

NLP and machine learning applied to forward guidance and market sentiment analysis

Paula Luvini, Facundo Marconi, Florencia Ludueña

Noviembre 2020

Abstracto

El papel de los discursos y declaraciones de los hacedores de política económica juega un rol fundamental en la formación de expectativas de los agentes, afectando la toma de decisiones de los mismos. En el caso de la política monetaria, además de los medios tradicionales, se utilizan herramientas no convencionales como las exposiciones de los presidentes de los Bancos Centrales para anclar e influenciar a los mercados hacia una dirección deseada. Esto es conocido en la literatura económica, como *forward guidance*.

En este trabajo, analizaremos los discursos de los chairman de la Reserva Federal y el contenido de las minutas del Federal Open Market Committee (FOMC) del período 2001-2019 y su impacto en los mercados financieros. Nuestra hipótesis parte de que los discursos son preparados para que los participantes de mercado puedan extraer complejas señales de los mismos. Específicamente, la curva de rendimiento y el comportamiento activos financieros como letras de Tesoro (guía de otros activos) dependen del contenido específico del discurso: outlook, certidumbre, etc.

Para ello utilizaremos diferentes metodologías: *web scrapping* para extraer el corpus del discurso, NLP para limpiar el texto, realizar un market sentiment analysis y vectorizar los discursos. Luego, una vez procesado utilizaremos como clasificador SVM para predecir qué cambio hubo en la volatilidad de los activos.

1 Introducción

En este trabajo buscamos realizar un análisis íntegro de los discursos de los Presidentes, Vicepresidentes y Gobernadores de la Reserva Federal de Estados Unidos (*FED*) y del Federal Open Market Committee (FOMC). Con el objetivo de fijar expectativas e influir en la coyuntura económica, a la cual también responden, las conferencias y declaraciones que brindan los *policymakers* pueden generar mucho impacto en diversos indicadores económicos y servir como señalización para los agentes si existe credibilidad. En palabras del ex-presidente de la FED Ben Bernanke: “La política monetaria es 98% de charla y 2% de acción”¹.

En línea con el objetivo del curso estudiamos diferentes algoritmos y técnicas que permiten abordar y estructurar información detallada del lenguaje, y que desarrollaremos en detalle en las siguientes secciones. El objetivo principal del trabajo es testear la hipótesis de que existe una relación entre la coyuntura económica, los discursos y los activos financieros al que los hacedores de política quieren influir.

2 Obtención y Pre-Procesamiento de los datos

En primer lugar, obtuvimos el corpus discursivo vía *webscrapping*², dado que los speeches están disponibles en la página de la reserva federal³ ⁴. Si bien los discursos son de libre disponibilidad, la forma en que se publicaron a lo largo del tiempo fue mutando, motivo por el cual el proceso

¹Ben Bernanke’s Blog, Brookings ([hyperlink](#))

²Nos basamos en <https://www.kaggle.com/natanm/federal-reserve-governors-speeches-1996-2020>

³<https://www.federalreserve.gov/newsevents/speeches.htm>

⁴<https://www.federalreserve.gov/boarddocs/speeches/>

involucró mezclar 3 scrapping diferentes: Cambia el link entre antes y después de 2006, y luego en 2010, cambia la estructura *html* de la página, de una donde todos títulos, links, orador y lugar se encuentran por separado en listas, a una lista que ya contiene todo junto. En particular, para los speakers fue necesario homogeneizar los nombres de los speeches ya que se escribían variantes del mismo.

