

CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.



¿ PUEDEN LOS RENDIMIENTOS DE LAS ACCIONES DEL DOW 30 SER
PRONOSTICADAS POR BÚSQUEDAS DE INTERNET ?

TESINA

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN ECONOMÍA

PRESENTA

Carlos Aranda González Guerra

DIRECTOR DE LA TESINA: Dr. Daniel Vaughan Caro

CIUDAD DE MÉXICO

SEPTIEMBRE, 2016

A las personas que me apoyaron durante la licenciatura.

Agradecimientos

Agradezco a mi asesor de tesina el Dr. Daniel Vaughan por apoyarme, guiarme y enseñarme tanto con referencia al tema.

Agradezco a mis compañeros por tantas buenas vivencias y aprendizajes. Sin duda, son parte fundamental de lo que el CIDE significa para mí.

Agradezco a mi familia por su apoyo incondicional.

Índice general

1. Introducción	1
2. Revisión de Literatura y Motivación	3
2.1. Revisión de Literatura	3
2.2. Motivación	5
3. Datos	6
3.1. Selección de la variable dependiente	6
3.1.1. Especificación de la variable dependiente	7
3.2. Variable Independiente: búsquedas en Google	9
3.2.1. Introducción a Google Trends	9
3.2.2. Ajustando la serie de Google Trends	13
3.2.3. ¿ Importan los fines de semana?	15
4. Pruebas de causalidad de Granger	17
4.1. Primera prueba: causalidad directa	17
4.2. Segunda prueba: causalidad inversa	20
4.3. Tercera prueba: variaciones diarias	22
4.3.1. Causalidad inversa	24
5. Metodología	26
5.1. Modelo en diferencias	26

5.2. Modelo en variaciones	29
5.3. Análisis de resultados	31
6. Conclusiones	32
7. Anexo Estadístico	34
7.1. Pruebas de Raíz Unitaria	34
Referencias	37

Capítulo 1

Introducción

Pronosticar las acciones de una Bolsa de Valores ha sido un fenómeno de mucho interés para el sector académico y financiero. Uno de los enfoques para predecir el valor de las acciones, utiliza noticias y el surgimiento de nueva información. Para el sector académico resulta de interés pues podría indicar cambios en la manera en que se realizan las inversiones. Mientras que para el sector financiero, el pronóstico implicaría ganancias para los agentes que implementen los pronósticos. En este sentido, este trabajo buscará relaciones entre los rendimientos de una acción y su demanda de información vía internet.

En los últimos años se ha desarrollado una importante literatura que permite pronosticar el presente o el cercano futuro a partir de búsquedas de internet o actividad en redes sociales. Esto debido a una creciente organización de la información por compañías como Google o Twitter; además del desarrollo computacional para analizar la cantidad de datos vinculado al internet. Al igual que el sector financiero, han existido avances para pronosticar otras cuestiones económicas importantes como el desempleo; o de negocios la venta de carros, la demanda por una determinada ciudad turística o los ingresos de una determinada política. La metodología consta de estimar un ARIMAX de cada una de las acciones, enfocando la demanda de información como la variable explicativa para el precio de la acción.

La plataforma Google Trends provee información de la popularidad de la búsqueda de cualquier

palabra ya sea en Google o páginas de noticias. Los rangos de tiempo para la información que van desde años hasta intervalos de 8 minutos del mismo día de la consulta. El análisis utilizará un enfoque diario. El análisis estará por cada empresa del índice industrial Dow Jones (DOW 30 ó DJI¹); en las que se analizarán nombres y códigos de tablero de la compañía. Se considera el Dow Jones, ya que contiene una cantidad pequeña de empresas y están son un buena muestra del mercado americano.

Este análisis verifica que no exista una causalidad inversa; es decir, que el precio de las acciones pasado impacte en el número de búsquedas futuras. Para ello, el análisis utiliza una prueba de causalidad de Granger para verificar que sí existe una posible causalidad de las búsquedas al valor de las acciones.

Los resultados de esta metodología indican que sí posibilidad de explicar algunas acciones DOW 30, a partir de la búsqueda de códigos de tablero² en la búsqueda diaria. Por lo tanto, la contribución más importante es que la demanda de información financiera tiene un impacto directo en el precio de algunas acciones. La estructura del trabajo comprende una revisión de literatura de este tipo de experimentos y de las técnicas econométricas, financieras y computacionales. Posteriormente, se analizan los datos de manera estadística junto con la prueba de causalidad de Granger. La explicación del modelo para pronosticar los valores de cada acción. Por último, los resultados del modelo serán discutidos.

¹Por sus siglas es en inglés Dow Jones Industrial Average. Además DJI es, también, el código de tablero

²Un código de tablero, es el acrónimo utilizado para buscar a un determinado valor financiero en un servicio de información financiero(Bloomberg, Reuters, Infocell, etc).

Capítulo 2

Revisión de Literatura y Motivación

2.1. Revisión de Literatura

Este trabajo se desarrolla entre dos corrientes de literatura: la primera, la económico-financiera que se ha dedicado a modelos teóricos para la correcta inversión y entendimiento de las series de tiempo aplicado a valores financieros; la segunda, es una nueva literatura de aplicaciones económicas a partir del desarrollo del internet y la información que surge a partir de está. Algunas publicaciones recientes indican el potencial que tiene internet para las teorías económicas y sociales. Para Choi y Varian (2012) el uso de Google Trends sirve para *predecir el presente*. Las búsquedas en Google sirven para generar indicadores sobre todo tipo de acontecimientos que están sucediendo en el presente. Ejemplos de este tipo de indicadores son: venta de vehículos y auto-partes, la confianza del consumidor, empleo, precio de las casas, destinos turísticos . Según McLaren y Shanbhogue (2011), este tipo de información puede traer beneficios a encuestas previamente realizadas, en el sentido de que se amplía la serie de tiempo hasta el momento actual.

En el sector financiero es relativamente fácil encontrar información del presente¹. Lo que la literatura muestra es que hay intentos de utilizar herramientas derivadas del internet para pronosticar el futuro rendimiento de los activos financieros. Un ejemplo de este poder predictivo

¹Basta con tener un servicio de información especializado o entrar a una pagina como Yahoo! Finanzas

es el de Bollen, Mao, y Zeng (2011) los cuales argumentan que el estado de ánimo de los usuarios de Twitter tiene poder predictivo en el rendimiento del mercado accionario. Para su análisis hacen un análisis semántico de los tuits catalogando el estado de ánimo predominante en un intervalo de tiempo. Los autores descubren que la calma² puede ser un predictor, con causalidad en el sentido de Granger, del índice industrial Dow Jones(DJI). Aplican una red neuronal para encontrar los valores, según sus resultados encuentran en su muestra de validación que su modelo tiene un 80 % de probabilidad de obtener la dirección correcta del DJI. La idea de que las palabras pueden en Twitter tienen poder predictivo es importante, Zheludev, Smith, y Aste (2014) demuestran que la intención de la palabra es más importante que los volúmenes. Según Souza, Kolchyna, Treleaven, y Aste (2015) las palabras en Twitter tienen poder predictivo sobre el valor accionario de empresas de comerciales³. Además muestran que Twitter tiene un mayor impacto que las noticias para este tipo de compañías.

Existen otras contribuciones en el sentido más relacionado con las finanzas y las noticias, que gracias al internet son mucho más fáciles de cuantificar. Alanyali, Moat, y Preis (2013) analizaron la relación entre las noticias del *Financial Times* por cuatro años y su relación con el impacto en el precio de la acción. Encontraron que efectivamente existe una relación positiva entre el número de menciones de una acción en el *Financial times* y su número de transacciones diarias. La volatilidad, también, ha sido explorada por Risteski y Davcev (2014), usando un modelo de Google Trends para pronosticar la volatilidad de *CAC40*. Siguiendo sobre esta línea de pensamiento, Kristoufek (2013b) proporciona la idea de un portafolio diversificado considerando las acciones populares como las posiblemente más riesgosas. Kristoufek hace un análisis a partir de una serie de búsquedas de internet obtenidas con Google Trends para cada una de las acciones del DJI, según los resultados del autor este método de ponderación genera mejores retornos fuera de muestra que el mismo DJI.

²Como sentimiento expresado en los tuits de los usuarios

³Empresas con marcas como Apple, Mattel o Nike que tienen trato directo con el consumidor

2.2. Motivación

La importancia de los artículos anteriores es que demuestran que la información que proviene del internet tiene un impacto en el mundo financiero. En particular, los avances en el mercado accionario se centran en utilizar las estrategias tecnológicas para pronosticar volatilidad o rendimiento de los índices. Este trabajo es relevante con respecto a los anteriores ya que busca dos contribuciones nuevas.

La primera contribución que este trabajo busca encontrar una relación causal entre las búsquedas en Google y los rendimientos de acciones individualmente. Para ser más preciso, se busca la relación causal de cambios en número de búsquedas de Google genera cambios en el precio de las acciones. Atrás de está búsqueda de causalidad, esta la idea de que una acción más buscada tiene mayor interés para los inversionistas. La primera contribución resulta relevante ya que implica que los inversionistas realizan ciertas acciones en común antes de tomar de decisiones relevantes antes de tomar acciones relevantes conforme a una posición financiera. En este caso la causalidad implicaría que dependiendo al número de búsquedas en Google cambiará las acciones de los individuos y por ende el precio de las acciones.

La segunda, este trabajo busca si las búsquedas en Google pueden ser utilizadas para pronosticar rendimientos accionarios. Este trabajo se diferencia de los sus anteriores por buscar un pronóstico individual para cada acción del DJI. Esta contribución sería valiosa ya que implicaría que las acciones la demanda de información, en este caso las búsquedas en Google, son no sólo determinantes si no tienen poder predictivo en los precios de las acciones.

