

# PROYECCIÓN DE PRODUCCIÓN Y VENTAS INDUSTRIALES CON PROCEDIMIENTOS AUTOMATIZADOS\*

Rómulo Chumacero E.\*\*

## I. INTRODUCCIÓN

La precisión es importante en las proyecciones, porque estas se usan para guiar decisiones. Dada la variedad de métodos de proyección disponibles, se hace necesario contar con un enfoque metodológico riguroso para evaluar y comparar sus fortalezas y debilidades.

En este sentido, este artículo provee un aporte al considerar una variedad de modelos de series de tiempo y sus respectivos procesos automatizados de selección.<sup>1</sup> Estos modelos se utilizan para proyectar la producción y las ventas industriales de Chile con información mensual de estas series para una muestra comprendida entre los años 1991 y 2003.

Aunque el principal objetivo del trabajo es el de extraer lecciones para realizar proyecciones, en casos, los resultados pueden servir como guía del tipo de modelos teóricos compatibles con los resultados aquí obtenidos.

Algunos de los modelos considerados incorporan características no lineales que pueden estar presentes en los datos. Es importante, sin embargo, destacar que en todo proceso de estimación o proyección que no esté estrictamente guiado por un modelo teórico, el investigador incurrirá muchas veces en una exhaustiva búsqueda entre modelos de complejidad variable. Por ello, es necesario que cuando se comparen proyecciones se utilice un enfoque robusto a este aspecto.

El resto del artículo se organiza como sigue: la sección II entrega una breve descripción de los datos. La sección III presenta los diversos tipos de modelo utilizados. La sección IV aplica varias metodologías para comparar la precisión de los modelos de proyección. La sección V resume los resultados encontrados.

## II. Los DATOS

En este trabajo se utilizan observaciones mensuales de las variaciones en doce meses de la producción y de las ventas industriales que publica el Instituto Nacional de Estadísticas (INE) para el período 1991:12 y 2003:11.<sup>2</sup>

**CUADRO 1**

**Resumen Estadístico: Producción y Ventas Industriales en Chile**  
(tasas de crecimiento entre 1991:12 y 2003:11)

	Producción	Ventas
Media (%)	3.27	3.46
Desviación estándar (%)	4.85	5.09
Autocorrelación de primer orden	0.56	0.50
Test de normalidad de Jarque-Bera (valor $p$ )	0.25	0.86
Test de raíz unitaria ADF (valor $p$ )	0.04	0.02

El cuadro 1 y el gráfico 1 presentan la evolución y principales estadísticas descriptivas de las series. Ambas presentan tasas de crecimiento positivas, aunque muy volátiles, en la mayor parte de la muestra. Si bien —como se esperaría— las raíces unitarias se rechazan con fuerza (test ADF), ambas series muestran

\* Agradezco a Álvaro Aguirre, Ernesto Pastén, Klaus Schmidt-Hebbel, Rodrigo Valdés y a un árbitro anónimo por sus útiles comentarios y sugerencias. También agradezco a Fondecyt (N° 1030681) por su valioso apoyo. Se aplican los descargos habituales.

\*\* Departamento de Economía, Universidad de Chile y Gerencia de Investigación Económica, Banco Central de Chile. Email: rchumace@bcentral.cl.

<sup>1</sup> Chumacero (2004) provee un marco teórico y metodológico general para realizar proyecciones.

<sup>2</sup> Si  $Y_t$  representa los datos originales, la variable de interés es  $y_t = \ln(Y_t/Y_{t-12})$ . No se realiza ninguna transformación adicional a los datos.

una persistencia relativamente alta para variables que ya están expresadas como tasas de crecimiento. Una muestra del desafío que implica proyectar estas series es que sus coeficientes de variación (la razón entre la desviación estándar y la media) exceden en mucho la unidad. Por último, ambas series están fuertemente correlacionadas (la correlación simple es 0.76) y existe evidencia marginal de causalidad unidireccional *à la* Granger desde producción hacia ventas (no reportada aquí).

### III. LOS MODELOS

Esta sección presenta los modelos de proyección utilizados. Todos ellos son modelos univariados aunque incluyen adicionalmente a la variación en doce meses del número de días laborales en el mes ( $d$ ).

Se sometieron a escrutinio cuatro tipos de modelos: un modelo lineal autorregresivo, uno de redes neuronales artificiales y un tercero autorregresivo de umbral; el cuarto es una combinación de los tres.

