

Estimación de la caída de los principales índices ante la suba de tasas



Daiana García Álvarez, Facundo García, Josué Yoon
Universidad de San Andrés

Trabajo final de Big Data

December 2018

Introducción

Los movimientos en la tasa de interés internacional pueden provocar fuertes shocks en los distintos mercados de la economía mundial. Los principales índices no son la excepción a la regla, por ende, para entender su comportamiento, debemos incluir a las tasas de interés internacionales dentro del análisis. A pesar de que es de público conocimiento el hecho de que ante una subida en las tasas los principales índices caerán, todavía no hay ningún modelo que pueda estimar con exactitud la magnitud de la caída del mercado. Lograr estimar esta caída es de suma importancia dado que los mercados emergentes, como la Argentina, se enfrentan ante una situación de vulnerabilidad macroeconómica cada vez que hay una subida de tasas dado que provocan una devaluación de las monedas locales. Estas caídas, fuerzan a los mercados emergentes a tener que subir las tasas locales para evitar la fuga de capitales.

El objetivo del presente trabajo es tratar de estimar la relación entre la magnitud del aumento porcentual de la tasa de interés de EEUU y la magnitud de la caída de los índices. Para ello, primero se presentará una metodología teórica tradicional para estimar esta relación, y luego, nos enfocaremos en presentar una metodología de Artificial Neural Networks que busque estimar dicha relación nuevamente. Esta metodología ha cobrado un nuevo “boom” en los últimos años y se sigue desarrollando e implementando en diversas áreas del conocimiento. Creemos que esta nueva metodología es de gran utilidad y que podría aportar mayor poder de predicción al modelo tradicional.

La presente propuesta se dividirá en las siguientes secciones. La primera sección introducirá una adaptación de un modelo teórico perteneciente a la teoría financiera que estima de esta relación. La segunda sección se enfocará en presentar una metodología computacional que buscara entrenar una red neural que logre predecir estos movimientos del mercado. Por último, y debido a la complejidad que presentan los modelos de redes y su programación, presentaremos un ejemplo y las posibles limitaciones del trabajo.

Modelo teórico

El modelo que se plantea a continuación es una adaptación del modelo de Gordon Pye (1974). Tanto el SP 500, el Dow Jones y el Nasdaq 100 tienen un price earning del orden de 22-23 . Asumimos que en promedio las empresas que forman parte de dichos índices están valuadas como un bono perpetuo que mantienen aproximadamente la misma relación precio ganancias. Es importante destacar que este supuesto solo vale para las empresas consolidadas tipo value, es decir para aquellas que alcanzaron su máximo potencial de crecimiento, no para aquellas que tienen todavía expectativas de ganancias crecientes (las denominadas del tipo grow). Si esto es así, la ecuación de estado es $P/G=22$; o lo que es lo mismo,

$$P = \frac{G}{(1/22)} = \frac{G}{(0,045)}$$

donde la última expresión indica que la tasa de interés implícita de esa relación G/P es de 4.5%. Actualmente la tasa del bono del tesoro americano de 30 años esta 3.40%. Eso quiere decir que el mercado de bonos y acciones esta en equilibrio con una sobretasa o spread de aproximadamente 1.0% de premio por el riesgo adicional que significa de invertir en acciones en lugar de bonos del tesoro.

¿Qué ocurrirá cuando la tasa de los bonos del tesoro suba medio punto al 3.92%? ¿Cuánto caerán las acciones? Si esto ocurriese ahora instantáneamente la respuesta seria simple. Asumiendo que el mercado sigue pidiendo el mismo spread, las acciones deberían rendir el 5.0% y, como sus ganancias esperadas siguen siendo las mismas, eso implica que su precio caería un 10% en forma proporcional al derrumbe de la tasa.

Ahora bien, ¿qué pasa si la tasa sube medio punto pero en el plazo de un año?. En ese caso, tenemos que tener en cuenta que la tasa sube como consecuencia de que la economía americana está creciendo a un ritmo, en promedio, de un 3% anual con pleno empleo. Por lo que las expectativas de ganancias de las empresas son crecientes. Si tenemos en cuenta que la torta de del crecimiento no se reparte proporcionalmente entre todos los agentes económicos, sino que las empresas se llevan una porción más grande que el resto, sus ganancias van a crecer de acá a una año como,

$$G(T + 1) = G(T)(1 + k * 0,03)$$

donde k es un factor que típicamente implica el crecimiento del PIB entre 2 y 3 veces (ver estimación de K en el Apéndice). Es decir, que si la suba de tasas ocurre dentro de un año las ganancias de las empresas van a estar por lo menos un 6 % arriba, con lo cual una caída instantánea del índice de un 10 % se amortiguaría del orden de 4 % si asumimos $k=2$:

$$P(T + 1, i = 5) = \frac{G(T + 1)}{0,05} P(T, i = 4,5) = \frac{G(T)}{0,045}$$

$$\frac{P(T + 1, 5)}{P(T, 4,5)} = (1 + k * 0,03) \frac{0,045}{0,05} = 0,954$$

y en forma generalizada,

$$\frac{P(T + 1, i_{t+1})}{P(T, i_t)} = \frac{G(T + 1)}{0,05} = (1 + k * g) \frac{i_t}{i_{t+1}} = (1 + k * g) \frac{i_0 t + \Delta}{i_{0t+1} + \Delta}$$

donde i_t es la tasa implícita de rendimiento del índice, $i_0 t$ es la tasa del bono del tesoro americano de 30 años, el spread o premio del índice y g es la tasa anual de crecimiento del PIB

Modelo de redes neuronales

Literatura relacionada

Existe una vasta cantidad de trabajos relacionados con el pronóstico de divisas y acciones utilizando redes neuronales. Esta metodología surge a partir del fracaso de los modelos de pronóstico tradicionales como los AR, ARMA y ARIMA dado que no capturan el comportamiento no lineal de las series temporales, es decir la no linealidad del valor de las acciones del mercado.

Dentro de la literatura de redes ANN aplicada a finanzas, se encuentran dos grandes enfoques: quienes buscan pronosticar el precio de una acción y quienes buscan pronosticar la rentabilidad de una acción. Sin embargo, En relación a la primera categoría se encuentra el trabajo de Cao et. al. (2005) que busca predecir el precio de las acciones del mercado de Shanghai utilizando como input el precio de las acciones. En

relación a la segunda categoría, se encuentra el de Zhu et al. (2008) que encuentran que las redes son útiles para pronosticar el movimiento de acciones tanto el corto como el largo plazo y entrenan a su red con las ganancias y el volumen operado. Por otro lado, también existen trabajos como el de Lin, Khan Chang (2008) que utilizan una Fuzzy Neural Network (FNN) para predecir la ocurrencia de las crisis de moneda.

Metodología

a. Introducción

A continuación utilizaremos la metodología expuesta en el trabajo de Velasquez, Fonegra y Villa (2013), para poder abordar el objetivo de nuestro trabajo. Se propone pronosticar el precio del índice utilizando una ANN del tipo RNA (Perceptron Multicapa) regularizada y con validación cruzada (CV).

Las redes de Perceptron Multicapa son utilizadas para series temporales porque tienen una amplia capacidad para aprender relaciones desconocidas a partir de un conjunto de ejemplos, una alta tolerancia a patrones extraños de ruido y componentes caóticas presentes en el conjunto de datos. También, son suficientemente robustas para procesar información incompleta o contaminada, y por sobre todo, no restringen el tipo de no linealidad de la serie de tiempo a la estructura matemática del modelo de red. Sin embargo, al igual que los modelos tradicionales, pueden presentar el problema de sobreajuste. Frente a este problema, Tikhonov introdujo la metodología de regularización, es decir establecer algún tipo de función que penalice a la función objetivo. Sin embargo, en el caso de redes el único problema no es seleccionar el método de regularización sino también decidir cuánto debe incidir la estrategia sobre el entrenamiento de la red. Una de las técnicas más utilizadas de regularización sobre de entrenamiento es la descomposición de pesos de Hinton. A continuación presentaremos el modelo de una red MLP.

b, Variables

Para nuestra propuesta utilizaremos, para entrenar la red MLP, las series de tiempo correspondientes al índice Dow Jones, y como único regresor el valor de la tasa de interés del Tesoro de Estados Unidos. Utilizaremos como umbral de variación de la tasa de interés los aumentos mayores o iguales a 0.25 %. También, se propone como una alternativa, dado que MLP puede operar bajo información incompleta, recortar la

serie de tiempo y dejar solamente los datos próximos temporalmente a una suba o baja de tasas correspondiente al umbral que definimos como una variación significativa.

c. Red MLP

Las redes MLP son un tipo de red que imita la estructura neural del cerebro. Es un conjunto de nodos que están lógicamente conectados en 3 o más capas, generalmente, posee una capa de entrada, una oculta y una capa de salida, y cada una de estas posee al menos una neurona. En el caso de MLP partimos de una de T observaciones ordenadas en el tiempo.