Capítulo 3

Datos

3.1. Selección de la variable dependiente

El objetivo de este trabajo es encontrar una relación entre el precio de las acciones del DJI y sus respectivas búsquedas en Google. Se elige el índice industrial Dow Jones[DJI], por ser un índice popular en el número de empresas que representa. El DJI representa una cartera diversificada de la bolsas de valores de los Estados Unidos. El hecho que el DJI tenga 30 acciones hace que sea de fácil manejo para investigar individualmente a comparación de otros índices populares como el S&P 500, NASDAQ composite y NYSE composite que, también, representan los rendimientos de las bolsas de valores. De acuerdo a la revisión de literatura el DJI es un índice utilizado en este tipo de experimentos, ejemplo de ello son los trabajos de Bollen y cols. (2011) y Kristoufek (2013b).

Los componentes del DJI han variado durante el tiempo para adecuarse ante los cambios en la estructura del mercado. Por lo que es necesario especificar que componentes tiene el DJI a la fecha de publicación de este trabajo. En la Tabla 1 encontramos las treinta empresas que componen el DJI en marzo del 2016. Estas empresas representan diversas industrias de la economía norteamericana como lo son: tecnologías de la información, banca, finanzas, entretenimiento, televisión, telecomunicaciones, energía, comerciales, farmacéutico, médico, entre otras. En ge-

neral las compañías son líderes en sus respectivos sectores y cuentan con presencia internacional. El DJI sintetiza los rendimientos del mercado eligiendo empresas de distintos rubros.

Tabla 1: Componentes del Índice Industrial Dow Jones DJI¹

# Acción	Nombre de la Compañía	Ticker/Symbol	# Acción	Nombre de la Compañía	Ticker/Symbol
1	Apple Inc.	AAPL	15	JPMorgan Chase & Co.	JPM
2	American Express Company	AXP	16	The Coca-Cola Company	KO
3	The Boeing Company	BA	17	McDonald's Corp.	MCD
4	Caterpillar Inc.	CAT	18	3M Company	MMM
5	Cisco Systems, Inc.	CSCO	19	Merck & Co. Inc.	MRK
6	Chevron Corporation	CVX	20	Microsoft Corporation	MSFT
7	E. I. du Pont de Nemours and Company	DD	21	NIKE, Inc.	NKE
8	The Walt Disney Company	DIS	22	Pfizer Inc.	PFE
9	General Electric Company	GE	23	The Procter & Gamble Company	PG
10	The Goldman Sachs Group, Inc.	GS	24	The Travelers Companies, Inc.	TRV
11	The Home Depot, Inc.	HD	25	UnitedHealth Group Incorporated	UNH
12	International Business Machines Corporation	IBM	26	United Technologies Corporation	UTX
13	Intel Corporation	INTC	27	Visa Inc.	V
14	Johnson & Johnson	JNJ	28	Verizon Communications Inc.	VZ
15	JPMorgan Chase & Co.	JPM	29	Wal-Mart Stores Inc.	WMT
16	The Coca-Cola Company	KO	30	Exxon Mobil Corporation	XOM

3.1.1. Especificación de la variable dependiente

La variable de interés son los precios al cierre de cada día. Estos datos se obtuvieron del servicio de información de Bloomberg para cada una de las acciones de la Tabla 1 desde el año 2000². El precio de las acciones no tiene un procesos estacionario. En el anexo estadístico de raíz unitaria, podemos apreciar que las acciones, son $I(1)$. Por lo tanto, deben ser diferenciadas para que podamos realizar inferencia con la metodología ARIMAX. En el Gráfico 1 se representa el comportamiento de precios de la acción de la empresa Johnson & Johnson [JNJ].

¹Fuente: <https://es-us.finanzas.yahoo.com/q/cp?s=%5EDJI+Componentes>, revisado Marzo del 2016

²Exceptuando, la empresa Visa (V), todas las demás empresas ya eran publicas para ese año. Visa inc. comienza hacer publica a partir del 20/03/2008

Gráfico 1: Comportamiento del precio de una acción



La primera serie del Gráfico 1, es el nivel de precios diarios al cierre de operaciones para la compañía Johnson & Johnson [JNJ]. El nivel se expresa como $P_{i,t}$. En el primer panel del Gráfico 1 sugiere que una presencia de raíz unitaria³. Por ello, se pueden tener dos alternativas para su análisis como variable independiente:

1. Diferenciar la variable bastará para que sea estacionaria. Esto implica que $\delta P_{i,t} = P_{i,t} - P_{i,t-1}$. La segunda serie del Gráfico 1 se realiza este ajuste, esta nueva serie sugiere ser $I(0)$. La evidencia de que el nivel del precio de las acciones es $I(1)$ está en el anexo estadístico.
2. La variación diaria resulta atractiva para el mercado financiero pues implica pronosticar los rendimientos en el corto plazo. La variación diaria queda definida de la siguiente forma: $\Delta P_{i,t} = \frac{P_{i,t}}{P_{i,t-1}} - 1$. Este tipo de transformación se encuentra en la tercera serie del Gráfico 1. La variación de precios es un proceso $I(0)$, esto se comprueba en el anexo estadístico, y por lo tanto se puede aplicar la metodología ARIMAX.

³Revisar anexo estadístico

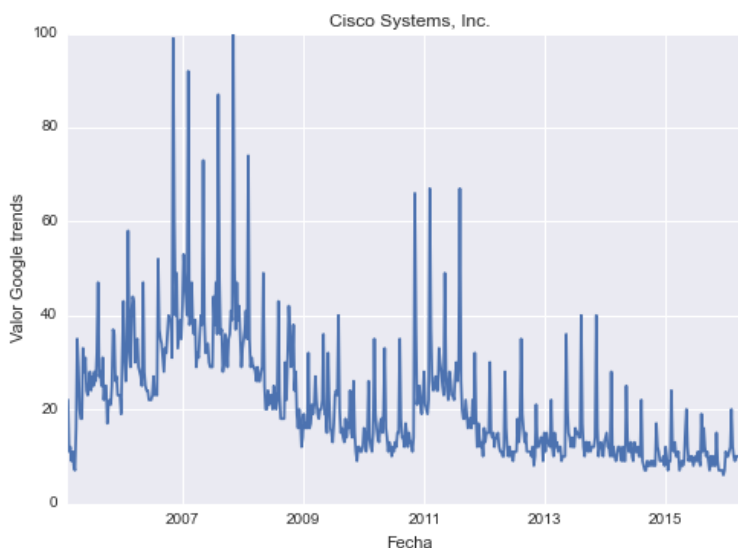
Ambos enfoques se realizarán en las estimaciones, pues ambas resultan útiles para los fines del documento. La primera versión nos dará parámetros en términos de dinero los cuales pueden resultar valiosos para la interpretación. El segundo enfoque es valioso pues implican los rendimientos diarios de la acción. Además, obtener un modelo preciso en el cual tenga poder predictivo fuera de muestra será el objetivo de la metodología.

3.2. Variable Independiente: búsquedas en Google

3.2.1. Introducción a Google Trends

Google Trends es un servicio de información de Google, que provee indicadores de cuántas veces se ha buscado un término o palabra en el buscador de Google. Idealmente, nos gustaría conocer el número de búsquedas totales de cada término dentro de un determinado intervalo de tiempo. Desafortunadamente, Google no reporta este tipo de información sino que reporta Google Trends es un índice como en el caso del Gráfico 2. El Gráfico 2 reporta el ticker de la empresa Cisco [CSCO] en la región de los Estados Unidos de América.

Gráfico 2: Serie de Google Trends



búsqueda tipo 1, valores semanales

El Gráfico 2 es un ejemplo de que, el indicador de Google Trends está definido como: $G_t \in [0, 100]$. Según las propias palabras de Google⁴:

"Los números reflejan el interés de búsqueda en relación con el mayor valor de un gráfico en una región y en un periodo determinados. Un valor de 100 indica la popularidad máxima de un término, mientras que 50 y 0 indican una popularidad que es la mitad o inferior al 1 %, respectivamente, en relación al mayor valor."

En el Gráfico 2, se observa que en dos ocasiones el índice llega a ser 100. En otros casos esto también sucede por lo tanto las reglas para definir los máximos no están bien definidas. Por lo tanto es posible de que en casos de más de 100 den G_t sean variables censuradas. En caso de que no hubiera búsquedas en esa temporalidad existen dos opciones: primera, que reporte ceros; en segundo caso que no reporte nada implicando que el término no es tan popular como para generar una tendencia.

