

ORACLE®

CRYSTAL BALL

**Modelación, Pronósticos, Back-Testing y Análisis del Precio Spot y
Demanda de la Energía Eléctrica Horaria en El Salvador**

Luis Francisco Zaldívar, MSE

Consultor en Finanzas, Minería de Datos y Riesgos

l.zaldivar@modelacionderiesgos.com

www.modelacionderiesgos.com

Noviembre 2015

**INVERSIONES
CARROUSEL**
S.A. DE C.V.

INDICE

- I. Propósito**
- II. Modelación de Precios Spot Horario de la Energía Eléctrica**
- III. Descomposición de la Serie de Tiempo & Visualización**
- IV. Etapas del Proceso de Pronósticos**
- V. ¿Es Estacionaria la Serie de Tiempo del Precio Spot Horario?**
- VI. Modelo # 1 Propuesto de Pronóstico del Precio Spot**
- VII. Modelo # 2 Propuesto de Pronóstico del Precio Spot Con ARIMA**
- VIII. Modelo # 3 Propuesto del Pronóstico del Precio Spot con Serie de Tiempo Avanzada**
- IX. ¿Es posible crear un Modelo de Pronóstico del Precio Spot Horario Combinado de un Modelo de Serie de Tiempo y Regresión Múltiple Auto regresiva?**
- X. Bibliografía**
- XI. Hoja de Vida de Luis Francisco Zaldívar, MSE**

I. Propósito

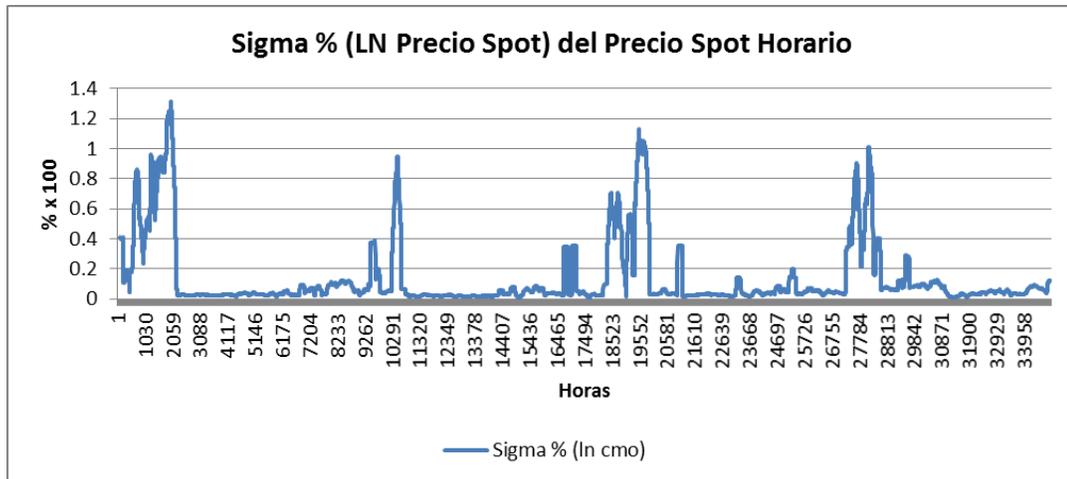
El presente trabajo brinda al lector la metodología, la tecnología de software, las técnicas de pronósticos y el análisis de riesgos que es necesario que el ejecutivo moderno conozca y aplique en sus trabajos empresariales. Los cálculos de estos pronósticos siempre son un reto, dado que los estimados calculados enfrentan riesgos e incertidumbres a medida que más se alejan del presente, principalmente si las volatilidades históricas son altas.

Es importante saber que existen dos análisis en el tema de series de tiempo, que son el Análisis Descriptivo y Predictivo. El Análisis Descriptivo es necesario siempre realizarlo para revisar que ha pasado o sucedido con la variable de nuestro interés a la fecha. Por el contrario, el Análisis Predictivo proyecta cual será la dirección, magnitud y variación que tomará la variable de interés en el futuro. Es crítico que cuando iniciemos un proyecto de pronósticos, siempre iniciemos con un buen Análisis Descriptivo.

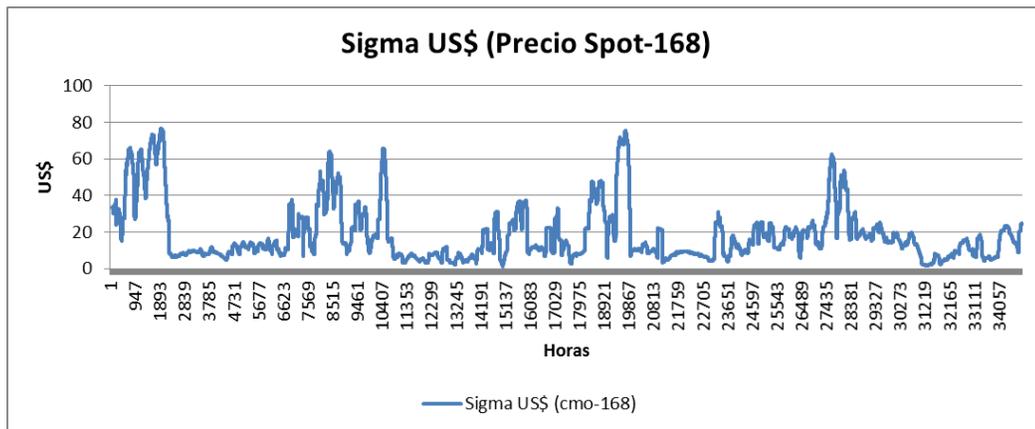
Se ha considerado seleccionar **los precios spot horario** (\$ por KWH) y su **demanda** (MWH) del Mercado Eléctrico de El Salvador, Centro América. Todos los valores empleados en el presente trabajo han sido tomados de la información de la Unidad de Transacciones S.A de C.V. (UT) de El Salvador, Centro América. Su sitio web es www.ut.com.sv.

El precio spot de la energía eléctrica es el producto que posee la mayor volatilidad, dado que estos precios experimentan muchos saltos o cambios extremos por la diferente demanda que existe durante las 24 horas del día. Adicionalmente, existen muchos efectos estacionales, cíclicos y exógenos que hacen que la modelación y pronóstico del precio spot de la electricidad sea un reto realizarla. Las volatilidades **diarias** de varios productos se detallan a continuación: Notas del Tesoro y Bonos menor del 0.5%; Índices Bursátiles entre 1.5% - 4.0%; Precio del Petróleo y Gas entre 1.5% - 4.0%; Precio de Acciones Volátiles menor del 4.0%; **sin embargo el precio de la electricidad tiene volatilidad extrema de hasta el 50.0% diaria: es el productos más volátil en el mundo.**

En el siguiente grafico se puede apreciar la tendencia horaria de la volatilidad de los precios spot en el Mercado de El Salvador, la cual cambia con mucha amplitud en el tiempo.



A continuación se presenta la volatilidad medida en US\$:



El aprendizaje del tema de pronósticos se hace emocionante cuando usamos el producto más volátil como el valor o precio spot horario de la electricidad, dado que si somos capaces de realizar este tipo de pronóstico, no será complicado realizar otros pronósticos de demanda o precios de otros productos o servicios que poseen menores volatilidades y son más estables en el tiempo. En el presente trabajo el horizonte de pronósticos será de corto plazo y estaremos empleando Modelos de Regresión, Serie de Tiempo y Combinación de ellos.

II. Modelación de Precios Spot de Energía Eléctrica

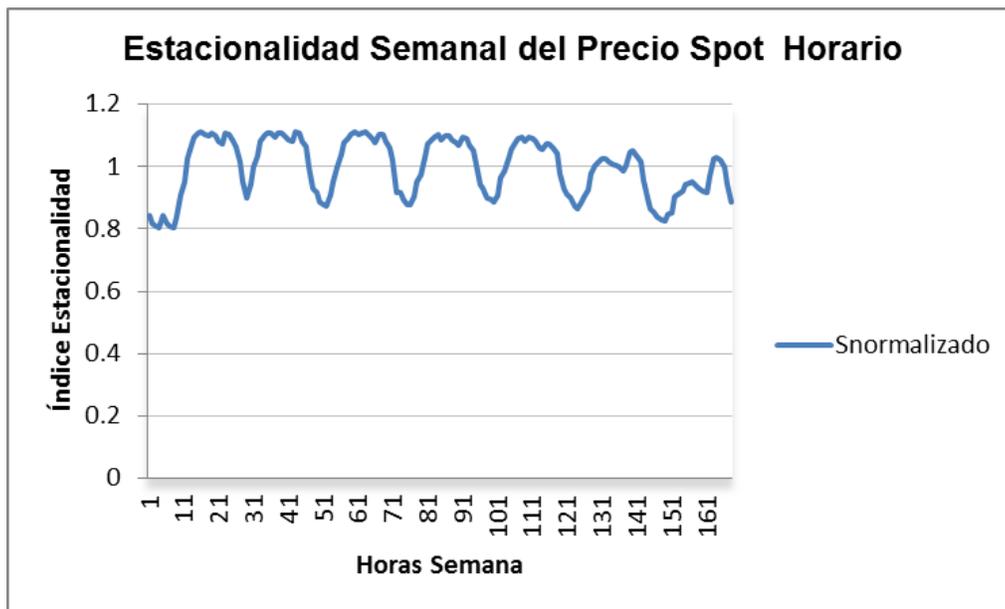
La modelación de los precios spot horarios es compleja debido a que para modelar la semana, 7 días o 168 horas, se deben de hacer tres modelos. Haciendo tres modelos la complejidad de modelación se reduce considerablemente. ¿Porqué se ha seleccionado la modelación de la semana o 168 horas? Esto se debe a que

realizando una estación de 168 horas hace que a las tendencias T y Ciclo C se les elimine el Factor Estacional e Irregular. Es por esto que las series de tiempo de la electricidad a nivel mundial se diferencian en 1,24 y finalmente 168 horas para encontrar el factor remanente que es el error aleatorio de la serie de tiempo. Si hacemos el análisis de 24 horas, el factor tendencia no se logra eliminar por ningún método empleado.

