

PREDICIENDO EL TIPO DE CAMBIO EN ALTA FRECUENCIA

Carlos R. Barrera Chaupis
DMM

Resumen

Un problema recurrente en el presente periodo de flotación es que los modelos estructurales de determinación del tipo de cambio no logran predecirlo con mayor precisión que un camino aleatorio. El objetivo de la presente investigación es verificar si es posible construir un grupo de modelos econométricos que generen proyecciones precisas del tipo de cambio diario sobre la base de la muestra disponible enero 2004 - mayo 2008. Los principales modelos a compararse en términos predictivos son: (a) camino aleatorio en el nivel del tipo de cambio; (b) auto-regresión con p rezagos en la variación del tipo de cambio; (c) perceptrones con p rezagos en la variación del tipo de cambio y (d) auto-regresión fraccional con p rezagos en el nivel del tipo de cambio. Los resultados uni-variados obtenidos hasta el momento confirman que los perceptrones poseen la capacidad para anticipar el patrón de los movimientos diarios en el tipo de cambio, especialmente cuando se utiliza el *spread* entre el tipo de cambio venta y compra como porcentaje del tipo de cambio promedio de estas dos cotizaciones, la depreciación diaria del yen contra el dólar americano y el diferencial de tasas domésticas de interés interbancarias en ambas monedas.

CONTENIDO TENTATIVO

1. Introducción
2. Los modelos
3. Resultados empíricos
 - 3.1 Para modelos uni-variados
 - 3.2 Evaluación ex post (todos los modelos)
4. Conclusiones

1. Introducción

Desde que Meese & Rogoff(1983) presentaron evidencia de que un camino aleatorio tenía predicciones de corto plazo más precisas que las de los modelos macroeconómicos, se ha presentado múltiple evidencia sobre el pobre desempeño de las teorías macroeconómicas para explicar (y predecir) el tipo de cambio nominal en el corto, mediano y largo plazos (ver por ejemplo, Cheung, Menzie & Pascual (2005))¹.

En algunos países se dispone de información sobre el flujo de órdenes de compra y venta en el mercado cambiario y de la simple observación de estos datos resultó evidente que no son flujos constantes. Ante esta evidencia surgió el enfoque de micro-estructura de Lyons(2001), que afirma que es el flujo de órdenes que resulta de las decisiones de inversionistas heterogéneos el que mejor explica las fluctuaciones del tipo de cambio. Evans y Lyons(2002) presentan evidencia a favor de este enfoque para una muestra de países donde se dispone de estos datos.

Otra dimensión que puede ser informativa es la heterogeneidad de los inversionistas que generan la secuencia temporal del flujo de órdenes. Bacchetta & Wincoop (2004) presentan un modelo para explicar cómo esta dispersión de la información afecta el grado de influencia de cada variable macroeconómica sobre el flujo de órdenes y por lo tanto sobre el tipo de cambio. Así, cuando alguna variable macroeconómica fundamental para el tipo de cambio se aleja de su nivel promedio histórico, ésta genera un movimiento importante (aunque temporal) en el flujo de órdenes. Como su impacto sobre el tipo de cambio es mayor que en otros periodos, se convierte en el “chivo expiatorio” durante algún periodo. Como resultado de estos oleajes en la información, el tipo de cambio se “desconecta” de sus determinantes fundamentales.

Por lo tanto, la dirección para predecir el tipo de cambio nominal es enfatizar las frecuencias altas (diarias, semanales e incluso mensuales). El objetivo del presente trabajo es construir modelos dinámicos que permitan predecir diariamente el tipo de cambio, así como algunas otras variables observables que sean relativamente importantes para predecirlo en esta frecuencia. En la segunda sección se describe los modelos dinámicos utilizados (modelos VAR, VARFIMA y VARNN) así como otros que se incluirán en la versión final (VAR-ARCH & VARNN-ARCH). En la tercera sección se describe los resultados tentativos obtenidos, que corroboran las ventajas de los modelos VARNN, basados en redes neuronales artificiales (ANN). La cuarta sección concluye.

¹ Aunque el horizonte temporal podría ser una variable importante para calificar esta afirmación, pues hay evidencia a favor de los determinantes fundamentales en el largo plazo (entre 2 y 4 años según Mark (2005); ver también MacDonald & Taylor(1994)), Groen(1999) encuentra que el poder predictivo *ex post* de largo plazo es muy débil cuando el periodo de evaluación *ex post* es mayor que en los documentos anteriores. Groen atribuye esta evidencia a la ausencia de co-integración entre el tipo de cambio y sus determinantes fundamentales.

2. Los Modelos

Los modelos a utilizar en el presente estudio incluyen modelos VAR, **VARCH**, VARFIMA y VARNN, los que se describen estilizadamente a continuación. Esta descripción utiliza sólo una variable informativa (bi-variados) o sólo dos variables informativas (tri-variados), a discreción, para facilitar una comprensión de las características y particularidades principales de los modelos².

2.1. Modelos VAR

Estos modelos dinámicos son una buena referencia para el desempeño predictivo. Por ejemplo, un modelo VAR(K=2,p=2) consiste en dos ecuaciones

$$\begin{aligned} y_{1t} &= a_{10} + a_{11}^1 y_{1t-1} + a_{12}^1 y_{2t-1} + a_{11}^2 y_{1t-2} + a_{12}^2 y_{2t-2} + \varepsilon_{1t} \\ y_{2t} &= a_{20} + a_{21}^1 y_{1t-1} + a_{22}^1 y_{2t-1} + a_{21}^2 y_{1t-2} + a_{22}^2 y_{2t-2} + \varepsilon_{2t} \end{aligned} \quad \varepsilon_t \equiv \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad \varepsilon_t | \Omega_{t-1} \square N(0, \Sigma)$$

donde $y'_t \equiv \{y_{1t}, y_{2t}\}$ es un vector de variables estacionarias, $\Omega_{t-1} \equiv \{y'_{t-1}, y'_{t-2}\}$ es el conjunto de información relevante y Σ es la matriz de co-varianzas condicionales, invariantes en el tiempo. Los supra-índices en los coeficientes facilitarán su agrupamiento en matrices según los rezagos,

$$y_t - \mu + \Phi_1(y_{t-1} - \mu) + \Phi_2(y_{t-2} - \mu) = \varepsilon_t \rightarrow \Phi(L)(y_t - \mu) = \varepsilon_t$$

el cual puede estimarse usando el método de quasi-máxima verosimilitud o mediante MCO.

2.2. Modelos VAR-ARCH

Para caracterizar este tipo de modelos partiremos de una representación muy cercana a un modelo VAR(K=3;P=1)

$$\begin{aligned} y_{1t} &= a_{10} + a_{11} y_{1t-1} + a_{12} y_{2t-1} + a_{13} y_{3t-1} + \varepsilon_{1t} \\ y_{2t} &= a_{20} + a_{21} y_{1t-1} + a_{22} y_{2t-1} + a_{23} y_{3t-1} + \varepsilon_{2t} \\ y_{3t} &= a_{30} + a_{31} y_{1t-1} + a_{32} y_{2t-1} + a_{33} y_{3t-1} + \varepsilon_{3t} \end{aligned} \quad \varepsilon_t \equiv \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \quad \varepsilon_t | \Omega_{t-1} \square N(0, \Sigma_t)$$

donde $\Omega_{t-1} = \{y'_{t-1}\}$ es el conjunto de información relevante y Σ_t es la matriz de co-varianzas condicionales en el periodo t, donde

² En la siguiente sección se describe los resultados obtenidos con las versiones uni-variadas de los modelos descritos aquí de forma más general.

$$\Sigma_t \equiv \begin{bmatrix} \sigma_t^{11} & \sigma_t^{12} & \sigma_t^{13} \\ \sigma_t^{21} & \sigma_t^{22} & \sigma_t^{23} \\ \sigma_t^{31} & \sigma_t^{32} & \sigma_t^{33} \end{bmatrix}$$

Esta matriz se puede definir como un proceso GARCH(1,1) multi-variado con una parametrización que la restrinja a ser una matriz definida semi-positiva á la BEKK (ver Engle & Kroner(1995)) donde las correlaciones entre las variables puede ser variantes en el tiempo. Por ejemplo, Ebrahim(2000) plantea la siguiente parametrización :

$$\begin{aligned} \Sigma_t = & C'C + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'\Sigma_{t-1}B \\ & + G'u_{t-1}u'_{t-1}G + S'\psi_{t-1}\psi'_{t-1}S + P'\xi_{t-1}\xi'_{t-1}P \\ & + T'\eta_{t-1}\eta'_{t-1}T + Q'\zeta_{t-1}\zeta'_{t-1}Q \end{aligned}$$

donde las matrices (3,3) del lado derecho son matrices de parámetros: C es triangular superior, mientras que A,B,G,S,P,T y Q son diagonales. La matriz C captura el efecto sobre los componentes diferentes en Σ_t de un intercepto común (para considerar *dummies* estacionales basta definir matrices triangulares superiores sólo para determinados t's). La matriz A captura los impactos sobre las varianzas en Σ_t de las innovaciones pasadas (al cuadrado) mientras que la matriz B, los impactos de las varianzas del periodo anterior.

