

Parte 1

Modelos de pronóstico de corto plazo para precios del petróleo WTI spot

Por **Claudio Araya Sassi**

Ingeniero de estudios, Empresa Nacional del Petróleo, Chile.

Aspectos generales y antecedentes

La relevancia del petróleo en la economía mundial es indiscutible. Según el Departamento de Energía de Estados Unidos la producción mundial de petróleo alcanzó 84,8 millones de barriles por día (Mbd) el año 2007, de los cuales el 41,9% fue producido por los países de la OPEP, esto es, 35,5 (Mbd). Si los países de la OPEP mantienen alrededor del 80% de las reservas mundiales de petróleo, los países de la OCDE y los europeos pueden contar solamente con el 7% y 0,8%, respectivamente. Sin embargo, alrededor del 60% del consumo mundial de petróleo se origina en los países de la OCDE. El impacto del petróleo en los mercados financieros es a lo menos igual de importante. El promedio semanal del volumen del *open interest* (OI) de crudo WTI, que corresponde a la totalidad de contratos de futuros de petróleo fue igual a 634.207 contratos durante el período 2002-2005, el cual se incrementó a 1.069.938 durante 2006 y a 1.395.269 durante 2007 (fuente: *Commodity Futures Trading Commission*, 2007).

La relevancia del petróleo en las economías mundiales hace necesario evaluar el desempeño de los pronósticos de los diferentes modelos del petróleo WTI spot presentes en la literatura. También importa proponer un nuevo modelo que permita pronosticar los precios del crudo WTI en el corto plazo. Este estudio, que se publicará dividido en tres números consecutivos de *Petrotecnia*, evalúa diferentes modelos econométricos de corto plazo de petróleo WTI spot en función de distintas variables financieras y estructurales. La primera parte, que aquí presentamos, tratará los aspectos generales y los antecedentes; la segunda ahondará en las metodologías utilizables y la tercera incluirá los resultados y las conclusiones, como así también la bibliografía y los sitios consultados para llevar a cabo la investigación.

Antecedentes

1. Modelos de pronóstico de corto plazo de precios del petróleo

La literatura sobre los precios del petróleo se ha enfocado en dos principales clases de modelos econométricos de ecuación singular de forma reducida. El primer grupo corresponde a los *modelos financieros* que incluyen aquellos que están inspirados en la teoría económica financiera y basados en la Hipótesis de Eficiencia del Mercado (HEM), mientras que los modelos que pertenecen a la segunda clase corresponden a los *modelos estructurales* que consideran los efectos de los agentes del mercado del petróleo y variables reales sobre los precios del petróleo. Ambos modelos, financieros y estructurales, a menudo usan especificaciones de series de tiempo puras para la comparación referencial o benchmarking.

1.1. Modelos financieros y de series de tiempo

En general, los modelos financieros para pronóstico de los precios del petróleo examinan la relación entre

los precios del petróleo spot y los del petróleo futuros, y analizan en particular si los precios de futuros son predictores eficientes e insesgados de los precios spot (Longo *et al.*, 2007). El modelo posee la siguiente especificación:

$$S_t = \beta_0 + \beta_1 F_{t-1} + \varepsilon_t \quad (1)$$

Donde:

S_t : precio del petróleo spot en el tiempo t .

F_{t-1} : precio del futuro de petróleo en el tiempo $t-1$ con vencimiento T .

β_0, β_1 : parámetros.

ε_t : término de error aleatorio o estocástico.

Donde la hipótesis nula conjunta de insesgamiento ($\beta_0=0$ y $\beta_1=1$) no debería ser rechazada y no se debería encontrar autocorrelación en los términos de error (eficiencia). Un rechazo de la hipótesis nula conjunta sobre los coeficientes β_0 y β_1 es usualmente racionalizada por la literatura en términos de la presencia de un premio/castigo por riesgo en función del tiempo.

Un subgrupo de modelos, los cuales también están basados en la teoría financiera pero han sido menos investigados, explota la siguiente relación de arbitraje entre los precios spot y futuros:

$$F_t = S_t e^{(r + \omega - \delta)(T-t)} \quad (2)$$

Donde:

S_t : precio del petróleo spot en el tiempo t .

