



**UNIVERSIDAD
TORCUATO DI TELLA**

Master in Management & Analytics

**Industria de la Energía en Datos:
Modelos de estimación para la
planificación del negocio**

Fecha: Julio 2019
Tutor: Marcelo Celani
Alumno: Emiliano Serrano

Índice

	Página
Resumen Ejecutivo	3
1 Consideraciones Iniciales	
I. Objeto de Estudio.....	4
II. Administración y empleo de los datos: Técnicas, formatos y herramientas utilizadas	5
III. Desarrollo.....	6
2 Análisis del contexto geo-político	
I. Pre-Revolución del Shale	7
II. Auge del Shale y Norteamérica.....	7
III. Impacto en el resto del mundo	8
IV. Conclusión Capítulo	9
3 Variables de la industria de la Energía	
I. Precio del Petróleo y Gas	11
II. Otros indicadores de la industria	12
III. Estudiando la relación entre indicadores: Fundamentos Estadísticos y técnicas utiliza	15
IV. Explicaciones de negocio a los valores de correlación.....	19
V. Conclusión Capítulo	20
4 Indicadores de Tenaris	
I. Fluctuaciones de las ventas OCTG	21
II. Distribución ventas Premium / No-Premium	23
III. Line Pipe	25
IV. Días de servicio de entubación y asistencia técnica on-site	27
5 Modelado para estimación de ventas	
I. Preparación de las variables seleccionadas	28
II. Fundamento teórico del test Augmented Dickey-Fuller	29
III. Resultados del Test e interpretación	30
IV. Modelización	33
V. Conclusión Capítulo	44
6 Tomando decisiones de negocio	
I. Decisiones de negocio en base la estimación de ventas	46
II. Decisiones de negocio en base la estimación de Días de Servicio.....	49
7 Conclusión	50
Anexos	52
Fuentes Bibliográficas	56

Resumen Ejecutivo

Hoy en día es imperioso para cualquier Compañía que provee materiales o servicios entender el comportamiento de las variables que pueden tener un impacto en la industria en la cual opera. Este comportamiento lo definen efectos endógenos como también exógenos, que terminan influyendo en el contexto y panorama final del mercado. Las variables que sufren estas fluctuaciones deben estar correctamente definidas para que sea posible estudiar su comportamiento y de esta manera definir estrategias en base a niveles estimados y escenarios futuros.

Al estudiar la variabilidad a lo largo del tiempo y entendiendo como eso impacta en los resultados del negocio es posible plantear, basándose en la historia y en fundamentos estadísticos, modelos que tomen valores futuros de dichas variables explicativas para estimar resultados de negocio. Por este motivo, el principal objetivo de este trabajo es analizar los valores históricos de las principales variables de la industria del gas y petróleo para cruzarlos con indicadores pasados de actividad de Tenaris logrando así encontrar modelos que permitan estimar, con cierto intervalo de confianza, una proporción de actividad futura.

De esta manera, con dicha estimación, desarrollar estrategias de negocio a mediano/largo plazo para distintas áreas de la Compañía que permitan planificar acciones tales como inversiones, contrataciones, previsiones, acopio de materia prima, etc. permitiendo así cumplir con dicho horizonte estimado de la manera más eficiente posible y maximizando resultados.

1. Consideraciones iniciales

I. Objeto de estudio

El estudio de este trabajo se focalizará en el mercado e industria Norteamericana ya que permitirá obtener apreciaciones respecto al mercado que, lenta pero sostenidamente, se transforma en la principal potencia en términos de generación de hidrocarburos.

En los gráficos a continuación se evidencia la creciente relevancia que el Norte de América tiene para Tenaris en los últimos años:

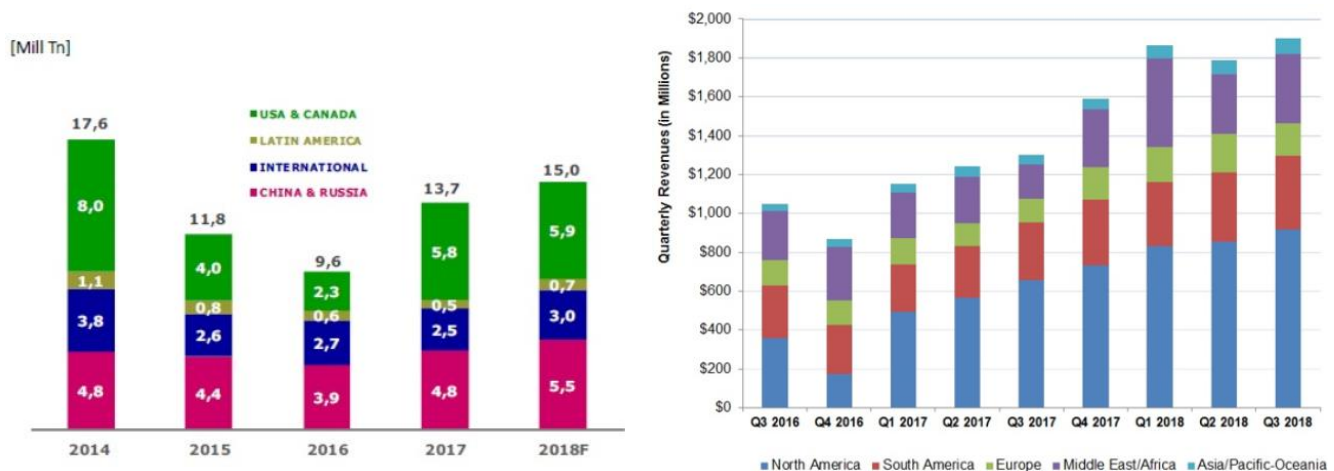


Figure 1: Actividad en Toneladas (izquierda) y en millones de USD (derecha) de Tenaris a nivel global.
Fuente: TS investor Presentation November 2018 y TS S.A Fillings

Por este motivo, no es casualidad que, las inversiones más grandes de dinero por parte de Tenaris en los últimos 3 años fueron puestas en Estados Unidos. Por un lado, la construcción de la planta de laminación y terminación en Bay City, Texas que significó una inversión de casi 2 billones de dólares sumado al reciente anuncio respecto a la compra de TMK IPSCO (sujeto a aprobación final de Anti Trust) por 1.2 billones. Lo que no hace más que demostrar la importancia y relevancia de este mercado para la compañía y evidencia como el foco está puesto allí.

II. Administración y empleo de los datos: Técnicas, formatos y herramientas utilizadas

Este trabajo cuenta con 3 grandes grupos de información:

- El primero, utilizado a lo largo de todo el trabajo, refiere a la bibliografía en temas de economía global, contexto geo-político, econometría, estadística descriptiva, métodos estadísticos, métodos de análisis de datos y fundamentos de programación.
- Por otro lado, información histórica de carácter público que refieren las principales variables de la industria del petróleo desde el 2013 en adelante (siendo Enero 2013 el punto de partida del estudio; hasta Marzo 2019).
- Por último, información de actividad histórica de Tenaris del mismo período. Por el Acuerdo de Confidencialidad firmado oportunamente con la Cía., no es posible mostrar los valores históricos reales pero si se demostrará gráficamente cómo fue la actividad a través de tendencias, comparaciones o análisis cualitativos y utilizar estas series como input para los modelos a desarrollar.

Una vez recabados los datos de sus distintas fuentes, la homologación y homogeneización se hizo en archivos .xls y .csv con formato tipo bases de datos para una mayor facilidad de manipulación. No fue necesario emplear técnicas de programación en R o Python para esta primer parte recién comentada ya que el volumen de datos no es excesivamente extenso.

De todas formas, siguiendo con el volumen de datos, puede determinarse que es suficiente para realizar el análisis que tiene como objetivo este trabajo teniendo en cuenta que cada serie cuenta con 72 valores ordenados cronológicamente de manera mensual.

Las técnicas de interpretación de tendencias y análisis de correlación fueron realizadas con distintas librerías en Python y tanto los test estadísticos como también los distintos modelos de regresión fueron hechos en R. Durante el desarrollo de la tesis se explicará más en detalle cada librería y el principal funcionamiento de los algoritmos utilizados.

Para las visualizaciones y gráficos, se utiliza el software PowerBI de Microsoft al cual se le introducen las bases en formato .xls directamente al programa.

Por último, los papers y los links a la información y metodologías utilizadas a lo largo de la tesis se encuentran disponibles en la bibliografía detallada al final del trabajo.

III. Desarrollo

El recorrido del trabajo comienza con un análisis geopolítico de la situación actual respecto a la industria de la energía, lo que sirve para brindar un panorama inicial de cuales han sido los principales drivers a nivel mundial en los últimos años para llegar al contexto en el cual la industria se encuentra hoy en día. Se describirán a los principales “jugadores” y la situación actual de cada uno para luego ahondar en el análisis de un mercado en particular que sea representativo para la industria y también para Tenaris.

Luego, en el siguiente capítulo, se analizarán las variables históricas de la actividad de la industria, desde comienzos de 2013, para entender la relación que guardan entre si y a la vez poder tener indicios de cuáles son las que afectan la actividad de una compañía proveedora.

En el capítulo 4, se expondrán las tendencias y fluctuaciones que ha tenido históricamente Tenaris en tres unidades de negocio que serán luego los indicadores que se intentarán estimar, en el capítulo 5, mediante la búsqueda y posterior análisis de modelos de regresión.

Una vez obtenidos estos modelos se detallarán las consideraciones de negocio que podrían hacerse por el hecho de contar con estimaciones de información futura y por último, en la conclusión, se volverá sobre las salvedades y excepciones para la correcta interpretación de este análisis permitiendo así sentar las bases para un análisis más profundo.

2. Análisis del contexto geo-político

I. Pre-Revolución del Shale

En 2008, el Consejo Nacional de Inteligencia (CNI) de los Estados Unidos publicó el informe Global Trends 2025 el cual tenía como premisa central el endurecimiento de la competencia internacional por la energía. China, acompañando al crecimiento de su industria y capacidad de producción para exportación, mantenía su alza en la demanda de hidrocarburos y regiones que proveían a dicha demanda (no pertenecientes a la OPEP*; como el Mar del Norte) se estaban agotando.

Tras dos décadas de precios bajos y estables, el valor de un barril de petróleo era de más de 100usd. En este contexto, se creía que se había llegado a un límite en la explotación, donde se afirmaba que la disponibilidad de reservas se estaba agotando y que quedaban alrededor de 50 años de extracción de hidrocarburos. A todo esto, la OPEP anticipaba que la producción se concentraría en países del Medio Oriente como Arabia Saudita, Irán e Irak aunque se creía incluso, que Arabia Saudita ya estaba totalmente explorada y que era muy poco probable que se hallaran nuevos grandes yacimientos. En este contexto, probado erróneo con el paso del tiempo, el CNI ponía a USA como un país que perdería influencia geopolítica por tener una creciente demanda de las importaciones de energía que venían encareciéndose.

II. Auge del Shale y Norteamérica

En aquel momento, debido a una tendencia global de energía verde y empujada por diferentes movimientos que iban tomando fuerza, los analistas del CNI consideraron la posibilidad de una revolución tecnológica. La realidad luego indicó que estaban pensando en la revolución equivocada. Al concentrarse en el potencial de fuentes renovables como la solar, la eólica y la hídrica, no vieron la verdadera revolución tecnológica que venía de la mano del esquisto (shale).

*OPEP: Organización de Países Exportadores de Petróleo.

A partir de esta tecnología, cambia la tendencia global: Estados Unidos realiza grandes inversiones de dinero para transformarse y dejar de ser un importador de energía para transformarse en exportador. El Depto. de Energía de USA estima que el país cuenta con 25 billones de metros cúbicos de shale gas, que combinados con otras fuentes de gas y petróleo, pueden durar dos siglos. Se prevé que Norteamérica alcance la autosuficiencia en la década de 2020. A su vez, instalaciones que se construyeron para recibir gas natural licuado se han reconvertido para procesarlo y generar exportaciones. En contraposición a los yacimientos tradicionales, los pozos y operaciones de shale son de menor escala, lo cual los hace más flexibles a variaciones de los precios regidos en el mercado. Esto significa que una inversión millonaria en un pozo tradicional no se puede detener y reiniciar fácilmente; por el contrario, los pozos de esquisto son de menores dimensiones, más baratos y más fáciles de activar/desactivar en respuesta a variaciones de precios. Esto implica que USA pasa a ser el productor bisagra o según el término utilizado en la industria el “dueño de la canilla”, capaz de equilibrar la oferta y demanda de los mercados globales de hidrocarburos.

Por último, distintos autores sostienen que esta nueva abundancia de energía aumenta el poder de Estados Unidos: impulsa la economía y crea empleo; mejora la balanza de pagos porque reduce las importaciones y aumenta la recaudación impositiva; fortaleciendo la competitividad, particularmente en industrias con alto consumo de energía como las petroquímicas, la del aluminio, la del acero, entre otras. No menor, y aparejado a este impulso que brinda la industria energética en la política Norteamericana, se evidencia una ola proteccionista llevada adelante por el gobierno actual que está generando una revolución del mercado interno, asegurando la protección del mercado local y así asegurando inversiones en sectores productivos que antes estaban acostumbrados a la dependencia de actores internacionales.

III. Impacto en el resto del mundo

Este shock no conoce fronteras. La transformación alcanza a todos los mercados. En épocas donde no se explotaban yacimientos del tipo shale, la dependencia de gasoductos limitaba el mercado del gas a lo que las distintas geografías podían ofrecer. Rusia tenía un alto poder de mercado, que usó para ejercer influencia política y económica sobre sus vecinos europeos. Estas

nuevas tecnologías, circuncidadas a licuar el gas extraído con plantas acorde, significan una flexibilización del mercado de GNL que redujo la influencia rusa. En números concretos, en 2005 sólo quince países importaban GNL; hoy son más del triple. Los beneficios encubiertos o que vienen derivados, gracias al shale, para la política exterior Norteamericana incluyen que países como Venezuela ya no puedan utilizar el petróleo para ejercer influencia en las Naciones Unidas, y que Rusia ya no pueda, al menos, tan fácilmente amenazar a sus vecinos con cortes de suministro de gas. Resumiendo, y en otras palabras, se evidencia un claro corrimiento en la cúpula del poder de la energía.

IV. Conclusión Capítulo

Los recursos energéticos como el petróleo y gas no son más considerados recursos escasos como lo eran antes. La “batalla” geopolítica no es más por poder acceder al mismo si no que lo que está en disputa es el Market Share. Fueron claros los intentos que tuvieron países como Arabia Saudita, allá por 2016, en inundar el mercado con la idea de perjudicar a nuevos competidores como USA.

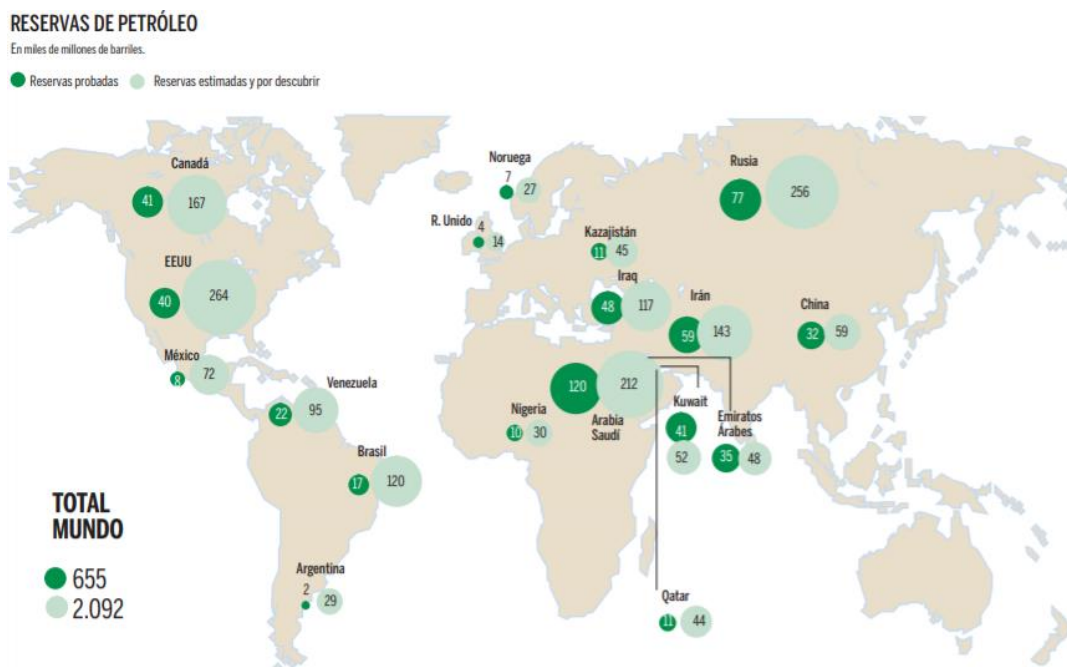


Figure 2: Reservas probadas y estimadas a nivel Mundial. Año 2016. [Fuente](#)

Por otro lado, La OPEP (cuyos países que la componen producen aproximadamente el 40% del crudo que se consume globalmente) se denota expectante y a la defensiva, registrando en Marzo 2019 los niveles de producción más bajos desde plena época de crisis en Febrero del 2015. Esto se debe principalmente a fuertes bajas voluntarias de producción en Arabia Saudita, tal como prometido por el ministro de energía Saudí con el fin de reducir los niveles de inventarios y así evitar una nueva baja en los precios, y a la vez acrecentado por los apagones y deficiencias estructurales en Venezuela que redujeron un 35% la producción del país latino.

