

Modelos estocásticos para predecir la demanda de gas licuado de petróleo en Uruguay

Silvia Rodríguez
Natalia da Silva

Junio 2010

Serie DT (10 /01)
ISSN : 1688-6453

Modelos estocásticos para predecir la demanda de gas licuado de petróleo en Uruguay

Silvia Rodríguez,^{1*} Natalia da Silva,²

¹*Instituto de Estadística, FCEA, Uruguay.*

²*Instituto de Estadística, FCEA, Uruguay.*

Resumen

En Uruguay el gas licuado de petróleo (GLP) se destina principalmente al uso domiciliario en calefacción y cocción. Este trabajo surge como una extensión a un trabajo originado a partir de un convenio con ANCAP en que se elaboraron un conjunto de modelos de demanda de combustibles líquidos y GLP.

Este documento tiene dos objetivos, primero obtener pronósticos de demanda de GLP diaria y mensual mediante modelos de series temporales y segundo, analizar la variación del desempeño predictivo de estos modelos de demanda, al incorporar la variable temperatura como variable exógena.

Se plantean diferentes aproximaciones metodológicas para obtener las proyecciones de demanda de GLP. Se estiman modelos univariados (SARIMA-IA) y multivariados (VEC). A partir de los primeros, se obtienen pronósticos de demanda con frecuencia diaria y mensual, mientras que con los modelos multivariados se obtienen pronósticos de demanda mensuales. Se evalúa la performance predictiva de los distintos modelos en diferentes horizontes de proyección y su sensibilidad ante la incorporación de información adicional. Se estudia el papel que juega la variable temperatura, considerando diversas alternativas, en el desempeño predictivo del modelo.

El modelo de predicción diario, por su flexibilidad y buen desempeño predictivo es una herramienta adecuada para planificar la demanda de corto plazo. Al incorporar la temperatura, el error de predicción del modelo disminuye, pero en menor medida que en el modelo mensual. La inclusión de la variable temperatura en el modelo de demanda mensual mejora el desempeño predictivo en el período más sensible de la planificación, los primeros meses de frío. En el modelo multivariado la mejora en el desempeño predictivo se asocia con toda la información adicional incorporada y no únicamente por la variable temperatura, aunque sin ella el modelo empeora su desempeño predictiva.

Los modelos presentados son complementarios en lo que a horizontes de predicción se refiere, son útiles para la tarea de planificar la demanda de GLP brindando recursos para proyectar la misma apoyada en modelos estocásticos, con diferentes requerimientos de información.

Palabras clave: Demanda de GLP, ARIMA, Modelos de corrección del error .

*E-mail: silvia@iesta.edu.uy

1. Introducción

Este trabajo surge como una extensión a un trabajo originado a partir de un convenio entre el Instituto de Estadística y ANCAP donde se elaboraron un conjunto de modelos de demanda de combustibles líquidos y GLP.

El objetivo general de este trabajo es predecir la demanda diaria y mensual de GLP, el objetivo específico es analizar el efecto en la performance predictiva al incluir la temperatura como variable exógena en los modelos univariados y multivariados. Se ensayan diferentes formas para incluir la variable temperatura y se releva el resultado tanto en términos de calidad del modelo como de la reducción del error de predicción.

De este trabajo se deduce que la incorporación de la temperatura como variable exógena en los modelos analizados, mejora los indicadores de calidad de ajuste y de performance predictiva. Adicionalmente, las distintas especificaciones para la variable exógena temperatura dieron lugar a resultados diferentes, en particular en el modelo mensual, lo que justifica la búsqueda de otras alternativas para incluir dicha variable en el modelo, no se han agotado las posibilidades de especificación de la variable, por lo que queda abierto el camino a mejorar los resultados predictivos del modelo mediante otras especificaciones alternativas.

El documento se organiza de la siguiente manera, en la segunda sección se comenta el antecedente inmediato a este trabajo, en la tercera sección se presenta la aproximación metodológica realizada en los modelos univariados (diario y mensual) y multivariados. En la sección cuarta se presentan los resultados para los modelos univariados, mensuales y diarios, así como la comparación entre ellos, en la sección quinta se presenta el modelo multivariado y la comparación con el modelo mensual y finalmente en la sección seis se presentan algunos comentarios finales.

2. Antecedentes

El GLP, en términos generales se utiliza durante todo el año para cocción y para calefacción fundamentalmente en los meses de frío. La estacionalidad en el consumo de GLP determina a su vez una fuerte estacionalidad en las ventas que realiza ANCAP del producto a las empresas distribuidoras. La variable que se modeliza en este trabajo son las ventas de ANCAP, mensuales y diarias de GLP.

Los modelos estimados en el marco del convenio, incorporaban la representación de la estacionalidad, en forma tanto estocástica como determinística, los mejores resultados en términos predictivos, se obtuvieron al incorporar la dinámica estacional de forma estocástica en el modelo mensual y en el modelo diario ambas alternativas presentaron resultados similares.

Estos modelos, cuyo objetivo fundamental es lograr buenas predicciones de la demanda futura que enfrenta ANCAP, a distintos horizontes temporales (diario, semanal, quincenal, mensual, trimestral, semestral y anual) permiten obtener una muy buena aproximación a la demanda agregada, anual en el caso del modelo de frecuencia mensual y hasta un mes con el modelo diario.

En lo que refiere al modelo univariado mensual, y dadas las necesidades de planificación anual del Sector Comercialización, se consideró el desempeño predictivo hasta 12 meses. Al agregar la demanda anualmente, los errores relativos de predicción fueron bajos, pero se registró un comportamiento dispar de dichos errores en algunos meses de invierno y no en forma regular, los errores llegaron a más de un dígito,

esto motivó que el equipo continuara explorando diferentes opciones de especificación de los modelos.

Los modelos a los que hacemos referencia son originariamente univariados, que por su sencillez y parquedad en términos de información los hace muy útiles a la hora de mantenerlos activos y utilizarlos para planificar la demanda futura. Además se trabajó con modelos multivariados para obtener predicciones mensuales, estos modelos mostraron mejor desempeño predictivo que los modelos univariados, los mismos utilizan más información, son más complejos de analizar y elaborar que los anteriores.

La serie mensual de demanda se obtiene como la suma de la demanda diaria y las ventas de GLP se realizan fundamentalmente de lunes a sábados y sólo en casos excepcionales se vende el producto los domingos o los feriados, cómo en el caso del invierno de 2007, estos modelos incorporan el efecto calendario. En términos del modelo mensual, tanto univariado como multivariado, esto implica considerar que no todos los días del mes son iguales en términos de demanda y por tanto cuánto mayor cantidad de domingos hay en el mes, menor es la demanda. En el modelos mensual multivariado se trabajó con un modelo VEC con dos ecuaciones (logaritmo de la demanda de GLP y logaritmo del IMS) además se consideró la temperatura media máxima y el precio real del GLP como variables exógenas, se identificaron atípicos y se modeló la estacionalidad de forma determinística.

En el modelo diario además de considerar estacionalidades múltiples, dentro de la semana y dentro del año, se incorpora el efecto de los días donde no se vende el producto tanto por feriados u otras causas.

Los modelos univariados estimados en el marco del convenio, no consideraban la inclusión de la variable temperatura, mientras que, como se mencionó anteriormente, en el modelo multivariado se consideró la temperatura media máxima como una variable exógena, y los resultados en la performance predictiva eran mejores cuando se consideraba esta variable que cuando no, sólo se ensayó la opción de incluir la variable temperatura como la temperatura media máxima.