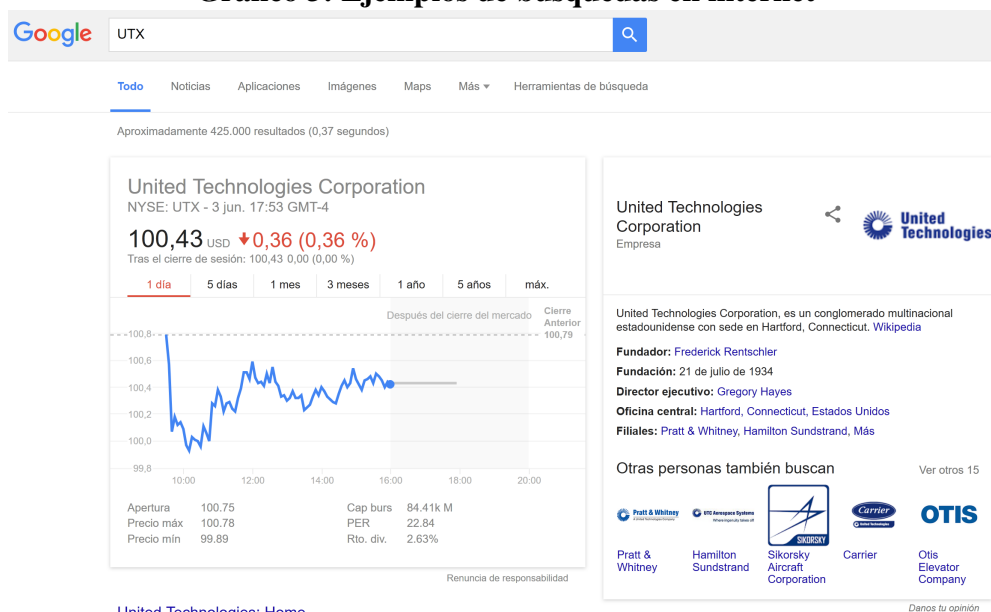
Tipos de búsquedas La variable explicativa son las búsquedas de Google de cada acción obtenidas a partir de la plataforma Google Trends. Para fines de la investigación se utilizan únicamente las búsquedas realizadas en Estados Unidos. Al igual que la metodología de Kristoufek (2013b) y Risteski y Davcev (2014), consta en buscar el *ticker* o código de tablero de cada acción y añadir otro tipo de palabras clave. Para este trabajo se usaron los siguientes tipos de búsquedas:

1. **ticker** : esto implica que se buscó el código de tablero de la acción. Ejemplo para Disney : "**DIS**"
2. **ticker + Stock** : esto implica que se buscó el código de tablero pero se añadió la palabra "stock". Ejemplo para 3M company: "**MMM Stock**"
3. **Stock + ticker**: similar al caso anterior, aunque tiene índices de Google Trends distintos. Ejemplo para 3M company: "**Stock MMM**"
4. **ticker + News**: es posible que se busquen noticias relacionadas a la acción. Ejemplo de Goldman Sachs : "**GS News** "

⁴Fuente de datos: Tendencias de búsqueda de Google (www.google.com/trends) Revisado en agosto del 2016

Es de notar que los términos de Google Trends no son sensibles a mayúsculas y minúsculas⁵. Este tipo de búsquedas, suponemos sólo son realizadas por personas interesadas con el sector financiero. Lo anterior, con el argumento que un usuario que realiza este tipo de búsquedas es por que es un potencial inversionista. El Gráfico 3 ejemplifica el tipo de información que provee Google en algunos tipos de búsquedas. En el Gráfico 3 se busca el código de tablero de la empresa United Technologies Corporation [UTX], Google da información acerca de la empresa y de su rendimiento accionario. Ante esto, se puede suponer que alguien que busca una cosa tan específica como código de tablero o *ticker* entonces le interesa el mercado accionario. Esto se refuerza aún más, en los tipos de búsquedas 2 y 3 donde se usa la palabra "Stock"

Gráfico 3: Ejemplos de búsquedas en internet



Existencia de la serie No en todos los casos Google Trends provee la serie de información. La Tabla 2 es un ejemplo desde donde empieza la serie para cada acción de DJI. Debido a larga historia se considera empezar las regresiones a partir de 2008 o empezar desde donde exista información para cada tipo de búsqueda.

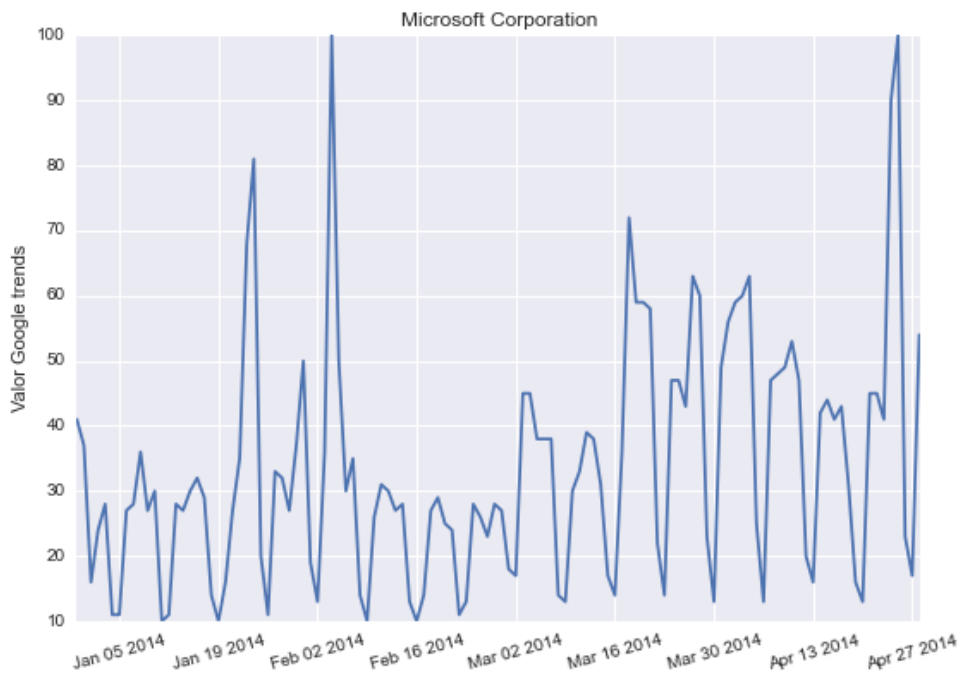
⁵Esto implica que no importan si se escribió con minúsculas o mayúsculas. Ejemplo: "mMm" es equivalente a "MMM" y a "mMM".

Tabla 2: Existencia de la serie en Google Trends

# Acción	Ticker	Búsqueda 1	Búsqueda 2	Búsqueda 3	Búsqueda 4	Fecha de inicio
1	AAPL	2006	2006	2006	2006	2008
2	AXP	2006	2006	2006	x	2008
3	BA	2006	2006	2006	2006	2008
4	CAT	2006	2006	2006	2006	2008
5	CSCO	2009	2009	2009	2009	2009
6	CVX	2011	2011	2011	2011	2011
7	DD	2006	2006	2006	2006	2008
8	DIS	2006	2006	2006	2006	2008
9	GE	2006	2006	2006	2006	2008
10	GS	2006	2006	2006	2006	2008
11	HD	2006	2006	2006	2006	2008
12	IBM	2006	2006	2006	2006	2008
13	INTC	2006	2006	2006	2006	2008
14	JNJ	2007	2007	2007	2007	2008
15	JPM	2008	2008	2008	2008	2008
16	KO	2006	2006	2006	2006	2008
17	MCD	2006	2006	2006	2008	2008
18	MMM	2008	2008	2008	x	2008
19	MRK	2006	2006	2006	x	2008
20	MSFT	2006	2006	2011	2006	2011
21	NKE	2013	2013	2013	2013	2013
22	PFE	2006	2006	2006	x	2008
23	PG	2006	2006	2006	2006	2008
24	TRV	2006	2009	x	x	2009
25	UNH	2006	2006	2006	2006	2008
26	UTX	2006	2006	2006	x	2008
27	V	2006	2006	2006	2006	2008
28	VZ	2006	2006	2006	2008	2008
29	WMT	2006	2006	2006	2006	2008
30	XOM	2006	2006	2006	2007	2008

Problemas con series diarias Google Trends no provee datos diarios para intervalos mayores que 90 días. Como las series de Google Trends están normalizadas dentro de cada intervalo no son comparables entre intervalos. Ejemplo de esta problemática es el Gráfico 4. El Gráfico 4 es una serie compuesta por 2 bloques de 60 días. En cada uno de los bloques existe un máximo relativo, el cual es determinante para el nivel de cada serie. Además, los máximos relativos de cada bloque de información no son comparables.

Gráfico 4: Ejemplo Google Trends *Gt*



búsqueda tipo 1, dos bloques de información

3.2.2. Ajustando la serie de Google Trends

Modificaremos un poco la metodología de Johansson (2014) para generar una serie donde los valores diarios estarán ponderados por la serie semanal. La solución implica estos cambios:

1. Obtener los distintos conjuntos de datos diarios y los datos semanales. Ya que los valores semanales sí están ponderados globalmente.

2. Para obtener el factor de ajuste se divide el dato de la semana entre el primer diario de la misma semana. Fa es el factor de ajuste, Gts es el valor de Google Trends semanal y Gtd es el valor de Google Trends diario.

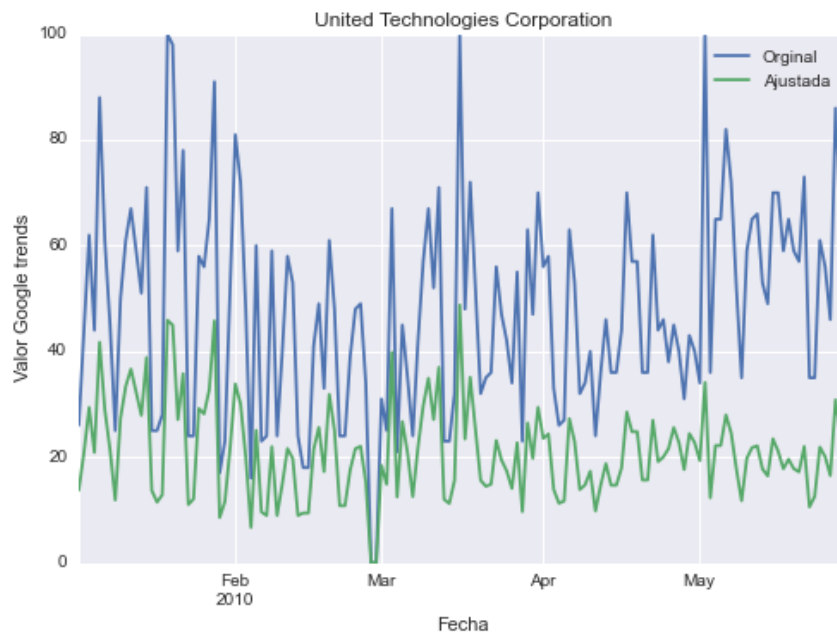
$$Fa = \frac{Gts}{\frac{\sum_i^7 Gtd_i}{7}}$$

3. Multiplicar el factor de ajuste (Fa) por cada miembro de la semana. Ga es el valor de Google Trends ajustado

$$Ga_i = Fa \times Gd$$

De esta manera se puede generar una serie de Google Trends diaria, donde estos valores sí son comparables a través de la serie. El Gráfico 5 es ejemplo de utilizar esta metodología. La bondad de este ajuste es que utiliza las series semanales que sí tienen una posición relativa entre toda la historia de búsquedas. La serie ajustada sí es comparable en todo momento.