#### 1. El Modelo Autorregresivo Lineal

El modelo autorregresivo lineal (AR) para la serie  $y$  se formula como sigue:

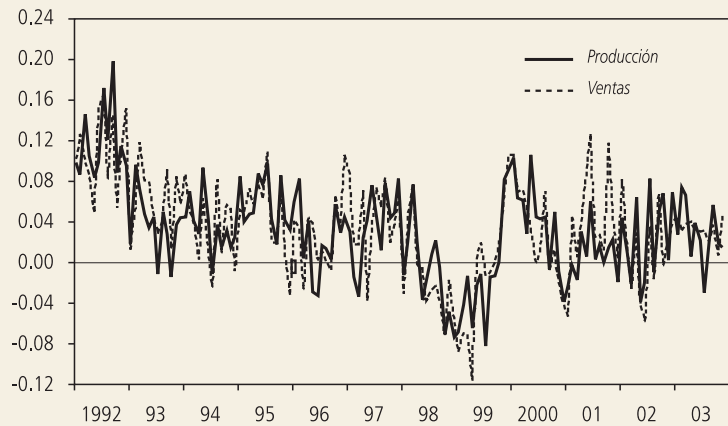
$$y_t = \alpha + \sum_{j=1}^p \beta_j y_{t-j} + \delta d_t + u_t, \quad (1)$$

donde  $y$  es la variable de interés (producción o ventas),  $u$  es un ruido blanco y  $\alpha, \beta, \delta$  son parámetros a ser estimados.

Este tipo de modelos tiene varias ventajas como instrumento para realizar proyecciones. Cuando se utiliza con series estacionarias, un modelo AR puede aproximar procesos ARMA más generales. A su vez, estos modelos son simples de estimar e interpretar, pudiendo muchas veces capturar la dinámica de corto plazo y tener implicancias de largo plazo. En casos, estas especificaciones pueden ser motivadas por modelos teóricos generales, aunque obviamente tienen la desventaja de no ser capaces de capturar no linealidades en los datos.

### GRÁFICO 1

#### Producción y Ventas Industriales en Chile (tasas de crecimiento entre 1991:12 y 2003:11)



Para determinar el orden del proceso, se selecciona  $p$  minimizando el criterio de información de Hannan-Quinn ( $HQ$ )

$$HQ_i \cong \ln(\hat{\sigma}_i^2) + 2\frac{k_i}{T} \ln(\ln(T)), \quad (2)$$

donde  $k$  es el número de parámetros estimados,  $T$  es el tamaño de la muestra, y  $\hat{\sigma}_i^2$  es un estimador consistente de la varianza de  $u$  para el modelo  $i$ .<sup>3</sup> Este criterio tiene algunas propiedades deseables por sobre otros candidatos, pues se sitúa entre el criterio de Akaike (que tiende a sobreajustar) y el criterio de Schwarz (que escoge modelos más reducidos).<sup>4</sup>

En la aplicación empírica, los órdenes mínimo y máximo de  $p$  se fijan en 1 y 36. Utilizando (2), los modelos escogidos fueron un AR(12) para la producción industrial y un AR(3) para las ventas industriales. En ambos casos, los residuos resultantes se pueden caracterizar como ruido blanco.

#### 2. El Modelo de Redes Neuronales Artificiales

El modelo de redes neuronales artificiales (NN) con

<sup>3</sup> Como todo criterio de selección, este contrapesa el hecho de que modelos "más grandes" (con más parámetros estimados) reducirán la volatilidad de los residuos (primer término) a costa de reducir los grados de libertad (segundo término).

<sup>4</sup> Más aun, este criterio de información es consistente en el sentido de que escoge el modelo "correcto" con probabilidad 1 a medida que  $T \rightarrow \infty$ . Como es bien sabido, el criterio de Akaike no cumple con este requisito (Inoue y Kilian, 2003).

$K$  unidades (capas) ocultas se define como

$$(3)$$

donde  $\Psi(v) = (1 + e^{-v})^{-1}$  es una función de activación logística,  $\phi_k$  es el peso con que se pondera la unidad oculta  $k$  que debe ser estimada junto a los parámetros  $\alpha, \beta, \delta$  de cada componente.

Como señala Tkacz (2001), los modelos NN tiene varias ventajas. En primer lugar, las redes neuronales son herramientas flexibles y empujadas por los datos, que resultan particularmente útiles cuando no hay ideas preconcebidas respecto de formas funcionales. A su vez, cuando la especificación es la apropiada, las NN son aproximaciones universales de distintas formas funcionales, colapsando en modelos lineales como casos particulares cuando los coeficientes  $\phi_k$  son todos cero. Finalmente, la función de activación logística es de una función suave y fácil de operar desde el punto de vista computacional.

Para un valor dado de  $K$ , se puede estimar (3) usando mínimos cuadrados no lineales (Kuan y White, 1994). Puede ser difícil estimar modelos con valores elevados de  $K$  pues el número de parámetros a estimar aumenta en forma lineal. La elección de  $K$  también se hace minimizando (2). Para hacer manejable la búsqueda, el valor de  $p$  se determina igual que en el modelo lineal, forzando así el modelo NN a anidar el modelo AR.

En la aplicación empírica, los órdenes mínimo y máximo de  $K$  se fijan en 1 y 3. En el caso de la producción industrial, el modelo escogido tenía  $K=3$ . Para las ventas, el modelo escogido tenía  $K=1$ . Esto puede interpretarse como que, de existir no linealidades en la media condicional, estas serían más pronunciadas en el caso de la producción que en el de las ventas.