F_t : precio del petróleo futuro en el tiempo $t-1$ con maduración T .

r : tasa de interés.

ω : costo de almacenamiento (tasa).

δ : tasa de dividendos.

Esta relación de arbitraje significa que los precios futuros deben ser iguales al costo de financiamiento de la adquisición del activo spot hoy y mantenerlo hasta la fecha de vencimiento de los futuros (lo cual incluye costo de la deuda para la adquisición inicial, o tasa de interés, y cualquier costo de almacenamiento) una vez que los continuos dividendos que retribuyen el activo subyacente (rentabilidad de conveniencia) hayan sido tomados en cuenta.

Asimismo, Zeng y Swanson (1998), realizan una comparación entre modelos financieros y de series de tiempo, usando datos diarios sobre el período desde el 4 de enero de 1990 al 31 de octubre de 1991, especificando una caminata aleatoria (*random walk*), un modelo autorregresivo y dos modelos alternativos de corrección de error (MCE), cada uno con una definición diferente de equilibrio de largo plazo, que se presentan en detalle más adelante.

1.2. Modelos estructurales y de series de tiempo

Los modelos estructurales enfatizan la importancia de variables explicatorias que describen las características peculiares del mercado del petróleo. Algunos ejemplos son ofrecidos por variables que son estratégicas para este mercado como los niveles de inventario comercial y gubernamental, variables "reales", tales como consumo y producción, y variables que toman en cuenta el papel desempeñado por la OPEP en el mercado internacional del petróleo.

1.2.1. Modelos estructurales de mercado de petróleo

Recientemente, Déas *et al.* (2007) especifican un modelo de pronóstico trimestral que considera una muestra desde el año 1984 a 2002. Este modelo consiste en tres bloques dados por las ecuaciones de demanda, oferta y precios. En este estudio los autores ponen particular atención en el comportamiento de la OPEP, usando como regresores estructurales la cuota de la OPEP (definida como la cantidad de petróleo para ser producida por los miembros de la OPEP), capacidad de utilización y la razón entre los stocks de petróleo y demanda de la OCDE (Organización para la Cooperación y Desarrollo Económico). Usando un modelo de corrección de errores, los autores demuestran que la OPEP es capaz de influenciar los precios reales del petróleo, mientras que su especificación econométrica es capaz de producir pronósticos estáticos y dinámicos exactos dentro de la muestra.

1.2.2. Modelos estructurales de inventario de petróleo

Algunos autores introducen el rol del nivel de inventario relativo de petróleo, definido como la desviación de los inventarios de petróleo desde su nivel normal, como un determinante adicional de los precios. Debido a que los niveles de inventario de petróleo son una medida del balance, o desbalance, entre la demanda y producción de petróleo, ellos reflejan presiones cambiantes del mercado sobre los precios del crudo, y así proveen un buen barómetro del precio del petróleo que varían en el corto plazo.

Ye *et al.* (2002) presentan un modelo de pronóstico mensual de corto

plazo para los precios spot del petróleo West Texas Intermediate usando niveles de inventario de petróleo de la OCDE. Este modelo fue construido basado en una comprensión de los aspectos fundamentales del mercado del petróleo y el comportamiento observado del mercado durante el período post Guerra del Golfo desde enero de 1992 a febrero de 2001. Los resultados del modelo de pronóstico demuestran empíricamente que los inventarios de petróleo son un buen indicador de mercado de los cambios del precio del petróleo.

El análisis comienza con el supuesto de que existe una relación entre precio y fundamentos de oferta-demanda. La variable inventario relativo es una medida explícita del balance entre producción y demanda. Así, el inventario relativo es una buena variable para indicar el estado de balance de la oferta-demanda y cómo impacta el precio.

El nivel de inventario relativo, denotado por RIN_t , es definido como:

$$RIN_t = IN_t - IN_t^* \quad (3)$$

Donde:

IN_t : es el nivel de inventario de petróleo total OCDE, gubernamental y comercial, en el mes t , y
 IN_t^* : es el nivel normal, o deseado.

