UNIVERSIDAD NACIONAL DE ROSARIO
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA

CARRERA DE POSGRADO

MAESTRÍA EN FINANZAS

Tema: La inversión de la curva de rendimientos en Estados Unidos y su relación con las recesiones. Una estimación de su impacto a través de distintos análisis de regresión con aplicaciones en programación

Autor: Lic. Esteban Hernán Smolarz

Director: Mag. Diego Hernán Iaccarino

Rosario, lunes 10 de octubre de 2022

A mi familia

Resumen

La relación que existe entre la inversión de la curva de rendimientos y la posterior aparición de recesiones es un fenómeno conocido en Estados Unidos. Desde hace cuarenta años distintos académicos e instituciones han propuesto modelos que pretenden explicar esta relación y calcular las probabilidades de una recesión. Partiendo de este basamento teórico, es factible explorar otras dimensiones de este patrón que se repite entre el diferencial negativo de tasas entre bonos de corto y largo plazo, por un lado, y el advenimiento de recesiones, por el otro. Puntualmente, esta Tesis Final intenta descubrir dos posibles cualidades. En primer lugar, qué conexión puede establecerse entre la profundidad de la inversión de la curva de rendimientos (cuán negativo es el diferencial de tasas) y la gravedad de la recesión posterior (cuánto cae la economía). En segundo lugar, qué vínculo subyace entre la persistencia de una curva invertida y la duración de una recesión. Para la primera de estas cuestiones no se descubrió ninguna relación significativa. Pero para la segunda, la evidencia permite afirmar que existe una correlación positiva entre la extensión temporal de una inversión de la curva de rendimientos y la posterior recesión. Para arribar a estas conclusiones, metodológicamente se efectuaron distintos análisis de regresión. Los modelos fueron implementados en lenguaje de programación R, lo que aporta un valor adicional.

Palabras clave: curva de rendimientos – curva de rendimientos invertida – recesiones – modelos de regresión – programación en R

Abstract

The relationship that exists between a yield curve inversion and the subsequent emergence of a recession is a well-known phenomenon in the United States. For the last 40 years several academicians and institutions have proposed models that attempt to explain this relationship and calculate the probabilities of a recession. Starting from this theoretical base, it is feasible to explore other dimensions of this pattern that repeats between the negative spread between short-term and long-term bonds, on one hand, and the advent of recessions, on the other hand. More precisely, this Final Thesis tries to discover two possible qualities. First, what connection can be established between the depth of a yield curve inversion (how negative the spread is) and the severity of the subsequent recession (how much does the economy fall). Second, what link underlies between the persistence of an inverted yield curve and the length of a recession. For the first of these questions, no meaningful relationship was found out. But for the second, the evidence enables to claim that there exists a positive correlation between the temporal extension of a yield curve inversion and the subsequent recession. To come to these conclusions, methodologically speaking different regression analyses were performed. The models were carried out in R programming language, which gives an additional value to this thesis.

Keywords: yield curve – inverted yield curve – recessions – regression models – programming in R

Tabla de contenidos

I. Introducción... 7

II. Preguntas e hipótesis de investigación 10

III. Estado de la cuestión 14

III.1 Curva de rendimientos... 14

III.2 Inversión de la curva de rendimientos 16

III.3 Recesión.. 20

III.4 *Spreads* negativos y recesión... 21

IV. Metodología y conceptos estadísticos 24

IV.1 Análisis de regresión lineal.. 24

IV.2 Conceptos estadísticos relevantes.. 31

IV.2.1 Bondad de ajuste y coeficiente de determinación.................................. 32

IV.2.2 R cuadrada ajustada... 33

IV.2.3 Error estándar.. 34

IV.2.4 Valor t 36

IV.2.5 Valor p 36

IV.2.6 Estadístico F.. 36

V. Análisis de datos y resultados 38

V.1 Información original.. 38

V.2 Primer análisis de regresión.. 56

V.2.1 Modelo de regresión lineal simple 1 56

V.2.2 Modelo de regresión lineal simple 2 64

V.2.3 Modelo de regresión lineal múltiple 72

V.3 Segundo análisis de regresión 79

VI. Conclusiones 87

VII. Referencias bibliográficas 91

VIII. Siglas y acrónimos 93

IX. Apéndices 94

IX.1 Código en R 94

IX.2 Código en R con *output* 98

IX.3 *Environment* en R 109

I. Introducción

Desde mediados de 2022, una inquietud comenzó a transmitirse entre los agentes de los principales mercados financieros del mundo. La curva de rendimientos[1], que grafica el retorno de los bonos soberanos de Estados Unidos (EEUU) para distintos plazos, comenzaba a mostrar signos de estar invirtiéndose (Rennison, 2022). En su forma "normal", la *yield curve* presenta rendimientos más altos para plazos más lejanos. Cuando se invierte, ocurre lo opuesto: los intereses más altos se encuentran entre los bonos de más corto plazo, mientras que los bonos con vencimiento[2] más lejano en el tiempo ven caer su rendimiento.

Desde la década de 1980, a medida que EEUU empieza a desindustrializarse en términos relativos, ganando peso el sector financiero, la curva de rendimientos se ha convertido en un indicador clave para realizar vaticinios acerca de la marcha futura de la economía. Los operadores de bancos y empresas siguen constantemente la evolución del diferencial[3] entre bonos de distinto plazo, atentos a cualquier acontecimiento que pueda señalar que dicho *spread* está por tornarse negativo.

Igualmente, las autoridades monetarias, especialmente las de la Reserva Federal, realizan un seguimiento de estas variables, teniendo preparadas medidas que apuntan específicamente a corregir desviaciones en la curva de rendimientos. Las herramientas empleadas suelen ser la manipulación de la tasa de interés de referencia, así como el control de los agregados monetarios.

Teniendo en cuenta la centralidad indiscutible de la economía estadounidense para el conjunto de la economía mundial, es imperioso entender cuáles son las principales variables que la afectan. Las alteraciones en la actividad económica norteamericana repercuten en todos los mercados y en todos los países. Del mismo modo, las decisiones políticas, sobre todo en materia monetaria y crediticia, ejercen un impacto

[1] *Yield curve* en inglés.
[2] *Maturity* en inglés.
[3] *Spread* en inglés.

sin parangón sobre el escenario en el cual deben desenvolverse empresas, hogares y gobiernos de terceros países.

Por lo tanto, entender cuál es la relación que existe entre la inversión de la curva de rendimientos por un lado, y las recesiones por el otro, es una tarea obligada para discernir acabadamente cómo funciona la economía moderna. De manera análoga, constituye un tema de investigación interesante para una Tesis Final de Maestría en Finanzas.

Una motivación adicional para encarar este trabajo de investigación, es que hay disponible información pública para llevar a cabo los análisis. El gobierno norteamericano, a través de distintos organismos y agencias, como la misma Reserva Federal o el National Bureau of Economic Research (NBER), se encarga de facilitar el acceso a estos datos.

Metodológicamente, se llevarán a cabo distintos análisis de regresión que buscarán encontrar correlaciones entre distintas propiedades de las dos variables identificadas (es decir, entre la inversión de la curva de rendimientos y las recesiones). Se trata de modelos de regresión lineal. Aunque la regresión lineal por definición presenta limitaciones para modelar acabadamente la realidad, es sin embargo un instrumento útil en tanto habilita interpretaciones transparentes de las propiedades y relaciones que son de interés en la investigación.

Además de la ventaja que entrañan a la hora de testear hipótesis, la regresión lineal también se presta con facilidad para ser ejecutado con herramientas informáticas. Especialmente, constituye un ejercicio valioso para ser replicado en entornos de programación. En este caso, el lenguaje escogido es R, y la aplicación correspondientes es RStudio. La implementación de análisis estadístico aplicado a finanzas en entornos de programación es un valor agregado de esta Tesis Final.

A modo de aclaración, en este texto se entiende por curva de rendimientos la *yield curve* de los bonos soberanos del gobierno de EEUU. El *spread* al que se hace mención repetidas veces es el que vincula a bonos soberanos de corto y largo plazo del Departamento del Tesoro de ese país. Temporalmente, el análisis se circunscribe al período posterior a la Segunda Guerra Mundial (SGM). Por último, por recesión se entiende las sucesivas caídas en el producto bruto interno (PBI) a lo largo de un número de períodos consecutivos en el tiempo.

II. Preguntas e hipótesis de investigación

Como ya se comentó en la Introducción, esta Tesis Final de Maestría tiene como objeto desarrollar la temática de la relación entre la inversión de la curva de rendimientos y las recesiones en EEUU desde la SGM. Esta relación ya ha sido teorizada y constatada. Un vistazo superficial a la información histórica permite comprobar lo que los modelos proponen: que hay una alta probabilidad de que a una inversión de la curva de rendimientos le suceda una recesión.

La pregunta general que orienta este trabajo de investigación es la siguiente:

¿Qué tipo de correlaciones existen entre la inversión de la curva de rendimientos y las recesiones en EEUU desde la SGM, y cómo pueden modelarse estas correlaciones de manera operativa?

La hipótesis general que intenta responderla es la siguiente:

Las correlaciones existentes entre la inversión de la curva de rendimientos y las recesiones en EEUU desde la SGM surgen de características tales como su intensidad (en términos de cuán profunda es una inversión de la curva y cuán intensa es una recesión) y su duración. Operativamente, dichas correlaciones pueden modelarse recurriendo a análisis de regresión que identifiquen propiedades relevantes, implementando estos análisis en lenguaje de programación R.

La variable independiente es la curva de rendimientos en EEUU (limitada a los casos en que está invertida) en sus distintas dimensiones, tales como el valor numérico del *spread* o la duración en el tiempo de la inversión. La variable dependiente es el crecimiento de la economía norteamericana medido a través del PBI. Al igual que con la variable dependiente, se toman sólo los valores para los cuales el crecimiento del PBI es negativo.

De esta pregunta general inicial se derivan dos preguntas específicas, con sus correspondientes hipótesis. Al igual que con la pregunta y la hipótesis generales, las dos preguntas específicas y sus respectivas hipótesis deben pensarse como geográficamente ubicadas en EEUU y cronológicamente situadas después de la SGM. En términos operativos, cada hipótesis específica intenta ser respondida mediante un modelo de regresión lineal. En algunos casos, distintas variantes de un mismo modelo de regresión lineal, sea simple o múltiple, son implementadas.

La primera pregunta de investigación específica está formulada de la siguiente manera:

¿De qué manera repercute la magnitud de la inversión de la curva de rendimientos en la profundidad de la recesión que acontece con posterioridad a la aparición de dicha inversión?

La hipótesis específica que intenta responder de manera tentativa dicho interrogante es la siguiente:

Existe una correlación positiva y significativa entre la magnitud de la inversión de la curva de rendimientos, por un lado, y la profundidad de la recesión que acontece con posterioridad a la aparición de dicha inversión, por el otro.

Es decir, la hipótesis afirma que cuánto más amplio y negativo es el *spread* entre bonos de corto y largo plazo, más grave es la recesión que le sucede (el PBI cae con más fuerza). Está primera hipótesis específica es testeada mediante un modelo de regresión que cuenta con tres variantes.

La primera variante se trata de un modelo de regresión lineal simple o bivariado, es decir, con una única variable explicativa. La segunda variante mantiene la forma funcional del primero, pero altera el *data set* mediante la eliminación de las observaciones más extremas. La tercera variante conserva este conjunto de datos más restringido, pero empleando una forma funcional cuadráticas. Esto es, esta tercera y última variante del primer modelo de regresión cuenta con dos variables independientes, una de ellas la variable explicativa original en su forma lineal, y la otra el cuadrado de aquella. En consecuencia, consiste en un modelo ya no simple, sino múltiple.

Para las tres variantes de este primer modelo de regresión, la variable independiente es la curva de rendimientos analizada desde el punto de vista de los puntos porcentuales negativos que caracterizan su inversión. Incluso para la tercera variante, la del modelo de regresión lineal múltiple con dos variables independientes, éste también es el caso. Esto es así porque la segunda variable explicativa es el cuadrado de la otra, por lo que en esencia se trata de la misma variable. La variable dependiente es el crecimiento de la economía en términos porcentuales negativos.

La segunda pregunta de investigación específica se propone de la siguiente forma:

¿De qué manera impacta la duración de una inversión de la curva de rendimientos en la duración de la recesión que acontece con posterioridad a la aparición de dicha inversión?

La hipótesis específica que intenta responder de manera tentativa dicho interrogante es la siguiente:

Existe una correlación positiva y significativa entre la duración de una inversión de la curva de rendimientos, por un lado, y la duración de la recesión que acontece con posterioridad a la aparición de dicha inversión, por el otro.

En otras palabras, esta segunda hipótesis específica sostiene que cuanto más prolongado es el *spread* (cuánto mayor es el tiempo que transcurre desde que la curva se invierte hasta que retorna a su forma normal), más extensa temporalmente es la recesión que ocurre posteriormente. Para poner a prueba esta hipótesis, es que se lleva a cabo un único análisis de regresión que consiste en un modelo lineal simple.

Como es evidente, ambos pares de preguntas e hipótesis específicas intentan ser respondidos mediante análisis de regresión. En el caso de la primera pregunta de investigación específica, le corresponden tres análisis de regresión que son variantes de un mismo modelo. Dos de ellos son simples y el tercero es múltiple. En el caso de la segunda pregunta de investigación específica, se trata de un único modelo de regresión lineal simple. En total, se realizan tres análisis de regresión simple y uno múltiple.

Retornando a la pregunta general, estos cuatro análisis de regresión son modelados recurriendo al lenguaje de programación R. Específicamente, se emplea la aplicación RStudio, que tiene la ventaja de no sólo dejar registrados el código y el *output*, sino también los gráficos, los modelos y los *data sets* (estos dos últimos en el *environment*). Este aspecto, de índole operativa, es indispensable para poder implementar en la práctica la cuestión de la identificación de las propiedades relevantes de las variables independiente y dependiente.

III. Estado de la cuestión

En esta sección se desarrollan los principales conceptos y teorías que intentan explicar la relación entre la inversión de la curva de rendimientos y la aparición de recesiones. Este repaso de la literatura relevante en el área permite introducirse en el tema y saber cuáles propiedades de la correlación entre ambas variables ya han sido probadas. En combinación con los conceptos y procedimientos metodológicos, que se abordan en la siguiente sección, proveen los insumos necesarios para guiar el análisis desde el punto de vista teórico (que acontece en paralelo a la ejecución operativa).

III.1 Curva de rendimientos

El primer concepto que debe ser desarrollado para introducirse al tema de esta Tesis Final concierne a lo que se denomina curva de rendimientos o *yield curve*. La curva de rendimientos, de referencia obligada para quienes trabajan en el sector financiero y los *policy makers* de los principales gobiernos, es la representación gráfica de lo que en términos más formales se conoce como estructura temporal de tasas de interés (ETTI).

La ETTI relaciona dos variables fundamentales de un bono, como son el tiempo y el rendimiento, a través de una función. Los indicadores concretos a los que se acude para dar cuenta de conceptos abstractos como el tiempo y el rendimiento son generalmente la *duration* y la tasa interna de retorno (TIR)[4], respectivamente. No obstante, en otras ocasiones se recurre a otros factores que también están atados al tiempo al vencimiento y a la tasa de interés.

La curva de rendimientos exhibe la íntima ligazón que existe entre el tiempo al vencimiento y el rendimiento en un instrumento de deuda. Es decir, para cada plazo

[4] *Internal Rate of Return* (IRR) en inglés.

es esperable una determinada tasa de interés. La metodología más idónea para construir una *yield curve* no emplea una única TIR para todos los plazos (Johnson, 1984). En vista de ello, se aplican distintos procedimientos para aproximar con la mayor exactitud posible la forma real de la ETTI para un determinado instrumento de deuda.

Un requisito obvio para elaborar una curva de rendimientos consistente estriba en que los bonos que se asignen a los distintos horizontes temporales deben poseer todos similar *rating* crediticio. Si así no fuese, los diferenciales de calidad crediticia imprimirían un sesgo indeseable a la comparación entre tasas de descuento. Consecuentemente, es usual que se utilicen bonos soberanos, principalmente los emitidos por el Departamento del Tesoro de Estados Unidos, para evaluar la curva de rendimientos (Stradi, 2005).

Adicionalmente, los bonos soberanos, especialmente los respaldados por el gobierno estadounidense, son objeto de un volumen significativo de operaciones diarias, de lo cual se derivan dos ventajas. La primera, que esto genera un caudal de información valioso para efectuar análisis y testear hipótesis. La segunda, que no interfieren *spreads* de liquidez que podrían distorsionar la tasa de descuento (Nymand-Andersen, 2018).

Las propiedades de la *yield curve* le permiten cumplir numerosos roles prácticos para realizar análisis financieros y económicos. El que aquí interesa atañe a la forma que asume la curva, que se erige en una referencia ineludible para descifrar el estado de la economía en el presente y aventurar proyecciones acerca de su evolución futura (McCallum, 2005).

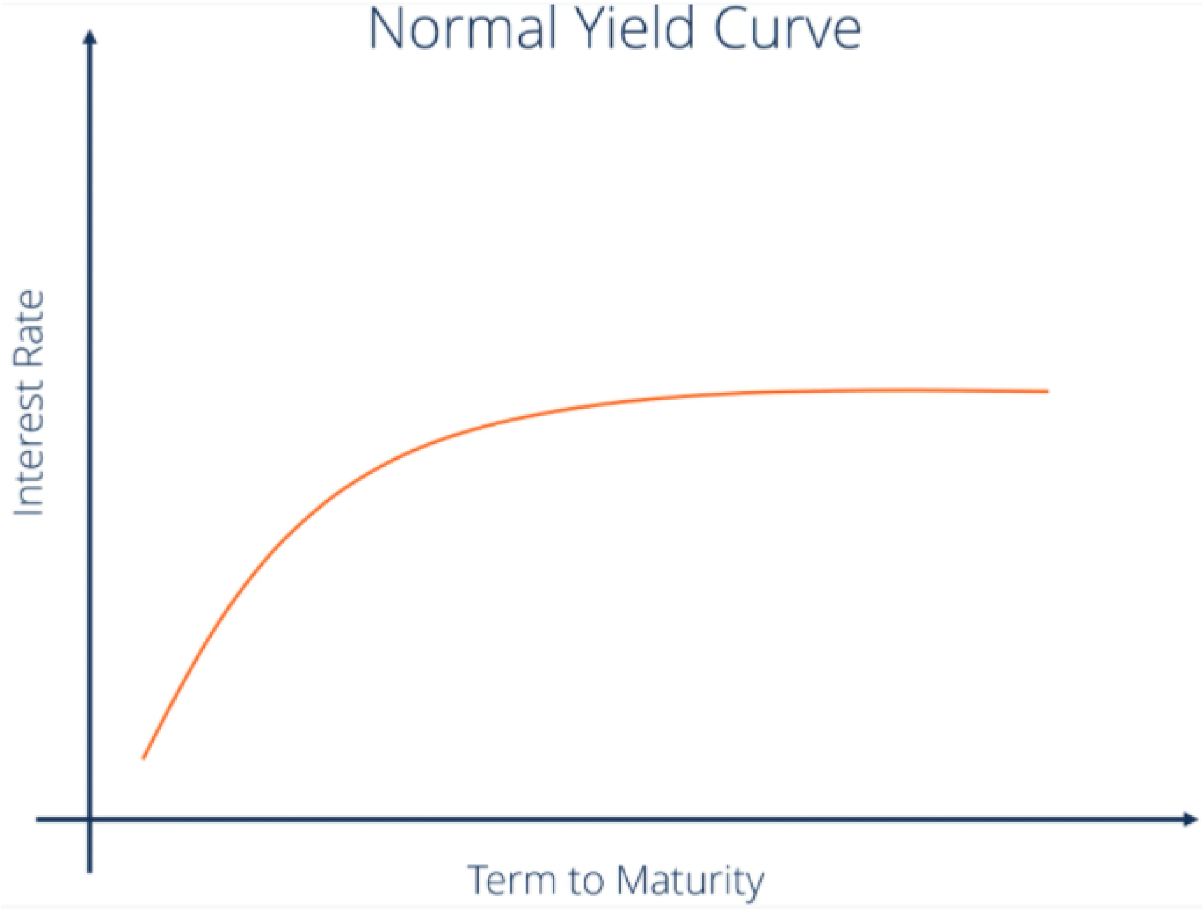
Esencialmente, la forma que adopta la curva y los diferenciales de rendimientos entre distintos períodos aportan datos que facilitan la formulación de conjeturas en relación al ritmo de crecimiento (o decrecimiento) de la economía. En otras palabras, provee indicios sobre si un país entrará en un ciclo expansivo o recesivo. De allí que la ETTI sea constantemente monitoreada por operadores financieros, empresas y autoridades gubernamentales.

III.2 Inversión de la curva de rendimientos

Empíricamente, la curva de rendimientos puede exhibir distintos aspectos. La teoría ha desarrollado una taxonomía de estos perfiles de la ETTI, de los cuales aquí interesan dos. El primero es lo que se llama curva normal o ascendente. Como su nombre lo indica, se trata de una *yield curve* en la que, a mayor plazo al vencimiento, mayor tasa de interés (Fisher, 2004). Matemáticamente, se parece a una función logarítmica, con una pendiente que se suaviza a medida que transcurre el tiempo. Es decir, es una función con una primera derivada positiva y una segunda derivada negativa.

En su dimensión analítica, la curva de rendimientos ascendente entraña perspectivas promisorias respecto de la economía. Es decir, implica una expectativa de que el PBI se expandirá, y que congruentemente la inflación se incrementará. Paralelamente, esto lleva a suponer que el banco central se verá presionado a subir las tasas de interés de más corto plazo para morigerar las tendencias inflacionarias (Clinton, 1995). La curva de rendimientos normal en su forma estándar puede visualizarse en la Ilustración 1.

Ilustración 1: yield curve normal o ascendente



Fuente: Corporate Finance Institute. Yield curve – Definition, diagrams, types of yield curves

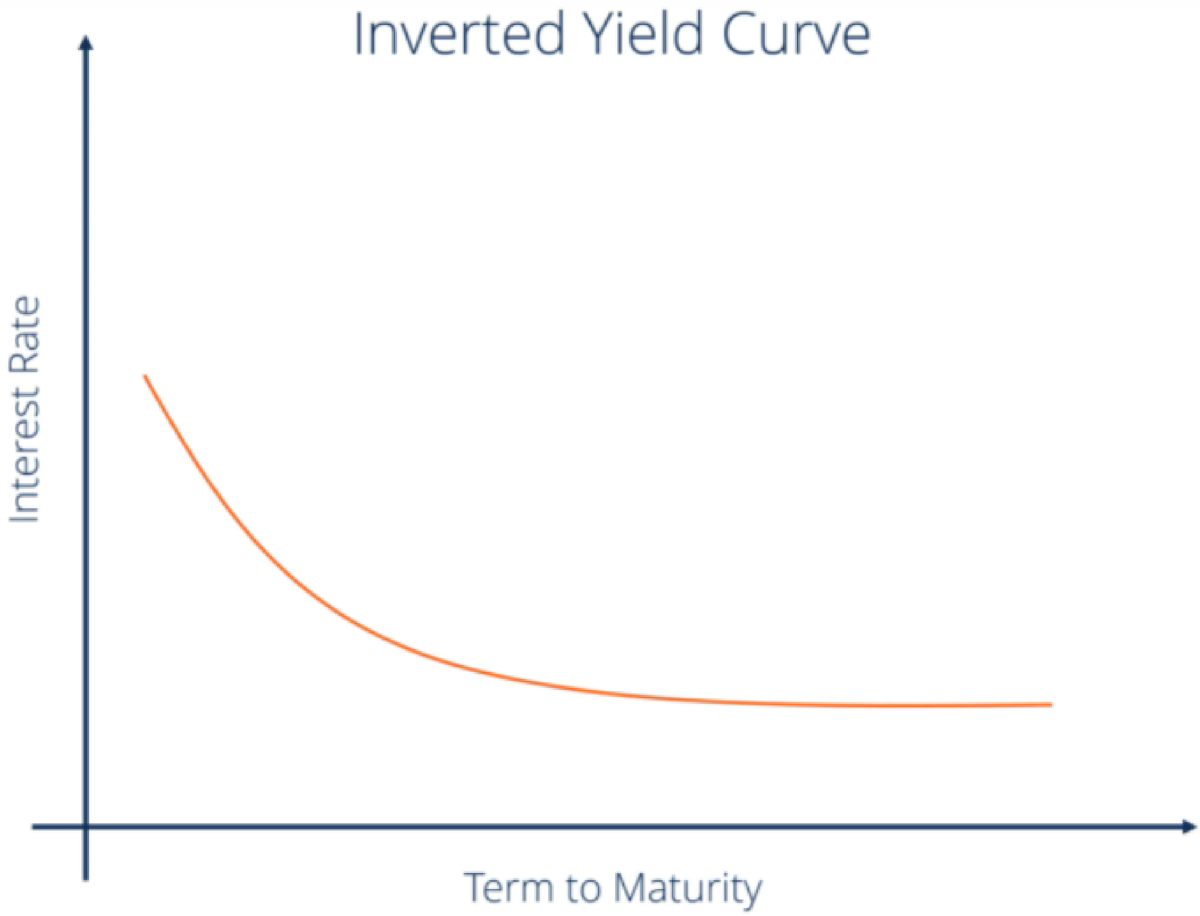
La otra modalidad de la curva de rendimientos que es relevante para estudiar su nexo con el ciclo económico es lo que se conoce como curva invertida o descendente. En contraste con la curva normal, en la curva descendente las tasas de corto plazo son más altas que las de largo plazo. Por lo tanto, cuanto mayor es el horizonte temporal, más baja es la tasa (Wang & Yang, 2011). Comparte con la curva ascendente la propiedad de que la pendiente es más empinada para plazos menores, tornándose más horizontal para plazos mayores.

En los hechos, cuando la *yield curve* presenta esta característica de inversión, acontece que tras un plazo que es fluctuante entre un evento y otro, la economía entra en recesión. Es lo que ha ocurrido en EEUU en todos los casos de inversión de la curva de rendimientos (con una sola excepción) con posterioridad a la SGM (Lonski, 2019). En ocasiones, la inversión de la curva puede verse generada en parte por expectativas inflacionarias a la baja, lo cual suele indicar un desaceleramiento de la economía (Wright, 2006).

El comportamiento de los inversores está naturalmente ligado a los ciclos económicos. Cuando la economía se expande, los agentes exigen mayores *yields* de bonos de más largo plazo por dos motivos principalmente. Primero, como una manera de cubrir el costo de oportunidad de invertir en un activo de renta fija como un bono en lugar de activos de renta variable como acciones. Segundo, para resguardarse ante la inflación.

Cuando la curva comienza a aplanarse y luego a invertirse, con el correlato de contracción económica que ello trae aparejado, es esperable que la *performance* de las acciones empeore en términos relativos. Los bonos de largo plazo se erigen así en una alternativa atractiva. La mayor demanda por estos instrumentos de renta fija de largo plazo incrementa su precio y disminuye su rendimiento, lo que eventualmente deriva en la inversión de la curva. Un ejemplo de curva de rendimientos descendente puede apreciarse en la Ilustración 2.

