸ Por ejemplo, Granger (1981) comenta: “Si $f_{t+h,t}$ es un pronóstico óptimo [bajo ECM] basado en un conjunto de información propio, entonces la varianza del error de pronóstico h -pasos hacia delante será una sucesión no-decreciente conforme aumenta h ... si los pronósticos no hacen un uso óptimo del conjunto de información, las varianzas de los errores de pronóstico no necesariamente tienen que incrementarse conforme uno pronostica más lejos en el tiempo” (traducción propia) o más recientemente, en su libro de texto sobre pronósticos, Diebold

⁷Para una revisión sobre los diferentes usos, y sobre las pruebas que se han desarrollado, véase Diebold y Lopez (1996).

⁸Existen algunas pruebas para la forma “convergente” de la propiedad, esto es, cuando se aplica a pronósticos de eventos-fijos. Estas pruebas se discuten más adelante.

(2001) escribe: “Hemos aprendido numerosas lecciones acerca de la optimalidad de pronósticos mientras ignoramos la incertidumbre de la estimación de los parámetros, por ejemplo: La varianza del error de pronóstico aumenta conforme el horizonte de pronóstico se alarga,... Tales lecciones proveen intuición y entendimiento en cuanto a la mecánica de modelos de pronóstico y proveen un valioso punto de referencia para evaluar pronósticos...” (Traducción propia).

Para el ejemplo del principio de esta sección (Cuadro 1), la literatura indica que la optimalidad de los pronosticadores 1, 3, 96 y 452 (los no-monotónicos) debe ponerse en duda, aunque no se ha propuesto un método formal para hacer esto.

Quizá, la razón principal por la cual no se han desarrollado pruebas de la tercera propiedad sea porque cuando se relaja el supuesto de certidumbre en los parámetros, la propiedad se desvanece. Este resultado data, dentro de la literatura de pronósticos económicos, desde por lo menos Schmidt (1974), y es conocido como la “no-monotonicidad del ECM”.

El problema es que, cuando existe incertidumbre en los parámetros, la varianza condicional tiene un segundo término que depende de la diferencia entre el parámetro estimado y el poblacional, el cual es creciente en h . No se han desarrollado resultados generales acerca de este segundo término, sin embargo, se han estudiado algunos casos especiales. Por ejemplo, para el $AR(1)$: $y_t = \phi y_{t-1} + \varepsilon_t$, $|\phi| < 1$, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, Clements y Hendry (1998) presentan el siguiente resultado:

$$var(\hat{e}_{T+h,T}^* | \Omega_T) = \sigma_\varepsilon^2 \frac{(1 - \phi^{2h})}{1 - \phi^2} + E \left[\left(\phi^h - \hat{\phi}^h \right)^2 \right] y_T^2,$$

el cual aproximan con:

$$var(\hat{e}_{T+h,T}^* | \Omega_T) = \sigma_\varepsilon^2 \frac{(1 - \phi^{2h})}{1 - \phi^2} + h^2 \phi^{2(h-1)} T^{-1} (1 - \phi^2) y_T^2,$$

donde T es el número total de observaciones utilizadas en la estimación por MCO del modelo $AR(1)$ (véase también Box et al., 1994, p. 304). El primer término, la varianza asintótica del error, es el mismo que en el caso de certidumbre en los parámetros, por lo que es no-decreciente en h . En el segundo término es donde surge el problema. Chong y Hendry (1986) encontraron que $h^2 \phi^{2(h-1)}$ es no-monotónico en h ya que tiene un máximo. Por lo tanto, el segundo término puede hacer que la varianza condicional sea no-monotónica en h . Este resultado sólo es válido en muestras finitas, ya que el término se desvanece asintóticamente ($T \rightarrow \infty$). Esto implica que con muestras finitas es más complicado generar pruebas de la tercera propiedad.

Este resultado genera conclusiones como la siguiente, nuevamente del libro de texto de

Diebold (2001): “[Sobre las lecciones aprendidas bajo parámetros conocidos] algunas veces necesitan modificarse cuando se reconoce incertidumbre en la estimación de los parámetros: La varianza del error de pronóstico no necesariamente crece monotónicamente con el horizonte. Típicamente *esperamos* que la varianza del error de pronóstico crezca monotónicamente con el horizonte, pero no *tiene* por que hacerlo.” (Cursiva en el original, traducción propia). Regresando al Cuadro 1, bajo incertidumbre en los parámetros no queda claro como debe interpretarse la no-monotonicidad de los pronosticadores 1, 3, 96 y 452. Algo puede estar mal, pero debido a que la varianza *no tiene que aumentar* (de acuerdo a Diebold), parecería que uno no puede ser definitivo.

2.3 Pronósticos de Eventos-Fijos

Swidler y Ketcher (1990), al analizar pronósticos de eventos-fijos, encontraron que debido a que el conjunto de información es no-decreciente con el horizonte de pronóstico, el pronóstico óptimo del evento τ hecho h pasos hacia atrás debe ser más preciso que un pronóstico óptimo del mismo evento hecho $h + j$ pasos hacia atrás ($j > 0$). Esto es lo mismo que la propiedad 3 mencionada anteriormente, solamente aplicado a pronósticos de eventos-fijos. Para probar esta propiedad se observaron las R^2 s de las regresiones de eficiencia ($Y_\tau = \alpha_h + \beta_h f_{\tau, \tau-h} + u_{\tau, \tau-h}$) hechas para diferentes horizontes. Su argumento es que si la propiedad se satisface, las R^2 s deben ser no-crecientes cuando h aumenta.⁹ Al analizar las encuestas de Blue Chip de pronósticos sobre crecimiento del PNB real e inflación, encuentran que, aparentemente, para ambas variables los pronosticadores logran hacer mejor pronósticos conforme se acerca el evento de interés. El problema con este enfoque es que no toman en cuenta la distribución muestral de las R^2 s.

Batchelor y Dua (1991) nombran esta propiedad como convergencia. La escriben como:

$$E(e_{\tau, \tau-h}^{*2} | \Omega_{\tau-h-j}) \leq 0, \quad j > 0,$$

y proponen dos pruebas. La primera consiste en correr una regresión de los errores cuadráticos contra una constante y h , y probar si el coeficiente de h es no-negativo. El problema con esta prueba, es que la relación no tiene porque ser lineal, como ellos asumen. También, debe de tomarse en cuenta la autocorrelación de los residuales, como en la propiedad 2, ya que aún los errores de un pronóstico óptimo presentan correlación serial. La segunda prueba es una prueba binomial no-paramétrica, donde el estadístico de prueba es el número de incrementos de los errores cuadráticos mes a mes. Sin embargo, aplican esta prueba utilizando

⁹Esto tiene sentido en el esquema lineal-cuadrático porque la R^2 es una transformación monótona de la varianza del error de pronóstico.

solamente los pronósticos para un evento a la vez, perdiendo la información contenida al tomar en cuenta los pronósticos entre eventos. Al observar el PNB real, el deflactor de precios del PNB, el desempleo, y las tasas T-bill, encuentran que el pronóstico del consenso en cada caso parece ser racional, pero que el análisis de los pronosticadores individuales genera evidencia de irracionalidad (lo cual equiparan con falta de optimalidad, al igual que muchos en la literatura) para algunos pronosticadores.

A partir de esta literatura, los pronosticadores no-monotónicos del Cuadro 1 serían etiquetados como no-óptimos (o incluso, irracionales), ya que, aparentemente, no logran satisfacer la propiedad de convergencia.

3 La Propiedad de la Pérdida Esperada No-Decreciente

En años recientes, se han realizado varias generalizaciones del esquema lineal-cuadrático tradicional de optimalidad de pronósticos (e.g., Granger (1999)). Se han considerado tanto modelos no-lineales, como en Granger y Terasvirta (1993), Clements y Hendry (1998), y Patton y Timmermann (2007), como otras funciones de pérdida, incluyendo funciones asimétricas, por ejemplo Granger (1969), Christoffersen y Diebold (1996) y Patton y Timmermann (2007).

A continuación, se presenta, de forma general, la teoría detrás de la propiedad de convergencia, o de precisión decreciente, de los pronósticos óptimos, construyendo sobre el trabajo de Granger (1999) y Patton y Timmermann (próxima publicación, 2007). Se permiten funciones de pérdida asimétricas al igual que modelos diferentes, posiblemente no-lineales, para diferentes horizontes. Así, se utiliza la propiedad para determinar la hipótesis nula de interés.