En función de estos modelos antecedentes, se elabora la hipótesis de que la incorporación de la variable temperatura en los modelos univariados mejoraría la performance predictiva de los mismos, sobre todo en los meses de picos de demanda. Pero además se considera que las distintas maneras de incorporar la variable temperatura, podrían modificar los resultados, en particular en el modelo multivariado que ya considera este efecto. Por lo que a continuación se detalla los nuevos ensayos, en cuanto a las especificaciones realizadas.

3. Aproximación metodológica

En esta etapa además de las intervenciones y la modelización de los puntos raros, se incorpora la variable temperatura como variable exógena en los modelos.

Se incorpora la media de la temperatura máxima mensual relevada por la estación experimental de Las Brujas. Esta selección se fundamenta en el hecho de que los picos de demanda de GLP tienen como fin la calefacción, si las temperaturas máximas durante el día son muy bajas, se enciende la calefacción por más tiempo lo que implica una demanda mayor. No se utilizan las temperaturas mínimas en el entendido que se registran fundamentalmente en la madrugada y la hipótesis de trabajo es que quienes se calefaccionan con GLP lo hacen durante el día.

Hay diferentes formas de introducir la variable temperatura en el modelo. Las opciones ensayadas se especifican a continuación para cada caso.

A los efectos de evaluar la performance predictiva de los modelos se trabaja con los valores observados de la temperatura según las distintas variantes de la misma.

3.1. Modelo de demanda univariado mensual

Se estiman modelos de tipo SARIMA-IA, se consideran distintas especificaciones y se considera la modelización de la estacionalidad anual tanto de forma determinística como estocástica. Finalmente se selecciona modelar la estacionalidad de forma estocástica. Por otro lado se modelizan los puntos raros y se considera el efecto de la temperatura considerando las siguientes opciones:

1. Incorporar la serie completa de temperatura media máxima.
2. Incorporar una variable que toma el valor de la temperatura media máxima durante los meses de mayo a agosto y cero en el resto de los meses.
3. Incorporar la temperatura con su efecto en las diferentes estaciones del año, mediante tres variables, una para otoño, otra para invierno y otra para primavera. La variable otoño toma el valor de la temperatura media máxima durante los meses de abril y mayo y cero en el resto. La variable invierno toma el valor de la temperatura media máxima durante los meses de junio, julio y agosto y cero en el resto. La variable primavera toma el valor de la temperatura media máxima durante los meses de setiembre y octubre y cero en el resto.

3.2. Modelo de demanda univariado diario

El modelo presenta características diferentes al anterior dada la mayor frecuencia de los datos. En los modelos para representar la demanda diaria, deben contemplarse las diferentes estacionalidades superpuestas que se dan en el año. La modelización ARIMA permite incorporar dos, una de naturaleza estocástica y otra de naturaleza determinística.

La demanda de GLP tiene seis observaciones por semana, una para la demanda de cada día de lunes a viernes y otra para la demanda agregada de sábados y domingos. Al definir de esta forma la serie, la estacionalidad de naturaleza estocástica es la que se da dentro de la semana y la estacionalidad dentro del año se especifica como determinística, esto es se incorporan 11 variables indicatrices del mes, de enero a noviembre. Es posible, también modelizar todas las estacionalidades de forma determinística.

En los modelos diarios se incorpora también como variable exógena la variable temperatura diaria máxima de la estación experimental de Las Brujas. Las opciones ensayadas para incorporar esta variable de forma diaria fueron las siguientes:

1. Incorporar la serie completa de temperatura máxima.
2. Incorporar una variable que toma el valor de la temperatura máxima durante los meses de mayo a agosto y cero en el resto de los meses.

3.3. Modelo de demanda multivariado mensual

Se ajustó un modelo de corrección del error para estimar la demanda mensual de GLP. Las variables involucradas para estimar el modelo fueron: el logaritmo de la demanda de GLP, el logaritmo del consumo

de energía eléctrica residencial, el logaritmo del IMSR, el logaritmo del precio relativo entre el GLP y la energía eléctrica, efecto calendario e intervenciones, finalmente se consideraron dos opciones para incorporar el efecto de la temperatura. Como se verá posteriormente estas variables entran en el modelo de distinta manera, quedando finalmente un modelos con dos ecuaciones y el resto de la información es usada para modelar el corto o el largo plazo.

Las opciones ensayadas para la variable temperatura son:

1. Incorporar la serie completa de temperatura media máxima.
2. Incorporar una variable que toma el valor de la temperatura media máxima durante los meses de mayo a agosto y cero en el resto de los meses.

4. Modelos univariados

4.1. Modelo mensual

4.1.1. Principales características de la serie

La muestra comprende el período que va de enero de 1998 hasta abril de 2009 y en base a la performance predictiva se selecciona el modelo. Se selecciona abril del 2009 como punto de corte, ya que predecir los meses de frío es lo más importante para la planificación comercial del producto y en esos meses es que se evalúa el desempeño predictivo del modelo.

Con el objetivo de predecir la evolución mensual futura de la demanda de GLP se elaboran modelos SARIMA-IA al que se le incluye la variable temperatura como exógena, incorporada en forma aditiva.

En el gráfico 1 se presenta la evolución que ha tenido la demanda de GLP para el período analizado. En la dinámica de la serie se destaca su fuerte comportamiento estacional, períodos que se repiten año a año, marcados por una fuerte demanda en los meses de invierno seguidos por períodos de baja. No se puede decir que en la ventana temporal analizada haya existido una demanda creciente del producto, sino que la misma se ha mantenido aproximadamente entre un mínimo de 8.900.000 litros y un máximo de casi 27.900.000 litros, con dos máximos claros en los años 2000 y 2007.

A los efectos de observar el componente estacional se presenta en el gráfico 2 la evolución de la demanda mensual para los últimos siete años completos de la muestra, 2003-2009. Al analizar la estacionalidad de la serie, se observan dos características fundamentales, por un lado el comportamiento diferencial de los distintos meses dentro de un mismo año, por otro la similitud en los mismos meses de los distintos años. Esta información se puede leer también en el autocorrelograma y autocorrelograma parcial que aparece en los gráficos 3 y 4.