### 3. El Modelo de Umbrales

El modelo autorregresivo de umbrales con dos regímenes (TAR) se formula como

$$y_t = \begin{cases} \alpha_1 + \sum_{j=1}^p \beta_{j,1} y_{t-j} + \delta_1 d_t + u_t & \text{si } y_{t-r} \leq \theta \\ \alpha_2 + \sum_{j=1}^p \beta_{j,2} y_{t-j} + \delta_2 d_t + u_t & \text{si } y_{t-r} > \theta \end{cases} \quad (4)$$

donde  $\theta$  es el valor del umbral e  $y_{t-r}$  es la variable del umbral. Esta ecuación dice que cuando la variable rezagada (utilizada como variable de umbral) alcanza y excede un cierto valor crítico, su influencia sobre su valor presente es significativamente distinta que cuando no lo hace. Este tipo de modelos puede servir para modelar fenómenos en los cuales se considere que existen importantes asimetrías. Por ejemplo, si  $\theta$  fuese igual a cero, la ecuación (4) nos diría que la serie tiene comportamientos radicalmente distintos en períodos de expansión respecto de períodos de contracción.

La popularidad de este modelo proviene de que entrega una alternativa fácil de estimar frente al modelo de cambio de régimen (de hecho, es un caso especial de este último) y es coherente con las características no lineales de los datos, además de proporcionar funciones asimétricas de impulso-respuesta.<sup>5</sup>

Para un valor de  $r$ , el estimador de  $\theta$  en (4) puede obtenerse mediante un proceso directo de búsqueda que minimice la suma de cuadrados de los residuos totales.<sup>6</sup> Como el valor de  $r$  es desconocido, su elección se hace minimizando (2). Esto es, para cada posible candidato a variable de umbral, se minimiza la suma de cuadrados de los residuos y se calcula el valor del estadístico  $HQ$ , escogiendo finalmente el valor de  $r$  para el candidato a variable de umbral que entregue un valor menor de este estadístico.

En la aplicación, nuevamente hacemos coincidir el valor de  $p$  con aquel encontrado para el modelo lineal y permitimos variar los valores de  $r$  entre 1 y 12. Tanto para la producción como para las ventas, el valor resultante para  $r$  fue de 1.

<sup>5</sup> Para más detalle, véase Hansen (1997) o Siliverstovs y van Dijk (2003) y las referencias allí contenidas. Una variante de los modelos de umbral corresponde a modelos de transición suave (smooth transition) en los que no se consideran cambios discretos de régimen. Esta opción metodológica también permite introducir asimetrías y no linealidades aunque difiere en términos de teoría asintótica. En términos de proyecciones, la profesión parece haberse movido hacia modelos TAR. Pese a que este trabajo no lo hace, puede ser interesante comparar los méritos relativos de estos modelos.

<sup>6</sup> Esto es, considerar todos los valores de  $\theta$  observados para  $y_{t-r}$  con los que sea posible realizar la estimación de los parámetros en cada régimen.

#### 4. Proyección Combinada

En palabras de Fang (2003), los modelos de proyección difieren por su estructura y los datos que empleen. Si las proyecciones no están perfectamente correlacionadas entre sí, cada una puede capturar diferentes dimensiones de la dinámica de una serie. Por ello, a menudo, la combinación de proyecciones en competencia permite hacer proyecciones más precisas.

El modelo de proyecciones combinadas (C) es

$$\hat{y}_t^c = \sum_{l=1}^L \hat{\omega}_l \hat{y}_t^l, \tag{5}$$

donde  $\hat{\omega}_l$  es la ponderación asociada con la proyección obtenida con el método  $l$

Dadas  $L$  proyecciones, las ponderaciones se obtienen como sigue:

Esto es, las ponderaciones son obtenidas mediante la estimación de un problema de mínimos cuadrados restringidos entre el valor efectivo de una serie y las proyecciones obtenidas con los distintos modelos.

En esta aplicación, definimos  $L$  igual a 3 y combinamos las proyecciones de los tres modelos descritos arriba.

#### IV. EVALUACIÓN DE LAS PROYECCIONES

Existen varios criterios de evaluación para juzgar el comportamiento de un modelo de proyecciones. Aquí nos concentramos en una evaluación de las proyecciones puntuales  $h$  períodos adelante.<sup>7</sup>

Para evaluar las proyecciones puntuales, se consideran la Raíz del Error Cuadrático Medio (RMSE) y el Promedio de Valores Absolutos de los Errores (MAFE).<sup>8</sup> Sea  $\{\hat{u}_{i,t}\}_{t=T_0}^T$  la secuencia de los errores de proyección  $h$  períodos adelante obtenidos con el modelo  $i$  y sea  $T^* = T - T_0 + 1$ . Definimos RMSE y MAFE del modelo  $i$  como

$$RMSE_i = \sqrt{\frac{1}{T^*} \sum_{t=T_0}^T \hat{u}_{i,t}^2}$$

$$MAFE_i = \frac{1}{T^*} \sum_{t=T_0}^T |\hat{u}_{i,t}|$$

Al comparar varios modelos de proyección, es inevitable preguntarse cuál es la probabilidad de que el resultado que un modelo tenga efectivamente un RMSE error se deba a la casualidad. Diebold y Mariano (1995) comparan las proyecciones en un esquema como este.