3.1 La Propiedad de Interés

Cuando la variable de interés es Y_t y el pronóstico se hace h pasos hacia delante, entonces la variable a pronosticar es Y_{t+h} . La distribución condicional de esta variable aleatoria es $F_{Y_{t+h}|\Omega_t}(y) = \Pr(Y_{t+h} \leq y | \Omega_t)$, el condicionamiento hecho con respecto a la información conocida al tiempo t , denotada por Ω_t . El conjunto de información es tal que contiene el pasado y el presente de la variable de interés y de cualquier variable disponible al pronosticador en el momento en que se produce el pronóstico, $\Omega_t : Y_{t-j}, X_{i,t-j} \quad i = 1, \dots, l \quad j = 0, 1, \dots$. Se asume que cualquier parte determinística ha sido eliminada de Y_t . La función de pérdida $L : \mathbb{R} \longrightarrow \mathbb{R}^+$ toma un error de pronóstico y regresa su pérdida. Esto es lo que Christoffersen y Diebold (1997) llaman una función de pérdida de predicción-error, $L(y_{t+h} - f_{t+h,t}) = L(e_{t+h,t})$. Los únicos requerimientos de esta función de pérdida es que

sea débilmente convexa, con la normalización $L(0) = 0$, y que sea diferenciable, al menos una vez. En este contexto, un pronóstico óptimo se define como sigue:

Definición 1 (Pronóstico Óptimo) *Un pronóstico óptimo $f_{t+h,t}^*$ es el pronóstico que minimiza la pérdida esperada. En un esquema paramétrico, esto se puede escribir como:*

$$f_{t+h,t}^* \equiv g_h(x_t, \beta_h^*)$$

$$\beta_h^* \equiv \arg \min_{g_h} E[L(y_{t+h} - g_h(x_t, \beta_h)) | \Omega_t],$$

donde

$$E[L(y_{t+h} - g_h(x_t, \beta_h)) | \Omega_t] = \int_e L(y_{t+h} - g_h(x_t, \beta_h)) dF_{e|\Omega_t}(e),$$

$$x_t \in \Omega_t, \quad \beta_h \text{ es un parámetro.}$$

Si un pronosticador quiere pronosticar conjuntamente más de un paso hacia delante al mismo tiempo, el problema puede ser mejor modelado utilizando una función de pérdida de horizontes múltiples. Sin embargo, esto no es común en la literatura. Para este documento, uno puede pensar en el pronosticador como alguien que tiene una función de pérdida sobre horizontes múltiples pero que es separable entre ellos. En este caso los resultados son equivalentes a los aquí presentados.¹⁰

Ya que el pronóstico óptimo se define como la solución al problema de minimización, los dos objetos matemáticos usuales que se observan para examinar sus propiedades son la condición de primer orden (CPO) y la función de valor.

Definición 2 (CPO) *La condición de primer orden que hay que resolver para el pronóstico óptimo es:*¹¹

$$E \left[\left. \frac{\partial L(y_{t+h} - g_h(x_t, \beta_h))}{\partial \beta_h} \right|_{\beta_h^*} \middle| \Omega_t \right] = 0$$

La literatura de evaluación de pronósticos se ha concentrado en probar la optimalidad de un conjunto de pronósticos utilizando esta condición, que usualmente se conoce como la

¹⁰Por ejemplo, usando una pérdida de error ECM, si se quiere generalizar la pérdida definiendo $e_H = (e_{t,t+1}, \dots, e_{t,t+H})'$, se puede replantear el problema como $\min e_H' W e_H$. El problema es equivalente a minimizar la pérdida un horizonte a la vez (i.e. el problema es separable), siempre y cuando W sea una matriz diagonal positiva definida (Capistrán (2006)).

¹¹Esto es cierto siempre y cuando F satisfaga algunas condiciones de regularidad que permita el intercambio de diferenciación e integración. La CPO es necesaria para obtener un óptimo. Si L es estrictamente convexa, entonces la CPO también será suficiente.

propiedad de la secuencia de martingalas en diferencia (smd) (cuando $h = 1$) (e.g., Granger (1999); Patton y Timmermann (próxima publicación)). Cualquier prueba que se base en esta propiedad tiene que especificar la función de pérdida y el conjunto de información utilizados por el productor del pronóstico. Por ejemplo, cuando el conjunto de información incluye el pasado de las series, entonces la propiedad smd implica que la pérdida marginal h -pasos hacia delante, cuando se evalúa en el óptimo, tiene una media condicional e incondicional igual a cero y una estructura de promedios móviles que se comporta como un $MA(h - 1)$. Si la pérdida es ECM, entonces la pérdida marginal es el error de pronóstico y estas propiedades se aplican a este, demostrando que realmente esta es una generalización del esquema lineal-cuadrático. Como se mencionó anteriormente, estas propiedades sólo concierren a un horizonte de pronóstico a la vez. Todas las pruebas de eficiencia existentes en la literatura se pueden ver como variantes de esta propiedad. La propiedad de insesgamiento se obtiene cuando se considera una constante (ya que siempre forma parte del conjunto de información), y la propiedad de eficiencia se obtiene con cualquier otra variable que pertenezca al conjunto de información.

Definición 3 (Función de valor) *La función de valor es una función del tiempo para el cual se hace el pronóstico, del tiempo en que se hace el pronóstico, y del conjunto de información.*¹²

$$V(t + h, t, \Omega_t) = E[L(y_{t+h} - g_h(x_t, \beta_h^*)) | \Omega_t].$$

Hasta ahora, el marco conceptual corresponde a pronósticos de eventos-móviles, para el cual la propiedad de precisión decreciente es la relevante. Patton y Timmermann (próxima publicación) la probaron utilizando la Definición 3. La manera de relacionar pronósticos óptimos para diferentes horizontes es comparando sus funciones de valor, manteniendo todo lo demás constante. Su resultado se resume en la siguiente proposición, tomado de Patton y Timmermann (próxima publicación), pero repetido aquí para complementar y para definirlo en la notación del documento:

Proposición 4 (Precisión Decreciente) *(De Patton y Timmermann (próxima publicación)). Si el error de pronóstico óptimo es estrictamente estacionario y la función de pérdida es del tipo predicción-error y no varía en el tiempo, entonces, la pérdida esperada incondicional de un pronóstico óptimo, para un conjunto de información dado Ω_t , es una función no-decreciente del horizonte de pronóstico, h , i.e.*

$$V(t + h, t) \leq V(t + h + j, t) \quad \forall j > 0.$$

¹²En terminología estadística, la función de valor es el riesgo evaluado en el óptimo.

Es posible interpretar la Proposición 4 diciendo que en general es más difícil pronosticar mientras más largo sea el horizonte (i.e. la precisión de los pronósticos es decreciente en h , de aquí el nombre de precisión decreciente).

La propiedad es lo suficientemente general para permitir funciones de pérdida asimétricas. Aquí se plantea que se asume una función de pérdida de predicción-error, sin embargo, la propiedad se puede probar para funciones de pérdida más generales, siempre y cuando uno haga los supuestos apropiados sobre F , por ejemplo, que existe estacionariedad estricta de Y_{t+h} . En cambio, cuando uno hace supuestos más fuertes sobre la función de pérdida, es posible relajar los requerimientos sobre F , por ejemplo, bajo la pérdida ECM sólo es necesario que Y_{t+h} sea un proceso estacionario en covarianza.

El requerimiento de que el error de pronóstico óptimo sea estrictamente estacionario puede obtenerse de varias formas. Una es asumiendo directamente que Y_{t+h} es una variable aleatoria estrictamente estacionaria. Sin embargo, como se mencionó anteriormente, la proposición 4 permite una Y_{t+h} que no sea estrictamente estacionaria. Esto es teóricamente posible ya que el pronóstico óptimo cancela cualquier propiedad dominante de la variable de interés. De este modo, lo que es relevante es que el error de pronóstico satisfaga el supuesto de estacionariedad estricta.

Cuando se analizan pronósticos de horizontes múltiples como pronósticos de eventos-fijos, la propiedad relevante es la de convergencia. La propiedad también puede ser probada utilizando la definición 3 (cambiando la notación de la manera obvia).¹³

Proposición 5 (Convergencia) *Si la función de pérdida no varía en el tiempo, y el conjunto de información es una filtración, entonces la pérdida esperada incondicional de un pronóstico óptimo, para un conjunto de información dado, $\Omega_{\tau-h}$, es una función no-decreciente del horizonte de pronóstico h , i.e.*

$$V(\tau, \tau - h) \leq V(\tau, \tau - h - j) \quad \forall j > 0.$$

Se puede interpretar la Proposición 5 diciendo que en general es más fácil pronosticar mientras más cerca se este del evento (i.e. los pronósticos convergen al resultado del evento de interés, de ahí el nombre de convergencia).

La propiedad es lo suficientemente general para permitir funciones de pérdida asimétricas. La ventaja de utilizar convergencia es que no es necesario suponer estacionariedad estricta. También, la interpretación de que los pronosticadores convergen al valor actual de la variable parece ser más intuitiva. Asimismo, la insuficiencia de los pronósticos para cumplir la propiedad de convergencia puede interpretarse fácilmente como un fracaso de los pronosti-

¹³Las pruebas de las proposiciones se presentan en el Apéndice.

cadore en incorporar informaci3n nueva Para el resto del documento, se utiliza la forma de convergencia de la propiedad, aunque la discusi3n y las pruebas que se presentan a continuaci3n aplican para ambas formas.

3.2 Comentarios sobre la Propiedad de Inter3s

Lo primero a tener en cuenta es que la propiedad concierne a la p3rdida esperada incondicional, pero que el pron3stico 3ptimo empleado en su c3lculo es funci3n del conjunto de informaci3n. Adicionalmente, n3tese que la p3rdida esperada *condicional* no necesariamente es una funci3n no-decreciente de h .

Cuando la funci3n de p3rdida es ECM, la funci3n de valor incondicional es igual a la varianza incondicional del error de pron3stico 3ptimo, $E [L (y_\tau - f_{\tau,\tau-h}^*)] = \text{var} (e_{\tau,\tau-h}^*)$. En este caso, la Propiedad 3 implica que la varianza incondicional es no-decreciente en h , mostrando nuevamente que el esquema lineal-cuadr3tico tradicional es un caso particular. En algunas situaciones el resultado tambi3n aplica a la varianza condicional (e.g., si los pron3sticos son generados por un $AR(1)$).