Las extensiones hechas por Ebrahim (2000) a la representación BEKK de Engle & Kroner (1995) son tres. La primera captura los impactos asimétricos de las innovaciones pasadas mediante la matriz G y la definición de

$$u_t \equiv \begin{bmatrix} \min(\varepsilon_{1t}, 0) \\ \min(\varepsilon_{2t}, 0) \\ \min(\varepsilon_{3t}, 0) \end{bmatrix}$$

que tiene entradas diferentes de cero cuando las innovaciones son negativas. La segunda captura los efectos sobre la varianza de un mercado de innovaciones pasadas en otros mercados (efectos conexos ó *spillovers*) vía las matrices S y P y la definición de

$$\psi_t \equiv \begin{bmatrix} \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{1t} \end{bmatrix} \quad \xi_t \equiv \begin{bmatrix} \varepsilon_{3t} \\ \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Finalmente, la tercera extensión³ captura la asimetría de los efectos conexos sobre las varianzas mediante las matrices T y Q y la definición de

³ Sin embargo, en Ebrahim(2000) estas extensiones se realizan *reduciendo* la generalidad de la representación BEKK. Por un lado, las varianzas condicionales no dependen de los productos cruzados de innovaciones pasadas (A es diagonal), de las co-varianzas condicionales pasadas en

$$\eta_t \equiv \begin{bmatrix} \min(\varepsilon_{2t}, 0) \\ \min(\varepsilon_{3t}, 0) \\ \min(\varepsilon_{1t}, 0) \end{bmatrix} \quad \zeta_t \equiv \begin{bmatrix} \min(\varepsilon_{3t}, 0) \\ \min(\varepsilon_{1t}, 0) \\ \min(\varepsilon_{2t}, 0) \end{bmatrix}$$

Si T es el número de observaciones, $y'_t \equiv \{y_{1t}, y_{2t}, \dots, y_{Kt}\}$ es la transpuesta del vector de K variables y Θ es el vector columna de todos los parámetros, la densidad condicional de $y_t | \Omega_{t-1}$ puede escribirse como

$$f(y_t | \Omega_{t-1}; \Theta) = \frac{1}{2\pi} |\Sigma_t|^{-1/2} e^{-\frac{1}{2}(\varepsilon'_t \Sigma_t^{-1} \varepsilon_t)}$$

y la función de quasi-verosimilitud resulta $l_Q = \sum_{t=1}^T \log(f(y_t | \Omega_{t-1}; \Theta))$ (ver).

2.3. Modelos VARFIMA

La forma general de un modelo VARFIMA(K;p,d,q) es básicamente un modelo VARMA aplicado a una transformación fraccionaria particular para cada una de las K variables no estacionarias incluidas que puede lograr que algunas presenten reversión a la media.

$$\Phi(L)\Delta^d(L)(y_t - \mu) = \theta(L)\varepsilon_t \quad \varepsilon_t | \Omega_{t-1} \square D(0, \Sigma)$$

donde $d = \{d_1, d_2, \dots, d_K\}$, $\Phi(L) = I - \sum_{j=1}^p \Phi_j L^j$, $\theta(L) = I + \sum_{j=1}^p \theta_j L^j$, D es una distribución multi-variada Normal o Student T, y $\Delta^d(L)$ es la matriz de transformaciones fraccionarias está definida como

$$\Delta^d(L) \equiv \begin{bmatrix} (1-L)^{d_1} & 0 & \dots & 0 \\ 0 & (1-L)^{d_2} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & (1-L)^{d_K} \end{bmatrix}$$

Como está establecido en la literatura desde Sowell(1992), el operador de la diferencia fraccional está definido como

otros mercados (B es diagonal) ni de las varianzas condicionales pasadas en otros mercados (cuya inclusión requiere de más matrices de coeficientes y de definir vectores con sólo algunos elementos de la diagonal de Σ_t). Por otro lado, las co-varianzas condicionales no dependen de las innovaciones pasadas al cuadrado ni de las varianzas condicionales pasadas en otros mercados.

$$(1-L)^{d_k} = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} L^j \quad \forall d_k \in \left(-\frac{1}{2}, \frac{1}{2}\right]$$

donde $\Gamma(\cdot)$ es la función Gamma, por lo que

$$x_{kt} \equiv (1-L)^{d_k} y_{kt} = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} y_{kt-j}$$

Como hace notar Sowell(1992), la serie x_{kt} puede obtenerse únicamente cuando se dispone de una serie infinita de y_{kt} . La solución que brinda a este problema para el dominio del tiempo es aplicar este filtro infinito (para un valor específico de d) a la serie finita disponible de y_{kt} aumentada con un vector de ceros para completar el pasado de la muestra⁴.

Una estrategia de estimación muy apropiada para estos modelos es el estimador CSS (asociado a la Suma de Cuadrados Condicional)⁵. La estimación multi-ecuacional toma como referencia los casos uni-ecuacionales. Cuando la distribución D es Normal, este estimador uni-ecuacional maximiza la función

$$S^N(\Theta) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2$$

donde $\varepsilon_t = \theta(L)^{-1} \Phi(L) \Delta^d(L)(y_t - \mu)$ ⁶, mientras que cuando la distribución D es Student T con ν grados de libertad, el estimador CSS uni-ecuacional maximiza la función

$$S^T(\Theta) = T \left[\log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2} \log(\nu-2) \right] - \frac{1}{2} \sum_{t=1}^T \left[\log(\sigma^2) + (\nu+1) \log\left(1 + \frac{\varepsilon_t^2}{(\nu-2)\sigma^2}\right) \right]$$

(ver Baillie & Cheng (1996)).

Olgun & Ozdemir (2008) son optimistas respecto a la principal ventaja de los modelos VARFIMA por sobre modelos como los VAR (y los VEC): la de aproximar mejor la dinámica de baja frecuencia presente en los datos, capturando características como memoria larga u oscilaciones largas, que pueden inutilizar

⁴ Nótese que es posible aplicar el filtro infinito sobre los datos como desviaciones de la media. Es en este último caso cuando es correcto completar el pasado de la muestra con un vector de ceros.

⁵ Ver Chung & Baillie(1993) para una prueba de que este estimador es asintóticamente equivalente al estimador MV cuando las observaciones iniciales son fijas. El estimador MV tiene la desventaja de tener un costo computacional prohibitivo.

⁶ En el caso de un modelo AR(p) para la serie transformada, se tiene $\varepsilon_t = \mu_x + \Phi(L)x_t$.

estas técnicas econométricas convencionales⁷. Más aún, comparten el optimismo en la literatura sobre los modelos ARFIMA sobre la ventaja de usar los datos en niveles sin sujetarlos a *ninguna* transformación porque se preserva el componente permanente (de baja frecuencia) de los datos. Esto no es exacto porque hemos puesto en evidencia que los datos sí están sujetos a una transformación, aunque ésta dependa de un parámetro estimado. La transformación específica es

$$x_t - \mu_x = g(y_t - \mu) \equiv \Delta^d(L)(y_t - \mu)$$

por lo que los modelos VARFIMA efectivamente seleccionan las transformaciones en el vector d de modo tal que obtenemos un modelo VARMA en los datos transformados

$$\Phi(L)(x_t - \mu_x) = \theta(L)\varepsilon_t \quad \varepsilon_t | \Omega_{t-1} \square D(0, \Sigma)$$

Como estas transformaciones basadas en los datos modifican la media, ello explicaría los resultados en las simulaciones de Cheng & Baillie(1993) sobre los considerables sesgos en los estimados de μ en muestras pequeñas, la corrupción implicada sobre los estimados CSS de los otros parámetros del modelo ARFIMA (especialmente los asociados a las medias móviles). Más aún, ello explicaría los resultados asintóticos de Yajima(1988) respecto a la inusualmente baja tasa de convergencia del estimador de μ , lo que desestima que la solución sea disponer de muestras de más de 150 observaciones porque la solución es más bien incluir criterios para seleccionar las transformaciones determinísticas más relevantes. Finalmente, estas consideraciones explican porqué Olgun & Ozdemir (2008) estiman su sistema con el método de MV con información completa y no muestran los estimados de su matriz Σ : estos estimados adolecen de una corrupción excesiva tal que resulta conveniente no presentarlos.