El nivel de inventario deseado es calculado por:

$$IN_t^* = a_0 + b_1 t + \sum_{k=2}^{12} b_k D_k \quad (4)$$

Donde:

D_k , $k = 2, 3, \dots, 12$ son 11 variables estacionales mensuales.

a_0 , b_1 , y b_k , $k = 2, \dots, 12$ son coeficientes estimados desde detendencialización y desestacionalización desde el inventario de petróleo OCDE total observado. Los resultados de la regresión muestran un patrón estacional estadísticamente significativo y también un consistente crecimiento positivo para este período.

El modelo de pronóstico lo especificó de la siguiente forma:

$$WTI_t = a + \sum_{i=0}^5 b_i RIN_{t-i} + \sum_{i=0}^5 c_i \Delta LIN_{t-i} + dAIN_t + eWTI_{t-1} + \varepsilon_t \quad (5)$$

Donde:

WTI_t : es el precio spot nominal de crudo WTI;

t : es para el t -ésimo mes;

i : es para el i -ésimo mes previo al t -ésimo mes;

a , b_i , c_i , d , y e , para $i = 0, 1, 2, \dots, 5$ son coeficientes para ser estimados;

RIN : como es definido en (3), es el nivel de inventario relativo total OCDE, esto es, las desviaciones de los niveles de stocks actuales, IN , desde sus correspondientes niveles normales deseados, IN^* .

LIN : es una variable de inventario bajo, definida para capturar el comportamiento del mercado asimétrico caracterizado por una respuesta de diferencia de precios ante cambios de inventario cuando el nivel de inventario es bajo al nivel normal deseado y cuando el nivel de inventario está sobre el normal, que es:

$$LIN_t = RIN_t, \text{ si } IN_t < IN_t^* \text{ y } LIN_t = 0 \quad (6)$$

o en cualquier otro caso;

$$AIN_t = IN_t - IN_{t-12} \quad (7)$$

AIN_t : corresponde a las diferencias anuales en el inventario mensual; esta variable se incorpora para reflejar el comportamiento cíclico del mercado no captado por la remoción de la estacionalidad y tendencia;

ε : es el término aleatorio.

Nótese que el modelo expresado en (5) puede ser visto en el marco de un modelo de ajuste parcial, donde RIN es el inventario relativo (la desviación de IN desde su correspondiente nivel normal deseado, IN^*).

Con el objetivo de reducir la multicolinealidad, (5) fue transformado isomórficamente y estimado como:

$$WTI_t = a + \sum_{i=0}^4 b_i \Delta RIN_{t-i} + b_5 RIN_{t-5} + \sum_{i=0}^4 c_i \Delta LIN_{t-i} + c_5 LIN_{t-5} + LIN_{t-5} + dAIN_t + eWTI_{t-1} + \varepsilon_t \quad (8)$$

Siguiendo a Ye *et al.* (2002), Merino y Ortiz (2005) especifican un MCE con el porcentaje de inventario relativo industrial de petróleo y "especulación", definidas como las posiciones "largas" mantenidas por los agentes no-comerciales de petróleo, gasolina y heating oil en el mercado de futuros de NYMEX, como variables explicatorias. Evidencia desde enero de 1992 a junio de 2004 demuestra que la especulación puede mejorar significativamente el modelo de inventario propuesto por Ye *et al.*, especialmente en la última parte de la muestra.

Zamani (2004) propone el siguiente modelo de pronóstico trimestral para el período 1988-2004:

$$WR_t = \alpha_1 + \alpha_2 QU_t + \alpha_3 OV_t + \alpha_4 SIS_t + \alpha_5 SGS_t + \alpha_6 DN_t + \alpha_7 D90 + u_t \quad (9)$$

Donde:

WR: precio real de petróleo WT1

QU: cuota OPEP

OV: sobreproducción de la OPEP

SIS: desviación desde el nivel normal del nivel de inventario comercial de la OCDE

SGS: Desviación desde el nivel normal del nivel de inventario estratégico (SPR)

D90: variable dicótoma o dummy para la guerra de Irak en el 3er y 4to trimestre de 1990.