En algunos casos es posible utilizar una prueba aplicada a la varianza como prueba de la propiedad de inter3s, aun cuando la funci3n de p3rdida no sea cuadr3tica. El resultado es resumido en la siguiente proposici3n:

Proposici3n 6 *Si la funci3n de p3rdida es del tipo de predicci3n-error, invariante en el tiempo, $E [e_{\tau,\tau-h}^*] = E [e_{\tau,\tau-h-j}^*] \quad \forall j > 0$, y siempre y cuando una de las siguiente dos condiciones se mantengan,*

1. *La funci3n de p3rdida es, m3ximo, diferenciable dos veces, o*
2. *Por lo menos los dos primeros momentos de la funci3n de distribuci3n condicional existen,*

entonces,

$$E [L (e_{\tau,\tau-h}^*)] - E [L (e_{\tau,\tau-h-j}^*)] \approx \text{var} (e_{\tau,\tau-h}^*) - \text{var} (e_{\tau,\tau-h-j}^*), \quad \forall j > 0.$$

¿Cu3ndo se cumplen las condiciones requeridas para la Proposici3n 6? Siempre que la funci3n de p3rdida sea invariante en el tiempo y $f_{\tau,\tau-h}^* = E [Y_t | \Omega_{\tau-h}]$, se cumplir3 el resultado, como es claro de Granger (1969) los requerimientos bajo los cuales el pron3stico 3ptimo es la media condicional. Las condiciones tambi3n se cumplen si la distribuci3n condicional es Gaussiana, a3n cuando la p3rdida sea asim3trica, porque en ese caso $f_{\tau,\tau-h}^* = E [Y_t | \Omega_{\tau-h}] + \gamma$, pero γ es independiente del conjunto de informaci3n (Granger (1999)). Las condiciones no se cumplen en general con funciones de p3rdida asim3tricas si la funci3n de distribuci3n condicional depende del conjunto de informaci3n, ya que en ese caso el sesgo 3ptimo es variable

en el tiempo (Granger (1999); Christoffersen y Diebold (1997)).

3.3 La Hipótesis de Interés

Con datos de pronósticos de horizontes múltiples, uno podría querer probar la optimalidad de los pronósticos utilizando la Proposición 5, la propiedad de convergencia. Esto sugiere la siguiente hipótesis nula compuesta:

$$H_0 : E [L (e_{\tau, \tau-1}^*)] \leq E [L (e_{\tau, \tau-2}^*)] \leq \dots \leq E [L (e_{\tau, \tau-H}^*)] , \quad (1)$$

donde H es el horizonte de pronóstico más grande. La hipótesis 1 es compuesta, incluye múltiples desigualdades, y se especifica en términos de los parámetros óptimos poblacionales.¹⁴

Por lo que resta del documento, a la hipótesis 1 se le denominará como la hipótesis de interés.

4 Probando Convergencia

4.1 Simplificación de la Hipótesis de Interés

La hipótesis de interés es un objeto complejo. Es una hipótesis compuesta que incluye múltiples desigualdades. Existen dos enfoques para manejarla. El primero es hacer una prueba simultánea. Este enfoque tiene la ventaja de que puede limitar el nivel de la prueba, pero en general es complicado (Gourieroux y Monfort (1995)).¹⁵ Adicionalmente, cuando se rechaza la nula, uno no puede distinguir cual horizonte es el que está originando el problema.

El segundo enfoque se conoce como comparación múltiple. Este enfoque es menos complicado y también limita el nivel de la prueba (aunque las regiones de rechazo son diferentes, véase Goldberger (1992)). La ventaja es que permite una detección del horizonte o los horizontes que causan el problema. Este es el enfoque que se sigue en este documento para probar la hipótesis de interés.

Se aplica el Principio de Unión-Intersección (PUI) para implementar una prueba de comparación múltiple (Roy (1953); Sen y Silvapulle (2002)).¹⁶ La hipótesis de interés puede ser

¹⁴La propiedad existe sólo para el parámetro poblacional pero no para el estimado (i.e., bajo incertidumbre de los parámetros), por lo tanto, uno sólo puede probar si la propiedad se mantiene en la población.

¹⁵Wolak (1987) propone pruebas simultaneas que pueden ser utilizadas en bajo marco. Las pruebas son pruebas de Wald con distribuciones complejas que son un promedio ponderado de distribuciones χ^2 , y que tienen que se calculadas caso a caso.

¹⁶Cuando la hipótesis nula puede ser expresada como una intersección, $H_0 : \theta \in_{i \in I} \Theta_\gamma$, donde I es un índice arbitrario, y existen pruebas para cada una de las hipótesis componentes: $H_{0,i} : \theta \in \Theta_i$ vs $H_{a,i} : \theta \in \Theta_i^c$, el PUI puede utilizarse para encontrar un procedimiento de prueba adecuado para H_0 . De acuerdo al PUI, si la región de rechazo de la i -ésima prueba es $\{x : T_i(x) \in R_i\}$, entonces la región de rechazo para H_0 es

escrita como la intersección de $H - 1$ hipótesis:

$$H_0 : E [L (e_{\tau, \tau-1}^*)] \leq E [L (e_{\tau, \tau-2}^*)] \cap \dots \cap E [L (e_{\tau, \tau-(H-1)}^*)] \leq E [L (e_{\tau, \tau-H}^*)], \quad (2)$$

por lo que el PUI puede ser aplicado para separarla en $H - 1$ hipótesis componentes, donde la i -ésima hipótesis es:

$$\begin{aligned} H_{0,i} & : E [L (e_{\tau, \tau-i}^*)] \leq E [L (e_{\tau, \tau-(i+1)}^*)] \\ H_{a,i} & : E [L (e_{\tau, \tau-i}^*)] > E [L (e_{\tau, \tau-(i+1)}^*)] \end{aligned}$$

4.2 Límites para el Tamaño de la Prueba Global

Cuando las pruebas están disponibles para cada una de las hipótesis componentes, el problema es controlar el nivel del procedimiento de prueba para la hipótesis de interés. Asíumase por ahora que dichas pruebas están disponibles. Las pruebas se construirán más adelante en el documento.

De acuerdo con el PUI, cada una de las hipótesis componentes tiene que ser probada, y la hipótesis de interés es rechazada cuando se rechaza al menos una de las hipótesis componentes. Para controlar el nivel global, cada una de las pruebas componentes debe hacerse con un nivel de significancia menor al nivel deseado para la hipótesis de interés.

Si las pruebas para las hipótesis componentes fueran independientes, entonces, para tener una prueba con nivel α , el nivel apropiado para cada prueba debería ser $\alpha_i = \left[1 - (1 - \alpha)^{\frac{1}{(H-1)}}\right]$. Sin embargo, es probable que las pruebas estén altamente correlacionadas debido a la propiedad $MA(h-1)$ de los errores (generalizados) de pronóstico. Las Cotas de Bonferroni proveen una manera de manejar esta dependencia. Los niveles de significancia individual bajo estos límites son $\alpha_i = \frac{\alpha}{(H-1)}$, para una prueba global con nivel α .

Sin embargo, es conocido que las Cotas de Bonferroni son muy conservadoras, especialmente cuando las pruebas son altamente dependientes. Para encargarse de esta cuestión, Simes (1986) propuso un procedimiento basado en una modificación de estos límites. La modificación de Simes es simple: Sea $pv_{(1)}, \dots, pv_{(H-1)}$ los valores-p ordenados para probar la hipótesis componente $H_{0,(1)}, \dots, H_{0,(H-1)}$, con $pv_{(1)}$ el valor más grande. Entonces la hipótesis de interés es rechazada si $pv_{(i)} \leq \frac{i\alpha}{(H+1)}$ para cualquier $i = 1, \dots, (H-1)$. Simes (1986) muestra que este procedimiento tiene nivel α para pruebas independientes y un nivel de significancia mucho más cercano al nivel nominal que el procedimiento Bonferroni clásico en

$i \in I \{x : T_i(x) \in R_i\}$. H_0 es rechazada si al menos una de las componentes nulas $H_{0,i}$ es rechazada.

el caso dependiente.¹⁷ Simes también muestra que el procedimiento modificado tiene mayor poder que las Cotas de Bonferroni cuando las pruebas están altamente correlacionadas.

El uso del PUI y la modificación de Simes a las cotas de Bonferroni sugieren el siguiente procedimiento de prueba para las hipótesis componentes.

1. Calcule los $(H - 1)$ valores-p de las hipótesis componentes.
2. Tome el valor-p más grande y compárelo con $\alpha_1 = \frac{\alpha}{(H-1)}$. Si el valor-p es menor que α_1 deténgase. Rechace la hipótesis de interés. Si no es menor, proceda,
3. Tome el segundo valor-p más grande y compárelo con $\alpha_2 = \frac{2\alpha}{(H-1)}$. Si el valor-p es menor que α_2 deténgase. Rechace la hipótesis de interés. Si no es menor, proceda,
4. El procedimiento continua hasta que el valor-p más pequeño es comparado con $\alpha_{(h-1)} = \frac{(H-1)\alpha}{(H-1)} = \alpha$. Si el valor-p es menor entonces la hipótesis de interés es rechazada. Si no, se concluye que no hay suficiente evidencia en la muestra para rechazar la hipótesis de interés.

El procedimiento es un poco complicado, pero hay algunos posibles atajos:

1. Si la pérdida esperada (muestral) es creciente con el horizonte de pronóstico para un pronosticador dado, entonces la conclusión inmediata para este pronosticador es que no hay suficiente evidencia para rechazar la hipótesis nula de interés.¹⁸ Por lo tanto, el procedimiento de Simes puede ser aplicado sólo a pronosticadores no-monotónicos. Esto debe generar los mismos resultados pero, posiblemente, con un incremento en el poder de la prueba.
2. Si H es pequeña, generando un número pequeño de hipótesis componentes, entonces es conocido que las Cotas de Bonferroni están cercanas a los niveles individuales de significancia que uno utilizaría bajo independencia (Miller (1980)). Por lo tanto, para H pequeña (de acuerdo a Miller $H < 6$ es lo suficientemente pequeña) uno puede utilizar sin problemas las clásicas Cotas de Bonferroni.¹⁹

Dada una prueba para cada una de las hipótesis componentes, los valores-p pueden ser calculados y la hipótesis de interés puede ser probada. Inclusive, al utilizar el procedimiento de Simes el tamaño global de la prueba puede ser acotado. La siguiente sección desarrolla pruebas que pueden ser aplicadas a cada una de la hipótesis componentes.

