

Departamento de Economía – Universidad Nacional del Sur

Trabajo de Grado de la Licenciatura en Economía

“Modelos de selección de variables predictivas. Un análisis
sobre el índice Merval”

Alumno: Juan Horgan

Directores:

Dr. Fernando Delbianco

PhD. Andrés Fioriti

Noviembre, 2018.

Resumen

Desde que existen los mercados de acciones se ha tratado de encontrar métodos que permitan predecir los precios futuros de los valores, todos ellos sin éxito. Por ello, el presente trabajo no pretende encontrar un conjunto de variables que permitan estimar los precios futuros de un índice bursátil, pero al menos intenta probar si existe una relación de corto plazo entre el *MERVAL* y determinadas variables de frecuencia diaria. Las variables elegidas pueden dividirse en dos grupos: variables macroeconómicas e indicadores técnicos. Se realiza un análisis econométrico sobre las series, utilizando para ello un modelo lineal conocido como *Lasso* (“least absolute shrinkage and selection operator”, por sus siglas en inglés) y se comparan los resultados con los obtenidos mediante una regresión de Mínimos Cuadrados Ordinarios (MCO).

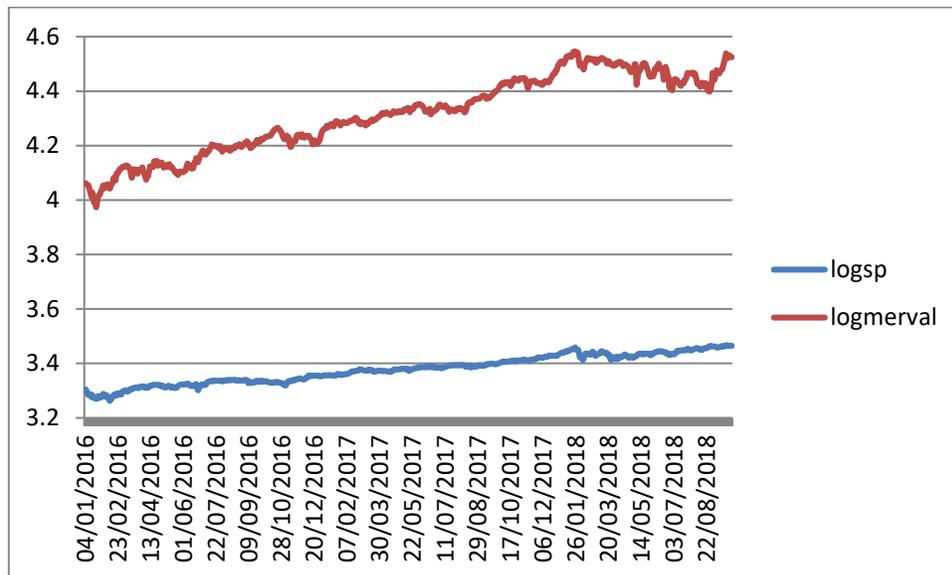
INTRODUCCIÓN	2
1. MODELOS	6
1.1 REGRESIÓN LINEAL	6
1.2 LASSO Y RIDGE	7
1.3 CROSS VALIDATION	10
2. VARIABLES	11
3. RESULTADOS	15
3.1 MODELO SIN REZAGOS	16
3.2 MODELO CON REZAGOS	20
3.3 PRESENCIA DE HETEROCEDASTICIDAD	25
3.4 CAPACIDAD PREDICTIVA	26
4. CONCLUSIÓN	28
BIBLIOGRAFÍA	30
APÉNDICE	32

Introducción

Desde que existen los mercados de acciones se ha tratado de encontrar métodos que permitan predecir los precios futuros de los valores, todos ellos sin éxito. Predecir los valores futuros de las acciones es una tarea extremadamente compleja, cuando no imposible, por la naturaleza misma del mercado. El mercado está compuesto por seres humanos que están dominados por emociones y que, a veces, presentan comportamientos irracionales (Murphy, 1999). Esto complejiza el análisis, dado que la cantidad de información que se necesita para al menos intentar predecir este comportamiento tiende a infinito. Por otro lado, Atsalakis y Valavanis (2009) dicen que la idea central de un modelo predictivo exitoso para el mercado bursátil es lograr los mejores resultados, utilizando la menor cantidad de información y un modelo lo más simple posible. Ésta es la idea que se trata de llevar adelante, generando un modelo lineal simple conocido como *Lasso* (“least absolute shrinkage and selection operator”, por sus siglas en inglés) que permita encontrar cuales de las variables propuestas son mejores para predecir el comportamiento del mercado, utilizando variables macroeconómicas e indicadores técnicos.

Es evidente que el precio de las acciones está influido por la economía real. Humpe y Macmillan (2009) testean la hipótesis de que las variables macroeconómicas estén cointegradas en el largo plazo con el precio de las acciones de Estados Unidos y Japón. En su análisis, dichos autores testean la cointegración de estos índices con el índice de la producción industrial, la inflación, M1 (monedas y billetes en manos del público y las reservas de los bancos) y la tasa de interés de largo plazo. Para ello, utilizan un modelo llamado *Vector Error Correction Model (VECM)*, y los test de Dickey-Fuller (aumentado), y Phillips-Perron. Encuentran que existe cointegración entre el índice de producción industrial, la inflación y la tasa de interés de largo plazo con los índices bursátiles.

Los mercados emergentes, como Argentina, suelen comportarse de manera más volátil que los mercados de países desarrollados. Esto puede observarse gráficamente:



Fuente: elaboración propia.

Donde \log_{sp} es el logaritmo de la serie del SP500 y \log_{merval} el logaritmo de la serie del Merval. La varianza de $\log_{sp} = 0,0027$, mientras que la varianza de $\log_{merval} = 0,0205$, casi 10 veces mayor. Además, existe la posibilidad de que, en el corto plazo, la influencia de variables macroeconómicas locales sobre el mercado sea no significativa, y el movimiento de los mercados esté determinado principalmente por los movimientos de los mercados desarrollados. A diferencia de esto, en el largo plazo si se podría esperar una significatividad de las variables macroeconómicas, porque la cotización del índice bursátil depende del precio de las acciones de las empresas, y este, a su vez, depende del desempeño que tengan las empresas en el período. Ésta es la hipótesis que desarrollan Bilson et al. (2001). En su estudio, estos autores testean la significatividad de variables macro sobre los mercados emergentes durante un período que va desde enero de 1985 a diciembre de 1997. Las variables que utilizaron los autores fueron: la M1, el índice de precios al consumidor (IPC), el índice de actividad, el tipo de cambio, el retorno de los mercados a nivel mundial, el riesgo país, la tasa de interés, el ratio de Precio/Ganancias de las empresas e indicadores regionales ponderados para América Latina, Europa, Asia y otros. Su análisis se basó sobre 20 mercados emergentes que incluyen 6 latinoamericanos (Argentina, Brasil, Chile, Colombia, Venezuela y Méjico), 8 asiáticos (India, Indonesia, Malasia, Pakistán, Filipinas, Corea del Sur, Taiwán y Tailandia), 3 europeos (Grecia, Portugal y Turquía), 1 de medio oriente (Jordania) y 2

africanos (Nigeria y Zimbabue). Los autores encuentran que las variables macroeconómicas de dichos países son, no solo significativas, sino que además tienen un poder predictivo mayor al del retorno de los mercados mundiales. En este sentido, el presente trabajo no pretende necesariamente encontrar un conjunto de variables que permitan estimar los precios futuros de un índice bursátil, pero al menos intenta probar si existe una relación de corto plazo entre el Merval y determinadas variables de frecuencia diaria.

En línea con las ideas de estos autores, Gan et al. (2006) examinan los efectos de las variables macroeconómicas sobre el mercado de acciones de Nueva Zelanda. En este estudio, se testea la hipótesis de que el índice de precios al consumidor (IPC), el tipo de cambio, la inflación, el producto bruto interno (PBI), las tasas de interés de corto y largo plazo, la M1 (monedas y billetes en manos del público y las reservas de los bancos) y el precio doméstico del barril de petróleo tiene influencia sobre el índice bursátil. Encuentran que todas las variables mencionadas están cointegradas con el índice, y por ende presentan una relación de largo plazo con el mismo. Además, indican que el PBI, la M1 y las tasas de interés de corto y largo plazo resultan estadísticamente significativas para explicar las variables del índice.

En el presente estudio también son incluidos algunos indicadores técnicos en las variables a testear. Los indicadores técnicos son cálculos matemáticos basados en precios, volumen, máximos, mínimos y cualquier otro tipo de información sobre los movimientos de una acción que pueda ser extraída de un gráfico. Este tipo de herramientas es utilizado en lo que los *traders* denominan análisis técnico. El análisis técnico se basa en la idea de que los precios se mueven entorno a tendencias que son determinadas por los cambios de posturas de los inversores en función de una variedad de situaciones posibles como cambios en la política monetaria, fiscal, etcétera. El análisis técnico está apoyado en la teoría de que el precio es un reflejo de lo que la masa de inversores espera de la evolución futura del mismo. En este sentido el análisis técnico intenta predecir el precio futuro asumiendo que los inversores se mueven en masa entre el pánico, el optimismo, la codicia y la confianza extrema. Este tipo de análisis es, fundamentalmente, un análisis de corto plazo.