2.4. Modelos VARNN-perceptrón

Una de las principales arquitecturas de las *redes neuronales artificiales* (ANN) aplicables a la modelación de series temporales corresponde a los perceptrones multi-capas (MLP) (ver Dorffner(200?)), también conocidos como *Feedforward ANNs* (ver Kuan & Liu (1995)). En vez de la función lineal postulada en los modelos VAR(K,p),

$$y_t = A_0 + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t = A_0 + \sum_{j=1}^p A(j) y_{t-j} + \varepsilon_t \quad \varepsilon_t | \Omega_{t-1} \square N(0, \Sigma)$$

⁷ El criterio común nos lleva a la pregunta sobre cuánta información sobre la dinámica de baja frecuencia puede existir en series de 150 datos mensuales? La respuesta sería: mucha información sobre la dinámica de “baja” frecuencia, pero en la vecindad de la frecuencia de los datos (es decir, mucha información sobre “tendencias locales”).

donde $y'_t \equiv \{y_{1t}, y_{2t}, \dots, y_{Kt}\}$ es un vector de K variables estacionarias, $\Omega_{t-1} \equiv \{y'_{t-1}, y'_{t-2}, \dots, y'_{t-p}\}$ es el conjunto de información, podría postularse un modelo VAR no lineal

$$y_t = g(\Omega_{t-1}) + \varepsilon_t \quad \varepsilon_t | \Omega_{t-1} \sim N(0, \Sigma)$$

donde $g(\cdot)$ es una función no lineal multi-valorada específica, por ejemplo, la asociada a un modelo de transición suave (VSTVAR) o a un modelo de auto-excitación con límite (SETVAR) (ver Granger & Terasvirta(1993)). En contraste, un MLP, o más específicamente, un modelo VARNN-perceptrón (alternativamente, VARMLP), aproximará una función no lineal multi-valorada $g(\cdot)$ general como elemento de una familia muy amplia de funciones, aquella que más se ajuste a los patrones en los datos. A diferencia de una aproximación Taylor, que requiere conocer la función específica así como de un punto de aproximación, este MLP aproximará esta función mediante una combinación (en general, una función no lineal) de un número finito de H funciones no lineales “básicas” estructuradas en un grafo multi-capas. Por ejemplo,

$$g(\Omega_{t-1}) \cong \beta_0 + \sum_{i=1}^H \beta_i h_i(\Omega_{t-1}) = \beta_0 + \sum_{i=1}^H \beta_i \Psi_i \left(\Delta_{0,i} + \sum_{j=1}^p \Delta_i(j) y_{t-j} \right)$$

donde las H unidades h_i se denominan “unidades escondidas”, cada una de las cuales es una función no lineal multi-valorada Ψ_i que usualmente se restringe de modo tal que todos sus componentes son funciones acotadas (p.ej., la función tanh o la función logística).

Shachmurove(2003) menciona las principales ventajas y desventajas de las ANN. Entre las primeras se tiene la habilidad de analizar rápidamente patrones complejos, con un alto grado de precisión y sin mantener supuestos sobre la distribución de los datos. Entre las segundas se tiene que no existe un método estructurado y estándar para elegir, desarrollar, entrenar y verificar una ANN, su pertenencia al conjunto de herramientas que suele denominarse “caja negra”, así como su tendencia a sobre-parametrizar y a sub-parametrizar.

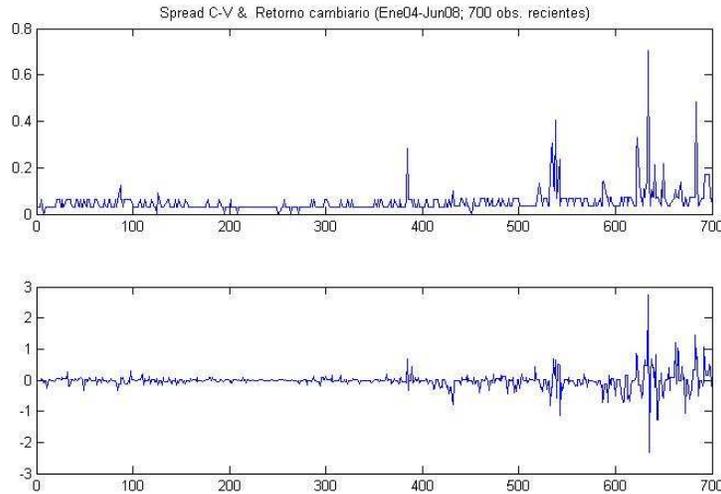
3. Resultados empíricos

3.1 Para modelos uni-variados

En primer lugar se realiza una exploración individual de las variables para ser usados como referencia de comparación con los modelos multi-variados. Las series temporales individuales estudiadas son (i) el *spread* de cotizaciones diarias venta-compra como porcentaje del promedio de ambas; y (ii) la variación diaria en

este tipo de cambio promedio diario. El Gráfico 1 muestra las 700 observaciones más recientes de la muestra disponible para ambas series⁸.

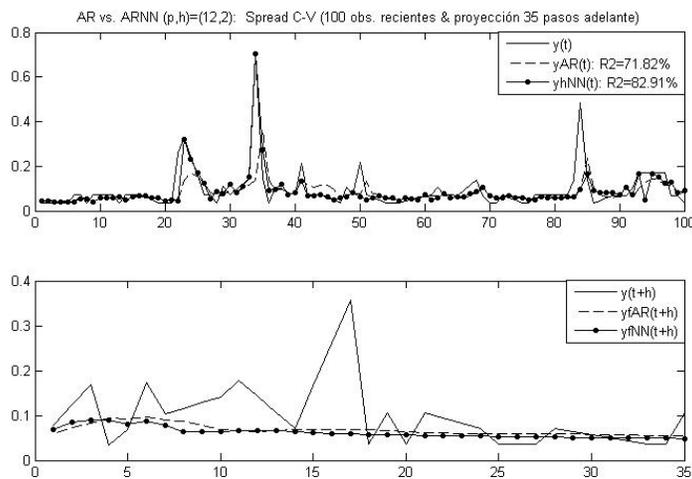
Gráfico 1



Spread: Modelos AR versus ARNN (y camino aleatorio)

La ventana superior del Gráfico 2 compara los 100 datos más recientes al interior de la muestra de estimación (marzo, abril y mayo 2008) con los *estimados* generados por los modelos y sus parámetros estimados (así como las correspondientes medidas de ajuste) para el caso del *spread*. La ventana inferior del mismo compara los datos más recientes *fuera de la muestra* (junio-julio 2008) con las *proyecciones* generadas por los modelos. Estos modelos utilizan **12 rezagos** y en el caso del ARNN, contiene 2 capas “escondidas”.

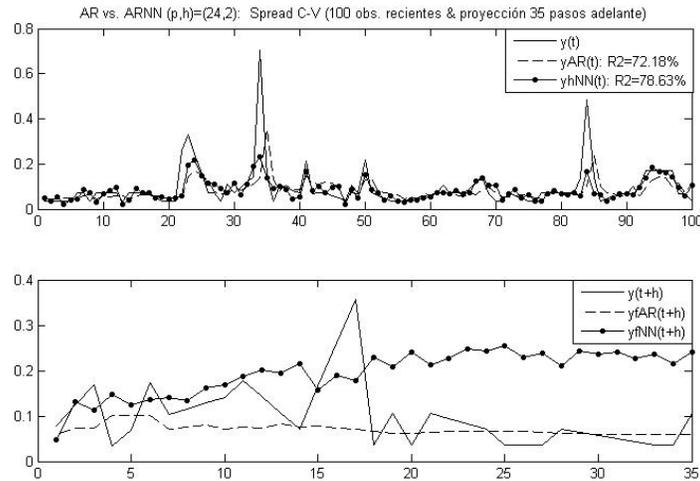
Gráfico 2



⁸ Dado que la muestra incluye más de 1500 observaciones, el problema de parsimonia puede tomarse como secundario.

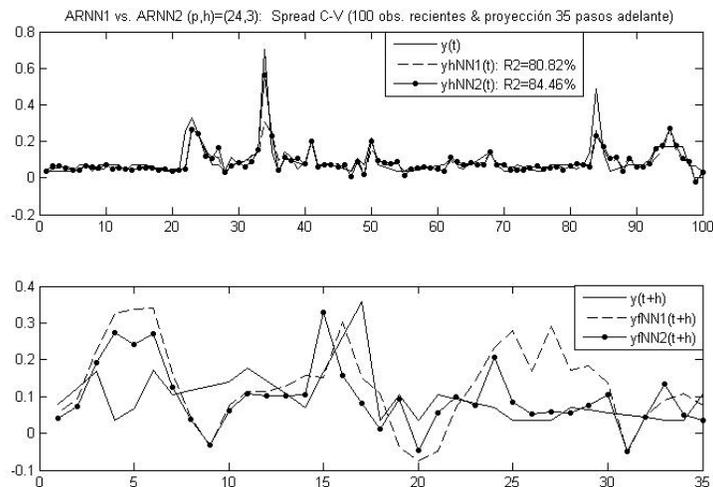
Puede observarse que, a pesar de un ajuste relativamente elevado, las dos proyecciones no logran capturar el patrón ejecutado fuera de la muestra. Más bien son muy similares a la proyección de un camino aleatorio para el *spread* (variaciones mensuales cero). El siguiente gráfico muestra los mismos resultados para el caso de modelos con **24 rezagos**.

Gráfico 3



Esta modificación aumenta marginalmente el ajuste del modelo AR pero reduce el ajuste del modelo ARNN. En términos de las proyecciones, la del modelo AR se mantiene similar al caso anterior, pero es la proyección del modelo ARNN la que logra capturar parte del patrón ejecutado fuera de la muestra, digamos los primeros 17 días adelante.