Evolución de la demanda mensual de GLP de 1998.01 a 2010.05

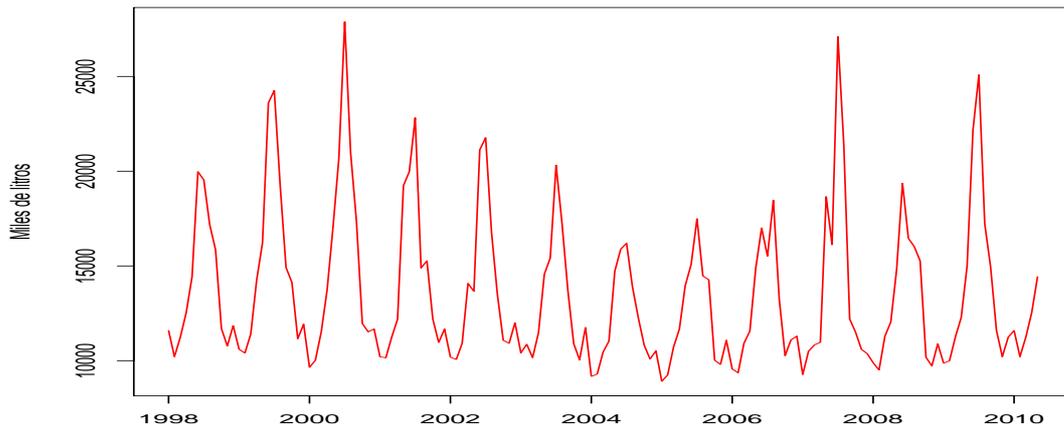


Figura 1: Gráfico de la demanda mensual de GLP

Comportamiento Estacional

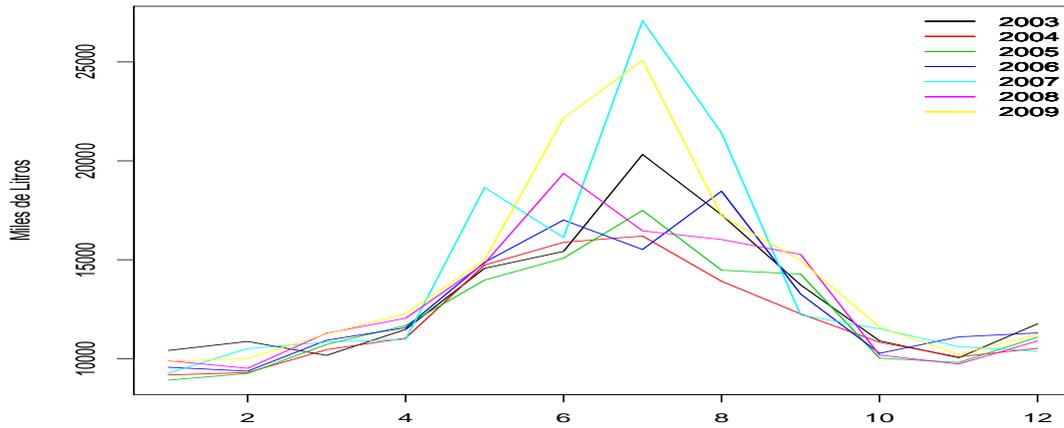


Figura 2: Comportamiento estacional mensual de la demanda mensual de GLP

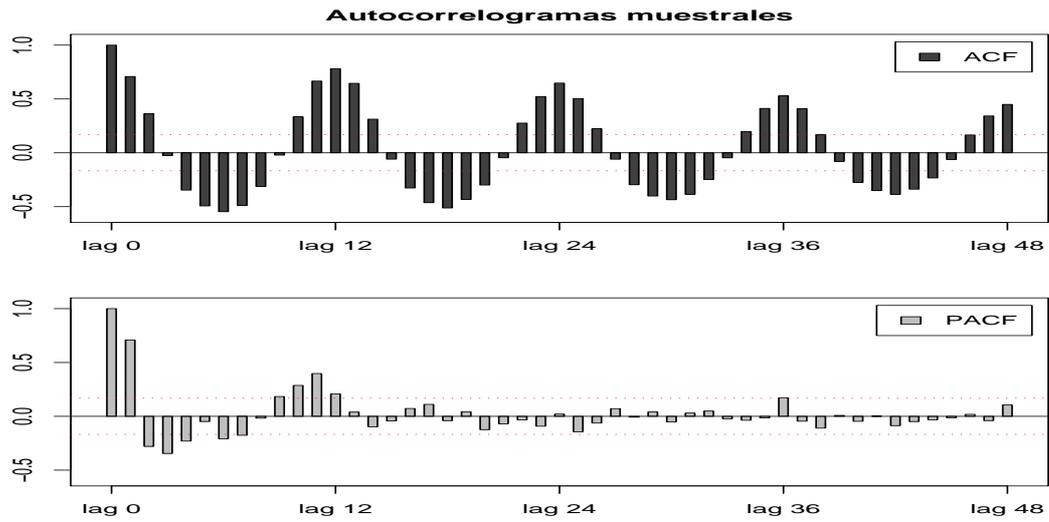


Figura 3: Correlogramas de la demanda mensual de GLP

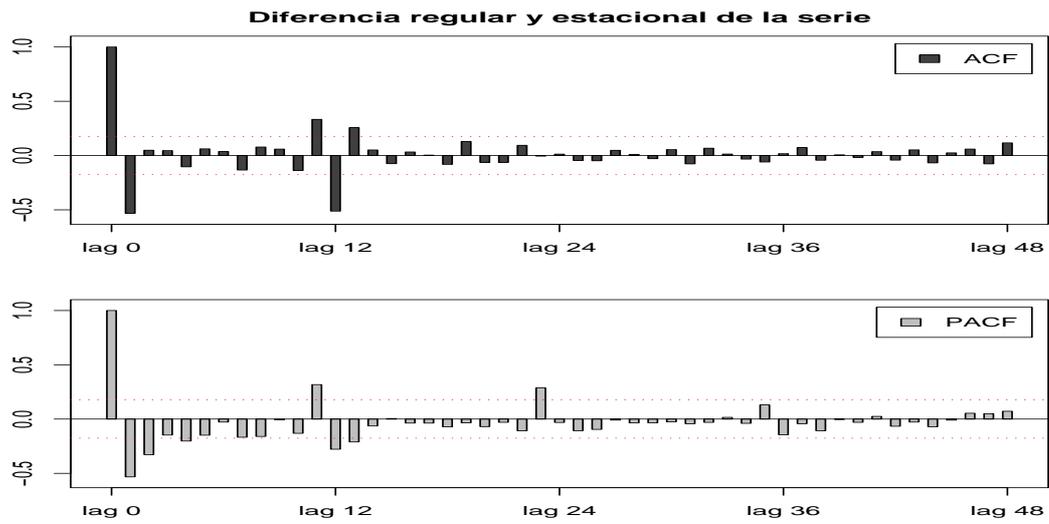


Figura 4: Correlogramas de la primera diferencia regular y estacional de la demanda mensual de GLP

4.1.2. Modelos estimados

En la dinámica de la serie se destaca su fuerte comportamiento estacional, períodos que se repiten año a año marcados por una fuerte demanda durante los meses de otoño e invierno, seguidos de períodos de baja. Se estimaron distintos modelos y las modificaciones más importantes en cuanto a la performance predictiva, se encuentran cuando la variable temperatura es incorporada como variable exógena en los mismos.

Se comienza considerando modelos univariados sin temperatura, los modelos son $SARIMA(0, 1, 1)(0, 1, 1)_{12}$ (contiene outliers) y el modelo $SARIMA(0, 0, 2)(0, 1, 1)_{12}$ (contiene outliers y un regresor para considerar el efecto calendario)¹. En base al menor error cuadrático medio se selecciona el primer modelo, que se puede especificar como sigue.

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-12} - \theta_3\epsilon_{t-13} + \phi_1AO_{200707} + \phi_2TC_{200007} + \epsilon_t$$

donde $Y_t = \log(GLP)$, AO_{200707} , TC_{200007} , son atípicos de tipo aditivo para julio del año 2007 y de tipo transitorio en julio del año 2000 respectivamente.

La performance predictiva del modelo ha sido evaluada en base a predicciones fuera de la muestra con datos hasta abril del 2009 prestando especial atención en los meses de frío.

Cuadro 1: Evaluación de la performance predictiva de la demanda de GLP mensual, del modelo sin temperatura, en miles de litros

Meses	Dato	Predicción	Error relativo	Inf 95 %	Sup 95 %
200905	14993	15680	-4,58 %	13039	18856
200906	22168	17647	20,40 %	14644	21264
200907	25082	17724	29,34 %	14678	21401
200908	17233	17374	-0,82 %	14360	21021
200909	14953	13949	6,72 %	11506	16910
200910	13148	10816	17,74 %	8904	13137
200911	10669	10383	2,68 %	8531	12637
200912	12195	11068	9,24 %	9077	13497
Acumulado	130442	114641	12,11 %	94739	138726

Los errores que se presentan en el cuadro 1 son errores relativos, los mismos muestran que en los primeros dos meses junio y julio se dan los mayores porcentajes de error. Incluso los valores observados en los meses de junio, julio y octubre, caen fuera de los intervalos de confianza, lo que significa que los niveles de demanda de esos meses no estaba dentro de los rangos esperados. Estos resultados y la hipótesis de que la temperatura juega un papel fundamental en el incremento de la demanda de GLP en los meses de invierno, justifican la incorporación en los siguientes modelos de la temperatura como regresora.