Ilustración 2: yield curve invertida o descendente



Fuente: Corporate Finance Institute. Yield curve – Definition, diagrams, types of yield curves

Teniendo en cuenta que la ETTI tiene como sustrato la relación que se establece entre los *maturities* de los bonos y su TIR, asignándose a cada plazo un rendimiento, se infiere que otra aproximación a esta interconexión entre tasas y tiempo puede formularse a través de los diferenciales de rendimiento. Efectivamente, el *spread* de tasas de interés es otra manera de examinar el comportamiento de la curva.

Cuando a la tasa de largo plazo se le sustrae la tasa de corto plazo y la diferencia es positiva, se está ante una curva ascendente. Si la diferencia continúa siendo positiva, pero el diferencial entre tasas aumenta, se está en presencia de una curva *steep* o empinada. Por el contrario, si la brecha se achica, significa que la curva se está aplanando (*flat curve*). La curva *flat* es el primer atisbo de que la curva podría invertirse. Cuando finalmente esto ocurre, el *spread* de tasas de interés se torna negativo.

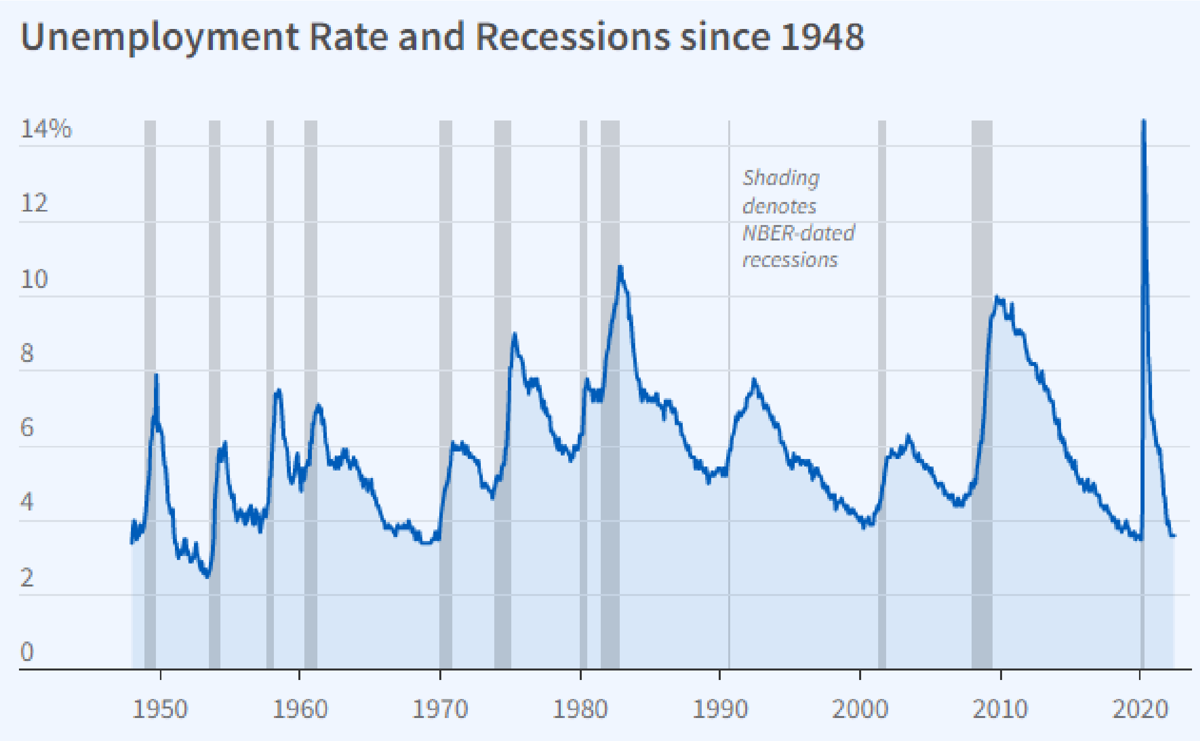
III.3 Recesión

La cuestión de qué se entiende por recesión no es interesante sólo como debate teórico, sino que adquiere fundamental importancia en el marco de este trabajo de investigación. En efecto, recesión es la variable dependiente de la relación que aquí se estudia entre inversión de la curva de rendimientos y ciclo económico. En consecuencia, se impone la necesidad de delimitar conceptualmente el término, pero sobre todo, de designar un indicador inequívoco para evitar inconvenientes metodológicos.

El hecho de que el NBER cuente con un *Business Cycle Dating Committee* encargado de delimitar la cronología de los ciclos económicos en EEUU (y por ende, de señalar los indicadores y variables a considerar), denota que el tema de aclarar qué significa el concepto de recesión es un asunto de interés público. En términos abstractos, una recesión es una caída generalizada y prolongada del nivel de actividad económica. Pero de esta definición pueden dimanar varios indicadores.

El criterio más aceptado para anunciar oficialmente una recesión es cuando el PBI de un país experimenta dos trimestres consecutivos de caída (Claessens, 2009). Sin embargo, el NBER y el gobierno norteamericano toman en cuenta otros indicadores además del PBI, tales como el ingreso real, el nivel de producción industrial, la inversión agregada y sobre todo la tasa de desempleo (National Bureau of Economic Research, 2020). La conexión entre tasa de desempleo y recesiones puede juzgarse en la Ilustración 3.

Ilustración 3: tasa de desempleo y recesiones en EEUU



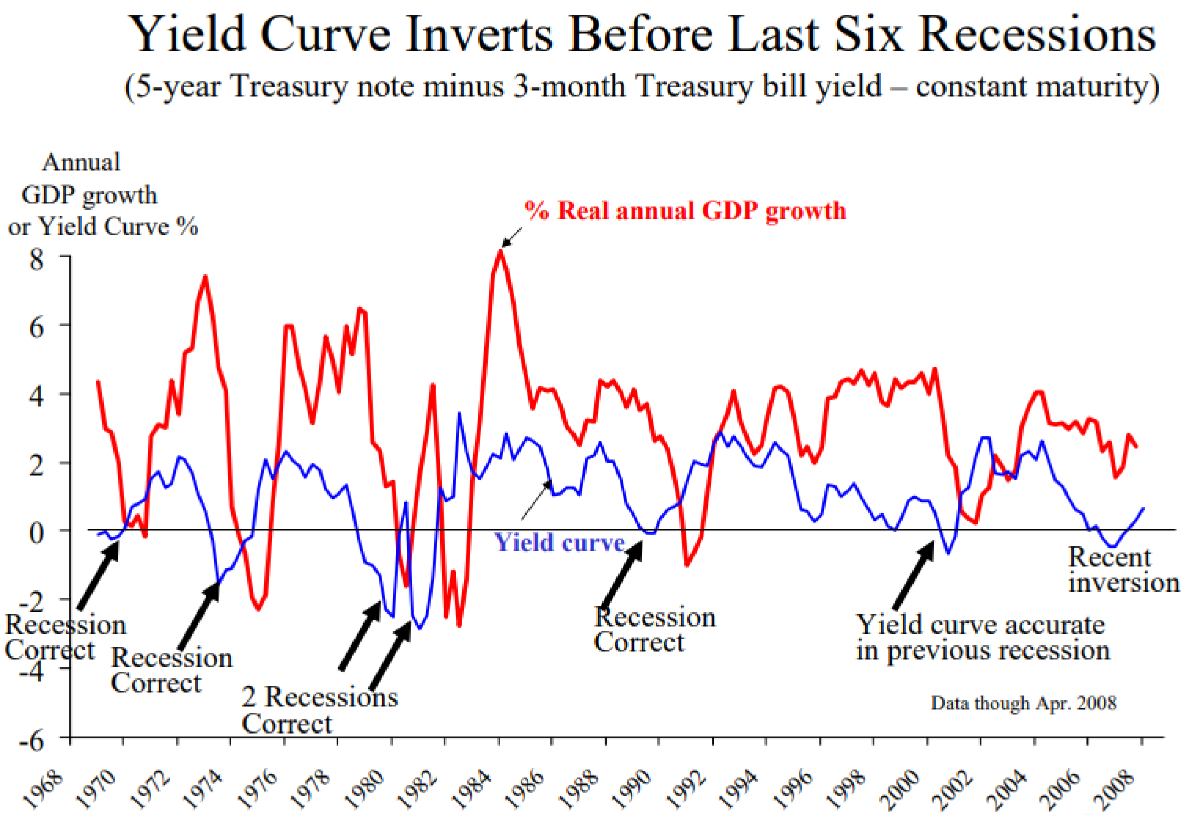
Fuente: NBER; Federal Reserve Bank of St. Louis

Hecha esta mención respecto al debate académico vigente en torno al concepto de recesión, y las definiciones oficiales correspondientes, vale aclarar que en esta tesis se toma un concepto restringido de recesión. Esto es así porque por motivos metodológicos, es conveniente que la variable de respuesta se reduzca a una sola dimensión. En consecuencia, se respeta la definición estándar de recesión, esto es, cuando la economía presenta crecimiento negativo por dos o más trimestres consecutivos.

III.4 *Spreads* negativos y recesión

Aunque hoy en día es un lugar común afirmar que existe una correlación positiva entre el *spread* de tasas de interés y el ciclo económico, o que "a una inversión de la curva de rendimientos le sigue una recesión", hubo un economista que fue precursor en este sentido. Se trata de Campbell R. Harvey, quien en su disertación doctoral (1986) elaboró un modelo predictivo de recesiones basado en la inversión de la curva de rendimientos. No es necesario aducir lo preciso que es este modelo, algo esperable en función de la información de público conocimiento y que se refleja en la Ilustración 4.

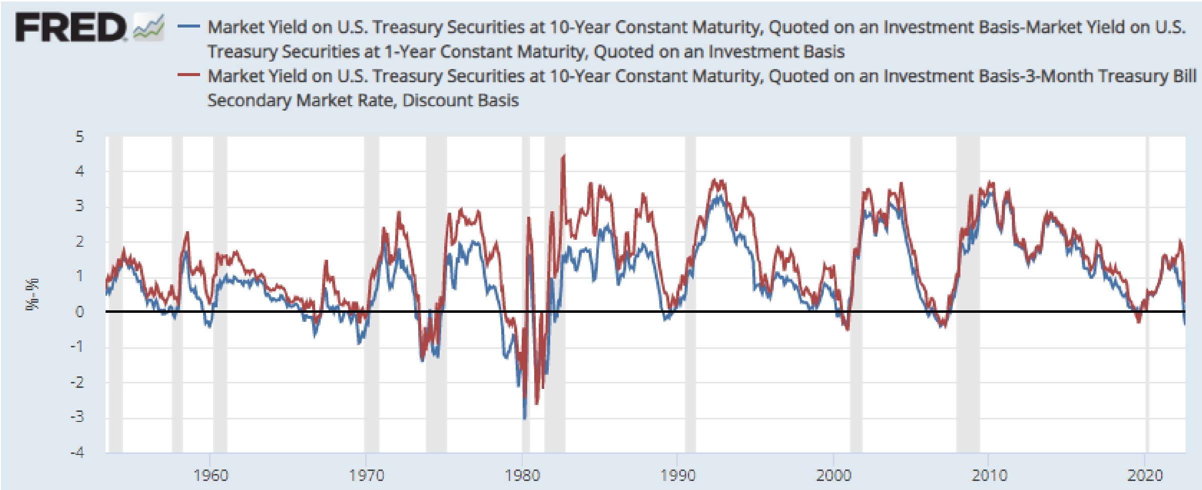
Ilustración 4: recesiones precedidas por inversiones de la yield curve



Fuente: Campbell R. Harvey

De acuerdo a la información oficial publicada por organismos tales como el *Board of Governors of the Federal Reserve System*, el NBER y el *U.S. Bureau of Economic Analysis*, desde 1948 EEUU ha sufrido doce recesiones. La información es analizada en detalle en la subsección V.1, pero en este momento basta con apuntar que es factible detectar el antedicho patrón por el cual (con excepciones puntuales), a una inversión de la curva o *spread* negativo de tasas de interés, le sucede una recesión. Esto puede verse en la Ilustración 5.

Ilustración 5: inversiones de la yield curve y recesiones en EEUU tras la SGM



Fuente: Board of Governors of the Federal Reserve System

Como ya se mencionó, esta relación entre *spreads* negativos y recesiones es el objeto de investigación de esta Tesis Final. Estando suficientemente probada la correlación (con un *lag* temporal) entre inversión de la curva y contracción del ciclo económico, un trabajo de indagación en esta área de estudios debe formularse otras preguntas si es que se pretende realizar un aporte original. Entre ellas, de qué manera se establece esa correlación, qué factores intervienen y cuáles herramientas pueden instrumentalizarse para estudiarla.

IV. Metodología y conceptos estadísticos

El tipo de información que se emplea en esta Tesis Final es lo que se denomina serie temporal[5]. Por definición, los datos de una serie temporal ya están ordenados cronológicamente. Tanto la información relativa a las recesiones, como aquella concerniente a las inversiones de la curva de rendimientos, constituyen series temporales.

Sin embargo, cabe una aclaración. Por un lado, la frecuencia de la serie temporal sobre la tasa de crecimiento del PBI está expresada en trimestres o *quarters*. Por el otro, la frecuencia de la serie temporal de *spreads* entre bonos de corto y largo plazo se encuentra disponible en formato mensual. Aunque pareciera que esto podría acarrear dificultades metodológicas, en la sección V se explica cómo puede superarse este inconveniente.

IV.1 Análisis de regresión lineal

El modo en que están planteadas las preguntas e hipótesis de investigación, tanto generales como específicas, impone ciertas limitaciones metodológicas. La más importante radica en elegir las herramientas apropiadas para dar respuesta a las preguntas y testear las hipótesis. Ya se anticipó que se efectúan análisis de regresión para llevar a cabo la investigación, ya que por definición son el instrumento idóneo para encontrar correlaciones entre variables.

Una motivación adicional para esta elección es que pueden ser directamente replicados en un entorno de programación informático con un lenguaje específico. En este caso, el entorno de programación es RStudio y el lenguaje es R. Una parte apreciable del valor agregado de esta Tesis Final y del aporte original que pretende

[5] *Time series* en inglés.

realizar al campo de estudios en cuestión, es combinar conceptos de finanzas con aplicaciones informáticas y de programación.

Por análisis de regresión se entiende el conjunto de procedimientos estadísticos que vinculan dos o más variables, entre las cuales se encuentra una variable dependiente, y una o más variables independientes. Entre las distintas alternativas de análisis de regresión, se encuentra la regresión lineal, la cual puede ser simple o múltiple dependiendo de si intervienen una o más variables independientes, respectivamente.

Un modelo de regresión lineal toma como inputs observaciones de la variable dependiente y de la(s) variable(s) independiente(s) y arroja como resultado una fórmula. Esta fórmula es llamada ecuación de regresión, teniendo la particularidad de que sus parámetros (a priori desconocidos y estimados justamente a través del modelo de regresión) son lineales. La propiedad de linearidad no se extiende a la(s) variable(s) independiente(s), que pueden o no ser lineales (Hansen, 2020).

El modelo de regresión lineal simple, también denominado regresión lineal bivariado, se define por poseer sólo una variable independiente. El modelo de regresión lineal múltiple, por el contrario, se halla condicionado por dos o más variables explicativas. En esta Tesis Final, se implementan tanto modelos de regresión lineal simple como modelos de regresión lineal múltiple. La fórmula general que relaciona la variable explicada con la variable independiente en el modelo de regresión lineal simple está contenida en la Ecuación 1.

Ecuación 1: regresión lineal simple

$$y = \beta_0 + \beta_1 x + u$$

Donde:

y es el valor de la variable dependiente

β_0 es el intercepto, constante u ordenada al origen

β_1 es el parámetro de la variable explicativa o parámetro de la pendiente

x es el valor de la variable independiente *u* es el término de error

Este último concepto cumple un rol atado al hecho que el carácter lineal del modelo de regresión simple implica que, si los factores que intervienen en el término de error se mantienen fijos, puede afirmarse que la variable predictiva tiene un efecto lineal sobre la variable predicha. Esta afirmación se ve reflejada en la Ecuación 2.

Ecuación 2: efecto lineal de la variable predictiva

$$\Delta u = 0 \rightarrow \Delta y = \beta_1 \Delta x$$

Donde:
Δu es el cambio en el término de error
Δy es el cambio en la variable dependiente
β_1 es el parámetro de la variable explicativa o parámetro de la pendiente
Δx es el cambio en la variable independiente

De lo anterior se deriva que el parámetro β_1 es igual a la ratio entre el cambio en la variable dependiente y el cambio en la variable independiente, *ceteris paribus*, tal como se muestra en la Ecuación 3.

Ecuación 3: parámetro β1

$$\beta_1 = \frac{\Delta y}{\Delta x}$$

Cualquiera sea la manera en que se lo exprese, matemáticamente el cambio en la variable dependiente es igual al cambio en la variable independiente multiplicado por el parámetro de la pendiente β_1 (Wooldridge, 2015).

Por definición, el modelo de regresión lineal simple padece de limitaciones. Es poco probable que una relación entre variables pueda ser modelada de manera realista con un análisis de regresión lineal simple. Sin embargo, los modelos lineales aproximan (de manera lineal precisamente) las interacciones no lineales que en la vida real se establecen entre las variables.

La gran ventaja del modelo lineal estriba en que su estimación resulta sencilla, y los parámetros obtenidos pueden interpretarse sin mayores dificultades. Asimismo, los parámetros se prestan para un análisis claro de sus distintas propiedades estadísticas, lo que permite probar hipótesis a través de ellos.

Además, es un medio útil para implementar en un entorno de programación informático (en este caso, el lenguaje R) conceptos específicos, además de poner a prueba (de manera muy restringida) ciertas hipótesis relativas al tema de investigación escogido.

El método estándar para estimar los parámetros de un modelo de regresión es el de mínimos cuadrados ordinarios[6] (MCO). Este procedimiento minimiza la suma de los residuales cuadrados, entendiéndose por residual la diferencia entre el valor real de una observación y su valor ajustado. El valor ajustado es aquel que cae sobre la línea de regresión para el correspondiente valor de *x* del punto de dato.

Gráficamente, el residual es la distancia vertical entre los valores de la muestra y los valores estimados por el modelo. La fórmula matemática (un problema de optimización) que refleja la lógica del método de MCO se reproduce en la Ecuación 4.

Ecuación 4: parámetro β̂

$$\hat{\beta} = \min_{\beta} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 = \min_{\beta} \sum_{i=1}^{n} \varepsilon_i^2$$

[6] *Ordinary least squares (OLS)* en inglés.

Donde:

$\hat{\beta}$ es el estimador de MCO (la β que minimiza la expresión contenida dentro del índice sumatorio)

y_i es la observación *i* de la variable dependiente

α es el parámetro (no observable) de la variable dependiente cuando el valor de la variable independiente es cero

β es el parámetro (no observable) que da cuenta del cambio en la variable de respuesta cuando la variable explicativa varía de forma unitaria

x_i es la observación *i* de la variable independiente

ε_i es el *i*-término de error

El intercepto α también puede denotarse como β_0, y se determina por medio de la Ecuación 5.

Ecuación 5: parámetro $\hat{\beta}0$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Donde:

$\hat{\beta}_0$ es el estimador de MCO del intercepto

$\bar{y}$ es el promedio o valor esperado[7] de la variable independiente

$\hat{\beta}_1$ es el estimador de MCO del parámetro de la variable independiente

$\bar{x}$ es el promedio o valor esperado de la variable dependiente

Por último, el parámetro de la pendiente de la línea de regresión, que afecta a la variable independiente *X* y determina en cuánto se altera la variable dependiente *Y* ante cambios en aquella, se calcula a través de la Ecuación 6.

[7] El valor esperado es lo que se llama una “medida de tendencia central” de una variable aleatoria. En la práctica, es común tomar el promedio como indicador de medida de tendencia central.

Ecuación 6: parámetro $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Donde:
$\hat{\beta}_1$ es el estimador de MCO del parámetro de la variable independiente
x_i es la observación *i* de la variable independiente
$\bar{x}$ es el promedio o valor esperado de la variable dependiente
y_i es la observación *i* de la variable dependiente
$\bar{y}$ es el promedio o valor esperado de la variable independiente

Cuando un análisis de regresión cuenta con más de una variable independiente, se está ante un modelo de regresión lineal múltiple. Las ventajas de la regresión múltiple son evidentes. Permite añadir más factores explicativos, lo que en principio dotaría de mayor poder predictivo al modelo. A la vez, la presencia de varias variables independientes permite mantener fijos o sin cambios a las demás variables independientes excepto una (esto es la condición de *ceteris paribus*), de la cual se intentan descubrir sus efectos parciales sobre la variable de respuesta.

Finalmente, los modelos de regresión múltiple admiten el empleo de distintos recursos que revisten de mayor versatilidad a las fórmulas de regresión. En oposición, la forma funcional en la regresión simple, esto es, la expresión algebraica ya descripta que relaciona la variable independiente con la variable dependiente, es siempre la misma.

Gran parte de la utilidad de la regresión múltiple radica en que es un mecanismo para combinar simultáneamente diversas variables independientes que pueden estar correlacionadas entre sí. El no incluir una variable que posee un alto grado de correlación con otra, puede conducir a sobreestimar los efectos sobre la variable dependiente de la variable independiente que sí está incluida en el modelo. Esto se vincula con el problema del sesgo de variable omitida[8].

[8] *Omitted variable bias (OVB)* en inglés.

En esta Tesis Final se llevan a cabo tres análisis de regresión. Dos de ellos se tratan de modelos de regresión lineal simple, mientras que el tercero consiste en un modelo de regresión lineal múltiple. En particular, el modelo de regresión lineal múltiple se basa en una forma funcional que explica la profundidad de una recesión como una función cuadrática de la gravedad de la inversión de la curva de rendimientos.

Es menester aclarar que, en este caso, el cambio en la variable independiente no sólo depende de las fluctuaciones en la variable independiente, sino también del valor previamente existente de la misma variable explicativa. La Ecuación 7 representa este modelo de regresión lineal múltiple con una función cuadrática.

Ecuación 7: regresión lineal múltiple con forma funcional cuadrática

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

Donde:
y es el valor de la variable dependiente
β_0 es el intercepto, constante u ordenada al origen
β_1 es el parámetro de la variable explicativa o parámetro de la pendiente
β_2 es el parámetro de la variable explicativa elevada al cuadrado
x es el valor de la variable independiente
u es el término de error

Para determinar cómo cambia la variable dependiente en función de los cambios en la variable dependiente, debe tomarse la primera derivada de esta última, tal como se muestra en la Ecuación 8.

Ecuación 8: tasa de cambio con forma funcional cuadrática

$$\frac{\Delta y}{\Delta x} = \beta_1 + 2\beta_2 x$$

Los modelos de regresión lineal múltiple permiten realizar análisis *ceteris paribus* en relación a una determinada variable independiente justamente porque habilitan un control preciso de las demás variables independientes. En otras palabras, facultan una interpretación de los efectos sobre la variable dependiente de los cambios en una de las variables independientes, manteniendo constantes todas las demás variables. Esto es posible incluso si existe correlación entre la variable analizada y las variables que se mantienen fijas. Matemáticamente, eso implica tomar la primera derivada parcial de la variable en cuestión. La sencilla fórmula que da cuenta de esta relación se halla reproducida en la Ecuación 9.

Ecuación 9: parámetro βj

$$\beta_j = \frac{\Delta y}{\Delta x_j}$$

Donde:
β_j es el parámetro de la variable explicativa en cuestión
Δy es el cambio en la variable dependiente
Δx_j es el cambio en la variable explicativa en cuestión

IV.2 Conceptos estadísticos relevantes

Otros conceptos relevantes que deben ser explicados, porque se recurre a ellos a la hora de interpretar los resultados de los análisis de regresión, se detallan a continuación. Comprender cabalmente estos conceptos es imprescindible para a la vez analizar correctamente el *output* del código en R, lo que permitirá dar una respuesta argumentada y sólida a las preguntas de investigación.

IV.2.1 Bondad de ajuste y coeficiente de determinación

Un modelo de regresión lineal relaciona una variable dependiente con una o más variables independientes. Esta relación se estima por medio de una ecuación de regresión que arroja un intercepto y los parámetros que se corresponden con cada variable dependiente y determinan la pendiente de la función.

Sin embargo, en sí misma esta ecuación no revela cuán fuerte es el poder explicativo de la(s) variable(s) independiente(s). En otras palabras, es preciso un indicador que de cuenta de la bondad de ajuste[9] de la línea de regresión calculada con el método de MCO. Ese indicador se conoce como coeficiente de determinación, R cuadrada o R^2, y es justamente la proporción de la variación en la variable dependiente que es explicada por la variable independiente. La fórmula para calcularla se comparte en la Ecuación 10.

Ecuación 10: coeficiente de determinación

$$R^2 \equiv \frac{SEC}{STC} = 1 - \frac{SRC}{STC}$$

Donde:
R^2 es el coeficiente de determinación
SEC es la suma explicada de cuadrados
STC es la suma total de cuadrados
SRC es la suma residual de cuadrados

Como aclaración, la SEC, STC y SRC están vinculadas al hecho de que el método de MCO se fundamenta en separar cada observación en la muestra (cada y_i) en un valor ajustado y un valor residual, que se supone no se encuentran correlacionados. Mientras que la SEC o variación explicada es una medición de la variación de los valores ajustados $\hat{y}_i$, la SRC o variación no explicada mide la variación de los términos de error. La Ecuación 11 demuestra que la STC no es más que la suma de ambas.

[9] *Goodness of fit* en inglés.

Ecuación 11: suma de totales cuadrados

$$STC = SEC + SRC$$

Por lo tanto, el coeficiente de determinación no es sino la ratio de la variación explicada sobre la variación total, de allí que se lo entienda como la proporción de la variación muestral en *y* que es explicada por *x* (Wooldridge, 2015). La R cuadrada tiene las mismas propiedades tanto en el modelo de regresión lineal simple como en el múltiple. Sin embargo, debe tenerse en cuenta que al agregarse más variables independientes, el R^2 necesariamente se incrementa.

La R^2 puede tomar valores entre 0 y 1, generalmente expresados de manera porcentual. Cuanto mayor es el coeficiente de determinación, es decir, cuanto más cercano a 1, mayor es la proporción de la variación en la variable dependiente que es explicada por la variable independiente. Una R cuadrada de 1 implica una correlación perfecta y una R cuadrada de 0 implica la ausencia de toda correlación.

IV.2.2 R cuadrada ajustada

Teniendo en cuenta que de entre los análisis de regresión que se exponen en esta tesis, uno de ellos es múltiple, se torna conveniente introducir el concepto de R cuadrada ajustada o $\bar{R}^2$. La R cuadrada ajustada se calcula de la manera que se muestra en la Ecuación 12.