Gráfico 4



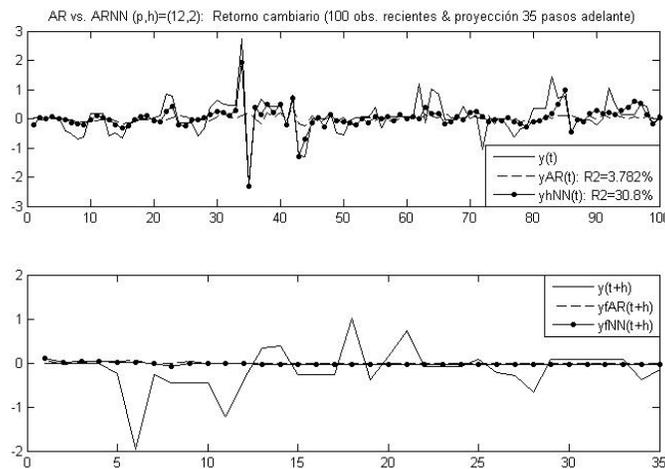
Los resultados anteriores señalan la posibilidad de que un mayor número de capas "escondidas" en el modelo ARNN mejore su capacidad predictiva. Efectivamente, este es el caso. El Gráfico 4 considera dos modelos ARNN entrenados uno después del otro (sin sobre-entrenamiento). Se puede observar

que capturan bien el patrón fuera de la muestra hasta 20 días adelante, pero sólo el segundo modelo ARNN sobrepasa las expectativas al capturar adecuadamente los siguientes días de la secuencia ejecutada.

Retorno: Modelos AR y ARNN (y camino aleatorio)

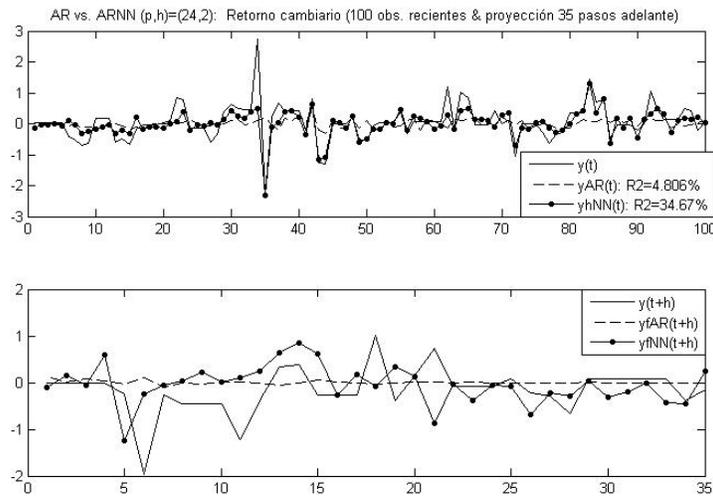
Los siguientes gráficos presentan el mismo ejercicio para el caso del retorno cambiario por mantener efectivo en dólares (la variación mensual del tipo de cambio promedio diario). El Gráfico 5 corresponde al mismo tipo de modelos del Gráfico 2, que utilizan 12 rezagos y en el caso del ARNN, 2 capas “escondidas”.

Gráfico 5



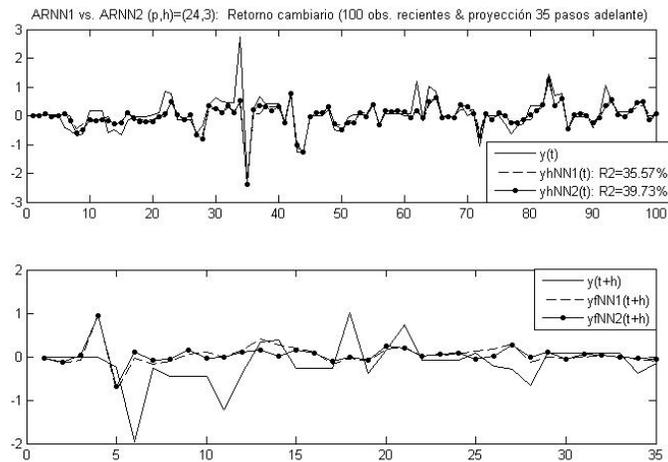
El ajuste no es tan elevado si tomamos como referencia los resultados análogos obtenidos para el *spread*, por lo que no es extraño que las dos proyecciones no logren capturar el patrón ejecutado del retorno fuera de la muestra (siendo equivalentes a la proyección de un camino aleatorio). Los modelos con 24 rezagos se presentan en el siguiente gráfico, el análogo al Gráfico 3.

Gráfico 6



Mientras que esta modificación aumenta marginalmente el ajuste de ambos modelos y, en términos de las proyecciones, las del modelo AR se mantienen similares al caso anterior, nuevamente son las proyecciones del modelo ARNN las que logran capturar dos segmentos del patrón ejecutado fuera de la muestra (los primeros 15 días adelante y los últimos 12 días de la secuencia de 35 días adelante).

Gráfico 7



Los resultados anteriores señalan la posibilidad de que un mayor número de capas “escondidas” en el modelo ARNN mejore su capacidad para predecir el retorno cambiario por mantener efectivo en dólares. Esta posibilidad no se confirma. El Gráfico 7 muestra dos modelos ARNN entrenados uno después del otro (sin sobre-entrenamiento), donde ambas proyecciones son muy similares y capturan el patrón fuera de la muestra hasta 25 días adelante, pero sin sobrepasar las expectativas.

Modelos ARFIMA(p,d,0)

Las proyecciones de la versión uni-variada de estos modelos no capturaron el patrón de corto plazo a pesar de lograr un ajuste dentro de la muestra de 99.99% con valores estimados del parámetro de integración fraccionario d cercanos a cero y $p=12$ rezagos para el nivel del tipo de cambio promedio diario (lo que indicaría estacionariedad).

Aunque estos resultados son tentativos, indican que usar un modelo de memoria larga cuando esta característica no está presente en los datos no impide obtener un ajuste elevado dentro de la muestra (el parámetro d puede generar sobreajuste) junto con una pobre capacidad predictiva⁹.

3.2 Evaluación *ex post* de todos los modelos

Las capacidades relativas de los modelos uni-variados descritos en la sección 2 fueron ilustradas con una proyección de 45 días como horizonte mediante su comparación gráfica con los correspondientes datos observados *ex post*, como en la sub-sección 3.1. En esta sub-sección se presenta los resultados de una evaluación de la precisión de una secuencia de proyecciones generadas por dichos modelos respecto a los datos observados *ex post* (para los modelos).

La motivación básica para realizar esta evaluación puede resultar demasiado evidente para un banco central: buenas proyecciones llevan a buenas decisiones (Diebold & López(1996)). Sin embargo, es importante enfatizar que, en dominios donde las consecuencias financieras de los errores de predicción son críticos, la evaluación de las fortalezas y las debilidades de los modelos utilizados para generar proyecciones resulta fundamental (Pollock *et al* (2005)). Más aún, incluso en ámbitos académicos se acepta con objetividad que es inútil disponer de modelos teóricamente sofisticados, y con un ajuste aparente en la muestra de estimación, para generar proyecciones para diversos horizontes (pues siempre es posible generarlas) cuando no se calcula su precisión (RECME) ni se compara con la de modelos alternativos usados como referencia.

La evaluación que sigue parte del criterio determinístico de la raíz del error cuadrático medio ejecutado (RECME), para luego presentar las pruebas estadísticas sobre las principales propiedades de las proyecciones óptimas, sobre la comparación del RECME entre los pares de modelos competitivos y finalmente sobre la dirección del cambio en las proyecciones (*market timing*).

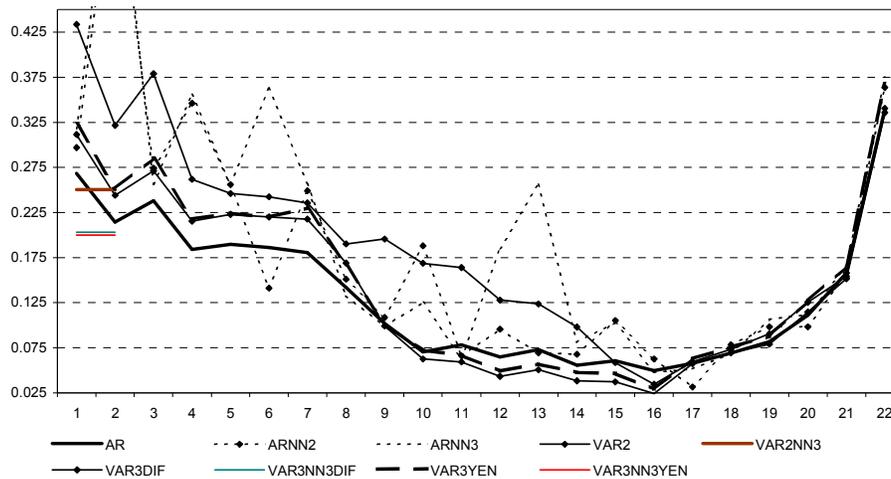
⁹ Se espera probar un modelo con 24 rezagos (sugeridos por los mejores modelos anteriores) tan pronto como el excesivo uso de tiempo de computadora que su estimación requiere pueda ser cubierto.

RECME como una medida de precisión determinística

El RECME es el criterio más objetivo de evaluación cuando se calcula para una muestra particular de datos observados *ex post* (muestra de evaluación), descartando tratarlo como un parámetro teórico o poblacional que deba estimarse. Una ventaja del RECME es que puede ser descompuesto en la suma de una medida de variabilidad y una medida de errores sistemáticos (la varianza del error de predicción y el sesgo). La única desventaja del RECME es que depende de todos los segundos momentos de la distribución conjunta de la proyección $y_{t+h,t}^F$ y el dato ejecutado y_{t+h} , por lo que un estimador de esta distribución puede ser más informativo (aunque sólo cuando la muestra de evaluación disponible es considerablemente amplia).