Tam y Cuong (2018) examinan la efectividad que poseen algunos indicadores técnicos (MACD, RSI y MA)¹ para operar en el mercado de acciones de Vietnam. Ellos encuentran que estos indicadores (especialmente el RSI) pueden maximizar los retornos durante los momentos de alza, y minimizar pérdidas durante los momentos de bajas.

Antecedentes de aplicación del modelo Lasso para la predicción de valores bursátiles pueden encontrarse en trabajos como el de Roy et al. (2015). Los autores proponen utilizar este modelo sobre las acciones de Goldman Sachs Group Inc. y comparar su rendimiento con el del modelo Ridge² cuya operatoria si bien es muy similar, difiere en la norma vectorial que cada modelo utiliza, lo que puede producir resultados muy distintos. Al comparar los resultados de los modelos mediante RMSE (root mean squared error) y MAPE (mean absolute percentage error), los autores encuentran que el modelo Lasso produjo mejores estimaciones que el modelo Ridge.

Para facilitar la comprensión, este trabajo se encuentra dividido de la siguiente manera: En la *primera sección* se presentan modelos que son utilizados para analizar las series temporales propuestas. La *segunda sección* consiste en un análisis de las variables seleccionadas para explicar las variaciones en el índice Merval. En la *tercera sección* se muestrann los resultados de las estimaciones realizadas. Finalmente, en la cuarta *sección* se exponen las conclusiones y alcances de este trabajo.

¹ MACD significa moving average convergence/divergence, RSI significa Relative Strength Index y MA significa moving average. Los dos primeros serán explicados con detenimiento más adelante en este trabajo, ya que forman parte de las variables seleccionadas.

² Lasso y Ridge son casos extremos de un modelo conocido como Elastic Net. Estos modelos serán descriptos con mayor precisión más adelante en este trabajo.

1. Modelos

En esta sección se explican los diferentes modelos relevantes para el trabajo. Empezando por una descripción del modelo de regresión lineal, se hace mención a los problemas que este modelo puede presentar en determinadas circunstancias (algunas muy frecuentes en el ámbito de las finanzas) que hacen necesaria la aplicación otros modelos, como el Lasso.

Luego se presenta el modelo Lasso, con un análisis profundo de las características de este modelo y las particularidades que lo distinguen del modelo Ridge, haciendo especial énfasis en aquellas que resultan relevantes para la aplicación del Lasso en finanzas en general, y en este trabajo en particular.

1.1 Regresión lineal

En la regresión lineal se intenta explicar y predecir el comportamiento de una variable $y \in \mathbb{R}$, a partir de un vector de variables explicativas x . El modelo expresa la esperanza condicionada de la variable explicada como combinación lineal de las explicativas, donde:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i, \text{ donde } \varepsilon_i \sim N(0, \sigma^2) \text{ e } (y_i, x_i) \ i = 1, \dots, n$$

El objetivo del método de mínimos cuadrados ordinarios (MCO) es la minimización del cuadrado de los residuos, en adelante RSS (por sus siglas en inglés), de modo que:

$$RSS(\beta) = \|Y - X\beta\|_2^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij} \right)^2$$

De manera tal que, si se cumplen los supuestos de Gauss – Markov, los estimadores son MELI (mejor estimador lineal insesgado) y, por lo tanto, son insesgados y poseen varianza mínima.

Sin embargo, y de no cumplirse alguno de los supuestos del teorema de Gauss – Markov, el método de MCO puede tener algunos problemas. En primer lugar, y como señala Tibshirani (1996), los estimadores normalmente tienen poco (o directamente no tienen) sesgo, pero poseen una gran varianza. Esto genera que la estimación pierda precisión, característica vital para efectuar predicciones. En finanzas es común encontrarse con problemas de este tipo, dado que la naturaleza misma de los datos provoca que las series no sean homocedásticas (es decir que $Var(y_t|x_t) = Var(\epsilon_t) \neq \sigma^2$), o que los residuos estén correlacionados ($Cov(\epsilon_i, \epsilon_j) \neq 0$). Si algo de esto sucede, los estimadores ya no son MELI.

En segundo lugar, existe el problema de colinealidad. Si la colinealidad es perfecta, esto es: una variable es una combinación lineal de otra, genera que el determinante de la matriz ($X'X$) sea igual a cero y, por lo tanto, la matriz no se puede invertir, lo que provoca que el modelo no se pueda estimar. Si bien es infrecuente encontrar casos de colinealidad perfecta (siempre y cuando se haya tenido cuidado con la recolección de datos), es frecuente encontrar problemas de colinealidad severa o casi perfecta (principalmente cuando se trabaja en finanzas). Esto es, cuando una variable no es exactamente una combinación lineal de otra, pero *casi*. Un ejemplo de esto podría ser incluir en el conjunto de variables al salario de una persona, y a su mismo salario en dólares (asumiendo que el tipo de cambio no está fijo, sino que fluctúa, pero no demasiado). Lo que provoca que el determinante de la matriz ($X'X$) no sea cero, pero tienda a cero. Esto significa que la varianza, además de los tiempos computacionales, aumentará enormemente, con los efectos antes mencionados.

1.2 Lasso y Ridge

El *Lasso* fue propuesto por primera vez por Robert Tibshirani en 1996. Este modelo minimiza la suma de los cuadrados de los residuos, sujeto a que la suma del valor absoluto de los coeficientes sea menor que una constante. La naturaleza de la restricción (dada la norma vectorial con la que opera el modelo) produce que algunos coeficientes sean exactamente

iguales a cero (Tibshirani, 1996), lo que permite quedarse solamente con aquellas variables que mejor explican el comportamiento de la variable objetivo.

Lasso y Ridge son los casos extremos de otro modelo conocido como *elastic net*:

$$\hat{\beta} = \arg \min_{\beta} \|y - X\beta\|^2 + \lambda [(1 - \alpha) \|\beta\|_2^2 + \alpha \|\beta\|_1]$$

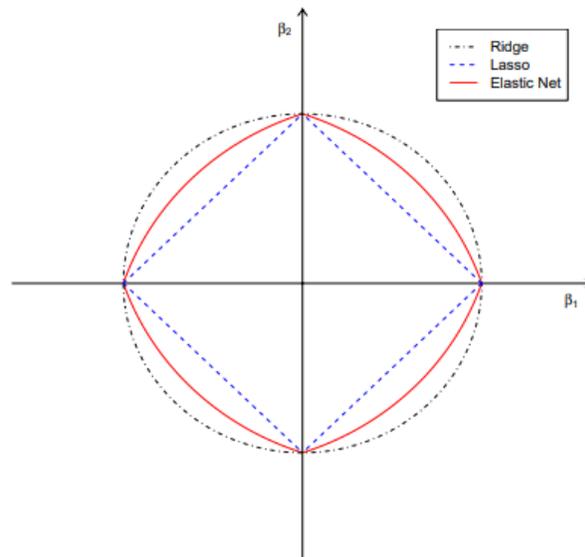
Si $\lambda = 0$, entonces el modelo de vuelve un modelo de MCO, porque el término de penalización queda invalidado.

Si $\lambda > 0$ y $0 < \alpha < 1$, el modelo es *elastic net*.

Si $\lambda > 0$ y $\alpha = 0$, entonces el modelo es Ridge.

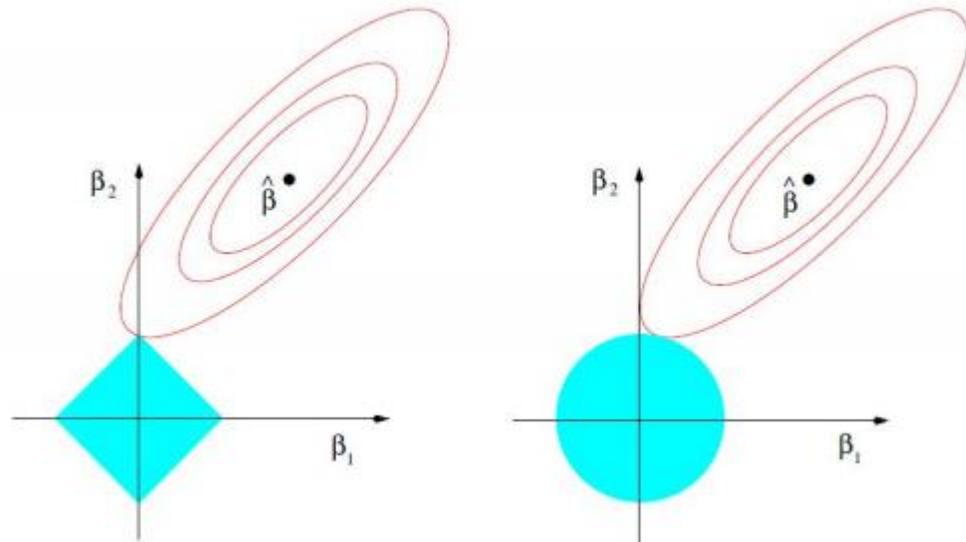
Si $\lambda > 0$ y $\alpha = 1$, entonces el modelo es Lasso.