Considérese el par de proyecciones  $h$  períodos adelante de los modelos  $i$  y  $j$  ( $\hat{u}_{i,t}, \hat{u}_{j,t}$ ) para  $t = T_0, \dots, T$ ; cuya calidad se juzgará mediante la función de pérdida.<sup>9</sup> Definimos por lo que bajo la hipótesis nula de igual precisión en las proyecciones de los modelos  $i$  y  $j$ , tenemos que el valor esperado de  $x$  debe ser igual a cero. Si este proceso es estacionario en covarianza, es natural basar un test en la media muestral observada:

$$\bar{x} = \frac{1}{T^*} \sum_{t=T_0}^T x_t.$$

Como es sabido, aún realizando proyecciones óptimas  $h$  períodos adelante, la secuencia de errores de proyección seguirá un proceso MA( $h - 1$ ).<sup>10</sup> En tal caso, las autocorrelaciones de orden  $h$  y superiores son cero, por lo que la varianza ( ) de puede obtenerse del modo siguiente:

donde es una estimación de la  $j$ -ava autocovarianza de  $x_t$ .

El estadígrafo de Diebold Mariano (DM) viene dado por

<sup>7</sup> La capacidad predictiva de los modelos también se puede juzgar en otras dimensiones, tales como las proyecciones de intervalos y de densidad (Silverstovs y van Kijk, 2002). Chumacero (2004) presenta un marco general para realizar estas comparaciones y describe métodos automáticos diseñados para estos efectos.

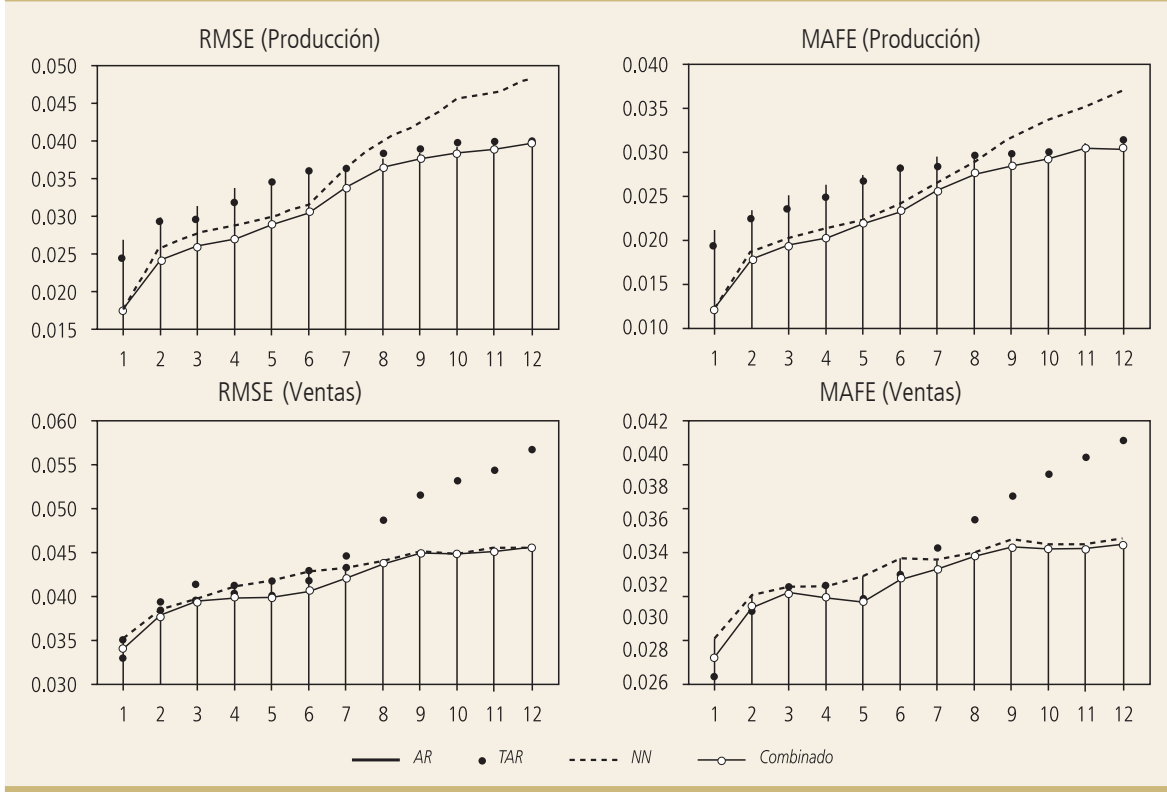
<sup>8</sup> Al comparar modelos con ambos criterios no debieran presentarse grandes discrepancias cuando la distribución de los errores es simétrica y no presenta colas gruesas o "outliers."

<sup>9</sup> Por ejemplo, en el caso de la comparación del RMSE tendremos la función cuadrática de pérdida  $g(\hat{u}_{i,t}) = \hat{u}_{i,t}^2$  y, en el caso de MAFE, la función de pérdida

<sup>10</sup> Esto quiere decir que el único cosa en que la variable  $x$  se comportaría como ruido blanco es un proyecciones un periodo adelante (ver Patton y Timmermann, 2002).