Es cierto que existe una multiplicidad teórica de funciones de pérdida y pruebas estadísticas en la literatura sobre evaluación de proyecciones (Diebold & Mariano(1995), Diebold & Lopez(1996)). En primer lugar, la elección de una función de pérdida diferente a la RECME debe estar justificada explícitamente. En segundo lugar, la variedad de pruebas estadísticas es reflejo de la variedad de hipótesis mantenidas, de las cuales la menos considerada es que el proceso generador de los datos es conocido, lo que ha sido aceptado recientemente (McCracken(2007)). Por esta razón, una vez calculado el RECME, no es necesario utilizar todos los criterios poblacionales y las pruebas estadísticas existentes, más aún cuando algunas requieren incurrir en costos computacionales prohibitivos¹⁰.

Gráfico No. 8
SECUENCIAS DEL RECME(h) DEL CONJUNTO DE MODELOS



¹⁰ Existen simulaciones que muestran que al modificar la muestra de estimación se puede concluir que el RECM teórico estimado para un modelo cambia tanto que al compararse con el de otros modelos, sus ordenamientos también varían, desvirtuando cualquier pretensión respecto al propio RECM teórico y por ende respecto al conocimiento del proceso generador de los datos (). Lo mismo se aplica a cualquier otra función de pérdida alternativa, pero con el agravante que tendrá que justificarse la elección de la misma por sobre el RECME.

El Gráfico No. 8 muestra que las ventajas de la modelación no lineal (VAR3YEN) en términos de precisión pueden llegar hasta un horizonte de entre 10 y 15 días útiles adelante (aunque muy similares al VAR2 y al AR), pero especialmente sobresalen por su precisión 1 día adelante (VAR3NN3YEN, VAR3NN3DIF y VAR2NN3). Ello se explica por su eficiencia en términos de procesar la información no lineal asociada al SPREAD¹¹.

Evaluación estadística de modelos individuales

Además del *ranking* de RECME, la evaluación estadística individual se refiere a las propiedades principales asociadas a proyecciones óptimas, propiedades que dependen de h (ver Diebold & Lopez (1996)). En primer lugar, para h=1, que tienen media cero y que son ruido blanco (independientes) y en segundo lugar, para h>1, que tienen media cero y que son, a lo más, procesos MA(h-1)¹².

- La prueba del signo y la prueba del ordenamiento de signos de Wilcoxon: son pruebas no-paramétricas para la hipótesis nula de que los errores de predicción tienen la mediana cero y no tienen supuestos mantenidos sobre la distribución de los errores $e_{t+h,t} \equiv y_{t+h} - y_{t+h,t}^F$ más allá de independencia serial¹³, lo que hace robustas estas pruebas a una variedad de distribuciones y a la heterocedasticidad de errores independientes y no distribuidos idénticamente, al menos para h=1.
- Sin embargo, incluso los errores óptimos mostrarán correlación serial para h>1, por lo que debemos asumir que los errores tienen una dependencia de orden h-1 (a lo más), lo que significará que las siguientes series deben ser independientes:

$$\begin{aligned} & \{e_{1+h,1}, e_{1+2h,1+h}, e_{1+3h,1+2h}, \dots\} \\ & \{e_{2+h,2}, e_{2+2h,2+h}, e_{2+3h,2+2h}, \dots\} \\ & \{e_{3+h,3}, e_{3+2h,3+h}, e_{3+3h,3+2h}, \dots\} \\ & \vdots \\ & \{e_{2h,h}, e_{3h,2h}, e_{4h,3h}, \dots\} \end{aligned}$$

¹¹ Debe notarse que al aumentar el horizonte h, las comparaciones entre los RECME(h) son indicativas pues el número de errores de predicción ejecutados utilizados en su cálculo disminuye.

¹² Diebold & Lopez(1996) establecen estas propiedades para modelos de proyección lineales. La tercera propiedad es que la varianza poblacional del error de proyección (óptimo) es no decreciente en el horizonte de proyección h. Evidentemente, las propiedades de los errores de proyección óptimos dependen de las características del modelo de proyección sólo si éste es el proceso generador de los datos, por lo que sólo las dos primeras propiedades son importantes para cualquier modelo mantenido.

¹³ Sin embargo, la hipótesis nula de la prueba Wilcoxon asume simetría de las distribuciones.

Esto significa que para $h=2$ obtenemos dos pruebas del signo y dos pruebas de Wilcoxon porque en este caso las siguientes dos series deben ser independientes:

$$\{e_{3,1}, e_{5,3}, e_{7,5}, \dots\}$$

$$\{e_{4,2}, e_{6,4}, e_{8,6}, \dots\}$$

mientras que para $h=3$, obtenemos tres pruebas del signo y tres pruebas de Wilcoxon, cada par asociado a las siguientes tres series que deben ser independientes:

$$\{e_{4,1}, e_{7,4}, e_{10,7}, \dots\}$$

$$\{e_{5,2}, e_{8,5}, e_{11,8}, \dots\}$$

$$\{e_{6,3}, e_{9,6}, e_{12,9}, \dots\}$$

y es evidente que la muestra de errores en estas series disminuye rápidamente a medida que h crece, lo que reduce el poder de las pruebas.

Tabla No. 1
PRUEBAS INDIVIDUALES PARA H_0 : MEDIANA=0
 (NO-PARAMETRICAS, P-VALUES)

		AR	ARNN2	ARNN3	VAR2 (SPD)	VAR2NN3 (SPD)	VAR3 (SPD-DIF)	VAR3NN3 (SPD-DIF)	VAR3 (SPD-YEN)	VAR3NN3 (SPD-YEN)
h=1	Sign	0.33491	0.33491	0.19688	0.33491	0.33491	0.50000	0.50000	0.33491	0.33491
	Wilcoxon	0.36659	0.22275	0.16104	0.35444	0.44192	0.36659	0.48058	0.28497	0.44192
h=2 (impar)	Sign	0.03141	0.08622	0.13214	0.08622	0.35491	0.05343	0.26757	0.08622	0.26757
	Wilcoxon	0.15168	0.03748	0.14562	0.11779	0.48827	0.20607	0.33194	0.11023	0.35598
h=2 (par)	Sign	0.03141	0.08622	0.08622	0.05343	0.03141	0.08622	0.13214	0.08622	0.01749
	Wilcoxon	0.15477	0.11397	0.01417	0.12707	0.00318	0.18795	0.19149	0.12303	0.02224

La Tabla No. 1 muestra los valores de probabilidad de estas pruebas para $h=1$ y $h=2$. Todas las pruebas son de una cola y (salvo un caso) no pueden rechazar la hipótesis de que la mediana es cero para un nivel de significancia de 1 por ciento.

Evaluación estadística de la precisión de proyecciones competidoras

Como se mencionó, una primera evaluación de la precisión de las proyecciones es comparar los RECME(h) para cada horizonte h . Sin embargo, la naturaleza estadística de todo estimador implica la necesidad de controlar por un error estadístico en los RECME(h) estimados.

Siguiendo a Diebold & Mariano (1995), todos los supuestos mantenidos posibles¹⁴ restringen sobremanera la relevancia de la prueba F . Para relajar el supuesto de correlación contemporánea, Granger y Newbold aplican la transformación octogonal de Morgan, definen los vectores $x(h) = e^i(h) + e^j(h)$ y $z(h) = e^i(h) - e^j(h)$ y

¹⁴ Que la función de pérdida es cuadrática y que los errores de predicción de dos modelos (ordenados como dos vectores $\{e^i(h), e^j(h)\}$ de dimensión $T \times 1$) tienen media cero, son normales y no correlacionados (serial y contemporáneamente).

demuestran que la hipótesis nula de igual precisión predictiva es equivalente a la hipótesis nula de correlación cero entre $x(h)$ y $z(h)$. El estadístico de la prueba de hipótesis es

$$MGN_{calc}(h) = \hat{\rho}_{xz}(h) / \sqrt{[1 - \hat{\rho}_{xz}^2(h)] / (T - 2)}$$

tiene una distribución t de Student con T-2 grados de libertad y el estimador de la correlación entre $x(h)$ y $z(h)$. es $\hat{\rho}_{xz}(h) \equiv \bar{x}'(h)\bar{z}(h) / \sqrt{[\bar{x}'(h)\bar{x}(h)][\bar{z}'(h)\bar{z}(h)]}$, donde cada vector se expresa en desviaciones respecto a la correspondiente media muestral. Sin embargo, la distribución del coeficiente de correlación muestral puede derivarse a partir del teorema del límite central más apropiado, la aplicación del método delta y la transformación z de Fisher aplicada a la correlación $\rho(h)$,

$$Z_{\hat{\rho}_{xz}}(h) = \arctan h(\hat{\rho}_{xz}(h)) = \frac{1}{2} \log \left(\frac{1 + \hat{\rho}_{xz}(h)}{1 - \hat{\rho}_{xz}(h)} \right)$$

con la ventaja de que converge más rápidamente a la distribución normal estándar que la distribución anterior para el coeficiente de correlación muestral (Bond & Richardson(2004)),

$$\sqrt{T-2} [Z_{\hat{\rho}_{xz}}(h) - Z_{\rho_{xz}}(h)] \xrightarrow{D} N(0,1)$$

La Tabla No. 2 presenta los valores de probabilidad de estas pruebas para horizontes de hasta 3 días adelante. Se confirma que algunas de estas pruebas pueden rechazar la hipótesis de igualdad de precisión (con un nivel de significancia de 1 por ciento) justo en los casos en los que su desempeño es evidentemente superado por el mínimo RECM (resaltados en negra).