Este contexto sigue beneficiando a USA como productor ya que gana Market Share y se apunta lentamente en la primera posición de productor de petróleo y gas a nivel global. Durante los próximos años, puede esperarse que la estrategia y política Norteamericana se enfocará en la incesante búsqueda del autoabastecimiento energético para, entre otros motivos, evitar depender de las importaciones. Por este motivo, y dada la relevancia que este mercado tiene para Tenaris, más adelante se enfocará el análisis del trabajo en Estados Unidos.

3. Variables de la industria de la Energía

A continuación se expone el comportamiento histórico de las principales variables de la industria de la Energía. El objetivo no será explicar los valores en si ni los motivos de la evolución de estos indicadores sino más bien exponer con argumentos de negocios la potencial relación entre las variables que mueven el mercado.

I. Precio del Petróleo y Gas

A lo largo del análisis y exposición de la información de este trabajo, se estará considerando el precio del barril según el índice WTI (West Texas Intermediate) que es el precio por el cual se rige el mercado Norteamericano. El mismo, refiere a una clase de petróleo bruto que los economistas también utilizan como valor estándar para determinar el precio como materia prima en los mercados de intercambio, con contratos a largo plazo. Esta cotización se realiza en el New York Mercantile Exchange, conocida como la principal bolsa de materias primas a nivel mundial. Los otros dos tipos de petróleo bruto que sirven de referencia son el Brent, del Mar del Norte y el de Dubai, que rige para los países de la OPEP. Con respecto al Gas, el valor de referencia será el Henry Hub que se mide por millón de BTU (British Thermal Unit).

En los gráficos a continuación, se muestra la evolución del precio del barril de petróleo y del gas para poder dar una primera visión del comportamiento del mercado durante los últimos 6 años.

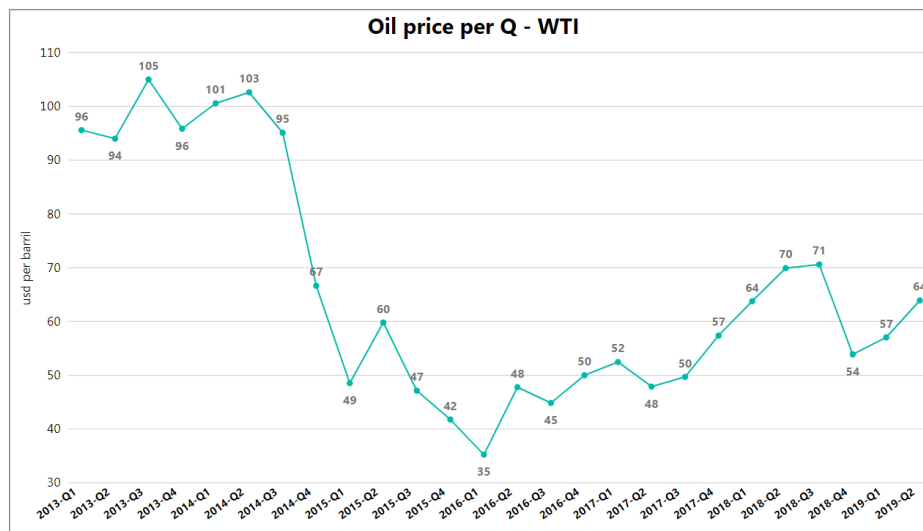


Figure 3: Evolución precio del barril WTI

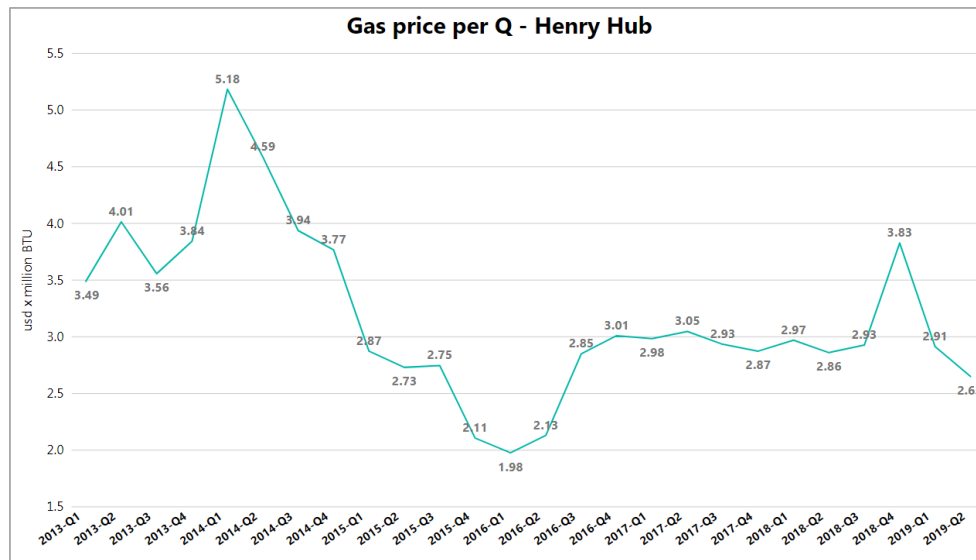


Figure 4: Evolución del precio del Gas

En ambos gráficos, puede apreciarse que el punto más bajo se encuentra en el primer cuarto del 2016 y que a fines del 2017 los valores (sobre todo del WTI) comienzan a subir por el reflote de la actividad de USA y la explotación de la nueva tecnología de extracción de hidrocarburos gracias al Shale.

II. Otros indicadores de la industria

Estos valores de referencia, que se encuentran entre los principales indicadores del mercado, luego impactan “aguas abajo” en las distintas variables que deben tenerse en cuenta y que dirigen las inversiones, estrategias y modelos de negocios que emplean las compañías que proveen a las petroleras o empresas de energía.

Una de las variables más significativas y observadas por el mercado es la cantidad de taladros (rigs) o plataformas que hay en actividad. Por rig se entiende a la torre de perforación que puede estar ya sea en etapa de perforación propiamente dicha o en producción. Durante la etapa de perforación es cuando se van colocando los tubos que permitirán luego pasar a la etapa de producción, que es cuando se extrae el hidrocarburo.

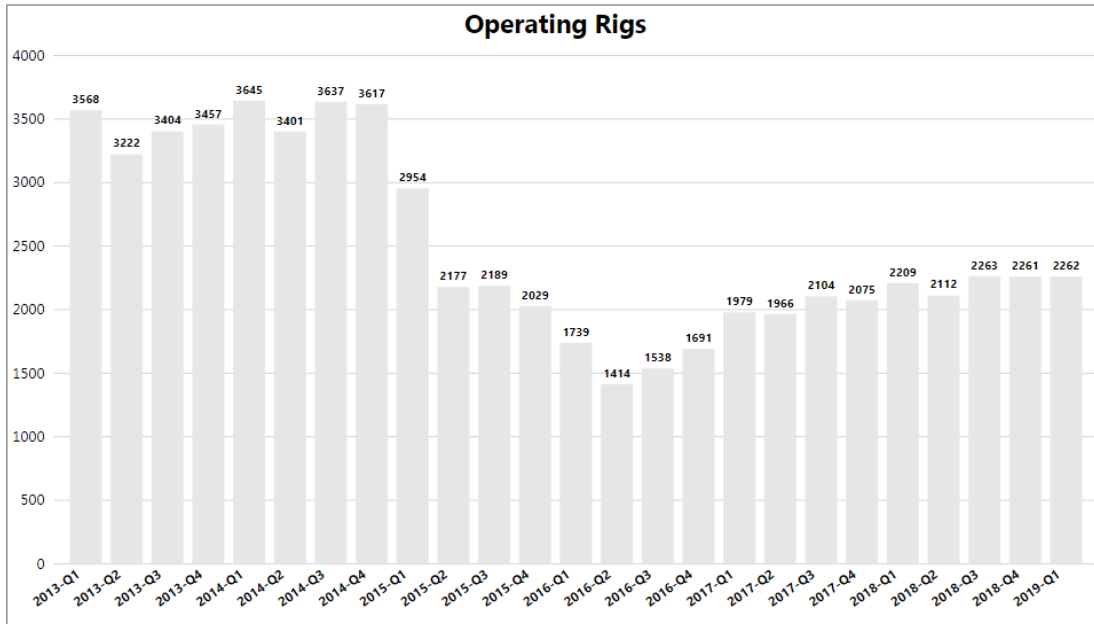


Figure 5: Evolución de Rigs en operación

A simple vista se ve que la caída en el precio del apartado anterior tiene una relación directa (que se estudiará luego con métodos estadísticos) con la cantidad de rigs que están en operación. Esto, visto desde el lado del negocio y de los inversores, es completamente entendible ya que un precio más bajo del barril se traduce a menos ganancia y por ende menos inversión en operaciones de perforación.

Al analizarse como fue esta evolución a nivel regional, y de esta manera adentrarse en el estudio de las variables del trabajo, a continuación se demuestra la cantidad de rigs en operación por período a nivel mundial, debe considerarse que no se muestra información de Rusia, de países del Caspio (Kazakstán, Azerbaiyán, Georgia, etc.) ni de China.

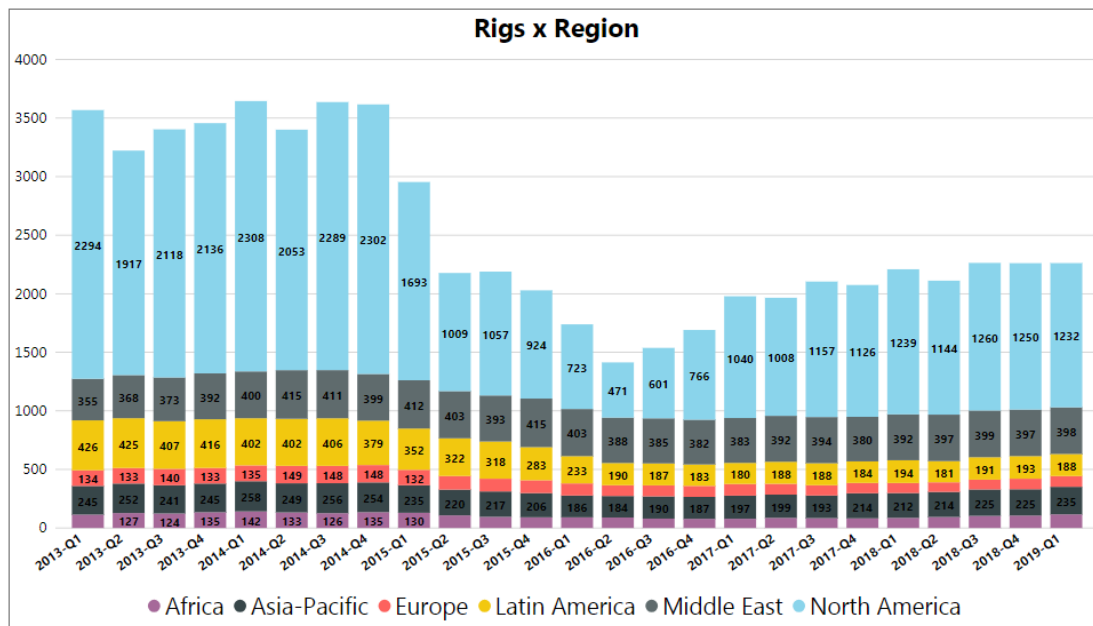


Figure 6: Evolución Rigs por región

El índice Pipe Logix es otro indicador que es sumamente importante para proveedores de la industria, especialmente para aquellos como Tenaris que proveen de tubería OCTG (Oil Country Tubular Goods). Este indicador provee de un índice medido en USD/tonelada en el cual confluyen las 38 categorías más populares de tubos entre las empresas petroleras. A su vez, se infiere que esto está directamente relacionado a precios vigentes de materia prima, siendo algunos de estos el valor del transporte de carbón, de las distintas aleaciones de acero, entre otras pero más que nada con la complejidad de proyectos que cada petrolera tenga en curso. Proyectos con condiciones extremas como marítimos, en el ártico o en superficies corrosivas requieren material más resistente y a la vez más complejos que en definitiva son más costosos. A continuación, la evolución del precio de productos OCTG utilizados en la industria.

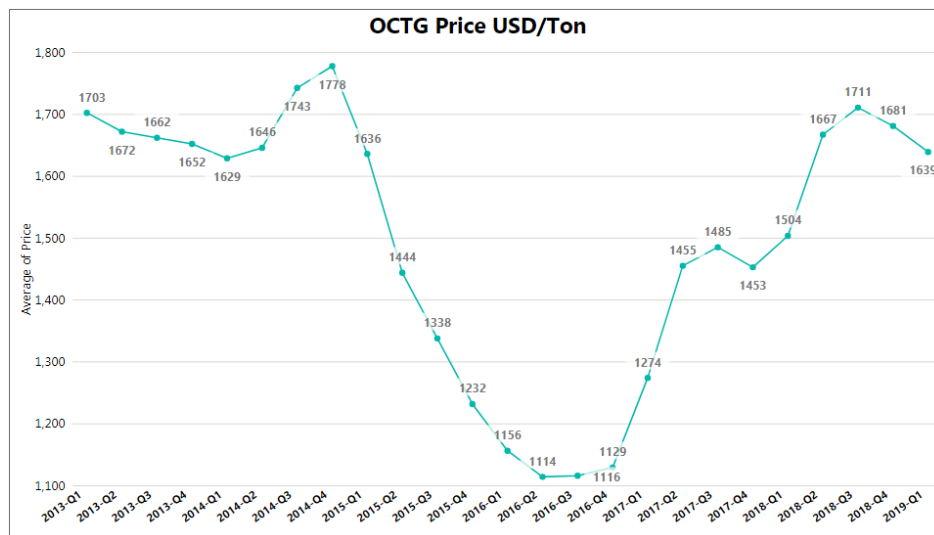


Figure 7: Evolución precio tubería OCTG

El principal comentario que surge de visualizar los gráficos respecto a estas variables expuestas es que la baja más grande se evidencia en el período donde la crisis del petróleo, mediados del 2015 y durante todo el 2016, se acentuó en mayor medida. Se evidencia que antes de la crisis, las compañías petroleras gastaban aproximadamente 1700 USD por tonelada y en época de crisis esos precios caen un 36%. En el próximo apartado se intentará demostrar esta relación mediante el uso de técnicas de análisis de datos.

III. Estudiando la relación entre indicadores: Fundamentos Estadísticos y técnicas utilizadas

La relación estadística entre dos, o más, variables se llama correlación. Esta correlación puede ser:

- Positiva: si las variables se mueven (incrementan o decrecen) en la misma dirección;
- Negativa: En el caso que las variables crezcan en direcciones opuestas; o
- Neutra: Las variables no tienen relación alguna.

Cuantitativamente hablando, se estará buscando el coeficiente de correlación de Pearson (ρ) el cual da una medida lineal entre dos variables. Este coeficiente nos permite independizarnos de la escala de medida de las variables a utilizar.

La forma en la cual se estudiará la relación entre las variables será por medio del armado de una matriz de correlación. En primera instancia, se estudiará cuán fuerte es la correlación entre las

variables del mercado, sin incluir en el análisis los resultados del negocio. De esta forma, partiendo en lo que de manera visual puede apreciarse en los gráficos históricos anteriormente presentados será posible demostrar estadísticamente que existen variables que sirven para explicar otras, y que por lo tanto pueden servir para luego cruzarlos con resultados de negocio tales como volumen de ventas, despachos, tipos de tubería vendida obtenidos a lo largo de los años de este estudio para así entender tendencias y definir posibles estrategias.

Los datos están expuestos con sus valores reales (no tasa de cambio) y en formato de serie de tiempos ordenados cronológicamente desde Enero del 2013 a Marzo de 2019 de forma mensual. Para realizar este análisis, se emplearon librerías de Python que facilitan el análisis y gestión de los datos como Pandas y Numpy. Para la generación de gráficos se utilizó la librería Matplotlib. En el anexo se encuentra el running code utilizado.

Debe tenerse en consideración que el dataset utilizado contiene muchos más datos que los demostrados anteriormente en los gráficos. El principal objetivo será investigar también si variables, a priori, menos relevantes sirven para entender cómo se comportan otras.