GRÁFICO 2

Evaluación de Proyección  $h$  Períodos Adelante Dentro de la Muestra



$\hat{\beta}_T$

bajo la hipótesis nula de igual precisión en las proyecciones. Harvey et al. (1997) sugieren modificar el test  $DM$  y usar en su lugar:

$$HLN = DM \cdot \left[ \frac{T^* + 1 - 2h + h(h-1)/T^*}{T^*} \right]^{1/2}$$

para corregir problemas de tamaño de  $DM$ . Ellos también sugieren utilizar una distribución  $t$  de Student con  $T^* - 1$  grados de libertad en lugar de una normal estándar para dar cuenta de posibles errores de cola gruesa.

Para testear si el modelo  $i$  no está dominado por el modelo  $j$  en cuanto a la precisión de la proyección para la función de pérdida  $g(\cdot)$ , se puede realizar un test de una cola para  $DM$  o  $HLN$ , con la hipótesis nula  $E x_t < 0$ , donde  $E$  denota la esperanza condicional.

De rechazarse la hipótesis nula, concluiremos que existe evidencia estadística de que el modelo  $j$  predomina sobre el modelo  $i$ .

1. Evaluación dentro de la Muestra

Una forma de evaluar modelos en competencia es estimar la precisión de sus proyecciones. Para cada proyección dentro de la muestra (IS), se estima cada modelo con la muestra completa. El “mejor” de cada categoría se escoge minimizando (2), y se usan los coeficientes así obtenidos para hacer proyecciones  $h$  períodos adelante. Por ejemplo, una proyección tres períodos adelante para el período  $y_t$  utiliza los coeficientes estimados con la muestra completa, pero con las observaciones de  $y$  solo hasta el período  $t-3$ .

El gráfico 2 muestra el patrón típico de las proyecciones  $h$  períodos adelante (presentadas en el eje horizontal). Esto es, tanto RMSE como MAPE aumentan a medida que aumenta el período de proyección. Los mismos resultados sugieren que, en términos de proyecciones dentro de muestra, no existen ventajas en utilizar modelos de proyección no lineales respecto al modelo AR (que se presentan como líneas verticales) en el caso de las ventas

industriales y que estas se disipan en horizontes de proyección prolongados en el caso de la producción industrial.

Dado que los resultados del gráfico 2 corresponden a los estadísticos obtenidos con proyecciones puntuales, estos no sirven para concluir si las diferencias en desempeño entre los distintos métodos son estadísticamente significativas. Por ello, el cuadro 2 muestra el número de horizontes de proyección (de los doce posibles) en los que existe evidencia estadísticamente significativa de que un modelo domine a otro. Por ejemplo, utilizando RMSE como criterio de comparación, en el caso de la producción industrial, el modelo AR es dominado por el modelo de redes neuronales en 2 de los 12 horizontes de proyección considerados. Mientras tanto, bajo la misma métrica, el modelo NN es dominado por el modelo AR solo en seis de los doce horizontes.

Concluyendo, al evaluar el desempeño de los distintos modelos en proyecciones dentro de muestra, encontramos que los modelos no lineales parecen superar al modelo lineal solo en proyecciones de muy corto plazo. Finalmente, es importante destacar que, en términos de proyecciones dentro de muestra, existen ventajas en combinar proyecciones. Esto porque el modelo que combina proyecciones nunca es dominado en proyecciones por ninguno de los

otros modelos en el caso de las ventas industriales y solo en uno de los doce horizontes de proyección por el modelo NN (usando como criterio de comparación a MAFE) en el caso de la producción.

Si bien son útiles como herramientas de evaluación (Inoue y Kilian, 2002), las proyecciones dentro de la muestra no son la razón por la que en la práctica se usan los modelos. A continuación, evaluamos el desempeño de los modelos presentados en cuanto a las proyecciones fuera de la muestra.

## 2. Evaluación Fuera de la Muestra

La muestra completa para ambas series comprende 144 observaciones. Para las proyecciones fuera de la muestra (OS), obtenemos estimaciones para cada modelo partiendo con las primeras 100 observaciones, producimos una proyección para el horizonte correspondiente, agregamos una observación, volvemos a estimar cada modelo, producimos sus nuevas proyecciones, y así sucesivamente hasta agotar la muestra. El modelo usado en cada categoría es el mejor obtenido con la muestra completa, y no es escogido con cada nueva observación.

Los resultados de la comparación del comportamiento de los distintos modelos son bastante robustos. El gráfico 3 muestra que las proyecciones de la producción industrial para un período adelante todavía se pueden capturar mejor con modelos NN (no lineales) que, una vez más, se deterioran cuando aumenta el horizonte de proyección.<sup>11</sup> En tanto que ningún modelo no lineal parece proveer mejor información que los modelos lineales a la hora de proyectar las ventas industriales.