Se incorpora un regresor que toma los valores de la temperatura máxima media en cada mes de la muestra. Cuando se estima igual modelo, $SARIMA - IA(0, 1, 1)(0, 1, 1)_{12}$, incluyendo un regresor para recoger el efecto calendario. Este modelo presenta un error relativo acumulado para los 8 meses menor que el modelo

¹Estos mismos modelos fueron estimados pero modelando la estacionalidad de forma determinística, y finalmente se seleccionaron modelos donde la estacionalidad es estocástica ya que los mismos mostraron mejor performance predictiva.

sin temperatura, pero presenta residuos correlacionados, por lo que esta especificación se descarta².

Como se quiere analizar la sensibilidad de los modelos a la incorporación de alguna variable que recoja el efecto de la temperatura, se consideraron dos opciones adicionales; una variable que toma los valores de la temperatura media máxima para los meses de invierno (mayo a agosto) y cero en el resto de los meses³. En segundo lugar se incluyen tres variables regresoras, una que incorpora la temperatura de los meses de junio a agosto y cero en el resto de los meses, otra que toma los valores de la temperatura máxima media durante los meses de abril y mayo y cero en el resto y finalmente otra que toma los valores de la temperatura en setiembre y octubre y cero en el resto.

La siguiente ecuación corresponde a un modelo de tipo $SARIMA - IA(0, 1, 1)(0, 1, 1)_{12}$ con efecto calendario, modelización de datos atípicos y considerando la temperatura en los meses de frío:

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-12} - \theta_3\epsilon_{t-13} + \phi_1D_{LV} + \phi_2AO_{200706} + \phi_3IO_{200307} + \phi_4T_t + \epsilon_t$$

donde $Y_t = \log(GLP)$, D_{LV} recoge el efecto calendario, es la sumatoria de la diferencia entre la cantidad de días hábiles de la semana y el número de sábados y domingos. AO_{200706} , IO_{200307} , son atípicos de tipo aditivo e innovativo respectivamente y finalmente T_t es una variable que releva la temperatura media máxima en los meses de mayo a agosto y en el resto toma valores cero.

La performance predictiva del modelo ha sido evaluada en base a predicciones fuera de la muestra con datos hasta abril del 2009. Este modelo presenta una mejora en los errores relativos de predicción en los meses de frío respecto a modelos similares que no tomaban en cuenta como variable exógena la temperatura. A modo de ejemplo con un modelo $SARIMA - IA(0, 1, 1)(0, 1, 1)_{12}$ con las mismas intervenciones que el modelo anterior pero sin considerar la variable temperatura, da como resultado un error relativo de predicción para junio del 17 %, julio 29 %, agosto 8 % y setiembre de un 7 %. Estos son los meses claves para la planificación comercial por lo que reducir el error de predicción en ellos es muy importante.

Cuadro 2: Evaluación de la performance predictiva del modelo de demanda de GLP con la variable temperatura de mayo a agosto, expresado en miles de litros

Meses	Dato	Predicción	Error relativo	Inf 95 %	Sup 95 %
200905	14993	14115	5,86 %	12433	16025
200906	22168	21151	4,59 %	18602	24049
200907	25082	25687	-2,41 %	22558	29250
200908	17233	14381	16,55 %	12611	16400
200909	14953	14149	5,38 %	12389	16158
200910	13148	10855	17,44 %	9491	12415
200911	10669	10402	2,50 %	9083	11914
200912	12195	11322	7,16 %	9871	12985
Acumulado	130442	122062	6,42 %	107039	139196

Al incluir la temperatura durante los meses de mayo a agosto como variable regresora, se puede observar en el cuadro 2 una reducción en los errores relativos de predicción, con excepción de los meses de mayo

²Se probaron diversas especificaciones con la variable temperatura en todos los meses pero todos los modelos analizados presentaron residuos correlacionados.

³Para predecir fuera de la muestra se consideran los datos de temperatura reales

y agosto. La reducción del error relativo, para el acumulado de 8 meses es sustancial, pasando de 12 % a 6 % al incorporar esta regresora. A pesar de ello, la demanda registrada en los meses de agosto y octubre continua cayeron fuera del rango esperado.

Finalmente, como forma alternativa de modelizar la temperatura, se incluyen tres regresores que permitan diferenciar los incrementos en la demanda de GLP en las diferentes épocas del año, lo que llamamos otoño (abril y mayo), lo que llamamos invierno (junio, julio y agosto) y lo que llamamos primavera (setiembre y octubre). Se estima nuevamente el modelo $SARIMA - IA(0, 1, 1)(0, 1, 1)_{12}$ con regresores de tres tipos, el efecto calendario especificado mediante D_{LV} (la sumatoria de la diferencia de entre la cantidad de días hábiles de la semana y el número de sábados y domingos para cada mes), los puntos anómalos (AO en junio de 2007, un IO en junio de 2003 y un TC en mayo de 2001), y la temperatura como ya se explicitó (T_t).

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-12} - \theta_3 \epsilon_{t-13} + \phi_1 D_{LV} + \phi_2 AO_{200706} + \phi_3 IO_{200306} + \phi_4 TC_{200105} + \phi_5 T_t + \epsilon_t$$

donde $Y_t = \log(GLP)$

En el cuadro 3 se presenta la evaluación de la performance predictiva del modelo seleccionado.

Cuadro 3: Evaluación de la performance predictiva del modelo de demanda de GLP con la variable temperatura en tres períodos del año, expresado en miles de litros

Meses	Dato	Predicción	Error relativo	Inf 95 %	Sup 95 %
200905	14993	14573	2,81 %	13012	16321
200906	22168	22246	-0,35 %	19827	24960
200907	25082	26030	-3,78 %	23158	29257
200908	17233	14330	16,85 %	12727	16135
200909	14953	15972	-6,82 %	14161	18016
200910	13148	11954	9,08 %	10580	13506
200911	10669	10413	2,39 %	9201	11785
200912	12195	11341	7,01 %	10004	12857
Acumulado	130442	126859	2,75 %	112671	142837

La inclusión de la variable temperatura modelizada de esta forma da lugar a una reducción adicional del error relativo de predicción en los 8 meses, pasando de 12 % al no incluir la temperatura, un 6,42 % con la temperatura en los meses de frío y un 2,75 % con la última especificación. La demanda de GLP del mes de agosto sigue quedando fuera del rango esperado.

Durante el proceso de selección de los modelos el indicador considerado fue el error cuadrático medio (ECM) y la raíz del error cuadrático medio (RECM), a la hora de evaluar la performance predictiva y el efecto de la incorporación de la temperatura en el modelo se observa la reducción en los errores relativos, lo que implica poner especial énfasis en las predicciones en el punto.

A modo de realizar la comparación entre los modelos presentados, en el cuadro 4 se presentan los errores relativos a 8 pasos del modelo sin temperatura, con temperatura en los meses de frío y con temperatura en las distintas estaciones.