Gráfico 5: Comparación ajustada Vs. original



búsqueda tipo 1

La serie ajustada pierde dos cualidades importantes. Primero, la series de Google Trends ajustada pueden ser mayor a 100. Segundo, la serie ya no esta compuesta por números enteros. Estos cambios no afectan en la estimación del modelo. Además, estos cambios son el motivo por el que se puedan comparar en cada momento del tiempo.

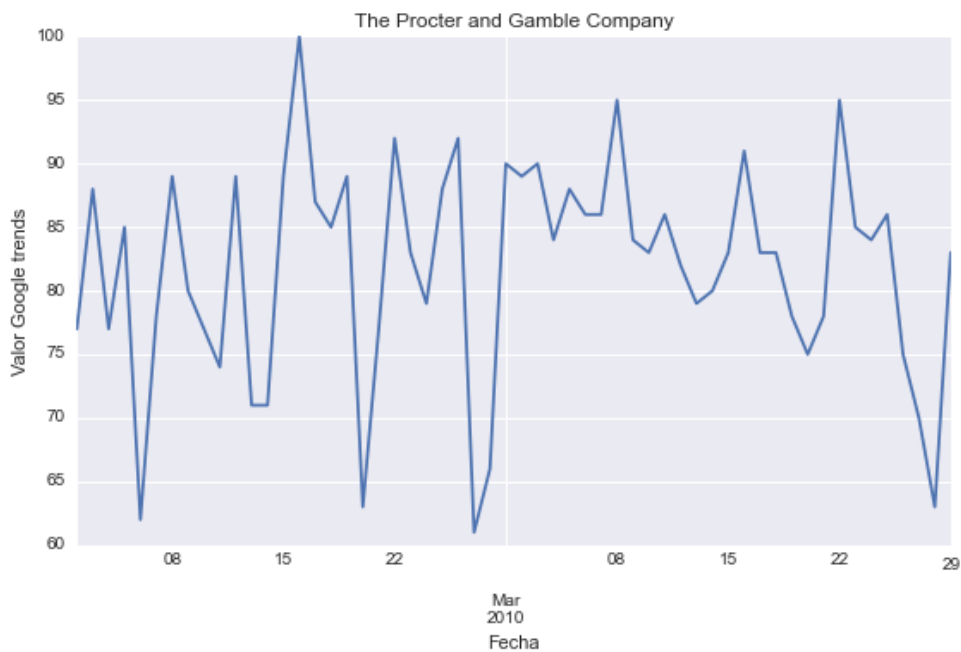
3.2.3. ¿ Importan los fines de semana?

Por un lado la variable dependiente, los precios de las acciones varían diario de lunes a viernes. Por otro lado las variables independientes, las series de Google Trends tienen datos de domingo a sábado. Más adelante se especificará los modelos con diferencias diarias y por lo tanto surge la pregunta si usar los valores de fin de semana. La principal diferencia es utilizar datos de viernes - lunes o domingo-lunes para las series de Google Trends.

Por lo tanto se analizaran las diferencias entre estos casos. Para determinar, que tipo de serie usar ilustrar las dos series. El Gráfico 6 y 7 son la búsqueda tipo 1 para la empresa Procter & Gamble [PG] con una series para febrero y marzo del 2010. El Gráfico 6 está definido desde $G_t \in [61, 100]$ los valores inferiores hacen 60 son únicamente compuestos de fin de semana. El Gráfico 7 no contiene valores de fin de semana y es una serie con un menor rango relativo a la del Gráfico 6. El hecho de que las personas busquen menos en fines de semana, implica que existe mucho menos interés por los valores accionarios cuando los mercados están cerrados. Por lo tanto, en valores en lunes (ante un variable de diferencias diarias) usualmente seran más grandes y positivas relativo al resto de los valores en la semana.

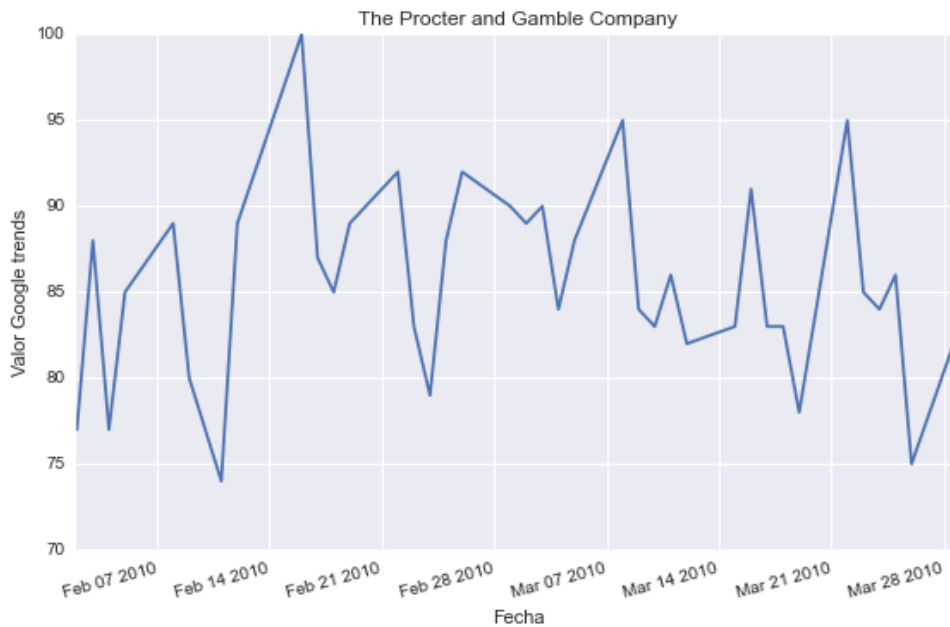
Una metodología similar fue utilizada por Risteski y Davcev (2014) en el que resuelven los valores en fines de semana para pronosticar la volatilidad de un índice financiero *CAC40*. Similarmente, Alanyali y cols. (2013) utilizan información de menciones en el *Financial Times* y descartan los fines fines de semana y los días que no abren las bolsas. Por ello utilizaremos, únicamente los días que están abiertas las bolsas de valores.

Gráfico 6: Google Trends con valores de fin de semana



búsqueda tipo 1 sin ajuste semanal

Gráfico 7: Google Trends sin valores de fin de semana



búsqueda tipo 1 sin ajuste semanal

Capítulo 4

Pruebas de causalidad de Granger

En la motivación se aclaró que uno de los objetivos era comprobar si las búsquedas de internet tienen impacto en los rendimientos accionarios. Surge el problema de la causalidad, ya que existen motivos para pensar que el cambio en el precio de las acciones sea el que genera las búsquedas. Es decir que derivado de una escena de volatilidad accionaria provoque un mayor número de búsquedas en Google.

4.1. Primera prueba: causalidad directa

Se plantea una prueba de causalidad de Granger (1969) para responder ¿ si las búsquedas de internet pueden afectar los precios de las acciones ? Lo que hace la prueba de causalidad de Granger, es realizar una regresión donde se incluye el pasado de la variable dependiente e independiente. Si los parámetros de la variable independiente son significativos a pesar de controlar por la historia de las variables dependientes entonces obtenemos causalidad de Granger. La causalidad de Granger implica que la variable independiente sí tiene efecto al definir los valores de la variable dependiente. En este primer caso se especifica el siguiente modelo:

$$\delta P_{C_{i,t}} = \alpha + \sum_r^H \phi_r \delta P_{C_{i,t-r}} + \sum_r^H \beta_r \delta B_{m,t-r} + \varepsilon$$

$\delta P_{c_{i,t}}$ se refiera al la diferencia de precio de cierre de la acción i en el tiempo t , α se refiere a la constante, ϕ se refiere al parámetro asociado al pasado de $\delta P_{c_{i,t}}$, $\delta B_{i,m,t-r}$ se refiera a la diferencia Google Trends del índice ajustado, β se refiere al parámetro asociado a $\delta B_{i,m,t-r}$, ε es el error. $H = 6$ al ser el de una semana habil más uno. Además, se planea calcular la siguiente ecuación:

$$\delta P_{c_{i,t}} = \alpha + \sum_r^H \phi_r \delta P_{c_{i,t-r}} + \sum_r^H \beta_r B_{m,t-r} + \varepsilon$$

La única diferencia es que se considera el valor de Google Trends ajustado en vez de por diferencias.