$$Y_T = [Y_i]_1^T \quad (1)$$

La cual se pretende estimar una función que permita explicar Y_T en función de sus rezagos, por lo tanto reescribimos la función Y_T ,

$$Y_T = \beta_* + \sum_{h=1}^H \beta_h * g \left(\alpha_h + \sum_{p=1}^P w_{p,h} * y_{t-p} \right) + \varepsilon_t \quad (2)$$

Esta última expresión es equivalente a un modelo estadístico no paramétrico de regresión no lineal donde t sigue una distribución normal con varianza cero, H representa el número de neuronas que se encuentran en la capa oculta, P es el número máximo de rezagos introducidos (neuronas de entrada), $g()$ es la función de activación de la capa oculta y los parámetros $(\beta_*, \beta_h, \alpha_h, w(p, h))$ son estimados usando el principio de máxima verosimilitud (minimización de ECM).

A partir de lo presentado, los pasos a seguir son:

- Seleccionar cuáles son el número de capas de entrada (rezagos)
- Determinar la cantidad de neuronas en la capa oculta.
- Seleccionar la función de activación.
- Seleccionar cuál es la función objetivo que se desea optimizar (ECM).
- Estimar los parámetros del modelo con alguna técnica de optimización.
- Evaluar la capacidad predictiva del modelo.

Cabe aclarar que, los parámetros óptimos no son únicos para cierto diseño elegido del modelo (especificación – es decir n grados de rezagos, n grados de neuronas en la capa oculta, función de activación, etc.).

d. Causas del sobre ajuste

Como mencionamos anteriormente, las redes MLP padecen del mismo problema que los modelos tradicionales: el sobreajuste. Este puede ser causado en primer lugar por la presencia de outliers en los rezagos de la capa de entrada, esto genera que la varianza de los parámetros de la red sea alta. En segundo lugar, el tamaño óptimo de la red, es decir si sobre parametrizamos la neurona (exceso de rezagos de entrada) esta memoriza los datos de entrenamiento en vez de aprender su comportamiento. Este problema se ve reflejado cuando tenemos un error de entrenamiento chico y un error de validación alto. En tercer lugar, que el subconjunto de entrenamiento no posee la cantidad suficiente de datos para lograr representar el proceso de la serie. Los primeros dos problemas se resuelven con la implementación de un método de regularización, en nuestro caso implementaremos el Weight Decay. El último problema se resuelve utilizando el método de validación cruzada (CV).

e. Regularización y Metodo Weight Decay

Como hemos visto a lo largo de la materia, los métodos de regularización buscan el equilibrio entre la minimización del ECM (maximización de ajuste) y la penalización de la complejidad.

$$R(W) = \sum_{t=1}^T (\hat{Y}_t - Y_t)^2 + \mu \varepsilon(W) \quad (3)$$

Donde $\mu \in (W)$ es el termino de regularización que está dado por la integral de suavizado de orden k y μ es el factor de regularización que controla la incidencia de (W) sobre el entrenamiento de la red donde (W) tiene la forma,

$$\varepsilon(w, k) = \frac{1}{2} \int \left\| \frac{\partial^k}{\partial w^2} F(w, m) \right\|^2 * \mu(w) dw \quad (4)$$

$F(w, m)$ es el mapeo de entrada-salida del modelo, (w) es una función que pondera el espacio sobre el cual se suavizara al función $F(w, m)$.

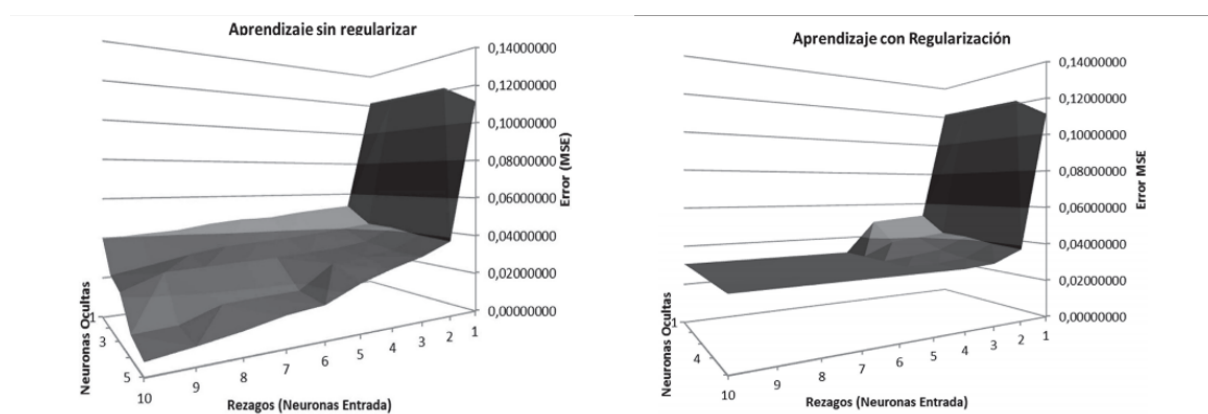
En conclusión, la regularización (4) es la penalización que se impone sobre la función MLP (3). A continuación describiremos la técnica elegida: la descomposición de pesos de Hinton.