Los modelos serían:

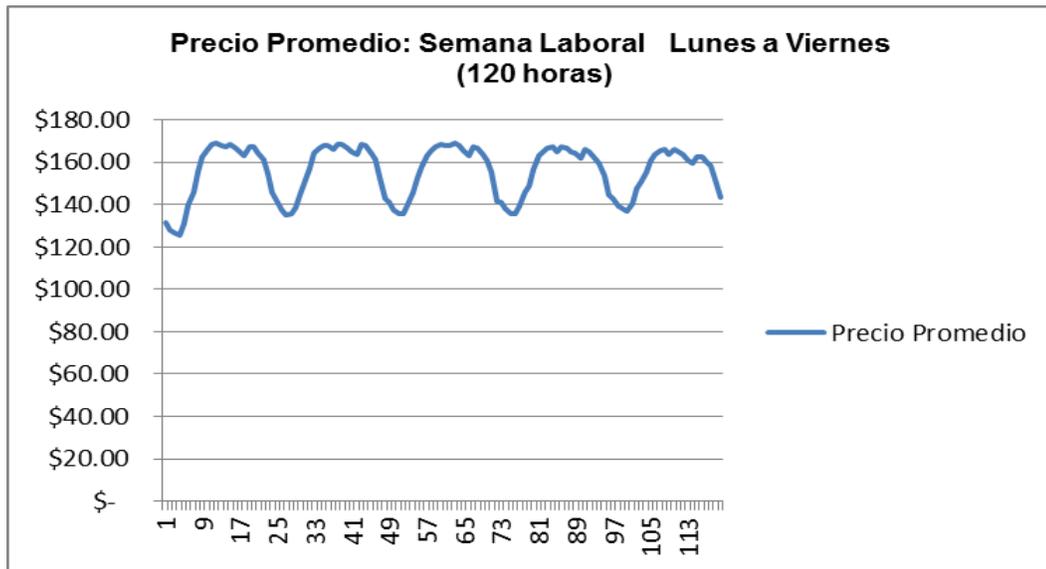
1. Modelo semana laboral de lunes a viernes, es decir 120 horas.
2. Modelo de sábado que tomaría 24 horas.
3. Modelo de domingo que tomaría 24 horas.

Esto se debe principalmente a las características de estacionalidad principalmente que tienen los precios de la energía eléctrica. Durante la semana laboral los precios tienen la misma estacionalidad. Luego, el fin de semana se presenta diferente estacionalidad y niveles de precios. Siendo el domingo el día que los precios de la energía son los más bajos por la baja demanda. Lo descrito en el párrafo anterior lo podemos descubrir en el siguiente gráfico de Índices de Estacionalidad Horaria, durante una semana, o sea 168 horas. Se observa que los primeros ciclos hasta las 120 horas son muy similares, luego los de fin de semana cambian en su forma y niveles.



A continuación presentamos una gráfica donde se demuestra que el primer modelo sería la semana laboral, es decir 120 horas. Se puede apreciar en la gráfica que los mismos ciclos se repiten durante los días laborales de lunes a viernes teniendo los

mayores precios al inicio de la jornada laboral y cuando todos los trabajadores llegan a sus casas a finales de la tarde.



II. Descomposición de la Serie de Tiempos & Visualización

Existen muchas formas de realizar la descomposición de las series de tiempo. En el presente trabajo hemos empleado el **Modelo Clásico Multiplicativo**. Este modelo se formula a continuación:

$$Y_t = T_t \times C_t \times S_t \times I_t$$

Donde,

Y_t = Precio Spot Horario en tiempo t

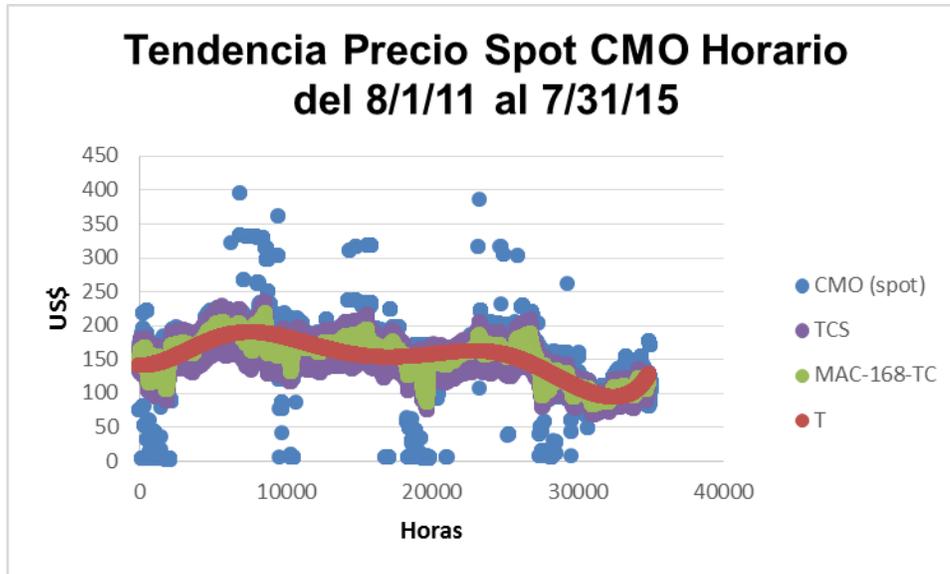
T_t = Tendencia en el tiempo t

C_t = Factor Cíclico en el tiempo t

I_t = Factor Aleatorio en tiempo t

Si la información a emplear es anual, la variable S_t se elimina del modelo.

Se ha aplicado la descomposición al precio spot horario desde el 8 de enero de 2011 hasta el 31 de julio de 2015. A continuación presentamos el gráfico de descomposición.



El color azul son los precios spot históricos extremos que no han sido capturados por el **Modelo Clásico Multiplicativo** descrito anteriormente (CMO Spot). Luego, todos los demás precios si han sido capturados por el modelo. La interpretación es la siguiente:

1. La Tendencia (T) de color ladrillo muestra un movimiento de reversión al promedio. Es decir tiene inflexión y velocidad de cambio.
2. El Factor Estacional es el que mayor banda de variación posee, superior a la Tendencia y Factor Ciclo (MAC-168-TC). Pueden, comparar las bandas morada versus la verde.
3. Se hace vital modelar tanto el error extremo de los puntos azules y la estacionalidad de esta serie de tiempo.
4. Sin embargo, lo que más riesgo e incertidumbre posee es la Tendencia y Ciclo (MAC-168-TC) dado que esto poseen punto de inflexión o cambio. Luego, los factores de estacionalidad son más predictivos que las otras variables en la descomposición. La diferenciación de 168 horas elimina el factor estacional y aleatorio como se muestra en la función en el gráfico anterior.

Tomando en cuenta lo antes expuesto en los literales anteriores del 1-4, la modelación debe de buscar un modelo que tenga las siguientes características:

1. Capturar estacionalidad
2. Poder ajustarse a un cambio de dirección o inflexión

3. El error del pronóstico deberá capturar los valores extremos que se tienen cuando hay baja o extrema demanda de energía. La modelación del error será muy crítica para reducir el error de pronóstico.

Durante el horizonte analizado, se ha podido clasificar la variación de los Precios Spot Horarios como se destalla en el cuadro siguiente:

Análisis de Precios Spot

Escenario	Numero Horas	%
Bajo	845	2.42%
Banda - 2 Sigma - aprox 95%	33,534	96.09%
Extremo	518	1.48%
Total	34,897	100%

El Precio Spot Horario de la energía eléctrica tiene aproximadamente un 96.09% de Comportamiento normal. Sin embargo, tiene precios extremos, saltos o Spikes con dirección hacia abajo del 2.42% y arriba de 1.48%.

Esto es lo que ha motivado en mercados eléctricos en países desarrollados a ofrecer opciones put-call para que los participantes puedan mitigar los riesgos de mercado de este producto.

III. Etapas del Proceso de Pronósticos

Se debe de presentar el proceso básico para poder elaborar pronósticos de cualquier tipo. A continuación se muestra el proceso y una descripción de la etapa;

1. **Objetivo:** Lo más importante es conocer qué se desea pronosticar, y el horizonte de tiempo requerido, exactitud y frecuencia.
2. **Base de datos:** Tradicionalmente necesitábamos información. Sin embargo, hoy día también se puede realizar pronósticos sin necesidad de tener información histórica. La técnica que más se emplea cuando no hay información o la calidad es dudosa, es la Simulación Monte Carlo. La aplicación de la Simulación Monte Carlo se hace con mucha facilidad con software como Oracle Crystal Ball en ambiente de MS Excel.
3. **Análisis de los datos:** Usualmente la información hay que revisarla para ver si todos los campos vienen completos y si no existe valores extremos ocasionados por errores. Hoy en día, los modeladores estamos interesados en los valores extremos aunque ocasionan dificultad para encontrar los modelos, sin embargo "El demonio está en las colas o valores extremos de la

información". No podemos dejarla de usar dado que también son forma de estresar todo.

4. **Visualización:** Se hace muy crítica esta actividad dado que de analizar la información en un gráfico podemos descubrir sus patrones de tendencia, estación, ciclo, magnitudes y cambios.
5. **Pre-procesamiento:** Después de desarrollar la Visualización, podemos pensar en realizar transformaciones a los datos. Usualmente, la transformación que más se emplea es transformar al logaritmo natural para darle menos variación, hacer los datos lineales y tratar de hacer los datos más normales. También se puede en esta etapa ver el Delta Tiempo de nuestro análisis.
6. **Partición:** Esta etapa es crítica dado que se debe de separar información para construir el modelo y luego probarlo con otra información no tomada en cuenta para probar el modelo. Usualmente, como son series de tiempo, se usa un 90% de la información disponible para construir el modelo y el 10% restante para probar el modelo. Si se generan pronósticos sin esta partición, se tendrá el llamado Error de Sobre Ajuste. Es decir, cuando desarrollamos el modelo el error de pronóstico es bien bajo, pero cuando se aplique a la información real, el error será magnificado. También es muy frecuente emplear el período de prueba o back-testing que sea igual al ciclo de la estacionalidad de la información. Es decir, si tenemos estación de 12 meses, podemos emplear 1 mes. Si la estacionalidad es de 168 horas, podemos emplear 168 horas. Si la estacionalidad es de trimestres, podemos emplear un trimestre.
7. **Aplicación de modelos de pronósticos:** Teniendo todas las etapas descritas en los literales 1-6, debemos de probar varios modelos. En el caso del el software Oracle Crystal Ball, ya el módulo de Predictor tiene una gran gama que pueden ser su inicio de modelos a emplear. También, con el simulador de este software puede crear sus propios modelos de pronósticos.
8. **Evaluación de resultados de los modelos de pronósticos:** La evaluación debe ser en por lo menos dos dimensiones: La primera es minimizar el error de pronósticos. Existen muchas métricas como RMSE, MAD, MAPE, BIC, AIC, etc. Sin embargo, la segunda dimensión es la calidad, es decir el supuesto que tienen los modelos. Ambas dimensiones van de la mano. Por ejemplo, en las serie de tiempo usualmente las métricas de calidad son dos: Durbin-Watson (DW) y THIEL. El primero DW mide el nivel de auto correlación y el segundo THIEL mide si es beneficioso emplear el modelo técnico versus un simple o pronóstico naive.
9. **Pronósticos:** Ya habiendo seleccionado el mejor modelo de pronóstico. Se procede a realizar el pronóstico deseado.

10. **Administración, control y ajuste al modelo de pronóstico empleado:** La parte más crucial del pronóstico es administrarlo y controlarlo. Es decir, una vez hemos instalado un modelo de pronóstico debemos de medir su efectividad según la frecuencia del back-testing o ajuste y cuando este salga de los límites determinados. Se debe de volver a realizar el proceso o etapa del 6-8. Esto se debe a que muchos productos son complejos que no solo cambian los supuestos sino que hacen que cambien el modelo o la técnica de pronósticos.