Ecuación 12: R cuadrada ajustada

$$\bar{R}^2 = 1 - \frac{SRC/(n-k-1)}{STC/(n-1)}$$

Donde:
$\bar{R}^2$ es la R cuadrada ajustada
STC es la suma total de cuadrados
SRC es la suma residual de cuadrados
n es el número de observaciones contenidas en la muestra
k es la cantidad de variables explicativas

Como puede apreciarse en la fórmula, la R cuadrada ajustada implica una penalización derivada de agregar más variables independientes al modelo. Esto la diferencia del coeficiente de determinación, ya que la R^2 no disminuye cuando se adicionan nuevas variables explicativas. Esto no significa que automáticamente la R cuadrada ajustada disminuya sólo por adicionar nuevas variables.

De hecho, el valor numérico de esta propiedad estadística, que frecuentemente es positivo pero puede ser negativo, puede tanto aumentar como disminuir si se agregan más variables al modelo. Eso sí, la R cuadrada ajustada siempre será menor a la R cuadrada, lo cual es manifiesto al compararse ambas fórmulas, y se debe justamente a la inclusión de forma sustractiva de las variables explicativas *k* en el numerador.

El recurso de la R cuadrada ajustada es útil porque una tentación que existe para dotar de mayor poder explicativo a un modelo es añadir más variables explicativas, bajo el supuesto de que un incremento mecánico en la R cuadrada es de por sí mejor. El riesgo que esto entraña es lo que se denomina sobreajuste[10]. Para enfrentar ese riesgo es que se apela a la R cuadrada ajustada, que delimita el nivel de confianza que puede depositarse en la correlación, y la proporción del coeficiente de determinación que es producto de la incorporación de nuevas variables explicativas.

IV.2.3 Error estándar

El error estándar (EE) de un coeficiente de regresión no es sino una estimación de la desviación estándar del coeficiente. A su vez, la desviación estándar es una medida

[10] *Overfitting* en inglés.

de la dispersión de los valores que adopta una variable aleatoria. Es una manera de calibrar cuán lejos se hallan los valores de un conjunto de su promedio. Matemáticamente, es igual a la raíz cuadrada de la varianza. Su fórmula es la de la Ecuación 13.

Ecuación 13: desviación estándar

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}}$$

Donde:
s es la desviación estándar de la muestra
x_i es el valor de la observación *i*
$\bar{x}$ es el promedio de todas las observaciones
n es el número de observaciones

El EE se computa como la ratio de la desviación estándar sobre el cuadrado del número de observaciones, de acuerdo con la exposición de la Ecuación 14.

Ecuación 14: error estándar

$$EE = \frac{s}{\sqrt{n}}$$

Donde:
EE es el error estándar de la muestra
s es la desviación estándar de la muestra
n es el número de observaciones

El error estándar expresa cuán incierto es el coeficiente de regresión. Es útil para crear intervalos de confianza y para inferir el valor t (que se explica a continuación). También constituye una aproximación al significado estadístico del modelo.

IV.2.4 Valor t

El valor t o estadístico t es lisa y llanamente el resultado de dividir el coeficiente de la variable por su error estándar. Se emplea para poner a prueba hipótesis relativas a los parámetros de un modelo de regresión. Al tratarse de una ratio, su valor numérico será mayor cuanto mayor sea el coeficiente y menor sea el error estándar. En consecuencia, es deseable desde el punto de vista estadístico que el estimador de una variable tenga un valor t alto, porque es una señal de que el error estándar es bajo en relación al coeficiente. Si el estadístico t es alto, eso provee mayor certeza de que el coeficiente no es cero. El valor t indica cuántos errores estándar se halla alejada la variable independiente desde cero. Además, juega un papel en el cálculo del valor p.

IV.2.5 Valor p

El valor p se computa recurriendo al estadístico t bajo una distribución *t-Student* (un tipo de probabilidad de distribución para los valores de una población). El valor p es una herramienta que permite conjeturar cuán significativos son los coeficientes de un modelo. Generalmente, se considera que el valor p es significativo si se encuentra por debajo de 0.05. Si el valor p es significativo, esto conduce a deducir con un alto grado de certidumbre que el coeficiente es distinto de cero, y por lo tanto la variable independiente detenta un grado de poder explicativo respecto a los cambios en la variable dependiente.

IV.2.6 Estadístico F

Además de evaluar hipótesis acerca de los parámetros de las variables individuales, también es posible realizar una prueba de hipótesis para el modelo de regresión en su conjunto. La hipótesis nula enuncia que no existe entre la variable de respuesta y la(s) variable(s) explicativa(s). La hipótesis alternativa, por el contrario, afirma que sí existe una relación. Naturalmente, la expectativa que subyace al definir un modelo de regresión es que la hipótesis nula planteada de esta manera podrá ser rechazada con cierto grado de certeza.

En términos numéricos, la hipótesis nula sostiene que todos los coeficientes de las variables independientes (o el coeficiente de la única variable explicativa si se trata de un modelo simple) son iguales a cero. La hipótesis alternativa manifiesta que al menos uno de estos coeficientes es distinto de cero. El estadístico F justo a su correspondiente valor p contribuye a precisar los resultados de esta prueba. Cuanto mayor es el valor del estadístico F, mayor es la capacidad explicativa del modelo.

No obstante, es necesaria una precaución al trabajar con modelos de regresión múltiples. Si se agregan demasiadas variables, podría arribarse a un estadístico F muy bajo junto a un valor p que habilitaría a rechazar la hipótesis nula, lo que sería engañoso. Afortunadamente, esto no aplica al único modelo de regresión múltiple que aquí se desarrolla, ya que cuenta con sólo dos variables independientes. En este caso, si el estadístico F es alto, es un indicio apropiado de que se puede rechazar la hipótesis nula.

V. Análisis de datos y resultados

Esta sección concierne directamente al análisis de la *raw data*, su procesamiento, y la generación de conclusiones. Como ya se explicó previamente, el método escogido es el de análisis de regresión lineal. Se implementan cuatro modelos. Tres de ellos corresponden a la primera dupla de pregunta e hipótesis específica, siendo dos de tipo simple y uno de tipo múltiple, mientras que el cuarto, de carácter simple, refiere a la segunda dupla de pregunta e hipótesis específica.

La información se recabó de fuentes públicas del gobierno de EEUU. Más concretamente, del *Federal Reserve Bank of St. Louis*. Un primer filtrado se realiza con aplicaciones convencionales (Microsoft Excel). Los análisis estadísticos y el procesamiento de la información se ejecutan con lenguaje de programación R. El entorno de programación empleado es RStudio.

V.1 Información original

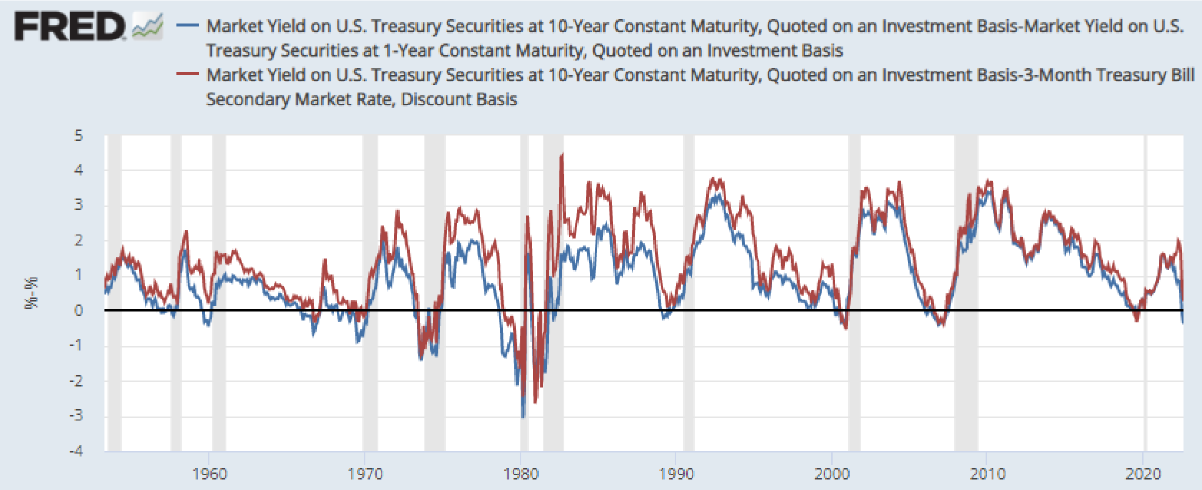
El primer paso para poder efectuar el análisis de regresión estriba en reunir y ordenar la información apropiada. Esto requiere efectuar la búsqueda preferentemente en bases de datos oficiales. Las fuentes aquí consultadas son todas gubernamentales, e incluyen al NBER, la Reserva Federal y principalmente el *Federal Reserve Bank of St. Louis*. De este último es que se extraen las estadísticas y los datos de mayor importancia.

El primero de ellos refiere al *spread* entre bonos de corto y largo plazo (Federal Reserve Bank of St. Louis, 2018). Esta fuente ofrece una comparación entre dos *spreads*. Por un lado, entre bonos del Tesoro (*treasuries*) a 10 años y a un año, etiquetados como T10Y1Y. Por el otro, entre *treasuries* a 10 años y a 3 meses[11],

[11] A modo de aclaración, *treasuries* denomina genéricamente a todos los bonos emitidos por el Departamento del Tesoro de EEUU. Pero el nombre varía de acuerdo con el horizonte temporal del bono. Los instrumentos de corto plazo, no mayores a un año, son las llamadas *T-bills* o *Treasury bills*.

rotulados como T10Y3M. La gráfica de ambos spreads ya fue presentada en la Ilustración 5.

Ilustración 5: inversiones de la yield curve y recesiones en EEUU tras la SGM



Fuente: Board of Governors of the Federal Reserve System

En primera instancia se nota que ambas curvas están íntimamente correlacionadas. Su forma y evolución son prácticamente idénticas, sin que exista un *lag* en su comportamiento (ambas se mueven simultáneamente). Sí es cierto que los *spreads* entre ambas curvas, que en realidad manifiestan el diferencial entre bonos a un año y *T-bills* a tres meses (ya que ambas toman como base para el cálculo las *T-notes* a 10 años), no son constantes. En algunos casos se estrechan, hasta coincidir (es decir, el *spread* es cero), mientras que en otros se separan de manera considerable. Pero la curva T10Y3M siempre está "por arriba" de la T10Y1Y. Es decir, es más acentuada o *steep*.

Esto es perfectamente lógico, ya que un bono a tres meses es un instrumento de muy corto plazo, por lo que es razonable que su tasa sea bastante reducida. Subsecuentemente, es esperable que el *gap* en T10Y3M supere al de T10Y1Y. Por estas razones, no siempre que la T10Y1Y pasa a tener un diferencial negativo, la T10Y3M también lo hace. Por lo tanto, históricamente la T10Y3M se ha invertido en un número menor de ocasiones que la T10Y1Y.

Los bonos de mediano plazo, entre dos y diez años, son designados como *T-notes* o *Treasury notes*. Por último, los bonos de largo plazo, de entre veinte y treinta años, son apodados como *T-bonds o Treasury bonds*.

Ese es un primer motivo para preferir la T10Y1Y a la T10Y3M para efectuar el análisis recurriendo a la primera de estas curvas. Ya de por sí los eventos de inversión de la curva no son tan numerosos desde un punto de vista estadístico. Si la muestra se redujese aún más, cualquier conclusión que se intentase extraer perdería demasiado valor.

El segundo motivo es que las inversiones de la curva y las recesiones están relacionadas de manera asincrónica. Este efecto temporal demorado, o *lag*, ha sido en la práctica de unos trece meses promedio. Es decir, en EEUU, cada vez que la *yield curve* se ha invertido, la subsecuente recesión ha irrumpido trece meses más tarde, o poco más de un año. Es decir que los agentes prevén (nuevamente, en promedio) que una recesión acaecerá aproximadamente un año después de que la curva se ve trastocada. Esto refuerza el argumento a favor de emplear los *treasuries* a un año como indicador de instrumentos de deuda de corto plazo.

La base de datos del *Federal Reserve Bank of St. Louis* contiene un total de 832 puntos de dato tanto para la T10Y1Y (referenciada en la página oficial como GS10_GS1) como para la T10Y3M (referenciada en la página oficial como GS10_TB3MS). Es decir, un total de un mil seiscientos sesenta y cuatro observaciones. Las mismas tienen frecuencia mensual, comenzando en abril de 1953 y llegando hasta julio de 2022.

La T10Y1Y se invierte un total de ciento treinta y nueve veces en dicho período, mientras que la T10Y3M lo hace en setenta y dos ocasiones. En sesenta y ocho oportunidades ambas curvas están invertidas al mismo tiempo, por lo que sólo en cuatro meses a lo largo de los sesenta y nueve años considerados la T10Y3M sí se invierte mientras que la T10Y1Y no lo hace[12], pero lo opuesto acontece setenta y un veces. La Tabla 1 contiene la información de todos los meses en que la T10Y1Y exhibe un *spread* negativo.

[12] Puntualmente, en febrero de 1974, junio de 2019, julio de 2019 y febrero de 2020. Es decir que es un fenómeno que además de poco frecuente, es muy reciente.

Tabla 1: spreads negativos de la T10Y1Y

Mes - Año	**T10Y1Y**	**T10Y3M**
dic.-56	-0.09%	0.38%
feb.-57	-0.04%	0.24%
mar.-57	-0.01%	0.33%
abr.-57	-0.01%	0.41%
ago.-57	-0.08%	0.56%
sep.-57	-0.15%	0.39%
oct.-57	-0.04%	0.39%
sep.-59	-0.32%	0.64%
oct.-59	-0.27%	0.48%
nov.-59	-0.28%	0.38%
dic.-59	-0.45%	0.20%
ene.-60	-0.31%	0.37%
feb.-60	-0.17%	0.53%
dic.-65	-0.10%	0.24%
ene.-66	-0.27%	0.02%
feb.-66	-0.11%	0.18%
mar.-66	-0.10%	0.28%
abr.-66	-0.15%	0.13%
may.-66	-0.15%	0.14%
jun.-66	-0.16%	0.31%
jul.-66	-0.15%	0.22%
ago.-66	-0.32%	0.26%
sep.-66	-0.64%	-0.19%
oct.-66	-0.57%	-0.34%
nov.-66	-0.38%	-0.16%
dic.-66	-0.36%	-0.12%
ene.-67	-0.17%	-0.14%
feb.-67	-0.08%	0.07%
dic.-67	-0.01%	0.73%
abr.-68	-0.07%	0.26%

Mes - Año	T10Y1Y	T10Y3M
may.-68	-0.27%	0.21%
jun.-68	-0.26%	0.20%
jul.-68	-0.15%	0.19%
ago.-68	-0.01%	0.33%
nov.-68	-0.05%	0.25%
dic.-68	-0.16%	0.07%
ene.-69	-0.30%	-0.10%
feb.-69	-0.22%	0.07%
mar.-69	-0.04%	0.28%
abr.-69	-0.09%	0.06%
may.-69	-0.10%	0.28%
jun.-69	-0.47%	0.13%
jul.-69	-0.88%	-0.28%
ago.-69	-0.85%	-0.29%
sep.-69	-0.66%	0.07%
oct.-69	-0.54%	0.10%
nov.-69	-0.75%	-0.10%
dic.-69	-0.52%	-0.17%
ene.-70	-0.31%	-0.08%
feb.-70	-0.35%	0.11%
mar.-73	-0.14%	0.62%
abr.-73	-0.18%	0.41%
may.-73	-0.04%	0.49%
jun.-73	-0.41%	-0.29%
jul.-73	-1.26%	-0.88%
ago.-73	-1.42%	-1.27%
sep.-73	-1.22%	-1.20%
oct.-73	-0.61%	-0.43%
nov.-73	-0.84%	-1.10%
dic.-73	-0.53%	-0.71%
ene.-74	-0.43%	-0.78%

Mes - Año	T10Y1Y	T10Y3M
mar.-74	-0.55%	-0.75%
abr.-74	-1.11%	-0.82%
may.-74	-1.20%	-0.65%
jun.-74	-1.13%	-0.36%
jul.-74	-0.99%	0.26%
ago.-74	-1.32%	-0.92%
sep.-74	-0.83%	-0.02%
oct.-74	-0.15%	0.44%
sep.-78	-0.22%	0.57%
oct.-78	-0.50%	0.65%
nov.-78	-1.20%	0.17%
dic.-78	-1.29%	-0.07%
ene.-79	-1.31%	-0.25%
feb.-79	-1.14%	-0.22%
mar.-79	-1.13%	-0.36%
abr.-79	-0.94%	-0.28%
may.-79	-0.87%	-0.36%
jun.-79	-0.66%	-0.15%
jul.-79	-0.69%	-0.29%
ago.-79	-0.95%	-0.49%
sep.-79	-1.51%	-0.93%
oct.-79	-2.14%	-1.40%
nov.-79	-1.74%	-1.14%
dic.-79	-1.59%	-1.65%
ene.-80	-1.26%	-1.20%
feb.-80	-1.51%	-0.45%
mar.-80	-3.07%	-2.45%
abr.-80	-1.83%	-1.73%
sep.-80	-0.01%	1.24%
oct.-80	-0.74%	0.13%
nov.-80	-1.47%	-1.05%

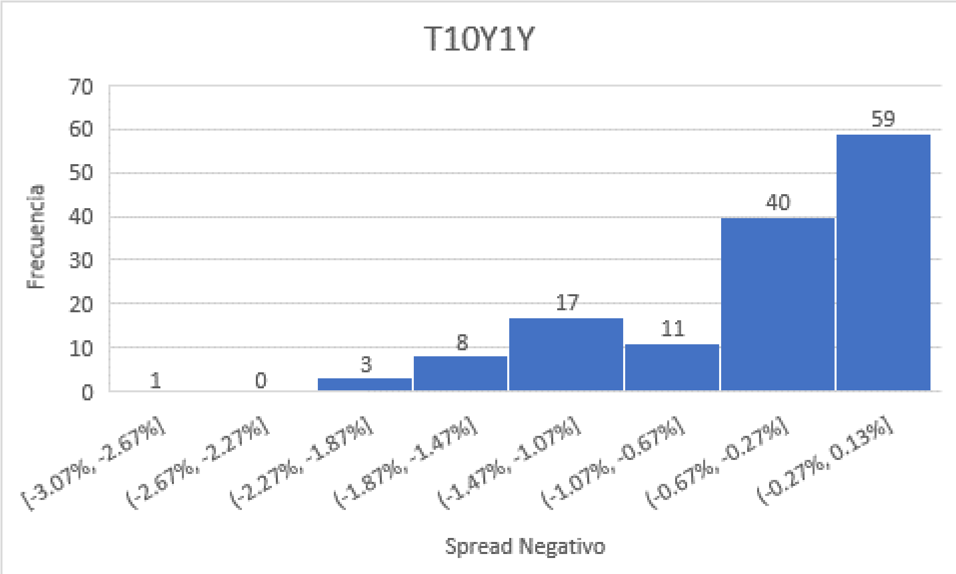
Mes - Año	T10Y1Y	T10Y3M
dic.-80	-2.04%	-2.65%
ene.-81	-1.51%	-2.45%
feb.-81	-1.38%	-1.60%
mar.-81	-0.59%	-0.24%
abr.-81	-0.64%	-0.01%
may.-81	-2.10%	-2.20%
jun.-81	-1.39%	-1.26%
jul.-81	-1.44%	-0.67%
ago.-81	-1.78%	-0.57%
sep.-81	-1.20%	0.62%
oct.-81	-0.23%	1.61%
feb.-82	-0.30%	0.95%
mar.-82	-0.09%	1.18%
abr.-82	-0.11%	1.17%
feb.-89	-0.08%	0.64%
mar.-89	-0.21%	0.54%
abr.-89	-0.18%	0.53%
may.-89	-0.12%	0.43%
jun.-89	-0.16%	0.13%
ago.-89	-0.07%	0.21%
sep.-89	-0.03%	0.44%
abr.-00	-0.16%	0.33%
jun.-00	-0.07%	0.41%
jul.-00	-0.03%	0.09%
ago.-00	-0.35%	-0.26%
sep.-00	-0.33%	-0.20%
oct.-00	-0.27%	-0.37%
nov.-00	-0.37%	-0.45%
dic.-00	-0.36%	-0.53%
ene.-06	-0.03%	0.18%
feb.-06	-0.11%	0.14%

Mes - Año	T10Y1Y	T10Y3M
mar.-06	-0.05%	0.21%
jun.-06	-0.05%	0.32%
jul.-06	-0.13%	0.14%
ago.-06	-0.20%	-0.08%
sep.-06	-0.25%	-0.09%
oct.-06	-0.28%	-0.19%
nov.-06	-0.41%	-0.34%
dic.-06	-0.38%	-0.29%
ene.-07	-0.30%	-0.22%
feb.-07	-0.33%	-0.31%
mar.-07	-0.36%	-0.38%
abr.-07	-0.24%	-0.18%
may.-07	-0.16%	0.02%
ago.-19	-0.14%	-0.32%
sep.-19	-0.10%	-0.19%
jul.-22	-0.12%	0.67%

Fuente: Board of Governors of the Federal Reserve System

Con el fin de obtener una mirada más global sobre cuán profundas han sido las inversiones de la curva T10Y1Y, puede observarse el histograma de la Ilustración 6. Allí puede apreciarse que la mayoría de los *gaps* son más bien reducidos. Un 71% del total (noventa y nueve meses) han sido menores (en términos absolutos) al -0,67%.

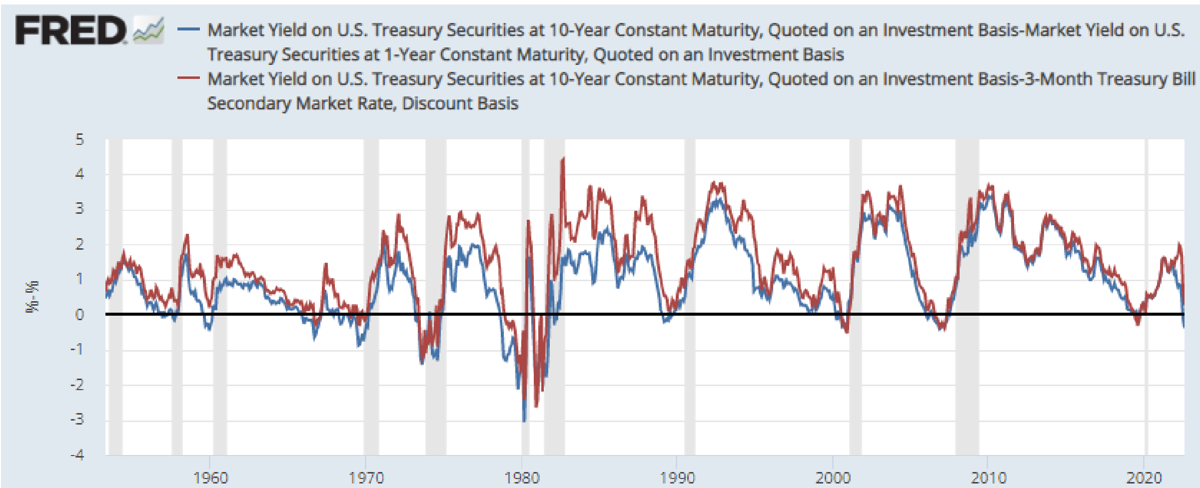
Ilustración 6: histograma T10Y1Y (a)



Fuente: Board of Governors of the Federal Reserve System

Por otra parte, si se vuelve a la Ilustración 5, puede advertirse que, por una gran diferencia, los *spreads* más pronunciados se dieron en la “doble inversión” de 1980-1982. No sólo eso, todas las inversiones de la curva hasta principios de la década de 1980 fueron más acentuadas que las que ocurren en los últimos cuarenta años. Otro fenómeno curioso es que las inversiones de la curva, así como sus subsecuentes recesiones, se tornan más esporádicas.

Ilustración 5: inversiones de la yield curve y recesiones en EEUU tras la SGM



Fuente: Board of Governors of the Federal Reserve System

Una vez efectuado este filtrado inicial de los datos relativos a los *spreads* entre bonos de corto y largo plazo, que suministra la información concerniente a la variable independiente, debe realizarse la misma operación con los datos pertinentes a la variable dependiente. Nuevamente recurriendo a la base de datos del *Federal Reserve Bank of St. Louis*, se hallan disponibles trescientos un puntos de dato sobre la tasa de crecimiento del PBI de la economía norteamericana, desde 1947 hasta 2022.

Las observaciones tienen una periodicidad trimestral[13]. Si se consideran sólo los trimestres en los que la economía cae, la muestra se reduce a cuarenta y cinco observaciones, que abarcan el mismo período, desde el segundo trimestre de 1947 hasta el segundo trimestre de 2022. La Tabla 2 muestra todos los trimestres en que la economía estadounidense padece crecimiento negativo.

Tabla 2: trimestres con crecimiento negativo en la economía de EEUU

Trimestre - Año	**Tasa PBI %**
Trim. 2 - 1947	-1.10%
Trim. 3 - 1947	-0.80%
Trim. 1 - 1949	-5.40%
Trim. 2 - 1949	-1.40%
Trim. 4 - 1949	-3.30%
Trim. 3 - 1953	-2.20%
Trim. 4 - 1953	-5.90%
Trim. 1 - 1954	-1.90%
Trim. 1 - 1956	-1.50%
Trim. 3 - 1956	-0.40%
Trim. 2 - 1957	-0.90%
Trim. 4 - 1957	-4.10%
Trim. 1 - 1958	-10.00%

[13] Antes de esa fecha, es decir, desde que se comienza a medir la actividad económica por medio del concepto de PBI (durante la década de 1930, en el marco de la Gran Depresión y a propuesta del economista Simon Kuznets) y hasta el fin de la SGM, los datos sobre el PBI tienen una frecuencia menor, anual.