Cuadro 4: Errores relativos para 8 meses de los modelos con diferentes alternativas de inclusión de la variable temperatura

Meses	Mod sin temp	Mod con temp en inv	Mod con temp por estación
200905	-4,58 %	5,86 %	2,81 %
200906	20,40 %	4,59 %	-0,35 %
200907	29,34 %	-2,41 %	-3,78 %
200908	-0,82 %	16,55 %	16,85 %
200909	6,72 %	5,38 %	-6,82 %
200910	17,74 %	17,44 %	9,08 %
200911	2,68 %	2,50 %	2,39 %
200912	9,24 %	7,16 %	7,01 %
Acumulado	12,11 %	6,42 %	2,75 %

A modo de resumen, se puede decir que la incorporación de la temperatura como variable exógena en el modelo univariado mensual, mejora los indicadores de calidad de ajuste y de performance predictiva (error relativo acumulado en las predicciones de los siguientes 8 meses del 2009 que van de mayo a diciembre). El modelo con mejor desempeño predictivo, dentro de las alternativas ensayadas, es aquel que incorpora la temperatura mediante tres variables regresoras, una para invierno (junio, julio y agosto), otra para otoño (abril y mayo) y otra para primavera (setiembre y octubre).

De estas pruebas se deduce que las distintas especificaciones para la variable exógena temperatura, da lugar a resultados diferentes, lo que justifica la búsqueda de especificaciones alternativas para incluir la variable en el modelo. Naturalmente que en este trabajo no se han agotado las posibilidades de especificación de la variable por lo que queda abierto el camino a mejorar los resultados predictivos del modelo mediante especificaciones alternativas.

Para analizar la sensibilidad de los modelos a la incorporación de la variable temperatura se trabajó con valores conocidos de la misma. Para que estos modelos sean utilizados para mejorar las predicciones de la demanda de GLP, es necesario tener predicciones de la temperatura media máxima. Se ha comenzado a trabajar con modelos univariados de series de tiempo para predecir la temperatura media máxima a nivel mensual, como primera aproximación se seleccionó un modelo que tiene sólo componente estacional y se utilizó como predicción el intervalo inferior al 95 % de confianza. Utilizando esto como insumo en el modelo con temperatura en los meses de mayo a agosto los errores relativos de predicción son similares a los obtenidos con el verdadero valor de temperatura, por lo que los resultados nos animan a seguir explorando modelos alternativos para la temperatura media máxima.

4.2. Modelo diario

4.2.1. Principales características de la serie

El análisis que se realiza en esta sección refiere a la demanda diaria de GLP. Se trabajó con las ventas diarias de GLP expresadas en litros, la información disponible abarca el periodo 1998-01-02 al 2010-05-31 (4533 datos). Como en general la empresa no vende GLP los domingos se trabajó con la serie considerando semanas de seis días, y en caso que los domingos tengan ventas distintas de cero se adicionan a las ventas de los sábados, lo que da lugar a que la muestra pase a tener 3885 datos.

Para evaluar la performance predictiva de los modelos se trabajó con la información hasta abril del 2009

y se realizaron predicciones a distintos horizontes. A continuación se realiza una breve descripción de los datos.

En el gráfico 5 se presenta la serie diaria de GLP desde el año 2005 al 2010 considerando semanas de

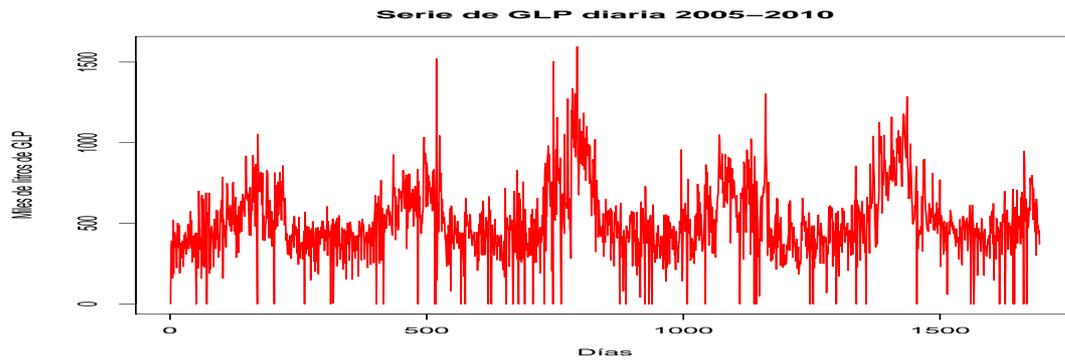


Figura 5: Serie Diaria de GLP

seis días, en dicho gráfico se visualiza un comportamiento regular al interior del año, mostrando como es de esperar mayores niveles de ventas en los períodos de invierno, además se pueden ver varios puntos a lo largo del tiempo que son cero, los mismos corresponden a días donde no se vendió, por ejemplo feriados, paros, etc. Por otro lado tanto el autocorrelograma simple como en el parcial que se presentan en el gráfico 6, indican que la serie tiene un componente estacional no estacionario. En el correlograma simple se pueden visualizar correlaciones más altas cada seis días esto podría estar indicando un comportamiento estacional al interior de la semana.

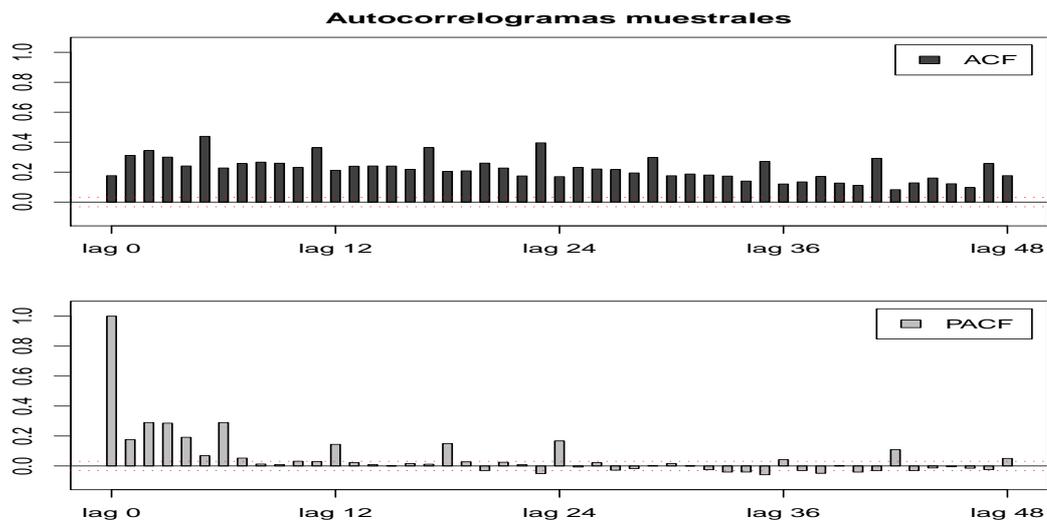


Figura 6: Correlograma Serie Diaria de GLP

En el gráfico 7 se presentan los autocorrelogramas simple y parcial de la diferencia estacional $(1 - L^6)$ de la serie de GLP. Esta transformación refleja la variación cada seis días, la variación de lunes contra lunes, martes contra martes y así con el resto de los días. Este gráfico nos permite identificar un modelo con componente de medias móviles tanto en la parte estacional como regular.