La principal diferencia entre los modelos Lasso y Ridge es que el primero opera con norma vectorial de tipo L1, mientras que Ridge lo hace con norma vectorial de tipo L2. Esta diferencia genera que el contorno contra el que se minimiza la suma del cuadrado de los residuos sea convexo o lineal:



Fuente: Friedman *et al.* (2008). Con un $\alpha = 0.5$ *elastic net* es una solución intermedia.

Por la propia característica del contorno de Lasso, las soluciones tienden a ser de esquina. Esto genera que algunos coeficientes sean iguales a cero. En el caso opuesto, Ridge penaliza los coeficientes, pero tiende a no hacerlos iguales a 0:



Fuente: Friedman *et al.* (2008). Lasso (izquierda) y Ridge (derecha). Las áreas azules se corresponden con las restricciones $|\beta_1| + |\beta_2| \leq t$, y $\sqrt{\beta_1^2 + \beta_2^2} \leq t$. Las elipses moradas son los contornos de la función de error mínimo cuadrático.

Lasso tiende a eliminar los grupos de variables, quedándose solo con la más relevante. Esto facilita la interpretación de los modelos, al contar con menos variables explicativas. Otra ventaja del modelo Lasso es que tiende a ser computacionalmente eficiente para calcular los estimadores, lo que es una ventaja en problemas con conjuntos de variables grandes, y su capacidad predictiva suele ser buena (Varian, 2014).

La elección de un modelo u otro depende principalmente del objeto de estudio. Si el área de estudio es, por ejemplo, la biología, entonces es probable que se quieran observar los efectos de todas las variables, por más pequeños que estos sean, o que se tengan variables de control que se quieran introducir en el modelo. En el caso de este trabajo en particular, lo importante es el poder predictivo de las variables, y no si es una, la otra o un conjunto de ellas. Por eso,

la igualación de alguno o algunos coeficientes a cero es bienvenida, siempre que potencie el poder predictivo de las variables elegidas.

1.3 Cross validation

Como puede observarse todas estas técnicas de mínimos cuadrados penalizados dependen de un parámetro de penalización λ , que controla la importancia dada a la penalización en el proceso de optimización. Cuanto mayor es el valor del parámetro λ mayor es la penalización en los coeficientes de regresión y más son contraídos éstos hacia cero. Un método para estimar λ es validación cruzada o *cross validation*.

El método de validación cruzada consiste en dividir los datos del modelo en un set de para ajustar el modelo y un set de prueba para evaluar su capacidad predictiva, mediante el error de predicción u otra medida. La forma en que se aplica la validación cruzada es mediante la división del conjunto de datos disponibles de manera aleatoria en K subconjuntos de igual tamaño y mutuamente excluyentes. Uno de los subconjuntos se utiliza como datos de prueba y el resto ($K - 1$) como datos de entrenamiento. El proceso de validación cruzada es repetido durante K iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. Finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado. Este método es muy preciso puesto que se evalúa a partir de K combinaciones de datos de entrenamiento y de prueba, pero aun así tiene una desventaja, y es que es lento desde el punto de vista computacional. Nuestro valor del parámetro será el que arroje el mínimo error. Matemáticamente, dividimos el conjunto de datos D en K subconjuntos de igual tamaño, de manera tal que D_1, \dots, D_k , con $k = 1, \dots, K$. Se ajusta el modelo $\hat{f}_{-k}^{(\lambda)}(z)$ con el conjunto de entrenamiento. Se calcula el error por *cross validation*:

$$(CV\ Error)_k^{(\lambda)} = \frac{1}{|D_k|} \sum_{z \in D_k} [y - \hat{f}_{-k}^{(\lambda)}(z)]^2$$

Minimizando el error global de *cross validation*:

$$\lambda^* = \arg \min_{\lambda} (CV \text{ Error})^{(\lambda)} = \arg \min_{\lambda} \left[\frac{1}{K} \sum_{k=1}^K (CV \text{ Error})_k^{(\lambda)} \right]$$

2. Variables

Las variables descriptas a continuación fueron seleccionadas para testear su impacto sobre el índice Merval. Los datos van desde el 04/01/2016 al 28/09/2018. Este espectro temporal fue seleccionado por dos motivos:

En primer lugar, porque lo que se busca en este trabajo es generar predicciones de corto plazo, por lo tanto no interesan los efectos de largo plazo, sino los impactos inmediatos. De esta manera, no se necesita tener una serie extremadamente larga.

En segundo lugar, porque las elecciones presidenciales de 2015 en Argentina generaron un impacto sobre el índice de extrema volatilidad, que podría provocar un fuerte aumento en la varianza de los estimadores.

A continuación se describen las variables seleccionadas para este trabajo:

1. Merval. Fuente de los datos: www.bolsar.com.ar. El Merval es el principal índice del Mercado de Valores de Buenos Aires. Este índice mide el valor en pesos de una cartera de acciones que cotizan en la Bolsa de Comercio de Buenos Aires. El criterio de selección de estas acciones es en base al volumen operado y al número de transacciones en los últimos seis meses, bajo la condición necesaria de una negociación en al menos el 80% de las ruedas consideradas. El rebalanceo del Merval se hace trimestralmente.

2. Volumen. Fuente de los datos: elaboración propia en base a www.bolsar.com.ar. Es la cantidad nominal de pesos negociada durante un día en acciones pertenecientes al panel líder (índice Merval).

3. Tipo de cambio nominal. Fuente: Banco Central de la República Argentina. Expresada como la cantidad de pesos necesarios para adquirir un dólar estadounidense.
4. Stock de reservas internacionales (expresado en miles de millones de dólares). Fuente: Banco Central de la República Argentina.
5. M2 (expresada en miles de millones de pesos). Fuente: Banco Central de la República Argentina. La M2 es un agregado monetario que incluye la M1 (monedas y billetes en manos del público y las reservas de los bancos) y a ello le suma los depósitos a corto plazo (hasta dos años).
6. Base Monetaria (expresada en miles de millones de pesos). Fuente: Banco Central de la República Argentina. Constituido por las monedas y billetes en manos del público y las reservas de los bancos.
7. EMBI Argentina. Fuente: JP Morgan Chase. El EMBI (Emerging Markets Bonds Index o Indicador de Bonos de Mercados Emergentes) es el principal indicador de riesgo país y está calculado por JP Morgan Chase. Es la diferencia de tasa de interés que pagan los bonos denominados en dólares, emitidos por países subdesarrollados, y los Bonos del Tesoro de Estados Unidos, que se consideran libres de riesgo. Esta calculado en *puntos básicos*, donde 100 puntos básicos equivalen a un 1% de *sobre tasa o spread*.
8. Tasa de interés (plazo fijo en pesos 30-44 días). Fuente: Banco Central de la República Argentina.
9. Tasa de interés USA. Fuente: www.investing.com. Es la tasa de interés de la FED (Reserva Federal de Estados Unidos).
10. S&P 500. Fuente: www.investing.com. El índice Standard & Poor's 500 es uno de los índices bursátiles más importantes de Estados Unidos. Al S&P 500 se lo considera el índice más representativo de la situación real del mercado. El índice se basa en la capitalización bursátil de 500 grandes empresas que poseen acciones que cotizan en las bolsas NYSE o NASDAQ, y captura aproximadamente el 80% de toda la capitalización de mercado en Estados Unidos.
11. VIX. Fuente: www.investing.com. VIX es el código del oficialmente llamado Chicago Board Options Exchange Market Volatility Index. En el momento en que hay alta

volatilidad, el VIX alcanza una cifra elevada y se correlacionan con caídas del S&P 500, indicando que en el mercado hay miedo y pesimismo.

12. ETF de mercados emergentes. Fuente: www.investing.com. Un ETF (Exchange Traded Fund – por sus siglas en inglés) es un fondo de inversión con cotización pública. Se considera al mismo como un activo cotizante que replica el comportamiento de un índice, un commodity, bonos o una canasta de activos como lo es un fondo indexado. Los ETF cotizan como si fuesen una acción más del mercado. Es necesario mencionar que el Merval es parte de este ETF, sin embargo, representa una porción muy pequeña del mismo. El objetivo de incluir esta variable reside en captar el efecto que genera un aumento en el flujo de capitales a acciones de mercados emergentes en general, y no a las acciones Argentinas específicamente.

13. TIR del bono Y10 USA. Fuente: www.investing.com. Utilizada como proxy de la tasa de interés de largo plazo, es la tasa interna de retorno de los bonos a 10 años de Estados Unidos.