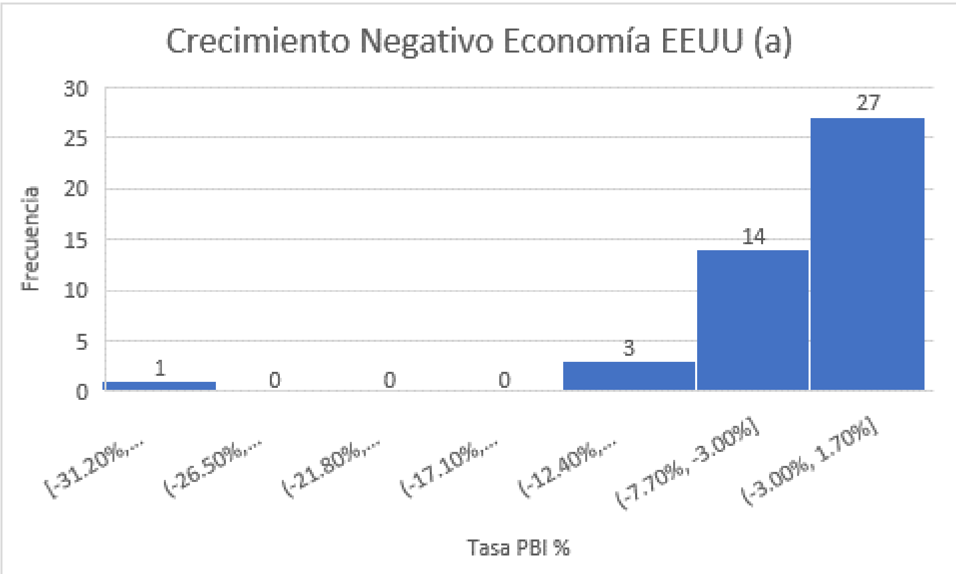
Trimestre - Año	**Tasa PBI %**
Trim. 2 - 1960	-2.10%
Trim. 4 - 1960	-5.00%
Trim. 4 - 1969	-1.90%
Trim. 1 - 1970	-0.60%
Trim. 4 - 1970	-4.20%
Trim. 3 - 1973	-2.10%
Trim. 1 - 1974	-3.40%
Trim. 3 - 1974	-3.70%
Trim. 4 - 1974	-1.50%
Trim. 1 - 1975	-4.80%
Trim. 2 - 1980	-8.00%
Trim. 3 - 1980	-0.50%
Trim. 2 - 1981	-2.90%
Trim. 4 - 1981	-4.30%
Trim. 1 - 1982	-6.10%
Trim. 3 - 1982	-1.50%
Trim. 4 - 1990	-3.60%
Trim. 1 - 1991	-1.90%
Trim. 1 - 2001	-1.30%
Trim. 3 - 2001	-1.60%
Trim. 1 - 2008	-1.60%
Trim. 3 - 2008	-2.10%
Trim. 4 - 2008	-8.50%
Trim. 1 - 2009	-4.60%
Trim. 2 - 2009	-0.70%
Trim. 1 - 2011	-1.00%
Trim. 3 - 2011	-0.20%
Trim. 1 - 2014	-1.40%
Trim. 1 - 2020	-5.10%
Trim. 2 - 2020	-31.20%
Trim. 1 - 2022	-1.60%

Trimestre - Año	Tasa PBI %
Trim. 2 - 2022	-0.60%

Fuente: Board of Governors of the Federal Reserve System

Si se grafican en un histograma todos los trimestres en que la economía de EEUU decreció, nuevamente se observa que la distribución está sesgada hacia la izquierda. Es decir, en términos relativos, las recesiones tienden a no ser tan graves. No obstante, debe tenerse en cuenta que los intervalos son aquí distintos de los usados en la Ilustración 6. Mientras en el histograma de los spreads el rango era de 0,4%, aquí es casi doce veces mayor, de 4,7%.

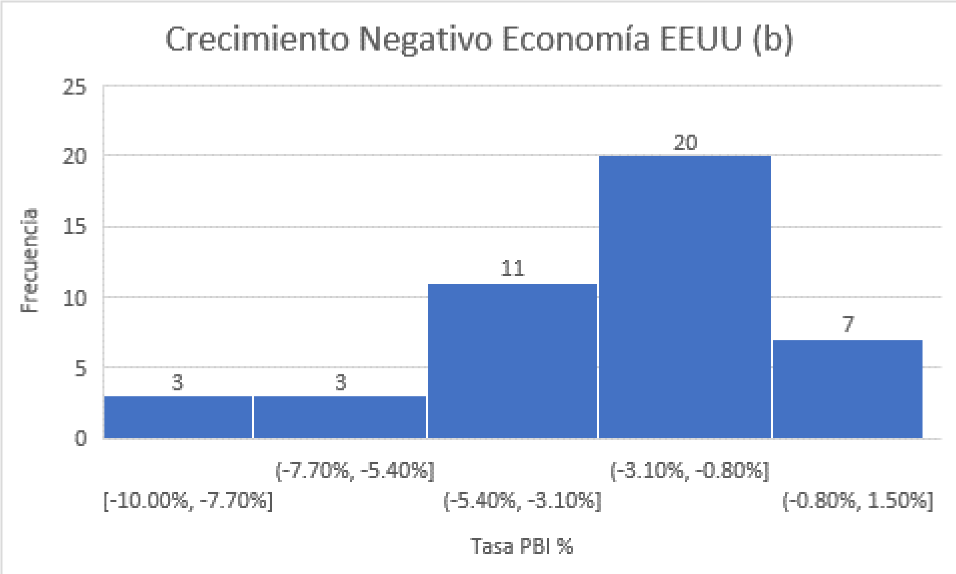
Ilustración 7: histograma crecimiento negativo PBI EEUU (a)



Fuente: Board of Governors of the Federal Reserve System

Además, la observación que se encuentra más a la izquierda (correspondiente a la recesión causada por la pandemia de Covid-19) altera notablemente la distribución. Si se decide excluir este evento, se obtiene una nueva distribución que, aunque continúa indudablemente sesgada hacia la izquierda, está algo más normalizada. Los intervalos son aquí más reducidos, de 2,3%, tal como exhibe la Ilustración 8.

Ilustración 8: histograma crecimiento negativo PBI EEUU (b)



Fuente: Board of Governors of the Federal Reserve System

El siguiente paso consiste en agrupar los datos en función de las recesiones identificadas por *Business Cycle Dating Committee* de la NBER. Como ya se mencionó antes, el criterio estándar que se emplea en EEUU (y muchos otros países) para determinar la aparición de una recesión es cuando el PBI registra dos trimestres consecutivos de caída.

Sin embargo, esta no es la única variable que las autoridades tienen en cuenta, y la decisión oficial respecto a qué ciclos han sido de expansión y cuáles de contracción es un proceso complejo. De acuerdo a la periodización gubernamental, EEUU ha registrado doce recesiones desde el fin de la SGM, tal como se puede ver en la Tabla 3.

Tabla 3: recesiones en EEUU tras la SGM

Inicio	**Fin**	**Duración**
Trim. 1 - 1949	Trim. 4 - 1949	4 trimestres
Trim. 3 - 1953	Trim. 2 - 1954	4 trimestres
Trim. 4 - 1957	Trim. 2 - 1958	3 trimestres
Trim. 2 - 1960	Trim. 1 - 1961	4 trimestres

Inicio	**Fin**	**Duración**
Trim. 1 - 1970	Trim. 4 - 1970	4 trimestres
Trim. 1 - 1974	Trim. 1 - 1975	5 trimestres
Trim. 2 - 1980	Trim. 3 - 1980	2 trimestres
Trim. 4 - 1981	Trim. 4 - 1982	5 trimestres
Trim. 3 - 1990	Trim. 1 - 1991	3 trimestres
Trim. 2 - 2001	Trim. 4 - 2001	3 trimestres
Trim. 1 - 2008	Trim. 2 - 2009	6 trimestres
Trim. 1 - 2020	Trim. 2 - 2020	2 trimestres

Fuente: Board of Governors of the Federal Reserve System

De estas doce recesiones, las dos primeras, la de 1949 y la de 1953-1954, no están relacionadas con una inversión de la curva de rendimientos. La primera, porque se carece de información, y la segunda, porque ciertamente no se registró un *spread* negativo. Esto deja una muestra con diez observaciones, justo lo suficiente para poder realizar un análisis de regresión. A fin de evitar confusiones, estos diez acontecimientos serán identificados como:

- Recesión de 1957-1958
- Recesión de 1960-1961
- Recesión de 1970
- Recesión de 1974-1975
- Recesión de 1980
- Recesión de 1981-1982
- Recesión de 1990-1991
- Recesión de 2001
- Recesión de 2008-2009
- Recesión de 2020

En lo que respecta a la curva de rendimientos, la misma se ha invertido también en doce oportunidades, a partir de abril de 1953, según se observa en la Tabla 4.

Tabla 4: inversiones de la yield curve en EEUU tras la SGM

Inicio	**Fin**	**Duración**
dic.-56	oct.-57	7 meses
sep.-59	feb.-60	6 meses
dic.-65	feb.-67	15 meses
abr.-68	feb.-70	22 meses
mar.-73	oct.-74	19 meses
sep.-78	abr.-80	20 meses
sep.-80	abr.-82	17 meses
feb.-89	sep.-89	7 meses
abr.-00	dic.-00	8 meses
ene.-06	may.-07	15 meses
ago.-19	sep.-19	2 meses
jul.-22	jul.-22[14]	1 mes

Fuente: Board of Governors of the Federal Reserve System

De estos doce sucesos en los que la curva se invirtió, dos deben ser excluidos. El primero es la inversión de 1965-1967, que no se vio seguida de una recesión. Es el único caso que contradice la regla. Por tener implicancias para la validez del análisis aquí llevado a cabo, este hecho va a ser mencionado en las conclusiones.

La otra inversión que no debe ser tenida en cuenta es la más reciente, que comienza en julio de 2022. En este caso, se trata de un acontecimiento en curso al momento de escribir esta tesis. Por otra parte, la eventual recesión que le seguiría obviamente aún no ha ocurrido por tratarse de un evento futuro.

Esto deja un total de diez inversiones de la curva en el período considerado, exactamente igual al número de recesiones estudiadas. Para eludir malentendidos, estas diez inversiones de la curva serán designadas como:

- Inversión de 1956-1957

[14] La información más reciente sólo llega hasta julio de 2022.

- Inversión de 1959-1960
- Inversión de 1968-1970
- Inversión de 1973-1974
- Inversión de 1978-1980
- Inversión de 1980-1982
- Inversión de 1989
- Inversión de 2000
- Inversión de 2006-2007
- Inversión de 2019

Por último, tanto las inversiones de la curva como las recesiones deben ser enlazadas respetando el orden cronológico, lo que se expone en la Tabla 5.

Tabla 5: años de inversiones de la yield curve y recesiones en EEUU tras la SGM

Inversión de la Curva de Rendimientos	**Recesión**
1956-1957	1957-1958
1959-1960	1960-1961
1968-1970	1970
1973-1974	1974-1975
1978-1980	1980
1980-1982	1981-1982
1989	1990-1991
2000	2001
2006-2007	2008-2009
2019	2020

Fuente: Board of Governors of the Federal Reserve System

La última fase del procesamiento de la información original estriba en definir cuáles datos van a ser finalmente empleados como *input* en los análisis de regresión. Inmediatamente aparecen dificultades, las cuales sin embargo pueden ser superadas.

El que tanto la tasa de variación del PBI como el *spread* de tasas de interés se encuentren expresados en porcentajes, no entraña mayor problema. Esto es así porque ambas variables están expresadas en los mismos términos. Si ese no fuese el caso, y una de las variables estuviese en valores porcentuales mientras la otra no, sería necesario realizar ciertas operaciones en los datos, por ejemplo, una transformación logarítmica. Pero como ambas variables son porcentajes, el modelo de regresión lineal no se ve afectado.

Sí hay una dificultad por el lado de qué valor final se termina tomando. Las tasas de variación del PBI son interanuales, es decir, comparan trimestres respectivos entre años consecutivos (por ejemplo, el tercer trimestre de 1963 contra el tercer trimestre de 1962). Dentro de un un período identificado como recesión, los porcentajes de caída del producto no son de un trimestre a otro, ni tampoco acumulados dentro de la fase recesiva (es decir, desde que se entra hasta que se sale de la recesión).

Por lo tanto, debe tomarse un valor estadístico único para cada recesión. La mínima y la mediana deber ser desechadas. Esto es así porque dentro de una recesión existen fluctuaciones. Es decir, más allá de que una contracción "toca piso" al llegar a lo que se llama *trough*, si se piensa a la curva de la tasa de variación del PBI dentro de los límites de una recesión, pueden existir varios "mínimos locales" (el piso o *trough* es el mínimo global de la recesión). En consecuencia, la única propiedad estadística remanente que puede tomarse para reseñar un valor único de caída del PBI para cada recesión es el promedio.

Una lógica similar se aplica a las inversiones de la curva. Dentro de cada período, aunque el spread permanece negativo, su valor numérico oscila, acercándose o alejándose del punto de paridad (una vez alcanzado y atravesado el punto de paridad de tasas, el spread se torna positivo nuevamente). Por lo tanto, ni el mínimo ni la mediana pueden ser juzgados como indicadores apropiados. Esto conlleva que también deba calcularse el promedio.

Las tasas de caída promedio del PBI y el nivel promedio del spread negativo para cada una de las diez recesiones y las diez inversiones de curva identificadas se muestran en la Tabla 6. Estos serán los datos que se introducirán como *input* en el análisis de regresión que buscará dar respuesta a la primera pregunta de investigación.

Tabla 6: inversiones de la yield curve y recesiones en EEUU tras la SGM

Inversión de la Curva de Rendimientos	***Spread* Negativo %**	**Recesión**	**Caída PBI %**
1956-1957	-0.06%	1957-1958	-7.05%
1959-1960	-0.30%	1960-1961	-1.70%
1968-1970	-0.32%	1970	-0.48%
1973-1974	-0.76%	1974-1975	-2.48%
1978-1980	-1.28%	1980	-4.25%
1980-1982	-1.00%	1981-1982	-2.53%
1989	-0.12%	1990-1991	-2.75%
2000	-0.24%	2001	-0.13%
2006-2007	-0.22%	2008-2009	-2.53%
2019	-0.12%	2020	-18.15%

Fuente: Board of Governors of the Federal Reserve System

Finalmente, para dar respuesta a la segunda hipótesis, deben demarcarse la duración tanto de las recesiones como de las inversiones de la curva. La traba aquí radica en que mientras las primeras se miden en trimestres, las segundas se cuentan en meses. La solución es sencilla: transformar los trimestres en meses. La Tabla 7 expone la duración en meses de las 10 inversiones de curva y las 10 recesiones evaluadas.

Tabla 7: inversiones de la yield curve y recesiones en EEUU tras la SGM en meses

Inversión de la Curva de Rendimientos	**Duración**	**Recesión**	**Duración**
1956-1957	7	1957-1958	6
1959-1960	6	1960-1961	9
1968-1970	22	1970	15
1973-1974	19	1974-1975	15
1978-1980	20	1980	6

Inversión de la Curva de Rendimientos	**Duración**	**Recesión**	**Duración**
1980-1982	17	1981-1982	12
1989	7	1990-1991	6
2000	8	2001	9
2006-2007	15	2008-2009	18
2019	2	2020	6

Fuente: Board of Governors of the Federal Reserve System

Ahora sí, se dispone de los datos requeridos para poder correr los análisis de regresión, los cuales se desarrollan en las siguientes subsecciones.

V.2 Primer análisis de regresión

V.2.1 Modelo de regresión lineal simple 1

Este primer análisis de regresión procura dar respuesta a la primera pregunta de investigación, formulada de la siguiente manera:

¿De qué manera repercute la magnitud de la inversión de la curva de rendimientos en la profundidad de la recesión que acontece con posterioridad a la aparición de dicha inversión?

La hipótesis que intenta responder de manera tentativa dicho interrogante es la siguiente:

Existe una correlación positiva y significativa entre la magnitud de la inversión de la curva de rendimientos, por un lado, y la profundidad de la recesión que acontece con posterioridad a la aparición de dicha inversión, por el otro.

En otras palabras, la hipótesis aquí planteada afirma que cuanto más negativo es el *spread* (cuánto mayor sea, en términos absolutos, el diferencial entre la tasa de bonos de corto y largo plazo), más profunda es la recesión que sobreviene con posterioridad (mayor es la caída en la producción). Para poner a prueba esta hipótesis, es que se realiza un análisis de regresión.

La variable independiente son los *spreads* negativos promedio para cada instancia de inversión de la curva, mientras que la variable dependiente son las variaciones negativas promedio del PBI para cada recesión posterior. La información ya fue presentada en la Tabla 6.

Tabla 6: inversiones de la yield curve y recesiones en EEUU tras la SGM

Inversión de la Curva de Rendimientos	***Spread* Negativo %**	**Recesión**	**Caída PBI %**
1956-1957	-0.06%	1957-1958	-7.05%
1959-1960	-0.30%	1960-1961	-1.70%
1968-1970	-0.32%	1970	-0.48%
1973-1974	-0.76%	1974-1975	-2.48%
1978-1980	-1.28%	1980	-4.25%
1980-1982	-1.00%	1981-1982	-2.53%
1989	-0.12%	1990-1991	-2.75%
2000	-0.24%	2001	-0.13%
2006-2007	-0.22%	2008-2009	-2.53%
2019	-0.12%	2020	-18.15%

Fuente: Board of Governors of the Federal Reserve System

Como este análisis de regresión contempla sólo una variable independiente, se trata de un modelo simple. La regresión busca establecer una relación lineal entre las variables. La Ecuación 1 introducida en la subsección IV.1 sintetiza la forma funcional de este tipo de regresión:

$$y = \beta_0 + \beta_1 x + u$$

Ya definido el tipo de regresión, la forma funcional, las variables, y contando con los datos necesarios, es momento de llevar a cabo la regresión. Operativamente, eso significa correr el código en lenguaje de programación R desarrollado explícitamente para ese fin.

En primer lugar debe cargarse el *data set* que contiene la información sobre los *spreads* entre los bonos de corto y largo plazo, así como las tasas de crecimiento (negativo) del PBI. Una vez que el *data set* aparece en el *environment* de RStudio, debe procederse a etiquetarse las columnas, que es como se ordenan las variables. El *output* del código se observa en la Ilustración 9.

Ilustración 9: valores del modelo de regresión lineal simple 1

```
> colnames(datosfed) <- c("pbi", "spread")
> datosfed
       pbi  spread
1  -0.0705 -0.0006
2  -0.0170 -0.0030
3  -0.0048 -0.0032
4  -0.0248 -0.0076
5  -0.0425 -0.0128
6  -0.0253 -0.0100
7  -0.0275 -0.0012
8  -0.0013 -0.0024
9  -0.0253 -0.0022
10 -0.1815 -0.0012
```

Código ejecutado en lenguaje de programación R

A continuación, pueden obtenerse las principales propiedades estadísticas tanto de la variable independiente como de la variable dependiente, incluyendo promedio, varianza y desvío estándar. El *output* se expone en la Ilustración 10.

Ilustración 10: propiedades estadísticas del modelo de regresión lineal simple 1 (a)

```
> mean(datosfed$pbi)
[1] -0.04205
> mean(datosfed$spread)
[1] -0.00442
> var(datosfed$pbi)
[1] 0.002777992
> var(datosfed$spread)
[1] 1.770178e-05
> sd(datosfed$pbi)
[1] 0.05270666
> sd(datosfed$spread)
[1] 0.004207348
```

Código ejecutado en lenguaje de programación R

La característica más sobresaliente es que la varianza de la variable dependiente es bastante mayor que la de la variable independiente. Esto debe ser tenido en cuenta a medida que se desarrolle el análisis, y se ve corroborado si se calcula otras propiedades estadísticas, como el mínimo, el máximo, la mediana y el rango intercuartil de ambas variables, tal como se muestra en la Ilustración 11.

Ilustración 11: propiedades estadísticas del modelo de regresión lineal simple 1 (b)

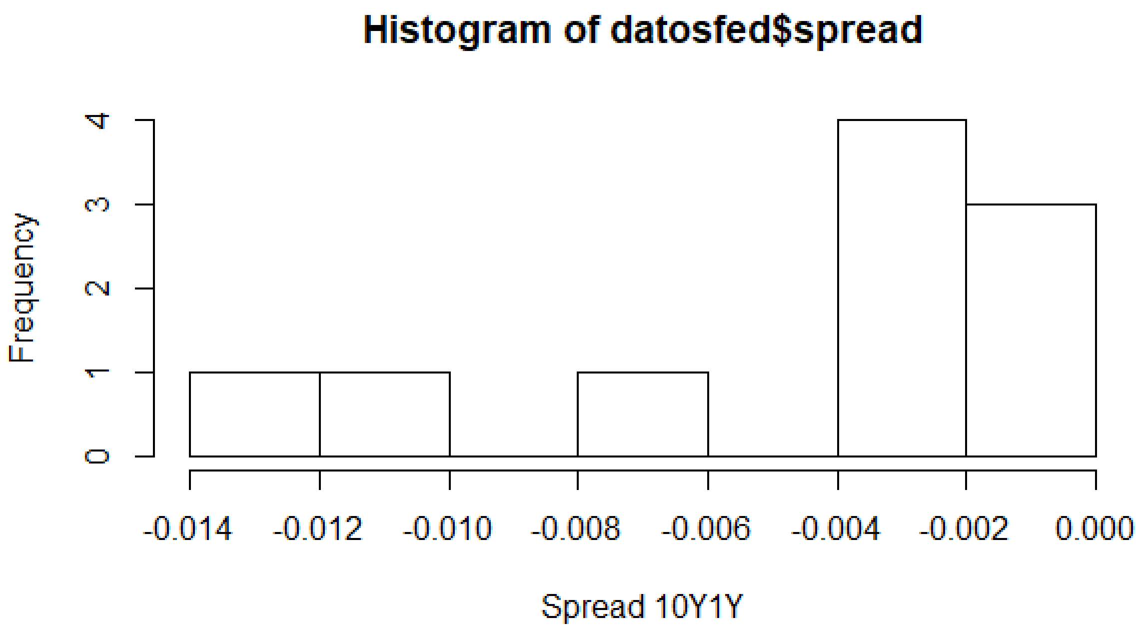
```
> summary(datosfed)
      pbi                 spread
 Min.   :-0.18150   Min.   :-0.01280
 1st Qu.:-0.03875   1st Qu.:-0.00650
 Median :-0.02530   Median :-0.00270
 Mean   :-0.04205   Mean   :-0.00442
 3rd Qu.:-0.01895   3rd Qu.:-0.00145
 Max.   :-0.00130   Max.   :-0.00060
```

Código ejecutado en lenguaje de programación R

Efectivamente, la distancia entre la mínima y la máxima es bastante más amplia para la variable dependiente que para la variable independiente.

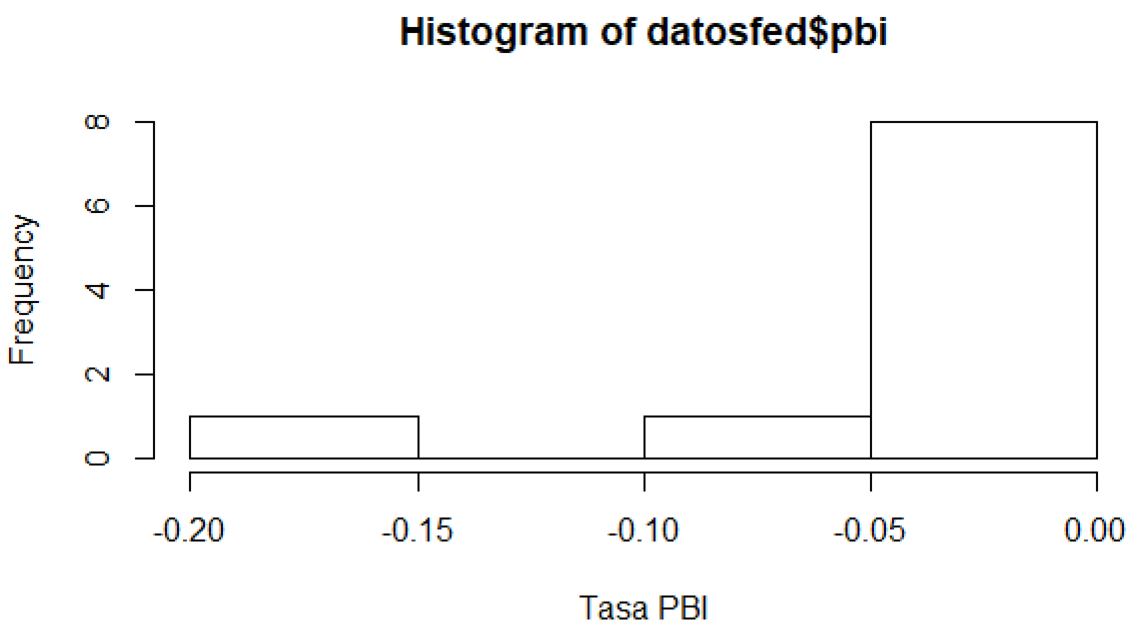
Con fines de visualización, se pueden graficar los histogramas de ambas variables, tal como se expone en la Ilustración 12 y en la Ilustración 13.

Ilustración 12: histograma T10Y1Y (b)



Código ejecutado en lenguaje de programación R

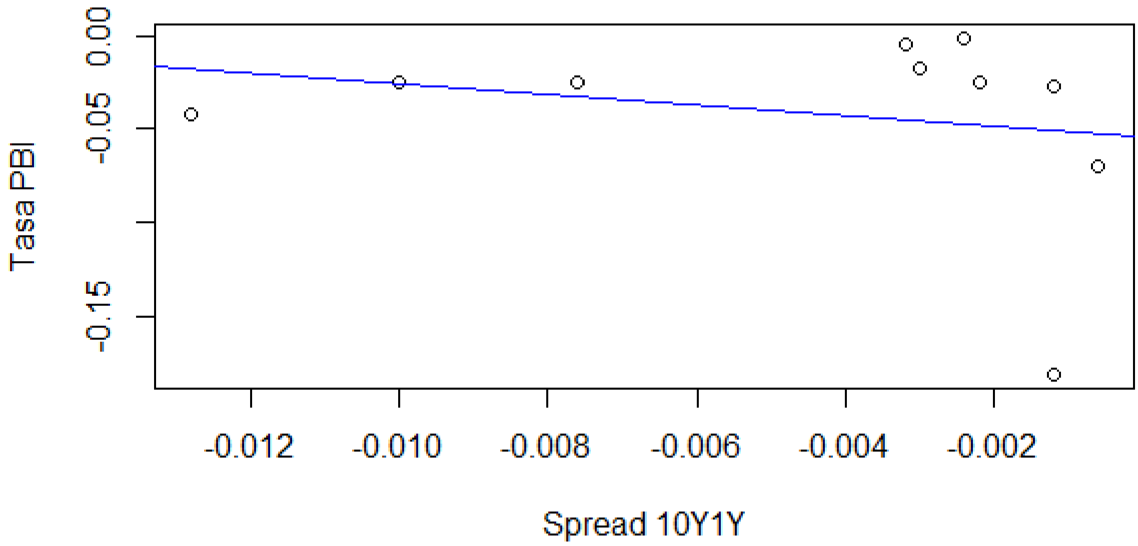
Ilustración 13: histograma crecimiento negativo PBI EEUU (c)



Código ejecutado en lenguaje de programación R

Para prever cómo luce el modelo, puede trazarse un diagrama de dispersión que relacione ambas variables, al que se agrega una línea de regresión, que se reproduce en la Ilustración 14.

Ilustración 14: inversión de la yield curve y recesiones en EEUU



Código ejecutado en lenguaje de programación R

La pendiente negativa de la línea de regresión ya anticipa una respuesta a la pregunta de investigación, y permite intuir si la hipótesis será rechazada o no. Pero previo a eso, se define formalmente el modelo de regresión lineal simple 1 por medio de la Ecuación 15 (el modelo aparece en el *environment*).

Ecuación 15: modelo de regresión lineal simple 1 (a)

$$tasaPBI = \beta_0 + \beta_1 spreadT10Y1Y + u$$

El *output* del modelo de regresión arrojado por R se exhibe en la Ilustración 15.