IV. ¿Es Estacionaria la Serie de Tiempo del Precio Spot Horario?

Es muy importante estar seguro que la serie de tiempo es estacionaria. ¿Por qué? Desde el punto de vista práctico, si la serie de tiempo es estacionaria, el método de pronóstico que hemos desarrollado será útil para emplearlo para proyectar el futuro. Si la serie no es estacionaria, quiere decir que ese método solo sirve para la situación o momento actual. La buena noticia es que toda serie no estacionaria se puede hacer estacionaria, entre los métodos más prácticos es hacerle diferenciación a la serie de tiempo o emplear transformaciones. El ser estacionaria la serie de tiempo también hace que las variables promedio y varianza sean estables. Hoy en día con la gama de software disponible, se puede calcular la prueba estadística de Dickey-Fuller Aumentada (ADF Test) o prueba de Raíz Unitaria. Se ha empleado el software muy idóneo para análisis de series de tiempo NumXL®, el cual brinda los resultados favorables que si es estacionaria la serie del precio spot horario, los resultados de la prueba ADF Test al nivel del 5.0 % son:

Stationary Test					
Test	Stat	P-Value	C.V.	Stationary?	5.0%
ADF					
<i>No Const</i>	-0.6	47.1%	-1.9	FALSO	
<i>Const-Only</i>	-6.4	0.1%	-2.9	VERDADERO	
<i>Const + Trend</i>	-6.4	0.0%	-1.6	VERDADERO	
<i>Const+Trend+Trend^2</i>	-8.4	0.0%	-1.6	VERDADERO	

En el cuadro a de arriba podemos comprobar con la prueba estadística ADF Aumentada que existen tres modelos que son estacionarios, los cuales se presenta el resultado de la prueba en color amarillo.

V. Modelo Propuesto de Pronóstico del Precio Spot

Después de haber realizado un Análisis Descriptivo del precio spot horario de la energía eléctrica se desea hacer un Modelo Predictivo del precio spot horario pero deseamos saber el nivel de error y la calidad del modelo a emplear. **El modelo propuesto es solamente para la semana laboral de lunes a viernes o sea 120 horas.**

El modelo predictivo que se propone es el siguiente:

$$P_{t+1} = F(D_t, P_{t-1}, E_t)$$

Donde,

P_{t+1} = Precio a estimar en la hora $t+1$

D_t = Demanda en la hora t

P_{t-1} = Precio retrasado una hora $t-1$

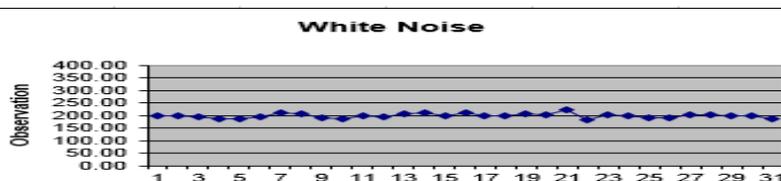
E_t = Error de pronóstico en tiempo t

Recordemos que en el tema de pronósticos es muy importante el **Error de Pronóstico**:

El Error de pronósticos se define como:

E_t = Precio Actual o Real - Precio Estimado que brinda el modelo.

Una prueba importante al modelo de pronóstico es que el error no debería llevar ningún efecto de tendencia T , estación S y ciclo C para que podamos decir que el modelo propuesto captura todos los elementos endógenos y exógenos que inciden en el valor del precio spot horario. Idealmente, el error de pronóstico debería no tener ninguna relación con los demás errores. Es decir su auto correlación medida por Durbin-Watson debería ser 2, el coeficiente de correlación retrasado un periodo ρ_{t-1} debería de ser 0, o el Factor Z absoluto menor de 2. Si estas condiciones se presentan, se dice que el Error Posee Ruido Blanco. Gráficamente, el Ruido Blanco se puede apreciar en la siguiente gráfica:



Pueden apreciar que su varianza es constante y su tendencia tiende a ser constante, cuya pendiente es muy cerca de 0.

El Factor Z Absoluto es muy frecuentemente usado para probar que existe Ruido Blanco en el Error de Pronóstico. Su cálculo es básico y práctico:

Factor Z ABS= Raíz (# de observaciones) x Rho t-1

Se dice que si el valor de la ecuación anterior es menor que 2, los datos del error de pronóstico siguen Ruido Blanco o son independientes uno del otro.

VI. Modelo Predictivo # 1 Propuesto

Estamos empleando la fecha para construir el modelo del 12/31/2014 al 5/22/2015. Luego la prueba del modelo es del 5/25/015 al 5/29/2015.

A continuación se muestra la información en la hoja de cálculo de MS Excel:

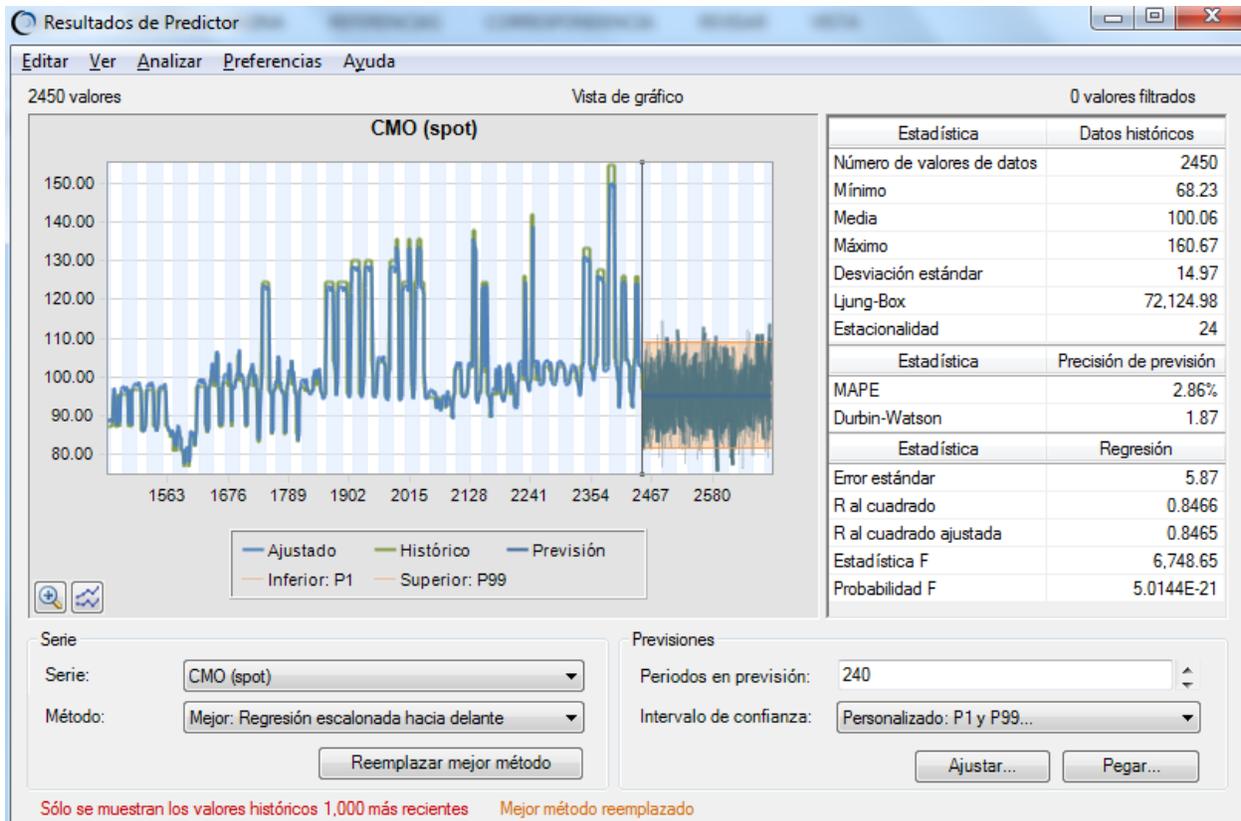
# Datos	FECHA	HORA	Estación Ciclo	# Datos	Precio Spot	Demanda MWH	Lag 1 Precio Spot
1	12/31/2014	22	71	1	\$ 86.58	604.63	\$ 86.58
2	12/31/2014	23	72	2	\$ 86.58	582.82	\$ 86.58
3	1/1/2015	0	73	3	\$ 86.15	547.85	\$ 86.15
4	1/1/2015	1	74	4	\$ 83.12	485.16	\$ 83.12
5	1/1/2015	2	75	5	\$ 83.12	438.59	\$ 83.12
6	1/1/2015	3	76	6	\$ 78.89	411.58	\$ 78.89
7	1/1/2015	4	77	7	\$ 78.89	399.55	\$ 78.89
8	1/1/2015	5	78	8	\$ 78.89	395.22	\$ 78.89

La variable de pronóstico el precio spot (color celeste) y las variables independiente son la demanda (MWH) y Lag 1 precio spot retrasado una hora (color amarillo). En el cuadro anterior se puede observar que tanto el precio spot y lag 1 precio spot son iguales, dado que el Módulo de Predictor del software Oracle Crystal Ball internamente y automáticamente hace los ajustes del precio spot retrasado.

Estaremos empleando el Módulo de Predictor de Oracle Crystal Ball para facilitarnos el pronóstico.

Una de las facilidades que tiene el Módulo de CB Predictor es que se puede construir modelos de regresión con más de 15 variables independientes con mucha facilidad retasar variables sin necesidad de modificar los datos en la hoja de MS Excel y calcula los valores de las variables independientes.

Los resultados del Módulo de Predictor, lo presentamos a continuación:



El modelo de regresión encontrado es el siguiente:

$$P_{t+1} = 6.23 + 0.0070 D_t + 0.8848 P_{t-1} + E_t$$

D-W 1.87 , R² 84.65% y MAPE del 2.86%

Donde:

P_{t+1} = Precio Estimado

D_t = Demanda en tiempo t

P_{t-1} = Precio retrasado en tiempo t-1

E_t = Error de pronóstico

En la pantalla de Módulo Predictor, podemos observar lo siguiente:

1. La confiabilidad del modelo de regresión con la Probabilidad F que es menor del 1% de tasa de corte, lo que indica que las dos variables independientes la demanda y precio spot retrasado una hora tienen poder predictivo para estimar el futuro precio spot horario.

2. El R^2 ajustado es igual al 84.55% que es un nivel muy bueno. Esto indica que la demanda y el precio spot retrasado explican altamente el comportamiento de la variable de pronóstico.
3. El error estándar de la regresión es del US\$5.87, lo que equivale a un 8.60% a 3.65% de los precios spot mínimos y máximos de la serie de tiempo.
4. El error del pronóstico medido por MAPE es de solamente el 2.86%.
5. La auto correlación medida por métrica estadística Durbin-Watson asciende a 1.87. Los rangos de esta prueba deben de ser entre 1.5 - 2.5, pero idealmente debería tender a ser cerca de 2.0. Lo que significa en el error solamente ha quedado como componente aleatorio. Más adelante someteremos el error residual a las pruebas del Ruido Blanco.