14. MFI. Fuente: Elaboración propia. El MFI (money flow index) es un indicador que utiliza las variaciones en el precio y el volumen para mostrar períodos de sobrecompra o sobreventa, indicando el *momentum* de venta o compra respectivamente. Normalmente, el MFI está calculado sobre los últimos 14 períodos. La manera de construirlo es la siguiente:

$$\text{Typical Price} = \frac{\text{High} + \text{Low} + \text{Close}}{3}$$

$$\text{Raw Money Flow} = \text{Typical Price} \times \text{Volume}$$

$$\text{Positive Raw Money Flow}_i = \begin{cases} \text{Raw Money Flow} & \text{si } \text{Merval}_t - \text{Merval}_{t-1} > 0 \\ 0 & \text{si } \text{Merval}_t - \text{Merval}_{t-1} \leq 0 \end{cases}$$

$$\text{Negative Raw Money Flow}_i = \begin{cases} \text{Raw Money Flow} & \text{si } \text{Merval}_t - \text{Merval}_{t-1} < 0 \\ 0 & \text{si } \text{Merval}_t - \text{Merval}_{t-1} \geq 0 \end{cases}$$

$$\text{Money Flow Ratio} = \frac{\sum_{i=1}^{14} \text{Positive Raw Money Flow}_i}{\sum_{i=1}^{14} \text{Negative Raw Money Flow}_i}$$

$$Money Flow Index = 100 - \frac{100}{1 + Money Flow Ratio}$$

15. RSI. Fuente: Elaboración propia. El RSI (relative strength index) es un indicador técnico que mide la fuerza de la tendencia para encontrar momentos de sobrecompra o sobreventa. El indicador se construye de la siguiente manera:

$$Average Gain = \sum_{i=1}^{14} \text{cambios positivos en el precio}_i$$

$$Average Loss = \sum_{i=1}^{14} \text{cambios negativos en el precio}_i$$

$$RS = \frac{Average Gain}{Average Loss}$$

$$RSI = 100 - \frac{100}{1 + RS}$$

El signo esperado de cada una de las variables mencionadas se muestra a continuación:

Variable	Signo esperado
volumen	+
TCN	+/-
Stock de reservas	+
M2	+
Base Monetaria	+
EMBI Argentina	-
Tasa de interés	-
Tasa de interés USA	-
S&P 500	+
VIX	-
ETF	+
TIR bono Y10 USA	-
MFI	+
RSI	+

El signo esperado del tipo de cambio depende de la circunstancia. Si el tipo de cambio aumenta mientras que el valor en dólares de las acciones se mantiene constante, podríamos esperar un aumento en el índice Merval, dado que la mayoría de las acciones que componen el índice cotizan a su vez en Estado Unidos y debe aumentar su valor en pesos para mantener la paridad y no quedar desarbitradas. Si el aumento del tipo de cambio se da por, por ejemplo una corrida cambiaria provocada por una caída abrupta de la confianza de los inversores en la moneda local, podríamos esperar que también vendan los activos del país, generando una caída en el precio en dólares de las acciones que no llegue a ser compensada por el aumento en el tipo de cambio, provocando una caída en el índice Merval.

Además de las variables descriptas, se instrumentan rezagos de 1 días y 5 días para las variables, entendiendo que es posible que los efectos de cambios en dichas variables sobre el índice Merval no sean instantáneos.

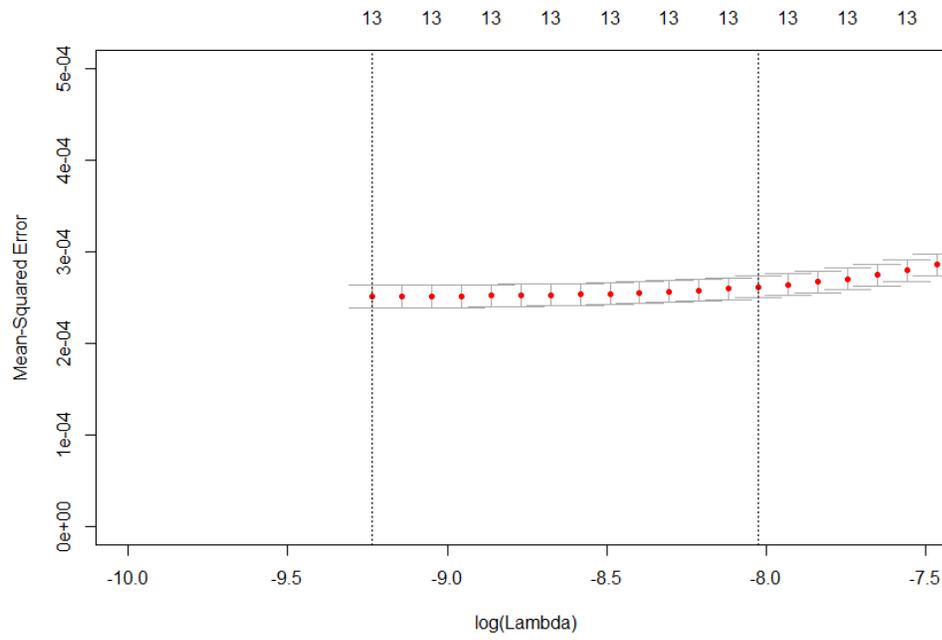
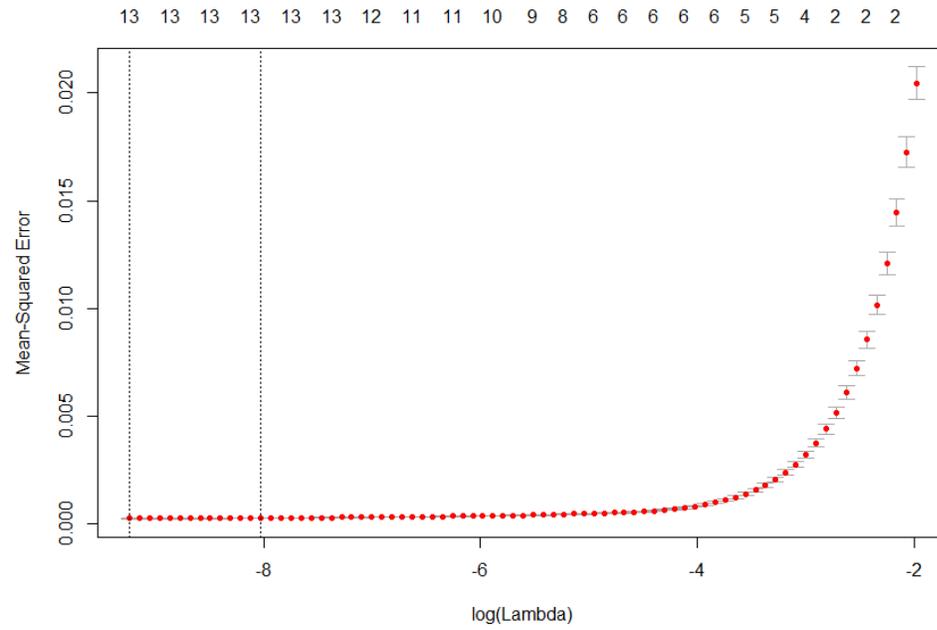
3. Resultados

Se trabajó con la base de datos presentada, generando logaritmos de todas las variables para una mejor interpretación. Los modelos logarítmicos (o log-log), son modelos donde la interpretación de los coeficientes de las variables se lee en forma porcentual (utilizada por ejemplo, para medir elasticidades). Es decir, que el coeficiente β_i que acompaña a la variable x_i mide el impacto porcentual en y de un cambio de un 1% en x_i .

También se trabajó con dos modelos. El primero de ellos, sin rezagos. El segundo, contiene rezagos de 1 y 5 días respectivamente.

3.1 Modelo sin rezagos

En primer lugar, se obtuvieron los valores óptimos del parámetro λ :



Se observa que el valor de λ que minimiza el error medio cuadrático (MSE) es muy pequeño:

$$\log(\lambda) = -9.234616$$

$$\lambda = 9.760166 * e^{-5} = 0.0000976$$

Esto significa que el término de penalización tiene muy poco peso en el modelo.

El modelo Lasso arrojó los siguientes coeficientes³:

	Coefficiente	p-value	significativo al 0.05	significativo al 0.01
(Intercept)	-0.8956		*	***
Volumen	0.0106	1.18E-08	*	***
TCN	0.1305	1.38E-25	*	***
Reservas	0.0619	9.39E-11	*	***
M2	.			
BM	0.0962	9.64E-06	*	***
EMBI	-0.0942	7.45E-22	*	***
interescp	-0.0499	6.48E-14	*	***
interesUScp	-0.0397	1.69E-17	*	***
sp500	0.3282	5.45E-16	*	***
VIX	0.034	2.16E-19	*	***
ETF	0.3834	2.56E-48	*	***
TIRY10	0.0303	0.00025	*	***
MFI	0.0172	7.68E-13	*	***
RSI	0.0107	0.00075	*	***

Sobre la misma base, se aplicó una regresión por mínimos cuadrados ordinarios:

Residuals:					
Min	1Q	Median	3Q	Max	
-0.056423	-0.009577	0.000705	0.009232	0.044102	

³ Lasso no posee P-value en sus resultados. Por ello, se utilizó un paquete de R llamado *hdi* que calcula dichos estadísticos. Más información sobre el test puede ser encontrada en el trabajo de *Dezeure, R et al.*, (2015) citado en la bibliografía.