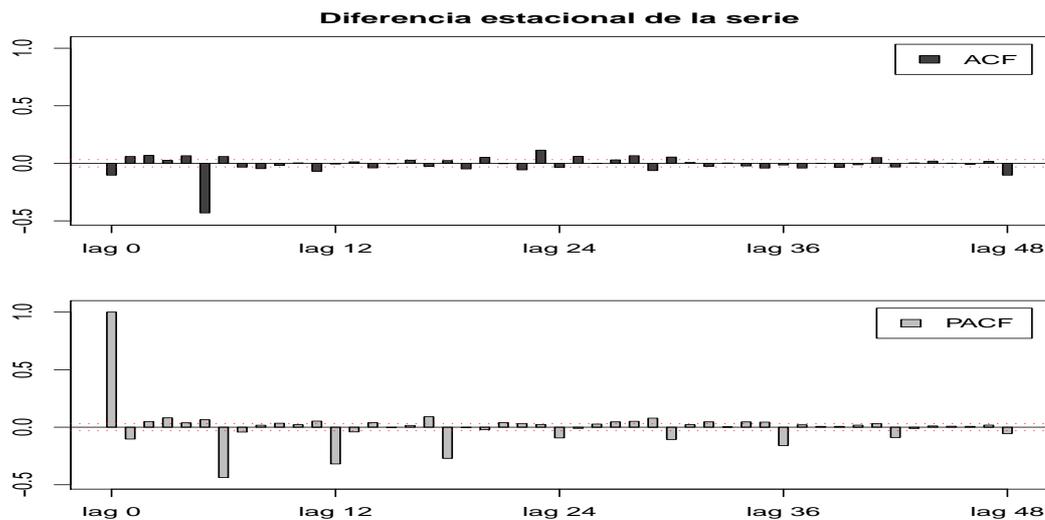


Figura 7: Correlogramma Serie Diaria de GLP con diferencia estacional de orden 6

4.2.2. Modelos estimados

Al igual que en los modelos mensuales, se estimaron distintas alternativas que consisten en el modelo sin la temperatura como variable exógena, con la temperatura máxima para cada día y finalmente la temperatura máxima sólo en los meses de frío (mayo a agosto) y cero en el resto.

En los distintos modelos que se estimaron fue necesaria la construcción de variables adicionales que fueron utilizadas como regresores en dichos modelos. En primer lugar se construyeron indicatrices que identifican los feriados laborables y los no laborables. Se incluyen variables que recogen el efecto estacional dentro de la semana y el que se da mensualmente, dentro del año, estas variables indican el día de la semana e indican el mes del año respectivamente. Por otro lado, se creó la variable turismo que identifica el momento en que se registra la semana de turismo.

En ninguno de los modelos analizados la semana de turismo y los feriados laborables fueron significativas para un nivel de significación del 5%, por tanto dichos regresores no se incluyen.

Luego de analizar una serie de modelos, se seleccionan dos modelos cuya diferencia sustancial es la forma en que se modeliza la estacionalidad semanal, en un caso es de tipo determinística y en otro de tipo estocástica:

1. Modelo 1: $ARIMA - IA(0, 0, 5)(0, 0, 1)_6$ con indicatrices estacionales que indican los días de la semana.
2. Modelo 2: $ARIMA - IA(0, 0, 5)(0, 1, 1)_6$

En el modelo 1 se incorporaron, 5 indicatrices que marcan los días de la semana (la variable omitida es la correspondiente al día viernes). Están construídas para semanas de seis días y recogen el efecto de la estacionalidad de tipo determinístico que se da dentro de la semana. Se incluyen además, 11 indicatrices que marcan los meses (la omitida es el mes de diciembre) para recoger la estacionalidad anual. Finalmente se incorporó una variable que marca los días donde no hay ventas de GLP (ya sea que la causa sea un día feriado, paro, o cualquier otra razón). En el modelo 2 se trabajó con las mismas variables mencionadas anteriormente como regresoras, menos las 5 indicatrices que recogen la estacionalidad semanal de forma determinística, en esta alternativa la estacionalidad de la semana es modelada de forma estocástica.

Como ambos modelos presentan un desempeño muy similar se decidió trabajar con el modelo 1 incorporando las dos alternativas para incluir la variable exógena temperatura (la temperatura máxima diaria y la temperatura sólo en los meses de mayo a agosto).

Para comparar la sensibilidad del modelo ante la incorporación de la variable temperatura, primero se estimó el modelo sin la misma. En el cuadro 5 se presentan las predicciones del modelo 1 diario sin temperatura, las predicciones presentadas son el resultado de cortar la muestra al final de cada mes y realizar predicciones a 26 pasos ⁴.

Cuadro 5: Evaluación de la performance predictiva de la demanda de GLP diaria en miles de litros, sin temperatura

Meses	Dato	Predicción	Error relativo	Inf 95 %	Sup 95 %
200905	14993	14333	4.40 %	6777	30315
200906	22168	18441	16.81 %	8752	38912
200907	25082	20562	18.02 %	9758	43328
200908	17233	16027	7.00 %	26066	33720
200909	14953	13548	9.40 %	6441	28496
Acumulado	94429	82910	12.20 %	57793	174771

En el cuadro 5 se puede ver que los meses con mayor error relativo acumulado son junio y julio que fueron los meses mayor frío en el 2009. Es importante resaltar que realizar predicciones a 26 pasos implica un nivel alto de exigencia para los modelos univariados de series temporales. Se realizaron predicciones a doce pasos (dos semanas) y los errores son menores, por ejemplo en el mes de julio a 12 pasos el error relativo es del 14 %. El modelo diario permite obtener proyecciones a corto plazo siendo el mes el horizonte máximo de predicción para el mismo, los errores de predicción a 12 pasos son menores que cuando se proyecta el mes. Considerar el mes como horizonte de proyección ocasiona que los intervalos de confianza sean muy amplios y por tanto poco informativos.

Como segunda alternativa se estima el mismo modelo incorporando la temperatura máxima en todos los meses. En el cuadro 6 se presentan la evaluación de la performance predictiva de este nuevo modelo a 26 pasos en cada mes. El error acumulado en los 5 meses es inferior que en el modelo sin temperatura, pasando de un 12 % a un 10 %, se logra disminuir el error relativo en los meses de junio, julio y setiembre aunque se incrementa el error en mayo y agosto aunque siguen siendo pequeños. Incorporar la variable temperatura tiene efectos pero no tan marcado como en el modelo mensual.

⁴Como estamos trabajando con semanas de 6 días las predicciones son a 26 pasos para completar el mes, en el caso del mes de julio se predice a 27 pasos

Cuadro 6: Evaluación de la performance predictiva de la demanda de GLP diaria en miles de litros, con temperatura

Meses	Dato	Predicción	Error relativo	Inf 95 %	Sup 95 %
200905	14993	14198	5.31 %	6788	29698
200906	22168	19090	13.88 %	9149	39836
200907	25082	21463	14.43 %	10302	44717
200908	17233	15772	8.48 %	15538	10367
200909	14953	14332	4.15 %	14332	8886
Acumulado	94430	84855	10.14 %	56108	133504

Finalmente se probó modelar la temperatura con los valores máximos de la misma en los meses de frío (de mayo a agosto) y cero en el resto. Los resultados de la performance predictiva se presentan en el cuadro 7

Cuadro 7: Evaluación de la performance predictiva de la demanda de GLP diaria en miles de litros, con temperatura de mayo a agosto

Meses	Dato	Predicción	Error relativo	Inf 95 %	Sup 95 %
200905	14993	13919	7.16 %	6618	29277
200906	22168	19068	13.98 %	9087	40011
200907	25082	21434	14.54 %	10231	44908
200908	17233	15699	8.90 %	7509	32823
200909	14953	13770	7.90 %	6591	28769
Acumulado	94429	83893	11.16 %	40038	175791

Los resultados de este modelo presentan un error relativo acumulado en los 5 meses un poco mayor al modelo con la temperatura máxima en todos los meses, sin embargo si se miran los errores al interior de cada mes las diferencias no son tan claras.