Recordemos, que los modelos de regresión son muy valiosos debido a que podemos obtener con mucha exactitud la tendencia en cualquier serie de tiempo. Adicionalmente, los modelos de regresión aceptan que podamos tener varios factores o variables independientes que expliquen el comportamiento o se pueda estimar la variable de pronóstico.

Sin embargo, los modelos de regresión deben de cumplir varios requisitos para que los análisis que resultan de la aplicación correcta de los modelos sea beneficiosa para predecir el futuro o determinar relaciones de causa y efecto con información transversal.

En regresión múltiple, los requerimientos o condiciones se detallan a continuación:

1. La relación entre las variables dependientes e independientes deberá ser aproximadamente lineal.
2. No deberá existir variables independientes redundantes o superfluas para eliminar el problema de Multicolinealidad.
3. La Varianza del Pronóstico Y que brinde el modelo versus el error residual absoluto o al cuadrado deberá no demostrar varianza variable. Es decir, para los diferentes niveles de pronósticos, el error de pronósticos absoluto o al cuadrado deberá no ser cambiante. Lo mismo se aplica a la relación entre el Pronóstico Y del modelo versus los valores de las variables independientes.
4. El error de pronóstico o los residuales deberían ser modelados por la Distribución Normal. Sin embargo, la violación de este requerimiento no es tan grave siempre que las distribuciones tengan Sesgo de 0 y la Curtosis o el excedente de Curtosis superior a los niveles de la Distribución Normal que son 3 en Crystal Ball o 0 en MS Excel. Usualmente, las distribuciones que suelen resultar son la Student's T o Logística cuyas Curtosis son más densas que la de la curva de Gauss.
5. El error de pronóstico debería de no tener auto correlación, es decir tener un Durbin-Watson de 2 o una Prueba Z absoluta menor o igual de 2.

Dado que el modelo a evaluar presenta inicialmente resultados buenos, vamos a probar el cumplimiento de los requisitos o condiciones para posteriormente aplicarlo a los datos que fueron sacados del análisis para poder hacer el back-testing o prueba del modelo.

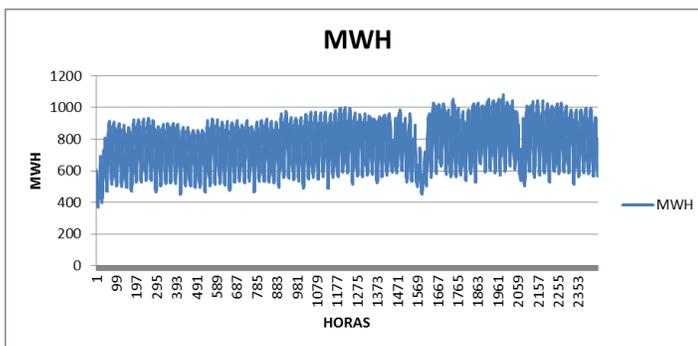
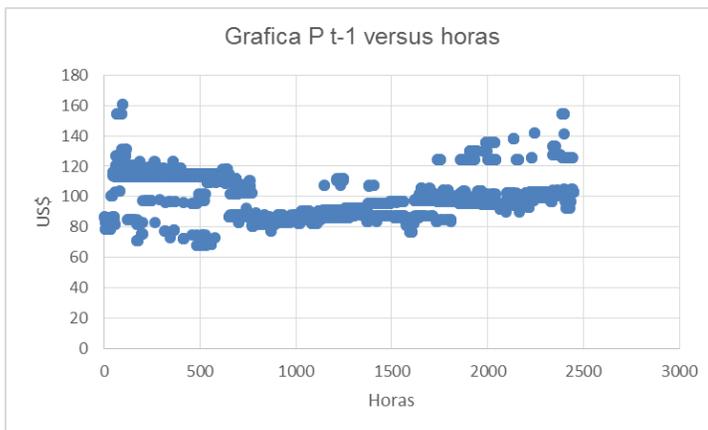
ANÁLISIS DE SUPUESTOS DE LOS MODELOS DE REGRESIÓN

Los resultados son los siguientes:

1. ¿Es Y/ Factores X Lineal ?

Si se obtiene un R^2 alto con en este modelo, lo más seguro es que se cumpla esta condición. Si no se cumpliera, se pueden realizar transformaciones.

Los gráficos de las dos variables independientes versus la variable de pronóstico se presentan a continuación:



Ambas gráficas pasan las pruebas y consideramos que no se hace necesario realizar una transformación.

2. ¿ Existen variables independientes redundantes Multicolinearidad?

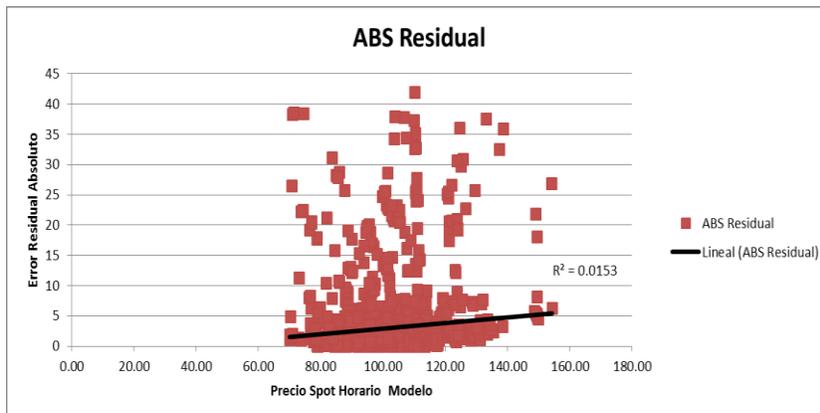
El análisis de redundancia de variables o evaluación de la Multicolinearidad la brinda el Factor de Inflación VIF del software Oracle Crystal Ball. A continuación, presentamos del reporte que brinda el Módulo Predictor:

Variable	Coefficient	VIF
Constant	6.23	-
MWH	0.0070	1.24
Lag CMO (spot) (Lag 1)	0.8848	1.24

Se puede apreciar que el VIF es menor de 10, con lo cual las dos variables independientes son importantes y no sobra ninguna.

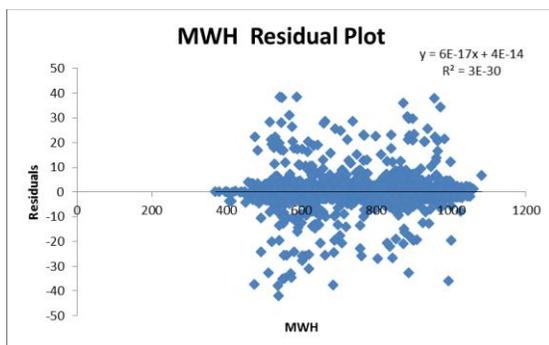
3. ¿ Tiene Varianza Constante o Homocedasticidad?

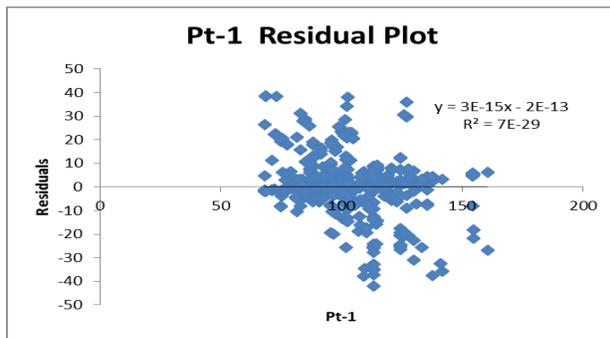
La prueba de Homocedasticidad o Varianza Constante se presenta empleando gráficos básicos elaborados en MS Excel:



Podemos observar en la gráfica anterior que su R^2 es menor al 0.30 o 30%, lo cual nos indica que se cumple que la variación del error es homocedástica.

Las gráficas de las dos variables independientes se presentan a continuación y sus resultados demuestran también el cumplimiento del requisito o condición de homocedasticidad, ambas tienen un R^2 menor del 0.30 o 30%.



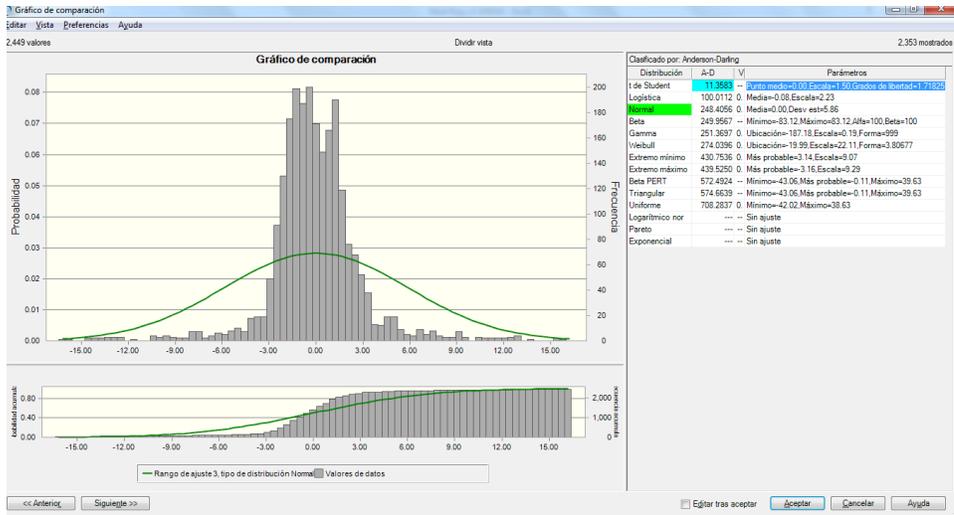


4. ¿Cuál es la distribución de error de pronóstico o residual?