La lista de variables a analizar es:

Variable	Concepto
Pipe Logix (USD/Ton)	Precio Promedio
Average Scrap price (USD/Ton)	Precio de tonelada de chatarra. Principal materia prima para el acero
Average HRC (USD/Ton)	Hot Rolled Coil. Indicador del acero como materia prima
Financing Libor 60D (%)	Tasa de Interés a corto plazo entre bancos
Pig Iron USA (USD/Ton)	Precio de tonelada de Arrabio.
Iron Ore (USD/DMT)	Precio de mineral de hierro
Freight Coal Panamax Platts (USD/Mt)	Precio de transporte de buque de carbón
Gas Price (SD/MM Btu)	Precio por millón de BTU del gas
Oil WTI (USD/Bbl)	Precio del barril de petróleo WTI
Rig Count	Cantidad de Rigs en operación
Oil Brent (USD/Bbl)	Precio del barril de petróleo Brent

Figure 8: Variables de la industria de la energía

Las técnicas fueron codificadas en un notebook de Jupyter, código en anexo 1, arrojando los siguientes resultados:

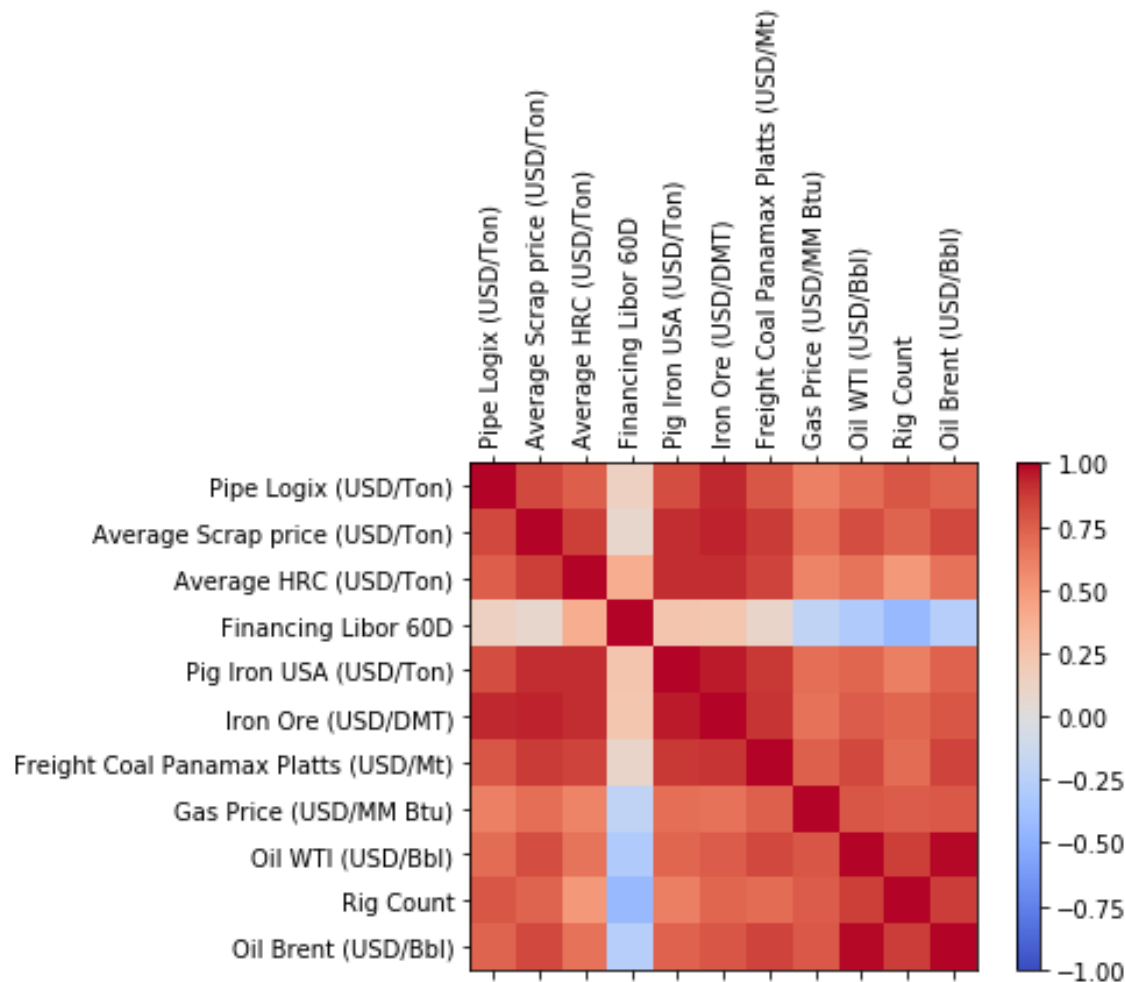


Figure 9: Matriz de correlación

En primera instancia, se presenta una matriz la cual se detalla la correlación según la tonalidad del color en la celda que une ambas variables. Como primera aclaración, en la diagonal el valor de correlación es 1 ya que es donde se intersecta la misma variable entre sí. Luego, cuánto más cerca del color rojo oscuro se encuentra coloreada la celda significa que las variables se mueven en la misma dirección y que la correlación es más cercana a 1, por lo tanto más fuerte. Por el contrario, la gama de los azules significa que la correlación tiene signo contrario por ende a medida que una variable crece la otra crece en dirección contraria. Así, los valores más cercanos a -1 indican una correlación negativa más fuerte. Por último los colores más claros, cercanos al

blanco, indican que los valores son cercanos a cero por lo que puede inferirse que la correlación es baja o nula.

Para una comparación más cuantitativa entre las variables, se expone debajo la grilla con los coeficientes de Pearson obtenidos. De esta manera será más sencillo explicar la correlación entre las variables estudiadas:

	Pipe Logix (USD/Ton)	Average Scrap price (USD/Ton)	Average HRC (USD/Ton)	Financing Libor 60D	Pig Iron USA (USD/Ton)	Iron Ore (USD/DMT)	Freight Coal Panamax Platts (USD/Mt)	Gas Price (USD/MM Btu)	Oil WTI (USD/Bbl)	Rig Count	Oil Brent (USD/Bbl)
Pipe Logix (USD/Ton)	1.000000	0.835214	0.750886	0.143044	0.818952	0.937091	0.782712	0.613477	0.701617	0.784736	0.731815
Average Scrap price (USD/Ton)	0.835214	1.000000	0.861851	0.088214	0.915972	0.939199	0.880619	0.688076	0.812588	0.729175	0.831415
Average HRC (USD/Ton)	0.750886	0.861851	1.000000	0.391910	0.916443	0.918391	0.849822	0.599549	0.671002	0.502279	0.676859
Financing Libor 60D	0.143044	0.088214	0.391910	1.000000	0.237617	0.228980	0.106945	-0.198187	-0.299235	-0.431724	-0.263981
Pig Iron USA (USD/Ton)	0.818952	0.915972	0.916443	0.237617	1.000000	0.956418	0.885832	0.691274	0.726362	0.621829	0.736373
Iron Ore (USD/DMT)	0.937091	0.939199	0.918391	0.228980	0.956418	1.000000	0.892630	0.679008	0.763488	0.719084	0.783089
Freight Coal Panamax Platts (USD/Mt)	0.782712	0.880619	0.849822	0.106945	0.885832	0.892630	1.000000	0.745047	0.830237	0.701423	0.849950
Gas Price (USD/MM Btu)	0.613477	0.688076	0.599549	-0.198187	0.691274	0.679008	0.745047	1.000000	0.783361	0.761566	0.779853
Oil WTI (USD/Bbl)	0.701617	0.812588	0.671002	-0.299235	0.726362	0.763488	0.830237	0.783361	1.000000	0.863562	0.991306
Rig Count	0.784736	0.729175	0.502279	-0.431724	0.621829	0.719084	0.701423	0.761566	0.863562	1.000000	0.869697
Oil Brent (USD/Bbl)	0.731815	0.831415	0.676859	-0.263981	0.736373	0.783089	0.849950	0.779853	0.991306	0.869697	1.000000

Figure 10: Matriz de correlación con coeficientes de Pearson

Se resalta en recuadros rojos aquellos valores con correlación directa más fuerte. En azul cuando la correlación es inversa y naranja donde no se ha encontrado un coeficiente lo suficientemente representativo.

Para poder ahondar en el análisis es necesario primero aclarar que hay una serie de variables que han dado una correlación fuerte pero que no merecen una explicación de negocio o mercado. Esto probablemente se deba a que ambas variables comparten tendencia estocástica pero no así correlación explicativa por hechos puntuales. Por otro lado, se ha dejado de lado un potencial efecto de time lag ya que la comparación se hace mes a mes entre los valores reales.

IV. Explicaciones de negocio a los valores de correlación

A continuación se detalla una explicación para todas aquellas correlaciones (tanto positivas fuertes como negativas) que se han encontrado, con el objetivo último de intentar dar una explicación de negocio y/o mercado que soporte el resultado estadístico obtenido.

ρ -> letra griega “rho” que detalla el coeficiente de Pearson entre esa variable y la que se está intentando explicar.

- Correlaciones de Pipe Logix

Como primer insight, respecto al negocio de la energía se ve una alta correlación ($\rho = 0.78$) con la cantidad de Rigs operando. Esto se puede entender como que a mayor cantidad de Rigs, se infiere no solo una mayor complejidad de los mismos ya sea por licitaciones más complejas del tipo HP/HT (high pressure/high temperatura) como también operaciones en plataformas marítimas o proyectos que requieren de una tubería más sofisticada.

En mucha menor medida, se podría conjeturar que el valor estándar del precio de las tuberías (Pipe Logix) está fuertemente relacionado con los indicadores que miden el precio de las materias primas necesarias para su producción y fabricación: el valor de la chatarra ($\rho = 0.84$), del mineral de hierro ($\rho = 0.82$) y del arrabio ($\rho = 0.93$), todos componentes fundamentales en una acería para producir el acero necesario. En este punto, es necesario aclarar que estas 3 variables seguramente no expliquen en una medida relevante el índice Pipe Logix ya que estos precios suelen derivar más bien de una relación comercial/contractual entre el proveedor y comprador por necesidades puntuales, que luego disparen una fabricación que viceversa.

Por último, sorprendentemente se evidencia una muy baja correlación con la tasa Libor ($\rho = 0.14$) por lo que no puede inferirse nada, al menos estadísticamente hablando.

- Correlaciones de Rig Count

Como era de esperarse, se evidencia una correlación muy fuerte ($\rho = 0.86$) con el valor del precio del barril WTI. Siguiendo el razonamiento que soporta la estadística, se espera que a mayor precio del petróleo más empresas estén invirtiendo dinero en operaciones de perforación, traduciendo esto a más rigs en operación intentando así maximizar la extracción del crudo para acrecentar

utilidades. En menor medida se evidencia una correlación ($\rho = 0.76$) con el precio del gas, explicado por los mismos motivos que fueron recién mencionados.

La correlación negativa ($\rho = -0.43$) con la tasa Libor a 60 días tiene que ver con que a menor tasa de interés de corto plazo, baja el dólar respecto a otras monedas y por ende suben los precios de los commodities que se miden en dólares y por lo tanto puede ser más provechoso extraer el petróleo. Por eso, y comentado un poco más a detalle al final del capítulo, cuando baja el dólar tiende a subir la actividad petrolera.

- **Correlaciones de precios de Barril WTI, Brent y Gas.**

La correlación más fuerte en el cuadro se da entre el precio del barril de WTI y el barril de Brent ($\rho = 0.99$) lo cual es de esperarse porque es sabido que son dos valores que se mueven prácticamente de la mano, aunque a escalas distintas.

Estas 3 variables tienen una correlación negativa con la Tasa Libor lo que induciría a afirmar que a mayor tasa, los precios del gas y petróleo tienden a caer.

V. Conclusión Capítulo

Como dicho al comienzo de este apartado, se busca dar una explicación de negocio a lo que se halló mediante el uso de técnicas estadísticas. Estas correlaciones pueden tener mayor o menor injerencia en el dato real y las tasas de cambio entre unas y otras, a sabiendas de que hoy por hoy muchos de estos valores no están solamente afectados por oferta, demanda y precios de materia prima sino también por decisiones comerciales, geo-políticas, hasta a veces arbitrarias, y por lo tanto sería de cierta forma incorrecto dar conclusiones deterministas meramente por resultados estadísticos. Como último comentario de este capítulo, basándose en un estudio* del 2017, vale recalcar la estrecha relación que hay entre los cambios en el precio del petróleo y la tasa de cambio de la moneda norteamericana. El estudio ahonda en el análisis del impacto a corto y largo plazo como también en el impacto por canales directos (dictado por las variaciones de la moneda) e indirectos (factores macroeconómicos y financieros). Dicha relación no se ha estudiado a detalle en el capítulo y por lo tanto el valor del barril será un input para el modelo que se quiere plantear.

*The Relationship between Oil Prices and Exchange Rates: Theory and Evidence. Beckman, Czudaj & Arora

4. Indicadores de Tenaris

En este capítulo se ahondará en la exposición y el análisis descriptivo de la actividad de Tenaris de manera de empezar a delinear su relación con las variables presentadas en el capítulo anterior. Esto permitirá afirmar las bases del análisis referido al impacto que han tenido las fluctuaciones históricas de las variables estudiadas con el desempeño de una de las compañías líder que proveen a la industria.

Con el fin de proveer un espectro más amplio del negocio, se expondrán los resultados (siempre con tendencias y niveles; no valores reales) de 4 facetas distintas de la actividad histórica de Tenaris:

- Ventas OCTG
- Distribución Premium / No- Premium
- Ventas Line Pipe
- Días de servicio de Entubación y asistencia técnica brindado por equipos de Tenaris on-site

La compañía cuenta con otras unidades de negocio que no serán tenidas en cuenta para este análisis como: varillas de bombeo, soluciones industriales, generación de energía y soluciones para el mercado automotriz. Los primeros son en cierta forma productos derivados que acompañan la venta de tubería que no serán incluidos en este análisis. Los otros tres son pertenecientes a otras industrias que no hacen al segmento que se está estudiando para el trabajo.

Para la exposición de los valores de la actividad de las ventas OCTG, distribución Premium / No-Premium y ventas de Line Pipe, se muestra a continuación la evolución de las ventas en toneladas y no en valores monetarios. Aunque la relación con las ganancias en términos económicos es prácticamente directa, de esta manera permite dar un mayor entendimiento del volumen de negocio que significarán luego los resultados obtenidos.

I. Fluctuaciones de las ventas OCTG

Para Tenaris, el negocio de OCTG es el más redituable y por lo tanto el que más recursos consume. A su vez, es donde está invertida la mayor parte de capital con respecto a la estructura de la Cía.

con respecto a instalaciones industriales, Research & development como en personal. Por otro lado, también es el más complejo de planificar por, como ha sido expuesto, la cantidad de variables exógenas que impactan y sobre todo porque los potenciales compradores en un mercado como el Norteamericano aparecen y desaparecen con mucha facilidad. Las principales compañías petroleras son más fáciles de predecir, sobre todo con las que se tienen acuerdos marcos firmados a largo plazo pero no así con pequeños productores que individualmente no son relevantes pero al agregarse si generan un impacto en el volumen total.

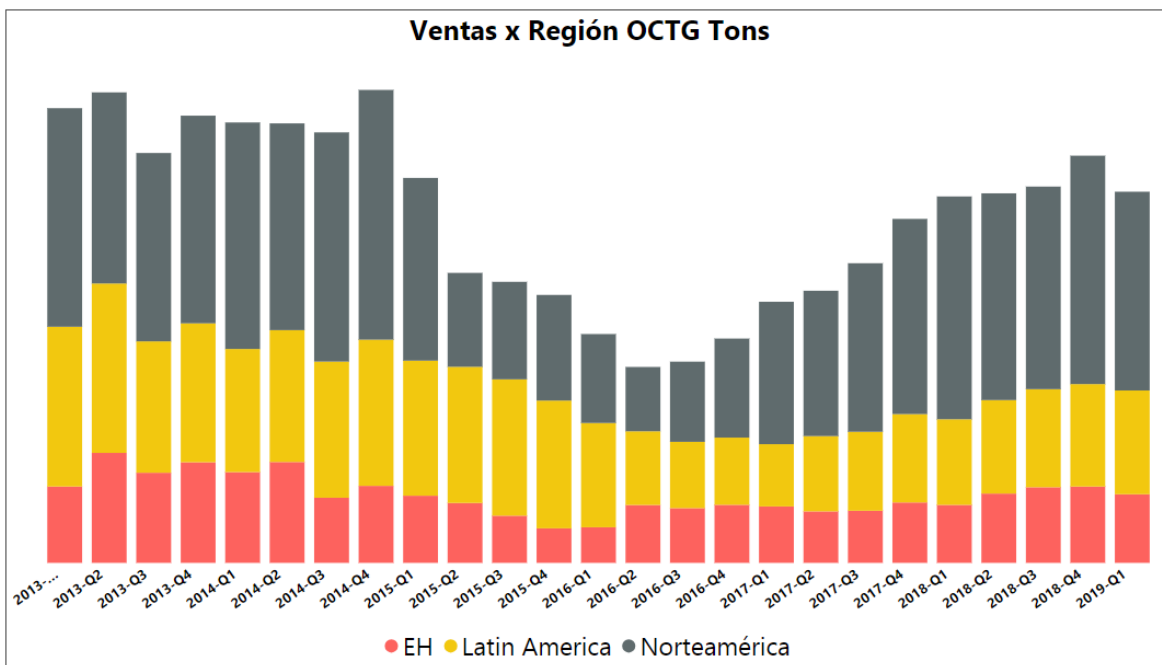


Figure 11: Evolución de las ventas de OCTG por región

Es necesario aclarar que dentro del grupo Norteamérica se encuentra también Canadá con el fin de facilitar el agrupamiento. Salvo en los años de crisis previo a la revolución del shale se puede ver claramente cómo, durante estos últimos 6 años, el mayor volumen de ventas para Tenaris se concentra en los países del norte, en línea con lo expuesto en el capítulo 2.

A su vez, no solo el volumen de ventas totales va en aumento en Norte América sino que también la proporción de lo que representa este mercado para la compañía va creciendo también, como puede verse en la siguiente figura:

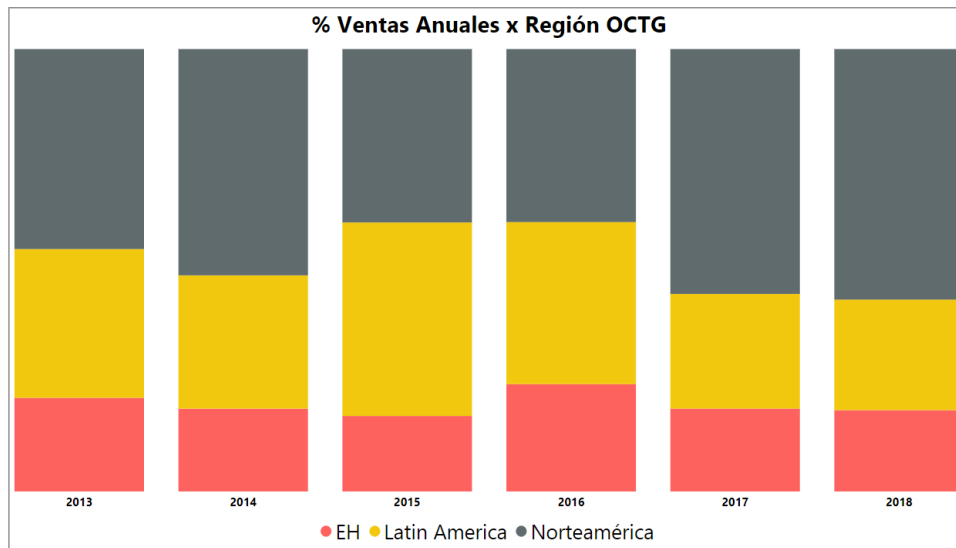


Figure 12: Evolución anual de la proporción por región de ventas OCTG

II. Distribución ventas Premium / No-Premium

En lo que respecta al resultado económico, no solo es interesante analizar el volumen de toneladas que Tenaris vende si no también la distribución de los productos que solicitan los clientes. Para simplificar y por temas de confidencialidad, se agrupan la gama de productos OCTG que Tenaris comercializa en dos: Premium y No-Premium.