α_i : parámetros

Luego verifica que existen relaciones de cointegración entre todas las variables por lo que reconstruye la ecuación (9) en un modelo de corrección de error, el cual separa los comportamientos de corto y largo plazo. Este modelo lo estableció de la siguiente forma:

$$\Delta WR_t = \beta_0 + \sum_{i=0}^5 \beta_{1i} \Delta QU_{t-i} + \sum_{i=1}^m \beta_{2i} \Delta OV_{t-i} + \sum_{i=1}^m \beta_{3i} \Delta SIS_{t-i} + \sum_{i=1}^m \beta_{4i} \Delta SGS_{t-i} + \sum_{i=1}^m \beta_{5i} \Delta DN_{t-i} + \beta_6 D90_{t-1} + \beta_7 \mu_{t-1} + \varepsilon_t \quad (10)$$

Donde:

Δ indica el operador de primera diferencia en las variables,

μ_{t-1} : es el error o residual rezagado un período del modelo (9), el cual captura el efecto de largo plazo y corrige el desequilibrio de corto plazo hacia el equilibrio de largo plazo.

2. Teoría de series de tiempo

2.1. No estacionariedad de series de tiempo

Toda la teoría de regresión está basada en series de tiempo estacionarias. Una serie se dice estacionaria si la media y autocovarianza de ésta no dependen del tiempo.

Un ejemplo común de una serie no estacionaria es la caminata aleatoria:

$$y_{t-1} + \varepsilon_t \quad (11)$$

Donde:

ε_t es el término estacionario de perturbación aleatoria.

La serie y tiene un valor de pronóstico constante, condicional en t , y la varianza se incrementa sobre el tiempo. La caminata aleatoria es una serie estacionaria de diferencia cuando la primera diferencia de y es estacionaria:

$$y_t - y_{t-1} = \varepsilon_t \quad (12)$$

Una serie estacionaria de diferencia se dice que está integrada y es denotada como $I(d)$, donde d es el orden de integración. El orden de integración

es el número de raíces unitarias contenidas en las series, o el número de operaciones de diferenciación que toma hacer la serie estacionaria. Para la caminata aleatoria de arriba, existe una raíz unitaria, así que es una serie (1). Similarmente, una serie estacionaria es $I(0)$.

Los procedimientos estándares de inferencia no se aplican a regresiones que contienen una variable dependiente integrada o regresores integrados. Por lo tanto, es importante chequear si una serie es estacionaria o no antes de usarla en una regresión. El método formal para probar la estacionariedad de una serie es la prueba de raíz unitaria.

2.2. Prueba de raíz unitaria sobre estacionariedad

La forma más fácil de introducir esta prueba es considerar el siguiente modelo:

$$y_t = y_{t-1} + \varepsilon_t \quad (13)$$

Donde:

ε_t : término de error estocástico que sigue los supuestos clásicos, a saber: tiene media cero, varianza constante σ^2 y no está autocorrelacionado (término de error ruido blanco).

La ecuación (13) es una regresión de primer orden, o AR(1), en la cual se efectúa la regresión del valor de y en el tiempo t sobre su valor en el tiempo $(t-1)$.

Ahora bien, si el coeficiente de y_{t-1} es igual a 1, surge lo que se conoce como el problema de raíz unitaria, es decir, una situación de no estacionariedad. Por consiguiente, si se efectúa la regresión

$$y_t = \rho y_{t-1} + \varepsilon_t \quad (14)$$

y se encuentra que $\rho = 1$, entonces se dice que la variable estocástica y_t tiene una raíz unitaria o corresponde a una caminata aleatoria.