Por otro lado, también recopilamos la información de las minutas expedidas por el Federal Open Market Committee (FOMC). El comité se reúne un total de ocho veces en el año para de tomar las decisiones de cambio de tasa de interés. Al finalizar estas reuniones, el comité emite un comunicado mejor conocido como minuta en el cual detalla las decisiones tomadas y los motivos de las mismas, en general haciendo hincapié en datos de empleo y de precios. Las minutas están también disponibles en la página de la reserva federal⁵ en la sección de Política Monetaria y bajo el título de *Federal Reserve issues FOMC statement*. Al igual que en el caso de los discursos, los links del scraping también cambian dependiendo del año así como el formato de los mismos. En el último caso, lo que cambia en algunos textos son la frase o palabras claves que se utilizan en el *re.search* para determinar el inicio y el final de la minuta.

El período contempla los años 2001-2019, período lo suficientemente extenso como para incluir diferentes partes del ciclo económico (de estabilidad y crisis).⁶

Antes de comenzar el análisis, realizamos un preprocesamiento en los textos para remover puntuación, y homogeneizar todo a minúsculas. Luego, aplicamos una tokenización, una lematización y eliminamos los stopwords.

Por otro lado, como detallaremos más adelante, utilizamos un índice de tasas del tesoro para 2, 5, 10 y 30 años de maturity para analizar la estructura de tasas que son guía de otros activos; y un índice de volatilidad basado en opciones denominado VIX para medir el stress del mercado. Ambos fueron obtenidos de Bloomberg.

3 Exploración de los datos

Contamos con un total de 1185 discursos en el período seleccionado que fueron brindados por un total de 27 oradores. Si bien en estos años la Reserva Federal contó con cuatro presidentes, debemos tener en cuenta que los discursos incluyen las declaraciones de vicepresidentes y gobernadores de la FED. En términos de cantidad de discursos brindados por cada funcionario Ben Bernanke como chairman en función entre el año 2006 al 2014, es el que más tiene. Alan Greenspan quien fue presidente en todos los años anteriores es el segundo en tener más discursos. Los discursos tienen lugar en diferentes locaciones. Si bien la mayoría ocurren en Washington D.C., sede de la FED, o en Nueva York, sede del Banco de la Reserva Federal de Nueva York que nuclea a muchos bancos importantes del país, también hay discursos brindados en el extranjero. Los discursos tienen en promedio una longitud de 3245 palabras, encontrándose la mayoría en el rango de entre 2000 y 4500 palabras y llegando a haber casos con discursos más de 10000.

Por otro lado, en las minutas la información es algo distinta. En primer lugar, la temática de las mismas es distinta a la de los discursos, estando acotada a exponer las decisiones del comité para lograr los objetivos de la Reserva Federal de maximizar el empleo y mantener la estabilidad de precios. Asimismo, están restringidas geográfica y temporalmente. Las mismas se expiden desde la sede de la FED en Washington y con una periodicidad establecida que depende de la cantidad de veces que se reúne el comité, es decir, ocho veces al año. En total contamos con 152 minutas.

Las minutas son mucho más cortas que los discursos, nunca superando las 1000 palabras. Son declaraciones cortas y concretas que fueron tomando una estructura invariante después de Bernanke⁷. Esto queda en evidencia en la figura 1, al analizar *wordcloud* de las minutas de cada mandato. Mientras durante el mandato de Greenspan el discurso es más desordenado,

⁵<https://www.federalreserve.gov/newsevents/speeches.htm>

⁶Se eliminó el año 2020 (ciclo lectivo del presente trabajo) dada la situación atípica de un fenómeno cuyos efectos todavía no se terminan de dimensionar.