- Tubería **Premium** engloban aquellos productos para los cuales Tenaris es propietario y posee las patentes de comercialización por haber desarrollado ya sea el grado de acero, la conexión, entre otras. En definitiva, es un producto que se desarrolló in-house en el departamento de Research & Development y que generalmente se utiliza para solicitudes de pozos más complejas, proyectos marítimos de aguas profundas o condiciones geográficas con alto grado de corrosión para la tubería. Esta tubería es más costosa y supone mejores márgenes para la compañía. Por otro lado, las plantas que desarrollan estos productos deben estar correctamente diseñadas y preparadas para cumplir con dicha producción y, por supuesto, contar con personal capacitado para operar.
- Por otro lado, por tubería **No-Premium** se entiende a que aquella que se fabrica según la norma API (American Petroleum Institute). Esta norma, mundialmente conocida, fija las condiciones y estándares de calidad bajo las cuales debe producirse. Se supone esta

tubería será solicitada a proyectos de menor complejidad, pozos menos profundos y por ende para operaciones menos riesgosas. Por este motivo, se asocia a un menor margen de ganancias por precios más bajos en relación a la tubería Premium.

El anterior gráfico representa la distribución de las ventas entre Tubería Premium y No-Premium solamente para Estados Unidos, región testigo para el análisis:

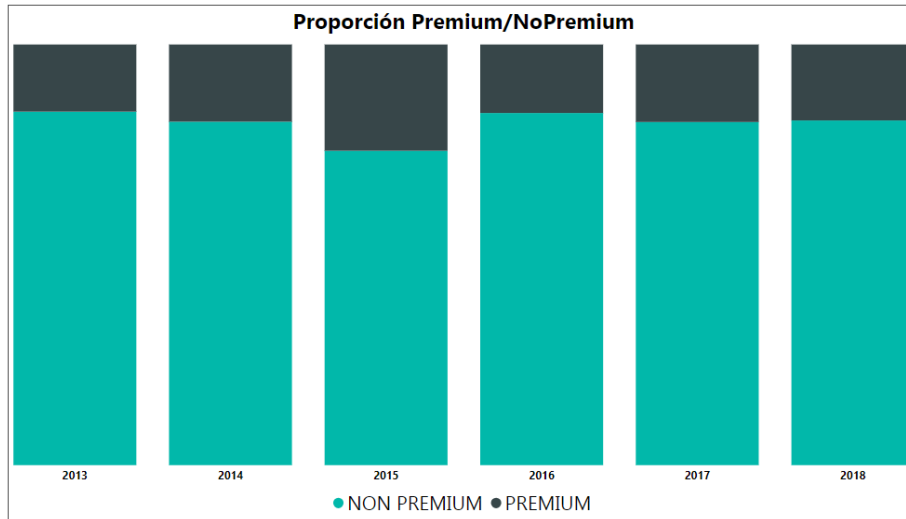


Figure 13: Evolución ventas y proporción de material Premium vs No-Premium

Es interesante ver como el mix entre la proporción de tubería Premium y No-Premium vendida a lo largo de estos años prácticamente se mantiene.

Más adelante se estudiará si es posible sacar alguna relación entre los valores históricos de la industria y la distribución de las ventas entre estos dos grupos para poder predecir también que tipo de planta, materias primas y capacitación de personal se necesitará en un análisis de actividad futuro.

III. Line Pipe

La unidad de negocio que se encarga de la venta de materiales que permite la distribución de los hidrocarburos extraídos se conoce como Line Pipe.

Estos tubos, fabricados también de acero, pueden ir enterrados o no para el transporte de gas o petróleo desde el yacimiento hacia líneas más grandes de distribución o directamente a locaciones de almacenamiento, refinerías, plantas petroquímicas, etc. Para obtener una magnitud del tamaño del negocio que significa, actualmente en Estados Unidos se encuentran instalados 3.8 millones de kilómetros de tubería para el transporte de hidrocarburos.



Figure 14: Instalación de tubería Line Pipe para distribución de hidrocarburos

Es importante destacar tres aspectos respecto a este negocio, por un lado es de suponer que un incremento de la actividad petrolera es condición necesaria para que las ventas de este negocio aumenten (ya que cuando crecen los niveles de extracción es necesario contar con la distribución de lo extraído) pero también, la capacidad instalada actualmente en mercados desarrollados como el de Estados Unidos ya es de tal magnitud que no necesariamente se requieren grandes inversiones en líneas de distribución por parte de las petroleras que se traduzcan en compras hacia Tenaris. Caso contrario sucede con yacimientos en primeras etapas de explotación, por ejemplo Vaca Muerta, en los cuales es necesario fabricar e instalar miles de kilómetros que permitan el correcto desarrollo de la operación y la posterior distribución necesaria para insertarla en las redes domésticas, industriales y que permitan llegar hasta plantas de licuefacción permitiendo así la exportación a países compradores.

El último aspecto de este negocio a considerar es que las plantas necesarias para fabricar estos tubos no son las mismas ya que las características de los tubos de Line Pipe tales como grado de acero, peso por metro, conexión, diámetros, entre otras, no son las mismas que para la tubería OCTG. A su vez, estas tuberías llevan un revestimiento tanto interno como externo que hacen que esencialmente el proceso de fabricación sea otro.

Para el posterior análisis de estimación del capítulo 5, se tratará a esta unidad de negocios como un único negocio global, independientemente de donde se utilice/fabrique ya que los datos que se cuentan no están discriminados por región geográfica.

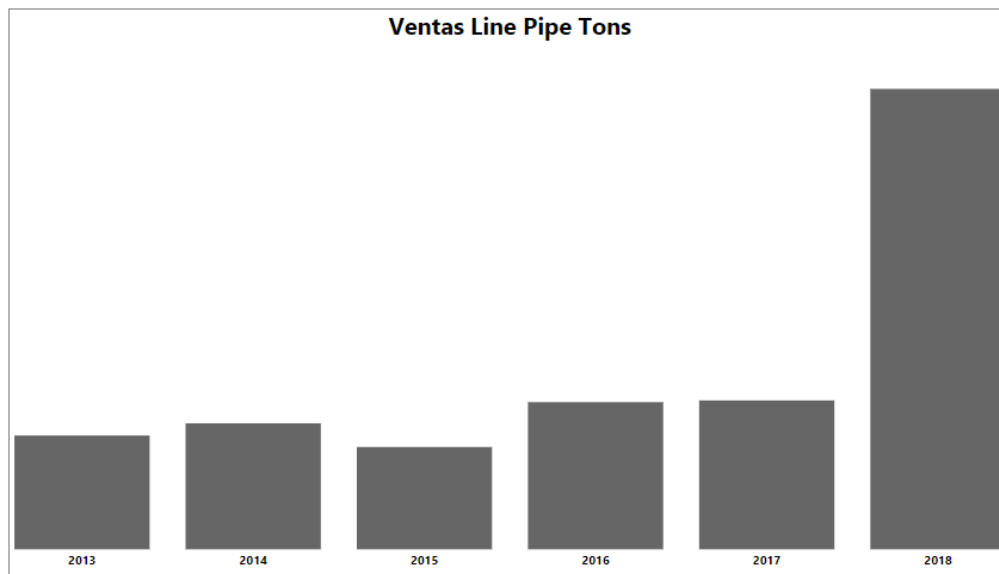


Figure 15: Ventas anuales Line Pipe

El pico de ventas del 2018 se debe a numerosos proyectos puntuales que surgieron durante ese año que incrementaron exponencialmente las ventas en esta unidad. Luego será cuestión de definir si con la información disponible corresponde o no adecuar un modelo que intente estimar las ventas en esta unidad de negocios.

IV. Días de servicio de entubación y asistencia técnica on-site

Sumado a la venta de tuberías, OCTG o Line Pipe, Tenaris ofrece una amplia gama de servicios on-site con personal capacitado que brinda soporte técnico y asistencia en la inspección, puesta a punto y bajada de la tubería comprada por el cliente. A diferencia de la unidad de negocio de Line Pipe, estos servicios van muy de la mano con los resultados de las ventas de tubería. Esto no significa una relación directa ya que no necesariamente todo cliente que compra tubería a la vez contrata el servicio, pero si ante un mayor volumen de ventas se espera que la cantidad de servicios de campo contratados aumente.

Este servicio se mide en días, generalmente son turnos de 12 horas por persona que significan un día, y lo brinda personal que se dedica específicamente a visitar la locación del cliente ya sea una base, un patio de tubería o un rig en operación. El personal cuenta con perfil técnico que conlleva un considerable programa de capacitación y experiencia necesaria por ende una significativa inversión de tiempo y recursos para lograr independencia, y excelencia, en el campo.

A continuación se muestra la evolución de la actividad para esta unidad de negocio en Estados Unidos:

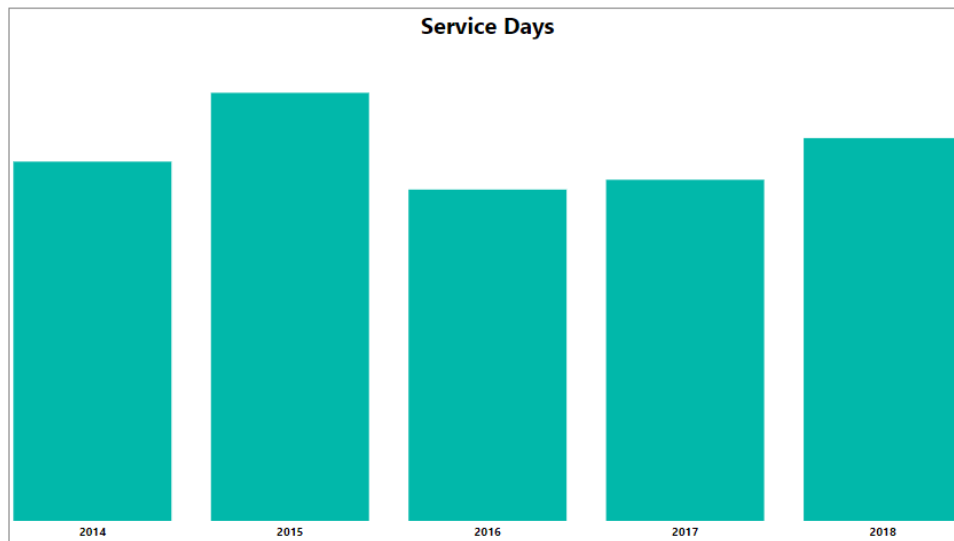


Figure 16: Días de Servicios de campo brindados

5. Modelado para estimación de ventas

En los capítulos anteriores se ha estudiado (i) la situación geopolítica y la injerencia en la industria; (ii) las principales variables endógenas que con mayor o menor proporción dictan como se mueve el mercado y por último (iii) la actividad histórica de Tenaris desde el punto de vista de los principales resultados de negocio.

Para el desarrollo de este capítulo, según la literatura existente, consultas con expertos en el área, lo expuesto en capítulos anteriores pero sobre todo con la información disponible recabada oportunamente, se tendrá en cuenta las siguientes cuatro variables para la definición del modelo que estime las ventas futuras, llamadas de ahora en más variables explicativas:

- Rig Count
- Gas Price
- WTI Price
- Tasa Libor 60D

En definitiva, se incluyen 3 variables de la industria y una variable macroeconómica para aportar al modelo una sensibilidad de la situación económica puntual de ese momento. En el caso de los modelos de Line Pipe y días de servicio, se utilizarán a su vez como variables explicativas las ventas totales ya que en base a esos resultados, de alguna manera, se desprenden estos otros dos negocios.

Para todas estas variables explicativas, se definirán los coeficientes que mejor ajustarán un modelo que permita con valores futuros aún no conocidos tener una estimación de la actividad en el mercado norteamericano. Esta estimación, o modelo prescriptivo, servirá no solo para predecir potenciales ganancias o el nivel de actividad de la Compañía sino también para contar con cierta precisión del “horizonte” (fundado con argumentos estadísticos basados en la historia) que servirá para poder tomar, con tiempo suficiente, decisiones de negocio que permitan cumplir con dicha demanda estimada.

I. Preparación de las variables seleccionadas

Es imperioso entender la forma y las características en la cual se mueven las variables que serán insertadas al modelo para poder remover aquellos efectos que puedan perjudicar la performance

del mismo. Esto significa que al estudiar variables estocásticas (variables que evolucionan en el tiempo) deben tenerse ciertas precauciones y consideraciones ya que de lo contrario puede incurrirse en lo que econometría se conoce como relaciones espurias. En otras palabras, la relación estadística encontrada no sirve para aprender ni para describir una relación y se estaría cometiendo un error en caso de incluirlas luego en un modelo porque las series no guardan relación verdadera, o de negocio, entre sí.

La metodología para estudiar y prevenir que este efecto llegue al modelo será por medio del estudio de la tendencia de las variables que a priori son relevantes para describir dicho modelo. En concreto, se aplica el test de prueba estacionaria o ADF (Augmented Dickey-Fuller) siendo este el más tradicional para el estudio de raíces unitarias. En función de los resultados del test, se definirá a que variables se les removerá la tendencia y a cuales no, para poder armar el modelo. A su vez, los resultados de negocio presentados en el Capítulo 4 también se someterán al análisis para poder ajustar un modelo con datos que se encuentren libres de tendencia.

II. Fundamento teórico del test Augmented Dickey-Fuller

El test se basa en rechazar o no la hipótesis nula de que para una serie de tiempo existe una raíz unitaria para cierto nivel de confianza. La existencia de una raíz unitaria se asocia al hecho de que dicha variable posee una tendencia estacionaria que puede derivar en una degradación del modelo. En términos estadísticos, la existencia de esta raíz unitaria significa que el anterior nivel de la serie, llámese Y_{t-1} , no proveerá información relevante para predecir los cambios en Y_t aparte de los que ya han sido explicados en ΔY_{t-k} siendo k , algún momento en el tiempo anterior. Por otro lado, cuando la serie no tiene una raíz unitaria, se define como estacionario y por lo tanto muestra una reversión a la media, por lo que niveles anteriores Y_{t-k} proporcionarán información relevante para predecir el cambio de la serie y por ende se rechaza la hipótesis nula.

III. Resultados del Test e interpretación

Para realizar este test, es necesario emplear un paquete en R llamado “urca” que permite utilizar librerías con manipulación de series de tiempo. En concreto la función que calcula el test ADF se llama `ur.df` en alusión a Dickey-Fuller. En el anexo 2 y 3 se encuentran los running code utilizados. Como resultado del test, a continuación se muestra un output a modo de ejemplo para la variable **Rig Count** para explicar lo obtenido y de esta manera se hace extensivo al resto de las variables analizadas:

```
[*] "Rig Count"

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

(1) Residuals:
    Min       1Q   Median       3Q      Max
-251.495  -37.333    9.131   48.494  168.262

(2) Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  119.06019   73.50560    1.620   0.1098
z.lag.1      -0.04157    0.02012   -2.066   0.0426 *
tt           -0.62824    0.71066   -0.884   0.3798
z.diff.lag    0.65716    0.09046    7.265  4.4e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(3) Residual standard error: 86.63 on 69 degrees of freedom
Multiple R-squared:  0.4563,    Adjusted R-squared:  0.4327
F-statistic:  19.3 on 3 and 69 DF,  p-value: 3.408e-09

Value of test-statistic is: -2.0661 1.9682 2.5389

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.04 -3.45 -3.15
phi2  6.50  4.88  4.16
phi3  8.73  6.49  5.47
```

Tomando como ejemplo la serie de Rig Count, los resultados del test pueden dividirse en cuatro grupos de información:

- (1) La fórmula bajo la cual se corre el modelo. Los valores “lag” indican que la modelización de tendencia toma valores de meses anteriores para estudiar el impacto de la historia en meses subsiguientes.
- (2) Valores residuales del test, no son utilizados para este análisis
- (3) Los coeficientes contienen la información de la recta obtenida. Básicamente el valor de la ordenada al origen y la pendiente de la variable a estudiar, sumado al error standard y su p-value asociado, que indica cuan significativo es dicho coeficiente. En este momento como no se está mirando la recta resultante en si sino la existencia o no de tendencia, no son valores que se toman en cuenta para este momento del análisis
- (4) Por último, y el que en definitiva interesa para este test, se encuentra la información del resultado del test estadístico. En este ejemplo el valor dio -2.0661 y como este es mayor a los valores críticos para 1%, 5% y 10% de significancia entonces no es posible rechazar la hipótesis nula y por ende se asume que existe una raíz unitaria y por lo tanto se concluye que la serie tiene tendencia.