El cuadro 3 presenta evidencia estadística robusta de lo mostrado en el gráfico 3. En particular, el modelo lineal no es dominado por los dos modelos no lineales considerados para ningún horizonte de proyección (salvo en el caso de la producción industrial cuando se compara al modelo lineal con el de redes neuronales para proyecciones un período

**CUADRO 2**

### Evaluación de Proyecciones Puntuales con Test HLN\*

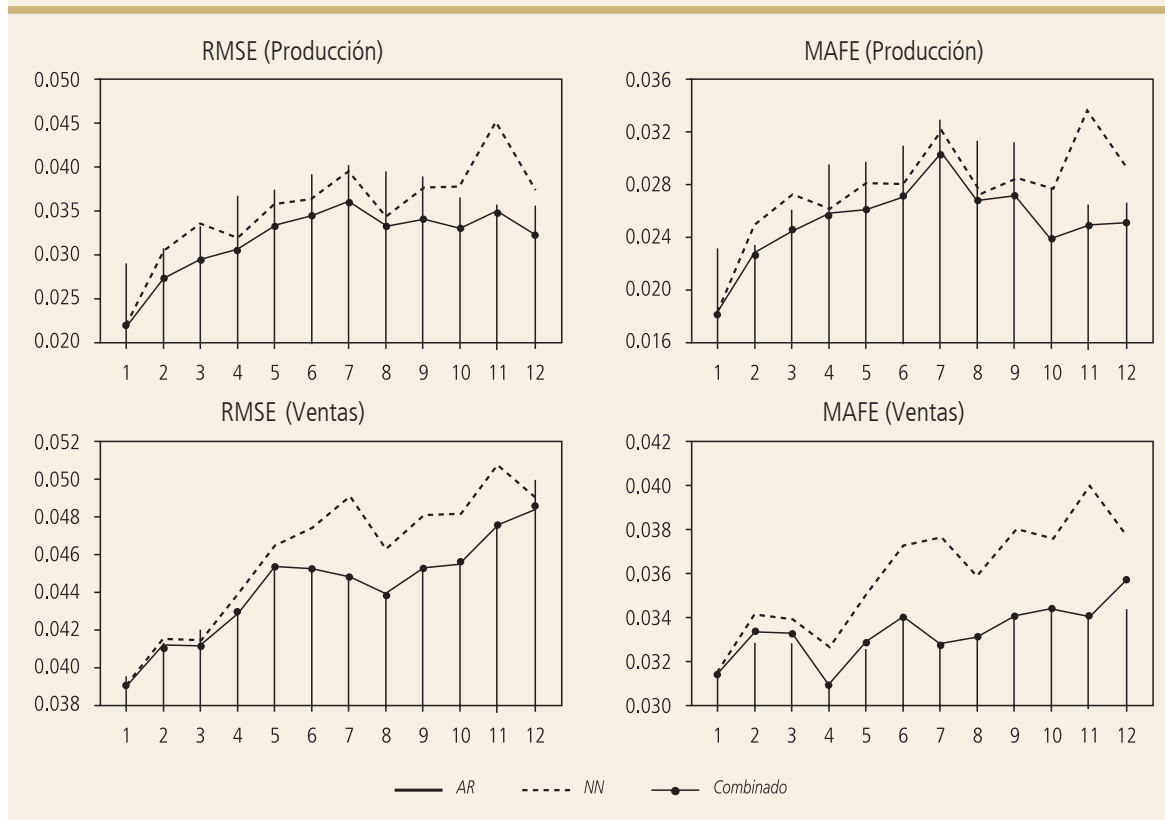
Modelo <i>i</i> – Modelo <i>j</i>	Producción				Ventas			
	RMSE	MAFE	RMSE	MAFE	RMSE	MAFE	RMSE	MAFE
<b>NN</b>	2	<b>6</b>	2	<b>5</b>	1	<b>0</b>	1	<b>0</b>
AR - <b>TAR</b>	2	<b>0</b>	1	<b>0</b>	1	<b>0</b>	1	<b>0</b>
<b>Combinado</b>	4	<b>0</b>	5	<b>0</b>	2	<b>0</b>	2	<b>0</b>
NN - <b>TAR</b>	7	<b>1</b>	4	<b>2</b>	0	<b>0</b>	0	<b>0</b>
<b>Combinado</b>	10	<b>0</b>	8	<b>1</b>	0	<b>0</b>	0	<b>0</b>
<b>TAR - Combinado</b>	4	<b>0</b>	5	<b>0</b>	3	<b>0</b>	1	<b>0</b>

\* Los números en color regular indican el número de horizontes de proyección (de los doce posibles) en los que se rechaza la hipótesis nula que el modelo *i* no está dominado por el modelo *j*. En negrita, la hipótesis inversa.

<sup>11</sup> El desempeño de los modelos TAR se deteriora con mucha rapidez para los horizontes de proyección a cuatro o más periodos. Por esta razón, ellos se excluyen del gráfico 3 para poder hacer una mejor comparación visual. Nuevamente, el RMSE del modelo AR se presenta en líneas verticales.