Para responder este requerimiento de la regresión hemos empleado uno de los reportes que genera el Módulo de Predictor del software Oracle Crystal Ball, el cual detallamos a continuación:

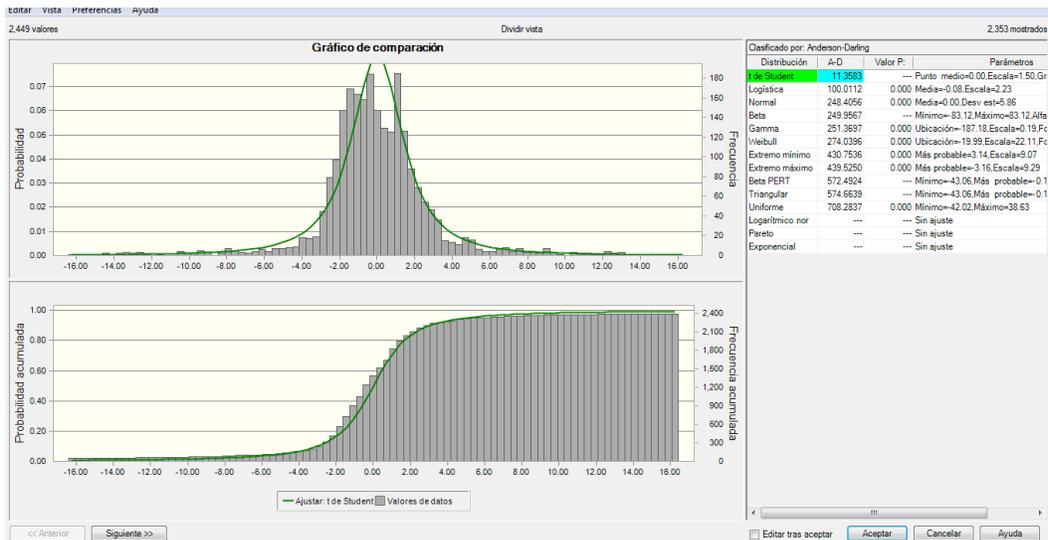
Results Table for Demo Reg.xlsx!Datos					
Created 11/17/2015 at 1:00 PM					
Series	CMO (spot) <input type="checkbox"/>				
	Data				
Date	Historical Data	Lower: 15%	Fit & Forecast	Upper: 85%	Residuals
1	86.58				
2	86.58		86.91		-0.3295
3	86.15		86.67		-0.5150
4	83.12		85.85		-2.7264
5	83.12		82.84		0.2800
6	78.89		82.65		-3.7612
7	78.89		78.82		0.0654
8	78.89		78.79		0.0957
9	78.89		78.61		0.2777
10	78.89		78.69		0.2048
11	78.89		78.92		-0.0316
12	78.89		79.16		-0.2690

Tomando todos los residuales y usando el ajuste de bondad individual que está disponible en la pantalla de configuración del supuesto en el software Oracle Crystal Ball, obtenemos los siguientes resultados:



En el reporte anterior vemos que se tienen dos distribuciones de probabilidad adelante de la Distribución Normal que son la t de Student y la Logística. Esto se debe a que modelan mejor los valores extremos que se presentan en la conducta de los precios extremos de la electricidad. La función verde que pueden ver en la gráfica anterior es la Distribución Normal aplicada a los datos. Puede verse que existe una sub estimación en la parte del medio de la distribución.

El mejor ajuste de bondad, que es la distribución T de Student puede verse que tiene un ajuste casi perfecto y no demuestra mucha sub-estimación en el centro. A continuación les presentamos el ajuste de bondad T de Student:

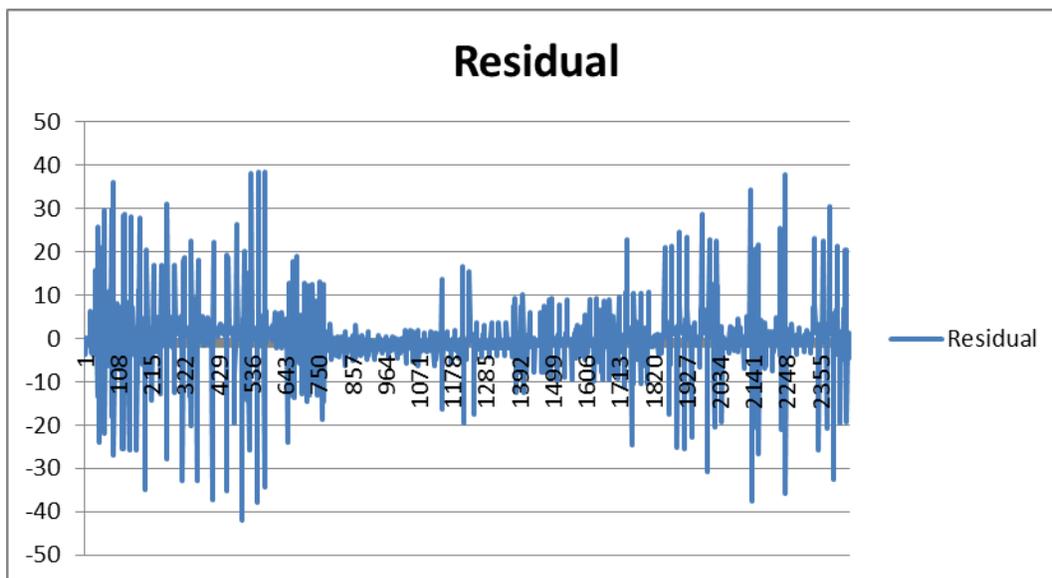


5. ¿ Existe Ruido Blanco o No Auto Correlación?

Los resultados de este modelo en cuanto al Error de Pronóstico se exportan de uno de los reportes de Módulo de Predictor de Oracle Crystal Ball, el cual es es el siguiente:

Results Table for Demo Reg.xlsx!Datos					
Created 11/17/2015 at 1:00 PM					
Series	CMO (spot) ▾				
Data					
Date	Historical Data	Lower: 15%	Fit & Forecast	Upper: 85%	Residuals
1	86.58				
2	86.58		86.91		-0.3295
3	86.15		86.67		-0.5150
4	83.12		85.85		-2.7264
5	83.12		82.84		0.2800
6	78.89		82.65		-3.7612
7	78.89		78.82		0.0654
8	78.89		78.79		0.0957
9	78.89		78.61		0.2777
10	78.89		78.69		0.2048
11	78.89		78.92		-0.0316
12	78.89		79.16		-0.2690

Luego, debemos hacer un gráfico del Error de Pronósticos y hacerle las pruebas si el error de pronósticos tiene Ruido Blanco.



Podemos, observar en el gráfico de error de pronóstico que la dispersión no es constante en el tiempo.

Sin embargo, debemos hacerle la prueba si el error de pronósticos sigue el proceso de Ruido Blanco.

La prueba más fácil y práctica en MS Excel es el **Factor Absoluto Z**, que debería ser menor o igual a 2 para tener Ruido Blanco y poder así continuar con los supuestos de regresión que se deben de cumplir.

Los detalles de las pruebas son los siguientes:

PRUEBA A TRAVES DEL FACTOR ABS Z

# Observaciones	2,449	CONTAR (A3:C2,451)
Rho t-1	0.07	COEF.DE.CORREL(C3:C2450,C4:C2451)
FACTOR Z ABSOLUTO	3.22	ABS(# de Observaciones* Rho t-1)

Podemos observar en el cuadro anterior que el Error de Pronóstico no posee Ruido Blanco.

Adicionalmente, presentamos las pruebas de Ruido Blanco del software NumXL, que concluye que el error de pronóstico no tiene Ruido Blanco:

También podemos hacer dos pruebas estadísticas empleando el software muy útil en las series de tiempo que es el NumXL. Estas pruebas son la de Ruido Blanco y Efecto ARCH.

Los resultados de ambas pruebas se presentan a continuación:

White-noise Test

Lag	Score	C.V.	P-Value	Pass?	5.0%
1	10.37	3.84	0.1%	FALSO	
2	11.89	5.99	0.3%	FALSO	
3	16.73	7.81	0.1%	FALSO	
4	18.94	9.49	0.1%	FALSO	
5	21.86	11.07	0.1%	FALSO	
6	22.23	12.59	0.1%	FALSO	
7	26.72	14.07	0.0%	FALSO	
8	26.88	15.51	0.1%	FALSO	

Podemos, observar en el cuadro anterior que se tiene Ruido blanco.

Luego, lo deseable es que la prueba Efecto ARCH sea falsa. Es decir, se buscaría que no existiera mucha volatilidad extrema en en el error de pronóstico o residual.

Los resultados de esta prueba son los siguientes:

ARCH Effect Test

Lag	Score	C.V.	P-Value	Present?	5.0%
1	11.51	3.84	0.1%	VERDADERO	
2	17.36	5.99	0.0%	VERDADERO	
3	18.19	7.81	0.0%	VERDADERO	
4	18.31	9.49	0.1%	VERDADERO	
5	58.11	11.07	0.0%	VERDADERO	
6	58.15	12.59	0.0%	VERDADERO	
7	64.17	14.07	0.0%	VERDADERO	
8	64.27	15.51	0.0%	VERDADERO	

A continuación, se presenta la estadística descriptiva del error de pronóstico y se puede comprender esta prueba. Si existe curtosis alta, esta prueba siempre será verdadera y no falsa. La estadística descriptiva del error es la siguiente:

Descriptive Statistics

AVERAGE:	7.10531E-13
STD DEV:	5.864291578
SKEW:	-0.05
EXCESS-KURTOSIS:	16.39
MEDIAN:	-0.205698007
MIN:	-41.98926913
MAX:	38.60058122
Q 1:	-1.498082535
Q 3:	1.294493188

Se puede apreciar el exceso de curtosis de 16.39, el cual si fuese una Distribución Normal sería de 0. Es por eso que el software NumXI concluye lo siguiente:

	Test	p-value	SIG?
	White-noise	0.07%	FALSO
Normal Distributed?		0.00%	FALSO
	ARCH Effect?	0.00%	VERDADERO

La conclusión es que no existe Ruido Blanco, no Existe Normalidad y el Efecto ARCH existe por la alta volatilidad que existe en el error.

VII. Modelo de Pronóstico Modificado Con Error Familia ARIMA

Una de las grandes facilidades de emplear el software Oracle Crystal Ball no solo es que podemos hacer un número ilimitado de modelos de pronósticos sino también que cuando nos encontramos con error de pronóstico complicado como el encontrado en este trabajo, se hace muy fácil modelar el error de pronósticos con el mejor modelo de la Familia ARIMA/SARIMA, que se ajuste mejor. En el presente trabajo, el mejor modelo es el ARIMA (0, 1,1).

En consecuencia, el nuevo modelo toma la siguiente forma:

$$P_{t+1} = F(DT, P_{t-1}, ET = \text{ARIMA}(0, 1, 1))$$

La ecuación de regresión es la siguiente:

$$P_{t+1} = 6.23 + 0.0070 Dt + 0.8848 P_{t-1} + \text{ARIMA ERROR}(0, 1, 1)$$

El modelo de regresión arriba formulado tiene error de Ruido Blanco como se puede apreciar con la Prueba Factor ABS Z, cuyo valor es menor o igual que 2.

PRUEBA A TRAVES DEL FACTOR ABS Z

# Observaciones	2,448	CONTAR (B2:C2449)
Rho t-1	0.034806	COEF.DE.CORREL(B2:C2448,B3:B2449)
FACTOR Z ABSOLUTO	1.72	ABS(# de Observaciones* Rho t-1)

Así mismo, las pruebas estadísticas empleando el software NumXL se detallan a continuación:

Lag	Score	C.V.	P-Value	Pass?	5.0%
1	3.02	3.84	8.2%	VERDADERO	
2	3.11	5.99	21.1%	VERDADERO	
3	3.59	7.81	31.0%	VERDADERO	
4	3.59	9.49	46.5%	VERDADERO	
5	14.78	11.07	1.1%	FALSO	
6	15.70	12.59	1.5%	FALSO	
7	29.90	14.07	0.0%	FALSO	
8	33.86	15.51	0.0%	FALSO	

La prueba ARCH EFFECT sigue demostrando que existe volatilidad extrema, como se explicó anteriormente con la Estadística Descriptiva del Error de Pronóstico o Residuales.