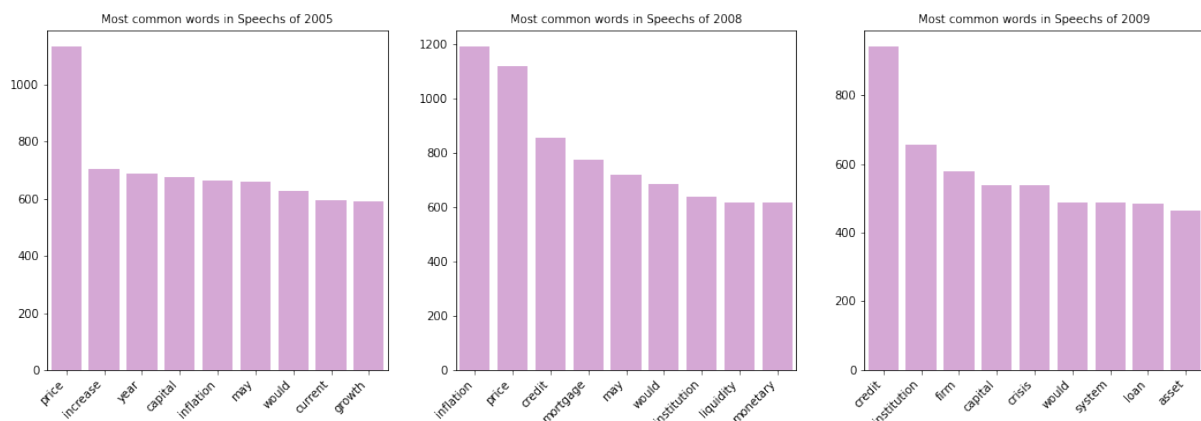
⁷Pese a que siempre se prestó atención a la redacción de los comunicados, Bernanke fue uno de los impulsores de su implementación directa como parte de herramientas no convencionales para la política monetaria.

luego de Bernanke se acentúa y explicita verbalmente el objetivo de la institución con frecuencia e importancia.

Figure 1: WordCloud Minutas FOMC



Por su parte, los discursos tienen una temática y formas más libres. En algunos casos, son dados en universidades, conferencias o reuniones y pueden tratar de temas distintos a los de política económica, como educación o innovación. También tienen la libertad de ocurrir en cualquier día de la semana, teniendo casos en los que los discursos ocurren los fines de semana. De todas formas los discursos son relevantes porque son utilizados por los hacedores de política económica para enfatizar su *forward guidance* y poder enviar señales claras a los mercados con una flexibilidad mayor a la que se tiene en las minutas. Pero debemos tener en cuenta que pueden traer más “ruido” a ciertos análisis, tema en el que nos explayaremos en siguientes secciones. Se observa que al analizar a los “speeches” agrupados por año en distintos momentos del tiempo, si bien hay palabras comunes que se repiten, se ven reflejadas particularidades del ciclo. En periodos de más estabilidad como el año 2005, aparece la palabra *growth* (crecimiento), mientras que en 2009 tras desencadenarse la crisis aparece propiamente la palabra *crisis* junto con *system*, *liquidity* asociadas a las inyecciones de liquidez para prevenir la crisis sistémica y *loan*, *firm* por el salvataje a las firmas privadas. En 2008 finalmente se ve la palabra *morgage* y *credit* asociado a un factor desencadenante de la crisis por el elevado apalancamiento y derivados sobre hipotecas presente en hojas de balances de bancos y fondos.