¹⁷Para un análisis bajo dependencia véase Sarkar y Chang (1997).

¹⁸En este caso todos los valores-p serán $\gg \alpha$.

¹⁹Por ejemplo, bajo independencia, para $H = 4$ y $\alpha = 0.10$, $\alpha_i = 0.0345$, en tanto que utilizando las cotas de Bonferroni, $\alpha_i = 0.0333$ genera un $\alpha \leq 0.10$.

5 Pruebas para cada una de las Hipótesis Componentes

Se necesita definir la estructura de probabilidad del problema para construir pruebas para cada una de las hipótesis componentes. Utilizando esa estructura, se proponen dos pruebas basadas en resultados asintóticos. También, se discute una prueba no-paramétrica. Finalmente, se examinan las propiedades de las pruebas para muestras finitas con un experimento tipo Monte Carlo.

5.1 Descripción del Entorno

Está disponible una matriz $P \times H$ de pronósticos de eventos-fijos, donde P es el número de pronósticos y H el número de horizontes de pronóstico disponible. Si $f_{t,t-h}$ es el pronóstico de y_t hecho en $t - h$, la matriz contiene datos $\{f_{t,t-1}, \dots, f_{t,t-H}\}_{t=1}^P$. Los pronósticos pueden estar basados en estimadores de regresiones $\hat{\beta}_{h,R}$, como se presenta en West (1996) y White (2000), en cuyo caso se hacen los siguientes supuestos: se utiliza una estimación recursiva, hay un total de $T + 1$ observaciones disponibles, la primera muestra que se emplea es de tamaño $R - h + 1$ para el primer pronóstico h -pasos hacia delante por lo que $T = R + P - 1$ es el último periodo para el que se realiza un pronóstico un-paso hacia delante, y los datos observados son generados por una sucesión $(\alpha-)$ mixta $\{X_i\}$ fuertemente estacionaria.²⁰

Siguiendo a West (1996) y a White (2000), el objeto de interés es el vector de momentos de $H \times 1$, $E(v_t^*)$ donde $v_t^* \equiv L(y_t - f_{t+h,t}^*) \equiv g(X_t, \beta^*)$ es un vector de $H \times 1$, X_t es un vector aleatorio con distribuciones marginales iguales a aquellas de $\{X_i\}$, y bajo la hipótesis nula de interés, $p \lim \hat{\beta}_R \equiv \beta^*$.

Dada la función de pérdida, los valores actuales, y los pronósticos, uno puede crear un vector de $H \times 1$, $v_t(\hat{\beta}_n) = [L(\hat{e}_{t,t-1}), \dots, (\hat{e}_{t,t-H})]'$, donde se hace explícita la dependencia de los estimadores, y puede calcular el estadístico de $H \times 1$:

$$\bar{v} \equiv P^{-1} \sum_{t=1}^P v_t(\hat{\beta}_i).$$

El Teorema 4.1 en West (1996) genera condiciones de regularidad en los momentos, dependencia y heterogeneidad de la sucesión de vectores $\left\{v_t(\hat{\beta}_n)\right\}_{t=1}^P$, tal que un TCL existe:

$$P^{\frac{1}{2}}(\bar{v} - E(v^*)) \xrightarrow{d} N(0, S). \quad (3)$$

²⁰El tiempo en que sea realizan los pronósticos es un poco diferente al caso usual para un-paso hacia delante. Esto se debe a que en el caso usual se emplean pronósticos de eventos-móviles.

La matriz $H \times H$

$$S \equiv \lim_{P \rightarrow \infty} \text{var} \left[P^{-\frac{1}{2}} \sum_{t=1}^P v_t^* \right] \quad (4)$$

es la densidad espectral de $(\bar{v} - E(v^*))$ en la frecuencia cero escalada por 2π , siempre que $K \equiv E \left[\frac{\partial v_t^*}{\partial \beta} \right] = 0$ ó $P/R \rightarrow 0$ conforme $T \rightarrow \infty$. Si ninguna de estas condiciones se mantiene, la expresión de la varianza es más compleja (West (1996)). K será cero siempre que la función de pérdida utilizada por el productor de los pronósticos sea la misma que la pérdida utilizada para la evaluación, y que los pronósticos sean óptimos (nótese que la última parte es cierta bajo la nula). La segunda condición requiere que el número de observaciones dentro-de-la-muestra crezca más rápido que el número de observaciones fuera-de-la-muestra.²¹

Debido a que es probable que exista autocorrelación, especialmente cuando H es grande, y porque en general no se tiene más información al respecto, se debe ser cuidadoso al estimar S .²²

El TCL de West sólo aplica a funciones de pérdida que son dos veces diferenciables, a modelos que son lineales en los parámetros, y a los modelos (utilizados para producir los pronósticos para diferentes horizontes) que no estén anidados.²³ Por lo tanto, aún cuando el marco presentado en la sección 3 permite modelos más complejos, el TCL disponible restringe la aplicación de los resultados.

5.2 Pruebas para las Hipótesis Componentes

Para elaborar una prueba para cada una de las hipótesis componentes es suficiente con construir una prueba para la primera componente, aquella que involucra pronósticos uno y dos-pasos hacia delante:

$$\begin{aligned} H_{0,1} : E [L(e_{t,t-1}^*)] &\leq E [L(e_{t,t-2}^*)] \\ H_{a,1} : E [L(e_{t,t-1}^*)] &> E [L(e_{t,t-2}^*)] . \end{aligned}$$

Las pruebas para las otras componentes surgen como una generalización obvia.

²¹Para una teoría asintótica alternativa, que utilice diferentes supuestos, véase Giacomini y White (2006).

²²Diebold y Mariano (1995) proponen utilizar la información de que un pronóstico óptimo h -pasos hacia delante tiene la misma autocorrelación que (a lo más) un $MA(h-1)$. Sin embargo, esto es engañoso, ya que esta autocorrelación se refiere a la propiedad *smd* de la *derivada* de la función de costos, no de la función de costos en sí.

²³McCracken (2000) ha extendido los resultados de West para casos donde la función de costos no es diferenciable, pero $E(v_t^*)$ sí. Esto permite el uso del costo EMA y, en general, los costos Lin-lin.

Utilizando el TCL en 3, es posible construir dos procedimientos de prueba, uno similar a la prueba de Diebold y Mariano (1995) y otro inspirado en una prueba de razón de varianzas.

5.2.1 Una prueba basada en la diferencia de las pérdidas esperadas

Las hipótesis se pueden plantear como:

$$\begin{aligned} H_{0,1} : E [L (e_{t,t-1}^*) - L (e_{t,t-2}^*)] &\leq 0 \\ H_{a,1} : E [L (e_{t,t-1}^*) - L (e_{t,t-2}^*)] &> 0, \end{aligned}$$

y un estadístico de prueba se puede basar en el promedio:

$$\bar{d}_{P,1} \equiv \frac{1}{P} \sum_{t=1}^P d_{t,1} = \frac{1}{P} \sum_{t=1}^P [L (\hat{e}_{t+1,t}) - L (\hat{e}_{t+2,t})].$$

$H_{0,1}$ se rechaza para valores grandes de $\bar{d}_{P,1}$. La siguiente proposición declara que las propiedades asintóticas de la prueba propuesta son estándar.

Proposición 7 *Si $\hat{S}$ es un estimador consistente de S (una matriz de 2×2 en este caso) y el TLC en 3 se mantiene, entonces, utilizando el elemento en la nula que es menos favorable para la alternativa,*

$$T_1 = \frac{P^{\frac{1}{2}} \bar{d}_{p,1}}{\widehat{se_d}} \stackrel{a}{\sim} N(0, 1),$$

donde $\widehat{se_d} = \left(G' \hat{S} G\right)^{\frac{1}{2}}$ y G es una matriz de 2×1 con 1 en la posición (1,1), -1 en la posición (2,1) y ceros en el resto de la posiciones.

El valor-p es $\Pr(Z \geq T_1)$, donde Z se distribuye de manera normal estándar.

5.2.2 Una prueba basada en la razón de las pérdidas esperadas

Alternativamente, la hipótesis se puede plantear como:

$$\begin{aligned} H_{0,1} : \frac{E [L (e_{t,t-1}^*)]}{E [L (e_{t,t-2}^*)]} &\leq 1 \\ H_{a,1} : \frac{E [L (e_{t,t-1}^*)]}{E [L (e_{t,t-2}^*)]} &> 1, \end{aligned}$$

mientras que $E [L (e_{t,t-2}^*)] \neq 0$, y un estadístico de prueba pueda basarse en la razón de las medias muestrales

$$\bar{R}_{P,1} = \frac{\frac{1}{P} \sum_{t=1}^P L(\hat{e}_{t,t-1})}{\frac{1}{P} \sum_{t=1}^P L(\hat{e}_{t,t-2})} = \frac{\bar{v}_1}{\bar{v}_2}.$$

Como $\bar{R}_{P,1}$ tiende a ser grande cuando $E [L (e_{t,t-1}^*)] > E [L (e_{t,t-1}^*)]$, $H_{0,1}$ es rechazada para valores grandes de $\bar{R}_{T,1}$. Mientras que para su distribución:

Proposición 8 *Si $\hat{S}$ es un estimador consistente de S (una matriz de 2×2 en este caso) y el TLC en 3 se mantiene entonces, utilizando el elemento en la nula que es menos favorable para la alternativa,*

$$TR_1 = \frac{P^{\frac{1}{2}} (\bar{R}_{P,1} - 1)}{\widehat{se_R}} \underset{a}{\sim} N(0, 1),$$

donde $\widehat{se_R} = \left(G' \hat{S} G\right)^{\frac{1}{2}}$ y G es una matriz de 2×1 con $\frac{1}{\bar{v}_2}$ en la posición $(1, 1)$, $-\frac{\bar{v}_1}{(\bar{v}_2)^2}$ en la posición $(2, 1)$ y ceros en el resto de la posiciones.