En la Tabla 3 se hacen referencia a estos primeros modelos y se reportan los valores- P de la prueba F de las regresiones. Recordando que abajo de 0.10 se considera que es una prueba significativa. La Tabla 3 reporta el nombre de la acción con base en los símbolos (revisar Tabla 1) y reporta una columna de resultados. Las columnas y el renglon de resultado indican cuantas veces resultó el valor-P con un valor significativo. Los resultados de la Tabla 3 indican que en la mayoría de los casos, las búsquedas en Google sí tienen un efecto el cambios de los precios de las acciones en al menos una de las variables independientes. No obstante no existe ninguna acción la cual sea susceptible a los cuatro métodos de búsqueda. Las empresas que son más susceptibles son: Apple (7), IBM(6), JP Morgan Chase (5) y Disney(5) . La Tabla 3, también, indica cuales son las mejores búsquedas que tienen más impacto frente al cambio de precio de acciones. Ningún tipo de búsqueda resulta contundente que pueda afectar el precio de la acciones. La búsqueda 2 (**ticker+"Stock"**) resulta la más efectiva obteniendo poco más de un tercio de las acciones del DJI como significativo. La búsqueda 4 (**ticker+"News"**) tampoco obtiene un resultado de casi un tercio de las acciones. La Tabla 3 sugiere que es mejor la diferencias en los índices ajustados son mejores explicando comparado con el valor del índice ajustado; esto sugiere que los cambios diarios de solicitud de información son más importantes que la relevancia general de una acción.

Tabla 3: Resultados de la prueba de causalidad de Granger Valores- P

Acción	B1-Nivel	B1-Diferencia	B2-Nivel	B2-Diferencia	B3-Nivel	B3-Diferencia	B4-Nivel	B4-Diferencia	Resultado
AAPL	0.129531	0.093470	0.000377	0.000176	0.001670	0.000909	0.098558	0.000909	7
AXP	0.019747	0.029034	0.630971	0.840626	0.808506	0.843106	1.000000	1.000000	2
BA	0.914843	0.914547	0.513059	0.037347	0.576341	0.264036	0.910409	0.264036	1
CAT	0.202920	0.556960	0.563580	0.031350	0.474024	0.012322	0.321752	0.012322	3
CSCO	0.003452	0.032995	0.484162	0.394630	0.396805	0.493689	0.665062	0.493689	2
CVX	0.151654	0.441452	0.621516	0.413605	0.436436	0.488661	0.460622	0.488661	0
DD	0.272666	0.704283	0.253055	0.910812	0.208906	0.613324	0.332709	0.613324	0
DIS	0.504372	0.523462	0.000625	0.000687	0.000655	0.005205	0.550594	0.005205	5
GE	0.703440	0.740048	0.313354	0.284872	0.009167	0.111645	0.403424	0.111645	1
GS	0.438244	0.200928	0.382684	0.149251	0.600517	0.503387	0.172763	0.503387	0
HD	0.002535	0.200321	0.102661	0.199126	0.040246	0.006984	0.186682	0.006984	4
IBM	0.617737	0.158631	0.000136	0.006055	0.000233	0.002709	0.000024	0.002709	6
INTC	0.632020	0.066984	0.096521	0.060591	0.566893	0.535129	0.701556	0.535129	3
JNJ	0.839645	0.883847	0.268722	0.071458	0.060439	0.001318	0.651978	0.001318	4
JPM	0.131292	0.539268	0.000851	0.039686	0.000279	0.000886	0.168679	0.000886	5
KO	0.391113	0.301589	0.748117	0.684955	0.944065	0.734189	0.706837	0.734189	0
MCD	0.124531	0.002623	0.760934	0.961749	0.690865	0.772648	0.036368	0.772648	2
MMM	0.378889	0.170890	0.280535	0.968767	0.243796	0.705742	1.000000	1.000000	0
MRK	0.870650	0.901266	0.806276	0.542215	0.769105	0.948216	1.000000	1.000000	0
MSFT	0.063526	0.138861	0.197699	0.394753	0.246525	0.342296	0.319039	0.342296	1
NKE	0.558417	0.686517	0.342303	0.697065	0.027518	0.071952	0.161287	0.071952	3
PFE	0.512414	0.854763	0.805691	0.633429	0.811555	0.275323	0.008545	0.275323	1
PG	0.842647	0.862970	0.558580	0.108956	0.938235	0.375082	0.487707	0.375082	0
TRV	0.903320	0.801310	0.060488	0.014161	1.000000	1.000000	1.000000	1.000000	2
UNH	0.909881	0.765479	0.817765	0.758859	0.307187	0.688287	0.677119	0.688287	0
UTX	0.620280	0.913920	0.748864	0.256540	0.701573	0.547440	0.014638	0.547440	1
V	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0
VZ	0.358402	0.291315	0.149562	0.047054	0.366697	0.426384	0.317075	0.426384	1
WMT	0.501158	0.280540	0.753483	0.044029	0.228866	0.059133	0.753699	0.059133	3
XOM	0.723646	0.007699	0.725442	0.946304	0.252174	0.710735	0.033484	0.710735	2
Resultado	4	6	6	11	8	9	6	9	-

4.2. Segunda prueba: causalidad inversa

Como planteamos al inicio del capítulo, es importante conocer si los efectos no son inversos. Es decir, que los cambios en los precios de las acciones sean generadoras de búsquedas interesa saber. Para responder la segunda pregunta es necesario especificar el siguiente modelo:

$$\delta B_{m,t} = \alpha + \sum_r^H \phi_r \delta B_{m,t-r} + \sum_r^H \beta_r \delta PC_{i,t-r} + \varepsilon$$

La diferencia principal con respecto a la primera prueba es que, ahora, la variable dependiente son los índices de Google Trends ajustados y los cambios en los precios la independiente. Al igual que en el caso anterior consideraremos los valores en niveles:

$$B_{m,t} = \alpha + \sum_r^H \phi_r B_{m,t-r} + \sum_r^H \beta_r \delta PC_{i,t-r} + \varepsilon$$

En la Tabla 4 encontramos los resultados de los valores-P de las pruebas de causalidad de Granger, los resultados se presentan de forma similar a la Tabla 3 exceptuando que ahora la causalidad va en la dirección opuesta. Los resultados de la Tabla 4 indican que existe una mayor causalidad por parte del cambio en el precio de las acciones. Principalmente, tienen efecto sobre las búsquedas tipo 2 y 3 (*Stock + ticker*). Al igual que la causalidad directa, la metodología tiene más impacto cuando es en diferencias. El cambio en precios de las acciones es causa de que exista un cambio en número búsquedas tipo 2, en más de 50 % de las ocasiones. Los resultados de la Tabla 4 indican que las acciones que más generan búsquedas son: Du pont (6), General Electric (7) y Goldman Sachs (7). La Tabla 4 provee solida evidencia, que también, el cambio en el precio de las acciones genera un cambio en el número de búsquedas. Por lo tanto sí existe causalidad inversa.

Tabla 4: Pruebas de causalidad Granger Valores-P causalidad inversa

Acción	B1-Nivel	B1-Diferencia	B2-Nivel	B2-Diferencia	B3-Nivel	B3-Diferencia	B4-Nivel	B4-Diferencia	Resultado
AAPL	0.820	0.255	0.029	0.000	0.107	0.000	0.413	0.099	4
AXP	0.492	0.700	0.881	0.875	0.578	0.920	1.000	1.000	0
BA	0.706	0.517	0.020	0.038	0.061	0.010	0.982	0.910	4
CAT	0.310	0.516	0.150	0.474	0.467	0.300	0.567	0.322	0
CSCO	0.459	0.521	0.921	0.655	0.937	0.816	0.690	0.665	0
CVX	0.001	0.039	0.002	0.015	0.001	0.004	0.456	0.461	6
DD	0.751	0.527	0.189	0.086	0.977	0.335	0.120	0.333	1
DIS	0.270	0.637	0.000	0.000	0.000	0.000	0.876	0.551	4
GE	0.001	0.001	0.000	0.000	0.000	0.000	0.019	0.403	7
GS	0.016	0.031	0.000	0.000	0.000	0.002	0.073	0.173	7
HD	0.169	0.137	0.599	0.071	0.880	0.553	0.947	0.187	1
IBM	1.000	0.675	0.207	0.359	0.380	0.181	0.127	0.000	1
INTC	0.019	0.488	0.000	0.093	0.001	0.112	0.522	0.702	4
JNJ	0.852	0.862	0.037	0.015	0.058	0.028	0.499	0.652	4
JPM	0.208	0.495	0.007	0.238	0.000	0.018	0.614	0.169	3
KO	0.975	0.777	0.415	0.385	0.305	0.084	0.950	0.707	1
MCD	0.964	0.665	0.002	0.383	0.000	0.011	0.036	0.036	5
MMM	0.960	0.707	0.722	0.367	0.948	0.418	1.000	1.000	0
MRK	0.930	0.830	0.387	0.694	0.890	0.928	1.000	1.000	0
MSFT	0.790	0.062	0.970	0.006	0.023	0.001	0.434	0.319	4
NKE	0.675	0.995	0.849	0.033	0.796	0.031	0.483	0.161	2
PFE	0.348	0.077	0.949	0.916	0.052	0.108	0.829	0.009	3
PG	0.768	0.807	0.721	0.699	0.064	0.674	0.875	0.488	1
TRV	0.944	0.928	0.051	0.040	1.000	1.000	1.000	1.000	2
UNH	0.634	0.571	0.171	0.679	0.933	0.490	0.885	0.677	0
UTX	0.635	0.951	0.366	0.314	0.539	0.812	1.000	1.000	0
V	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0
VZ	0.273	0.098	0.213	0.045	0.469	0.355	0.738	0.317	2
WMT	0.334	0.314	0.891	0.066	0.803	0.223	0.572	0.754	1
XOM	0.564	0.231	0.004	0.006	0.109	0.032	0.087	0.033	5
Resultado	4	6	12	16	12	13	4	5	-