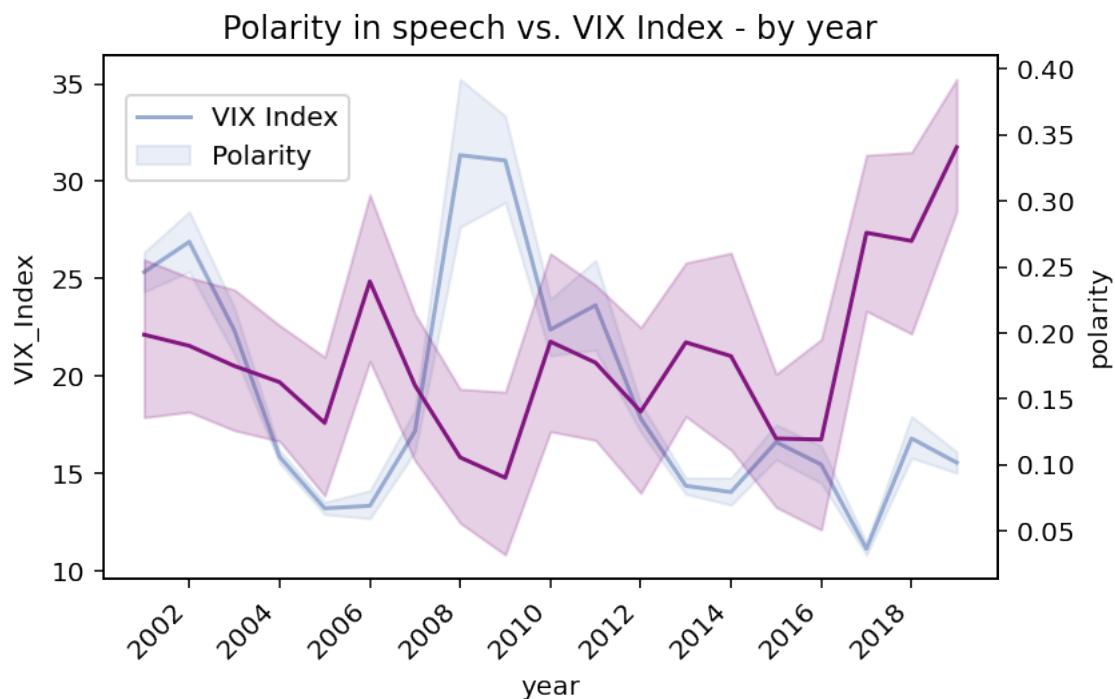
4 Sentiment Analysis

Para realizar este análisis utilizamos el paquete *polyglot* ⁸ con su función *polarity*. La misma permite tomar un texto y calificar las palabras del mismo como negativas, positivas o neutras mediante un scoring de -1, 1 y 0, respectivamente. Luego de haber clasificado cada palabra saca el *scoring* total del texto mediante un promedio simple, sin tener en cuenta palabras neutras. De esta manera, obtenemos para todos los discursos y minutas un scoring entre -1 y 1 denotando la negatividad y positividad total del mismo. Observamos que la mayoría de los mismos tienen un puntaje mayor a 0, con sólo un 15% del total de minutas teniendo scoring negativos, mientras

⁸<https://polyglot.readthedocs.io/en/latest/>

que en los discursos el porcentaje es de 20%.

Con esta herramienta podemos evaluar el tono de los discursos a lo largo de los años estudiados y ver si se corresponde con los sucesos económicos que se sucedieron en los mismos. Observamos que el score de polarity de 2004 comienza a aumentar en línea con una suba de tasas dado la performance económica optimista ⁹, mientras se mantiene un bajo nivel de VIX Index. En 2007 se interrumpe la suba de tasas dado ciertos signos de debilidad en la performance económica. Es en ese momento donde el discurso comienza a tomar un tono más negativo, las tasas bajan y se percibe mayor volatilidad llegando a un pico en 2008-2009. Pese a la mejora posterior, las tasas se mantuvieron a un bajo nivel por varios años, para luego volver a subir junto al scoring en 2016.



5 Entrenamiento y Clasificación

Por otro lado, siguiendo el análisis de *forward guidance* nos pareció oportuno poder medir si hay un verdadero impacto por parte de las declaraciones de los funcionarios de la Fed en algunos indicadores macroeconómicos. Es decir, hasta ahora hemos analizado y explorado los datos de una manera más descriptiva y con el conocimiento previo que teníamos de los sucesos económicos de estos años.

Siguiendo la línea del paper de Zadeh y Zollmann [2009] vamos a plantear primero el impacto que tienen las declaraciones de política monetaria en un indicador de volatilidad. Para ello tomamos en cuenta el índice de volatilidad conocido como *VIX*, datos que extraemos de la plataforma Bloomberg. Con esta información vamos a entrenar un clasificador de SVM que intentará predecir si la volatilidad aumenta o disminuye tras las declaraciones de la Fed.