Coefficients:	Estimate	Std.Error	tvalue	Pr(> t)	
(Intercept)	-1.240926	0.289114	-4.292	2.04E-05	***
Volumen	0.009854	0.001725	5.714	1.68E-08	***
TCN	0.141151	0.013676	10.321	2.00E-16	***
Reservas	0.063286	0.009699	6.525	1.36E-10	***
M2	0.018933	0.022992	0.823	0.41054	
BM	0.075203	0.01947	3.862	0.000123	***
EMBI	-0.099212	0.010306	-9.627	2.00E-16	***
interescp	-0.054957	0.007295	-7.534	1.63E-13	***
interesUScp	-0.05036	0.005875	-8.572	2.00E-16	***
sp500	0.376019	0.05425	6.931	9.96E-12	***
VIX	0.036623	0.004036	9.075	2.00E-16	***
ETF	0.373717	0.02642	14.145	2.00E-16	***
TIRY10	0.029674	0.008378	3.542	0.000425	***
MFI	0.017086	0.002354	7.259	1.10E-12	***
RSI	0.010686	0.003142	3.401	0.000711	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01563 on 659 degrees of freedom
Multiple R-squared: 0.9884, Adjusted R-squared: 0.9881
F-statistic: 3995 on 14 and 659 DF, p-value: < 2.2e-16

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance = 0.05

Call:
gvlma(x = lmodel)

	Value	p-value	Decision
Global Stat	17.226	0.0017467	Assumptions NOT satisfied!
Skewness	2.643	0.1039962	Assumptions acceptable.
Kurtosis	11.649	0.0006424	Assumptions NOT satisfied!
Link Function	1.152	0.2831998	Assumptions acceptable.
Heteroscedasticity	1.783	0.1818201	Assumptions acceptable.

Los resultados muestran que solo la variable M2 fue eliminada por el modelo. Los signos de casi todos los coeficientes se corresponden con lo esperado, a excepción de VIX y TIRY10. Un aumento en el VIX indica un aumento en la volatilidad en los mercados de Estados Unidos. Sería esperable que este aumento de la volatilidad lleve a los inversores a refugiarse en activos menos riesgosos y con menor volatilidad (como los bonos de la FED, el oro, etc.). Para tomar refugio en activos menos riesgosos, los inversores (que principalmente son muy

grandes y globalizados inversores institucionales, cuyos activos están diversificados entre varios países del globo) venden parte de los activos que poseen, principalmente los pertenecientes a los mercados más frágiles, como los mercados emergentes. Estas grandes ventas de activos en mercados pequeños como el Merval, provocan que la demanda por los activos no alcance para cubrir toda la oferta, haciendo caer los precios. Por ello, la relación entre el Merval y el VIX es negativa, y esperaríamos que el signo de la variable también lo fuese. Algo similar sucede con TIRY10, o la TIR del bono a 10 años de Estados Unidos. Cuando aumenta la TIR del bono, está cayendo su precio. Esto nos indica que los inversores esperan un aumento de las tasas de interés de la FED y, por ende, venderán activos, dado que el costo de oportunidad ahora es mayor. Por ende, es esperable que un aumento de TIRY10 se correlacione con una caída del índice Merval.

Las 3 variables que reportan mayor impacto son ETF, SP500 y TCN (aunque en menor medida).

El Merval es parte del ETF de mercados emergentes, por ende es esperable que ante una suba de este, se genere una suba en el índice Merval, porque parte de los capitales que estarían entrando a los mercados emergentes va a ir al Merval.

Algo similar sucede con el SP500. En general, las variaciones de los índices de mercados emergentes tienen alta correlación con los movimientos del SP500. Ya sea a la baja, cuando los inversores salen de los activos riesgosos para refugiarse en otros menos riesgosos, o en las subas, cuando van en busca de rentabilidad

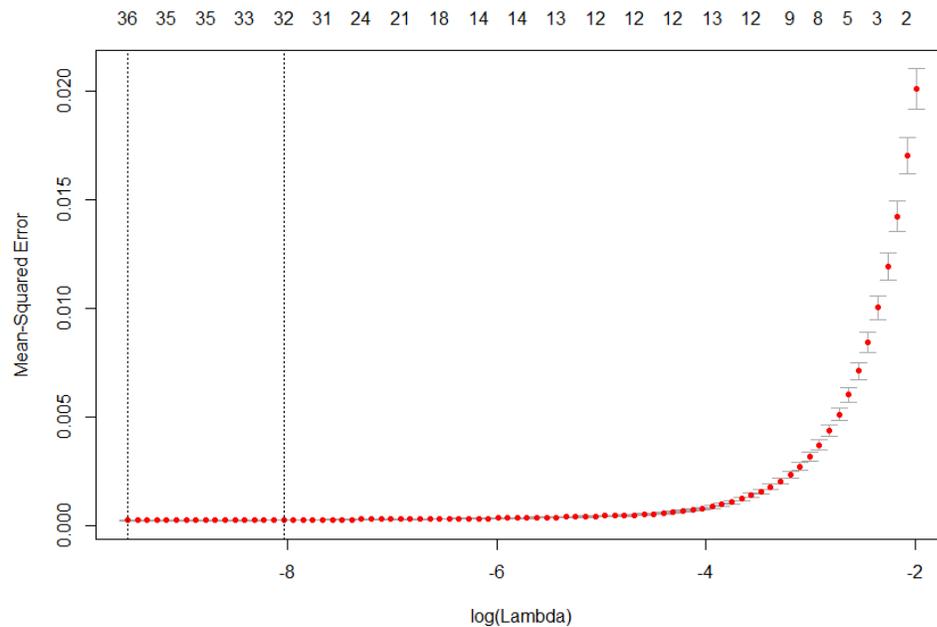
Con el TCN suceden efectos ambiguos. En general la suba del tipo de cambio (principalmente cuando se dan corridas cambiarias) está asociada a la pérdida de confianza de los inversores de muy corto plazo en el país, y una salida de capitales especulativos. Entonces el signo debería de ser negativo. Sin embargo, muchas de las acciones que componen el índice Merval, cotizan a su vez en Estados Unidos. Lo que significa que, *ceteris paribus* el precio en dólares de una acción, ante una suba del tipo de cambio, se espera una suba equivalente en el precio en pesos del papel, para no quedar *desarbitrado*. Por ende, ante aumentos pequeños y paulatinos del tipo de cambio, se espera que el precio de las acciones

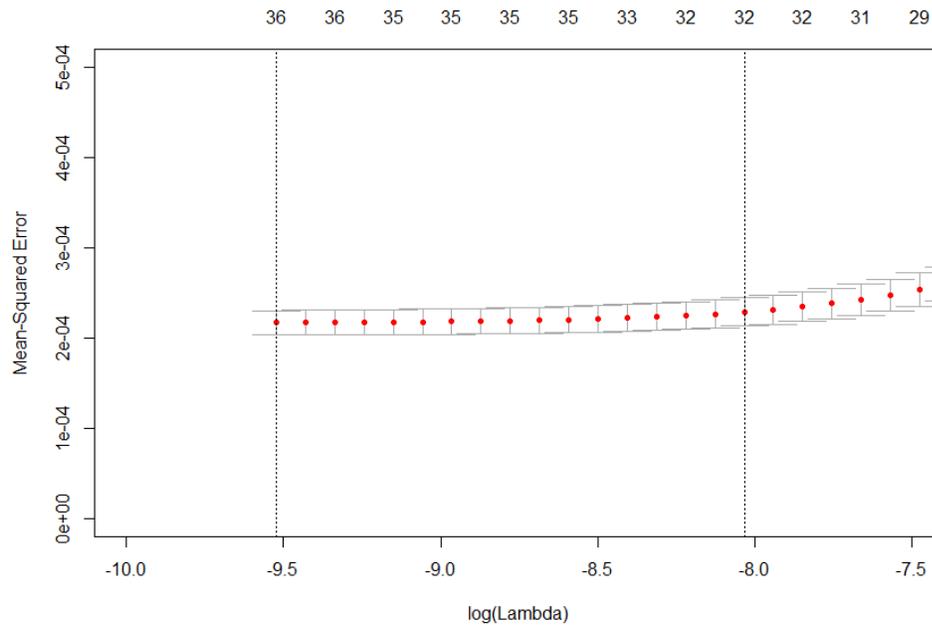
aumente en la misma proporción, siempre y cuando este aumento del TCN no se deba (o no genere) cambios en las expectativas de los inversores sobre los retornos de las acciones.