4.3. Tercera prueba: variaciones diarias

Como se señaló el capítulo 3 sección 2 puede que la metodología de diferencias en precios de cierre diarios no sea la mejor para resolver. El enfoque de variaciones diarias, resulta atractivo ya que implica los rendimientos diarios de las acciones, en vez del cambio en precios. En esta sección se aplicaran las pruebas de causalidad de Granger cambiando la variable dependiente. El primer modelo a especificar es:

$$\Delta P_{C_{i,t}} = \alpha + \sum_r^H \phi_r \Delta P_{C_{i,t-r}} + \sum_r^H \beta_r \delta B_{m,t-r} + \varepsilon$$

La única diferencia con el modelo de la primera prueba es que la variable dependiente está en variaciones diarias. Al igual que la prueba una se considera el nivel del índice ajustado de Google Trends como variable explicativa:

$$\Delta P_{C_{i,t}} = \alpha + \sum_r^H \phi_r \Delta P_{C_{i,t-r}} + \sum_r^H \beta_r B_{m,t-r} + \varepsilon$$

La Tabla 5 indica que los resultados son ligeramente inferiores de la Tabla 3, esto ya que en la Tabla 3 hay 59 resultados significativos relativo a los 56 de la Tabla 5. Los resultados sugieren no existir muchas diferencias relativo a los métodos, incluso la diferencia entre variable explicativa diferenciada y en nivel no es tan notoria. Las acciones más susceptibles bajo este esquema son: IBM (7), GE(5) y JP Morgan Chase(5). IBM y JPM ya habían aparecido como acciones susceptibles en la Tabla 3. En la Tabla 5 sólo 7 acciones no tienen ni un sola regresión significativa, por lo que un 76 % de las acciones al menos arrojan una búsqueda significativa. La Tabla 5 respalda el hecho de que algunas de las búsquedas en internet de acciones sí afectan el precio de las acciones, haciendo énfasis en particular en los rendimientos accionarios.

Tabla 5 : Prueba de causalidad de Granger Variable dependiente en variaciones diarias

Acción	B1-Nivel	B1-Diferencia	B2-Nivel	B2-Diferencia	B3-Nivel	B3-Diferencia	B4-Nivel	B4-Diferencia	Resultado
AAPL	0.040	0.416	0.023	0.344	0.032	0.417	0.053	0.417	4
AXP	0.000	0.002	0.992	0.983	0.951	0.964	1.000	1.000	2
BA	0.546	0.784	0.472	0.105	0.393	0.446	0.862	0.446	0
CAT	0.379	0.731	0.823	0.082	0.412	0.015	0.904	0.015	3
CSCO	0.058	0.334	0.449	0.357	0.269	0.525	0.483	0.525	1
CVX	0.059	0.209	0.451	0.276	0.259	0.326	0.473	0.326	1
DD	0.550	0.886	0.540	0.952	0.515	0.884	0.389	0.884	0
DIS	0.525	0.708	0.635	0.787	0.684	0.926	0.770	0.926	0
GE	0.912	0.380	0.019	0.003	0.000	0.003	0.583	0.003	5
GS	0.514	0.101	0.116	0.008	0.072	0.028	0.436	0.028	4
HD	0.036	0.286	0.157	0.344	0.118	0.060	0.048	0.060	4
IBM	0.520	0.047	0.004	0.067	0.006	0.054	0.002	0.054	7
INTC	0.499	0.041	0.194	0.064	0.564	0.351	0.512	0.351	2
JNJ	0.735	0.851	0.315	0.218	0.030	0.001	0.500	0.001	3
JPM	0.094	0.363	0.001	0.102	0.001	0.003	0.186	0.003	5
KO	0.374	0.545	0.733	0.794	0.835	0.708	0.785	0.708	0
MCD	0.003	0.000	0.972	0.993	0.775	0.793	0.001	0.793	3
MMM	0.200	0.069	0.649	0.988	0.290	0.706	1.000	1.000	1
MRK	0.515	0.664	0.636	0.156	0.819	0.934	1.000	1.000	0
MSFT	0.070	0.162	0.298	0.615	0.778	0.662	0.281	0.662	1
NKE	0.531	0.583	0.454	0.724	0.045	0.089	0.191	0.089	3
PFE	0.611	0.785	0.939	0.829	0.830	0.496	0.002	0.496	1
PG	0.790	0.894	0.670	0.184	0.935	0.467	0.489	0.467	0
TRV	0.911	0.791	0.031	0.006	1.000	1.000	1.000	1.000	2
UNH	0.789	0.274	0.962	0.907	0.201	0.517	0.517	0.517	0
UTX	0.497	0.922	0.738	0.440	0.638	0.657	0.025	0.657	1
V	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0
VZ	0.147	0.130	0.245	0.095	0.369	0.597	0.191	0.597	1
WMT	0.637	0.215	0.813	0.107	0.309	0.149	0.651	0.149	0
XOM	0.669	0.005	0.764	0.962	0.189	0.696	0.042	0.696	2
Resultado	8	6	5	7	7	8	7	8	

4.3.1. Causalidad inversa

Al igual que se hizo en la segunda sección de este capítulo. Es necesario probar que no existe causalidad inversa, es decir que los rendimientos accionarios afectar el número de búsquedas.

Al igual que la segunda parte se plantean los modelos a correr:

$$\delta B_{m,t} = \alpha + \sum_r^H \phi_r \delta B_{m,t-r} + \sum_r^H \beta_r \Delta P_{C_{i,t-r}} + \varepsilon$$

En esta primera ecuación la variable dependiente es la diferencia en el índice de Google Trends ajustado, y la independiente es el rendimiento accionario diario. Al igual que en las otras secciones se plantea el caso del índice ajustado de Google Trends diario.

$$B_{m,t} = \alpha + \sum_r^H \phi_r B_{m,t-r} + \sum_r^H \beta_r \Delta P_{C_{i,t-r}} + \varepsilon$$

En la Tabla 7 se encuentran los resultados de las dos ecuaciones anteriores. La Tabla 7 tiene más elementos significativos (62) comparados a la Tabla 6 (57). Relativamente, los tipos de búsqueda 1 y 4 son inferiores en resultados significativos de las búsquedas 2 y 3; este resultado difiere de los resultados de la Tabla 6. Las acciones más susceptibles a la estimación son: Chevron (6), General Electric (7), Goldman Sachs (6). En este caso Goldman Sachs y General Electric, también resultaron significativas en la Tabla 5. Los resultados indican que existen Granger causalidad en ambos sentidos. En la mayor parte de las acciones existe al menos un tipo de búsqueda que sea Granger causal. Al igual que en la sección 4.2 la tabla 5 provee evidencia de que existe causalidad inversa.

Tabla 7: Prueba de causalidad de Granger - causalidad inversa variación diaria

Acción	B1-Nivel	B1-Diferencia	B2-Nivel	B2-Diferencia	B3-Nivel	B3-Diferencia	B4-Nivel	B4-Diferencia	Resultado
AAPL	0.711	0.803	0.580	0.054	0.668	0.071	0.879	0.053	3
AXP	0.093	0.310	0.803	0.827	0.656	0.882	1.000	1.000	1
BA	0.621	0.109	0.100	0.102	0.363	0.245	0.999	0.862	1
CAT	0.605	0.769	0.027	0.098	0.062	0.015	0.846	0.904	4
CSCO	0.627	0.647	0.850	0.251	0.938	0.519	0.517	0.483	0
CVX	0.000	0.010	0.000	0.004	0.000	0.001	0.433	0.473	6
DD	0.837	0.780	0.329	0.272	0.955	0.611	0.534	0.389	0
DIS	0.849	0.922	0.170	0.282	0.319	0.293	0.910	0.770	0
GE	0.002	0.009	0.000	0.000	0.000	0.000	0.019	0.583	7
GS	0.015	0.023	0.000	0.000	0.000	0.000	0.237	0.436	6
HD	0.478	0.332	0.874	0.708	0.856	0.847	0.572	0.048	1
IBM	0.992	0.689	0.302	0.686	0.572	0.447	0.171	0.002	1
INTC	0.023	0.689	0.000	0.098	0.007	0.408	0.343	0.512	4
JNJ	0.716	0.762	0.028	0.014	0.034	0.014	0.196	0.500	4
JPM	0.544	0.658	0.032	0.208	0.000	0.002	0.689	0.186	3
KO	0.975	0.908	0.489	0.681	0.314	0.075	0.935	0.785	1
MCD	0.845	0.152	0.006	0.252	0.001	0.006	0.000	0.001	5
MMM	0.861	0.536	0.995	0.574	0.991	0.509	1.000	1.000	0
MRK	0.812	0.846	0.176	0.629	0.880	0.926	1.000	1.000	0
MSFT	0.637	0.340	0.973	0.031	0.383	0.030	0.182	0.281	2
NKE	0.518	0.941	0.879	0.145	0.866	0.180	0.550	0.191	0
PFE	0.401	0.180	0.952	0.996	0.174	0.227	0.806	0.002	1
PG	0.767	0.803	0.684	0.615	0.081	0.558	0.876	0.489	1
TRV	0.859	0.945	0.012	0.005	1.000	1.000	1.000	1.000	2
UNH	0.319	0.412	0.444	0.664	0.938	0.539	0.930	0.517	0
UTX	0.394	0.984	0.389	0.518	0.440	0.715	0.007	0.025	2
V	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0
VZ	0.310	0.150	0.180	0.060	0.609	0.357	0.825	0.191	1
WMT	0.434	0.367	0.952	0.096	0.877	0.257	0.528	0.651	1
XOM	0.443	0.132	0.001	0.003	0.068	0.023	0.110	0.042	5
Resultado	5	3	11	12	10	11	3	7	-

Capítulo 5

Metodología

En la motivación se estableció el interés de encontrar si estas series de búsquedas por internet podrían ser incorporadas a un modelo. La metodología consiste en comparar dos regresiones distintas y evaluarlas mediante indicadores de desempeño. El objetivo de este capítulo será demostrar que existe una mejora en la estimación intra-muestra. Este tipo de metodología es similar a la Risteski y Davcev (2014), en la que usa Google Trends para pronosticar la volatilidad del CAC40¹. El modelo planteado busca encontrar el mejor número de rezagos que minimicen el criterio de Akaike. Comparando una serie únicamente con su pasado, y otro incluyendo las series de Google Trends ajustado.