Sustrayendo y_{t-1} a ambos lados de la ecuación (14)

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \varepsilon_t$$

y arreglando:

$$\Delta y_t = (\rho - 1) y_{t-1} + \varepsilon_t \quad (15)$$

Se obtiene:

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t \quad (16)$$

Donde:

$$\delta = \rho - 1$$

Δ : operador de primera diferencia

Por lo tanto, la hipótesis nula es que $0 = \delta$. Si δ es en realidad 0, se puede escribir (16) como:

$$\Delta y_t = (y_t - y_{t-1}) = \varepsilon_t \quad (17)$$

La ecuación (17) dice que la primera diferencia de una serie de tiempo de caminata aleatoria ($= \varepsilon_t$) es una serie de tiempo estacionaria porque, por supuestos, ε_t es puramente aleatoria.

2.3. Regresión espuria

Al efectuar la regresión de una variable de serie de tiempo sobre otra variable de serie de tiempo, con frecuencia se obtiene un R^2 muy elevado aunque no haya una relación significativa entre las dos. Esta situación ejemplifica el problema de **regresión espuria**. Este problema surge porque si las dos series de tiempo involucradas presentan **tendencias** fuertes (movimientos sostenidos hacia arriba o hacia abajo), el alto R^2 observado se debe a la presencia de la tendencia y no a la verdadera relación entre las dos. Por consiguiente, es muy importante averiguar si la relación entre las variables económicas es verdadera o es espuria. Por ejemplo, cuando se efectúa una regresión entre dos series de tiempo, puede haberse efectuado una regresión de una serie de tiempo no estacionaria sobre otra igualmente no estacionaria (caminata aleatoria). En ese caso, los procedimientos estándares de pruebas t y F no son válidos. En este sentido, la regresión es espuria.

2.4. Cointegración

Debido a que muchas series de tiempo económicas parecen seguir caminatas aleatorias, esto sugiere que se desee, de manera típica, diferenciar una variable antes de usarla en una regresión. Aunque esto es aceptable, la diferenciación puede dar como resultado una pérdida de información acerca de la relación a largo plazo entre dos variables. La mayor parte de la teoría económica se postula con base en relaciones de largo plazo entre las variables en niveles, es decir, no en forma de primera diferencia. Así, la hipótesis del ingreso permanente de Milton Friedman postula que el nivel del consumo permanente es función del nivel del ingreso permanente; la relación no está planteada en términos de las primeras diferencias de estas variables.

El hallazgo de que muchas series de tiempo macroeconómicas pueden contener una raíz unitaria ha motivado el desarrollo de la teoría de análisis de series de tiempo no estacionarias. Engle y Granger (1987) señalan que una combinación lineal de dos o más series no estacionarias puede ser esta-

cionaria. Si tal combinación lineal estacionaria existe, las series de tiempo no estacionarias están **cointegradas**. La combinación lineal estacionaria es llamada la **ecuación de cointegración** y puede ser interpretada como la relación de equilibrio de largo plazo entre las variables.

Supongamos la siguiente regresión entre 2 variables:

$$y_t = \beta_1 + \beta_2 x_t + u_t \quad (18)$$

Donde:

y_t : variable dependiente

x_t : variable independiente o explicatorio

β_1, β_2 : parámetros estimados

u_t : error aleatorio

Además, supongamos que se ha determinado que ambas variables son procesos no estacionarios $I(1)$ o caminatas aleatorias. A pesar de esto, la combinación de estas dos variables podría ser estacionaria. Más específicamente, si se escribe (18) como:

$$u_t = y_t - \beta_1 - \beta_2 x_t \quad (19)$$

y se encuentra que u_t [es decir, la combinación lineal $(y_t - \beta_1 - \beta_2 x_t)$] es $I(0)$ o estacionaria, entonces se dice que las variables x_t e y_t están cointegradas; es decir, están "en la misma longitud de onda". Intuitivamente se observa que cuando u_t en (19) es $I(0)$, las "tendencias" en x_t e y_t se cancelan. Y éstas estarán "en la misma longitud de onda" si son integradas del mismo orden. Así, si una serie Y es $I(1)$ y otra serie X es también $I(1)$, ellas pueden estar cointegradas. En general, si Y es $I(d)$ y X es también $I(d)$, donde d es el mismo valor, estas dos series pueden estar cointegradas. Si ése es el caso, la regresión de las dos variables en niveles, como en (18), es significativa (es decir, no es espuria); y no se pierde información valiosa de largo plazo, lo cual sucedería si se utilizaran sus primeras diferencias.