Para este caso en particular, será necesario eliminar dicha tendencia por medio del cálculo de las tasas de cambio entre los valores mensuales y volver a correr el test para esta serie ahora referenciadas y así analizar la persistencia o no de tendencia. Será necesario realizar este cálculo de tasas tantas veces como sea necesario hasta obtener un valor estadístico más pequeño que los valores críticos que oscile entre el 90% y 99% de significancia.

A continuación se muestran todos los resultados en formato de tabla para facilidad de interpretación.

Resultados del test realizado a las variables explicativas en el primer grupo y dependientes en el segundo:

	Estimador	1pct	5pct	10pct	Conclusión
Libor 60D	-1.4181	-4.04	-3.45	-3.15	No es posible rechazar H_0
WTI Price	-1.8992	-4.04	-3.45	-3.15	No es posible rechazar H_0
Gas Price	-2.4046	-4.04	-3.45	-3.15	No es posible rechazar H_0
Rig Count	-2.2677	-4.04	-3.45	-3.15	No es posible rechazar H_0
Non Premium Tons	-2.1218	-4.04	-3.45	-3.15	No es posible rechazar H_0
Premium Tons	-2.0413	-4.04	-3.45	-3.15	No es posible rechazar H_0
Total Tons	-2.0238	-4.04	-3.45	-3.15	No es posible rechazar H_0
Line Pipe	-3.347	-4.04	-3.45	-3.15	No es posible rechazar H_0
Días Servicio	-3.5556	-4.04	-3.45	-3.15	No es posible rechazar H_0

Figure 17: Resultados test ADF a variables explicativas y dependientes con valores reales

Como puede verse, para ninguna variable es posible rechazar la hipótesis nula y por lo tanto deberá eliminarse la tendencia por medio del cálculo de tasas. Habiendo hecho esto, se vuelve a correr el test dando los siguientes resultados:

	Estimador	1pct	5pct	10pct	Conclusión
Libor 60D	-4.4731	-4.04	-3.45	-3.15	Rechazo H_0
WTI Price	-5.149	-4.04	-3.45	-3.15	Rechazo H_0
Gas Price	-5.8014	-4.04	-3.45	-3.15	Rechazo H_0
Rig Count	-6.6762	-4.04	-3.45	-3.15	Rechazo H_0
Non Premium Tons	-9.306	-4.04	-3.45	-3.15	Rechazo H_0
Premium Tons	-10.4639	-4.04	-3.45	-3.15	Rechazo H_0
Total Tons	-9.0161	-4.04	-3.45	-3.15	Rechazo H_0
Line Pipe	-6.8342	-4.04	-3.45	-3.15	Rechazo H_0
Días Servicio	-5.5731	-4.04	-3.45	-3.15	Rechazo H_0

Figure 18: Resultados test ADF con variables referenciadas

Ahora si es posible rechazar la hipótesis de existencia de raíz unitaria ya que para todas las variables el estimador ahora es más pequeño que el valor a comprar, por lo tanto es posible concluir que las series están libres de tendencia.

IV. Modelización

Teniendo las variables libres de tendencia que se incluirán al modelo y las series que lo componen, se procede a calcular las regresiones. Para alcanzar el mejor modelo posible para cada estimación, se irá incrementando la complejidad del algoritmo a utilizar para analizar cuál es el mejor método. En principio, se comenzará con el modelo lineal por mínimos cuadrados ordinarios (MCO), luego con un modelo dinámico que permite tener en cuenta los rezagos de las variables y por último se utilizará un algoritmo conocido como Autometrics que busca el mejor modelo en términos de significatividad probando distintas combinaciones de las variables incluidas. Para la estimación de los modelos se utilizarán librerías de R que permiten realizar este tipo de regresiones, en los anexos 4, 5 y 6 se encuentran los running code utilizados.

Es necesario aclarar, que para los primeros dos algoritmos, la inclusión o no de las variables fue completamente manual en función a lo que podría estimarse funcione mejor para el modelo.

(i) Regresión lineal (MCO)

Este primer método, consiste en minimizar el cuadrado de la distancia entre las series de tiempo para así encontrar la ecuación que mejor ajuste la variable dependiente. En detalle, se trata de encontrar una línea media que sintetice la dependencia entre la variable dependiente y las explicativas. Esta línea o ecuación resultante debe acompañarse siempre de alguna medida que demuestre el grado en el que el promedio puede sustituir a las observaciones individuales, que es lo que se conoce como dispersión. El análisis de esta dispersión, permite interpretar el coeficiente como la proporción o porcentaje de variación total en Y respecto a su media, que es explicada por el modelo de regresión.

Cuando esta proporción, conocida como R^2 , es igual a 1 significa un ajuste perfecto por el cual la variación total de la variable dependiente es explicada por el modelo de regresión. El R^2 igual a cero indica la no representatividad del modelo lineal, lo que concluye que el modelo no explica absolutamente nada de la variación total de dicha variable dependiente. Sin embargo, la literatura afirma que mirar solamente el R^2 puede ser un error ya que se pueden estar tomando conclusiones desacertadas. Esta medida de dispersión debe ser acompañada, en todos los casos,

de un análisis cualitativo respecto a lo que hay que tiene detrás el resultado con conocimiento y perspectiva del negocio.

Este modelo, al igual que el resto de las regresiones de este trabajo, se presenta matemáticamente de la siguiente forma:

$$Y = a + b.X_1 + c.X_2 + d.X_3 \quad (1)$$

Siendo:

- Y la variable dependiente o a estimar (Ventas, Días de Servicio, etc.)
- a ordenada al origen
- b, c, d, etc. coeficientes que se obtienen mediante los resultados de la regresión
- X_1, X_2, X_3 , etc. variables explicativas futuras

A continuación se muestran los resultados de esta primera regresión para cada combinación que se ha hecho y para cada variable dependiente a analizar y luego debajo algunos ejemplos del output del modelo empleado en R con el algoritmo "lm".

Modelo #	Variable Dependiente	Variables Explicativas					R ²
		Rig Count	Oil Price	Gas Price	Libor60D	Ventas	
1	Ventas Totales	x	x	x	x		-2.2%
2	Ventas Totales	x	x				-2.8%
3	Ventas Totales	x					1.4%
4	Ventas Totales		x				1.3%
5	Ventas Premium	x	x			x	41.1%
6	Ventas Premium					x	42.9%
7	Ventas Premium	x					-0.9%
8	Ventas Premium		x				-1.0%
9	Ventas No-Premium	x	x			x	95.3%
10	Ventas No-Premium					x	95.4%
11	Ventas No-Premium			x		x	95.3%
12	Ventas No-Premium			x			-1.3%
13	Ventas LinePipe	x	x	x	x	x	-2.7%
14	Ventas LinePipe	x				x	-0.8%
15	Ventas LinePipe					x	-1.2%
16	Ventas LinePipe		x				-0.8%
17	Días de Servicio	x	x	x	x	x	11.8%
18	Días de Servicio	x	x				7.3%
19	Días de Servicio					x	-1.1%
20	Días de Servicio	x					8.1%

Figure 19: Resultados modelado por MCO

Los modelos más representativos según el R^2 se encontraron para la estimación de ventas de tubería No-Premium. Guíandose únicamente por el R^2 , las variables seleccionadas en esos 3 modelos, explican aproximadamente el 95% de la variabilidad de las ventas No-Premium. A medida que se avance en la complejidad de los algoritmos que realizan la estimación se volverá sobre estos resultados para entender si es correcto este primer resultado. Como primera impresión, pareciera ser que los modelo 9, 10 y 11 alcanzan un R^2 mayor por el hecho de incluir la variable ventas. En cambio, para el modelo 12, donde ventas no se incluye, el resultado es claramente distinto.

A modo de ejemplo, el output de uno de los modelos obtenidos para la estimación de ventas No-Premium es el siguiente:

```
> summary(modeloMCO10)

Call:
lm(formula = nonpremium ~ totaltons)

Residuals:
    Min       1Q   Median       3Q      Max
-12.1334  -2.7772  -0.2964   3.2196  12.7068

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.17496    0.66744   0.262   0.794
totaltons    1.04153    0.02916  35.715 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.262 on 61 degrees of freedom
Multiple R-squared:  0.9544,    Adjusted R-squared:  0.9536
F-statistic: 1276 on 1 and 61 DF,  p-value: < 2.2e-16
```

Figure 20: Output Modelo 10 método MCO

Al analizar el resultado el resultado en detalle de la figura anterior, se evidencia un modelo que no es correcto para estimar ventas de material No-Premium ya que completando lo obtenido en la ecuación (1) se tendría:

$$\text{Ventas No Premium} = 0.175 + 1.042 * \text{Ventas Totales}$$

Lo cual es completamente ilógico desde una perspectiva de negocio ya que no es posible en la realidad que el valor de Ventas No Premium sea siempre mayor a las Ventas Totales siendo que las Ventas No Premium es un sub-conjunto de las totales.

En segundo lugar con respecto al ajuste del R^2 se encuentran los modelos 5 y 6, que estiman las ventas de material Premium. La ecuación resultante del modelo 6, con base en la ecuación (1) es:

$$\text{Ventas Premium} = 2.1327 + 0.8842 * \text{Ventas Totales}$$

```
> modeloMCO6

Call:
lm(formula = premium ~ totaltons)

Coefficients:
(Intercept)    totaltons
    2.1327         0.8848
```

Figure 21: Resultado modelo 6 MCO

Este resultado, una vez más, no tiene sentido ya que se contradice con los valores históricos de Tenaris. El modelo dice que básicamente las ventas de material Premium son cercanos al 90% de las ventas totales lo cual no es así según lo descrito en el capítulo anterior.

(ii) Modelo Dinámico

El principal supuesto del modelo anterior, es que la variable dependiente es impactada por las variables explicativas en el mismo mes, lo cual no es necesariamente cierto para aplicaciones de negocio. Puede haber sido el caso que, por ejemplo, una baja en la tasa Libor del mes anterior tenga impacto en la economía real en el mes siguiente, o que el impacto de la baja del precio del barril de petróleo en un mes se evidencie recién en meses subsiguientes (siempre y cuando se mantenga la tendencia a la baja). Por lo tanto, en este contexto, una variable dependiente podría estar impactada por valores rezagados de las variables explicativas. Estos modelos que consideran los rezagos de las variables son conocidos como Modelos Dinámicos.

Los modelos fueron planteados con la misma combinación de variables que en el caso de MCO pero con la diferencia de que siempre se incluyeron los rezagos de las variables que son sometidas a cada modelo para poder entender que efecto tienen en la performance y definición del mismo. También se incluyen los rezagos de las variables dependientes a estimar.

En R, los modelos se codifican gracias a la librería Dyn que contiene el algoritmo que permite incluir los rezagos de las variables.

A continuación los resultados:

Modelo #	Variable Dependiente	Variables Explicativas					R ²
		Rig Count	Oil Price	Gas Price	Libor60D	Ventas	
1	Ventas Totales	x	x	x	x		63.9%
2	Ventas Totales	x	x				52.2%
3	Ventas Totales	x					49.1%
4	Ventas Totales		x				52.7%
5	Ventas Premium	x	x			x	59.8%
6	Ventas Premium					x	51.9%
7	Ventas Premium	x					39.4%
8	Ventas Premium		x				42.7%
9	Ventas No-Premium	x	x			x	96.7%
10	Ventas No-Premium					x	95.9%
11	Ventas No-Premium			x		x	96.3%
12	Ventas No-Premium			x			43.8%
13	Ventas LinePipe	x	x	x	x	x	14.5%
14	Ventas LinePipe	x				x	5.8%
15	Ventas LinePipe					x	2.3%
16	Ventas LinePipe		x				13.2%
17	Dias de Servicio	x	x	x	x	x	14.0%
18	Dias de Servicio	x	x				12.4%
19	Dias de Servicio					x	1.6%
20	Dias de Servicio	x					10.5%

Figure 22: Resultados modelo Dinámico

Como primera impresión, se puede afirmar que este método mejora el ajuste que realiza el R² en comparación a los modelos obtenidos con el método anterior. Como lo indica la literatura, debe ahondarse en el análisis ya que el R² no es suficiente para poder dar un veredicto confiable acerca de la performance de un modelo*.

A continuación, se irán analizando los resultados de a grupos de modelos según las variables dependiente a estimar.

* Errores frecuentes en la interpretación del coeficiente de determinación lineal, E. Rodriguez

En primer lugar, debajo se encuentran los resultados del modelo dinámico 1 que estima las ventas totales:

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.5236     2.6260   2.484 0.01669 *
rigcount          0.6925     0.9551   0.725 0.47206
lag(rigcount, -1) -1.7383     0.9880  -1.759 0.08517 .
lag(rigcount, -2) -0.8040     1.0199  -0.788 0.43453
oil              -0.5643     1.8544  -0.304 0.76228
lag(oil, -1)      3.7815     1.7166   2.203 0.03265 *
lag(oil, -2)      2.9901     1.7568   1.702 0.09551 .
gas              -0.2150     0.2481  -0.867 0.39063
lag(gas, -1)     -0.2142     0.2443  -0.877 0.38521
lag(gas, -2)      0.8039     0.2600   3.092 0.00337 **
libor            0.7437     0.3496   2.127 0.03881 *
lag(libor, -1)   -0.8989     0.4019  -2.237 0.03018 *
lag(libor, -2)    0.7321     0.3507   2.088 0.04240 *
lag(totaltons, -1) -0.6713     0.1019  -6.589 3.77e-08 ***
lag(totaltons, -2) -0.6025     0.1038  -5.806 5.66e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.94 on 46 degrees of freedom
(4 observations deleted due to missingness)
Multiple R-squared:  0.7235,    Adjusted R-squared:  0.6394
F-statistic: 8.599 on 14 and 46 DF,  p-value: 1.081e-08
    
```

Figure 23: Resultados modelo dinámico 1

A partir de los resultados obtenidos, como hay muchas variables que no tienen injerencia estadística en la estimación se procede a remover las variables con un p-value (Columna Pr(>|t|)) no significativo* y se vuelve a correr el modelo. Esta iteración, se hará tantas veces como sea necesaria para obtener así un modelo que emplee solamente las variables que tienen un impacto significativo en el indicador que se quiere estimar.

* Para definir niveles de significativad se toman las variables con p-value menor a 10%.

Para la primer iteración se remueven las variables rigcount, lag (rigcount,-2), oil, gas, lag (gas,-1) y puede verse que el R² mejora en muy baja proporción pero si se evidencia que aún hay variables que deben removerse. De esta manera, en la siguiente iteración se remueven lag (rigcount, -1) y lag (libor, -2) y en la última iteración se remueve lag (libor, -1), obteniéndose así los siguientes resultados:

```
Call:
lm(formula = dyn(totaltons ~ lag(oil, -1) + lag(oil, -2) + lag(gas,
-2) + lag(totaltons, -1) + lag(totaltons, -2)))

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.28285    1.92027   4.313 6.73e-05 ***
lag(oil, -1)     1.37834    0.59849   2.303 0.025084 *
lag(oil, -2)     1.38138    0.63610   2.172 0.034212 *
lag(gas, -2)     0.87698    0.21762   4.030 0.000173 ***
lag(totaltons, -1) -0.71534    0.09408  -7.603 3.88e-10 ***
lag(totaltons, -2) -0.59062    0.09024  -6.545 2.09e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.31 on 55 degrees of freedom
(4 observations deleted due to missingness)
Multiple R-squared:  0.6518,    Adjusted R-squared:  0.6201
F-statistic: 20.59 on 5 and 55 DF,  p-value: 1.583e-11
```

Figure 24: Última iteración del modelo dinámico 1

Llegando de esta manera a un modelo que tiene 5 variables significativas con un intervalo de confianza del 90% que explican un 62% de la variabilidad de las Ventas Totales.

Modelo de estimación Ventas Totales:

$$V. Tot.t = 8.23 + 1.38 * Oil Price_{t-1} + 1.38 * Oil Price_{t-2} + 0.88 * Gas Price_{t-2} - 0.72 * V. Tot.t-1 - 0.59 * V. Tot.t-2$$

Siguiendo con el segundo grupo, para la estimación de las ventas Premium se toman los resultados del modelo 5, de más alto R^2 , y se procede a remover las variables no significativas de la misma forma que con el grupo anterior. Estas iteraciones resultan en un modelo en el cual se comenzó planteando 11 variables de las cuales solamente resultaron 2 significativas:

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.5029     2.8824   1.215  0.22911
totaltons      0.7974     0.1274   6.259 4.79e-08 ***
lag(premium, -1) -0.2597    0.0955  -2.720  0.00857 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.2 on 59 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared:  0.4982,    Adjusted R-squared:  0.4812
F-statistic: 29.29 on 2 and 59 DF,  p-value: 1.465e-09

```

Figure 25: Última iteración del modelo dinámico 5

Llegando de esta manera a un modelo que tiene 2 variables significativas con un intervalo de confianza del 90% que explican un 48% de la variabilidad de las Ventas Premium.

Modelo de estimación Ventas Premium:

$$\text{Ventas Premium}_t = 3.5 + 0.80 * \text{Ventas Totales} - 0.26 * \text{Ventas Premium}_{t-1}$$

Para la estimación de ventas No-Premium se siguen obteniendo valores muy altos cuando se incluye la variable “Ventas Totales” en el modelo, lo cual da indicio que puede estar ajustando por el hecho de que los valores reales de ambas variables se mueven prácticamente juntos.

Tomando en consideración el modelo 9 para estimar Ventas No-Premium y haciendo las iteraciones correspondientes se llega a un resultado que depende de las ventas totales del mes, las ventas totales del mes anterior y las ventas No-Premium del mes anterior.