Ilustración 15: output modelo de regresión lineal simple 1

```
Call:
lm(formula = pbi ~ spread, data = datosfed)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.13026 -0.01295  0.01563  0.02776  0.04652 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -0.05467    0.02568  -2.129   0.0659 .
spread      -2.85527    4.31249  -0.662   0.5265  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05443 on 8 degrees of freedom
Multiple R-squared:  0.05195,	Adjusted R-squared:  -0.06656 
F-statistic: 0.4384 on 1 and 8 DF,  p-value: 0.5265
```

Código ejecutado en lenguaje de programación R

El primer punto de la interpretación es sustituir los parámetros abstractos por coeficientes numéricos, para conseguir la fórmula concreta de la regresión, lo cual se muestra en la Ecuación 16.

Ecuación 16: modelo de regresión lineal simple 1 (b)

$$tasaPBI = -0{,}05467 - 2{,}85527 \cdot spreadT10Y1Y$$
$$n = 10, \qquad R^2 = 0{,}05195$$

El coeficiente negativo para la única variable independiente significa, lisa y llanamente, que no se detecta una correlación positiva entre el spread de tasas de interés y la caída del PBI. Es decir, no necesariamente cuánto más aguda es la inversión de la curva de rendimientos, más profunda es la recesión. Es más, si el modelo es correcto, estaría diciendo que cuánto más fuerte es la inversión de la *yield curve*, menos grave será la recesión. Esto llevaría a rechazar la hipótesis en cuestión.

Por medio del error estándar podemos crear un intervalo de confianza (usando una distribución *T-student*). Dicho intervalo de confianza del 95% y con 8 grados de libertad se encuentra entre -12,7999 y 7,08933. Es un intervalo excesivamente amplio (podría estrecharse, pero implicaría reducir el nivel de confianza). Esto se debe a que el error estándar es demasiado grande en relación con el valor del coeficiente. Estos inconvenientes son resultado de contar con una muestra demasiado pequeña.

Consecuentemente, no sorprende que el estadístico t sea tan bajo. Esto acarrea un cuestionamiento adicional a la validez del modelo, y permite preguntarse si existe certeza de que el coeficiente de la variable independiente es distinto de cero (la hipótesis nula justamente plantea que el coeficiente no es cero, pero los resultados no nos habilitan a rechazar la hipótesis nula con suficiente seguridad). De forma esperable, el valor p es significativamente mayor al umbral de 0,05 que conduciría a afirmar con certeza que el coeficiente es distinto de cero. Así, se refuerza la idea de que la hipótesis nula no puede ser rechazada (y que por lo tanto la hipótesis de investigación sí debería ser rechazada).

Los residuales suministran información adicional de interés. La distribución no es simétrica y se halla sesgada hacia la izquierda. El modelo no parece ser muy efectivo para predecir los valores más bajos (las recesiones más severas). En realidad, todo el modelo se ve gravemente afectado por un par de puntos de dato, principalmente los de las recesiones de 1957-1958 y 2020. Esto da una pista sobre qué decisión metodológica podría tomarse para intentar obtener un modelo que se ajuste más a las expectativas.

Por último, el coeficiente de determinación es bajo, de apenas 5,2%. Por lo tanto, la capacidad explicativa del modelo es bastante pobre. Restando aún más significado estadístico al análisis de regresión, el estadístico F también es muy bajo, lo cual refuerza la noción de que la hipótesis nula no puede ser rechazada. En otras palabras, no podemos afirmar que el coeficiente es distinto de cero y que por lo tanto existe una correlación entre las variables.

V.2.2 Modelo de regresión lineal simple 2

Al examinarse los residuales del modelo de regresión lineal simple 1, se descubrió que la distribución no era simétrica. Se intuyó que si las observaciones más extremas eran excluidas, tal vez los resultados serían distintos. Eso es lo que se hará en lo que llamamos modelo de regresión lineal simple 2.

Este segundo modelo procura dar respuesta a la misma pregunta de investigación que el primero, y comprobar si la correspondiente hipótesis de investigación es cierta o no. Las variables independiente y dependiente permanecen iguales en términos abstractos, pero sus valores, y por lo tanto sus propiedades estadísticas, son diferentes. La forma funcional, de la Ecuación 1, tampoco se ve alterada.

El primer paso es depurar el data set, que ahora contiene los valores que se muestran en la Ilustración 16.

Ilustración 16: valores del modelo de regresión lineal simple 2

```
> datosfed2 <- datosfed[-c(1, 10), ]
> datosfed2
      pbi  spread
2 -0.0170 -0.0030
3 -0.0048 -0.0032
4 -0.0248 -0.0076
5 -0.0425 -0.0128
6 -0.0253 -0.0100
7 -0.0275 -0.0012
8 -0.0013 -0.0024
9 -0.0253 -0.0022
```

Código ejecutado en lenguaje de programación R

Y a continuación se repite básicamente el mismo proceso que con el modelo de regresión lineal simple 1, corriendo el código en R. Las principales propiedades estadísticas tanto de la variable independiente como de la variable dependiente con sus nuevos valores, incluyendo promedio, varianza y desvío estándar, se exhiben en la Ilustración 17.

Ilustración 17: propiedades estadísticas del modelo de regresión lineal simple 2 (a)

```
> mean(datosfed2$pbi)
[1] -0.0210625
> mean(datosfed2$spread)
[1] -0.0053
> var(datosfed2$pbi)
[1] 0.0001746313
> var(datosfed2$spread)
[1] 1.830857e-05
> sd(datosfed2$pbi)
[1] 0.01321481
> sd(datosfed2$spread)
[1] 0.004278852
```

Código ejecutado en lenguaje de programación R

Lo que más de destaca, tal como se esperaba, es que la varianza de la tasa de cambio del PBI se reduce bastante (y lógicamente la desviación estándar). Las otras propiedades estadísticas naturalmente se ven trastocadas por la alteración en el *data set*, pero los cambios no parecen ser tan significativos.

Ilustración 18: propiedades estadísticas del modelo de regresión lineal simple 2 (b)

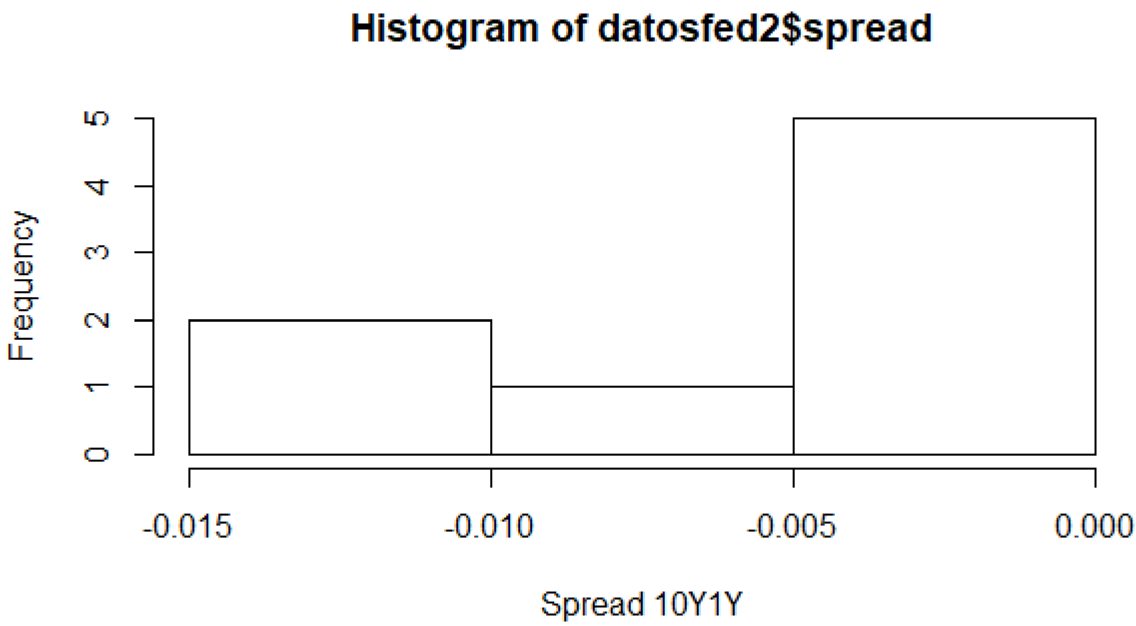
```
> summary(datosfed2)
      pbi                 spread
 Min.   :-0.04250   Min.   :-0.01280
 1st Qu.:-0.02585   1st Qu.:-0.00820
 Median :-0.02505   Median :-0.00310
 Mean   :-0.02106   Mean   :-0.00530
 3rd Qu.:-0.01395   3rd Qu.:-0.00235
 Max.   :-0.00130   Max.   :-0.00120
```

Código ejecutado en lenguaje de programación R

Al compararse otros valores estadísticos, como los que se muestran en la Ilustración 18, se verifica que la mínima de la variable dependiente es ahora bastante mayor (o menor en términos absolutos). La máxima se mantiene igual (ya que se sustrajeron los dos valores más bajos), pero la distancia entre la mínima y la máxima lógicamente se estrechó. Esto es coherente con el menor valor registrado para la varianza. Por otra parte, el promedio cae a la mitad (aumenta en términos absolutos). En lo que concierne a la variable independiente, no presenta alteraciones sustanciales.

La Ilustración 19 y la Ilustración 20 reproducen los histogramas de ambas variables. En el caso del *spread* entre bonos de corto y largo plazo, no se perciben modificaciones notables en la distribución de los datos. Por contraposición, si se comparan los histogramas de la tasa de cambio del PBI, se comprueba que la distribución de las observaciones, que antes estaba marcadamente sesgada hacia la izquierda, ahora ostenta un reparto algo más simétrico (teniendo en cuenta el escaso número de observaciones con que cuenta el *data set*).

Ilustración 19: histograma T10Y1Y (c)



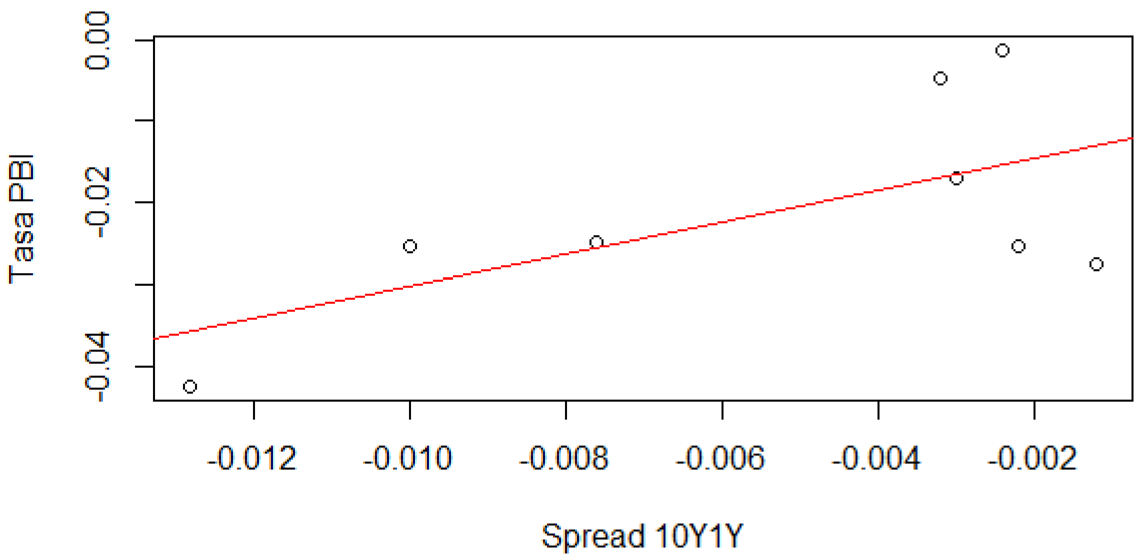
Código ejecutado en lenguaje de programación R

Ilustración 20: histograma crecimiento negativo PBI EEUU (d)

Código ejecutado en lenguaje de programación R

La Ilustración 21 revela cómo luce el diagrama de dispersión y la correspondiente línea de regresión de este nuevo modelo acotado en la cantidad de observaciones (pero que conserva las mismas variables).

Ilustración 21: inversión de la yield curve y recesiones en EEUU (b)



Código ejecutado en lenguaje de programación R

A continuación se interpreta el *output* producido por R para este nuevo análisis de regresión, pero sólo con observar la gráfica se advierte que la pendiente de la línea de regresión es ascendente, lo que nos permite anticipar que la correlación entre ambas variables es positiva. Vale recordar que el gráfico está representando el cuadrante III del plano cartesiano, de allí que los valores para ambos ejes son negativos (una característica que ya conocíamos del conjunto de datos).

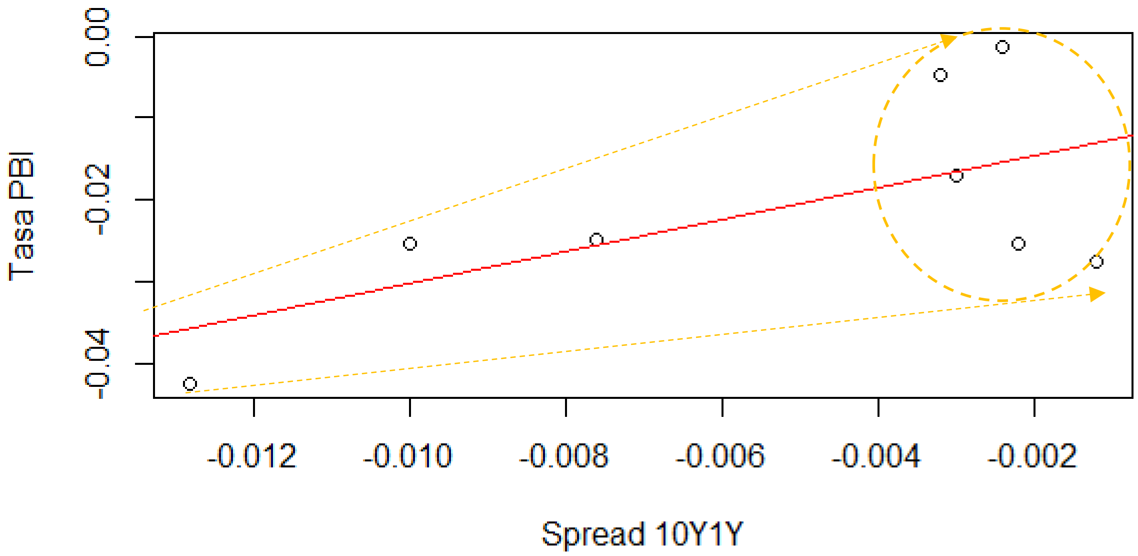
Asimismo, se advierte que la distancia de las observaciones respecto a la línea de regresión no es la misma para cada tramo del eje horizontal. Para valores más bajos de x, en la esquina superior derecha del gráfico, se contempla una diseminación mayor de los datos. Esta cualidad se conoce como heterocedasticidad, y surge cuando la varianza de los errores no se mantiene constante para todo el conjunto de datos.

Es cierto que el conjunto de datos con el que se está trabajando es reducido, pero también ha sido sometido a reiteradas manipulaciones. El inconveniente con la heterocedasticidad es que viola uno de los supuestos básicos del modelo de regresión lineal, que es justamente su opuesto, la homocedasticidad. Las implicancias de la presencia de heterocedasticidad son relevantes para las conclusiones que se pretendan extraer.

Desde el punto de vista metodológico, teniendo en cuenta que se está aplicando el MCO, lo que ocurre es que los estimadores de los coeficientes carecen de eficiencia. La eficiencia estadística acontece cuando los estimadores poseen mínima varianza. Los estimadores sí conservan la ventaja de ser insesgados, pero al no ser eficientes ya cesan de ser un mejor estimador lineal insesgado (MELI)[15]. La heterocedasticidad del modelo actual se encuentra señalada en la Ilustración 22.

[15] *Best linear unbiased estimator* (BLUE) en inglés.

Ilustración 22: heterocedasticidad modelo de regresión lineal simple 2



Código ejecutado en lenguaje de programación R

Como paso previo a analizar el *output* del código en R, se recuerda que formalmente el modelo de regresión lineal simple 2 se define de idéntica manera que el modelo de regresión lineal simple 1. Es decir, se mantiene la forma funcional de la Ecuación 15, pero en el *environment* aparece un nuevo modelo.

Ecuación 15: modelo de regresión lineal simple 1 (a)

$$tasaPBI = \beta_0 + \beta_1 spreadT10Y1Y + u$$

El output del modelo de regresión arrojado por R se exhibe en la Ilustración 23.

Ilustración 23: output modelo de regresión lineal simple 2

```
Call:
lm(formula = pbi ~ spread, data = datosfed2)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0144536 -0.0076551  0.0001625  0.0067529  0.0140926

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.010700   0.006482  -1.651    0.150
spread       1.955134   0.976019   2.003    0.092 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01105 on 6 degrees of freedom
Multiple R-squared:  0.4008,    Adjusted R-squared:  0.3009
F-statistic: 4.013 on 1 and 6 DF,  p-value: 0.09202
```

Código ejecutado en lenguaje de programación R

La primera tarea del procedimiento de análisis del *output* es precisar numéricamente los coeficientes de la ecuación de regresión, lo cual se muestra en la Ecuación 17.

Ecuación 17: modelo de regresión lineal simple 2

$$tasaPBI = -0{,}0107 + 1{,}955134 \cdot spreadT10Y1Y$$
$$n = 8, \qquad R^2 = 0{,}4008$$

La diferencia preeminente entre este segundo modelo de regresión y el primero, es que el coeficiente de la variable independiente ha pasado de ser negativo a ser positivo. Esto supone una correlación positiva entre la inversión de la curva de rendimientos y las recesiones. Dicho de otra manera, cuánto más acentuado es el *spread* negativo entre bonos de corto y largo plazo, más intensa es la recesión subsiguiente. Esto conduciría en principio a convalidar (en términos estadísticos, a no poder rechazar) la hipótesis de investigación, enunciada de la siguiente manera:

Existe una correlación positiva y significativa entre el spread negativo entre bonos de corto y largo plazo, por un lado, y la profundidad de la recesión que acontece con posterioridad a la inversión de la curva de vencimientos, por el otro.

Prosiguiendo con la interpretación del *output*, valiéndonos del error estándar es posible establecer un intervalo de confianza, empleando una distribución *T-student*. Este intervalo de confianza del 95% cuenta con 6 grados de libertad y se halla entre -0.4332 y 4.3435. Se trata de un intervalo sobradamente más acotado que el del modelo de regresión lineal simple 1. Eso es producto de que en este segundo modelo el error estándar es menor al coeficiente. Sin embargo, es menester señalar que la muestra es más restringida (ocho observaciones) y los grados de libertad son muy bajos. Se impone por lo tanto cierta prudencia a la hora de extraer conclusiones.

Por su parte, el estadístico t, que no es sino el coeficiente dividido por el error estándar, es más alto. No es demasiado elevado, es cierto (el dividendo es apenas dos veces el divisor), pero confiere más seguridad de que el coeficiente del spread de tasas es diferente de cero. Esto ofrece un segundo elemento (el primero era el signo positivo y el valor del coeficiente en sí mismo, reflejado en la pendiente positiva de la línea de regresión) para contestar afirmativamente la hipótesis de investigación. En contraste, el valor p se ha atenuado en comparación al primer modelo, pero permanece por encima del límite de 0,05. Esto contrarresta las inferencias previas de que es posible rechazar la hipótesis nula.

El análisis de los residuales es sugestivo. La mediana es prácticamente de cero, por lo que la distribución es más simétrica. Esto ya se anticipó al examinar el histograma de la tasa porcentual de cambio del PBI para este segundo modelo. Igualmente, la mínima y la máxima por un lado, y el primer intercuartil y el tercer intercuartil por el otro, son casi idénticos en términos absolutos. Es decir, son el mismo número pero con signo invertido. Consecuentemente, el modelo gozaría de efectividad para predecir tanto los valores más bajos como los más altos. En otras palabras, todas las recesiones independientemente de su gravedad.

Esto se ve empañado por la presencia de heterocedasticidad en los datos. Al haber suprimido las observaciones correspondientes a las recesiones de 1957-1958 y de 2020, fue factible reordenar un conjunto de datos que se ajusta mejor a la línea de regresión. La distribución es más simétrica, pero la varianza de los errores no es constante, lo que atenta directamente contra uno de los supuestos básicos del modelo de regresión lineal.

Finalmente, el coeficiente de determinación es bastante elevado, de 40,08%. Esto supondría que la variación en la inversión de la curva de rendimientos explica aproximadamente en un 40% la profundidad de la recesión[16]. El poder predictivo del modelo se ha acrecentado de manera indudable. El estadístico F también ha aumentado, lo que comportaría el poder rechazar la hipótesis nula con más sustento y afirmar con más fundamento que el coeficiente no es cero y existe una correlación positiva entre las variables. No obstante, se insiste en que el conjunto de datos es muy pequeño y ha sido manipulado arbitrariamente en repetidas ocasiones.

V.2.3 Modelo de regresión lineal múltiple

Este tercer análisis de regresión es el último que intenta responder a la primera pregunta de investigación. Pero a diferencia de los dos anteriores, éste es un análisis múltiple. Dispone de dos variables independientes, pero vinculadas entre sí, ya que una es el cuadrado de la otra. Esta forma funcional busca descubrir si el valor ya existente en la variable independiente ejerce alguna influencia adicional en el comportamiento de la variable dependiente. Puede proveer pistas para responder la segunda hipótesis de investigación, que pretende explicar la relación entre la duración de la inversión de la curva de rendimientos y la duración de la recesión.

La Ecuación 7 sintetizaba la formulación abstracta de esta forma funcional cuadrática.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

A su vez, cabe invocar la Ecuación 8, que da cuenta de la manera en que la variable dependiente cambia ante fluctuaciones en la variable independiente.

$$\frac{\Delta y}{\Delta x} = \beta_1 + 2\beta_2 x$$

[16] Aún en el caso de que el modelo fuese correcto desde el punto de vista estadístico, debe recordarse el riesgo de que se esté incurriendo en el sesgo de variable omitida, algo probable al intentar explicar un fenómeno tan complejo como la marcha de la economía en base a una única variable independiente.

Ya especificado el tipo de regresión, la pertinente forma funcional, las variables, resta sólo transformar el *data set*. Deben añadirse lo valores de la nueva variable independiente (el cuadrado de la variable previamente existente). Se parte del *data set* del modelo línea simple 2, renombrándoselo para que aparezca de manera independiente en el *environment* de RStudio, tal como se manifiesta en la Ilustración 24.

Ilustración 24: valores del modelo de regresión lineal múltiple (a)

```
> datosfed3 <- datosfed2
> datosfed3
      pbi  spread
2 -0.0170 -0.0030
3 -0.0048 -0.0032
4 -0.0248 -0.0076
5 -0.0425 -0.0128
6 -0.0253 -0.0100
7 -0.0275 -0.0012
8 -0.0013 -0.0024
9 -0.0253 -0.0022
```

Código ejecutado en lenguaje de programación R

Como este *data set* es una réplica del anterior, contiene las mismas variables con sus valores. Debe agregarse la nueva variable, por medio del comando indicado en la Ilustración 25.

Ilustración 25: nueva variable del modelo de regresión lineal múltiple

```
> datosfed3$spreadcuadrado = datosfed3$spread^2
> datosfed3$spreadcuadrado
[1] 0.00000900 0.00001024 0.00005776 0.00016384 0.00010000 0.00000144 0.00000576
[8] 0.00000484
```

Código ejecutado en lenguaje de programación R

El nuevo data set con el que se trabaja a partir de este momento se reproduce en la Ilustración 26.

Ilustración 26: valores del modelo de regresión lineal múltiple (b)

```
> datosfed3
      pbi  spread spreadcuadrado
2 -0.0170 -0.0030     0.00000900
3 -0.0048 -0.0032     0.00001024
4 -0.0248 -0.0076     0.00005776
5 -0.0425 -0.0128     0.00016384
6 -0.0253 -0.0100     0.00010000
7 -0.0275 -0.0012     0.00000144
8 -0.0013 -0.0024     0.00000576
9 -0.0253 -0.0022     0.00000484
```

Código ejecutado en lenguaje de programación R

Al igual que en las dos regresiones anteriores, pueden tomarse propiedades estadísticas tales como el promedio, la varianza y el desvío estándar tanto de la variable dependiente como de las variables independientes. Esto se presenta en la Ilustración 27.

Ilustración 27: propiedades estadísticas del modelo de regresión lineal múltiple (a)

```
> mean(datosfed3$pbi)
[1] -0.0210625
> mean(datosfed3$spread)
[1] -0.0053
> mean(datosfed3$spreadcuadrado)
[1] 4.411e-05
> var(datosfed3$pbi)
[1] 0.0001746313
> var(datosfed3$spread)
[1] 1.830857e-05
> var(datosfed3$spreadcuadrado)
[1] 3.551252e-09
> sd(datosfed3$pbi)
[1] 0.01321481
> sd(datosfed3$spread)
[1] 0.004278852
> sd(datosfed3$spreadcuadrado)
[1] 5.959238e-05
```

Código ejecutado en lenguaje de programación R

De forma complementaria, se computan otros valores tales como la mínima, la máxima, la mediana y el rango intercuartil. Los mismos se enseñan en la Ilustración 28.