Tabla No. 2
PRUEBAS DE MORGAN-GRANGER-NEWBOLD COMPARATIVAS PARA Ho: IGUALES RECM(h)
(Student-t & Aproximación z)

	AR	ARNN2	ARNN3	VAR2 (SPD)	VAR2NN3 (SPD)	VAR3 (SPD-DIF)	VAR3NN3 (SPD-DIF)	VAR3 (SPD-YEN)	VAR3NN3 (SPD-YEN)
RECM(h=1)	0.26847	0.29706	0.31785	0.31164	0.25033	0.43366	0.20312	0.32294	0.20018
[t-pval]	0.14815	0.16484	0.14089	0.07347	0.01651	0.01921	0.45487	0.06857	Mínimo
[z-pval]	0.14402	0.16085	0.13672	0.06921	0.01389	0.01640	0.45431	0.06434	RECM
RECM(h=2)	0.21437	0.62184	0.66067	0.24423	0.92352	0.32167	0.72662	0.25143	0.77290
[t-pval]	Mínimo	0.00227	0.00202	0.11975	0.00030	0.03418	0.00203	0.12587	0.00105
[z-pval]	RECM	0.00144	0.00126	0.11524	0.00013	0.03045	0.00126	0.12138	0.00058
RECM(h=3)	0.23792	0.16159	0.25619	0.27108	15.95136	0.37889	2.05728	0.28520	3.35051
[t-pval]	0.06654	Mínimo	0.09356	0.03876	0.00000	0.01349	0.00000	0.03666	0.00000
[z-pval]	0.06187	RECM	0.08874	0.03467	0.00000	0.01088	0.00000	0.03264	0.00000

Sin embargo, muchas de estas pruebas también significan que las ganancias en precisión determinísticas del Gráfico No. 1 son estadísticamente iguales a cero (en

particular, todas las pruebas para $h=1$). Estos resultados favorecen una interpretación determinística del $RECME(h)$.

Evaluación de la dirección del cambio implicado en las proyecciones

La prueba para las predicciones de cambio más importante es la de Henriksson & Merton (1990). El contexto original de esta prueba es la predicción del retorno $Z_{t+h,t}^M$ de un portafolio de acciones por encima del retorno $R_{t+h,t}$ de un portafolio de bonos (activos con rendimiento prefijado). Sin embargo, en vez de buscar predecir cuantitativamente este exceso en un modelo explícito, para construir una prueba no-paramétrica, se asume que los modelos únicamente predicen si el retorno de un activo será mayor o menor que el retorno del activo sin riesgo¹⁵. Si los modelos predicen estos eventos sistemáticamente, serán útiles para indicar la oportunidad de invertir en el mercado (*market timing*).

Definamos la variable que predice el *market timing*, $\gamma_{t+h,t}$ como

$$\gamma_{t+h,t} = \begin{cases} 1 & Z_{t+h,t}^M > R_{t+h,t} \\ 0 & Z_{t+h,t}^M \leq R_{t+h,t} \end{cases}$$

y sus probabilidades condicionadas a la realización de los eventos relevantes son

$$\begin{aligned} p_{t+h}^1 &= P[\gamma_{t+h,t} = 0 \mid Z_{t+h,t}^M \leq R_{t+h,t}] \\ 1 - p_{t+h}^1 &= P[\gamma_{t+h,t} = 1 \mid Z_{t+h,t}^M \leq R_{t+h,t}] \\ p_{t+h}^2 &= P[\gamma_{t+h,t} = 1 \mid Z_{t+h,t}^M > R_{t+h,t}] \\ 1 - p_{t+h}^2 &= P[\gamma_{t+h,t} = 0 \mid Z_{t+h,t}^M > R_{t+h,t}] \end{aligned}$$

es decir, p_{t+h}^j es la probabilidad condicional de una predicción correcta para el periodo $t+h$.

El supuesto de que las probabilidades de una predicción cualitativa correcta no dependen de la magnitud de $|Z_{t+h,t}^M - R_{t+h,t}|$, sino más bien del signo de esta diferencia permite deducir que $p_{t+h}^1 + p_{t+h}^2$ es un estadístico suficiente para evaluar la capacidad de predicción cualitativa. Bajo este considerando, una condición necesaria y suficiente para que estas predicciones cualitativas tengan valor o utilidad para indicar la oportunidad de invertir en el mercado (*market timing*) es que $p_{t+h}^1 + p_{t+h}^2 \neq 1$.

¹⁵ Nótese que esta simplificación es inocua y sin pérdida de generalidad pues cualquier proyección cuantitativa puede transformarse en una proyección cualitativa.

Usando esta condición poblacional puede construirse una prueba cuya hipótesis nula sea que las proyecciones no tienen valor en el sentido mencionado, $H_0: p_{t+h}^1 + p_{t+h}^2 = 1$ ¹⁶. Para ello determinamos la probabilidad de que una realización de la muestra disponible de proyecciones cualitativas satisfaga esta nula, lo que se define las siguientes variables:

- N_1 : el número de observaciones donde $Z_{t+h,t}^M \leq R_{t+h,t}$;
- N_2 : el número de observaciones donde $Z_{t+h,t}^M > R_{t+h,t}$;
- $N = N_1 + N_2$: el número total de observaciones;
- n_1 : el número de proyecciones **correctas** de $Z_{t+h,t}^M \leq R_{t+h,t}$;
- n_2 : el número de proyecciones **incorrectas** de $Z_{t+h,t}^M > R_{t+h,t}$; y
- $n = n_1 + n_2$: el número de proyecciones en las que $Z_{t+h,t}^M \leq R_{t+h,t}$.

Según la definiciones anteriores, $p_{t+h}^1 = E[n_1 / N_1]$ y $1 - p_{t+h}^2 = E[n_2 / N_2]$. Según H_0 , $E[n_1 / N_1] = E[n_2 / N_2] \Leftrightarrow p_{t+h}^1 = 1 - p_{t+h}^2 \Leftrightarrow p_{t+h}^1 + p_{t+h}^2 = 1$. Pero tenemos otro estimador para p_{t+h}^1 : $E[n / N] = p_{t+h}^1 \equiv p$. Bajo la nula, n_1 / N_1 y n_2 / N_2 tienen el mismo valor esperado, que denominamos p , y ambos son realizaciones de sub-muestras independientes, por lo que cualquiera de los dos puede usarse para estimar p .

Por su parte, n_1 y n_2 son sumas de variables aleatorias independientes e idénticamente distribuidas con distribución binomial, por lo que la probabilidad de que $n_j = x$ en una sub-muestra de tamaño N_j puede escribirse como

$$P[n_j = x | N_j, p] = \binom{N_j}{x} p^x (1-p)^{N_j-x} \quad j=1,2$$

Dada H_0 , podemos usar el teorema de Bayes para determinar la probabilidad de que $n_1 = x$ dados los parámetros, es decir, $P[n_1 = x | N_1, N_2, n, p]$. Para ello definimos los siguientes eventos:

- Evento A: el modelo predice m veces que $Z_{t+h,t}^M \leq R_{t+h,t}$ (es decir, $n = m$).
- Evento B: de las m veces que predice $Z_{t+h,t}^M \leq R_{t+h,t}$, el modelo predice correctamente x veces e incorrectamente $m-x$ veces (es decir, $n_1 = x$ y $n_2 = m - x$).

¹⁶ En esencia, esta es una prueba no-paramétrica de independencia entre la proyección cualitativa y los eventos cualitativos asociados a la diferencia de rendimientos, una prueba que utiliza el hecho de que las probabilidades condicionales asociadas a una predicción cualitativa correcta son estadísticos suficientes para medir su habilidad predictiva y sin embargo no dependen de la distribución de los retornos cuantitativos (por el supuesto mantenido de que ambas probabilidades no dependen de la magnitud de $|Z_{t+h,t}^M - R_{t+h,t}|$), sea ésta una distribución empírica o una distribución basada en un particular modelo teórico de valuación de los precios de los activos.

Entonces, $P[n_1 = x | N_1, N_2, n, p] = P[B | A]$ y por el teorema de Bayes,

$$P[B | A] = \frac{P[A \cap B]}{P[A]} = \frac{P[B]}{P[A]} = \frac{\binom{N_1}{x} \binom{N_2}{m-x} p^x (1-p)^{N_1-x} p^{m-x} (1-p)^{N_2-m+x}}{\binom{N}{m} p^m (1-p)^{N-m}} = \frac{\binom{N_1}{x} \binom{N_2}{m-x}}{\binom{N}{m}}$$

Bajo H_0 , $P[n_1 = x]$ tiene la forma de una distribución hipergeométrica y es independiente de p_1 y p_2 . Por lo tanto, para probar H_0 no se requiere estimar estas probabilidades sino disponer de las variables observables N_1, N_2, n que determinan la distribución $P[n_1 = x]$ bajo H_0 , que está determinada por $P[B | A]$ en el intervalo factible para n_1

$$\max\{0, n - N_2\} \equiv \underline{n}_1 \leq n_1 \leq \bar{n}_1 \equiv \min\{N_1, n\}$$

Una prueba estándar de dos colas para la hipótesis H_0 de que no existe habilidad predictiva y con un nivel de confianza c pre-determinado rechazará H_0 si $n_1 \geq \bar{x}(c)$ ó si $n_1 \leq \underline{x}(c)$, donde \bar{x} es el menor valor para el cual la sumatoria en la siguiente ecuación no excede $(1-c)/2$,

$$\sum_{x=\bar{x}}^{\bar{n}_1} \binom{N_1}{x} \binom{N_2}{n-x} / \binom{N}{n} = (1-c)/2$$

y \underline{x} es el mayor valor para el cual la sumatoria en la siguiente ecuación no excede $(1-c)/2$,

$$\sum_{x=\underline{x}}^x \binom{N_1}{x} \binom{N_2}{n-x} / \binom{N}{n} = (1-c)/2$$

(estas ecuaciones no pueden cumplirse con exactitud para un valor de c pre-determinado porque la distribución hipergeométrica es discreta).