Modelo de estimación Ventas No-Premium:

$$\text{Ventas No-Premium}_t = 0.24 + 1.05 * V. Tot. + 0.44 * V. Tot._{t-1} - 0.42 * \text{Ventas No-Premium}_{t-1}$$


```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.2435     0.6335   0.384 0.702055
totaltons      1.0467     0.0290  36.089 < 2e-16 ***
lag(totaltons, -1) 0.4421     0.1284   3.443 0.001076 **
lag(nonpremium, -1) -0.4241     0.1194  -3.553 0.000765 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.89 on 58 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared:  0.9623,    Adjusted R-squared:  0.9603
F-statistic: 493.4 on 3 and 58 DF,  p-value: < 2.2e-16
    
```

Figure 26: Última iteración del modelo dinámico 9

En el resultado anterior se evidencia que las 3 variables resultantes explican el 96% de la variabilidad de las ventas No-Premium siendo la variable Ventas Totales la que impacta con un nivel de significatividad mayor al 99% lo cual confirma lo planteado al comienzo de este análisis. La variable Ventas Totales incide directamente en el resultado de las Ventas No-Premium.

Por último, los modelos de Line Pipe y Días de Servicio siguen teniendo bajo R^2 , menor al 15% en los 8 modelos. Para el caso de Line Pipe, tal como se había dicho en el capítulo anterior, es un mercado complejo de predecir por depender de proyectos y licitaciones puntuales. En cambio, por el lado de los modelos de Días de Servicio, era de esperarse mejores resultados por la incidencia prácticamente directa que se tiene con las ventas en la realidad. En el próximo apartado, con el modelado más complejo, se intentará de obtener resultados de mayor impacto para estos dos indicadores de negocio.

Como conclusión de estos métodos dinámicos, es necesario aclarar que los tres modelos resultantes cuentan con valores de rezago de la misma variable que se está queriendo predecir (representadas en los modelos por $\text{lag}(\text{XXX}, -1)$ y $\text{lag}(\text{XXX}, -2)$):

$$V. Tot. t = 8.23 + 1.38 * Oil Price_{t-1} + 1.38 * Oil Price_{t-2} + 0.88 * Gas Price_{t-2} - 0.72 * V. Tot. t-1 - 0.59 * V. Tot. t-2$$

$$Ventas Premium_t = 3.5 + 0.80 * Ventas Totales - 0.26 * Ventas Premium_{t-1}$$

$$Ventas No-Premium_t = 0.24 + 1.05 * V. Tot. + 0.44 * V. Tot. t-1 - 0.42 * Ventas No-Premium_{t-1}$$

Motivo por el cual, es condición obligatoria tener dichos valores de meses anteriores para emplear un modelo sólido. En otras palabras, significa que si se está estimando las ventas de aquí a 6 meses entonces también debo contar con la estimación de lo que sucederá en el mes 4 y en el mes 5 porque dichos valores deberán necesariamente ser incluidos en el modelo.

Por otro lado, es interesante ver que cuando las variables de ventas no son incluidas y se vuelve a calcular la performance resultante, el R^2 de los modelos dinámicos cae considerablemente. En concreto, el modelo 1 cae a 26.2%, el modelo 5 cae a 15% y el modelo 9 a 9.1%. Estos resultados permiten afirmar que gran parte de la variabilidad de las ventas es explicada por lo que sucede en meses anteriores de las mismas ventas. Este efecto también puede evidenciarse al mirar los p-values de las variables significativas finales, donde en todos los modelos son siempre las de mayor significatividad en comparación al resto de las variables explicativas incluidas.

Por último, para los modelos 3 y 4, donde se estudia el impacto por separado de las variables Rig Count y Oil Price se evidencia que ambas no son significativas por si solas ya que los R^2 obtenidos se alcanzan por el rezago de la misma variable ventas. Aún más, ni la variable Rig Count ni tampoco la Tasa Libor terminan formando parte de alguno de los modelos finales obtenidos con niveles de significatividad mayores al 90%.

(iii) Modelo Autometrics

El algoritmo Autometrics trabaja de la metodología general a particular, lo que básicamente significa que prueba cuales son las mejores combinaciones (mediante un árbol de decisión) de las variables a incluir en el modelo y de esa manera busca el que mejor ajusta en términos de significatividad y que también cumpla con los supuestos subyacentes a este método de estimación. Luego de probar por una de estas combinaciones, almacena los resultados y comienza de nuevo a probar con otra combinación para así comparar resultados. El modelo parte de un *General Unrestricted Model* (GUM) que viene a ser el modelo más grande posible, en el cual se incluye la variable dependiente a estimar junto a todas las explicativas. Sobre este GUM se corre una función para que converja en un modelo que contenga solamente las variables explicativas significativas y que cumpla con las pruebas de diagnóstico tradicionales.

Partiendo del modelo 1, de Ventas Totales, hasta el de Días de Servicio que es el 5 se puede afirmar que debido a la multiplicidad de pruebas que realiza Autometrics, a diferencia de los métodos anteriores, se evita el hecho de realizar iteraciones que antes eran necesarias para remover las variables no significativas. Una vez obtenidos los resultados, será interesante ver qué relación se encuentra con los modelos obtenidos mediante las metodologías anteriores.

En R, primero es necesario que las variables tengan formato de series temporales utilizando la función que está en la librería “Zoo” y para el cálculo del modelo sobre el GUM se utiliza la función Getsm de la librería “Gets”.

Sin ahondar en los detalles estadísticos ni en particularidades que hacen al método en sí, es destacable la cantidad de iteraciones distintas que hace por variable a estimar. En el modelo que menos caminos emplea es en el 1 que utiliza 8 combinaciones distintas y para el que más utiliza es en el 5 que prueba con 17 combinaciones distintas de variables.

A fin de cuentas es un modelo dinámico, ya que se incluyen las variables rezago de cada una de las dependientes y explicativas, que se encarga de permutar las variables buscando la mejor combinación.

A continuación se muestran los coeficientes obtenidos y para cada variable en cada uno de los modelos a estimar:

Variable Dependiente	Variables Explicativas													Resultados	
	Rig Count	Rig Count t_{-2}	Oil Price	Oil Price t_{-1}	Oil Price t_{-2}	Gas Price t_{-2}	Libor 60D	Ventas	Ventas t_{-1}	Ventas t_{-2}	Prem. t_{-1}	Non Prem. t_{-1}	Días de Serv t_{-1}	R ²	Int. de Confianza
Ventas Totales	-	-	-	-	2.23	0.88	0.81	-	-0.71	-0.57	-	-	-	58%	99%
Ventas Premium	-	0.77	-	-	-	-0.72	-	0.97	0.43	-	-0.48	-	-	65%	95%
Ventas No-Premium	-	-	0.3	-	-0.57	-	-	1.06	0.59	-	-	-0.54	-	97%	95%
Ventas LinePipe	-11.22	-	-	29.03	-13.46	-	-	-	-	-	-	-	-	14%	95%
Días de Servicio	-	-	-	1.9	-	-	-	-	-	-	-	-	-0.28	12%	95%

Figure 27: Resultados modelo Autometrics

En el cuadro anterior con un color más oscuro se detallan las variables explicativas “puras” y con uno más claro aquellas que son de rezago. Vale destacar dos aspectos, por un lado la variable Gas Price del mes de estudio no termina figurando en ningún modelo pero si la de rezago de dos

meses anteriores, y por otro lado que en el único modelo final en el cual no toma variables de rezago propias de la variable a estimar es el de Line Pipe.

Con respecto al ajuste del R^2 y la descripción de la variabilidad que se obtiene para cada modelo, los resultados son prácticamente iguales a los obtenidos con el método dinámico anterior. Lo que si puede tomarse como certero es que este modelo no tiene el sesgo de la inclusión o no inclusión de las variables. Al incluir todas desde el comienzo es posible afirmar que estos resultados son más confiables que los obtenidos anteriormente ya que es el algoritmo quién decide cual variable y en qué momento incluirla, para ir comparando las distintas combinaciones, en pos de encontrar la mejor estimación en términos estadísticos.

V. Conclusión Capítulo

Al haber incrementado la complejidad del algoritmo a medida que se avanzaba en el análisis, fue posible entender cuáles son las desventajas o falencias que tiene cada uno de los modelos. Definitivamente, el modelo MCO alcanza bajos resultados por el hecho de no incluir variables de rezago mientras que el dinámico permite incluir un espectro más amplio por medio de los rezagos que le brinda sensatez al análisis desde el lado de negocio. En mayor o menor medida, para estos dos métodos, los resultados se ven afectados por la elección inicial de las variables a utilizar. En el caso de contar con más variables a analizar, hubiera sido más engorroso codificar los distintos modelos que se fueron implementando ya que para cada método se hicieron 20 combinaciones distintas, sin contar las iteraciones realizadas para el modelo Dinámico para eliminar variables no significativas.

El algoritmo de Autometrics, permite concluir que es el método más eficiente y sólido para realizar estimaciones en períodos cortos, donde se cuente con la mayor cantidad posible de información de las variables rezago. De todas maneras, es necesario aclarar que si se tienen en cuenta variables de rezago que pueden ser a su vez estimaciones, el riesgo de utilizar el modelo claramente crece.

Con respecto a la interpretación de los resultados obtenidos, se puede concluir que, tal como dice la literatura, es peligroso recaer solamente en el R^2 porque puede arribarse a resultados que son imposibles que se den en la realidad del negocio o estadísticamente hablando, tomar

coeficientes que no acompañan a variables significativas para el modelo lo que sería claramente perjudicial para el mismo. Respecto a los resultados en sí, se evidencia que los modelos más sólidos se obtienen para las Ventas Totales, Ventas Premium y Ventas No-Premium. De todas formas, los valores de Ventas Premium y No-Premium están muy condicionados por el de Ventas Totales, de esta forma, el más interesante para ahondar a detalle es el de Ventas Totales, que no solo cuenta con variables de la industria como el precio del Gas sino que también es el único que utiliza la variable económica de la Tasa Libor. Como fue mencionado durante el desarrollo de los modelos, era de esperarse un modelo más representativo respecto a los Días de Servicio el cual a priori, por experiencia de negocio, es el que más afectado se ve por las fluctuaciones de las ventas de Tenaris.

Por último, si en la realidad se quisiese realizar la estimación mediante alguno de estos métodos, la manera en la cual se obtendrían los valores futuros a insertar en el modelo podría ser por medio de recursos internos de la compañía, tales como el departamento de Marketing o de Business Insights o tercerizando el servicio con una consultora especializada que se contrate para dicho fin. Con respecto al horizonte de tiempo a estimar, sería razonable buscar un tiempo futuro que permita desarrollar un plan de negocios para la contratación, inversión, investigación necesaria siempre y cuando se tengan en cuenta que los resultados obtenidos mediante este estudio están muy relacionados a períodos anteriores. Por ejemplo, debajo una imagen de un estudio de valores futuros del precio del petróleo realizado por la consultora McKinsey:



Figure 28: Estudio de McKinsey realizado para la estimación futura de valores del barril del petróleo

6. Tomando decisiones de negocio

Para el desarrollo de este capítulo, no se utilizarán de manera directa los resultados que podrían obtenerse al emplear los modelos del capítulo anterior ya que el foco no debe ponerse en los resultados que puede tener Tenaris al estimar su actividad. Bajo la fundamental premisa de que la precisión con la que predecirá el modelo no es determinista ni un hecho concreto que significa sucederá, obliga a que cada estimación sea tomada como una aproximación que brinda a los equipos correspondientes dentro de la Compañía, la posibilidad de realizar dimensionamientos y preparar los recursos para satisfacer la potencial demanda obtenida por medio de dichos procesos de modelado.

De esta manera, lo que se planteará en este capítulo no es con respecto a los ventas en si ni a las ganancias que podrían obtenerse, sino más bien comentar acerca de que podrían hacer las distintas áreas de Tenaris en el caso de contar con dicha información.

I. Decisiones de negocio en base a la estimación de ventas

Conocer valores futuros de ventas permite a los departamentos de planificación de la demanda, operaciones, procurement de materia prima, servicios logísticos, costos (entre otros) prepararse acorde. Al ser Tenaris una compañía industrial y de manufactura, las plantas que abastezcan dicho mercado deben estar preparadas para poder cumplir con los niveles de producción esperados. Para encontrarse “preparadas”, entre otras consideraciones, debe tenerse en cuenta que:

- Procurement debe asegurar que la planta cuente con la materia prima necesaria para fabricar la cantidad de tubería que se estima se va a vender. A su vez, al conocer la cantidad necesaria a fabricar y la fecha de entrega, podrían surgir descuentos en compras a gran escala en commodities como el mineral de hierro o chatarra (aprovechando temporadas donde estos insumos son más económicos) siempre y cuando sea posible acopiar materia prima y así bajar el costo final del tubo para aumentar los márgenes de ganancia.

Operaciones, a cargo de la producción pero también del funcionamiento de la maquinaria y las líneas, junto al área de planificación e ingeniería deberán decidir si la capacidad

instalada que se cuenta permite cumplir con los niveles estimados o si es necesario realizar alguna inversión que signifique incrementar la capacidad de producción. A su vez, con respecto a la maquinaria deberá asegurar que los equipos se encuentren siempre en óptimas condiciones para ofrecer la calidad comprometida, minimizando errores y cumplir con los tiempos de entrega. Por ende, se pueden hacer más eficientes los esquemas de mantenimiento preventivo y conseguir los repuestos necesarios de antemano dado que se conocerán las licitaciones a las cuales estarán sometidas las líneas/equipos.

- Con respecto al personal y a la dotación a emplear, Recursos Humanos deberá dimensionar los equipos necesarios que permitan cumplir con los turnos que necesitará producción y así realizar contrataciones o re-asignaciones (en caso de capacidad ociosa) para producir de la manera más eficiente brindando el mayor rédito para la compañía. Lo que muchas veces se le pide a Producción puede traducirse a realizar el trabajo necesario con la menor cantidad de gente posible, así apuntando directamente al margen de cada producto. Un tema no menor, para el personal inexperto contratado, es sumamente importante considerar los tiempos de entrenamiento y capacitación que aseguren el conocimiento que debe tener un operario acerca de los aspectos de calidad y seguridad de su trabajo para no condicionar los tiempos y resultados de la producción.
- Para el equipo de costos y finanzas, realizar las correspondientes previsiones de capital y estimaciones de cash-flow que brinden la gimnasia que el negocio necesita para suplir la demanda con las estimaciones pretendidas.
- Por último, con respecto al costo logístico y distribución, siendo que ocupa gran proporción del total de los tubos, podrían buscarse contratos con los transportistas que sean de mejores condiciones por el hecho de comprometerse a mediano plazo a una utilización de trenes/camiones estimada que aseguran a la compañía proveedora una actividad estable durante un horizonte de tiempo. Obteniendo de esta manera, mejores costos por tarifas fijas entre ciertos rangos de actividad y luego valores variables si la actividad fluctúa.

Es necesario aclarar que bajo ningún punto de vista esto significa que hoy la compañía trabaje sin forecast de ventas o de actividad. El análisis que se realiza en torno a la planificación de la producción no es algo que se decide sobre la marcha si no que actualmente se cuenta con un “calendario” de las órdenes que tienen que fabricarse por medio del input de los equipos comerciales y de planificación de la demanda que están en contacto con el cliente. Lo que permitirían los modelos del capítulo anterior sería una planificación más proactiva versus reactiva para poder decidir así cual es la manera más eficiente de fabricar. Esto podría evitar cambios de último momento por órdenes que aparecen/desaparecen por ineficiencias comerciales, urgencias que signifiquen cambios de planes de producción (por ende mayores costos) o en casos extremos, falta de materia de prima.

Es una realidad a considerar, en caso que se quiera extrapolar este análisis a un caso real, que la distribución de productos realizado en este trabajo por medio de la agrupación básica en Premium y No-Premium significa una simplificación que está lejos de describir la realidad. De todas formas, si podría servir para dilucidar cuales deberían las materias primas, líneas de fabricación que se activarán y personal necesario para cumplir con las estimaciones de dos productos distintos (en este caso Premium y no Premium). En un desarrollo más a detalle, podría analizarse cuál sería la actividad por familia de productos significando así un resultado más provechoso para la realidad del negocio y un escenario más realista con respecto a los resultados comerciales.

Con respecto a la actividad de Line Pipe, por más que no se hayan obtenido resultados relevantes a fin de cuentas puede realizarla la simplificación similar al caso de la distribución de ventas Premium y No-Premium. En definitiva, son otro tipo de plantas y procesos industriales para los cuales deberían tenerse las mismas consideraciones respecto a la fabricación, personal y costos.

II. Decisiones de negocio en base a la estimación de Días de Servicio

El principal beneficio de estimar los días de Servicio que Tenaris brindará en el campo es el de poder dimensionar, de forma análoga que con la producción, la dotación de personal necesario. Al ser un trabajo de consultoría on-site y asesoría técnica, no es necesario invertir en capacidad instalada como maquinarias o equipos más allá de sencillos sets de herramientas. Por otro lado, siguiendo en términos de inversión, en caso de valores de actividad que requieran un mayor número de personas será necesario incurrir en la compra de vehículos que brinden movilidad a dichos empleados. Las compañías petroleras suelen contratar de a un consultor que trabaja en turnos de 12 horas supervisando la operación, y se mantiene en guardia el resto del tiempo por si surgen imprevistos. Generalmente, sus tiempos de descanso los pasan en trailers ubicados cerca de los pozos de perforación para los cuales se encuentran a minutos de distancia o en hoteles que pueden estar un poco más alejados. De esta manera, se entiende que trabajan solos y por ende necesitan movilidad propia las 24 horas del día. Por dicho motivo, volviendo a la necesidad de inversión, según el número de personas a contratar, puede ser significativa la cantidad de vehículos que se requieran adquirir. A mayor actividad, más movimiento de personal tendrá la compañía hacia y desde los distintos yacimientos para los cuales es condición irrestricta camionetas tipo 4x4 con altos estándares de seguridad, funcionamiento y mantenimiento.