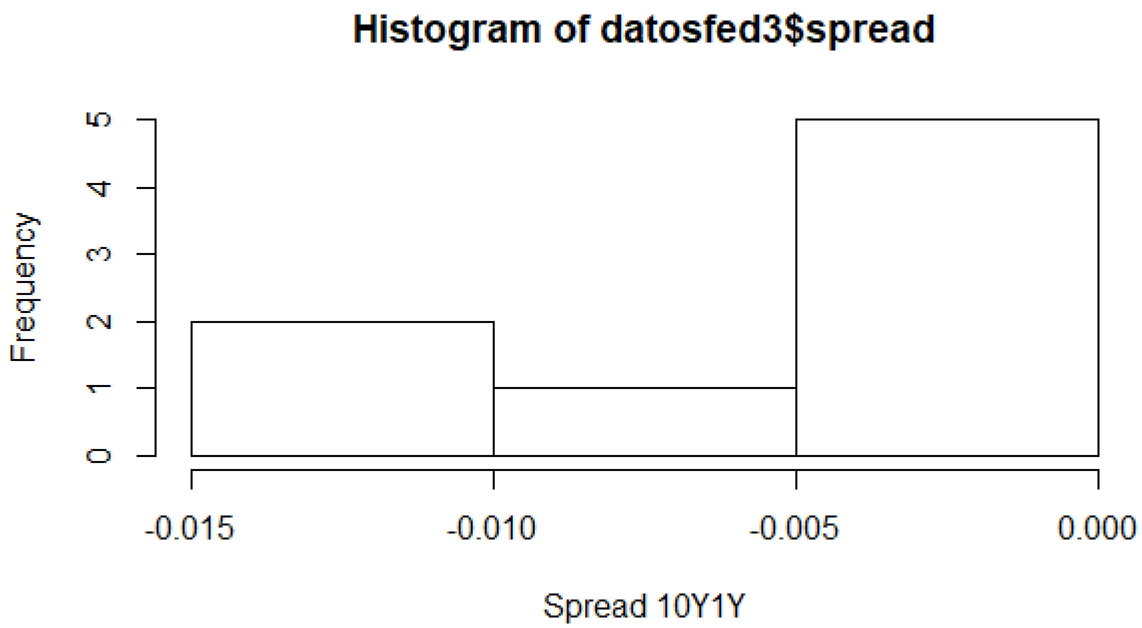
Ilustración 28: propiedades estadísticas del modelo de regresión lineal múltiple (b)

```
> summary(datosfed3)
      pbi                 spread            spreadcuadrado
 Min.    :-0.04250   Min.    :-0.01280   Min.    :1.440e-06
 1st Qu.:-0.02585   1st Qu.:-0.00820   1st Qu.:5.530e-06
 Median :-0.02505   Median :-0.00310   Median :9.620e-06
 Mean    :-0.02106   Mean    :-0.00530   Mean    :4.411e-05
 3rd Qu.:-0.01395   3rd Qu.:-0.00235   3rd Qu.:6.832e-05
 Max.    :-0.00130   Max.    :-0.00120   Max.    :1.638e-04
```

Código ejecutado en lenguaje de programación R

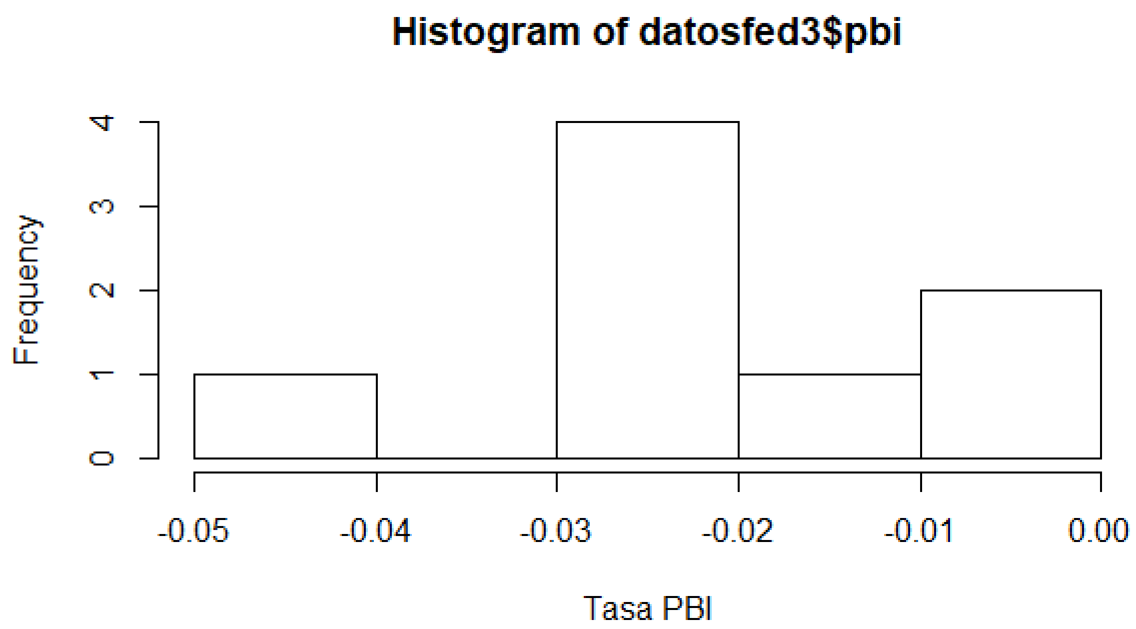
Los histogramas de la variable dependiente y de la variable dependiente en su forma original (lineal), lógicamente no se han distorsionado, tal como se muestra en la Ilustración 29 y en la Ilustración 30 respectivamente.

Ilustración 29: histograma T10Y1Y (d)



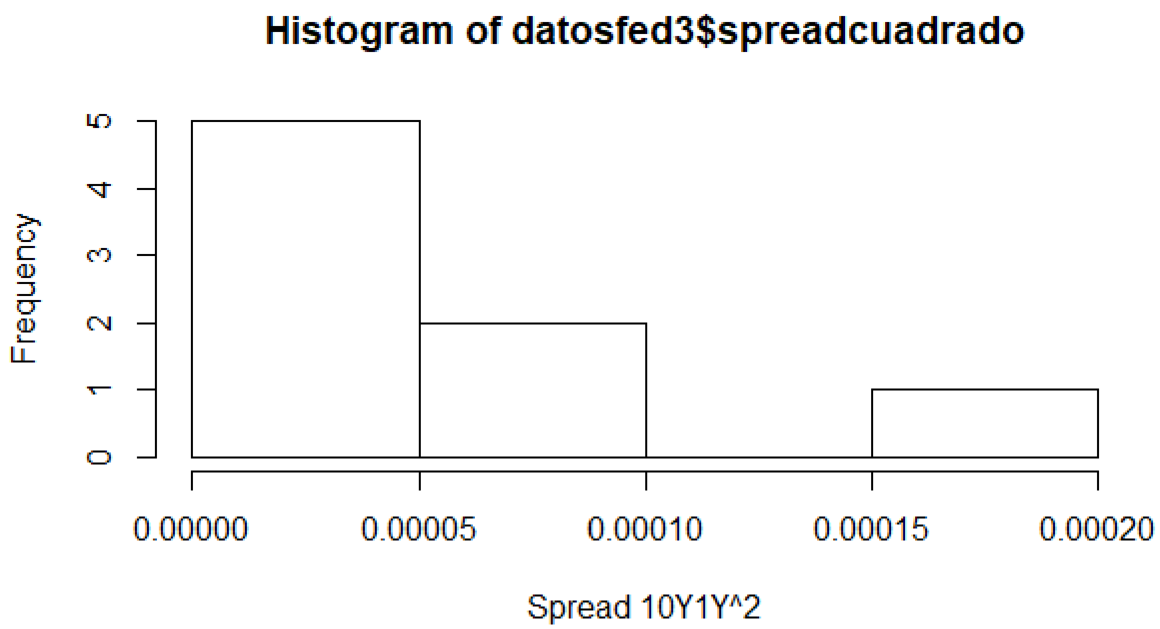
Código ejecutado en lenguaje de programación R

Ilustración 30: histograma crecimiento negativo PBI EEUU (e)



Código ejecutado en lenguaje de programación R

Ilustración 31: histograma T10Y1Y cuadrada



Código ejecutado en lenguaje de programación R

En lo que refiere al histograma de la variable independiente en su forma cuadrática (Ilustración 31), se observa un sesgo hacia la derecha en la distribución. Pero no es posible extraer conclusiones de esto por el hecho de que los datos son pocos, los valores son extremadamente bajos (por tratarse de cuadrados de decimales), y muy cercanos entre sí.

Al tratarse de un modelo de regresión múltiple, no puede graficarse la relación entre las variables en un plano bidimensional. Por eso es que se procede a definir formalmente este modelo de regresión lineal múltiple por medio de la Ecuación 18 (el modelo debe aparecer en el *environment*).

Ecuación 18: modelo de regresión lineal múltiple (a)

$$tasaPBI = \beta_0 + \beta_1 spreadT10Y1Y + \beta_2 spreadT10Y1Ycuadrada + u$$

El *output* del modelo de regresión múltiple generado por R se expone en la Ilustración 32.

Ilustración 32: output modelo de regresión lineal múltiple

```
Call:
lm(formula = pbi ~ spread + spreadcuadrado, data = datosfed3)

Residuals:
        2          3          4          5          6          7          8 
-0.0022028  0.0096685 -0.0078105  0.0015178  0.0009267 -0.0080109  0.0147143 
        9 
-0.0088030 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)  
(Intercept)      -0.02435    0.01129  -2.157   0.0835 .
spread           -4.62997    4.70410  -0.984   0.3702  
spreadcuadrado -481.75084  337.76376  -1.426   0.2131  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0102 on 5 degrees of freedom
Multiple R-squared:  0.5741,	Adjusted R-squared:  0.4037 
F-statistic: 3.369 on 2 and 5 DF,  p-value: 0.1184
```

Código ejecutado en lenguaje de programación R

La primera tarea del procedimiento de análisis del output es precisar numéricamente los coeficientes de la ecuación de regresión, lo cual se muestra en la Ecuación 19.

Ecuación 19: modelo de regresión lineal múltiple (b)

$$tasaPBI = -0{,}0244 - 4{,}63 \cdot spreadT10Y1Y - 481{,}7508 \cdot spreadT10Y1Ycuadrada$$
$$n = 8, \qquad R^2 = 0{,}5741$$

Al igual que en el modelo de regresión lineal simple 1, los *estimates* son negativos. En este caso, los coeficientes de ambas variables independientes son negativos. En consecuencia, al igual que en aquel modelo, no se descubre una correlación positiva entre la inversión de la curva de rendimientos y una recesión, incluso si se agrega un término cuadrático. Aún más, si se calcula el efecto que una variación de *x* tiene sobre *y*, resulta que los coeficientes son literalmente gigantescos (en sentido negativo), como se muestra en la Ecuación 20.

Ecuación 20: tasa de cambio del modelo de regresión lineal múltiple

$$\frac{\Delta y}{\Delta x} = -4{,}62997 - 963.50168x$$

Esto acarrearía rechazar rotundamente la hipótesis de investigación.

Recurriendo como siempre al error estándar, se pueden construir intervalos de confianza empleando la distribución *T-student*. Los intervalos de confianza con del 95% con 5 grados de libertad (se pierde un grado más de libertad en comparación a las regresiones anteriores porque se introdujo una nueva variable) son de entre -16,7242 y 7,4643 para la variable independiente lineal, y de entre -1.350,1415 y 386,6398 para la variable independiente cuadrática.

Son evidentemente intervalos de confianza excesivamente amplios, por la razón de que los errores estándares son demasiado grandes, y los coeficientes demasiado pequeños. Así, este modelo de regresión lineal múltiple presenta el mismo inconveniente que el modelo de regresión lineal simple 1.

De manera acorde, los estadísticos t de ambas variables son bajos, lo que arrima otro cuestionamiento al modelo. Como es esperable, ambos valores p son significativamente mayores a 0,05. Todos estos factores llevan a inferir que no existen certezas de que los coeficientes de ambas variables independientes son distintos de cero. En consecuencia, no es posible rechazar la hipótesis nula, mientras que la hipótesis de investigación sí debería ser rechazada.

Por último, el coeficiente de determinación es bastante alto, de 57,41%. Lamentablemente, el modelo carece de significancia estadística por los motivos ya mencionados. Además, incluir nuevas variables siempre incrementa el R cuadrado automáticamente (al menos que la nueva variable sea una combinación lineal de las anteriores).

Como el modelo de regresión lineal simple 1 y el modelo de regresión múltiple son *nested models* (cuando un modelo está contenido en otro, en este caso, el modelo de regresión lineal simple 1 está contenido en el modelo de regresión lineal múltiple porque este último contiene todas las variables de aquel), la R cuadrada ajustada es un criterio comparativo óptimo. Se observa que este indicador presenta diferencias marcadas entre ambos modelos. Lo mismo acontece con el estadístico F. A diferencia del modelo de regresión lineal simple 1, este valor es lo suficientemente alto en el modelo de regresión múltiple, al igual que el coeficiente de determinación. No obstante, los cuestionamientos ya desarrollados impiden depositar la suficiente confianza en la validez del modelo.

V.3 Segundo análisis de regresión

Este segundo análisis de regresión pretende dar respuesta a la segunda pregunta de investigación, propuesta de la siguiente manera:

¿De qué manera impacta la duración de una inversión de la curva de rendimientos en la duración de la recesión que acontece con posterioridad a la aparición de dicha inversión?

La hipótesis que intenta responder de manera tentativa dicho interrogante es la siguiente:

Existe una correlación positiva y significativa entre la duración de una inversión de la curva de rendimientos, por un lado, y la duración de la recesión que acontece con posterioridad a la aparición de dicha inversión, por el otro.

Dicho en otros términos, la hipótesis aquí presentada asegura que cuanto más duradero es el *spread* (cuanto mayor sea el tiempo que transcurre desde que el diferencial entre la tasa de bonos de corto y largo plazo se torna negativo hasta que vuelve a ser positivo), más dilatada temporalmente es la recesión que sobreviene con posterioridad. Para testear esta hipótesis, es que se realiza un análisis de regresión, siguiendo los mismos lineamientos generales que los diversos análisis efectuados para la primera hipótesis.

La variable independiente es la duración en meses de cada instancia de inversión de la curva, mientras que la variable dependiente es la duración en meses de cada recesión posterior. La información ya fue presentada en la Tabla 7.

Tabla 7: inversiones de la yield curve y recesiones en EEUU tras la SGM en meses

Inversión de la Curva de Rendimientos	**Duración**	**Recesión**	**Duración**
1956-1957	7	1957-1958	6
1959-1960	6	1960-1961	9
1968-1970	22	1970	15
1973-1974	19	1974-1975	15
1978-1980	20	1980	6
1980-1982	17	1981-1982	12
1989	7	1990-1991	6
2000	8	2001	9
2006-2007	15	2008-2009	18
2019	2	2020	6

Fuente: Board of Governors of the Federal Reserve System

Debido a que este análisis de regresión incorpora sólo una variable independiente, se trata de un modelo simple. La relación que se busca establecer entre las variables es de carácter lineal. La forma funcional de este tipo de regresión se consigue evocando la Ecuación 1.

$$y = \beta_0 + \beta_1 x + u$$

Ya precisados el tipo de regresión, la forma funcional, las variables, y disponiendo de los datos necesarios, es viable llevar a cabo la regresión, que se denomina modelo de regresión lineal simple 3. Concretamente, eso comprende ejecutar el código en lenguaje de programación R desarrollado explícitamente para este propósito.

En primera instancia debe cargarse el *data set* que contiene la información sobre la duración en meses de los *spreads* entre los bonos de corto y largo plazo, así como la duración en meses de las recesiones. Se trata de un conjunto de datos completamente nuevo, diferente del usado en los análisis de regresiones anteriores[17]. Por lo que los comandos a ejecutar son ligeramente distintos. Una vez que el nuevo *data set* aparece en el *environment* de RStudio, debe rotularse las columnas, que es como se ordenan las variables. El *output* del código se observa en la Ilustración 33.

Ilustración 33: valores del modelo de regresión lineal simple 3

```
> colnames(datosfed4) <- c("recesiones", "curvasinvertidas")
> datosfed4
   recesiones curvasinvertidas
1           6                7
2           9                6
3          15               22
4          15               19
5           6               20
6          12               17
7           6                7
8           9                8
9          18               15
10          6                2
```

Código ejecutado en lenguaje de programación R

[17] Aunque la fuente original de la información es la misma en ambos casos, sólo que ha sido procesada y combinada de diferente manera.

Pueden compendiarse las principales propiedades estadísticas tanto de la variable independiente como de la variable dependiente, incluyendo promedio, varianza y desvío estándar. El output se exhibe en la Ilustración 34.

Ilustración 34: propiedades estadísticas del modelo de regresión lineal simple 3 (a)

```
> mean(datosfed4$recesiones)
[1] 10.2
> mean(datosfed4$curvasinvertidas)
[1] 12.3
> var(datosfed4$recesiones)
[1] 20.4
> var(datosfed4$curvasinvertidas)
[1] 49.78889
> sd(datosfed4$recesiones)
[1] 4.516636
> sd(datosfed4$curvasinvertidas)
[1] 7.056124
```

Código ejecutado en lenguaje de programación R

El rasgo más destacado es que el promedio de ambas variables es bastante similar. Esto podría ofrecer una pista sobre los resultados finales del modelo de regresión. Si se calculan otras propiedades estadísticas, como el mínimo, el máximo, la mediana y el rango intercuartil de ambas variables, se encuentran otros parecidos, tal como se muestra en la Ilustración 35.

Ilustración 35: propiedades estadísticas del modelo de regresión lineal simple 3 (b)

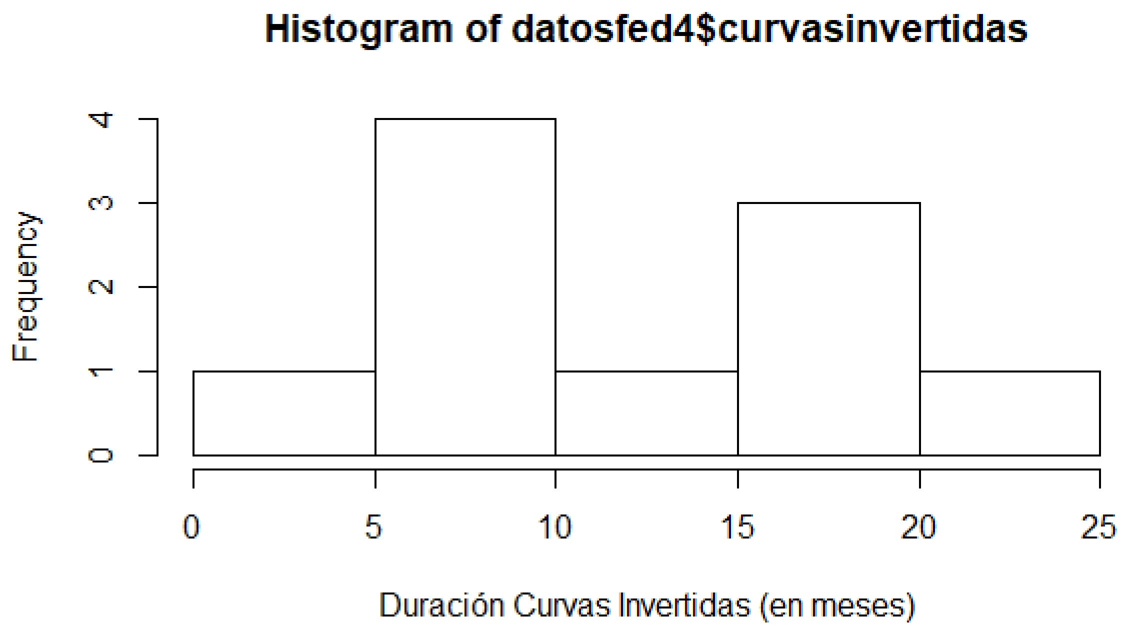
```
> summary(datosfed4)
   recesiones     curvasinvertidas
 Min.   : 6.00   Min.   : 2.0
 1st Qu.: 6.00   1st Qu.: 7.0
 Median : 9.00   Median :11.5
 Mean   :10.20   Mean   :12.3
 3rd Qu.:14.25   3rd Qu.:18.5
 Max.   :18.00   Max.   :22.0
```

Código ejecutado en lenguaje de programación R

Las mínimas y el tercer cuartil difieren en cierta medida, pero el primer cuartil, las medianas son bastante cercanas. Esto debe ser confrontado con el hecho de que la varianza de la duración de los *spreads* es mayor a la varianza de la duración de las recesiones (pero esta diferencia es menor a la encontrada en los análisis de regresión de la primera hipótesis).

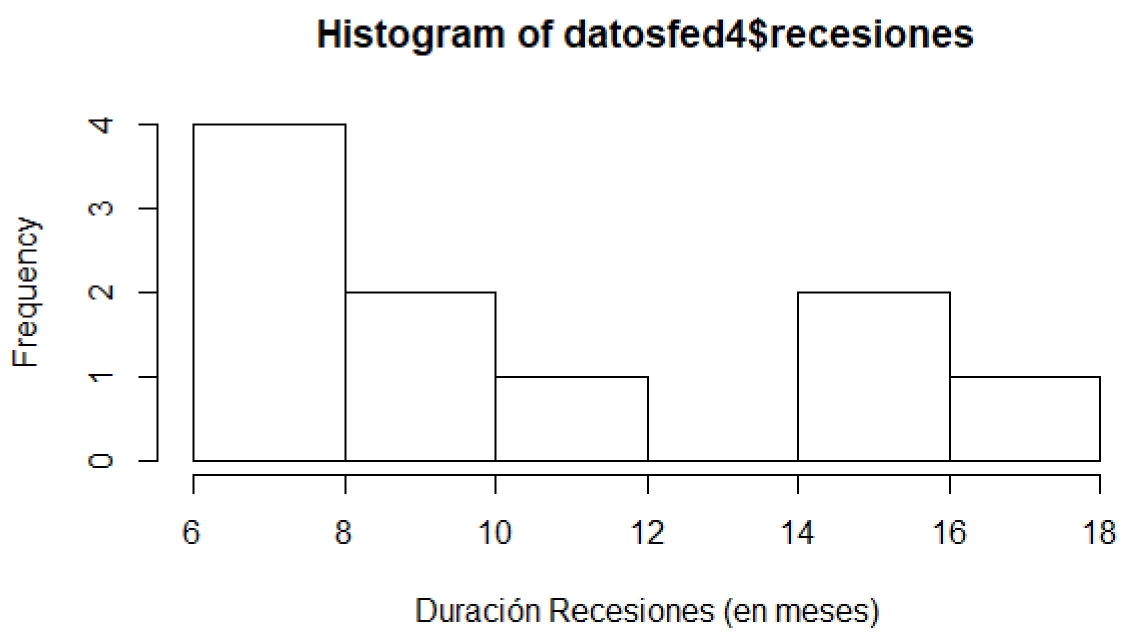
Con el objeto de visualizar ambas variables, se reproducen sus histogramas en la Ilustración 36 y en la Ilustración 37.

Ilustración 36: histograma T10Y1Y (e)



Código ejecutado en lenguaje de programación R

Ilustración 37: histograma crecimiento negativo PBI EEUU (f)

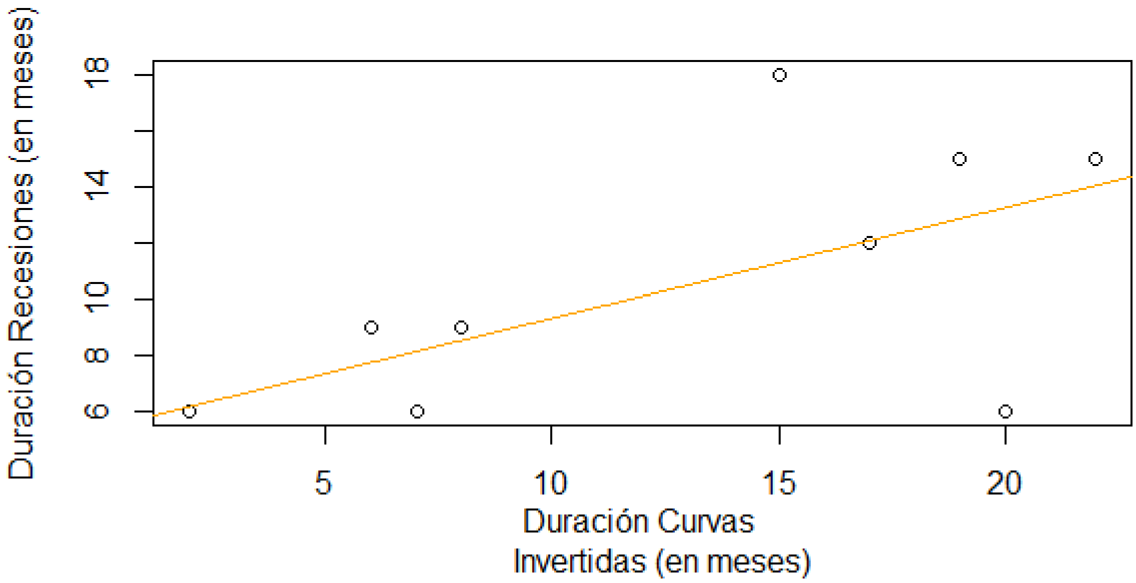


Código ejecutado en lenguaje de programación R

Ambos histogramas parecen presentar un patrón bimodal de distribución, aunque en el caso de la variable dependiente se nota cierto sesgo hacia la derecha.

El diagrama de dispersión y la respectiva línea de regresión se reproducen en la Ilustración 38.

Ilustración 38: duración inversión yield curve y recesiones



Código ejecutado en lenguaje de programación R

La pendiente positiva de la línea de regresión ya prevé una respuesta a la pregunta de investigación, y permitiendo atisbar si la hipótesis será rechazada o no. Como paso previo, se define formalmente el modelo de regresión lineal simple 3 por medio de la Ecuación 21 (el modelo aparece en el *environment*).

Ecuación 21: modelo de regresión lineal simple 3 (a)

$$duracionrecesion = \beta_0 + \beta_1 duracionspreadT10Y1Y + u$$

El *output* del modelo de regresión arrojado por R se exhibe en la Ilustración 39.

Ilustración 39: output modelo de regresión lineal simple 3

```
Call:
lm(formula = recesiones ~ curvasinvertidas, data = datosfed4)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.2312 -1.6215  0.2213  1.2054  6.7371 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)  
(Intercept)        5.3580     2.4990   2.144   0.0644 .
curvasinvertidas   0.3937     0.1785   2.206   0.0584 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.778 on 8 degrees of freedom
Multiple R-squared:  0.3782,	Adjusted R-squared:  0.3005 
F-statistic: 4.866 on 1 and 8 DF,  p-value: 0.05845
```

Código ejecutado en lenguaje de programación R

El primer paso de la interpretación es reemplazar los parámetros abstractos por valores numéricos, para obtener la fórmula concreta de la regresión, lo cual se muestra en la Ecuación 22.

Ecuación 22: modelo de regresión lineal simple 3 (b)

$$duracionrecesion = 5{,}358 + 0{,}3937 \cdot duracionspreadT10Y1Y$$
$$n = 10, \qquad R^2 = 0{,}3782$$

El coeficiente positivo de la única variable independiente implicaría la existencia de una correlación positiva entre la duración de una inversión en la curva de rendimientos y la duración de una posterior recesión. Debe notarse sin embargo que la pendiente, aunque positiva, es menor a la unidad. Esto significaría en lenguaje llano que, en promedio, una recesión no se prolongaría en el tiempo por un período tan extenso como el de la inversión de la curva de rendimientos que le precedió.

Además, el intercepto posee un valor relativamente alto. En teoría, ante la ausencia de un *spread* negativo en bonos de corto y largo plazo (es decir, cuando $x = 0$), la economía debería atravesar una recesión de más de 5 meses. Obviamente esto es un sinsentido. Por lo tanto, el intercepto está indicando la intervención de otras variables, no contempladas en el modelo, que condicionan la duración de una eventual recesión.

De todas maneras, el coeficiente positivo y con un valor relativamente alto en principio conduciría a aceptar la hipótesis de investigación. Esto es, el tiempo que la curva está invertida sí tiene un impacto en el tiempo que la economía está sumida en la recesión.

Apelando nuevamente al error estándar se puede delimitar un intervalo de confianza del 95% con 8 grados de libertad, siguiendo una distribución *T-student*. Este intervalo se ubica entre -0,0179 y 0,8053. Es una distancia razonable, fruto de que el error estándar es relativamente pequeño en relación al *estimate*. Esta propiedad refuerza aún más la confianza que se puede depositar en el modelo.

Paralelamente, el estadístico t también es relativamente alto, lo que implica una ventaja adicional intrínseca al modelo. Habría una certeza apreciable de que el coeficiente de la variable independiente es distinto de cero. Esto llevaría a rechazar la hipótesis nula. No obstante, el valor p, aunque bajo es ligeramente mayor a 0,05, lo que introduciría cierta dosis de incertidumbre respecto a si el coeficiente es cero o no.