Sin embargo, si los modelos no cometen errores sistemáticos (son “racionales”), nunca será verdad que $p_1 + p_2 < 1$ y un valor pequeño para n_1 significaría la “suerte de la realización”, sin importar qué tan poco probable sea. Entonces una prueba de una cola, la derecha (cola con mayor peso), es más conveniente y, para un nivel de confianza c pre-determinado, H_0 sería rechazada H_0 si $n_1 \geq x^*(c)$, donde este límite es el menor valor para el cual la sumatoria en la siguiente ecuación no excede $(1-c)$,

$$\sum_{x=x^*}^{\bar{n}_1} \binom{N_1}{x} \binom{N_2}{n-x} / \binom{N}{n} = (1-c)$$

y es evidente que $x^*(c) < \bar{x}(c)$, por lo que dado un valor calculado en la cola derecha, una prueba de una cola es más conservadora, es decir, rechazará H_0 más veces que una prueba de 2 colas, para cualquier valor prefijado de c .

La Tabla No. 3 presenta los valores de probabilidad de esta prueba para horizontes de hasta 17 días. Con un nivel de significancia de 10 por ciento no puede rechazarse que los modelos no lineales ARNN2, ARNN3, VAR2NN3, VAR3NN3DIF y VAR3NN3YEN tengan valor predictivo para capturar el patrón cualitativo (*market timing*) del tipo de cambio diario en un horizonte de hasta 10 días adelante. Estos resultados son compatibles con los resultados obtenidos gráficamente.

Tabla No. 3
PRUEBAS DE HENRIKSSON & MERTON PARA H_0 : NO HAY VALOR PREDICTIVO
 (NO-PARAMÉTRICAS, P-VALUES)

h	AR	ARNN2	ARNN3	VAR2 (SPD)	VAR2NN3 (SPD)	VAR3 (SPD-DIF)	VAR3NN3 (SPD-DIF)	VAR3 (SPD-YEN)	VAR3NN3 (SPD-YEN)
1	0.10991	0.20636	0.33496	0.63932	0.18492	0.33496	0.01499	0.63932	0.09919
2	0.42498	0.57502	0.71648	0.57502	0.57502	0.24494	0.75506	0.42498	0.75506
3	0.78328	0.18492	0.11920	0.89822	0.11920	0.84308	0.22136	0.65754	0.11920
4	0.70338	0.79174	0.55057	0.92896	0.66337	0.44943	0.79174	0.55057	0.14380
5	0.57240	0.50000	0.41686	0.86275	0.04739	0.72172	0.16742	0.41686	0.11652
6	0.66128	0.75339	0.24661	0.90126	0.15950	0.48416	0.07466	0.79460	0.03733
7	0.86126	0.63462	0.23077	0.86126	0.07692	0.50000	0.63462	0.96429	0.05769
8	0.83317	0.57343	0.81818	0.83317	0.18182	0.28671	0.00000	0.71329	0.04196
9	0.65734	0.50000	0.23077	0.50000	0.54545	0.34266	0.23077	0.65734	0.54545
10	0.58741	0.00000	0.00000	0.24942	0.19580	0.41259	0.31546	0.24942	0.09790
11	0.87879	0.24747	0.24747	0.50000	0.15909	0.31061	0.63636	0.50000	0.24747
12	0.39177	0.57576	0.34848	0.01515	0.39177	0.60823	0.54545	0.17532	0.80303
13	0.73810	0.33333	0.33333	0.26190	0.07143	0.45238	0.16667	0.45238	0.07143
14	0.35714	0.03968	0.35714	0.16667	0.16667	0.59524	0.16667	0.03968	0.64286
15	0.00000	0.50000	0.17857	0.17857	0.17857	0.35714	0.35714	0.17857	0.71429
16	0.00000	0.00000	0.00000	0.02857	0.14286	0.57143	0.28571	0.02857	0.37143
17	0.00000	0.40000	0.40000	0.06667	0.40000	0.00000	0.40000	0.33333	0.40000

Para horizontes mayores a 10 días adelante, los modelos lineales pueden tener valor predictivo en el sentido de estas pruebas, lo que resulta compatible con los resultados obtenidos en el Gráfico 1 (aunque para estos horizontes el número de errores de predicción ejecutados disponibles es menor).

4. Conclusiones

Los resultados obtenidos confirman que los perceptrones poseen la capacidad para anticipar los movimientos diarios en el tipo de cambio. Por su parte, los modelos ARFIMA(p,d,0) resultan demasiado costosos en términos de tiempo de computadora para los resultados tentativos inicialmente obtenidos y la restricción de utilidad de los modelos por construir¹⁷.

Para contrastar el desempeño predictivo del grupo de modelos disponible se utiliza los errores de predicción cometidos desde agosto hasta Setiembre 2008 para estimar secuencias de la raíz del error de predicción cuadrático medio (RECM(h), $h \leq 15$). Este criterio se elige para evaluar la precisión de las proyecciones y evitar

¹⁷ Que puedan re-estimarse diariamente para generar las proyecciones.

la subjetividad presente en algunas medidas utilizadas en la literatura, como aquellas que usan funciones de pérdida para reflejar las preferencias sobre la precisión demasiado específicas de algunos inversionistas. Complementariamente, se considera otras pruebas estadísticas para comparar las bondades predictivas¹⁸.

Las ventajas de los perceptrones (VAR3YEN) en términos de precisión pueden llegar hasta un horizonte de entre 10 y 15 días útiles adelante (aunque muy similares al VAR2 y al AR), pero especialmente sobresalen por su precisión 1 día adelante (VAR3NN3YEN, VAR3NN3DIF y VAR2NN3). Ello se explica por su eficiencia en términos de procesar la información no lineal asociada al SPREAD.

Todas las pruebas presentadas para la hipótesis de una mediana nula no pueden rechazar la hipótesis de que la mediana es cero para un nivel de significancia de 1 por ciento.

Los resultados de las pruebas para la comparación de la precisión de las proyecciones para horizontes de hasta 3 días adelante confirman que algunas de estas pruebas pueden rechazar la hipótesis de igualdad de precisión justo en los casos en los que su desempeño es evidentemente superado por el mínimo RECME. Sin embargo, muchas de estas pruebas también significan que las ganancias en precisión determinísticas del Gráfico No. 8 son estadísticamente iguales a cero (en particular, todas las pruebas para $h=1$). Estos resultados favorecen una interpretación determinística del RECME(h).

Respecto a la prueba de valor predictivo, con un nivel de significancia de 10 por ciento no puede rechazarse que los modelos no lineales ARNN2, ARNN3, VAR2NN3, VAR3NN3DIF y VAR3NN3YEN tengan valor predictivo para capturar el patrón cualitativo (*market timing*) del tipo de cambio diario en un horizonte de hasta 10 días adelante. Estos resultados son compatibles con los resultados obtenidos gráficamente en la sub-sección 3.1.

Para horizontes mayores a 10 días adelante, los modelos lineales pueden tener valor predictivo en el sentido de estas pruebas, lo que resulta compatible con los resultados obtenidos en el Gráfico 8 (aunque para estos horizontes el número de errores de predicción ejecutados disponibles es menor).

¹⁸ Asimismo, se prevé comparar el desempeño de estos modelos uni-variados con modelos similares que utilicen datos como desviaciones respecto a las tendencias locales, así como con versiones bi-variadas que incluyan información de variables individuales con potencial influencia sobre el tipo de cambio y disponibles en frecuencia diaria, como variables financieras, precios de *commodities*, etc. Estos modelos bi-variados serán utilizados para construir proyecciones combinadas como alternativa a la construcción de modelos multi-variados más complejos.

Anexo A: Descripción de datos, transformaciones y misceláneas

Datos del tipo de cambio

Los datos del tipo de cambio utilizados para estimar los modelos descritos corresponden a las cotizaciones horarias que diariamente están disponibles en Reuters hacia las 2 p.m., específicamente el tipo de cambio utilizado es el promedio de las cotizaciones del dólar para la compra y la venta en el mercado interbancario peruano a las 11 a.m. de cada día útil de la semana (FX).