5.1. Modelo en diferencias

Como bien se señaló en el capítulo 3.1 se buscará un modelo con diferencias. El modelo a contrastar es el siguiente:

$$\delta P_{i,t} = c + \sum_j^T \phi_j \Delta P_{i,t-j} + \varepsilon$$

¹Cotation Assistée en Continu es un índice que contiene 40 empresas listadas en la bolsa de París. Es un homólogo de DJI para Francia

Donde $\delta P_{i,t}$ es la diferencia en los cierres de precios, ϕ es el parámetro asociado al pasado del los cambios en precios y se elegirá T a partir del modelo que minimice el criterio de Akaike ².

Para el caso de diferencias se correrá la siguiente regresión:

$$\delta P_{i,t} = c + \sum_j^T \phi_j \Delta P_{i,t-j} + \sum_m^4 \sum_r^H \beta_{m,t-r} \delta B_{m,t-r} + \sum_m^4 \sum_r^H \gamma_{m,t-r} B_{m,t-r} + \varepsilon$$

El modelo a probar, incluye $\delta B_{m,t-r}$ que hacen referencia a la diferencia del índice de Google Trends ajustado, $B_{m,t-r}$ es el nivel de Google Trends ajustado, m representa los cuatro tipos de búsquedas³ y H al igual que T se escogerá por el modelo que minimice el criterio de Akaike. La Tabla 7 sintetiza los resultados de las regresiones. En primera instancia las acciones: CVX, CSCO, JPM y NKE los modelos no convergieron y por lo tanto estarán fuera de este análisis. La columna de resultado % reporta en porcentaje. El promedio de los resultados para las 25 acciones es de una mejora de 16 puntos base, esto a pesar de que en 14 ocasiones el modelo AR fuera superior al modelo retador. Solamente Pfizer [PFE] y Merck & Co [MRK]. tienen valores de mejores superiores al 1 %, curiosamente, también podrían ser considerado como datos anómalos si se observa el Gráfico 9. El último señala la relación entre los resultados de cada acción en la prueba de causalidad de Granger y los resultados de la comparativa entre los modelos. Lo que el Gráfico 9 muestra es si selecciona a las acciones que aprobaron en la causalidad de Granger entonces posiblemente mejoren el modelo.

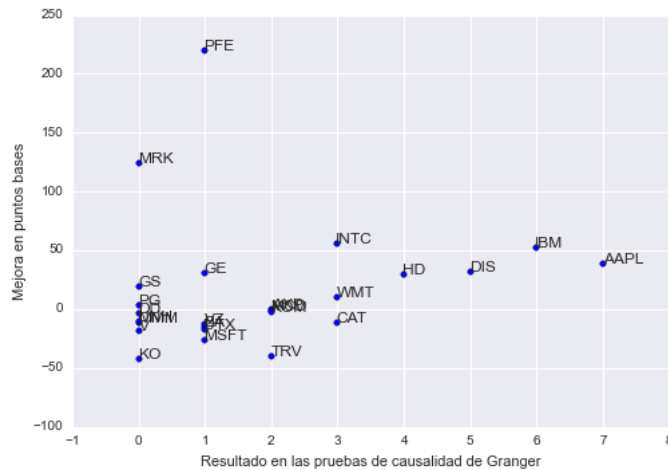
²Se eligió que el algoritmo iterará hasta 7

³Excepto en los casos que no exista serie. En esos bajara el número i. Esto sucede para los casos de: AXP, MMM, MRK, PFE, UTX y TRV

Tabla 8 : Resultados modelo en diferencias diarias

Acción	Mejor AR	Mejor Modelo	Resultado %	Resultados P.C.G.
AAPL	5609.408080	5588.001554	0.381618	7
AXP	5397.896178	5398.238915	-0.006349	2
BA	6540.157505	6549.779505	-0.147122	1
CAT	6836.233677	6843.931619	-0.112605	3
DD	4258.265093	4260.030159	-0.041450	0
DIS	4153.699645	4140.736410	0.312089	5
GE	1997.285257	1991.121922	0.308586	1
GS	9890.074125	9871.405104	0.188765	0
HD	4295.234952	4282.346473	0.300065	4
IBM	8037.032838	7994.663098	0.527181	6
INTC	2033.482877	2022.087623	0.560381	3
KO	1656.189181	1663.143345	-0.419889	0
MCD	7608.709405	7609.407403	-0.009174	2
MMM	6164.337495	6171.315559	-0.113201	0
MRK	3684.181092	3638.301600	1.245310	0
MSFT	1659.124441	1663.599199	-0.269706	1
PFE	448.444995	438.571572	2.201702	1
PG	4331.906253	4330.313040	0.036779	0
TRV	2733.681798	2744.716981	-0.403675	2
UNH	5137.844004	5143.282616	-0.105854	0
UTX	3448.371098	3454.499145	-0.177708	1
V	2586.585457	2591.506869	-0.190267	0
VZ	2756.507746	2760.067875	-0.129154	1
WMT	4142.642527	4138.624760	0.096986	3
XOM	6135.959934	6137.774891	-0.029579	2

Gráfico 9: Relación entre resultados del modelo diferencias y PCG



5.2. Modelo en variaciones

Como bien se señalo en el capítulo 3.1 se buscará un modelo con variaciones. El caso del modelo AR en variaciones es el siguiente:

$$\Delta P_{i,t} = c + \sum_j^T \phi_i \Delta P_{i,t-j} + \varepsilon$$

La única diferencia es que la variable independiente es una variación diaria⁴ relativo al modelo de la sección anterior. El modelo retardador para el caso de la variación es:

$$\Delta P_{i,t} = c + \sum_j^T \phi_i \Delta P_{i,t-j} + \sum_m^4 \sum_r^H \beta_{m,t-r} \delta B_{m,t-r} + \sum_i^4 \sum_r^H \gamma_{i,t-r} B_{i,t-r} + \varepsilon$$

Este modelo, también, tienen las mismas especificaciones que tuvo el modelo de la sección anterior. Los resultados del modelo se encuentran en la Tabla 9. En este tipo de especificación las mejores son menores a las señaladas por la Tabla 8. El promedio de los resultados da un mejora de sólo 3 puntos bases y 14 de los 25 no pudieron superar el mejor modelo AR. El Gráfico 10 señala una relación similar a la del gráfico 9, si los componentes fueron positivos en la prueba de causalidad Granger entonces es probable que sí mejoren el modelo AR.

⁴Esto se menciona en el capítulo 3.1

Tabla 9: Resultados del modelo en variaciones diarias

Acción	Mejor AR	Mejor Modelo	Mejora %	Resultados P.C.G.
AAPL	-9209.322732	-9222.611082	0.144292	4
AXP	-8226.389515	-8225.837060	-0.006716	2
BA	-9508.849260	-9502.385113	-0.067980	0
CAT	-8997.518632	-8985.990338	-0.128127	3
DD	-9532.663429	-9525.698349	-0.073065	0
DIS	-9653.790791	-9651.069187	-0.028192	0
GE	-9065.886463	-9081.818928	0.175741	5
GS	-8224.228248	-8241.699965	0.212442	4
HD	-9735.296313	-9744.739052	0.096995	4
IBM	-10614.677538	-10641.522736	0.252906	7
INTC	-9508.356409	-9519.008222	0.112026	2
KO	-11141.049623	-11131.564154	-0.085140	0
MCD	-24615.703656	-24626.644724	0.044448	3
MMM	-10510.812410	-10505.574048	-0.049838	1
MRK	-9930.557718	-9978.376132	0.481528	0
MSFT	-6363.622293	-6357.981336	-0.088644	1
PFE	-9479.097141	-9487.722987	0.090999	1
PG	-11429.690919	-11431.192704	0.013139	0
TRV	-6694.457688	-6688.139693	-0.094377	2
UNH	-8736.909057	-8734.259403	-0.030327	0
UTX	-6694.457688	-6688.139693	-0.094377	1
V	-8522.278260	-8512.958389	-0.109359	0
VZ	-10587.797633	-10587.507992	-0.002736	1
WMT	-11290.171118	-11290.496722	0.002884	0
XOM	-10245.797291	-10239.818751	-0.058351	2

Gráfico 10: Relación entre resultados del modelo variaciones y PCG



5.3. Análisis de resultados

El segundo objetivo de la motivación no se cumplió dado el plantamiento metodológico. Esto debido principalmente a una cuestión de y quizá de planteamiento del modelo. Los resultados de la metodología no sugieren probar nuestra motivación de que Google Trends pueda ser incorporado a modelos para definir rendimientos accionarios.

A pesar de que lo general no se cumpla, existen acciones que en ambas ocasiones mejoraron al utilizar la estimación. Las acciones que mejoraron son: APPL, IBM, PFE, MRK, INTC, GE y HD. Para futuros avances se recomienda trabajar con acciones que han pasado por las pruebas de causalidad de Granger. Además, ante el número reducido de acción candidatas para este tipo de tratamiento se recomienda utilizar un enfoque individualizado, donde la selección de variables y rezagos sea preciso dependiendo de cada acción. A partir de los resultados; también, se obtiene, que en general es más eficiente usar diferencias en precios de cierre en comparación a variaciones diarias. La selección de variable independiente, también es importante, lo que muestran las pruebas de Granger es que no basta con únicamente poner el *ticker* y la serie en bruto, sino falta modificar las series para obtener resultados.