El valor-p es $\Pr(Z \geq TR_1)$, donde Z se distribuye de manera normal estándar.

5.2.3 Una prueba no-paramétrica basada en la diferencia de las pérdidas esperadas

Pruebas no paramétricas han sido utilizadas en la literatura de pronósticos debido a la robustez que presentan ante desviaciones de ciertos supuestos como el de normalidad y también porque usualmente son más confiables para muestras pequeñas, típicamente disponibles en ejercicios de pronósticos (particularmente en macroeconomía), aunque es bien sabido que pueden ser muy sensibles a datos atípicos. Para pruebas relacionadas utilizadas en la evaluación de pronósticos véase Campbell y Ghysels (1995) y Diebold y Lopez (1996).

Cuando se está interesado en probar si una variable aleatoria en un par (X, Y) tiende a ser mayor que la otra variable, típicamente se utiliza la prueba del signo. Siguiendo a Conover (1980), es necesario que se cumplan tres supuestos para que la prueba del signo sea una prueba confiable: (i) las variables aleatorias (X_i, Y_i) , $i = 1, \dots, n$ deben ser mutuamente independientes, (ii) la medida de escala debe ser por lo menos ordinal, y (iii) que los pares (X_i, Y_i) sean internamente consistentes, en cuanto a que si $\Pr(X_i < Y_i) > \Pr(Y_i < X_i)$ es cierto para un i , entonces también será cierto $\forall i$. Bajo estos requerimientos, la prueba del signo puede utilizarse para probar:

$$\begin{aligned} H_0 &: E[X_i] \leq E[Y_i] \\ H_a &: E[X_i] > E[Y_i], \end{aligned}$$

usando como el estadístico de prueba el número total de veces que $Y_i < X_i$. Al sustituir $L(e_{t,t-1}^*)$ por X_i , y $L(e_{t,t-2}^*)$ por Y_i , la prueba del signo puede visualizarse como una prueba equivalente a la prueba de diferencias. En este caso, el estadístico de prueba puede basarse en:

$$sn_{P,1} \equiv \sum_{t=1}^P \mathbf{1}_{(L(e_{t,t-1}^*) > L(e_{t,t-2}^*))}, \text{ para } t = 1, \dots, P, \text{ donde } \mathbf{1} \text{ es la función indicadora.}$$

H_0^1 se rechaza para valores grandes de $sn_{P,1}$. Siguiendo a Conover (1980), bajo los tres supuestos planteados anteriormente, $sn_{P,1}$ se distribuye como una binomial con parámetros $\rho = \frac{1}{2}$ y $n = P$, y para valores grandes de P , ($P > 20$), $TS_1 = \frac{sn_{P,1} - \frac{P}{2}}{\frac{\sqrt{P}}{2}} \stackrel{a}{\sim} N(0, 1)$.

$$TS_1 = \frac{sn_{P,1} - \frac{P}{2}}{\frac{\sqrt{P}}{2}} \stackrel{a}{\sim} N(0, 1).$$

El valor-p puede calcularse como $p\text{-value}_1 = \Pr(Z \geq TS_1)$, donde Z se distribuye como una normal estándar.

De los tres supuestos requeridos para las pruebas no-paramétricas, el primero claramente no será satisfecho por las pérdidas esperadas, ya que se espera que exista alta autocorrelación. Sin embargo, la prueba del signo ha probado ser robusta a desviaciones de este supuesto (Diebold y Mariano (1995)), y de hecho puede ser útil bajo estas condiciones. La prueba del signo se considera aquí como una prueba alternativa debido a su sencillez.

5.2.4 Comentarios sobre las pruebas

En todas las pruebas se impone la nula al utilizar el elemento que es menos favorable para la alternativa. El problema con este enfoque es que cuando el momento poblacional está lejos de la igualdad, las pruebas propuestas serán más pequeñas que lo usual. Este es un efecto usual cuando hay una desigualdad presente bajo la hipótesis nula. Mientras más lejos está el momento poblacional de la igualdad, más grande será el problema del nivel.

Para implementar las dos primeras pruebas, es necesario obtener un estimador consistente de S . En la literatura existen varios estimadores consistentes que pueden ser aplicados. Por ejemplo, los no-paramétricos discutidos en Andrews (1991) o los paramétricos de Den Haan y Levin (1997). El problema con todos ellos es que aún cuando son consistentes, pueden tener un funcionamiento pobre en muestras finitas. Esto tiene como consecuencia que las pruebas serán más grandes que lo usual cuando se trabaja con muestras finitas.

Hay tres razones, *a priori*, por las cuales la primera prueba aparenta ser una mejor opción. La primera es que está basada en una hipótesis lineal, mientras que la segunda se basa en

una no-lineal, la segunda es que toma en cuenta la estructura de autocorrelación, mientras que la tercera no lo hace, y, finalmente, está relacionada con la prueba de Diebold-Mariano (1995), la cual es muy conocida dentro de la literatura de pronósticos.

5.3 Desempeño en Muestras Finitas

5.3.1 Diseño del experimento

Se realizaron varios experimentos Monte Carlo para evaluar el tamaño en muestras finitas de los estadísticos de prueba T_h , TR_h , y TS_h . Para todos los experimentos el proceso generador de datos (PGD) bajo la nula se tomó como un $AR(1)$ univariado y la función de pérdida, tanto para estimar el modelo como para evaluar los pronósticos, fue elegida con la forma ECM.²⁴ El estimador utilizado para estimar la matriz de covarianza de largo plazo cuando ésta se requiere es el VARHAC de Den Haan y Levin (1997).²⁵ Todas las pruebas se realizaron a un nivel de significancia $\alpha = 0.10$. Cada experimento se repitió 5,000 veces.

La primera división entre los experimentos se hace entre aquéllos donde el parámetro del $AR(1)$ se tomó como conocido y aquéllos donde tuvo que ser estimado. En el último caso, se realizó una estimación recursiva con MCO.

Para cada una de estas dos partes, se hicieron dos supuestos diferentes sobre la parte estocástica del PGD. El primer supuesto es que el error sigue una distribución normal estándar, este es el caso de los errores Gaussianos. El otro supuesto (siguiendo a Diebold y Mariano (1995)) es que el error sigue una distribución con colas gruesas. En este caso, la distribución que se usa para generar los errores es una distribución t con seis grados de libertad. Los errores fueron estandarizados dividiéndolos entre $\sqrt{\frac{3}{2}}$.

Los cuatro experimentos se muestran en los Cuadros 2 a 5. Un Cuadro por cada experimento.

En los Cuadros 2 y 3, cuando el parámetro del $AR(1)$ es conocido, se consideran tres números diferentes de pronósticos: $P = 50, 100, 500$, y tres valores para el parámetro del $AR(1)$: $\phi = 0.2, 0.5, 0.8$. Se presentan los resultados para las hipótesis componentes uno, tres y siete para las tres pruebas.

En los Cuadros 4 y 5, donde se tiene incertidumbre sobre los parámetros, se consideran tres números diferentes de pronósticos: $P = 50, 100, 500$, y tres casos para R , el tamaño de la primera muestra utilizada en la estimación: $R = 50, 100, 500$. Sólo se reportan las combinaciones relevantes (e.g., la combinación donde la muestra inicial de 50 se usa para

²⁴La condición inicial del $AR(1)$ es cero. Las primeras 500 observaciones no se tomaron en cuenta en cada repetición para evitar la dependencia de los resultados de la condición inicial.

²⁵Para el VARHAC, se utilizó el criterio de selección BIC para escoger los rezagos del VAR. El rezago máximo se determinó en la parte entera de $P^{\frac{1}{3}}$, como lo recomiendan Den Haan y Levin.

construir 500 pronósticos no se tomó en cuenta). El PGD en estos casos se tomó como un $AR(1)$ con $\phi = 0.2$. Se presentan los resultados para las hipótesis componentes uno, tres y siete.

5.3.2 Resultados del experimento

El primer resultado que se puede observar en los cuadros es que el nivel de las pruebas es el correcto, ya que el nivel actual es menor o cercano a 10 %.

Sobre los resultados específicos, para los casos donde el PGD es conocido (Cuadro 2 y 3) se pueden observar los siguientes efectos para las dos primeras pruebas:

1. Como se esperaba, el hecho de que la matriz de covarianza de largo plazo tenga que ser estimada provoca que las pruebas tengan un nivel mayor. Este efecto se identifica al observar una columna fija en el Cuadro 2. Cuando se mantiene todo lo demás constante, mientras el número de pronósticos disminuye, el tamaño incrementa.
2. También como se esperaba, las pruebas tienen un nivel menor al usual cuando el momento poblacional se encuentra lejos de la igualdad (en la nula). De hecho, mientras más lejos está el momento poblacional de la igualdad, el nivel de las pruebas es cada vez menor. Este efecto se puede ver en los cuadros de dos formas: Primero, mientras más persistente es el PGD, más pequeño es el nivel; y segundo, mientras más corto es el horizonte, más pequeño es el nivel. En ambos casos esto ocurre porque para un $AR(1)$ la diferencia en la varianza de los errores de pronóstico entre los horizontes h y $h + 1$ es: $var(h) - var(h + 1) = \frac{\phi^{2(h+1)} - \phi^{2h}}{1 - \phi^2} \sigma_\varepsilon^2$, y su magnitud se incrementa con ϕ y disminuye con h (un resultado similar se mantiene para la razón).