A modo de resumen, incorporar la variable temperatura en el modelo diario, reduce el error relativo principalmente en los meses de junio y julio donde se presentan los mayores errores en el modelo sin temperatura. Las dos alternativas ensayadas para considerar el efecto temperatura no presentan diferencias importantes en cuanto a desempeño predictivo.

El modelo diario permite obtener predicciones a corto plazo, aquí se presentaron predicciones a 26 pasos para poder comparar los resultados con el modelo mensual. Los resultados muestran que incorporar la variable temperatura en los meses de mayo a agosto, logra menores errores en los meses de junio y julio que si dicha variable no se incorporara. Es importante marcar que aunque existe un efecto en la disminución de los errores al incorporar la variable temperatura en el modelo diario, dicha mejora no es tan importante como la encontrada en el modelo mensual donde el error acumulado para los meses analizados logra reducirse a más de la mitad.

4.3. Comparación modelo diario y mensual

La diferente frecuencia de los datos permite que los modelos sean útiles para realizar pronósticos a distinto horizonte temporal. El modelo diario permite ajustar la planificación, diaria, semanal y quincenal. El modelo mensual se puede utilizar como herramienta de planificación mensual y anual. Por tanto ambos modelos se complementan en lo que a horizontes de planificación refiere.

Sin embargo en este bloque los compararemos en un lapso de tiempo que es factible de comparar, esto es, se consideran las proyecciones a un paso del modelo mensual y a 26 pasos, aproximadamente un mes en el diario. Se debe notar que el horizonte de predicción para el modelo diario es notoriamente mayor que para el modelo mensual, lo que puede hacer injusta la comparación en favor del modelo mensual. Se presentarán los errores relativos de predicción para ambos modelos sin la variable temperatura y con ella en los meses de mayo a agosto, y se calculan los errores para los meses de mayo a julio.

Cuadro 8: Comparación en performance predictiva del modelo diario y mensual sin temperatura

Fecha	Dato	Pred.Diario	Error.Diario	Pred.Mensual	Error.Mensual
200905	14993	14333	4,40 %	15680	-4,58 %
200906	22168	18441	16,81 %	17549	20,84 %
200907	25082	20562	18,02 %	18185	27,50 %
200908	17233	16027	7,00 %	17340	-0,62 %
200909	14953	13548	9,40 %	14757	1,31 %
Acumulado	94429	82910	12.20 %	83512	11.56 %

El modelo diario sin temperatura se desempeña mejor que el mensual en los meses de junio y julio donde los errores son más importantes, sin embargo en el acumulado ambos modelos presentan un comportamiento similar.

Cuadro 9: Comparación en performance predictiva del modelo diario y mensual con temperatura en los meses de mayo a agosto

Fecha	Dato	Pred.Diario	Error.Diario	Pred.Mensual	Error.Mensual
200905	14993	13920	7,16 %	14115	5,86 %
200906	22168	19068	13,98 %	21265	4,08 %
200907	25082	21435	14,54 %	26160	-4,30 %
200908	17233	15700	8,90 %	14550	15,57 %
200909	14953	13771	7,90 %	14681	1,82 %
Acumulado	94429	83893	11.16 %	90770.6	3.88 %

En cambio cuando se incorpora la variable temperatura durante los meses de mayo a agosto, ambos modelos mejoran su performance en los meses de junio y julio, sin embargo en mayo el error relativo del modelo diario es mayor. Incorporar la temperatura como variable exógena, tiene un efecto más claro en el modelo mensual que en el diario en cuanto a la mejora en la performance predictiva, sin embargo no

hay que olvidar que los horizontes de predicción de esta comparación son muy distintos.

En base a los resultados presentados se recomienda incluir la variable temperatura en los modelos univariados de GLP. Además en el caso del modelo mensual las distintas alternativas ensayadas, para considerar este efecto, mostraron diferencias en cuanto a la performance predictiva, por lo que es posible encontrar otras alternativas que capturen mejor el efecto temperatura en la demanda de GLP mensual.

5. Modelo multivariado

5.1. Modelos multivariados

Los modelos multivariados de series temporales se concentran en estimar y modelar la relación dinámica que existe entre s series temporales $Y_{1t}, Y_{2t} \dots Y_{mt}$ basado en observaciones de estas series en T puntos de tiempo igualmente espaciados ($t = 1 \dots T$) y también la relación entre estas series y variables exógenas $X_{1t}, x_{2t} \dots X_{nt}$ también observadas en el mismo periodo temporal.

Las posibles especificaciones de los modelos multivariados dependen de las características de las series involucradas, de forma simplificada:

- Si son **series estacionarias**, es posible utilizar modelos multiecuacionales con las variables expresadas en niveles (VAR).
- Si las series son **no estacionarias** y no hay relación de cointegración, es posible trabajar con las series en primeras diferencias (VAR en primeras diferencias).
- Un caso particular al anterior se da cuando las series son integradas pero es posible encontrar una combinación lineal entre ellas que sea estacionaria (cointegración). La especificación adecuada es un sistema VEC, un sistema de ecuaciones con las variables expresadas en primeras diferencias más un término que es la corrección del desvío del equilibrio de largo plazo.

$$\Delta Y_t = \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \pi Y_{t-1} + \epsilon_t$$

5.2. Modelo Estimado

Se ajustó un modelo de corrección del error para estimar la demanda mensual de GLP. Las variables involucradas en el modelo fueron: el logaritmo de la demanda de GLP, el logaritmo del consumo de energía eléctrica residencial, el logaritmo del IMSR, el logaritmo del precio relativo entre el GLP y la energía eléctrica, efecto calendario e intervenciones, finalmente se consideraron dos opciones para incorporar el efecto de la temperatura (serie completa de la temperatura media máxima y la temperatura media máxima en los meses de mayo a agosto).

La variable IMSR permite recoger el efecto que tienen las variaciones en el ingreso sobre la demanda del producto. Por otro lado, la demanda de energía eléctrica residencial se incorpora como bien sustituto de la demanda de GLP, mientras que la variable precio relativo recoge el efecto que las variaciones en el precio del GLP y el bien sustituto sobre la demanda. Finalmente la temperatura pretende recoger los picos de demanda que pueden producirse debido fundamentalmente al registro de bajas temperaturas.

De las variables antes mencionadas, algunas de ellas se incluyen en el modelo como variables exógenas, este es el caso de la diferencia del logaritmo del precio relativo entre el GLP y la energía eléctrica (en el corto plazo), la diferencia de la temperatura media máxima (se probaron dos opciones y ambas se incluyeron en el corto y en el largo plazo) y el logaritmo del IMSR (restringida a la relación de cointegración). Adicionalmente se incorporan variables que pretenden recoger la dinámica estacional, y un conjunto de variables, que como en los modelos univariados recogen el efecto de la cantidad de días hábiles (todas en el corto plazo). Por último se incorporan variables indicatrices que señalan la detección de puntos anómalos (en el corto plazo).

Se estimó un modelo de corrección del error (VEC), esto implica que las variables endógenas (logaritmo de la demanda de GLP y el logaritmo de la demanda de energía eléctrica residencial), son variables integradas y hay una relación de cointegración entre ellas, presentan una tendencia común la que representa el vínculo de largo plazo entre las mismas. El modelo adiciona la dinámica de corto plazo mediante las variaciones de las variables y el conjunto de variables exógenas.