Respecto a los test aplicados al modelo de MCO, se encontró que la serie tiene *curtosis*, lo cual es muy esperable en series financieras. En cuanto a los modelos utilizados, los resultados no arrojan evidencia de que utilizar Lasso aporte ventajas sobre la implementación del modelo de Mínimos Cuadrados Ordinarios

3.2 Modelo con rezagos

Se agregaron rezagos de 1 día y 5 días a la base de datos, y sobre ella se volvieron a instrumentar los modelos. Primero, se buscó el mínimo para el parámetro λ :





$$\log(\lambda) = -9.522685$$

$$\lambda = 7.31729 * e^{-5} = 0.0000732$$

Nuevamente, y como se esperaba, se obtuvo un λ muy pequeño.

Los resultados del modelo Lasso, junto a su p-value se muestran a continuación:

	Coeficiente	p-value	significativo o al 0.05	significativo o al 0.01
(Intercept)	-0.947			
x.Volumen	0.0053	0.00548	*	***
x.TCN	0.0099	0.58321		
x.Reservas	.	0.65864		
x.M2	0.0133	0.45240		
x.BM	0.0608	0.00663	*	***
x.EMBI	-0.0675	0.01116	*	
x.interescp	-0.01	0.39400		
x.interesUScp	-0.0245	0.00371	*	***
x.sp500	0.1346	0.03396	*	
x.VIX	.	0.71324		
x.ETF	0.1875	0.00014	*	***
x.TIRY10	0.0049	0.85248		

x.MFI	0.0184	0.00019	*	***
x.RSI	0.0124	0.00640	*	***
xlag1.Volumen	0.0073	0.00027	*	***
xlag1.TCN	0.0822	8.89E-07	*	***
xlag1.Reservas	.	0.26861		
xlag1.M2	0.0025	0.76506		
xlag1.BM	0.0168	0.14197		
xlag1.EMBI	-0.0092	0.66608		
xlag1.interescp	-0.0239	0.07480		
xlag1.interesUSc p	-0.0092	0.14550		
xlag1.sp500	0.1317	0.01852	*	
xlag1.VIX	0.0095	0.21032		
xlag1.ETF	0.0671	0.18140		
xlag1.TIRY10	0.0155	0.21207		
xlag1.MFI	.	0.75088		
xlag1.RSI	-0.0015	0.46895		
xlag5.Volumen	0.0085	1.53E-06	*	***
xlag5.TCN	0.0177	0.29083		
xlag5.Reservas	0.0712	6.36E-10	*	***
xlag5.M2	.	0.41104		
xlag5.BM	0.0187	0.27001		
xlag5.EMBI	-0.0087	0.32696		
xlag5.interescp	-0.0152	0.21752		
xlag5.interesUSc p	-0.0117	0.17453		
xlag5.sp500	0.0537	0.10883		
xlag5.VIX	0.0227	2.79E-05	*	***
xlag5.ETF	0.1125	0.00205	*	***
xlag5.TIRY10	.	0.85746		
xlag5.MFI	0.0011	0.71145		
xlag5.RSI	0.0033	0.18566		

Los resultados de MCO fueron los siguientes⁴:

Residuals:					
Min	1Q	Median	3Q	Max	
-0.051241	-0.008743	0.000671	0.008504	0.039486	

⁴ La siguiente tabla muestra solo las variables que resultaron significativas al 0.05 o menos. La tabla completa se muestra en el apéndice.

Coefficients:	Estimate	Std.Error	tvalue	Pr(> t)	
(Intercept)	-1.30425	0.321305	-4.059	0.0000555	***
x.Volumen	0.005573	0.001895	2.941	0.003391	**
x.EMBI	-0.062253	0.026552	-2.345	0.019359	*
x.ETF	0.200072	0.079639	2.512	0.012248	*
x.MFI	0.020998	0.005715	3.674	0.000259	***
x.RSI	0.014019	0.005641	2.485	0.013201	*
xlag1.Volumen	0.006827	0.001957	3.489	0.000519	***
xlag5.Volumen	0.008113	0.001708	4.749	0.00000254	***
xlag5.Reservas	0.091819	0.015388	5.967	4.04E-09	***
xlag5.VIX	0.027054	0.006758	4.003	0.0000699	***
xlag5.ETF	0.094128	0.04203	2.24	0.02547	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01428 on 626 degrees of freedom
Multiple R-squared: 0.9905, Adjusted R-squared: 0.9899
F-statistic: 1560 on 42 and 626 DF, p-value: < 2.2e-16

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance = 0.05

Call:
gvlma(x = lmodel)

	Value	p-value	Decision
Global Stat	11.4599	0.02185	Assumptions NOT satisfied!
Skewness	0.3368	0.56167	Assumptions acceptable.
Kurtosis	4.1448	0.04176	Assumptions NOT satisfied!
Link Function	1.0571	0.30387	Assumptions acceptable.
Heteroscedasticity	5.9211	0.01496	Assumptions NOT satisfied!

En este caso, los modelos no coincidieron en las mismas variables. A continuación se muestra una lista de las variables que resultaron significativas para ambos modelos, y las que solo fueron seleccionadas por Lasso. No hubo variables que resultaran significativas en el modelo de MCO que no fueran seleccionadas por Lasso.

Variable	Lasso		MCO	
	Coefficiente	pvalue	Coefficiente	pvalue
(Intercept)	-0.947		-1.30425	0.0000555
x.Volumen	0.0053	0.00548	0.005573	0.003391
x.EMBI	-0.0675	0.01116	-0.062253	0.019359

x.ETF	0.1875	0.00014	0.200072	0.012248
x.MFI	0.0184	0.00019	0.020998	0.000259
x.RSI	0.0124	0.0064	0.014019	0.013201
xlag1.Volumen	0.0073	0.00027	0.006827	0.000519
xlag5.Volumen	0.0085	0.00000153	0.008113	0.00000254
xlag5.Reservas	0.0712	6.36E-10	0.091819	4.04E-09
xlag5.VIX	0.0227	0.0000279	0.027054	0.0000699
xlag5.ETF	0.1125	0.00205	0.094128	0.02547

VARIABLES SELECCIONADAS SOLO POR LASSO

Variable	Coefficiente	pvalue	Coefficiente (MCO)	Pvalue (MCO)
x.BM	0.0608	0.00663	0.042931	0.139134
x.interesUScp	-0.0245	0.00371	-0.027051	0.25837
x.sp500	0.1346	0.03396	0.060303	0.712226
xlag1.TCN	0.0822	8.89E-07	0.040169	0.443341
xlag1.sp500	0.1317	0.01852	0.184274	0.302415

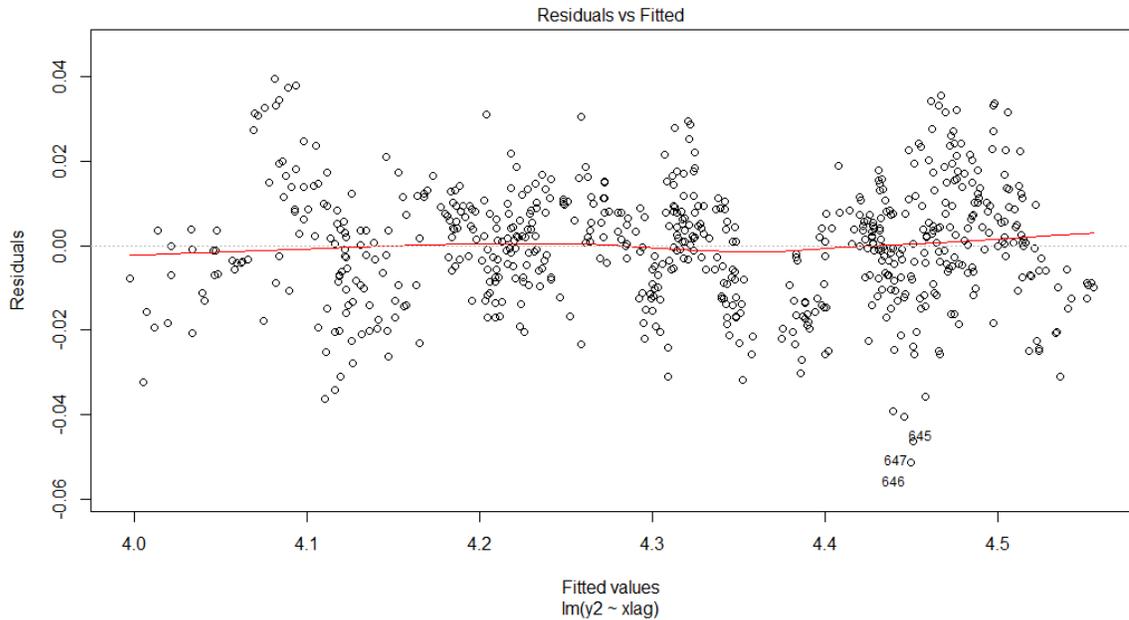
Respecto de los resultados, también se hallaron algunos efectos interesantes. En primer lugar, resulta llamativo que el efecto que tiene el volumen negociado sobre el índice Merval, aumente a medida que pasa el tiempo. Como se puede observar, el coeficiente del volumen con 5 días de rezago es mayor que el no rezagado. No sucede lo mismo con el ETF, que nuevamente se presenta como la variable de mayor importancia a tener en cuenta.