Los resultados muestran que el desempeño de las primeras dos pruebas es muy similar. El tamaño observado es un poco mejor cuando el PGD es Gaussiano, pero las pruebas muestran un buen desempeño cuando se aplican al PGD no-Gaussiano. Las pruebas no-paramétricas se desempeñan muy bien, mostrando que pueden ser utilizadas como una primera aproximación.

Para los casos en los que el parámetros del PGD tiene que ser estimado (Cuadros 4 y 5) el desempeño es muy similar al caso anterior. En particular, conforme el tamaño de la muestra utilizada en la estimación, R , aumenta, el tamaño se aproxima al tamaño de las pruebas cuando se conoce el parámetro. Esta conclusión muestra que, aún cuando existe incertidumbre sobre los parámetros, la propiedad de precisión decreciente puede ser probada utilizando una de las pruebas desarrolladas en este documento.

6 Un Ejemplo Ilustrativo

Cuando se quiere determinar si un conjunto de pronósticos de horizontes múltiples es óptimo, puede probarse la propiedad de precisión decreciente. Adicionalmente, ahora están disponibles un procedimiento de prueba y pruebas que pueden ser utilizadas para cada una de las hipótesis componentes. El siguiente paso es ilustrar la aplicación del procedimiento de prueba. Esto se hace aplicando las pruebas para contestar las preguntas generadas por el Cuadro 1 concerniente al desempeño de una muestra de pronosticadores de la Encuesta de Pronosticadores Profesionales (SPF por sus siglas en inglés).

La SPF es una encuesta trimestral conducida por el Banco de la Reserva Federal de Filadelfia. Esta encuesta se hace a economistas privados que producen pronósticos de variables económicas regularmente como parte de su trabajo. Los pronosticadores provienen de Wall Street, bancos privados, empresas consultoras, y centros de investigación, entre otros. Las variables pronosticadas incluyen producción y sus componentes, inflación y tasas de interés. Entre los pronósticos proporcionados, se hacen pronósticos trimestrales de uno a cuatro pasos hacia delante para la mayoría de las variables, donde en algunos casos las primeras predicciones datan de 1968. La identidad de los pronosticadores se mantiene en secreto, pero a cada pronosticador se le asocia un número.²⁶

Para la aplicación se toma una muestra de veinte pronosticadores de la encuesta. Los pronósticos se hacen para el nivel de producto nominal (PNB antes de 1992, PIB a partir de entonces) de uno a cinco pasos hacia delante. La muestra entre los pronosticadores de la encuesta se tomó de manera aleatoria, obteniendo al menos 10 observaciones por trimestre, para cada horizonte de pronóstico. No todos los pronosticadores están prediciendo los mismos periodos, y el número de observaciones por pronosticador es diferente entre pronosticadores, lo que hace que una comparación entre pronosticadores sea imposible. Sin embargo, no representa un problema para este ejercicio, ya que la prueba desarrollada en este documento concierne a la evaluación de un pronosticador a la vez, no a una comparación entre pronosticadores. Para el análisis los pronósticos se transforman de niveles a tasas de crecimiento (por lo tanto, sólo están disponible pronósticos de 1 a 4 pasos hacia delante).

Los valores observados se toman de la Oficina de Análisis Económico (BCA por sus siglas en inglés). La serie es el PIB trimestral (PNB antes de 1992), ajustado por estacionalidad, en billones de dólares. Las tasas de crecimiento se calcularon de la misma manera que para los pronósticos. Se asume que los pronosticadores intentan pronosticar los valores finales del producto nominal, y no el primer dato publicado.

²⁶Una descripción completa de esta encuesta se puede encontrar en Croushore (1993), o en la página de internet del Banco de la Reserva Federal: www.phil.frb.org/econ/spf.

El Cuadro 1 presenta la RECM de cada horizonte para cada pronosticador, donde se asume una pérdida ECM. Lo primero que se puede observar es que para la mayoría de los pronosticadores la RECM (la desviación estándar si los pronósticos son insesgados) es creciente con el horizonte. Esto es común para la mayoría de los pronosticadores de la encuesta, como se menciona en el exhaustivo estudio de Zarnowitz y Brown (1992): “Los errores de pronóstico generalmente aumentan conforme aumenta el número de periodos en el horizonte de pronóstico”. Pero se puede observar que para algunos pronosticadores esto no se cumple. Como se mencionó anteriormente, para los pronosticadores 1, 3, 96 y 452, en algún punto, la RECM disminuye con el horizonte. En algunos casos aumenta de nuevo. Se puede probar si esto implica que los pronósticos no son óptimos mediante la aplicación de una de las pruebas propuestas en este documento.

El primer paso es calcular los valores-p de cada una de las hipótesis componentes utilizando una de las pruebas propuestas anteriormente. En este ejemplo se usa la prueba paramétrica basada en la diferencia de las pérdidas.

Una forma sencilla de hacer esta prueba es haciendo la regresión de las diferencias muestrales $d_{t,h} = \hat{e}_{t,t-h}^2 - \hat{e}_{t,t-(h+1)}^2$ contra una constante y después hacer una prueba-t de una cola para ver si la constante es negativa o igual a cero. La autocorrelación de los errores en la regresión tiene que ser modelada en cada caso. Para hacer esto, después de que la primera regresión es estimada, se selecciona una estructura ARMA para los residuales mediante la observación de los autocorrelogramas. Después se estima nuevamente la regresión incluyendo el modelo ARMA seleccionado. El valor-p de interés es aquél asociado con la constante. De esta forma se obtienen tres valores-p para cada pronosticador.²⁷

El siguiente paso es ordenar los valores-p para cada pronosticador. Después, de eso, para efectuar una prueba con un nivel de significancia de 0.10, cada uno de los valores-p se compara con $\alpha_j = \frac{j(0.10)}{3}$, $j = 1, 2, 3$ donde $j = 1$ corresponde al valor-p más alto. La hipótesis nula de interés se rechaza para cada pronosticador si al menos uno de sus valores-p es menor que el α_j correspondiente.

Los resultados se muestran en el Cuadro 6. La prueba sólo rechaza la hipótesis nula de precisión decreciente para el pronosticador 452. Adicionalmente, es claro que d_2 es la diferencia que causa el problema, lo que significa que los pronósticos para dos o tres pasos hacia delante, o ambos, no están correctamente especificados o pueden ser mejorados. Por ejemplo, usando los pronósticos del pronosticador 452, un mejor pronóstico dos pasos hacia delante es simplemente, el pronóstico tres pasos hacia delante, pero empezando en $t - 1$ en

²⁷Esta es una forma alternativa para proseguir con la prueba. Tiene la ventaja de que todas las herramientas desarrolladas para la regresión pueden ser aplicadas, por ejemplo, aquí se modela explícitamente la autocorrelación.

lugar de empezar en t .²⁸

Todo lo que se necesita para la aplicación, aparte de los datos sobre los errores de pronóstico, es un supuesto sobre la función de pérdida del productor de los pronósticos. Esto contrasta con otras pruebas de optimalidad de pronósticos, ya que otras pruebas también necesitan un supuesto sobre la información utilizada por el pronosticador. Para este ejemplo se utiliza pérdida ECM. Pero aún cuando la pérdida verdadera no sea ECM, siempre y cuando los sesgos óptimos de pronóstico no varíen con el tiempo, el uso de errores cuadráticos puede dar cierta intuición sobre la optimalidad del pronóstico, ya que su movimiento puede ser proporcional al objeto de interés (proposición (6)).

7 Conclusiones

La propiedad de los pronósticos óptimos de que la pérdida esperada de los errores de pronóstico tiene que ser no-decreciente con el horizonte de pronóstico puede utilizarse para determinar la optimalidad de los pronósticos con horizontes múltiples.

La prueba puede aplicarse siempre que se determine la hipótesis nula de tal manera que relacione momentos poblacionales, y no momentos muestrales. La hipótesis de interés planteada de esta manera termina siendo una hipótesis componente con múltiples desigualdades. El Principio Unión-Intersección se usa para tratar esto. El tamaño global de la prueba está acotado por un procedimiento sugerido por Simes, el cual modifica las cotas de Bonferroni tradicionales de tal manera que las hace menos conservadoras para el caso de pruebas componentes dependientes. Se proveen tres pruebas diferentes para las hipótesis componentes. Todas las pruebas tienen un nivel correcto, pero la prueba basada en la diferencia de las pérdidas esperadas, parecida a la prueba de Diebold y Mariano, aparenta tener un mejor desempeño. Este procedimiento de prueba puede ser aplicado utilizando funciones de pérdida generales y cuando los pronósticos se toman como el producto de modelos estimados, siempre y cuando estos modelos sean lineales y no estén anidados.

Al aplicar la prueba se puede estudiar si un conjunto de pronósticos con horizontes múltiples está utilizando la información eficientemente y, por lo tanto, si se pueden basar decisiones en ellos con un cierto nivel de confianza, aunque esta prueba debe ser complementada con otra prueba de eficiencia, por ejemplo aquellas vistas por Diebold y Lopez (1996). En caso de rechazo, el procedimiento indica donde está el problema, encontrando así cuáles son los horizontes que tienen que ser mejorados.

Las ventajas del procedimiento de prueba desarrollado en este documento son que utiliza directamente la relación entre los pronósticos para diferentes horizontes, y que solamente

²⁸No se pudo encontrar ningún valor atípico utilizando un procedimiento simple.

necesita un supuesto sobre la función de pérdida del pronosticador, pero ninguno sobre el conjunto de información. De hecho, bajo ciertas condiciones se pueden relajar los supuestos acerca de la función de pérdida, ya que la prueba se puede aplicar para comparar la varianza de los errores de pronóstico entre horizontes, siempre y cuando se cumplan las condiciones requeridas.

Una dirección interesante para futura investigación es estudiar el grado de complementariedad entre las pruebas presentadas aquí y otras pruebas de optimalidad (i.e., el poder de diferentes pruebas). También se necesita mayor experiencia aplicando las pruebas.