**GRÁFICO 3**

**Evaluación de Proyección  $h$  Periodos Adelante  
Fuera de la Muestra**



**CUADRO 2**

**Evaluación de Proyecciones Puntuales  
Fuera de la Muestra con Test HLN\***

Modelo $i$ – Modelo $j$	Producción		Ventas					
	RMSE	MAFE	RMSE	MAFE				
<b>NN</b>	1	0	1	0	0	0	0	0
AR - <b>TAR</b>	0	5	0	3	0	5	0	5
<b>Combinado</b>	3	0	2	0	0	0	0	0
NN - <b>TAR</b>	0	3	0	3	0	2	0	1
<b>Combinado</b>	1	0	0	0	0	0	0	0
TAR - <b>Combinado</b>	9	0	9	0	6	0	6	0

\*Los números en color regular indican el número de horizontes de proyección (de los doce posibles) en los que se rechaza la hipótesis nula que el modelo  $i$  no está dominado por el modelo  $j$ . En negrita, la hipótesis inversa.

adelante). Esto demuestra que si bien existen algunas similitudes entre los resultados de proyecciones dentro y fuera de muestra, en el último caso, características no lineales no son necesariamente útiles para realizar proyecciones. Si algo, se muestra que las proyecciones del modelo TAR se deterioran bastante rápidamente.

Finalmente, los resultados son fuertemente coherentes con la ventaja de combinar proyecciones (sobre todo para horizontes de proyección cercanos). Esto porque ninguno de los modelos considerados es capaz de dominar a las proyecciones del modelo combinado y este domina en cuando menos algún horizonte de proyección a los demás modelos.

**3. El Test de la Realidad de White**

Tal como señala White (2000), cada vez que se obtiene un “buen” modelo de proyección mediante

una extensa búsqueda de especificación, existe el peligro de que el buen desempeño sea producto de la buena fortuna y no de una genuina capacidad predictiva. Aun si no existe una relación útil entre variables, si se mira un conjunto de datos por suficiente tiempo, normalmente se revelarán patrones inexistentes que al final no sirven de nada para realizar proyecciones. La práctica de realizar búsquedas extensas y derivar inferencias es conocida como “data snooping”, y puede llevar al investigador ingenuo a confundir lo espurio con lo sustantivo.

White (2000) proporciona un esquema formal para testear la hipótesis nula que el mejor modelo encontrado durante una especificación no es superior en su calidad predictiva que un modelo de referencia. Existe una diferencia sutil pero importante entre este test y los tests *DM* y *HLN* descritos previamente. Los últimos comparan las proyecciones resultantes de dos modelos dados, mientras que el test de White lo hace entre las proyecciones resultantes de un modelo obtenido luego de probar distintas especificaciones y un modelo referencial dado.

El test de White tiene el muy apropiado nombre de “test de realidad” (“*reality check*”) y se puede realizar en varias formas. Aquí realizamos y describimos brevemente el llamado “*bootstrap reality check*.” Consta de los siguientes pasos:<sup>12</sup>

- Obtener el test *DM* o *HLN* para cada modelo  $j$  contra un modelo de referencia, y representarlo por  $A_j$ . Por ejemplo, para obtener el valor de  $A_j$  para el modelo TAR, se obtiene el test *DM* o *HLN* que compara los errores de proyección del mejor modelo TAR (el que minimiza la ecuación 2) con los de un modelo referencial fijo (como un modelo AR(1)).
- Generar  $i = 1, \dots, M$  muestras artificiales de  $y$  y demás variables involucradas en la estimación de los modelos. En este caso, se necesita un procedimiento de re-muestreo aplicable a procesos dependientes. Aquí usamos el *bootstrap* estacionario de Politis y Romano (1994).<sup>13</sup>
- Realizar la búsqueda de especificación, minimizando (2) para seleccionar el mejor modelo de cada categoría  $y$  de cada muestra artificial.
- Para cada muestra artificial  $i$ , comparar la proyección obtenida por el modelo  $j$  con un modelo de referencia (fijo), usando ya sea un test

*DM* o *HLN*. Representar este resultado por  $F_j$ .<sup>14</sup> Luego de que para cada modelo  $j$  se obtienen los  $M$  valores de  $F_j$  correspondientes a igual número de muestras artificiales, ordenar estos valores (de menor a mayor) y denotarlos por  $F_{(k)}$ .

- Encontrar el valor de  $F$  tal que  $F_{(k)} = F$ . Luego, el valor  $p$  (de probabilidad) del test “*bootstrap reality check*” es  $1 - F/M$ . Los valores  $p$  evalúan la hipótesis nula que los errores de proyección del modelo  $j$  analizado no dominan a los errores del modelo referencial.

El modelo de referencia que escogemos es un AR(1). El cuadro 4 muestra los resultados del test de realidad cuando se compara a los distintos modelos con el modelo de referencia. Salvo para las proyecciones de muy corto plazo (uno y hasta dos períodos adelante), los mejores modelos AR y NN no superan (en error medio cuadrado) al modelo AR(1). Estos resultados confirman que las características no lineales son importantes solo para las proyecciones a muy corto plazo.<sup>15</sup>

## V. EN RESUMEN:

Contar con buenos modelos de proyección es vital para poder tomar decisiones informadas. Pese a su importancia, muchas veces no se evalúan rigurosamente las ventajas relativas de distintos modelos de proyección.