VIII. Pronósticos de Precios Spot Horarios de la Energía Eléctrica

En la sección anterior encontramos el mejor modelo de pronóstico de los precios spot horario de energía.

El modelo encontrado es en resumen el siguiente:

P_{t+1} = Pronóstico de Precio en el tiempo t+1

D_t = Pronóstico de Demanda t a riesgo

P_{t-1} = Pronóstico de Precio con retraso de 1 hora a riesgo

E_t = Error modelado a riesgos con ARIMA (0, 1, 1)

Cabe mencionar lo siguiente:

1. Recordar que partimos la serie de tiempo para hacer el modelo y para hacer el back-testing o prueba con datos nuevos no empleados para construir el modelo
2. El Módulo Predictor encontrará los supuestos de riesgo de las variables independientes o factores de la regresión: Demanda t y Precio retrasado 1 hora P_{t-1} . A ambas variables independientes se les aplica el riesgo.

- También el Módulo Predictor encontrará el supuesto riesgo del error ARIMA (0, 1,1). Es decir, también al error se le ha aplicado el riesgo.
- El MAPE del modelo inicial de regresión ascendió a 2.86% como se muestra en la pantalla de Módulo de Predictor en las Sección VI. El MAPE del back-testing será arriba de este.

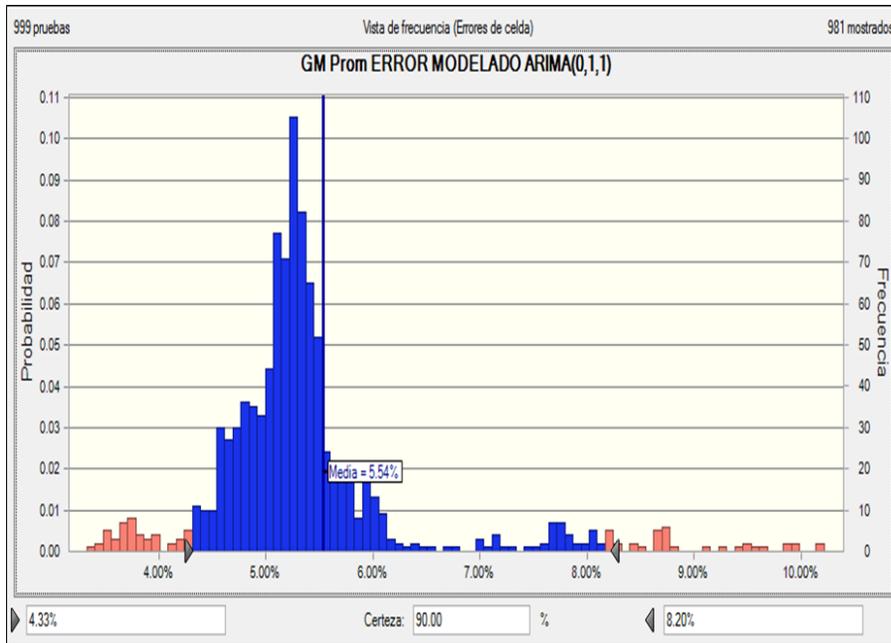
A continuación les presentamos la hoja de MS Excel de la prueba del modelo y de Back-testing:

							Error Normal			
							Error T de Student			
							Error ARIMA(0,1,1)			
HORA	# Datos	CMO (spot)	# Datos	Proyeccion Precio Spot Horario	Proyeccion de Demanda MWH	Lag1 CMO	Error ARIMA(0,1,1)	Referencia	Error Normal	Error t de Student
0	2451	\$ 92.03	2451	\$ 98.83	527.22	\$ 99.72	3.35	Back-testing 1	1.68	-1.02
1	2452	\$ 91.98	2452	\$ 101.73	487.59	\$ 91.46	3.86	Back-testing 2	-16.06	-2.98
2	2453	\$ 91.98	2453	\$ 87.91	467.49	\$ 93.33	-2.51	Back-testing 3	-3.81	-0.49
3	2454	\$ 91.98	2454	\$ 86.53	452.59	\$ 85.31	-5.44	Back-testing 4	-1.01	2.64
4	2455	\$ 92.03	2455	\$ 79.00	498.97	\$ 82.66	-6.20	Back-testing 5	4.55	1.09
5	2456	\$ 94.94	2456	\$ 87.21	550.78	\$ 85.71	3.99	Back-testing 6	-4.58	0.47
6	2457	\$ 101.17	2457	\$ 86.28	567.59	\$ 81.75	0.25	Back-testing 7	-2.95	-0.17
7	2458	\$ 104.43	2458	\$ 79.81	641.89	\$ 75.04	-3.24	Back-testing 8	-0.13	-0.68
8	2459	\$ 104.47	2459	\$ 84.39	735.07	\$ 67.73	6.62	Back-testing 9	0.42	0.96
9	2460	\$ 104.47	2460	\$ 78.97	773.01	\$ 66.75	7.41	Back-testing 10	-0.29	-0.74
10	2461	\$ 104.47	2461	\$ 82.06	798.81	\$ 85.63	11.18	Back-testing 11	-0.06	-0.69
11	2462	\$ 104.47	2462	\$ 99.47	810.46	\$ 90.35	11.80	Back-testing 12	-4.92	-1.85
12	2463	\$ 104.47	2463	\$ 95.92	793.78	\$ 76.49	4.21	Back-testing 13	0.28	0.96
13	2464	\$ 104.47	2464	\$ 82.38	807.24	\$ 81.46	2.83	Back-testing 14	10.74	-1.97

Se ha construido un modelo de pronósticos empleando tres tiempos de modelación de error de pronóstico como se puede apreciar en el cuadro anterior.

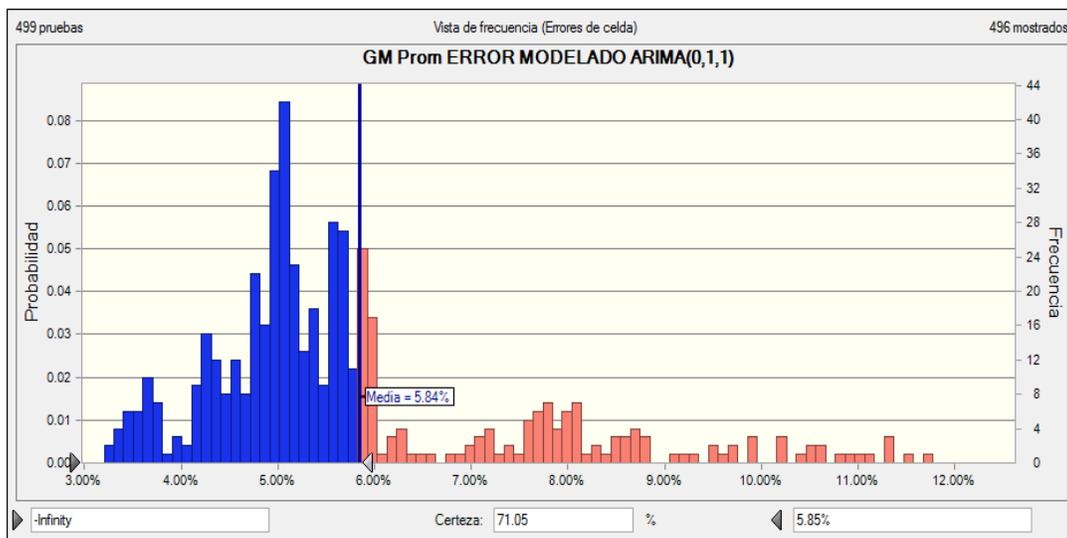
La celda celeste son los pronósticos que nos interesan y las celdas de color verde son los supuestos que ha encontrado y pegado el Modulo de Predictor del software Oracle Crystal Ball.

Los resultados obtenidos se presentan en los siguientes perfiles de error MAPE, siendo el modelo con menor MAPE, el que modela el error con ARIMA (0,1,1). A continuación, se presentan los resultados de la Simulación Monte Carlo:

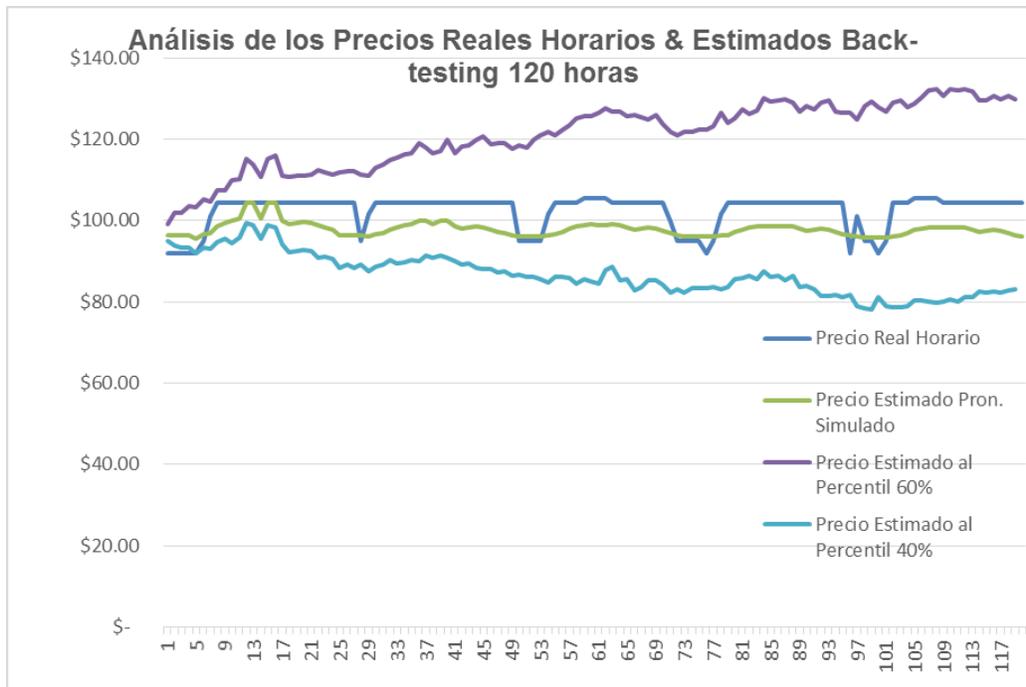


Los demás perfiles de riesgo de errores modelados con Distribución Logística y T de Estudiante poseen límites de certidumbre superiores.

El MAPE del mejor modelo de regresión tiene una probabilidad del 71.05% que es menor o igual al promedio de 5.84%. Usualmente, el error de pronóstico bajo cualquier métrica será mayor en el horizonte del back-testing que el de construcción del modelo.