Referencias

- [1] Andrews, D., 1991, “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817-858.
- [2] Bakhshi, H., G. Kapetanios, y A. Yates, 2005, “Rational Expectations and Fixed-events Forecasts: An Application to UK Inflation,” *Empirical Economics*, 30, 539-553.
- [3] Batchelor, R. y P. Dua, 1991, “Blue Chip Rationality Tests,” *Journal of Money Credit and Banking*, 33, 692-705.
- [4] Box, G.E.P., G.M. Jenkins, y G.C. Reinsel, 1994, *Time Series Analysis, Forecasting and Control*, 3rd edition. New Jersey: Prentice-Hall.
- [5] Campbell, B. y E. Ghysels, 1995, “Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency,” *Review of Economics and Statistics*, 17-31.
- [6] Capistrán, C., 2006, “On Comparing Multi-Horizon Forecasts,” *Economics Letters*, 93, 176-181.
- [7] Chong, Y.Y. y D.F. Hendry, 1986, “Econometric Evaluation of Linear Macroeconomic Models,” *Review of Economic Studies*, 53, 671-690.
- [8] Christoffersen, P. y F.X. Diebold, 1997, “Optimal Prediction under Asymmetric Loss,” *Econometric Theory*, 13, 806-817.
- [9] Clements, M.P. y D.F. Hendry, 1998, *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.
- [10] Conover, W.J., 1980, *Practical Nonparametric Statistics*, 2nd Edition. New York: Wiley & Sons.
- [11] Croushore, D., 1993, “Introducing: the Survey of Professional Forecasters,” *Business Review*, Noviembre/Diciembre, Federal Reserve Bank of Philadelphia, 3-13.
- [12] Den Haan, W.J. y A. Levin, 1997, “A Practitioner’s Guide to Robust Covariance Matrix Estimator,” en G.S. Maddala y C.S. Rao (eds.), *Handbook of Statistics*, 15, pp. 299-342. Amsterdam: North Holland.
- [13] Diebold, F.X., 2001, *Elements of Forecasting*, 2nd Edition. Ohio: South-Western.

- [14] Diebold, F.X. y J.A. Lopez, 1996, "Forecast Evaluation and Combination," en G.S. Maddala y C.S. Rao (eds.), *Handbook of Statistics*, 15, pp. 299-342. Amsterdam: North Holland.
- [15] Diebold, F.X. y R. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.
- [16] Elliott, G. y A. Timmermann, 2004. "Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions," *Journal of Econometrics*, 122, 47-79.
- [17] Elliott, G., I. Komunjer, y A. Timmermann, próxima publicación, "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss," *Journal of the European Economic Association*.
- [18] Giacomini, R., y H. White, 2006, "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545-1578.
- [19] Goldberger, A.S., 1992, "One-sided and Inequality Tests for a Pair of Means," en R. Bewley y T.V. Hoa (eds.), *Contributions to Consumer Demand and Econometrics*, pp. 140-162, New York: St. Martin's Press.
- [20] Gourieroux, C. y A. Monfort, 1995, *Statistics and Econometrics Models*, Volumes 2, (traducido por Quang Voun de la edición francesa de 1989) Cambridge: Cambridge University Press.
- [21] Granger, C.W.J., 1981, "The Comparison of Time Series and Econometric Forecasting Strategies," en J. Kmenta, y J.B. Ramsey (eds.), *Large-Scale Macro-Econometric Models*, pp. 123-128. North Holland Publishing Company.
- [22] Granger, C.W.J. 1969, "Prediction with a Generalized Cost Function" *Operational Research* 20, 199-207, reimpresso en E. Ghysels, N.R. Swanson y M.W. Watson (eds.), *Essays in Econometrics: Collected Papers of Clive W.J. Granger*, Volume I, 2001. Cambridge: Cambridge University Press.
- [23] Granger, C.W.J., 1999, "Outline of Forecast Theory using Generalized Cost Functions," *Spanish Economic Review*, 1, 161-173.
- [24] Granger, C.W.J. y P. Newbold, 1986, *Forecasting Economic Time Series*, 2nd edition. Orlando: Academic Press.
- [25] Granger, C.W.J. y T. Terasvirta, 1993, *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

- [26] Hamilton, J.D., 1994, *Time Series Analysis*. Princeton, N.J.: Princeton University Press.
- [27] Hendry, D.F. y M.P. Clements, 2001, "Economic Forecasting: Some Lessons from Recent Research," *European Central Bank, Working Paper* No. 82.
- [28] McCracken, M.W., 2000, "Robust Out-of-Sample Inference," *Journal of Econometrics*, 99, 195-223.
- [29] Miller, R.G. Jr., 1980, *Simultaneous Statistical Inference*. 2nd edition. New York: Springer-Verlag.
- [30] Mincer, J. y V. Zarnowitz, 1969, "The Evaluation of Economic Forecasts" en K. Mincer (ed.) *Economic Forecasts and Expectations*. New York: National Bureau of Economic Research.
- [31] Nordhaus, W.D., 1987, "Forecasting Efficiency: Concepts and Applications," *The Review of Economics and Statistics*, 69(4), 667-674.
- [32] Patton, A.J. y A. Timmermann, próxima publicación, "Testable Implications of Forecasts Optimality," *Journal of the American Statistical Association*.
- [33] Patton, A.J. y A. Timmermann, 2007, "Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity," *Journal of Econometrics*, 140, 884-918.
- [34] Roy, S.N., 1953, "On a Heuristic Method of Test Construction and its Use in Multivariate Analysis," *Annals of Mathematical Statistics*, 24, 220-238.
- [35] Sarkar, S.K., y C-K. Chang, 1997, "The Simes Method for Multiple Hypothesis Testing with Positively Dependent Test Statistics," *Journal of the American Statistical Association*, 92, 1601-1608.
- [36] Schmidt, P., 1974, "The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model," *Econometrica*, 42, 303-309.
- [37] Sen, P.K. y M.J. Silvapulle, 2002, "An Appraisal of Some Aspects of Statistical Inference under Inequality Constraints," *Journal of Statistical Planning and Inference*, 107, 3-43.
- [38] Simes, R.J., 1986, "An Improved Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 73, 751-754.
- [39] Stekler, H.O., 2002, "The Rationality and Efficiency of Individuals' Forecasts," capítulo 10 en M.P. Clements y D.F. Hendry (eds.), *A Companion to Economic Forecasting*. Oxford: Blackwell Publishers.

- [40] Swidler, S. y D. Ketcher, 1990, "Economic Forecasts, Rationality, and the Processing of New Information over Time," *Journal of Money, Credit, and Banking*, 22, 65-76.
- [41] West, K.D., 1996, "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067-1084.
- [42] White, H., 2000, "A Reality Check for Data Snooping," *Econometrica*, 68, 1097-1126.
- [43] White, H., 2001, *Asymptotic Theory for Econometricians*, edición revisada. San Diego: Academic Press.
- [44] Wolak, F.A., 1987, "An Exact Test for Multiple Inequality Restrictions and Equality Constraints in the Linear Regression Model," *Journal of the American Statistical Association*, 82, 782-793.
- [45] Zarnowitz, V., y P. Braum, 1992, "Twenty-Two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasts Performance," *NBER working paper* No. 3965.
- [46] Zellner, A., 1986, "Biased Predictors, Rationality and the Evaluation of Forecasts," *Economic Letters*, 21, 45-48.

Apéndice. Pruebas

Prueba de la proposición 5. Por optimalidad, $f_{\tau, \tau-h}^* \in \Omega_{\tau-h}$, y $f_{\tau, \tau-h-j}^* \in \Omega_{\tau-h-j} \forall j > 0$. Como $\Omega_{\tau-h}$ es una filtración, $\Omega_{\tau-h-j} \subset \Omega_{\tau-h}$, entonces $f_{\tau, \tau-h-j}^* \in \Omega_{\tau-h}$. Por lo tanto, si la función de pérdida no depende del tiempo,

$$V(\tau, \tau-h, \Omega_{\tau-h}) \leq V(\tau, \tau-h-j, \Omega_{\tau-h-j}) \quad \forall j > 0,$$

y por la Ley de Expectativas Iteradas,

$$E[V(\tau, \tau-h, \Omega_{\tau-h})] \leq E[V(\tau, \tau-h-j, \Omega_{\tau-h-j})] \quad \forall j > 0,$$

$$V(\tau, \tau-h) \leq V(\tau, \tau-h-j) \quad \forall j > 0.$$

■

Prueba de la proposición 6. Siguiendo a Elliot y Timmermann (2004), una expansión de Taylor alrededor de la media de los pronósticos, $\mu_e^{\tau-h} = E[e_{\tau, \tau-h}^*]$, da:

$$\begin{aligned} L(e_{\tau, \tau-h}^*) &= L(\mu_e^{\tau-h}) + L'(\mu_e^{\tau-h})(e_{\tau, \tau-h}^* - \mu_e^{\tau-h}) \\ &\quad + \frac{1}{2}L''(\mu_e^{\tau-h})(e_{\tau, \tau-h}^* - \mu_e^{\tau-h})^2 + \sum_{k=3}^{\infty} \frac{1}{k!}L^{(k)}(\mu_e^{\tau-h})(e_{\tau, \tau-h}^* - \mu_e^{\tau-h})^k, \end{aligned}$$

tomando la media,

$$\begin{aligned} E[L(e_{\tau, \tau-h}^*)] &= L(\mu_e^{\tau-h}) + \frac{1}{2}L''(\mu_e^{\tau-h})E[(e_{\tau, \tau-h}^* - \mu_e^{\tau-h})^2] \\ &\quad + \sum_{k=3}^{\infty} \frac{1}{k!}L^{(k)}(\mu_e^{\tau-h})E[(e_{\tau, \tau-h}^* - \mu_e^{\tau-h})^k]. \end{aligned}$$