En resumen, siempre y cuando se verifiquen que los residuales de las regresiones como (18) son $I(0)$ o estacionarios, la metodología tradicional de regresión (incluyendo las pruebas t y F) es aplicable a las series de tiempo. La contribución valiosa de los conceptos de estacionariedad, raíz unitaria y cointegración, es que obliga a determinar si los residuales de la regresión son estacionarios.

En el lenguaje de la teoría de la cointegración una regresión tal como (18) se conoce como **regresión cointegrada** y el parámetro β_2 , es conocido como **parámetro cointegrado**.

2.5. Mecanismo de corrección de errores (MCE)

Supongamos que se ha demostrado que X e Y están cointegradas; es decir, hay una relación de equilibrio de largo plazo entre las dos. Por supuesto, en el corto plazo, puede haber desequilibrio. En consecuencia, se puede tratar el término de error en (19) como el **error de equilibrio**. Y se puede utilizar este término de error para ligar el comportamiento de corto plazo de Y con su valor de largo plazo. El **mecanismo de corrección de errores** (MCE), popularizado por Engle y

Granger, corrige el desequilibrio. Como ejemplo sencillo del MCE, consideremos el siguiente modelo:

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_t + \beta_3 \hat{u}_{t-1} + \varepsilon_t \quad (20)$$

Donde:

Δ : denota la primera diferencia.

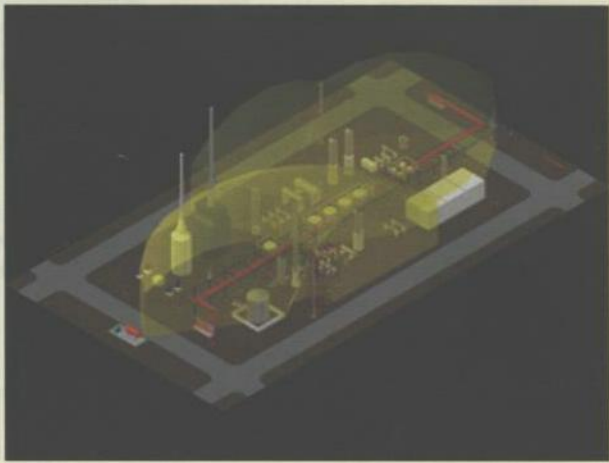
$\hat{u}_{t-1}$: es el valor del residual de la regresión (18) rezagado un período, la estimación empírica del término de error de equilibrio.

ε_t : es el término usual de error aleatorio.

Este trabajo técnico continúa en la próxima edición de *Petrotecnica*: "Modelos de pronóstico de corto plazo para precios del petróleo WTI spot". Segunda parte: Metodologías. ■


TEX
 PROTECCION CONTRA INCENDIOS



TEX ARGENTINA provee soluciones de Ingeniería, equipos y recursos, para cubrir los requerimientos específicos de Protección contra Incendio en la industria, esencialmente de Oil & Gas, tanto en operaciones aguas adentro (OFF SHORE), como terrestres, brindando los siguientes servicios:

- ▣ Proyectos de Ingeniería Conceptual, Básica y de Detalle
- ▣ Inspección, prueba y diagnóstico de instalaciones existentes
- ▣ Validación de Diseño e Ingenierías Reversas
- ▣ Diseño bajo normas y estándares nacionales, internacionales y corporativos
- ▣ Simulaciones hidráulicas y de eventos
- ▣ Desarrollo, construcción y suministro de equipamiento y soluciones especiales
- ▣ Comercialización de equipos y sistemas





Tex Argentina SRL
 Campos Salles 2154 2° piso - (C1429CFD) Capital Federal
 Tel: 54 11 4702-4002 / 4003 email: texar@texarsrl.com.ar