En lo que respecta a los residuales, la distribución es mayormente simétrica, aunque moderadamente sesgada hacia la izquierda. Debe tenerse en cuenta que este *data set* es “completo”, es decir, del mismo no se excluyeron las recesiones más severas (las de 1957-1958 y 2020) como sí se lo hizo en el modelo de regresión lineal simple 2.

Para concluir, el coeficiente de determinación es alto, de 37,82%. En consecuencia, el poder explicativo del modelo es relevante. Asimismo, el estadístico F también es alto, lo cual consolida la percepción de que la hipótesis nula puede ser rechazada. Podemos afirmar con una certeza considerable que el coeficiente de la variable independiente es distinto de cero, y que existe una correlación positiva con la variable dependiente. Así, se comprobaría la hipótesis de investigación.

VI. Conclusiones

En esta Tesis Final, la propuesta de investigación consiste en descubrir qué tipo de correlaciones pueden establecerse entre la inversión de la curva de rendimientos como variable independiente, y las recesiones como variable dependiente. A tal efecto, se plantean dos preguntas e hipótesis de investigación específicas.

La primera de ellas afirma que existe una correlación positiva entre la profundidad de una inversión de la curva de rendimientos y la severidad de la subsecuente recesión. Es decir, que cuánto más negativo es el *spread* negativo de los bonos de corto y largo plazo, más fuerte es la caída del PBI.

La segunda dupla de pregunta e hipótesis de investigación específicas sostiene que existe una correlación positiva entre la permanencia en el tiempo de una inversión de la curva y lo prolongada que es una recesión. En otras palabras, que cuanto más se extienda un *spread* negativo, mayor será la duración de la recesión posterior.

Se llevaron a cabo cuatro análisis de regresión, tres correspondientes al primer par de pregunta e hipótesis de investigación específicas, y uno pertinente al segundo par. Los modelos fueron desarrollados usando *data sets* de fuentes oficiales, procesados con código en lenguaje de programación R, y generando *outputs* y gráficas a través de RStudio. La interpretación se realizó recurriendo a conceptos y metodologías característicos de la estadística y la econometría.

No fue posible probar la primera hipótesis específica con suficiente solidez estadística. En la primera variante, la del modelo de regresión lineal simple 1, todas las propiedades estadísticas invalidaban el análisis. En vistas de estos pobres resultados, se manipuló el conjunto de datos excluyendo a las observaciones más extremas. Específicamente, se suprimieron del conjunto de datos las recesiones de 1957-1958 y 2020, de las más profundas que sufrió EEUU tras las SGM, pero que fueron precedidas por inversiones leves de la curva.

Aunque algunas cualidades estadísticas presentaron mejoras, especialmente el coeficiente de determinación y la pendiente positiva de la línea de regresión (que indicaría una correlación positiva), el conjunto de datos ya había sufrido demasiadas manipulaciones. En consecuencia, no podía responderse afirmativamente la pregunta de investigación.

Por otra parte, a pesar de que se ensayó un modelo alternativo que contemplaba múltiples variables, no fue posible tampoco encontrar un patrón. Tal como puede verificarse al revisar la información original (sea en formato gráfico o tabular), no parece posible detectar un patrón que vincule la profundidad del *spread* de bonos con la gravedad de las recesiones. Algunas inversiones de curva intensas se ven sucedidas por caídas relativamente suaves en el PBI, mientras que *spreads* insignificantes están asociados a recesiones muy severas.

En lo que respecta al segundo par de pregunta e hipótesis de investigación específicas, los resultados son más alentadores. Sin alterar el *data set* original, fue posible encontrar, a través de un modelo de regresión lineal bivariado, una correlación positiva y significativa entre ambas variables. Es posible enunciar con un razonable grado de certeza que cuánto más se extiende en el tiempo una inversión de la curva de rendimientos, más prolongada será la recesión que le sucederá.

Esta conclusión tiene importantes consecuencias a la hora de entender el modo de funcionamiento de la economía, y en lo que concierne a las políticas implementadas por las autoridades. Por ejemplo, podría implicar que al momento de darse una inversión de la curva (o incluso antes, cuando la misma se está aplanando), los gobiernos no deberían demorarse en invertir los recursos necesarios para que la curva adopte nuevamente una forma ascendente. De lo contario, cuánto más de dilate su intervención, más tiempo estará la economía sumida en una recesión.

Por último, cabría preguntarse si a futuro la *yield curve* seguirá siendo un indicador apropiado para pronosticar recesiones. Un primer cuestionamiento apareció en 2020, con la recesión que aconteció de manera simultánea a la pandemia de Covid-19. Quienes critican el valor de la curva de rendimientos como instrumento de seguimiento de la actividad económica arguyen que no era posible que la inversión de fines de 2019 haya anticipado la posterior caída en el PBI (tanto en EEUU como en el resto del mundo) el año siguiente. Sin embargo, la *repo crisis* de septiembre de 2019 parece sugerir lo contrario: la economía norteamericana ya presentaba evidentes problemas desde antes de que surgiesen los primeros casos de Covid-19.

También es cierto que existe el antecedente del “falso positivo” de 1965-1967, cuando la curva se invirtió por dos años, pero luego no sobrevino ninguna recesión. Este caso, como el anterior de la crisis de 2020, podría afectar la probabilidad de que una recesión sobrevenga tras una inversión de la curva. No obstante, ese no es el tema de investigación de esta Tesis Final. En este trabajo se plantearon dos hipótesis específicas, una de las cuales debería ser rechazada en función de los resultados obtenidos, y otra que en principio se vería ratificada.

En relación a esta última, sí podría objetarse que no puede establecerse un vínculo causal entre la duración de una inversión de la curva y la duración de las recesiones. Pero no es eso lo que aquí se afirma. De hecho, ya de por sí es prácticamente imposible poder crear un modelo de regresión con múltiples variables que dé cuenta de cuáles son las causas de (la duración de) una recesión. Un modelo simple estaría aún más alejado de este objetivo.

Es inevitable que se omitan numerosas variables. Y los análisis de regresión aquí realizados se ejecutaron con conciencia de que se verían afectados por un sesgo de variable omitida. Si no se tuviese presente esta cuestión, se plantearía una relación espuria entre las variables aquí consideradas.

Pero lo que concluye esta investigación es que existe una correlación positiva entre la duración de la inversión de la curva y la duración de las recesiones. Es decir, que cuánto más duradero es un *spread* negativo, es esperable que más prolongada será la recesión posterior. Es en definitiva una afirmación sencilla, pero llena de implicancias para la comprensión de la economía y la toma por parte de decisiones de los agentes económicos.

Por último, un especialista informado podría observar que en los últimos cuarenta años, las recesiones en EEUU se han tornado menos frecuentes que en los treinta y cinco años inmediatamente posteriores a la finalización de la SGM. Es cierto que la economía estadounidense, al igual que la de muchos otros países, experimentó un potente proceso de crecimiento en las tres décadas que siguieron a la última conflagración mundial. Ese proceso se caracterizó tanto por elevadas tasas de crecimiento, como por cambiantes ciclos económicos y relativamente altas tasas de inflación.

Bajo la dirección de Paul Volcker como presidente de la Reserva Federal, el crecimiento de la economía norteamericana se desaceleró en promedio. Pero este crecimiento se volvió menos fluctuante, y no se padeció una inflación tan alta. Asimismo, las recesiones efectivamente se volvieron menos asiduas.

Esta observación no hace sino reforzar la validez de las conclusiones a las que se ha arribado en esta investigación. Porque el período de tiempo considerado es extenso, aproximadamente coincidente con los tres cuartos de siglo posteriores a la finalización de la SGM en que la economía de EEUU se convierte en la más importante del sistema internacional.

Y ese período se divide a su vez en distintos subperíodos en los que tanto la economía norteamericana como la de los demás países experimentaron transformaciones estructurales. A pesar de que la frecuencia con la que EEUU sufre recesiones ha disminuido, todas estas recesiones del último medio siglo (desde 1969 en adelante) han acontecido con posterioridad a una inversión de la curva de rendimientos.

El conjunto de datos con el que se trabajó abarca toda esta etapa histórica de setenta años. Y en los que respecta a la segunda pregunta de investigación, no se lo manipuló (no se excluyeron observaciones). Por lo tanto, se consolida la hipótesis que concluye que existe una correlación positiva entre la duración de una inversión dela curva de rendimientos y la duración de la posterior recesión.

VII. Referencias bibliográficas

Claessens, S. (2009). What is a recession? *Finance & Development, 46*(1).

Clinton, K. (1995). The term structure of interest rates as a leading indicator of economic activity: a technical note. *Bank of Canada Review, 94-95*, 23-40.

Federal Reserve Bank of St. Louis. (11 de Octubre de 2018). *The data behind the fear of yield curve inversions*. Obtenido de Federal Reserve Economic Data. St. Louis Fed: https://fredblog.stlouisfed.org/2018/10/the-data-behind-the-fear-of-yield-curve-inversions/?utm_source=series_page&utm_medium=related_content&utm_term=related_resources&utm_campaign=fredblog

Fisher, M. (2004). Modeling the term structure of interest rates: an introduction. *Federal Reserve Bank of Atlanta Economic Review, 89*(3), 41-62.

Hansen, B. (2020). *Econometrics.* Princeton: Princeton University Press.

Harvey, C. R. (1986). *Recovering expectations of consumption growth from an equilibrium model of the term structure of interest rates.* Chicago: University of Chicago Booth School of Business.

Johnson, J. M. (1984). When are zero coupon bonds the better buy? *The Journal of Portfolio Management, 10*(3), 36-41.

Lonski, J. (2019). *The Fed cured 1998’s yield curve inversion.* Moody's Analytics Capital Markets Research Weekly Markets Outlook.

McCallum, B. T. (2005). Monetary policy and the term structure of interest rates. *Federal Reserve Bank of Richmond Economic Quarterly, 91*(4), 1-21.

National Bureau of Economic Research. (2020). *Business cycle dating*. Recuperado el 6 de Septiembre de 2022, de National Bureau of Economic Research: https://www.nber.org/research/business-cycle-dating

Nymand-Andersen, P. (2018). *Yield curve modelling and a conceptual framework for estimating yield curves: evidence from the European Central Bank's yield curves.* European Central Bank Statistics Series.

Rennison, J. (21 de Julio de 2022). A recession alarm is ringing on Wall Street. *The New York Times*.

Stradi, B. A. (2005). Term structure of interest rates. *Revista de Matemática: Teoría y Aplicaciones, 12*(1 & 2), 129-138.

Wang, X. H., & Yang, B. Z. (2011). A note on inverted yield curve and recession. *Management and Economics IPEDR, 3*, 72-75.

Wooldridge, J. M. (2015). *Introductory econometrics. A modern approach.* Boston: Cengage Learning EMEA.

Wright, J. H. (2006). *The yield curve and predicting recessions.* Board of Governors of the Federal Reserve System Finance and Economics Discussion Series.

VIII. Siglas y acrónimos

BLUE: *Best Linear Unbiased Estimator*

EE: Error Estándar

EEUU: Estados Unidos

ETTI: Estructura Temporal de Tasas de Interés

IRR: *Internal Rate of Return*

MCO: Mínimos Cuadrados Ordinarios

MELI: Mejor Estimador Lineal Insesgado

NBER: *National Bureau of Economic Research*

OLS: *Ordinary Least Squares*

OVB: *Omitted Variable Bias* (Sesgo de Variable Omitida)

PBI: Producto Bruto Interno

R^2: Coeficiente de Determinación

SEC: Suma Explicada de Cuadrados

SGM: Segunda Guerra Mundial

SRC: Suma Residual de Cuadrados

STC: Suma Total de Cuadrados

TIR: Tasa Interna de Retorno

IX. Apéndices

IX.1 Código en R

```r
# Universidad Nacional de Rosario
# Facultad de Ciencias Económicas y Estadística
# Maestría en Finanzas
# Tesis Final
# Esteban Hernán Smolarz

# Código en lenguaje R para la ejecución de análisis de regresión

# En primer lugar debe cargarse el data set que contiene la información sobre
# los spreads entre los bonos de corto y largo plazo, así como las tasas de
# crecimiento (negativo) del PBI.
# El data set del archivo en formato CSV es renombrado como "datosfed":

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
getwd()
datosfed <- read.csv("spreadsandgdp.csv",header=FALSE, sep=",")
datosfed

# Se asignan nombres a las columnas, correspondientes a las variables:

colnames(datosfed) <- c("pbi", "spread")
datosfed

# Se obtiene información estadística tanto de la variable independiente como de
# la variable dependiente, incluyendo promedio, varianza y desvío estándar:

mean(datosfed$pbi)
mean(datosfed$spread)
var(datosfed$pbi)
var(datosfed$spread)
sd(datosfed$pbi)
sd(datosfed$spread)

# Adicionalmente, se calcula el rango intercuartil de ambas variables:

```

```
summary(datosfed)

# Opcionalmente, puede graficarse el histograma de ambas variables:

hist(datosfed$spread, xlab = 'Spread 10Y1Y')
hist(datosfed$pbi, xlab = 'Tasa PBI')

# La relación entre ambas variables puede representarse por medio de un diagrama
# de dispersión:

plot(datosfed$spread,datosfed$pbi, xlab = 'Spread 10Y1Y', ylab = 'Tasa PBI')

# Se agrega una línea de regresión al diagrama de dispersión (lo cual ya implica
# predefinir el modelo de regresión):

abline(lm(pbi ~ spread, data = datosfed), col = "blue")

# Se define formalmente un modelo de regresión lineal simple:

modelo_simple_1 = lm(pbi ~ spread, data = datosfed)

# A continuación, se obtienen los principales valores estadísticos de este
# modelo (output del análisis de regresión):

summary(modelo_simple_1)

# Se advierte que la línea de regresión tiene pendiente negativa (o lo que
# es lo mismo, el coeficiente de la variable independiente tiene signo
# negativo), lo que indicaría la existencia de una correlación negativa entre
# ambas variables.

# Puede plantearse una variante del modelo que no contemple los datos más
# extremos.
```

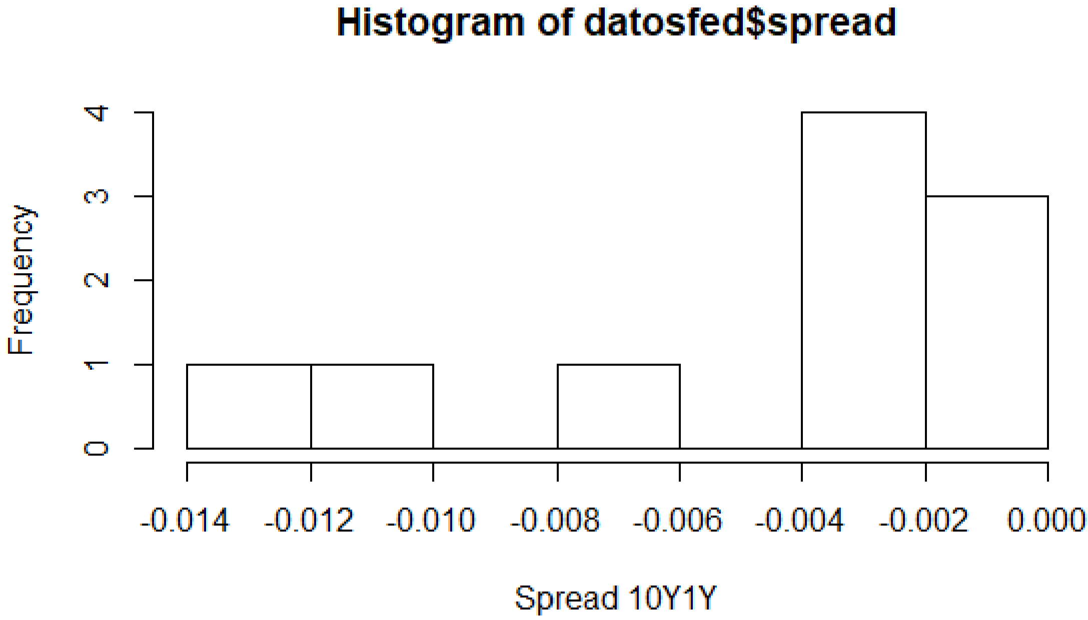
```r
# Con ese objeto, es necesario depurar el data set:

datosfed2 <- datosfed[-c(1, 10), ]
datosfed2

# Y repetir el proceso:

mean(datosfed2$pbi)
mean(datosfed2$spread)
var(datosfed2$pbi)
var(datosfed2$spread)
sd(datosfed2$pbi)
sd(datosfed2$spread)

summary(datosfed2)

hist(datosfed2$spread, xlab = 'Spread 10Y1Y')
hist(datosfed2$pbi, xlab = 'Tasa PBI')

plot(datosfed2$spread,datosfed2$pbi, xlab = 'Spread 10Y1Y', ylab = 'Tasa PBI')
abline(lm(pbi ~ spread, data = datosfed2), col = "red")

modelo_simple_2 = lm(pbi ~ spread, data = datosfed2)

summary(modelo_simple_2)

# Finalmente, cabe probar un tercer modelo, esta vez de regresión múltiple:

# Primero, es menester delimitar el nuevo data set:

datosfed3 <- datosfed2
datosfed3

# Debe añadirse la nueva variable al data set (el cuadrado de la variable
# independiente):

datosfed3$spreadcuadrado = datosfed3$spread^2
datosfed3$spreadcuadrado
datosfed3

mean(datosfed3$pbi)
mean(datosfed3$spread)
mean(datosfed3$spreadcuadrado)
var(datosfed3$pbi)
var(datosfed3$spread)
var(datosfed3$spreadcuadrado)
sd(datosfed3$pbi)
sd(datosfed3$spread)
sd(datosfed3$spreadcuadrado)

summary(datosfed3)

hist(datosfed3$spread, xlab = 'Spread 10Y1Y')
hist(datosfed3$spreadcuadrado, xlab = 'Spread 10Y1Y^2')
hist(datosfed3$pbi, xlab = 'Tasa PBI')

plot(datosfed3$spreadcuadrado,datosfed3$pbi, xlab = 'Spread 10Y1Y^2', ylab =
       'Tasa PBI')
abline(lm(pbi ~ spreadcuadrado, data = datosfed3), col = "green")

modelo_multiple = lm(pbi ~ spread + spreadcuadrado, data = datosfed3)

summary(modelo_multiple)
```

```
# Por último, se ejecuta el análisis de regresión que vincula el lapso durante
# el cual la curva de rendimientos está invertida, con la duración en el
# tiempo de la recesión que le sigue.

# En primera instancia debe cargarse el data set que contiene la información
# pertinente.
# El data set del archivo en formato CSV es renombrado como "datosfed4":

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
getwd()
datosfed4 <- read.csv("timerelation.csv",header=FALSE, sep=",")
datosfed4

# Nuevamente se asignan nombres a las columnas, correspondientes a las
# variables:

colnames(datosfed4) <- c("recesiones", "curvasinvertidas")
datosfed4

# Al igual que en los análisis anteriores, se obtiene información estadística
# tanto de la variable independiente como de la variable dependiente:

mean(datosfed4$recesiones)
mean(datosfed4$curvasinvertidas)
var(datosfed4$recesiones)
var(datosfed4$curvasinvertidas)
sd(datosfed4$recesiones)
sd(datosfed4$curvasinvertidas)

summary(datosfed4)

# Y se representa gráficamente la relación entre ambas variables:

hist(datosfed4$curvasinvertidas, xlab = 'Duración Curvas Invertidas (en meses)')
hist(datosfed4$recesiones, xlab = 'Duración Recesiones (en meses)')

plot(datosfed4$curvasinvertidas,datosfed4$recesiones, xlab = 'Duración Curvas
     Invertidas (en meses)', ylab = 'Duración Recesiones (en meses)')

abline(lm(recesiones ~ curvasinvertidas, data = datosfed4), col = "orange")

# Concluyendo, se define formalmente el modelo de regresión lineal simple:

modelo_simple_3 = lm(recesiones ~ curvasinvertidas, data = datosfed4)

# Y se muestra el output:

summary(modelo_simple_3)
```

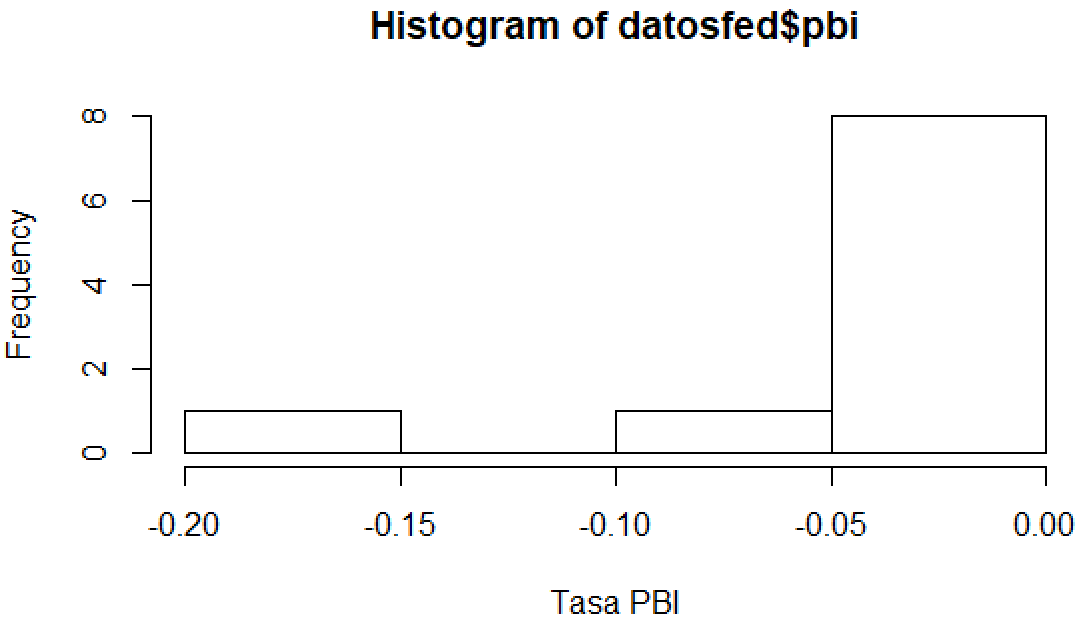
IX.2 Código en R con *output*

```
[Workspace loaded from C:/Esteban/Facultad/UNR/Ciencias Económicas y Estadística/Tesis/Dat
a/.RData]

> setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
> getwd()
[1] "C:/Esteban/Facultad/UNR/Ciencias Económicas y Estadística/Tesis/Data"
> datosfed <- read.csv("spreadsandgdp.csv",header=FALSE, sep=",")
> datosfed
        V1      V2
1  -0.0705 -0.0006
2  -0.0170 -0.0030
3  -0.0048 -0.0032
4  -0.0248 -0.0076
5  -0.0425 -0.0128
6  -0.0253 -0.0100
7  -0.0275 -0.0012
8  -0.0013 -0.0024
9  -0.0253 -0.0022
10 -0.1815 -0.0012

> colnames(datosfed) <- c("pbi", "spread")
> datosfed
       pbi  spread
1  -0.0705 -0.0006
2  -0.0170 -0.0030
3  -0.0048 -0.0032
4  -0.0248 -0.0076
5  -0.0425 -0.0128
6  -0.0253 -0.0100
7  -0.0275 -0.0012
8  -0.0013 -0.0024
9  -0.0253 -0.0022
10 -0.1815 -0.0012

> mean(datosfed$pbi)
[1] -0.04205
> mean(datosfed$spread)
[1] -0.00442
> var(datosfed$pbi)
[1] 0.002777992
> var(datosfed$spread)
[1] 1.770178e-05
> sd(datosfed$pbi)
[1] 0.05270666
> sd(datosfed$spread)
[1] 0.004207348
> summary(datosfed)
      pbi                spread
 Min.   :-0.18150   Min.   :-0.01280
 1st Qu.:-0.03875   1st Qu.:-0.00650
 Median :-0.02530   Median :-0.00270
 Mean   :-0.04205   Mean   :-0.00442
 3rd Qu.:-0.01895   3rd Qu.:-0.00145
 Max.   :-0.00130   Max.   :-0.00060
```

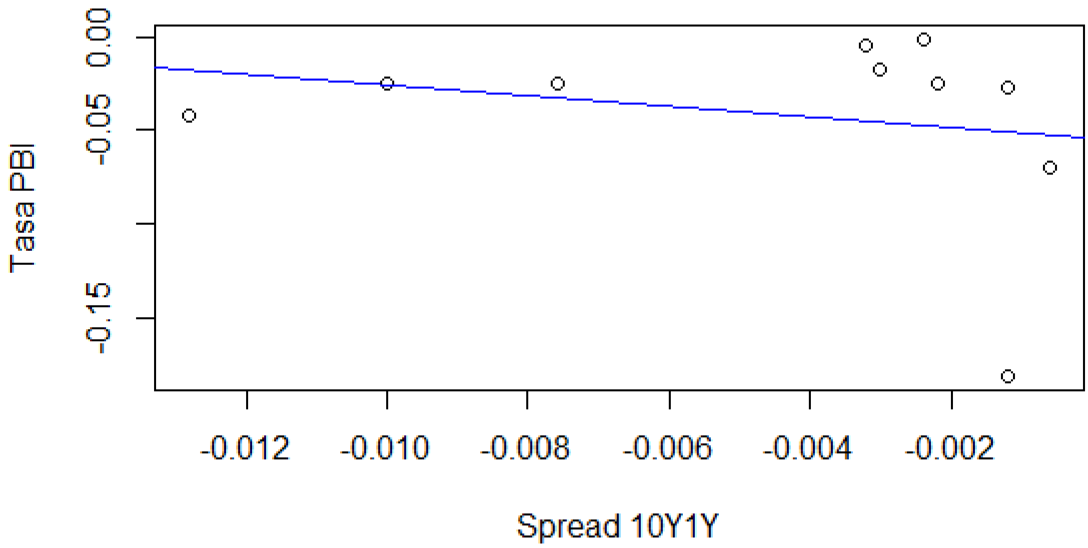
```
> hist(datosfed$spread, xlab = 'Spread 10Y1Y')
```




```
> hist(datosfed$pbi, xlab = 'Tasa PBI')
```