El modelo antes especificado sin considerar la variable temperatura presenta un error acumulado en los 5 meses de 6,6 %. Con las dos alternativas ensayadas para la temperatura, se presentó un mejor desempeño predictivo con la temperatura sólo en los meses de mayo a agosto y cero en el resto.⁵

El signo en la relación de cointegración de la demanda de energía eléctrica residencial es positivo, esto implica que un incremento en la misma, aumenta el consumo de GLP, tal vez porque el efecto sustitución es menor que otros efectos. En cuanto al nivel de ajuste, si consideramos exclusivamente el R^2 corregido, en la primera ecuación es 0,9 y en la segunda 0,87.

En el cuadro 10 se presentan las predicciones a 5 pasos del modelo multivariado considerando la variable temperatura media máxima sólo en los meses de mayo a agosto, se muestran los resultados sólo de la ecuación de GLP. El error de predicción acumulado es $-0,68\%$, es importante indicar que en este modelo todos los valores observados caen dentro del intervalo de confianza cosa que no pasaba en los modelos univariados mensuales.

La ecuación de demanda de energía eléctrica residencial mostró un muy buen desempeño predictivo siendo el error acumulado a cinco pasos del $0,17\%$, además todos los valores observados cayeron dentro del intervalo de confianza al 95% .

Cuadro 10: Evaluación de la performance predictiva de la demanda de GLP mensual, del modelo multivariado con temperatura, en miles de litros

Meses	Dato	Predicción	Error relativo	Inf 95 %	Sup 95 %
200905	14993	15235	-1.62 %	13801	16818
200906	22168	23564	-6.30 %	21302	26066
200907	25082	23418	6.63 %	21100	25991
200908	17233	17632	-2.32 %	15838	19630
200909	14952	15224	-1.82 %	13634	17000
Acumulado	94429	95076	-0.68 %	85677	105507

⁵En este modelo se rechaza la hipótesis de autocorrelación de los residuos al 10% para la ecuación de GLP, igual se decidió trabajar con el mismo para poder realizar las comparaciones con el modelo mensual.

En la tabla 11 se presentan los resultados de predecir dentro de la muestra a 5 pasos , esto es predecir el periodo mayo-setiembre de 2009 tanto con el modelo multivariado como con el modelo univariado. Se observa claramente que a nivel acumulado el error de predicción se reduce notoriamente al trabajar con el modelo multivariado en lugar del univariado. El error de predicción relativo para los 5 meses es de $-0,68\%$ mientras que el error acumulado con el modelo univariado es de $5,24\%$. Como mencionamos anterioremente en el modelo multivariado todos los datos observados caen dentro del intervalo de confianza al 95% , mientras que en el modelo univariado como se presentó en el cuadro 2 el dato observado del mes de agosto no cae dentro del intervalo, por lo que el modelo multivariado presenta mejores resultados.

Cuadro 11: Evaluación de la performance predictiva de la demanda de GLP multivariada y mensual en miles de litros, con temperaturam en los meses de frío

Fecha	Dato	Pred.multi	Err.multi	Pred.mes	Err.mes
200905	14993	15235	-1.62 %	14115	5,86 %
200906	22168	23564	-6.30 %	21151	4,59 %
200907	25082	23418	6.63 %	25687	-2,41 %
200908	17233	17632	-2.32 %	14381	16,55 %
200909	14953	15224	-1.82 %	14149	5,38 %
Acumulado	94429	95076	-0.68 %	89483	5.24 %

6. Comentarios Finales

- La incorporación de la temperatura como variable exógena en los modelos univariados y multivariados, mejora los indicadores de calidad de ajuste y de performance predictiva.
- En lo que refiere al modelo de demanda diaria y los errores de predicción relativos, durante los meses de junio y julio, donde se registran los errores relativos más importantes en el modelo sin temperatura, se reduce el error al incorporar la variable, las dos opciones ensayadas muestran resultados similares en cuanto a performance predictiva. Al incorporar el regresor el modelo proporciona proyecciones a 26 pasos para junio y julio más ajustadas en términos de errores relativos acumulados, más pequeños que si la variable exógena no se incorpora.
- El modelo mensual con mejor desempeño predictivo es $ARIMA-IA(0, 1, 1)(0, 1, 1)_{12}$ que incorpora la temperatura mediante tres variables regresoras y el efecto calendario mediante una regresora especificada como la sumatoria de la diferencia de entre la cantidad de días hábiles de la semana y el número de sábados y domingos y tres puntos anómalos, un AO en junio de 2007, un IO en junio de 2003 y un TC en mayo de 2001.
- El modelo multivariado presenta un buen desempeño predictivo que indica el aporte de incorporar información adicional para predecir la demanda de GLP. Las dos especificaciones de la variable temperatura muestran un mejor resultado en cuanto al desempeño predictivo que si no se consideraran, pero la opción con la temperatura de mayo a agosto es mejor.
- De estas pruebas se deduce que las distintas especificaciones para la variable exógena temperatura da lugar a resultados diferentes en el caso del modelos mensual, lo que justifica la búsqueda de

especificaciones alternativas para incluir la variable en dichos modelos, a partir de este trabajo no se han agotado las posibilidades de especificación de la variable, por lo que queda abierto el camino a mejorar los resultados predictivos del modelo mediante especificaciones alternativas.

- Para mejorar la performance predictiva en los modelos de demanda de GLP es importante obtener buenas predicciones de la temperatura, las pruebas realizadas con modelos univariados mensuales para la temperatura media máxima muestran resultados alentadores para seguir investigando en este sentido.

Referencias

- [1] Akinboade,O.;Ziramba,E.; Kumo,W. (2008) “The demand for gasoline in South Africa: An empirical analysis using cointegration techniques”.*Energy Economics* 30 (2008) 3222-3229
- [2] Box,G.; Jenkins, G.; Reinsel,G. (1994) “Time Series Analysis: Forecasting and Control”, Third Edition, Prentice Hall
- [3] Brockwell, P.J. y R.A.Davis (1991) “Time Series: Theory and Methods”, Second Edition, Springer-Verlag
- [4] Chen, C. y Liu, L. (1993) “Joint estimation of model parameters and outlier effects in time series”, *Journal of the American Statistical Association*, ISSN 0162-1459, Vol. 88, N° 421, 1993 , pags. 284-297
- [5] Enders,W. (1995) “Applied Econometric Time Series”. John Wiley and Sons.
- [6] Fuller, W. A. (1996) “Introduction to Statistical Time Series”, Second Edition, John Wiley and Sons.
- [7] Kaiser, R. y Maravall, A. (2001) “Seasonal outliers in time series”. Universidad Carlos III. Madrid
- [8] Hamilton, J. (1994) “Time Series Analysis”. Princeton University Press
- [9] Harvey, A.C. (1981) “Time Series Models”, MIT Press
- [10] Hendry,D.(1995)“Dynamic Econometrics” Oxford University Press
- [11] Kui Yin Cheung,Thomson,E. (2004) “ The Demand for Gasoline in China: a Cointegration Analysis”. *Journal of Applied Statistics*. Vol. 31.N° 5,533-544
- [12] Lutkepohl,H. (2004) “Recent Advances in Cointegration Analysis”. Economics Working Papers from European University Institute. No ECO2004/12.
- [13] Negro,M. (2008) “Estimación de la demanda de combustibles”, Informe final de pasantía-Licenciatura en Estadística
- [14] Rodríguez,M. (2006) “ Estimación de la demanda de combustibles en República Dominicana” Texto de discusión 6, Unidad de Análisis Económico. Secretariado Técnico de la Presidencia.