También resultaron significativos los indicadores técnicos (MFI y RSI) y por ende pueden ser tenidos en cuenta a la hora de realizar inversiones sobre el índice Merval. Es evidente que un aumento en alguno de estos indicadores no causa por sí mismo un aumento en el Merval, pero su sola significancia resulta relevante.

En cuanto a los modelos implementados, Lasso encontró significatividad en un mayor número de variables que mínimos cuadrados ordinarios.

3.3 Presencia de heterocedasticidad

Se realizó un análisis gráfico y estadístico sobre la presencia de heterocedasticidad en el modelo, encontrando evidencia de la misma:



studentized Breusch-Pagan test

```
data: lm1  
BP = 121.89, df = 42, p-value = 1.03e-09
```

El problema de la heterocedasticidad no invalida de ninguna manera el modelo. Los estimadores siguen siendo insesgados, aunque ya no son eficientes y, por lo tanto, no tendrán varianza mínima.

Para corregir el problema de la heterocedasticidad existen muchas formas distintas. Una de ellas es un modelo llamado *adaptive Lasso*. Este modelo es una versión mejorada del Lasso *clásico*. Contiene todas las bondades del Lasso, pero con mejoras en cuanto al método de selección de variables, lo que produce resultados más consistentes en determinadas situaciones. El objetivo del *adaptive Lasso* es minimizar:

$$RSS(\beta) + \lambda \sum_{j=1}^p \hat{\omega}_j |\beta_j|$$

Dónde:

$$\hat{\omega}_j = \frac{1}{(|\hat{\beta}_j^{ini}|)^\gamma}$$

El vector $\hat{\omega}_j$ genera distintas penalizaciones para cada coeficiente. $\hat{\beta}_j^{ini}$ puede ser estimado mediante el modelo Ridge, y γ puede ser dado subjetivamente, u obtenido mediante *cross validation*.

	Coefficientes adaptive Lasso
(Intercept)	1.097848921
x.BM	0.042037771
x.ETF	0.610781599
xlag1.TCN	0.098435454
xlag5.Reservas	0.106088668

Los resultados del modelo *adaptive Lasso* presentan el mismo valor de λ que el modelo *Lasso*, pero la cantidad de variables elegidas es menor, y sus coeficientes son mayores (especialmente x.ETF). Es probable que la diferencia en la selección de variables se deba a la presencia de heterocedasticidad. Dado que el modelo *adaptive Lasso* se comporta de mejor manera que el Lasso o el MCO ante la presencia de heterocedasticidad, se eligieron los resultados de este último modelo.

3.4 Capacidad predictiva

Interesa conocer si las variables seleccionadas por el modelo tienen capacidad predictiva. Para ello, se realizaron dos test. En primer lugar, se testea la estacionariedad de las series mediante el test de Phillips – Perron. En segundo lugar, se realiza un estudio sobre causalidad de Granger, mediante el test del mismo autor.

Que una serie de tiempo sea estacionaria, significa que su distribución de probabilidad se mantiene estable a lo largo del tiempo (Wooldridge, 2000). El test de Phillips – Perron arrojó los siguientes resultados⁵:

variable	p-value
y2	0.2248
d.y2	0.01
xbm	0.01
xetf	0.7025
d.xetf	0.01
x1tcn	0.99
d.x1tcn	0.01
x5res	0.6112
d.x5res	0.01

Dónde:

- $y2 = \text{Log}(\text{Merval})$
- $xbm = \text{base monetaria}$
- $xetf = \text{ETF}$
- $x1tcn = \text{TCN rezagado 1 día}$
- $x5res = \text{stock de reservas rezagado 5 días}$
- $d.\text{variable} = \text{la variable en diferencias}$

Que el *pvalue* sea inferior a 0.05 indica que la variable es estacionaria.

A continuación, se testeó *causalidad de Granger*. La causalidad de Granger implica que, si existe una relación causal en este sentido, una variable tiene poder predictivo sobre la otra. La causalidad puede ser unidireccional (A causa a B) o bidireccional (B también causa a A).

⁵ Los test completos pueden encontrarse en el apéndice.

En este trabajo no nos interesa la bidireccionalidad, porque el efecto que pueda o no causar el Merval sobre las variables macro excede los objetivos de este trabajo. Por ello, se testeó causalidad de Granger de las variables antes mencionadas sobre el Merval⁶:

Variable	Estadístico
Base monetaria	0.038
ETF	0.663
TCN (1 rezago)	0.091
Reservas (5 rezagos)	0.974

El resultado de los test muestra que solo la base monetaria muestra una relación causal en el sentido de Granger⁷ y, por ende, podría tener poder predictivo sobre el índice Merval. Es importante mencionar que el Merval es un índice bursátil medido en moneda local, es decir en pesos. Por ende, un aumento de la base monetaria llevará, casi con seguridad, a un aumento generalizado de precios (inflación), que afecta a todos los precios de la economía, incluidas las acciones bursátiles.

4. Conclusión

Se han analizado los efectos de corto plazo de un grupo de variables macroeconómicas e indicadores técnicos sobre el índice Merval. Durante el proceso se han encontrado varios resultados interesantes.

⁶ Los test completos pueden encontrarse en el apéndice.

⁷ La base monetaria resultó estacionaria y no necesito ser diferenciada. Sin embargo, para testear causalidad de Granger fue necesario quitarle una observación. Se testeó de tres maneras distintas: diferenciándola, quitándole la primera observación y quitándole la última; existiendo causalidad de todas las maneras.

En primer lugar se ha encontrado evidencia de que el modelo Lasso es sensible a problemas de heterocedasticidad. Cuando realizamos la regresión con el modelo *adaptive Lasso* hayamos resultados más robustos.

En segundo lugar se encontró que estimar mediante un conjunto de variables explicativas que contenga variables rezagadas generó mejores resultados que cuando no utilizamos rezagos. Esto nos indica que las variables en cuestión no tienen efectos inmediatos sobre el índice. También se halló evidencia de que solamente la base monetaria causa, en el sentido de Granger, al Merval.

Estos resultados no son menores. Nos muestran que en el corto plazo, el efecto que domina los movimientos del índice Merval es, muy probablemente, la aleatoriedad o variables latentes relacionadas con el “humor” de los inversores que resultan de muy compleja medición. En la actualidad se han desarrollado (y continúan en desarrollo continuo) modelos de *machine learning* que tratan de captar dicho efecto a través de, por ejemplo, Twitter.

Se encontró significatividad en algunas variables macroeconómicas que, si bien no pueden predecir los movimientos del índice, deberían ser tomadas en cuenta en modelos más complejos que pretendan realizar la misma tarea a través de un conjunto de variables explicativas más grande y heterogéneo.

Bibliografía

Atsalakis, G. y Valavanis, K. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 5932-5941.

Bilson, C., Brailsford, T. y Hooper, V. (2001). Selecting macroeconomic variables as explanatory factors of emerging stock market returns. *Pacific-Basin Finance Journal*, 9, 401–426.

Cuong, N. y Tam, P. (2018). Effectiveness of Investment Strategies Based on Technical Indicators: Evidence from Vietnamese Stock Markets. *Journal of Insurance and Financial Management*, 3(5), 55-68.

Daye, J., Chen, J. y Li, H. (2012). High-Dimensional Heteroscedastic Regression with an Application to eQTL Data Analysis. *Biometrics*, 62(1), 316-326.

Dezeure, R., Bühlmann, P., Meier, L. y Meinshausen, N. (2015) High-Dimensional Inference: Confidence, Intervals, p-Values and R-Software. *Statistical Science*, 30(4), 533-558.

Friedman, I., Hastie, T. y Tibshirani, R. (2008). The elements of statistical learning, Springer.

Fu, W. (1998). Penalized Regressions: The Bridge versus the Lasso. *Journal of Computational and Graphical Statistics*, 7(3), 397-416.

Gan, C., Lee, M., Yong, H. y Zhang, J. (2006). Macroeconomic Variables and Stock Market Interactions: New Zealand Evidence. *Investment Management and Financial Innovations*, 3(4), 89-101.

Granger, C. (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*. 37(3), 424-438.

Humpe, A. y Macmillan, P. (2009) Can macroeconomic variables explain long-term stock market movements? A comparison of the US and Japan, *Applied Financial Economics*, 19(2), 111-119.