De esta manera, vamos a plantear varios escenarios temporales en los cuales evaluaremos qué tan bien podemos predecir el comportamiento de la volatilidad. Comenzando con los cambios de 5 días después del discurso hasta llegar a los 100 días, vamos a codificar a la variable a predecir como 0 si la volatilidad aumentó en ese período y 1 si la misma bajó. Para el entrenamiento del texto vamos a utilizar como herramienta de similaridad semántica al *tf-idf*. Antes de comenzar dividiremos a la muestra en entrenamiento (80% de la misma) y clasificación (20%). La manera de evaluar la eficacia de nuestro algoritmo será en primer lugar cross-validando y mirando la accuracy resultante. Asimismo, se llevará a cabo un Grid Search para encontrar los parámetros óptimos a utilizar.

⁹<https://datosmacro.expansion.com/tipo-interes/usa>

En un principio, llevamos a cabo este análisis para los discursos de los miembros de la Fed. Lo primero que notamos es que, por la longitud y los diversos temas de los mismos, los discursos no eran una buena opción para predecir un aumento o una disminución de la volatilidad. Si bien nos dieron un análisis muy interesante en los temas repasados anteriormente, a la hora de clasificar es mejor tomar en cuenta a las minutas, que son los resultados que presentaremos a continuación.

Los resultados que encontramos están en parte alineados con los de Zadeh y Zollmann [2009]. En el caso de la volatilidad 5 o 20 días después de la reunión, los resultados no son buenos teniendo una accuracy que es peor que clasificar al azar. Recién cuando intentamos predecir períodos de tiempo más largos, como 50 o 100 días luego de la reunión esto mejora un poco.

Al mismo tiempo, nos pareció interesante ver cuánto y en qué tiempos impactaban las minutas en la determinación de las tasas de interés de la Reserva Federal. Dado que el objetivo de las reuniones de la FOMC son determinar los cambios en la tasa de interés de referencia, es bastante claro que va a existir una relación entre los comunicados que expidan y las tasas. De todas formas, el análisis es aún interesante para evaluar sobre todo los lags de la política monetaria entre que ciertas medidas son efectuadas y que se logra cambiar la variable de interés.

En el caso de las tasas de interés se utiliza la información diaria de la “Effective Federal Funds Rate” (EFFR), disponible libremente en la página de la Reserva Federal de St. Louis¹⁰. La EFFR es la tasa a la cual las instituciones bancarias se financian entre sí y está en última instancia determinada por el mercado de fondos bancarios, pero la influencia de la Fed en la misma es importante.

Los resultados en la predicción de las tasas de interés también son malos 5 días después de la reunión pero, a diferencia de en la volatilidad, estos mejoran gradualmente a medida que aumentamos el tiempo. Es decir, que podemos predecir mejor cómo se comportarán las tasas de interés cuando tomamos períodos más largos de tiempo. A 50 días, tanto en la muestra de testeo como en la cross-validación el accuracy mejora bastante y a 100 días podemos considerar que es muy bueno.

El comportamiento que vemos en las tasas de interés está alineado con lo que conocemos como *lags* de la política monetaria y que no es otra cosa que el tiempo en el que tarda en reflejarse la decisión del hacedor de política en las variables macroeconómicas. Como mencionamos anteriormente, si bien la FED modifica en sus reuniones la tasa de referencia de política monetaria, la tasa efectiva de transacciones interbancarias está determinada por los participantes de ese mercado y la duración de contratos que puede enlentecer la transmisión de tasa. De esta manera, podemos concluir que cuando el FOMC emite comunicados tomando la decisión de bajar o subir la tasa de interés deberemos esperar más de 50 días para poder ver reflejado el cambio con nuestro clasificador.

A continuación ponemos el resultado de Accuracy que tuvo el clasificador sobre la muestra separada como test en cada una de las variables y horizontes temporales, para el caso de las minutas.