Otros variables (indicadores anticipados)

VARIABLES asociadas cuyo comportamiento puede anticipar los movimientos en este tipo de cambio en el futuro son:

- el *spread cambiario* (SPREADC), definido como la diferencia entre las cotizaciones del dólar para la compra y la venta a las 11 a.m. de cada día útil de la semana, diferencia que es expresada como porcentaje del tipo de cambio promedio de ambas cotizaciones.
- el *spread financiero* (SPREADF), definido como el diferencial de tasas de interés interbancarias *overnight*.
- el tipo de cambio yen/dólar (YENUS), diariamente publicado por el Banco de la Reserva Federal de Estados Unidos.
- el índice multilateral de monedas de principales socios comerciales /dólar (MFXUS), diariamente publicado por el Banco de la Reserva Federal de Estados Unidos.
- **las compras netas de dólares (COMP), efectuadas por el Banco Central de Reserva del Perú en la mesa de negociaciones.**

Transformaciones a los datos

Las únicas transformaciones realizadas a los tipos de cambio utilizados mencionados son las requeridas para obtener el *spread* compra-venta o los retornos cambiarios (las variaciones porcentuales diarias). En este sentido estas variables transformadas son los datos primarios (tanto como por ejemplo podría ser el caso de la tasa de desempleo en macroeconomía aplicada).

Misceláneas

El lector atento se preguntará sobre el tratamiento de los fines de semana, feriados nacionales y feriados extranjeros, días para los cuales no se disponen de datos, por ejemplo, de cotizaciones de compra o venta para el tipo de cambio interbancario. Aunque los modelos podrían ser estimados utilizando únicamente

los “datos duros”, esta primera opción elimina no sólo la homogeneidad temporal que debería existir con la sucesión natural de los días sino también la escala de los retornos cambiarios¹⁹, por lo que con esta opción la estimación tendría que modificarse para considerar la irregularidad temporal de los “datos duros” (p.ej., Asai & McAleer(2007)), lo que sin embargo no soluciona la asociada irregularidad en la escala de los retornos cambiarios. Una segunda opción es que la estimación de modelos considere los “datos perdidos” como coeficientes aleatorios desconocidos a ser estimados conjuntamente con los parámetros usuales (p.ej., Kim(2005)).

Estas dos opciones pueden ser importantes *per se* pero son completamente secundarias si la evaluación *ex post* de las proyecciones se realiza únicamente con los “datos duros”. Este es el caso del presente estudio, por lo que detallamos aquí el procedimiento seguido para obtener los datos usados en la estimación y en la evaluación *ex post*:

- Muestra de estimación
 - Los datos para los fines de semana, los feriados nacionales y los feriados extranjeros se obtienen mediante una interpolación simple de los niveles. Por ejemplo, la cotización de una moneda entre dos días útiles se considera generada suavemente a partir de los dos extremos disponibles (sin saltos). De esta manera, las variaciones porcentuales diarias mantienen una escala homogénea.
 - La muestra de datos ampliada de esta manera es la que finalmente se transforma como se mencionó arriba, dando lugar a la muestra de estimación de los modelos. De esta manera se obtiene la homogeneidad temporal que facilita la estimación de los modelos arriba descritos.
 - Se crea una variable dicotómica que permite separar los “datos duros y cuasi-duros” de los datos puramente generados por interpolación en la muestra de estimación²⁰.
- Muestra de evaluación *ex post*
 - En contraste con la muestra de estimación, ya no es necesario que la muestra de evaluación tenga una homogeneidad temporal de las observaciones ni una homogeneidad en la escala de las variaciones porcentuales. Para evaluar objetivamente la precisión de las proyecciones de los modelos, se debe usar los “datos duros” (en niveles o en variaciones). El único costo de ello es que la dimensión temporal de la secuencia del RECM, usualmente indicado con el índice h , ya no es una simple extensión de la dimensión temporal de la muestra de estimación, indicado con el índice t (ver Anexo B).

¹⁹ Por ejemplo, con un día feriado entre dos días útiles, la variación porcentual del tipo de cambio entre estos últimos corresponde al retorno de dos días calendario en vez de sólo a uno. Para mantener la escala de los retornos, se tendría que calcular la variación porcentual diaria equivalente.

²⁰ Esta variable dicotómica puede ser incluida en el conjunto de información de los modelos y sus observaciones fuera de la muestra se generan de manera determinística.

Anexo B: Cambios realizados en el programa de evaluación *ex post*

El procedimiento típico de evaluación *ex post* considera datos ejecutados compatibles con las secuencias proyectadas por los modelos bajo evaluación. Esta compatibilidad se logra porque la dimensión temporal de la secuencia del RECM, usualmente indicado con el índice h , es una simple extensión de la dimensión temporal de la muestra de estimación, indicado con el índice t . Por esta razón puede usarse $T+h$ como índice de las proyecciones.

Usar los “datos duros” como la muestra de evaluación significa que h ya no será compatible con t , lo que implica los siguientes cambios en el procedimiento²¹:

- Como se dispone de proyecciones en variaciones diarias compatibles con el índice t , simplemente se obtienen los niveles del tipo de cambio aplicando estas variaciones proyectadas al último nivel del tipo de cambio de la muestra de estimación (que por construcción siempre es un “dato duro”).
- De esta secuencia de niveles proyectados se selecciona los niveles que correspondan a días útiles (“datos duros”), es decir, se descarta los niveles del tipo de cambio correspondientes a los fines de semana y los días feriados. Con estos niveles se calculan las variaciones porcentuales “diarias”.
- Los “datos duros” se expresan como variaciones porcentuales simples calculadas sobre los niveles ejecutados para los días útiles de la muestra de evaluación.

Referencias

Asai, Manabu & McAleer, Michael (2007) Non-trading day effects in asymmetric conditional and stochastic volatility models, *Econometrics Journal*, Vol. 10, pgs. 113-123.

Bacchetta & Wincoop (2004)

Baillie, Richard & Chung, Ching-Fan (1996) Analyzing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics*, Vol. 11, pgs. 23-40.

Bond, Charles & Richardson, Ken (2004) Seeing the Fisher's Z transformation, *Psychometrika*, Vol 69, No. 2, pgs. 291-303.

²¹ Para el caso de la secuencia de RECM(h) en variaciones porcentuales entre días útiles.

Chung, Ching-Fan & Baillie, Richard (1993) Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models, *Empirical Economics*, Vol. 18, pgs. 791-806.

Cheung, Menzie & Pascual (2005)

Diebold, Francis & Mariano, Robert (1995) Comparing predictive accuracy, *Journal of Business & Economics Statistics*, Vol. 13, pgs. 253-263.

Diebold, Francis & López, José (1996) Forecast evaluation and combination, Department of Economics, University of Pennsylvania.

Dorffner, Georg (200?) Neural networks for time series processing, Working Paper, Dept. of Medical Cybernetics and Artificial Intelligence, University of Vienna.

Ebrahim, Shafiq (2000) Volatility transmission between foreign exchange and money markets, Working Paper 2000-16, Bank of Canada.

Engle, Robert & Kroner, Kenneth (1995) Multivariate simultaneous generalized ARCH, *Econometric Theory*, Vol. 11, pgs. 122-150.

Evans & Lyons(2002)

Henriksson, Roy & Merton, Robert (1981) On market timing and investment performance II: Statistical procedures for evaluating forecasting skills, *Journal of Business*, Vol. 54, No. 4, pgs. 513-533.

Groen, Jan (1999) Long horizon predictability of exchange rates: is it for real? *Empirical Economics*, Vol. 24, pgs. 451-469.

*Jeantheau, Thierry (1998) Strong consistency of estimators for multivariate ARCH models, *Econometric Theory*, Vol. 14, pgs. 70-86.

Kim, Jeongeun (2005) Parameter estimation in stochastic volatility models with missing data, PhD Thesis, Department of Statistics, University of Pittsburgh.

Kuan, Chung-Ming & Liu, Tung (1995) Forecasting exchange rates using feedforward and recurrent neural networks, *Journal of Applied Econometrics*, Vol. 10, pgs. 347-364.

Lyons (2001)

*MacDonald, Ronald & Taylor, Mark (1994) The monetary theory of exchange rate: long-run relationships, short-run dynamics and how to beat a random walk, *Journal of International Money and Finance*, Vol. 13, pgs. 276-290.

*Mark, Nelson (1995) Exchange rates and fundamentals: evidence on long-horizon predictability, *American Economic Review*, Vol. 85, pgs. 201-218.

McCracken, Michael (2007) Asymptotics for out of sample test of Granger Causality, *Journal of Econometrics*, Vol. 140, pgs. 719-752.

Meese, R. & Rogoff, K. (1983) Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics*, Vol. 14, pgs. 3-74.

Olgun, Hasan & Ozdemir, Zeynel (2008) Linkages between the center and the periphery stock prices: evidence from the vector ARFIMA model, *Economic Modelling*, Vol. 25, pgs. 512-519.

Pollock et al (2005)

Rech, Gianluigi; Terasvita, Timo & Tschernig, Rolf (2000) A simple variable selection technique for nonlinear models, *Stockholm School of Economics & Humbolt-Universitat zu Berlin*.

Shachmurova, Yochanan (2003) *Applying Artificial Neural Networks to Business, Economics and Finance*, Dept. of Economics, University of Pennsylvania.

Teräsvirta, Timo & Granger, Clive (1993): *Modelling nonlinear economic relationships*. Oxford: Oxford University Press.

Yang, Kun & Shintani, Mototsugu (2006) Does the prediction horizon matter for the forward premium anomaly? Evidence from panel data, *Economics Letters*, Vol. 93, Issue 2, pgs. 255-260 (noviembre).