La descomposición de pesos de Hinton consiste en realizar una discriminación donde algunos pesos sinápticos toman valores cercanos a cero y otros valores altos. Esto implica que agruparemos los pesos que no tienen ninguna influencia sobre el modelo (los cercanos a cero) y los pesos que tienen una alta influencia (pesos de exceso). En base a esto, la penalización de complejidad se define como,

$$\varepsilon(W) = \|w_{p,h}\|^2 = \sum_{h=1}^H \sum_{p=1}^P w_{p,h}^2 \quad (5)$$

donde, los pesos $w(p,h)$ son los pesos que están entre la capa de entrada p a la neurona h (pesos entre la capa de entrada y la oculta). Cabe aclarar que se parte del supuesto de que la distribución de los pesos está centrada en 0.

A continuación podemos observar gráficamente el efecto del método de regularización de Weight Decay,



Como se puede observar en el siguiente ejemplo tomado del trabajo de Velasquez, et al (2013). Al igual que en Ridge, Lasso, y MCO, el ECM decrece a medida que se aumentan las neuronas de la capa de entrada y la capa oculta. Como podemos observar el volumen se estabiliza cuando regularizamos, esto permite controlar el sobreajuste.

Por último, se realiza K-fold cross validation para estimar la división óptima de la muestra en datos de evaluación y datos de entrenamiento.

Conclusión preliminar

A modo de conclusión del presente trabajo, podemos destacar que el modelo tradicional que estima la magnitud de la variación de los índices principales y las tasas internacionales muestra un poder predictivo eficaz pero a su vez incompleto. Para ello, el modelo de redes neuronales podría ser un complemento útil a la hora de captar el comportamiento no lineal de los mercados en cuestión.

Si bien, a la hora de llevar el modelo a la práctica podemos encontrar diversos problemas como sobreajuste del modelo o alta varianza, creemos que el modelo predictivo a través de redes neuronales demostrará resultados más acordes a los comportamientos vistos en los distintos mercados internacionales.

Bibliografía

- X. Zhu, H. Wang, L. Xu, and H. Li, “Predicting stock index increments by neural networks: The role of trading volume under different horizons,” *Expert Systems with Applications*, vol. 34, no. 4, pp. 3043–3054, 2008. Q. Cao, K. B. Leggio, and M. J. Schniederjans, “A comparison between Fama and French’s model and artificial neural networks in predicting the Chinese stock market,” *Computers Operations Research*, vol. 32, no. 10, pp. 2499–2512, 2005. J. D. Velasquez, Y. Fonnegra, F. A. Villa, “Pronóstico de series de tiempo con redes neuronales regularizadas y validación cruzada”, *Revista Vinculos*, vol. 10, pp. 267-293, 2013. G. Pye, “Gauging the Default Premium”, *Financial Analysts Journal*, vol 30, no. 1, pp. 49-52, 1974. C. Lin, H. A. Khan, Y. Chieh, R. Chang, “A new approach to modeling early warning systems for currency crisis: can a machine-learning fuzzy expert system predict the currency crises effectively?”, *Journal of International Money and Finance*, vol. 27, no. 7, pp. 1098-1121, 2006.

Apéndice

La relación entre el crecimiento anual de las acciones del Dow y el crecimiento del producto en los siguientes periodos fue,

Period	years	% Dow anual	% GDP anual	K=%Dow/%GDP
1982-1999	18	+15.9	+6.3	2.5
2000-2008	9	-1.5	+4.8	-3.2
2009-2015	7	+12.8	+2.7	4.7
1982-2015	34	+10.6	+5.1	2.07

El promedio historico de 34 años fue

$$\% \text{ Dow} = \frac{15.9 \times 18 - 1.5 \times 9 + 12.8 \times 7}{34} = 10.6$$

$$\% \text{ GDP} = \frac{6.3 \times 18 + 4.8 \times 9 + 2.7 \times 7}{34} = 5.1$$

Notemos que K mide el crecimiento de los precios cuando crece el GDP, solo si la tasa se mantiene constante el precio es proporcional a las ganancias y k es lo mismo que K: Uno podría argumentar que en un periodo de tiempo largo en donde la tasa fluctuó en torno a un promedio K me da una idea aproximada de k.