Desprendiéndose del punto anterior y a diferencia del caso de una planta donde la tarea a realizar puede ser manual, repetitiva o hasta de labor más físico, donde los tiempos para que los empleados obtengan autonomía con mayor celeridad pueden ser menores, el trabajo de consultoría on-site lleva un considerable tiempo de entrenamiento. Por eso, en caso de ser necesario contratar personal deberá también tenerse en cuenta los tiempos de entrenamiento que significan tanto capacitación teórica-técnica como también acompañamiento on-the-job-training junto a empleados más experimentados.

Independientemente del escenario futuro, también serán importantes las previsiones de costos que se hagan considerando dichos niveles de actividad. Una mala previsión puede significar grandes desvíos, gastando de más o de menos, que luego deben ser explicados a los departamentos de contabilidad y finanzas el hecho de contar con vehículos parados o personal con capacidad ociosa, que deberá ser reasignado.

7. Conclusión

Aunque la principal premisa para la generación de los modelos fue el mercado Norte Americano, el sistema entero de plantas de Tenaris localizadas en todo el mundo (incluidas las de USA) es el que luego se ve impactado por potenciales estimaciones de demanda. Dependiendo de los costos, impuestos de importación, cargas de planta, lead-times, entre otros, se define con constante periodicidad la asignación de las órdenes a las plantas encargadas de producir los pedidos de tubería que deben entregarse. Resulta más sensato, al ser un estudio de aproximación, haber centrado el análisis en una zona geográfica en particular antes que enfocarla en todo el mundo. Las variables a considerar para un análisis global serían complejas de definir y a la vez perdería foco, ya que cada unidad de negocios se maneja por separado dependiendo de su mercado, proveedores y particularidades.

Con respecto a los fundamentos y premisas utilizados a lo largo del trabajo para desarrollar los modelos de regresión, es necesario aclarar que las variables explicativas para estimar los modelos fueron aquellas que hacían más sentido desde una perspectiva de negocio. Hubiera sido interesante disponer de otras variables económicas tales como PBI, fluctuaciones del dólar respecto a otras monedas, y tasa de inflación para entender si el impacto/efecto sobre los resultados hubiera sido otro. Por otro lado, el haber tenido más variables hubiera sido más engorroso para la exposición de resultados y definición de modelos finales.

Respecto a los resultados obtenidos, a nivel de negocio y considerando la relevancia que supone para el mismo, el modelo de Ventas Totales es el modelo más rescatable, donde se puede concluir que la variación de las Ventas en un mes futuro t es explicada en un 58% por las variables Oil Price_{t-2} , Gas Price_{t-2} , Libor_t , Ventas_{t-1} y Ventas_{t-2} con un intervalo de confianza mayor al 99%. La parte no explicada por este modelo puede atribuirse a variables exógenas que en esta industria son más que relevantes tales como conflictos comerciales, contextos políticos, guerras contra el terrorismo, sanciones impuestas, etc. A su vez, está el efecto de contratos/licitaciones que Tenaris puede ganar o no que dependiendo el tamaño pueden tener un gran impacto para la unidad de negocios en cuestión. Estos contratos que se ganen pueden a la vez incluir las etapas de distribución y servicios de consultoría lo que necesariamente generará mayor actividad de

Line Pipe y Días de Servicio, cuyos modelos obtenidos no arrojaron resultados provechosos en este trabajo.

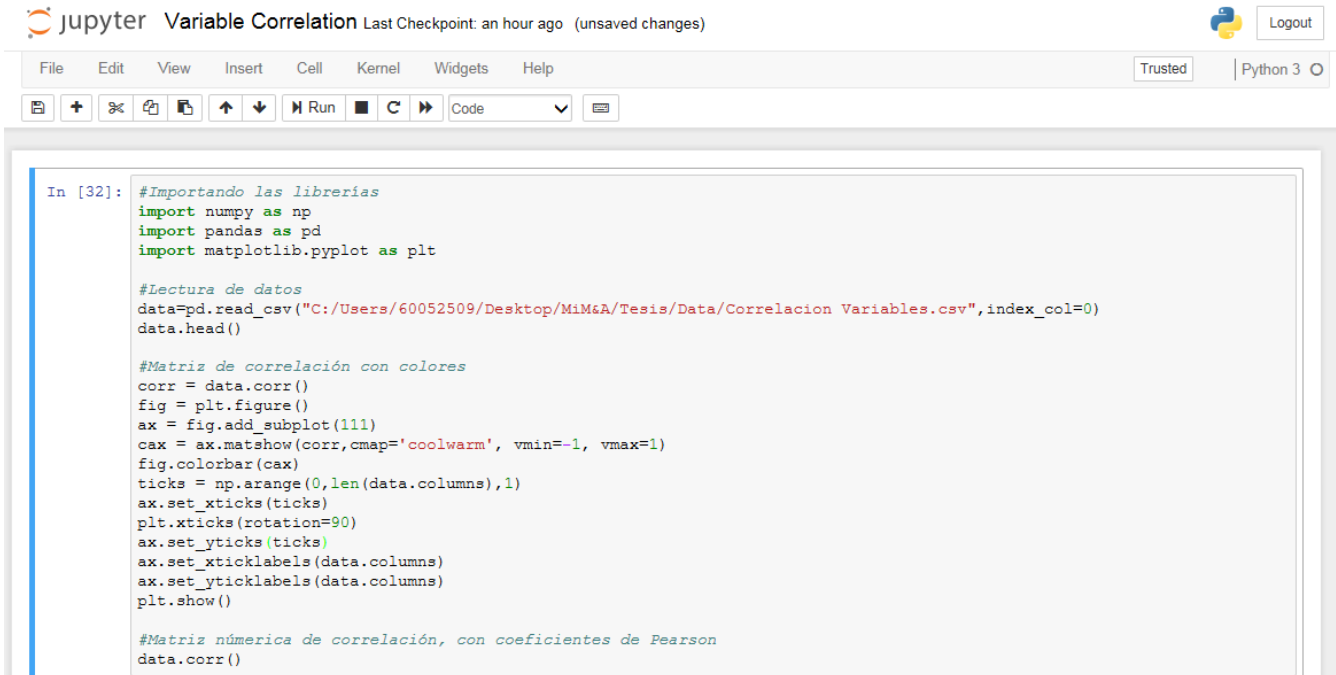
Era de esperarse que la variable Rig Count tenga injerencia en los modelos finales ya que básicamente dicta cuantos pozos están utilizando tubería para perforar. Se evidencia también que las variables que más peso tienen en los modelos son las de rezago de la variable que se quiere estimar.

El horizonte de tiempo a estimar, como ha sido determinado en el capítulo 5, debe ser lo más sensato posible ya que los modelos resultantes necesitan de valores que son de meses que serán potencialmente también desconocidos en caso que se estudie lo que sucede a 3, 6 o 12 meses. Por el contrario, en caso de estimar el nivel de ventas para el mes siguiente, donde si se cuenta con valores de $t-1$ o $t-2$, tiene como perjuicio el hecho de que será muy bajo el tiempo de reacción y preparación que permita el desarrollo y planificación de lo expuesto en el capítulo 6.

Por último, como dicho al principio de este capítulo, es necesario comentar que Tenaris es una compañía Global que opera en distintos mercados y que cada unidad de negocios tiene sus particularidades a nivel local (como aspectos institucionales e incentivos de política económica) como a su vez también del tipo internas como restricciones (i) de presupuesto; (ii) de recursos humanos; (iii) industriales y (iv) tecnológicos. Las decisiones estratégicas que deben tomarse siempre consideran los cuatro puntos anteriores en pos de la búsqueda del bien global con respecto al panorama completo de la cadena de valor de Tenaris. Por este motivo, el resultado obtenido en este trabajo indica que independientemente del modelo estadístico a utilizar o del análisis de mercado que pueda realizarse es posible concluir que la variabilidad de las ventas del mercado Norte Americano de Tenaris puede estimarse en una proporción cercana al 60% siguiendo las principales variables de la industria, lo sucedido en meses anteriores y que también deberá tenerse en cuenta un gran número de variables exógenas que muchas veces son impredecibles y a la vez inconexas unas con otras.

Anexos

Anexo 1 – Running Code en Python-Jupyter para el estudio de Correlación entre variables:



```

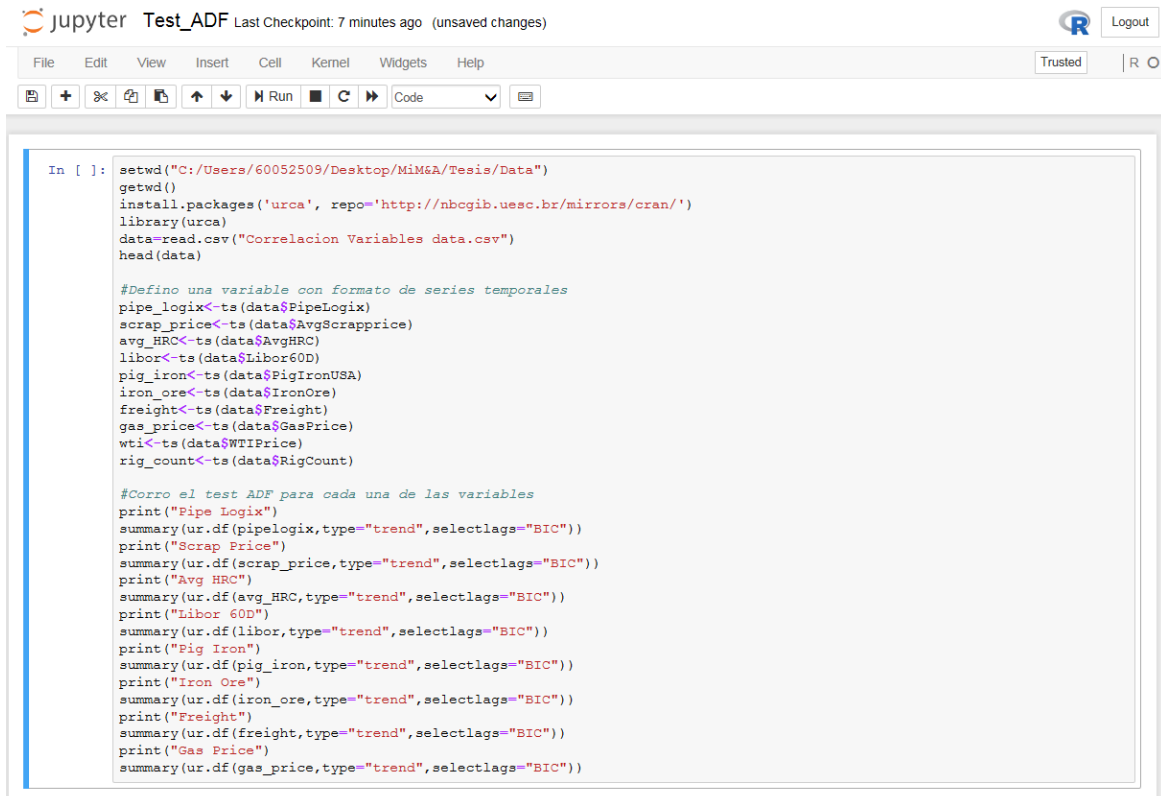
In [32]: #Importando las librerías
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

#Lectura de datos
data=pd.read_csv("C:/Users/60052509/Desktop/MiM&A/Tesis/Data/Correlacion Variables.csv",index_col=0)
data.head()

#Matriz de correlación con colores
corr = data.corr()
fig = plt.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(corr,cmap='coolwarm', vmin=-1, vmax=1)
fig.colorbar(cax)
ticks = np.arange(0,len(data.columns),1)
ax.set_xticks(ticks)
plt.xticks(rotation=90)
ax.set_yticks(ticks)
ax.set_xticklabels(data.columns)
ax.set_yticklabels(data.columns)
plt.show()

#Matriz numérica de correlación, con coeficientes de Pearson
data.corr()
    
```

Anexo 2 – Running Code en R-Jupyter para el test ADF de las variables dependientes:



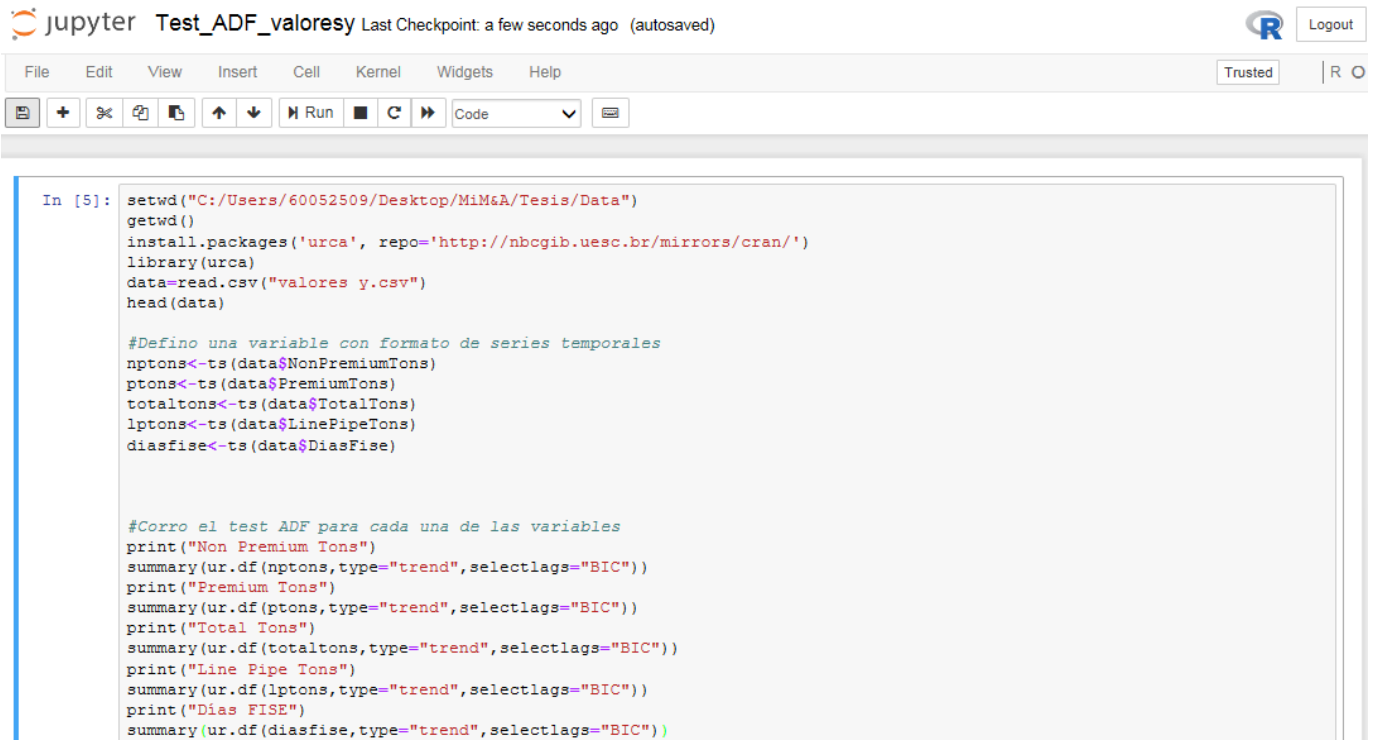
```

In [ ]: setwd("C:/Users/60052509/Desktop/MiM&A/Tesis/Data")
getwd()
install.packages('urca', repo='http://nbcgib.uesc.br/mirrors/cran/')
library(urca)
data=read.csv("Correlacion Variables data.csv")
head(data)

#Defino una variable con formato de series temporales
pipe_logix<-ts(data$PipeLogix)
scrap_price<-ts(data$AvgScrapprice)
avg_HRC<-ts(data$AvgHRC)
libor<-ts(data$Libor60D)
pig_iron<-ts(data$PigIronUSA)
iron_ore<-ts(data$IronOre)
freight<-ts(data$Freight)
gas_price<-ts(data$GasPrice)
wti<-ts(data$WTIPrice)
rig_count<-ts(data$RigCount)

#Corro el test ADF para cada una de las variables
print("Pipe Logix")
summary(ur.df(pipe_logix,type="trend",selectlags="BIC"))
print("Scrap Price")
summary(ur.df(scrap_price,type="trend",selectlags="BIC"))
print("Avg HRC")
summary(ur.df(avg_HRC,type="trend",selectlags="BIC"))
print("Libor 60D")
summary(ur.df(libor,type="trend",selectlags="BIC"))
print("Pig Iron")
summary(ur.df(pig_iron,type="trend",selectlags="BIC"))
print("Iron Ore")
summary(ur.df(iron_ore,type="trend",selectlags="BIC"))
print("Freight")
summary(ur.df(freight,type="trend",selectlags="BIC"))
print("Gas Price")
summary(ur.df(gas_price,type="trend",selectlags="BIC"))
    
```

Anexo 3 – Running Code en R - Jupyter para el test ADF de las variables dependientes:



Jupyter Test_ADF_valoresy Last Checkpoint: a few seconds ago (autosaved) Logout

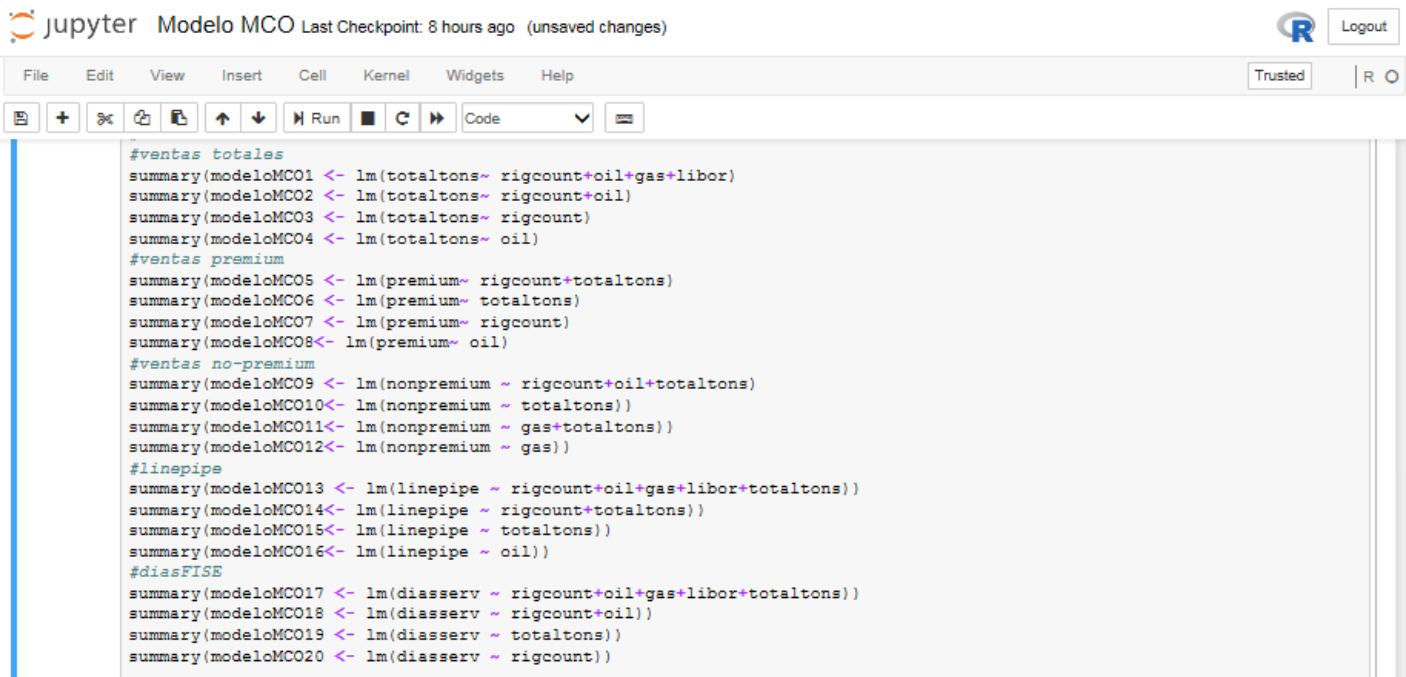
File Edit View Insert Cell Kernel Widgets Help Trusted | R O

```
In [5]: setwd("C:/Users/60052509/Desktop/MiM&A/Tesis/Data")
getwd()
install.packages('urca', repo='http://nbcgib.uesc.br/mirrors/cran/')
library(urca)
data=read.csv("valores y.csv")
head(data)

#Defino una variable con formato de series temporales
nptions<-ts(data$NonPremiumTons)
ptions<-ts(data$PremiumTons)
totaltons<-ts(data$TotalTons)
lptions<-ts(data$LinePipeTons)
diasfise<-ts(data$DiasFise)

#Corro el test ADF para cada una de las variables
print("Non Premium Tons")
summary(ur.df(nptions,type="trend",selectlags="BIC"))
print("Premium Tons")
summary(ur.df(ptions,type="trend",selectlags="BIC"))
print("Total Tons")
summary(ur.df(totaltons,type="trend",selectlags="BIC"))
print("Line Pipe Tons")
summary(ur.df(lptions,type="trend",selectlags="BIC"))
print("Dias FISE")
summary(ur.df(diasfise,type="trend",selectlags="BIC"))
```

Anexo 4 –Running Code en R - Jupyter para los modelos MCO:



Jupyter Modelo MCO Last Checkpoint: 8 hours ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted | R O

```
#ventas totales
summary(modeloMCO1 <- lm(totaltons~ rigcount+oil+gas+libor))
summary(modeloMCO2 <- lm(totaltons~ rigcount+oil))
summary(modeloMCO3 <- lm(totaltons~ rigcount))
summary(modeloMCO4 <- lm(totaltons~ oil))
#ventas premium
summary(modeloMCO5 <- lm(premium~ rigcount+totaltons))
summary(modeloMCO6 <- lm(premium~ totaltons))
summary(modeloMCO7 <- lm(premium~ rigcount))
summary(modeloMCO8<- lm(premium~ oil))
#ventas no-premium
summary(modeloMCO9 <- lm(nonpremium ~ rigcount+oil+totaltons))
summary(modeloMCO10<- lm(nonpremium ~ totaltons))
summary(modeloMCO11<- lm(nonpremium ~ gas+totaltons))
summary(modeloMCO12<- lm(nonpremium ~ gas))
#linepipe
summary(modeloMCO13 <- lm(linepipe ~ rigcount+oil+gas+libor+totaltons))
summary(modeloMCO14<- lm(linepipe ~ rigcount+totaltons))
summary(modeloMCO15<- lm(linepipe ~ totaltons))
summary(modeloMCO16<- lm(linepipe ~ oil))
#diasFISE
summary(modeloMCO17 <- lm(diasserv ~ rigcount+oil+gas+libor+totaltons))
summary(modeloMCO18 <- lm(diasserv ~ rigcount+oil))
summary(modeloMCO19 <- lm(diasserv ~ totaltons))
summary(modeloMCO20 <- lm(diasserv ~ rigcount))
```

Anexo 5 – Running Code en R - Jupyter para los modelos Dinámicos:

Jupyter Modelo Dinámico Last Checkpoint: 7 minutes ago (unsaved changes) R Logout

File Edit View Insert Cell Kernel Widgets Help Trusted | R O

```

In [ ]: summary(modelodin1 <- dyn$lm(totaltons ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ oil+ lag(oil, -1) +
      lag(oil, -2) + gas+ lag(gas, -1) + lag(gas, -2) + libor+ lag(libor, -1) + lag(libor, -2)
      + lag(totaltons, -1) + lag(totaltons,-2) ))
summary(modelodin2 <- dyn$lm(totaltons ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ oil+ lag(oil, -1) + lag(oil,
      + lag(totaltons, -1) + lag(totaltons,-2) ))
summary(modelodin3 <- dyn$lm(totaltons ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2) + lag(totaltons, -1) +
      lag(totaltons,-2) ))
summary(modelodin4 <- dyn$lm(totaltons ~ oil+ lag(oil, -1) + lag(oil, -2) + lag(totaltons, -1) + lag(totaltons,-2) ))
summary(modelodin5 <- dyn$lm(premium ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ oil+ lag(oil, -1)
      + lag(oil, -2) + totaltons + lag(totaltons, -1) + lag(totaltons,-2) + lag(premium, -1)
      + lag(premium,-2) ))
summary(modelodin6 <- dyn$lm(premium ~ totaltons + lag(totaltons, -1) + lag(totaltons,-2) + lag(premium, -1)
      + lag(premium,-2) ))
summary(modelodin7 <- dyn$lm(premium ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2) + lag(premium, -1) +
      lag(premium,-2) ))
summary(modelodin8 <- dyn$lm(premium ~ oil+ lag(oil, -1) + lag(oil, -2) + lag(premium, -1) + lag(premium,-2)))
summary(modelodin9 <- dyn$lm(nonpremium ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ oil+ lag(oil, -1) +
      lag(oil, -2) + totaltons + lag(totaltons, -1) + lag(totaltons,-2) + lag(nonpremium, -1)
      + lag(nonpremium,-2) ))
summary(modelodin10 <- dyn$lm(nonpremium ~ totaltons + lag(totaltons, -1) + lag(totaltons,-2) + lag(nonpremium, -1)
      + lag(nonpremium,-2) ))
summary(modelodin11 <- dyn$lm(nonpremium ~ gas+ lag(gas, -1) + lag(gas, -2) + totaltons + lag(totaltons, -1) +
      lag(totaltons,-2) + lag(nonpremium, -1) + lag(nonpremium,-2) ))
summary(modelodin12 <- dyn$lm(nonpremium ~ gas + lag(gas, -1) + lag(gas,-2) + lag(nonpremium, -1) + lag(nonpremium,-2))
summary(modelodin13 <- dyn$lm(linepipe ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ oil+ lag(oil, -1) + lag(oil,
      gas+ lag(gas, -1) + lag(gas, -2) + libor+ lag(libor, -1) + lag(libor, -2)+ totaltons
      + lag(totaltons, -1) + lag(totaltons,-2) + lag(linepipe, -1) + lag(linepipe,-2) ))
summary(modelodin14 <- dyn$lm(linepipe ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ totaltons +
      lag(totaltons, -1) + lag(totaltons,-2) + lag(linepipe, -1) + lag(linepipe,-2) ))
summary(modelodin15 <- dyn$lm(linepipe ~ totaltons + lag(totaltons, -1) + lag(totaltons,-2) + lag(linepipe, -1) +
      lag(linepipe,-2) ))
summary(modelodin16 <- dyn$lm(linepipe ~ oil + lag(oil, -1) + lag(oil,-2) + lag(linepipe, -1) + lag(linepipe,-2)))
summary(modelodin17 <- dyn$lm(diasserv ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ oil+ lag(oil, -1) +
      lag(oil, -2) + gas+ lag(gas, -1) + lag(gas, -2) + libor+ lag(libor, -1) +
      lag(libor, -2)+ totaltons + lag(totaltons, -1) + lag(totaltons,-2) + lag(diasserv, -1)
      + lag(diasserv,-2) ))
summary(modelodin18 <- dyn$lm(diasserv ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ oil+ lag(oil, -1)
      + lag(oil, -2) + lag(diasserv, -1) + lag(diasserv,-2) ))
summary(modelodin19 <- dyn$lm(diasserv ~ totaltons + lag(totaltons, -1) + lag(totaltons,-2) + lag(diasserv, -1)
      + lag(diasserv,-2) ))
summary(modelodin20 <- dyn$lm(diasserv ~ rigcount+ lag(rigcount, -1) + lag(rigcount, -2)+ lag(diasserv, -1) +
      lag(diasserv,-2) ))

#modelo significativo 1
summary(modelodin1_final <- dyn$lm(totaltons ~ lag(oil, -1) + lag(oil, -2) + lag(gas, -2) )+
      lag(totaltons, -1) + lag(totaltons,-2) ))
#modelo significativo 5
summary(modelodin5_final <- dyn$lm(premium ~ totaltons + lag(premium, -1)))
#modelo significativo 9
summary(modelodin9_final <- dyn$lm(nonpremium ~ totaltons + lag(totaltons, -1) + lag(nonpremium, -1) ))
    
```

Anexo 6 – Running Code en R - Jupyter para Modelos Autometrics

Jupyter **Modelo Autometrics** Last Checkpoint: Last Thursday at 10:48 PM (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted | R O

```

In [ ]: from <- as.Date("2014-01-01")
to <- as.Date("2019-03-01")
date <- seq.Date(from=from,to=to,by="month")
timeIndex <- as.Date(date)
explicativas=read.csv("explicativas modelo.csv")
dependientes=read.csv("dependientes modelo.csv")
attach(explicativas)
attach(dependientes)
#variables explicativas
rigcount <- (zoo(explicativas$RigCount,order.by=timeIndex))*100
gas <- (zoo(explicativas$GasPrice,order.by=timeIndex))*100
oil <- (zoo(explicativas$WTIPrice,order.by=timeIndex))*100
libor <- (zoo(explicativas$Libor60D,order.by=timeIndex))*100
#variables dependientes
totaltons <- (zoo(dependientes $ TotalTons,order.by=timeIndex))*100
nonpremium <- (zoo(dependientes $ NonPremiumTons,order.by=timeIndex))*100
premium <- (zoo(dependientes $ PremiumTons,order.by=timeIndex))*100
linepipe <- (zoo(dependientes $ LinePipeTons,order.by=timeIndex))*100
diasserv <- (zoo(dependientes $ DiasFISE,order.by=timeIndex))*100

#modelado

#ventas totales con las 4 variables
matriz_explicativas <- cbind(rigcount, lag(rigcount, -1), lag(rigcount, -2),oil, lag(oil, -1)
, lag(oil, -2), gas, lag(gas, -1) , lag(gas, -2) , libor, lag(libor, -1) ,lag(libor, -2))
gum <- arx(totaltons, ar=1:2, mxreg= matriz_explicativas)
getsm(gum)

#ventas no premium
matriz_explicativas <- cbind(rigcount, lag(rigcount, -1), lag(rigcount, -2),oil, lag(oil, -1) , lag(oil, -2), gas,
lag(gas, -1) , lag(gas, -2) , libor, lag(libor, -1) ,lag(libor, -2),totaltons,
lag(totaltons, -1), lag(totaltons,-2))
gum <- arx(nonpremium, ar=1:2, mxreg= matriz_explicativas)
getsm(gum)

#linepipe
matriz_explicativas <- cbind(rigcount, lag(rigcount, -1), lag(rigcount, -2),oil, lag(oil, -1) , lag(oil, -2), gas,
lag(gas, -1) , lag(gas, -2) , libor, lag(libor, -1) ,lag(libor, -2),totaltons,
lag(totaltons, -1), lag(totaltons,-2))
gum <- arx(linepipe, ar=1:2, mxreg= matriz_explicativas)
getsm(gum)

#dias FISE
matriz_explicativas <- cbind(rigcount, lag(rigcount, -1), lag(rigcount, -2),oil, lag(oil, -1) , lag(oil, -2), gas,
lag(gas, -1) , lag(gas, -2) , libor, lag(libor, -1) ,lag(libor, -2),totaltons,
lag(totaltons, -1), lag(totaltons,-2))
gum <- arx(diasserv, ar=1:2, mxreg= matriz_explicativas)
getsm(gum)
    
```

Fuentes bibliográficas:

Capítulo 2:

Global Trend 2025 – NIC USA:

https://www.files.ethz.ch/isn/94769/2008_11_Global_Trends_2025.pdf

New Geopolitics of Energy: <https://es.weforum.org/agenda/2017/11/la-nueva-geopolitica-de-la-energia/>

8 reasons why politics of oil have changed: <https://www.weforum.org/agenda/2016/02/eight-reasons-why-the-politics-of-oil-have-changed/>

Capítulo 3:

Datos precio histórico del Petróleo: <https://es.investing.com/commodities/crude-oil-historical-data>

Datos precio histórico del Gas: <https://www.indexmundi.com/es/precios-de-mercado/?mercancia=gas-natural&meses=120>

International Rig Count: <https://bakerhughesrigcount.gcs-web.com/intl-rig-count?c=79687&p=irol-rigcountsintl>

Pipe Logix: <https://www.pipe-logix.com/octg.html>

Coefficiente de Pearson:

- https://www.uv.es/webgid/Descriptiva/31_coeficiente_de_pearson.html
- Material visto en Métodos Estadísticos Aplicados a Negocios

Studying correlation in Python – Matriz con colores:

<https://medium.com/@sebastiannorena/finding-correlation-between-many-variables-multidimensional-dataset-with-python-5deb3f39ffb3>

Correlation in Python – Valores de correlación: <http://benalexkeen.com/correlation-in-python/>

Relationship between oil prices and Exchange rate:

https://www.eia.gov/workingpapers/pdf/oil_exchangerates_61317.pdf

Capítulo 4:

Tenaris OCTG: <http://www.tenaris.com/es-es/products/octg.aspx>

Tenaris Line Pipe:

<http://www.tenaris.com/en/Products/OnshoreLinePipe/OilAndGasPipeline.aspx>

Capítulo 5:

US Energy commission: <https://www.eia.gov/petroleum/>

Spurious relations: <https://www.econometrics-with-r.org/14-7-nit.html>

How does industry impact on companies:

<https://onlinelibrary.wiley.com/doi/epdf/10.1002/%28SICI%291097-0266%28199707%2918%3A1%2B%3C15%3A%3AAID-SMJ916%3E3.0.CO%3B2-1>

Test ADF: https://en.wikipedia.org/wiki/Augmented_Dickey%E2%80%93Fuller_test

Econometrics with R: <https://www.econometrics-with-r.org/14-7-nit.html>

Regresiones:

Explicación R²: https://www.rcumariacristina.com/wp-content/uploads/2010/12/11-Elena-Martinez_1.pdf

Time series in R: <https://www.statmethods.net/advstats/timeseries.html>

How to interpretate p-values in model regression:

<https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-interpret-regression-analysis-results-p-values-and-coefficients>

<https://infocenter.informationbuilders.com/wf80/index.jsp?topic=%2Fpubdocs%2FRStat16%2Fsource%2Ftopic41.htm>

Modelos Dinámicos:

https://repositorio.itesm.mx/bitstream/handle/11285/572320/DocsTec_2068.pdf?sequence=1

Fundamentos modelo Autometrics:

https://www.researchgate.net/profile/Alvaro_Veiga/publication/280894773_Comparing_variable_selection_techniques_for_linear_regression_LASSO_and_Autometrics/links/56094dcd08ae4d86bb11b61f.pdf

Capítulo 7:

Oil Industry insights - McKinsey: https://www.mckinsey.com/solutions/energy-insights/global-oil-supply-demand-outlook-to-2035/~/_media/352F6A10E20B4D85B0C426CBE2CFE88D.ashx

Demanda mercado global: <https://seekingalpha.com/article/4235846-faced-near-term-uncertainties-international-markets-drive-tenaris-growth>