Este artículo desarrolla una metodología para realizar comparaciones entre las proyecciones de distintos

<sup>12</sup> El “bootstrap” es un método de re-muestreo. Este procedimiento puede ser muy exigente en materia computacional. Por ejemplo, obtener el valor  $p$ , (definido más adelante) para el modelo NN toma más de 30 horas en un Pentium 4. Todos los cálculos de este artículo se hicieron utilizando código GAUSS que se describe en Chumacero (2004).

<sup>13</sup> Otro método de re-muestreo de uso común es el “bootstrap” de bloques, aunque no garantiza que la muestra así generada sea estacionaria. En tanto, el método de bootstrap estacionario utiliza bloques de datos con tamaño aleatorio (muestreado de una distribución geométrica) que garantiza que la serie resultante sea estacionaria.

<sup>14</sup> White (2000) también usa la variante del test DM de West (1996).

<sup>15</sup> Por ejemplo, el valor  $p$  del test que evalúa la hipótesis nula que el modelo NN es superior al modelo referencial para proyectar ventas industriales tres períodos adelante es 0.36. Esto indica que para niveles de confianza convencionales, no se rechaza la hipótesis nula que el modelo NN no domina a un AR(1).



CUADRO 4				
Valores <i>p</i> del <i>Bootstrap Reality Check</i> *				
Periodos adelante	Producción		Ventas	
	AR	NN	AR	NN
1	0.01	0.00	0.42	0.48
2	0.36	0.01	0.51	0.48
3	0.80	0.36	0.74	0.75
4	0.81	0.50	0.77	0.77
5	0.82	0.60	0.84	0.87
6	0.79	0.68	0.85	0.86
7	0.76	0.88	0.85	0.84
8	0.74	0.77	0.79	0.79
9	0.71	0.80	0.81	0.78
10	0.55	0.85	0.81	0.77
11	0.59	0.83	0.87	0.83
12	0.71	0.77	0.90	0.88

\* Como referencia, se usa un modelo AR(1).

modelos y la aplica a la proyección de la producción y de las ventas industriales de Chile. Los modelos considerados son univariados (es decir que dependen solo del pasado de la propia variable) pero pueden ser lineales y no lineales. Dentro de los segundos se toman en cuenta modelos de redes neuronales y modelos de umbral.

Los resultados encontrados en este trabajo sugieren que los modelos no lineales solo pueden ser importantes para realizar proyecciones de muy corto plazo y que su capacidad predictiva se deteriora rápidamente.

Otra enseñanza que surge de los ejercicios desarrollados es que existen ventajas en combinar proyecciones de distintos modelos. La lógica de esto es que distintos modelos son capaces de capturar distintas dimensiones de los datos, por lo que su combinación puede aportar información adicional.

Una última consideración que debe tomarse en cuenta es que el investigador suele “probar” distintas especificaciones antes de utilizar un modelo de proyección. Ignorar este proceso de búsqueda al momento de evaluar la capacidad predictiva de un modelo puede ser peligroso.

## REFERENCIAS

- Chumacero, R. (2004). “Forecasting Univariate Time Series Using Automated Procedures.” Mimeo, Banco Central de Chile.
- Diebold, F. y R. Mariano (1995). “Comparing Predictive Accuracy.” *Journal of Business and Economic Statistics* 13: 253-65.
- Fang, Y. (2003). “Forecasting Combination and Encompassing Tests.” *International Journal of Forecasting* 19: 87-94.
- Hansen, B. (1997). “Inferences in TAR Models.” *Studies in Nonlinear Dynamics and Econometrics* 2: 1-14.
- Harvey, D., S. Leybourne y P. Newbold (1997). “Testing the Equality of Prediction Mean Square Errors.” *International Journal of Forecasting* 13: 281-91.
- Inoue, A. y L. Kilian (2002). “In-sample or Out-of-sample Tests of Predictability: Which One Should We Use?” Mimeo, North Carolina State University.
- Inoue, A. y L. Kilian (2003). “On the Selection of Forecasting Models.” Mimeo, North Carolina State University.
- Kuan, C. y H. White (1994). “Artificial Neural Networks: An Econometric Perspective.” *Econometric Reviews* 13: 1-91.
- Patton, A. y A. Timmermann (2002). “Properties of Optimal Forecasts.” Mimeo, Universidad de California, San Diego.
- Politis, D. y J. Romano (1994). “The Stationary Bootstrap.” *Journal of the American Statistical Association* 89: 1303-13.
- Siliverstovs, B. y D. van Dijk (2002). “Forecasting Industrial Production with Linear, Non-linear, and Structural Breaks Models.” Mimeo, Erasmus University.
- Tkacz, G. (2001). “Neural Network Forecasting of Canadian GDP Growth.” *International Journal of Forecasting* 17: 57-69.
- West, K. (1996). “Asymptotic Inference About Predictive Ability.” *Econometrica* 64: 1067-84.
- White, H. (2000). “A Reality Check for Data Snooping.” *Econometrica* 68: 1097-126.