Clasificador	Accuracy
Volatilidad 5 días	0.48
Volatilidad 20 días	0.42
Volatilidad 50 días	0.68
Volatilidad 100 días	0.77
Fed Rate 5 días	0.42
Fed Rate 20 días	0.55
Fed Rate 50 días	0.77
Fed Rate 100 días	0.84

¹⁰<https://fred.stlouisfed.org/categories/118>

6 Conclusiones

Como reflexiones finales podemos decir que realizamos un análisis bastante exhaustivo e integral de las declaraciones públicas de la Reserva Federal de los Estados Unidos. Creemos que fue integral en el sentido de que pudimos utilizar una gran cantidad de herramientas que nos ofreció la materia y con ellas sacar conclusiones muy interesantes.

Por un lado, en lo que fue un análisis más descriptivo y exploratorio miramos patrones en las palabras más utilizadas, vimos cómo cambiaban a lo largo de las décadas que miramos y cómo evolucionaban conforme las condiciones económicas lo hacían. Observamos que por motivo de la crisis financiera mundial que abarcó varios años del 2007 en adelante el comportamiento de los textos cambió. Asimismo pudimos confirmar que no sólo cambiaban por la situación económica sino también por quién era el interlocutor.

Gracias al sentiment analysis también evaluamos el cambio de tono en los discursos a medida que sucedían noticias más negativas para la economía mundial. De esta manera, con una herramienta que parecía muy simple sacamos información muy potente, encontrando un comportamiento inverso en la positividad del discurso y el aumento de la volatilidad.

En último lugar planteamos un clasificador para ver qué tan fuertes eran los discursos para llevar a cabo *forward guidance* e influir en variables macroeconómicas. Allí nos topamos con la dificultad de que los discursos no eran tan buenos a la hora de clasificar, probablemente porque al ser tan largos y variados generaban mucho ruido en el algoritmo. Las minutas entonces fueron una mejor opción y nos permitieron comprobar algunas hipótesis interesantes respecto a la tardanza de la política monetaria en afectar las tasas de interés. También planteamos un clasificador para ver si las declaraciones que daba el FOMC podían ayudarnos a predecir si la volatilidad iba a subir o a bajar como consecuencia. Al respecto los resultados en general no nos parecen tan buenos, sólo sería prudente afirmar que a medida que pasa el tiempo desde que se efectúa la reunión los resultados son más confiables.

En síntesis, son muchos los ángulos desde donde podemos analizar o extraer información de estos textos. En la exploración de datos hicimos un análisis acerca de los oradores que creemos se puede seguir profundizando, de igual manera con el sentiment analysis. En términos de la clasificación definitivamente hay mucho terreno para mejorar, ya sea cambiando las variables de interés como probando con otros algoritmos.

7 Colab y Materiales

A continuación detallamos el link para acceder al colab y a una carpeta de drive en que se encuentran los archivos necesarios para correr el código. Estos son la base de Bloomberg con indicadores financieros, el archivo con las tasas de referencia de la Fed y un archivo con fechas que utiliza el scraping de las minutas para obtener las mismas.

https://colab.research.google.com/drive/1s_CzYGZqlPJm_ntb3IzZ4fQPYSbfWxDl?usp=sharing
<https://drive.google.com/drive/folders/1TvVNjnJk14Ns3ULSNh30B1r2eezrwWLg?usp=sharing>

Bibliografía

Bernanke, Ben S. *Blog*. URL: <https://www.brookings.edu/blog/ben-bernanke/>.

Boukus, Ellyn and Joshua V Rosenberg. “The information content of FOMC minutes”. In: *Available at SSRN 922312* (2006).

Wade, Garrett. *What the Fed said*.

Zadeh, Reza Bosagh and Andreas Zollmann. *Predicting Market-Volatility from Federal Reserve Board Meeting Minutes NLP for Finance*. 2009.