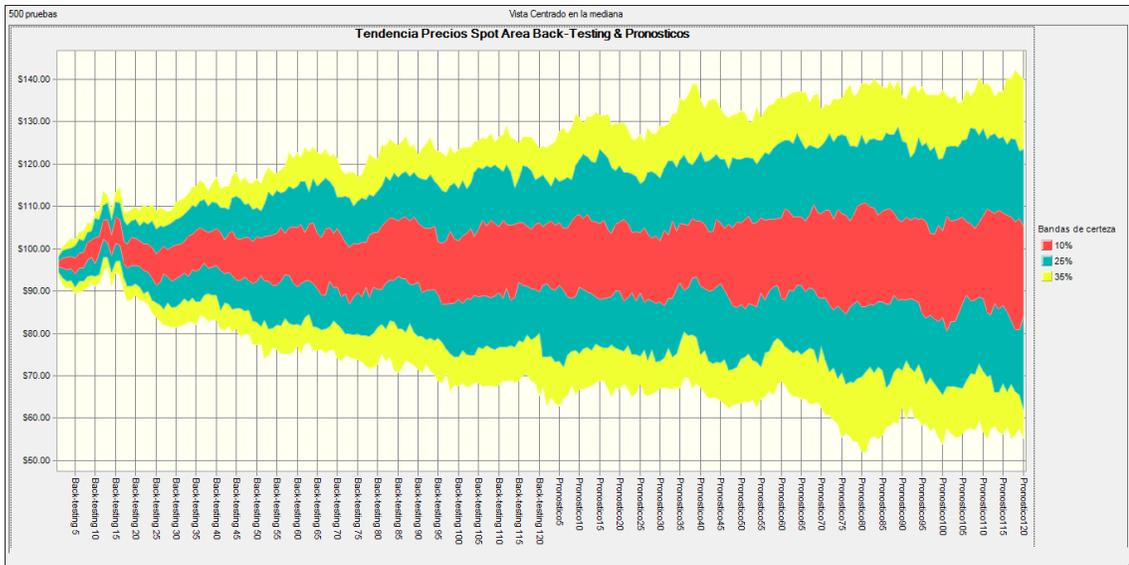
A continuación, se presenta el análisis comparativo de los precios spot horarios reales versus los estimados inferidos a través de la simulación del precio empleando el software Oracle Crystal Ball.



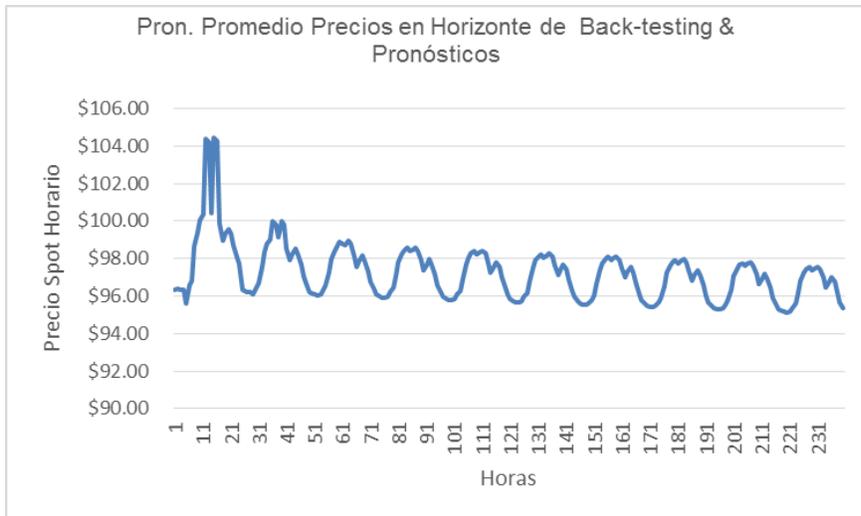
Podemos observar en la gráfica anterior que la estimación de precios es un poco abajo que el precio real, siendo este el motivo por el cual el error del pronóstico MAPE es de un promedio de 5.86% con una probabilidad de aproximadamente 71.05% que pueda ser más bajo.

Una pregunta que siempre nos hacemos los que realizamos pronósticos es ¿cuál debería ser el error máximo en un pronóstico? Esto siempre depende de qué pronóstico estamos hablando. Sin embargo, si no existen modelos tendríamos un 50% empleando una moneda con dos caras. En la práctica, los modelos que brindan confianza arriba del 70% son modelos aceptables. Hay sectores como el financiero que están empleando un nivel de confianza del 99.90%. Sin embargo, para usos prácticos arriba del 70% podemos decir que estaremos mejor que el emplear una moneda.

Finalmente, tomando en consideración el nivel de error de pronóstico, realizamos el siguiente pronóstico simulado para lo siguiente: 120 horas de back-testing simulado (Back-testing 1 al Back-testing 120 horas) y 120 horas de pronóstico simulado (Pronostico 1 hasta Pronostico 120), el cual lo detallamos en la gráfica de Tendencia del software Oracle Crystal Ball:



La proyección de Precios Spot para los horizontes de Back-testing y Pronósticos se detalla a continuación:



En la gráfica anterior se ha calculado el MAPE usando el estimado medio y el percentil de 40%, siendo el menor error de pronóstico empleando el estimado del precio spot horario simulado del promedio.

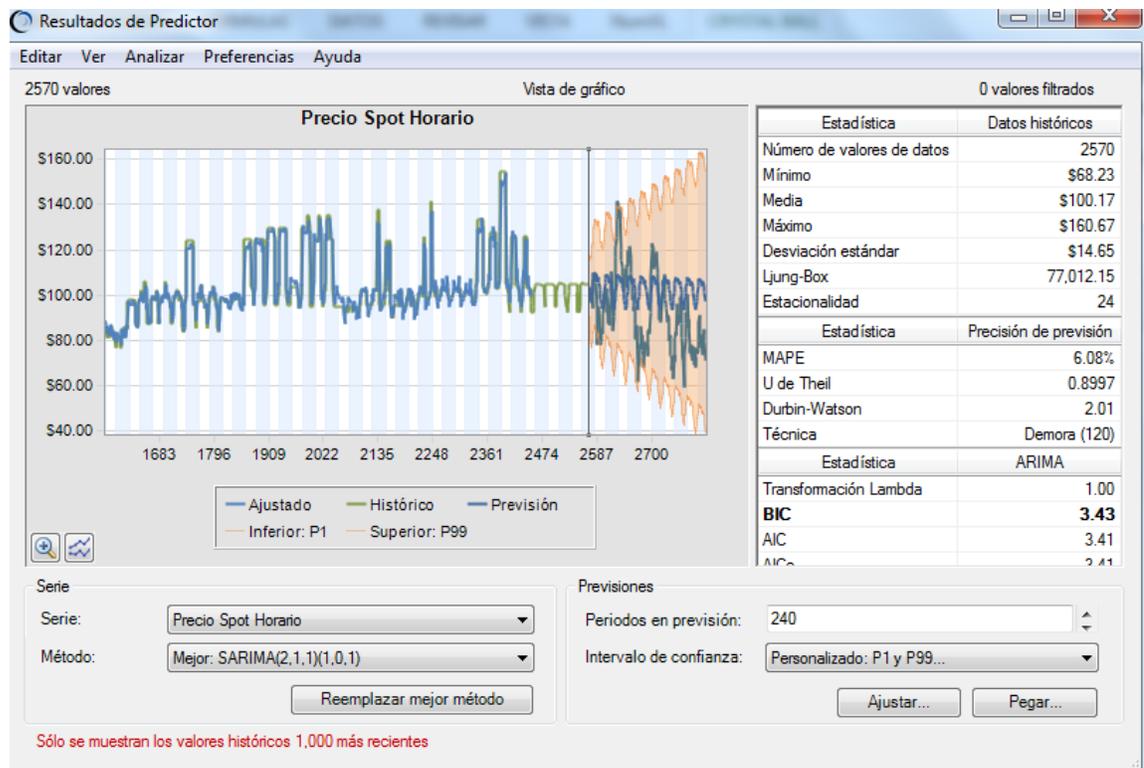
El éxito de generar el precio spot horario es de llevar el ciclo de actualización semanal por la manera que cambian los factores endógenos y exógenos del modelo.

VIII. Modelo # 3 Propuesto de Serie de Tiempo

A continuación presentamos el modelo de serie de tiempo que más se ajusta a la serie de tiempo. Estaremos empleando una retención o holdout para realizar el back-testing de 120 horas, la semana del 25-29 de mayo de los corrientes. Luego, el pronóstico será las próximas 240 horas o 2 semanas laborales. Recordemos que solamente necesitamos dos columnas, una columna de tiempo y la segunda del precio spot horario. La hoja de MS Excel empleada se presenta a continuación:

# Datos	FECHA	HORA	Estación Ciclo	# Datos	Precio Spot Horario
1	12/31/2014	22	71	1	\$ 86.58
2	12/31/2014	23	72	2	\$ 86.58
3	1/1/2015	0	73	3	\$ 86.15
4	1/1/2015	1	74	4	\$ 83.12
5	1/1/2015	2	75	5	\$ 83.12
6	1/1/2015	3	76	6	\$ 78.89
7	1/1/2015	4	77	7	\$ 78.89

Para lograr el mejor modelo de series de tiempo emplearemos el Modulo Predictor del software Oracle Crystal Ball, cuya pantalla con los resultados es la siguiente:



En el cuadro anterior podemos ver que el modelo idóneo de serie de tiempo avanzada es SARIMA (2,1,1) (1,0,1). La diferencia entre el modelo SARIMA versus el ARIMA es que el primero tiene el factor estacional, cuyos parámetros están en el segundo paréntesis.

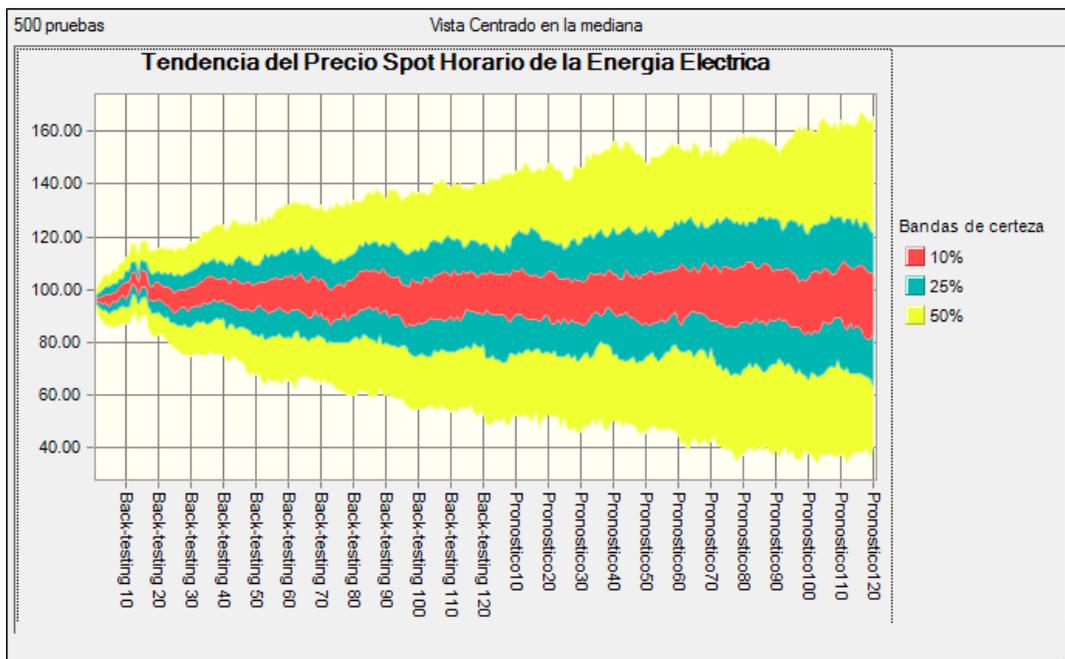
El modelo SARIMA encontrado brinda un error de pronóstico en el horizonte de back-testing del 6.08%. Se puede observar que el horizonte del back-testing no está cubierto del color azul del modelo encontrado por el software.