Capítulo 6

Conclusiones

En la revisión de literatura se recapitulan avances relevantes en la aplicación de la tecnologías de la información relacionadas con el sistema financiero. La literatura reciente indica que las aplicaciones derivadas de Google Trends son variadas y que tienen especial utilidad en cuestiones económicas y financieras. Se destaca el uso de Google Trends como una herramienta útil para pronosticar la volatilidad de índices accionarios.

Una contribución importante del trabajo, es el análisis referente a Google Trends como variable explicativa. Esto es relevante ya que la información no esta disponible de manera organizada. Los índices proporcionados por Google Trends , deben ser ajustados si se requieren utilizar diariamente. Este documento contribuye con un pequeño método para ajustar las series diarias.

Como primer objetivo se planteo verificar si las búsquedas de internet tenían un impacto en el cambio del precio de las acciones. De acuerdo a las pruebas de causalidad de Granger se encuentra evidencia de que lo primero sucede, en una situación moderada de veces. No todas los precios de acciones son susceptibles a las búsquedas. También, se encuentra que lo contrario sucede, el cambio en precio de las acciones del DJI afecta su número de búsquedas. Lo último es un poco más significativo que la primer objetivo; sin embargo, la utilidad está en saber en que casos particulares sucede.

Por último se realiza un modelo ARIMAX en diferencias y variaciones sobre los precios de las 30 acciones del DJI. Esto para probar si Google Trends debería ser utilizado para modelos predictivos. Dado el plantamiento, no se encuentra una respuesta definitiva en lo general. Sin embargo, encontramos acciones donde existe persistencia en la mejora del modelo. Esto sugiere que un enfoque individualizado podría funcionar. Esta mejora está vinculada con los resultados de la prueba de la causalidad de Granger. Concluimos que mediante un enfoque individualizado existe la posibilidad de mejorar las predicciones intramuestra.

Capítulo 7

Anexo Estadístico

7.1. Pruebas de Raíz Unitaria

Las pruebas de Raíz unitaria son indispensables para saber que metodología de series de tiempo hay que utilizar. En este anexo se proveen información de las pruebas de raíz unitaria por cada acción y cada tipo de búsqueda. En este anexo se reportan las pruebas de raíz unitaria, se reportan dos tablas. Una para el precio de las acciones y otra para las búsquedas. Se utilizaron una prueba de aumentada de Dickey-Fuller y se reporta el valor-p. El fin de este anexo es sólo para cuestión informativa, no se plantea hacer análisis en la sección.

Tabla Anexa 1: Pruebas de Dickey Fueller Aumentada sobre acciones

Acción	Nivel	Diferencia	Variación	Acción	Nivel	Diferencia	Variación
AAPL	0.983391	0	0	KO	0.862235	0	0
AXP	0.530915	0	0	MCD	0.993619	0	0
BA	0.764482	0	0	MMM	0.960468	0	0
CAT	0.442176	0	0	MRK	0.194134	0	0
CSCO	0.000002	0	0	MSFT	0.609813	0	0
CVX	0.580110	0	0	NKE	0.998653	0	0
DD	0.246834	0	0	PFE	0.336207	0	0
DIS	0.986142	0	0	PG	0.567987	0	0
GE	0.130544	0	0	TRV	0.976123	0	0
GS	0.280296	0	0	UNH	0.995193	0	0
HD	0.996953	0	0	UTX	0.686107	0	0
IBM	0.577737	0	0	V	0.990599	0	0
INTC	0.000062	0	0	VZ	0.113528	0	0
JNJ	0.863306	0	0	WMT	0.279485	0	0
JPM	0.096000	0	0	XOM	0.518132	0	0

Se reportan los valores-P

Tabla Anexa 2: Prueba de Dickey Fueller aumentada para índice de Google Trends

ajustado

Acción	B1-Nivel	B1- Diferencia	B2- Nivel	B2-Diferencia	B3- Nivel	B3-Diferencia	B4-Nivel	B4-Diferencia
AAPL	0.000000e+00	0	6.621557e-26	0	3.45994e-26	0	0.000000e+00	0.000000e+00
AXP	2.063809e-28	0	2.763747e-30	0	0	0	0.000000e+00	0.000000e+00
BA	3.130363e-06	0	3.901368e-28	0	3.63603e-28	0	0.000000e+00	0.000000e+00
CAT	2.715336e-13	0	2.757184e-29	0	2.71713e-29	0	0.000000e+00	0.000000e+00
CSCO	7.285627e-29	0	2.664071e-29	0	1.38058e-29	0	2.155674e-30	0.000000e+00
CVX	2.268535e-15	0	5.466730e-12	0	1.93153e-12	0	8.356543e-24	0.000000e+00
DD	2.040592e-30	0	0.000000e+00	0	0	0	0.000000e+00	0.000000e+00
DIS	3.393337e-18	0	2.777312e-20	0	2.10868e-19	0	0.000000e+00	0.000000e+00
GE	3.051006e-18	0	6.987056e-18	0	4.37446e-18	0	0.000000e+00	0.000000e+00
GS	1.356541e-23	0	3.164896e-27	0	5.36196e-28	0	0.000000e+00	0.000000e+00
HD	2.006484e-05	0	1.772136e-29	0	3.42038e-29	0	2.626920e-29	0.000000e+00
IBM	1.508607e-09	0	0.000000e+00	0	0	0	0.000000e+00	0.000000e+00
INTC	2.374156e-19	0	1.250208e-28	0	7.16118e-29	0	0.000000e+00	0.000000e+00
JNJ	0.000000e+00	0	1.024718e-24	0	5.9607e-25	0	0.000000e+00	0.000000e+00
JPM	9.767615e-25	0	2.751632e-29	0	1.38196e-29	0	4.671216e-30	0.000000e+00
KO	7.785270e-07	0	3.354471e-30	0	4.19353e-30	0	0.000000e+00	0.000000e+00
MCD	3.001972e-29	0	0.000000e+00	0	0	0	0.000000e+00	0.000000e+00
MMM	4.383785e-06	0	1.212010e-26	0	1.93449e-24	0	X	X
MRK	3.247817e-29	0	0.000000e+00	0	0	0	X	X
MSFT	2.818825e-21	0	7.191985e-29	0	8.77616e-18	0	0.000000e+00	0.000000e+00
NKE	2.187872e-30	0	9.424030e-08	0	6.5607e-09	0	4.574356e-18	2.808538e-29
PFE	8.227218e-27	0	8.153937e-29	0	7.12392e-28	0	0.000000e+00	0.000000e+00
PG	1.808536e-19	0	7.101691e-29	0	2.02178e-29	0	0.000000e+00	0.000000e+00
TRV	4.852850e-19	0	0.000000e+00	0	X	X	X	X
UNH	1.797350e-26	0	3.202272e-30	0	2.28324e-30	0	0.000000e+00	0.000000e+00
UTX	0.000000e+00	0	0.000000e+00	0	0	0	0.000000e+00	0.000000e+00
V	1.436025e-05	0	4.409040e-15	0	1.88789e-14	0	1.154846e-24	0.000000e+00
VZ	7.558244e-16	0	1.691782e-28	0	9.33256e-28	0	0.000000e+00	0.000000e+00
WMT	2.324204e-29	0	5.358683e-27	0	3.15576e-27	0	0.000000e+00	0.000000e+00
XOM	0.000000e+00	0	1.026795e-24	0	7.69886e-25	0	0.000000e+00	0.000000e+00

Se reportan los valores-P en la tabla.

En la Tabla sólo hay valores cercanos a ceros. Los valores X indican que no hay serie (vease la Tabla 2)

Referencias

- Alanyali, M., Moat, H. S., y Preis, T. (2013). “Quantifying the relationship between financial news and the stock market.” *Scientific reports*, 3.
- Bodie, Z., Kane, A., y Marcus, A. (2008). *Investments (7 uppl.)*. Boston: McGraw-Hill.
- Bollen, J., Mao, H., y Zeng, X. (2011). “Twitter mood predicts the stock market.” *Journal of Computational Science*, 2(1), 1–8.
- Choi, H., y Varian, H. (2012). “Predicting the present with google trends.” *Economic Record*, 88(s1), 2–9.
- Granger, C. W. (1969). “Investigating causal relations by econometric models and cross-spectral methods.” *Econometrica: Journal of the Econometric Society*, 424–438.
- Hunter, J. D. (2007). “Matplotlib: A 2d graphics environment.” *Computing In Science & Engineering*, 9(3), 90–95.
- Ince, H., y Trafalis, T. B. (2008). “Short term forecasting with support vector machines and application to stock price prediction.” *International Journal of General Systems*, 37(6), 677–687.
- Johansson, E. (2014, 12). *Creating daily search volume data from weekly and daily data*.
- Kristoufek, L. (2013a). “Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era.” *Scientific reports*, 3.
- Kristoufek, L. (2013b). “Can google trends search queries contribute to risk diversification?” *Scientific reports*, 3.
- McLaren, N., y Shanbhogue, R. (2011). “Using internet search data as economic indicators.”

Bank of England Quarterly Bulletin(2011), Q2.

Risteski, D., y Davcev, D. (2014). “Can we use daily internet search query data to improve predicting power of egarch models for financial time series volatility?”

Souza, T. T. P., Kolchyna, O., Treleaven, P. C., y Aste, T. (2015). “Twitter sentiment analysis applied to finance: A case study in the retail industry.” *arXiv preprint arXiv:1507.00784*.

Zheludev, I., Smith, R., y Aste, T. (2014). “When can social media lead financial markets?” *Scientific reports*, 4.