- Jia, J., Rohe, K. y Yu, B. (2010). The Lasso under Heteroscedasticity. *Working Paper*.
- Murphy, J. (1999). Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications.
- Phillips, P. y Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*. 75(2), 335-346.
- Roy, S., Mittal, D., Basu, A. y Abraham, A. (2015). Stock Market Forecasting Using LASSO Linear Regression Model. In: Abraham A., Krömer P., Snasel V. (eds) Afro-European Conference for Industrial Advancement. Advances in Intelligent Systems and Computing, vol 334. Springer, Cham
- Sharpnack, J. y Kolar, M. (2014). Mean and variance estimation in high-dimensional heteroscedastic models with non-convex penalties. *Working Paper*.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58 (1), 267-288.
- Varian, H. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2), 3–28.
- Wagener, J. y Dette, H. (2011). The adaptive Lasso in high dimensional sparse heteroscedastic models. *Mathematical Methods of Statistics*. 22 (2).
- Wang, H., Li, G. y Jiang, G. (2007). Robust Regression Shrinkage and Consistent Variable Selection Through the LAD-Lasso, *Journal of Business & Economic Statistics*, 25(3), 347-355.
- Wooldridge, J. (2010). Introducción a la econometría. Un enfoque moderno. Cengage Learnig.
- Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*. 101(476), 1418-1429.

Apéndice

Resultados completos de MCO sobre data set con rezagos:

Residuals:					
Min	1Q	Median	3Q	Max	
-0.051241	-0.008743	0.000671	0.008504	0.039486	

Coefficients:	Estimate	Std.Error	tvalue	Pr(> t)	
(Intercept)	-1.30425	0.321305	-4.059	0.0000555	***
x.Volumen	0.005573	0.001895	2.941	0.003391	**
x.TCN	0.040765	0.048669	0.838	0.402576	
x.Reservas	0.025322	0.028601	0.885	0.376309	
x.M2	0.04302	0.054036	0.796	0.426257	
x.BM	0.042931	0.02899	1.481	0.139134	
x.EMBI	-0.062253	0.026552	-2.345	0.019359	*
x.interescp	-0.010343	0.013752	-0.752	0.452242	
x.interesUScp	-0.027051	0.023912	-1.131	0.25837	
x.sp500	0.060303	0.163408	0.369	0.712226	
x.VIX	-0.007284	0.01156	-0.63	0.528862	
x.ETF	0.200072	0.079639	2.512	0.012248	*
x.TIRY10	-0.010667	0.031977	-0.334	0.738798	
x.MFI	0.020998	0.005715	3.674	0.000259	***
x.RSI	0.014019	0.005641	2.485	0.013201	*
xlag1.Volumen	0.006827	0.001957	3.489	0.000519	***
xlag1.TCN	0.040169	0.052368	0.767	0.443341	
xlag1.Reservas	-0.044059	0.031129	-1.415	0.15746	
xlag1.M2	0.019234	0.059052	0.326	0.744752	
xlag1.BM	0.007968	0.0299	0.266	0.789956	
xlag1.EMBI	-0.014974	0.029271	-0.512	0.609145	
xlag1.interescp	-0.021566	0.012745	-1.692	0.091112	.
xlag1.interesUScp	-0.017961	0.026135	-0.687	0.49219	
xlag1.sp500	0.184274	0.17854	1.032	0.302415	
xlag1.VIX	0.014128	0.012168	1.161	0.246047	
xlag1.ETF	0.062917	0.086451	0.728	0.467026	
xlag1.TIRY10	0.027292	0.035138	0.777	0.43762	
xlag1.MFI	-0.002625	0.006098	-0.43	0.667006	

xlag1.RSI	-0.004783	0.005977	-0.8	0.423859	
xlag5.Volumen	0.008113	0.001708	4.749	0.00000254	***
xlag5.TCN	0.040437	0.027355	1.478	0.139843	
xlag5.Reservas	0.091819	0.015388	5.967	4.04E-09	***
xlag5.M2	-0.019166	0.031389	-0.611	0.541695	
xlag5.BM	0.0182	0.02093	0.87	0.384877	
xlag5.EMBI	-0.01338	0.01441	-0.929	0.353497	
xlag5.interescp	-0.022052	0.013367	-1.65	0.099496	.
xlag5.interesUScp	-0.012157	0.012873	-0.944	0.345328	
xlag5.sp500	0.121995	0.089883	1.357	0.175186	
xlag5.VIX	0.027054	0.006758	4.003	0.0000699	***
xlag5.ETF	0.094128	0.04203	2.24	0.02547	*
xlag5.TIRY10	0.002623	0.017573	0.149	0.881405	
xlag5.MFI	0.001081	0.002907	0.372	0.710183	
xlag5.RSI	0.005224	0.003447	1.516	0.130146	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01428 on 626 degrees of freedom
Multiple R-squared: 0.9905, Adjusted R-squared: 0.9899
F-statistic: 1560 on 42 and 626 DF, p-value: < 2.2e-16

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance = 0.05

Call:
gvlma(x = lmodel)

	Value	p-value	Decision
Global Stat	11.4599	0.02185	Assumptions NOT satisfied!
Skewness	0.3368	0.56167	Assumptions acceptable.
Kurtosis	4.1448	0.04176	Assumptions NOT satisfied!
Link Function	1.0571	0.30387	Assumptions acceptable.
Heteroscedasticity	5.9211	0.01496	Assumptions NOT satisfied!

Test de Phillips-Perron:

- Log(Merval) = Variable Y2:

Phillips-Perron Unit Root Test

data: y2
Dickey-Fuller = -2.8352, Truncation lag parameter = 6, p-value
= 0.2248

Phillips-Perron Unit Root Test

data: d.y2
Dickey-Fuller = -24.211, Truncation lag parameter = 6, p-value = 0.01

Log(Merval) es estacionaria de orden 1.

- xbm es la variable de la base monetaria sin rezagos:

Phillips-Perron Unit Root Test

data: xbm
Dickey-Fuller = -5.8463, Truncation lag parameter = 6, p-value = 0.01

- xetf es la variable ETF sin rezagos.

Phillips-Perron Unit Root Test

data: xetf
Dickey-Fuller = -1.7066, Truncation lag parameter = 6, p-value = 0.7025

Phillips-Perron Unit Root Test

data: d.xetf
Dickey-Fuller = -26.238, Truncation lag parameter = 6, p-value = 0.01

- x1tcn es la variable de 1 rezago del tipo de cambio nominal.

Phillips-Perron Unit Root Test

data: x1tcn
Dickey-Fuller = 1.0044, Truncation lag parameter = 6, p-value = 0.99

Phillips-Perron Unit Root Test

data: d.x1tcn
Dickey-Fuller = -25.77, Truncation lag parameter = 6, p-value = 0.01

- x5res es la variable del 5to rezago de las reservas internacionales.

Phillips-Perron Unit Root Test

data: x5res
Dickey-Fuller = -1.9222, Truncation lag parameter = 6, p-value = 0.6112

Phillips-Perron Unit Root Test

data: d.x5res
Dickey-Fuller = -25.654, Truncation lag parameter = 6, p-value = 0.01

Causalidad de Granger

- Base monetaria sobre Merval:

Diferenciada:

Granger causality test

Model 1: d.y2 ~ Lags(d.y2, 1:5) + Lags(d.xbm, 1:5)

Model 2: d.y2 ~ Lags(d.y2, 1:5)

	Res.Df	Df	F	Pr(>F)
1	652			
2	657	-5	2.3646	0.03847 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sin la primera observación:

Granger causality test

Model 1: d.y2 ~ Lags(d.y2, 1:5) + Lags(xxbm, 1:5)

Model 2: d.y2 ~ Lags(d.y2, 1:5)

	Res.Df	Df	F	Pr(>F)
1	652			
2	657	-5	2.4202	0.03455 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sin la última observación:

Granger causality test

Model 1: d.y2 ~ Lags(d.y2, 1:5) + Lags(xxbm, 1:5)

Model 2: d.y2 ~ Lags(d.y2, 1:5)

	Res.Df	Df	F	Pr(>F)
1	652			
2	657	-5	2.4202	0.03455 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- ETF sobre Merval:

Granger causality test

Model 1: $d.y2 \sim \text{Lags}(d.y2, 1:5) + \text{Lags}(d.xetf, 1:5)$

Model 2: $d.y2 \sim \text{Lags}(d.y2, 1:5)$

	Res.Df	Df	F	Pr(>F)
1	652			
2	657	-5	0.6475	0.6635

- TCN (1 rezago) sobre Merval:

Granger causality test

Model 1: $d.y2 \sim \text{Lags}(d.y2, 1:5) + \text{Lags}(d.x1tcn, 1:5)$

Model 2: $d.y2 \sim \text{Lags}(d.y2, 1:5)$

	Res.Df	Df	F	Pr(>F)
1	652			
2	657	-5	1.91	0.09063

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Reservas (5 rezagos) sobre Merval:

Granger causality test

Model 1: $d.y2 \sim \text{Lags}(d.y2, 1:5) + \text{Lags}(d.x5res, 1:5)$

Model 2: $d.y2 \sim \text{Lags}(d.y2, 1:5)$

	Res.Df	Df	F	Pr(>F)
1	652			
2	657	-5	0.169	0.974