Cuando se cumple una de las dos condiciones indicadas en la proposición, el resultado se colapsa a:

$$E[L(e_{\tau, \tau-h}^*)] = L(\mu_e^{\tau-h}) + \frac{1}{2}L''(\mu_e^{\tau-h})E[(e_{\tau, \tau-h}^* - \mu_e^{\tau-h})^2],$$

en este caso, condicionado en que la función de pérdida sea estable en el tiempo,

$$\begin{aligned} E[L(e_{\tau, \tau-h}^*)] - E[L(e_{\tau, \tau-h-j}^*)] &= L(\mu_e^{\tau-h}) - L(\mu_e^{\tau-h-j}) + \\ &\quad \frac{1}{2}L''(\mu_e^{\tau-h})E[(e_{\tau, \tau-h}^* - \mu_e^{\tau-h})^2] - \\ &\quad \frac{1}{2}L''(\mu_e^{\tau-h-j})E[(e_{\tau, \tau-h-j}^* - \mu_e^{\tau-h-j})^2], \end{aligned}$$

adicionalmente, si $\mu_e^{\tau-h} = \mu_e^{\tau-h-j}$,

$$E [L (e_{\tau,\tau-h}^*)] - E [L (e_{\tau,\tau-h-j}^*)] = \frac{1}{2} L''(\mu_e^{\tau-h}) \left[E (e_{\tau,\tau-h}^* - \mu_e^{\tau-h})^2 - E (e_{\tau,\tau-h-j}^* - \mu_e^{\tau-h-j})^2 \right],$$

y sigue el resultado de la proposición. ■

Prueba de la proposición 7. El resultado se genera por la aplicación de la regla Delta y del teorema de Slutsky. ■

Prueba de la proposición 8. El resultado se genera por la aplicación de la regla Delta y del teorema de Slutsky. ■

Cuadro 1: Raíz del Error Cuadrático Medio para diferentes pronosticadores y horizontes.

Pronosticador	T	RECM ₁	RECM ₂	RECM ₃	RECM ₄
1	16	1.04	1.41	1.39	1.62
3	17	1.24	1.88	2.14	2.02
7	38	1.31	1.60	1.88	2.19
15	49	0.76	1.16	1.33	1.56
20	58	0.78	1.17	1.80	2.35
54	35	1.18	1.84	2.18	2.25
60	59	0.74	1.18	1.59	1.96
62	45	0.98	1.69	2.33	2.96
86	56	1.05	1.71	2.12	3.06
96	21	1.14	1.55	5.27	3.37
102	25	1.54	2.13	2.55	2.68
125	36	1.87	3.21	4.45	5.47
144	32	1.11	1.53	1.77	1.88
147	15	1.69	2.88	3.77	4.20
170	10	1.12	1.55	2.22	4.09
404	30	0.44	0.66	0.88	1.14
411	33	0.71	0.85	1.14	1.29
414	29	0.45	0.65	0.89	0.96
452	14	0.83	1.28	1.00	1.00
483	11	0.60	1.13	1.64	1.89

Notas: Muestra de 20 pronosticadores de la Encuesta de Pronosticadores Profesionales. Los pronósticos originales son para producción nominal entre 1968 y 2001. Se muestran las RECM de los pronósticos de 1 a 4 pasos hacia delante para la tasa de crecimiento anualizada. T es el número de pronósticos para cada horizonte.

Cuadro 2: % de rechazos de la nula verdadera cuando el $AR(1)$ es generado utilizando errores Gaussianos y los valores críticos son comparados con una $N(0, 1)$.

P	ϕ	d_1	d_3	d_7	r_1	r_3	r_7	s_1	s_3	s_7
50	0.20	2.36	9.78	10.44	2.52	9.52	10.86	4.20	10.00	9.94
50	0.50	0.06	4.76	8.00	0.08	4.60	7.84	0.70	6.84	9.76
50	0.80	0.00	0.26	1.00	0.00	0.26	0.94	0.08	1.62	5.54
100	0.20	1.18	9.40	10.38	1.14	9.10	10.36	2.50	9.44	9.86
100	0.50	0.00	2.56	6.70	0.00	2.44	6.60	0.12	5.00	9.18
100	0.80	0.00	0.00	0.18	0.00	0.02	0.14	0.00	0.46	4.02
500	0.20	0.00	9.16	9.76	0.00	9.16	9.74	0.18	9.28	10.12
500	0.50	0.00	0.14	3.66	0.00	0.12	3.66	0.00	1.84	9.04
500	0.80	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	1.06

Notas: Las pruebas son al 10 % de significancia. Se realizan 5,000 replicaciones Monte Carlo. P es el número de pronósticos disponible para cada horizonte. ϕ es el parámetro del $AR(1)$. d_h corresponde a la prueba de diferencias para la diferencia entre el horizonte h y $h + 1$. r_h corresponde a la prueba de la razón para la razón entre h y $h + 1$. s_h corresponde a la prueba de signo para la diferencia entre h y $h + 1$.

Cuadro 3: % de rechazos de la nula verdadera cuando el $AR(1)$ es generado utilizando errores con colas-gruesas y los valores críticos son comparados con una $N(0, 1)$.

P	ϕ	d_1	d_3	d_7	r_1	r_3	r_7	s_1	s_3	s_7
50	0.20	2.30	10.78	9.86	2.50	10.24	9.68	3.76	10.24	9.38
50	0.50	0.08	4.12	7.60	0.08	4.24	7.20	0.46	6.20	9.92
50	0.80	0.00	0.18	0.80	0.00	0.14	0.68	0.04	1.12	5.82
100	0.20	1.28	9.02	10.34	1.46	8.78	10.40	2.30	9.44	9.46
100	0.50	0.00	2.64	6.28	0.00	2.70	6.40	0.06	5.10	8.26
100	0.80	0.00	0.00	0.12	0.00	0.00	0.08	0.00	0.48	4.62
500	0.20	0.04	8.12	10.22	0.04	8.14	10.12	0.20	8.88	9.88
500	0.50	0.00	0.18	3.68	0.00	0.18	3.50	0.00	1.62	9.24
500	0.80	0.00	0.00	0.08	0.00	0.00	0.08	0.00	0.00	1.24

Notas: Véase el Cuadro 2.

Cuadro 4: % de rechazos de la nula verdadera cuando el $AR(1)$ es generado utilizando errores Gaussianos, el parámetro $(0, 2)$ es estimado y los valores críticos son comparados con una $N(0, 1)$.

P	R	d ₁	d ₃	d ₇	r ₁	r ₃	r ₇	s ₁	s ₃	s ₇
50	50	5.08	11.86	12.22	5.16	11.30	10.80	4.38	9.70	9.96
50	100	3.42	10.88	11.72	3.36	10.48	11.02	4.12	9.58	10.48
50	500	2.42	10.26	11.12	2.40	10.22	10.78	3.98	9.54	9.94
100	100	1.48	10.04	11.46	1.50	9.92	10.74	2.42	9.64	10.30
100	500	1.26	9.46	10.12	1.16	9.52	10.26	2.16	9.10	9.56
500	500	0.00	8.86	9.34	0.00	8.82	9.42	0.20	9.50	9.88

Notas: Véase el Cuadro 2. R es el número de observaciones de la primer muestra utilizada para estimar el $AR(1)$.

Cuadro 5: % de rechazos de la nula verdadera cuando el $AR(1)$ es generado utilizando errores con colas-gruesas, el parámetro $(0,2)$ es estimado y los valores críticos son comparados con una $N(0, 1)$.

P	R	d ₁	d ₃	d ₇	r ₁	r ₃	r ₇	s ₁	s ₃	s ₇
50	50	4.62	12.22	13.12	4.90	10.98	11.02	4.20	10.38	10.04
50	100	3.42	10.44	11.84	3.90	9.60	9.82	3.82	9.70	9.98
50	500	2.58	9.80	10.34	2.66	9.52	9.36	3.20	8.82	10.36
100	100	1.36	10.30	10.64	1.54	9.86	9.64	2.42	8.94	9.18
100	500	1.10	9.64	9.94	1.26	9.00	9.60	2.18	9.16	9.20
500	500	0.02	7.82	9.76	0.02	8.00	9.88	0.22	8.18	9.52

Notas: Véase el Cuadro 4

Cuadro 6: Valores-p, tamaño y conclusiones de la aplicación del estadístico-t y el procedimiento de Simes a los pronosticadores aparentemente no-monotónicos.

Componente	d_j	valor-p _j	α_j	Conclusión
Pronosticador 1				
$j = 1$	RECM ₁ –RECM ₂	0.84	0.03	no se rechaza
$j = 2$	RECM ₂ –RECM ₃	0.47	0.10	no se rechaza
$j = 3$	RECM ₃ –RECM ₄	0.77	0.07	no se rechaza
Pronosticador 3				
$j = 1$	RECM ₁ –RECM ₂	0.98	0.07	no se rechaza
$j = 2$	RECM ₂ –RECM ₃	0.99	0.03	no se rechaza
$j = 3$	RECM ₃ –RECM ₄	0.41	0.10	no se rechaza
Pronosticador 96				
$j = 1$	RECM ₁ –RECM ₂	0.99	0.03	no se rechaza
$j = 2$	RECM ₂ –RECM ₃	0.86	0.07	no se rechaza
$j = 3$	RECM ₃ –RECM ₄	0.20	0.10	no se rechaza
Pronosticador 452				
$j = 1$	RECM ₁ –RECM ₂	0.98	0.07	no se rechaza
$j = 2$	RECM ₂ –RECM ₃	0.00	0.10	se rechaza
$j = 3$	RECM ₃ –RECM ₄	0.99	0.03	no se rechaza