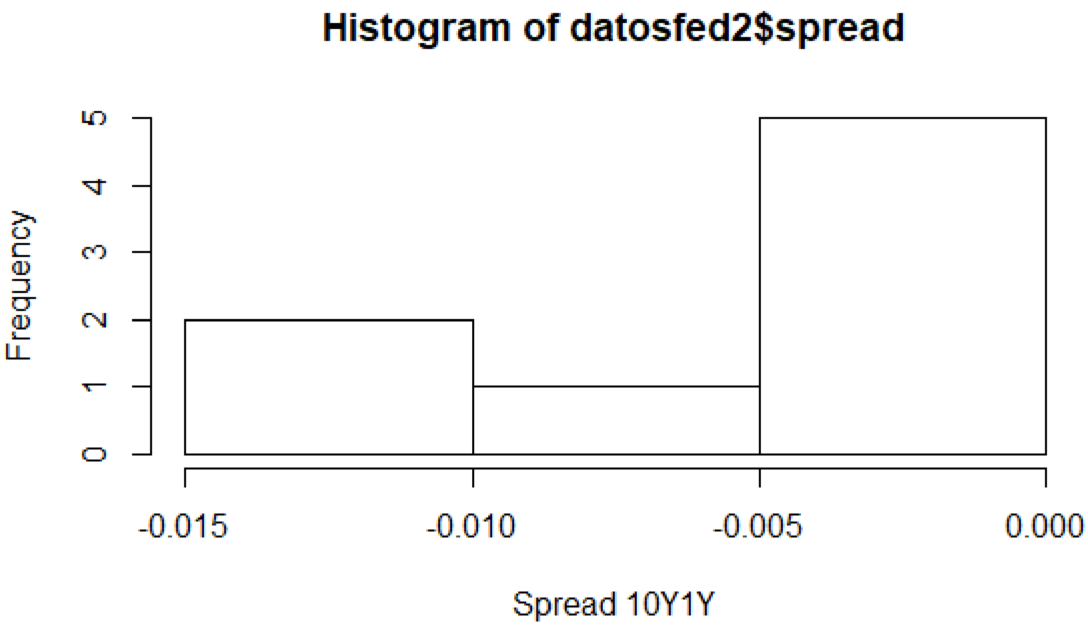


```
> plot(datosfed$spread,datosfed$pbi, xlab = 'Spread 10Y1Y', ylab = 'Tasa PBI')
> abline(lm(pbi ~ spread, data = datosfed), col = "blue")
```

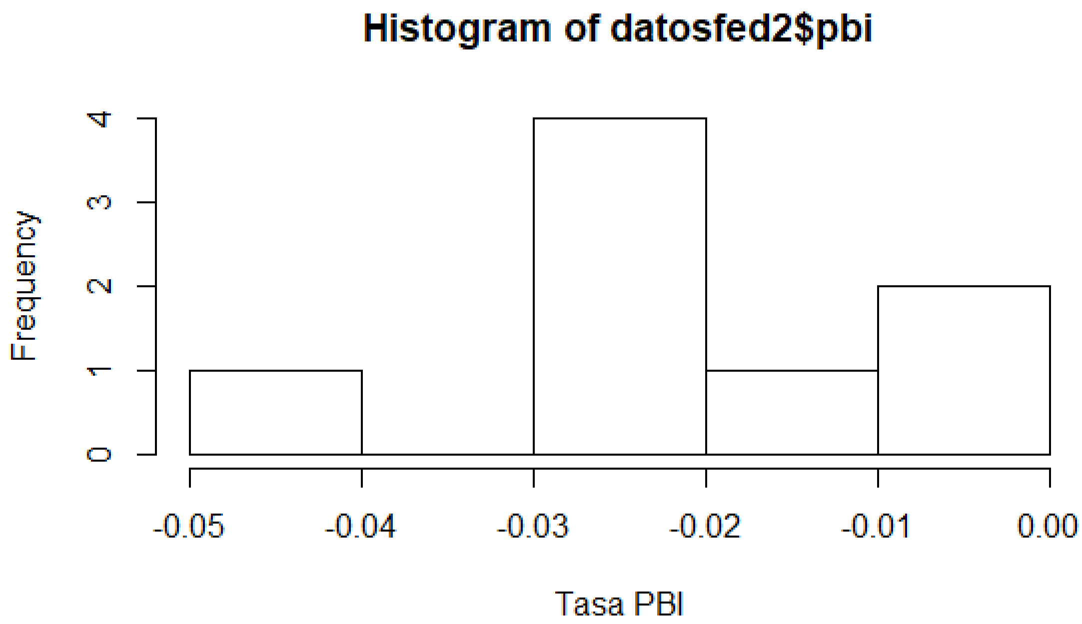



```
> modelo_simple_1 = lm(pbi ~ spread, data = datosfed)
> summary(modelo_simple_1)

Call:
lm(formula = pbi ~ spread, data = datosfed)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.13026 -0.01295  0.01563  0.02776  0.04652 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -0.05467    0.02568  -2.129   0.0659 .
spread      -2.85527    4.31249  -0.662   0.5265  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05443 on 8 degrees of freedom
Multiple R-squared:  0.05195,	Adjusted R-squared:  -0.06656 
F-statistic: 0.4384 on 1 and 8 DF,  p-value: 0.5265

> datosfed2 <- datosfed[-c(1, 10), ]
> datosfed2
      pbi  spread
2 -0.0170 -0.0030
3 -0.0048 -0.0032
4 -0.0248 -0.0076
5 -0.0425 -0.0128
6 -0.0253 -0.0100
7 -0.0275 -0.0012
8 -0.0013 -0.0024
9 -0.0253 -0.0022
```

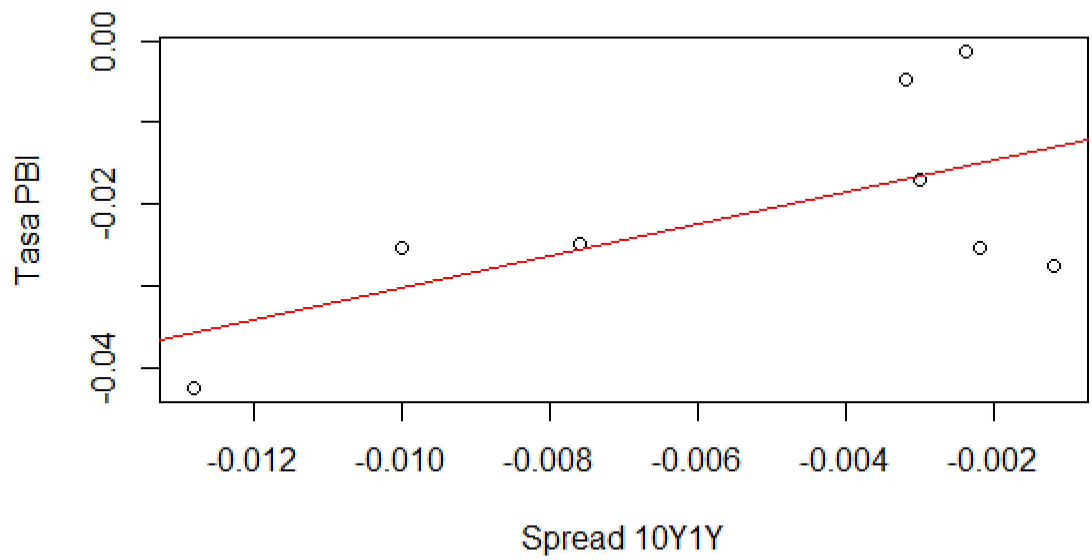
```
> mean(datosfed2$pbi)
[1] -0.0210625
> mean(datosfed2$spread)
[1] -0.0053
> var(datosfed2$pbi)
[1] 0.0001746313
> var(datosfed2$spread)
[1] 1.830857e-05
> sd(datosfed2$pbi)
[1] 0.01321481
> sd(datosfed2$spread)
[1] 0.004278852
> summary(datosfed2)
      pbi                  spread
 Min.   :-0.04250   Min.   :-0.01280
 1st Qu.:-0.02585   1st Qu.:-0.00820
 Median :-0.02505   Median :-0.00310
 Mean   :-0.02106   Mean   :-0.00530
 3rd Qu.:-0.01395   3rd Qu.:-0.00235
 Max.   :-0.00130   Max.   :-0.00120

> hist(datosfed2$spread, xlab = 'Spread 10Y1Y')
```



```
> hist(datosfed2$pbi, xlab = 'Tasa PBI')
```




```
> plot(datosfed2$spread,datosfed2$pbi, xlab = 'Spread 10Y1Y', ylab = 'Tasa PBI')
> abline(lm(pbi ~ spread, data = datosfed2), col = "red")
```

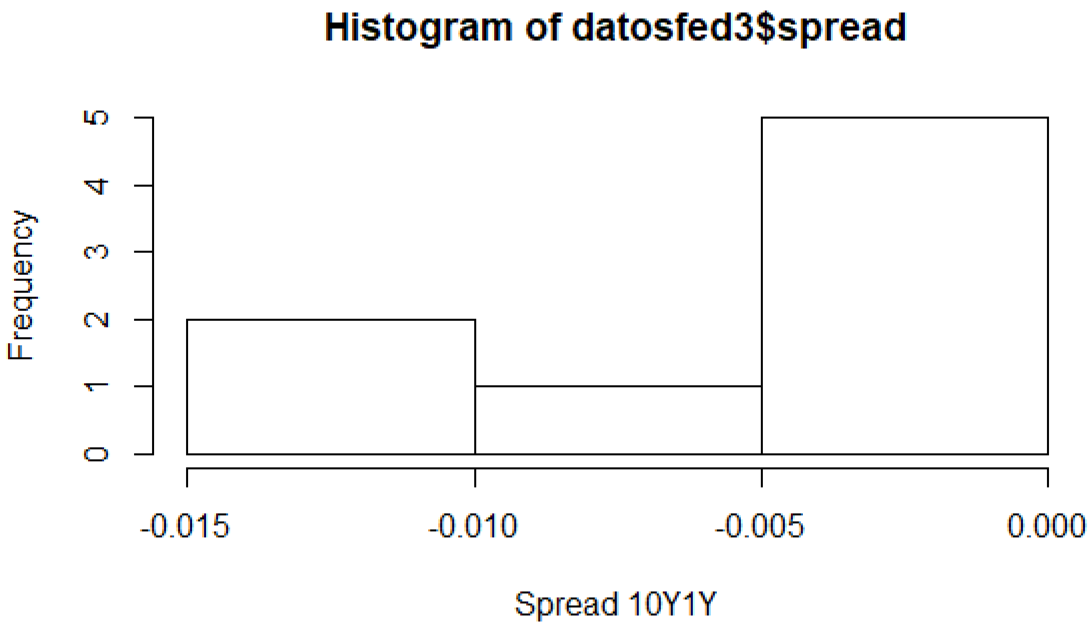


```
> modelo_simple_2 = lm(pbi ~ spread, data = datosfed2)
> summary(modelo_simple_2)

Call:
lm(formula = pbi ~ spread, data = datosfed2)

Residuals:
       Min         1Q     Median         3Q        Max 
-0.0144536 -0.0076551  0.0001625  0.0067529  0.0140926 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) -0.010700   0.006482  -1.651    0.150  
spread       1.955134   0.976019   2.003    0.092 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01105 on 6 degrees of freedom
Multiple R-squared:  0.4008,	Adjusted R-squared:  0.3009 
F-statistic: 4.013 on 1 and 6 DF,  p-value: 0.09202

> datosfed3 <- datosfed2
> datosfed3
      pbi  spread
2 -0.0170 -0.0030
3 -0.0048 -0.0032
4 -0.0248 -0.0076
5 -0.0425 -0.0128
6 -0.0253 -0.0100
7 -0.0275 -0.0012
8 -0.0013 -0.0024
9 -0.0253 -0.0022

> datosfed3$spreadcuadrado = datosfed3$spread^2
> datosfed3$spreadcuadrado
[1] 0.00000900 0.00001024 0.00005776 0.00016384 0.00010000 0.00000144 0.00000576
[8] 0.00000484
> datosfed3
      pbi  spread spreadcuadrado
2 -0.0170 -0.0030     0.00000900
3 -0.0048 -0.0032     0.00001024
4 -0.0248 -0.0076     0.00005776
5 -0.0425 -0.0128     0.00016384
6 -0.0253 -0.0100     0.00010000
7 -0.0275 -0.0012     0.00000144
8 -0.0013 -0.0024     0.00000576
9 -0.0253 -0.0022     0.00000484
```

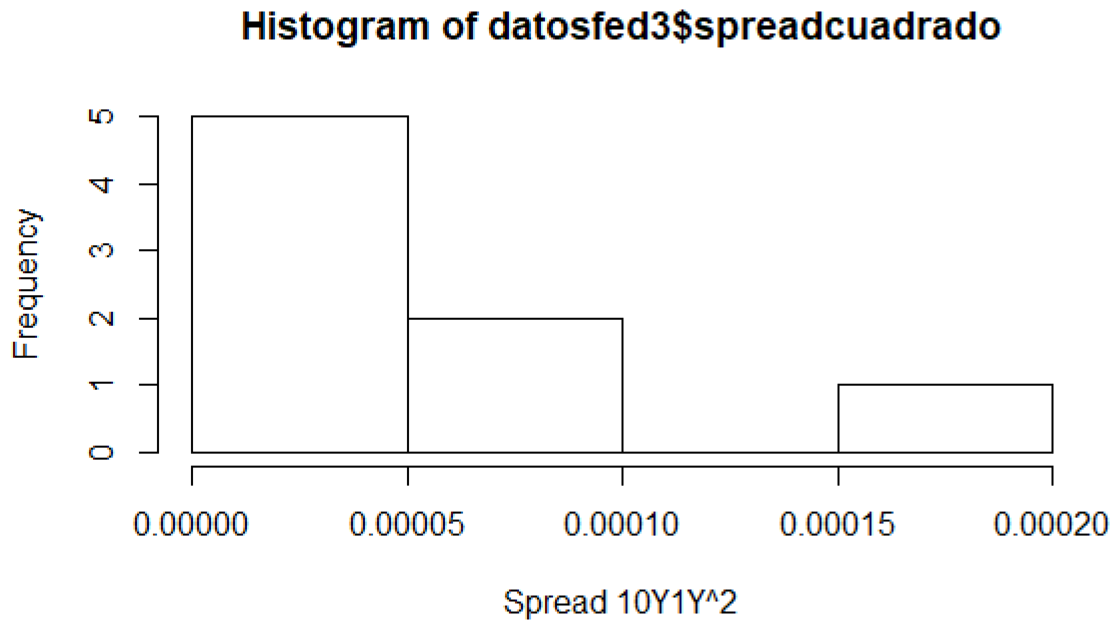
```
> mean(datosfed3$pbi)
[1] -0.0210625
> mean(datosfed3$spread)
[1] -0.0053
> mean(datosfed3$spreadcuadrado)
[1] 4.411e-05
> var(datosfed3$pbi)
[1] 0.0001746313
> var(datosfed3$spread)
[1] 1.830857e-05
> var(datosfed3$spreadcuadrado)
[1] 3.551252e-09
> sd(datosfed3$pbi)
[1] 0.01321481
> sd(datosfed3$spread)
[1] 0.004278852
> sd(datosfed3$spreadcuadrado)
[1] 5.959238e-05
```

```
> summary(datosfed3)
      pbi                spread          spreadcuadrado
 Min.   :-0.04250   Min.   :-0.01280   Min.   :1.440e-06
 1st Qu.:-0.02585   1st Qu.:-0.00820   1st Qu.:5.530e-06
 Median :-0.02505   Median :-0.00310   Median :9.620e-06
 Mean   :-0.02106   Mean   :-0.00530   Mean   :4.411e-05
 3rd Qu.:-0.01395   3rd Qu.:-0.00235   3rd Qu.:6.832e-05
 Max.   :-0.00130   Max.   :-0.00120   Max.   :1.638e-04
```

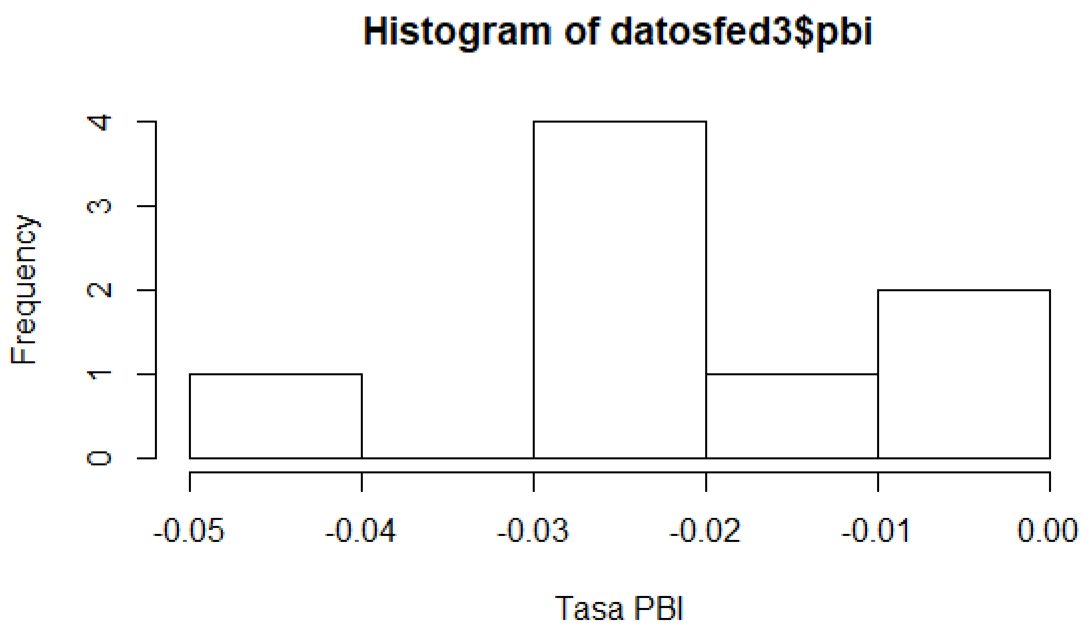
```
> hist(datosfed3$spread, xlab = 'Spread 10Y1Y')
```



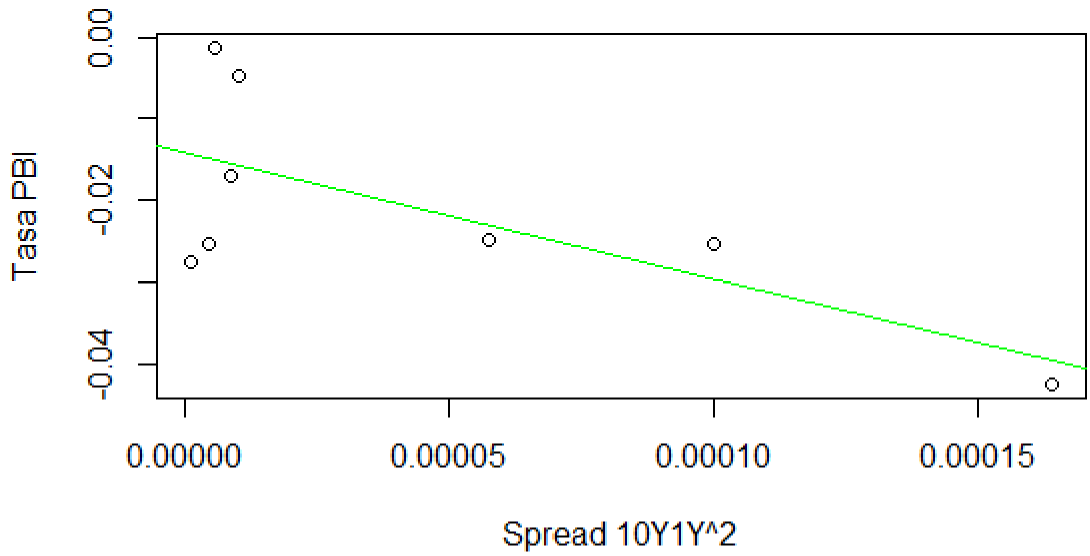
```
> hist(datosfed3$spreadcuadrado, xlab = 'Spread 10Y1Y^2')
```




```
> hist(datosfed3$pbi, xlab = 'Tasa PBI')
```

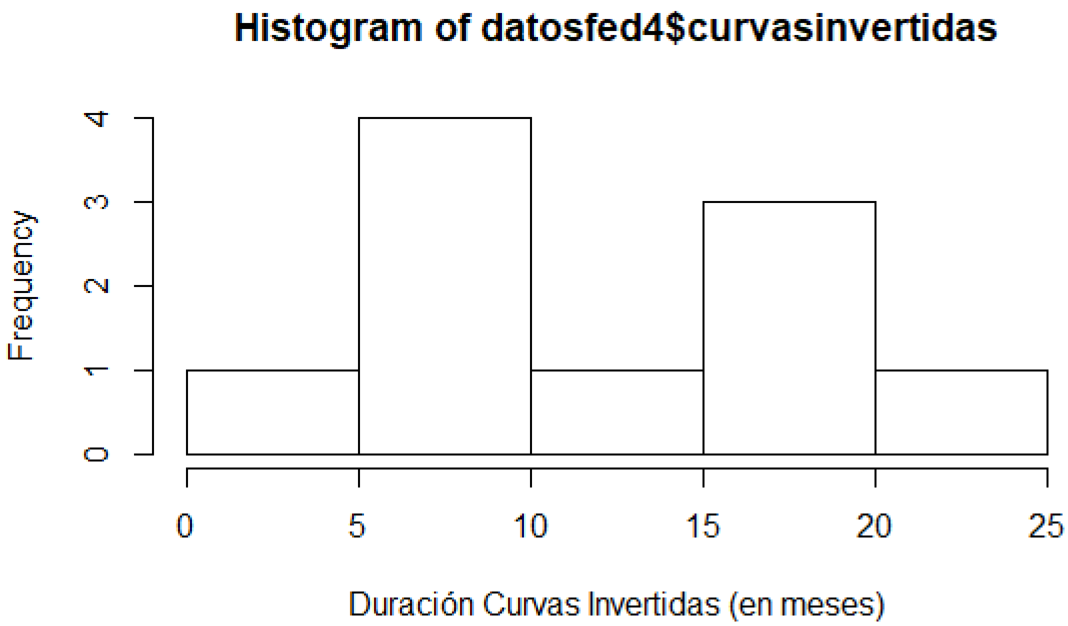


```
> plot(datosfed3$spreadcuadrado,datosfed3$pbi, xlab = 'Spread 10Y1Y^2', ylab =
+         'Tasa PBI')
> abline(lm(pbi ~ spreadcuadrado, data = datosfed3), col = "green")
```



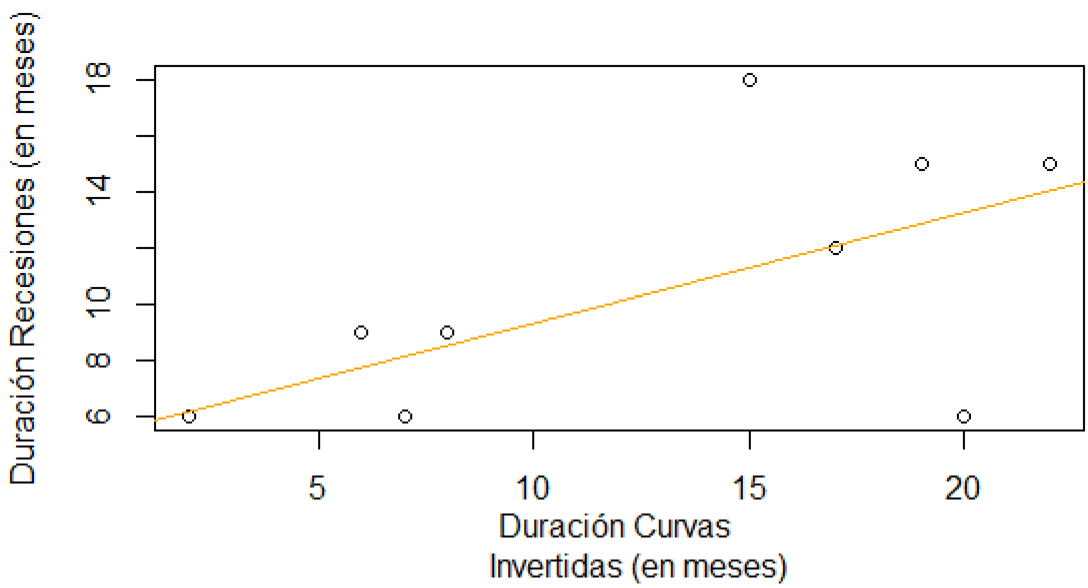

```
> modelo_multiple = lm(pbi ~ spread + spreadcuadrado, data = datosfed3)
> summary(modelo_multiple)

Call:
lm(formula = pbi ~ spread + spreadcuadrado, data = datosfed3)

Residuals:
         2          3          4          5          6          7          8          9
-0.0022028  0.0096685 -0.0078105  0.0015178  0.0009267 -0.0080109  0.0147143 -0.0088030

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.02435    0.01129  -2.157   0.0835 .
spread          -4.62997    4.70410  -0.984   0.3702
spreadcuadrado -481.75084  337.76376  -1.426   0.2131
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0102 on 5 degrees of freedom
Multiple R-squared:  0.5741,    Adjusted R-squared:  0.4037
F-statistic: 3.369 on 2 and 5 DF,  p-value: 0.1184

> setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
> getwd()
[1] "C:/Esteban/Facultad/UNR/Ciencias Económicas y Estadística/Tesis/Data"
> datosfed4 <- read.csv("timerelation.csv",header=FALSE, sep=",")
> datosfed4
   V1 V2
1   6  7
2   9  6
3  15 22
4  15 19
5   6 20
6  12 17
7   6  7
8   9  8
9  18 15
10  6  2
```

```
> colnames(datosfed4) <- c("recesiones", "curvasinvertidas")
> datosfed4
   recesiones curvasinvertidas
1           6                7
2           9                6
3          15               22
4          15               19
5           6               20
6          12               17
7           6                7
8           9                8
9          18               15
10          6                2
> mean(datosfed4$recesiones)
[1] 10.2
> mean(datosfed4$curvasinvertidas)
[1] 12.3
> var(datosfed4$recesiones)
[1] 20.4
> var(datosfed4$curvasinvertidas)
[1] 49.78889
> sd(datosfed4$recesiones)
[1] 4.516636
> sd(datosfed4$curvasinvertidas)
[1] 7.056124
> summary(datosfed4)
   recesiones     curvasinvertidas
 Min.   : 6.00   Min.   : 2.0
 1st Qu.: 6.00   1st Qu.: 7.0
 Median : 9.00   Median :11.5
 Mean   :10.20   Mean   :12.3
 3rd Qu.:14.25   3rd Qu.:18.5
 Max.   :18.00   Max.   :22.0

> hist(datosfed4$curvasinvertidas, xlab = 'Duración Curvas Invertidas (en meses)')
```



```
> hist(datosfed4$recesiones, xlab = 'Duración Recesiones (en meses)')
```


```
> plot(datosfed4$curvasinvertidas,datosfed4$recesiones, xlab = 'Duración Curvas
+      Invertidas (en meses)', ylab = 'Duración Recesiones (en meses)')
> abline(lm(recesiones ~ curvasinvertidas, data = datosfed4), col = "orange")
```



```
> modelo_simple_3 = lm(recesiones ~ curvasinvertidas, data = datosfed4)
> summary(modelo_simple_3)

Call:
lm(formula = recesiones ~ curvasinvertidas, data = datosfed4)

Residuals:
    Min      1Q  Median      3Q     Max
-7.2312 -1.6215  0.2213  1.2054  6.7371

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        5.3580     2.4990   2.144   0.0644 .
curvasinvertidas   0.3937     0.1785   2.206   0.0584 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.778 on 8 degrees of freedom
Multiple R-squared:  0.3782,    Adjusted R-squared:  0.3005
F-statistic: 4.866 on 1 and 8 DF,  p-value: 0.05845
```

IX.3 *Environment* en R