La calidad es el factor más importante en los pronósticos, y para las series de tiempo a diferencia de los modelos de regresión, solamente son dos supuestos los que deben cumplirse:

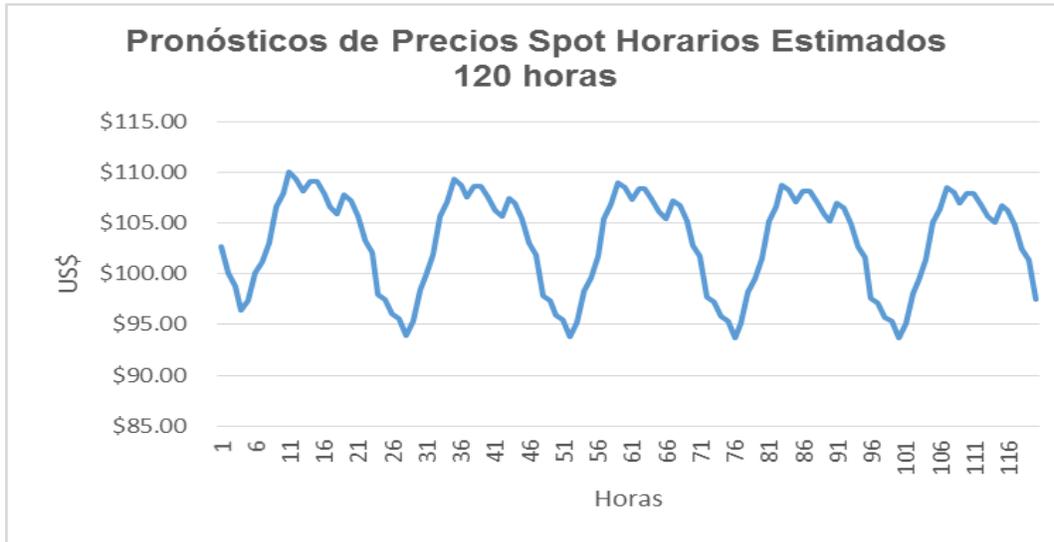
1. El Durbin-Watson (DW) debería ser de 2, o estar en los rangos de 1.5-2.5.
2. El Thiel no debería ser igual o mayor de 1 dado que esto indicaría que los pronósticos naive son superiores a las series de tiempos avanzadas que se tienen disponible en el Modulo Predictor.

Ambos supuestos encontrados en el modelo SARIMA pasan la prueba de los supuestos o requerimientos de los modelos de series de tiempo. La métrica de Durbin-Watson asciende a 2.01 y el valor de Thiel es de 0.8997. La retención o holdout ha sido de los precios spot horarios de las últimas 120 horas de información o precios reales o actuales.

Es de hacer notar que hemos pronosticado 240 horas. Sin embargo, estaremos empleando solamente la estimación de precios spot horaria de 120 horas.



La estimación de los próximos 120 precios spot horarios la presentamos en el siguiente gráfico:



IX. ¿Es posible crear un Modelo de Pronóstico del Precio Spot Horario Combinado de un Modelo de Serie de Tiempo y Regresión Múltiple?

Usualmente cuando tratamos el tema de pronósticos, se piensa en un solo modelo a emplear. Sin embargo, en la práctica se pueden emplear varios modelos combinados para poder ser más certero en el pronóstico, lo cual se pensaría que el error de pronóstico disminuya.

Para explorar la combinación de modelos se sugiere emplear las siguientes variables que se presentan en la hoja de MS Excel a continuación:

Fecha	Precio Real	Pronósticos Precios SARIMA	Pronósticos Modelo Regresión
2	\$ 86.58	\$ 82.10	\$ 86.91
3	\$ 86.15	\$ 84.61	\$ 86.67
4	\$ 83.12	\$ 83.76	\$ 85.85
5	\$ 83.12	\$ 81.98	\$ 82.84
6	\$ 78.89	\$ 82.54	\$ 82.65
7	\$ 78.89	\$ 81.13	\$ 78.82
8	\$ 78.89	\$ 83.51	\$ 78.79

La información ha sido extraída de los informes que brindan los datos de los pronósticos y residuales que se exportan del Módulo de Predictor del software Oracle Crystal Ball.

Luego, para probar que podemos combinar modelos se corre la siguiente regresión:

Pron. Precio = F (Resultados Modelo Series de Tiempo, Resultado Modelo de Regresión, Error).

Los resultados de correr esta regresión en MS Excel se presentan a continuación:

Resumen PRUEBA DEL MODELO COMBINADO					
Estadísticas de la regresión					
Coefficiente de correlación múltiple					0.93
Coefficiente de determinación R ²					0.87
R ² ajustado					0.87
Error típico					5.48
Observaciones					2449
ANÁLISIS DE VARIANZA					
	Grados de libertad	Suma de cuadrados	Varianza de los cua.	F	Valor crítico de F
Regresión	2	475344.4847	237672.242	7921.063063	0
Residuos	2446	73392.46009	30.0050941		
Total	2448	548736.9447			
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%
Intercepción	1.388733056	0.815053475	1.70385514	8.8535%	-0.209533271
Pronósticos Precios SARIMA	0.81901723	0.043181552	18.9668316	0.0000%	0.734341042
Pronósticos Modelo Regresión	0.166940484	0.044651136	3.73877352	0.0189%	0.07938254

Se puede apreciar en los resultados del análisis de regresión que el error asciende a US\$ 5.48 y que el intercepto de la regresión no es estadísticamente significativo dado que su probabilidad pasa en nivel de corte del 1%. Por consiguiente, sí podemos combinar los dos métodos y lograr una mejora en el error típico.

La comparación de errores típicos de los modelos de pronósticos individuales y colectivos se detalla a continuación:

Análisis de Error de Pronóstico	
Error de Pronóstico SARIMA	\$ 5.50
Error de Pronóstico Regresión	\$ 5.87
Error Modelo Combinado	\$ 5.48

Se puede apreciar que el modelo combinado tiene un menor error de pronóstico, en este caso su nivel no es tan diferente pero en futuros casos si lo puede ser. Es por eso que es valioso demostrar esta técnica de combinación.

El siguiente paso es encontrar el porcentaje que será distribuido a cada técnica de pronóstico. Es importante que la sumatoria de los porcentajes sean del 100%. Para lograr este objetivo, realizamos la segunda regresión, solo que se hace el cálculo sin considerar el intercepto que resultó ser no significativo estadísticamente.

Los resultados de esta regresión hecha en MS Excel se presentan a continuación:

Resumen DETERMINACION DE PORCENTAJES POR METODO

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.999
Coefficiente de determinación R^2	0.997
R^2 ajustado	0.997
Error típico	5.480
Observaciones	2449

ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Varianza de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	2	24995736.85	12497868.4	416201.1923	0
Residuos	2447	73479.56855	30.0284301		
Total	2449	25069216.42			

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>
Pronósticos Precios SARIMA	81%	0.043023682	18.8828035	2.37007E-74	0.72804113
Pronósticos Modelo Regresión	19%	0.043059789	4.34684061	1.43758E-05	0.102736638
	100%				

La distribución de los porcentajes sería como se muestran en el cuadro anterior de 81% para el estimado de SARIMA y el 19% para la Regresión.

X. Bibliografía

Hanke, John E.; Wichern, Dean W.: Pronósticos en los Negocios. Pearson.

Charnes, John: Financial Modeling with Oracle Crystal Ball and Excel. Wiley Finance.

Pardoe: Applied Regression Modeling A Business Approach. Wiley

Guajarati, Damodar N: Econometría. Mc-Graw Hill

Wiston, Wayne L .Marketing Analytics: Data Driven Techniques with Microsoft Excel. Wiley

Uriel, Esequiel y Aldas Joaquín. Analisis Multivariante Aplicado. Thomson.

De la Garza García, Jorge; Morales Serrano, Blanca Nieves y González Cavazos, Beatriz Adriana: Analisis Estadístico Multivariante. Mc Graw Hill

XI. Hoja de Vida

Luis Francisco Zaldívar, MSE

www.modelacionderiesgos.com

l.zaldivar@modelacionderiesgos.com

Graduado de Licenciatura en Administración de Empresas con concentración en Gerencia Industrial de The University of Tennessee, Knoxville, Tennessee.

Posee Maestría en Ciencias Económicas con concentración en Finanzas y Estadística Aplicada de North Carolina State University, Raleigh, North Carolina.

Entrenado por Oracle Crystal Ball en Denver en Crystal Ball Introductorio y Avanzado, Opciones Reales y Seis Sigma.

Presidente de Inversiones Carrousel, S.A. de C.V. fundada en 1992.

Facilitador y Consultor Certificado en Análisis de Riesgos empleando Crystal Ball por Crystal Ball Global Unit, Denver, Colorado.

Además, es Facilitador y Consultor Certificado usando el software de riesgo más avanzado ModelRisk® de Vose Software BVBA de Bélgica.

Profesor de Simulación Monte Carlo, Optimización, Valorización de Empresas, Minería de Datos, Riesgos, Estrategias Financieras y M&A en los programas de maestría en Finanzas y Administración de Empresas de la Universidad José Simeón Cañas of El Salvador (UCA), Universidad de El Salvador, y Universidad Rafael Landívar en Guatemala (URL), Universidad Autónoma Gabriel Rene Moreno (UAGRM BUSINESS SCHOOL) en Santa Cruz, Bolivia (UARGM).

Dirigió Bancos Comerciales por 8 años y empresas de manufactura de exportación por 16 años en El Salvador.

Participó en Juntas Directivas en empresas manufactureras de exportación.

Conferencista en Simulación, Pronósticos y Optimización Estocástica para Oracle USA en América Latina.

Consultor especializado en Finanzas, Inversiones, Análisis de Riesgos y en Minería de Datos en América Latina.

Ha realizado entrenamientos en Análisis de Riesgo con Crystal Ball aplicado a las siguientes industrias durante los últimos nueve años: Petróleo, Gas, Minería, Banca Comercial y Central, Educación y en Empresas Corporativas durante los últimos siete años en los siguientes países: México, Guatemala, El Salvador, Nicaragua, Costa Rica, Panamá, República Dominicana, Colombia, Perú, Chile, Bolivia, Paraguay y Venezuela.