

Facultad de Ciencias

Departamento de Estadística

Carrera Ingeniería Estadística

PROYECTO DE TITULO II

Asignatura : **PROYECTO TITULO II (220068)**
Título : "Elaboración de un Modelo de Predicción de Demanda de Gasolina a Nivel País"
Realizado por : Rodrigo Hernán Jiménez Morales
Profesor Guía : Dr. Sergio Contreras Espinoza
Co-Guía : Sr. Carlos Acevedo W.
Semestre : Primer Semestre 2014
Fecha : Concepción, Abril 2014



Resumen

El objetivo de este Proyecto de Título es realizar un Modelo de Predicción de Demanda de Gasolina en Chile, para esto fijaremos nuestro interés en la obtención de modelos con distintas herramientas estadísticas: la Regresión Lineal Múltiple (RLM) y los Modelos ARIMA para series de tiempo mediante la metodología de Box y Jenkins más algunas simulaciones con Redes Neuronales Artificiales (RNA) a través de una herramienta dispuesta en el software Matlab.

Una revisión general con las nociones básicas de la Regresión Lineal se expuso brevemente en el proyecto de título I que antecede a este informe, así como también se señalaron los conceptos fundamentales de la metodología de Box y Jenkins y de Redes Neuronales para poder diseñar RNA que sirvan para la predicción.

En el ajuste por mínimos cuadrados que utiliza la RLM se usarán las variables pib per cápita, precio de la gasolina, parque automotriz y rezagos como predictoras de la variable dependiente: Demanda de Gasolina. Nos detendremos un instante en realizar un estudio de la variable parque de vehículos a gasolina dividida en parque de vehículos de pasajeros y comerciales ya que de ésta, a diferencia de las otras, no se dispone de una serie larga con valores precisos.

Dada la serie mensualizada de demanda de gasolina, primero se obtendrá un modelo ARIMA para la predicción de valores futuros y luego se obtendrá otro modelo para este objetivo a través de redes neuronales.

El informe constará de tres partes, primero se hace una introducción al tema, se proponen los objetivos y justificación del tema. En el capítulo dos estudiamos el modelo lineal de regresión múltiple, finalmente se proponen los modelos de series de tiempo, ARIMA y de redes neuronales.

Cuando quieres algo,
todo el universo conspira para que realices tu deseo

Agradecimientos

A mi padre Humberto Jiménez Muñoz, por entregar su vida para que yo haga lo que quiera con la mía.

A mi madre Palmira Morales Yañez, por todo el amor y comprensión que ha tenido conmigo los últimos 30 años de su vida.

A mi hermano Fernando Jiménez Morales, por las lecciones de vida que me ha dado.

A Mara y en especial a Jessica, por estar ahí siempre, las quiero mucho.

A mi abuelo Humberto Jiménez Bustamante, ejemplo de vida, un ser intachable.

A mi hija Isidora Jiménez Saez, gracias por todo lo que me entregas en esas pequeñas ventanitas de tiempo.

Índice general

1. INTRODUCCIÓN	6
1.1. Origen del Tema	6
1.2. Justificación	7
1.3. Objetivos del Estudio	7
1.3.1. Objetivo General	7
1.3.2. Objetivos Específicos	7
2. MODELO DE REGRESIÓN LINEAL MÚLTIPLE	9
2.1. Modelo en Uso	9
2.2. Modelo propuesto	11
2.2.1. Parque de vehículos a gasolina	11
2.2.2. Obtención del modelo	11
3. MODELOS PARA LA SERIE DEMANDA DE GASOLINA MENSUALIZADA	14
3.1. Metodología de Box y Jenkins, ARIMA	14
3.2. Experimentos y Resultados para Redes Neuronales Artificiales	21
3.2.1. Algoritmo de Levenberg - Marquardt	21
3.2.2. Ajuste	22
3.2.3. Predicción	24
4. CONCLUSIONES	25

Índice de figuras

2.1. Modelo Enap reestimado con datos de 1986 a 2007	10
2.2. Ajuste y estimación modelo 2.1	10
2.3. Parque de vehículos a gasolina y suma de comerciales y pasajeros 2001-2012	12
2.4. Demanda de gasolina y ajuste (1986 a 2007)/pronóstico (2008 a 2012)	13
2.5. Estadísticos para el modelo 2.2	13
3.1. Secuencia de la demanda de gasolina mensual desde enero de 1992 hasta diciembre de 2011	14
3.2. Diagrama de caja por año para la demanda mensual entre 1992 y 2011	15
3.3. Tabla de resultados de la prueba de homogeneidad de varianzas de Levene para la demanda de gasolina	15
3.4. Serie demanda de gasolina diferenciada una vez: $Y_t = y_t - y_{t-1}$, enero 1992-diciembre 2011	16
3.5. Correlograma simple para la serie demanda de gasolina diferenciada entre enero 1992 a diciembre 2009	17
3.6. Correlograma parcial de la serie demanda1	17
3.7. Correlograma simple con las transformaciones de primera diferencia en la parte regular y primera diferencia en la parte estacional	18
3.8. Correlograma parcial con las transformaciones primera diferencia en la parte regular y primera diferencia en la parte estacional	18
3.9. Demanda real y estimada para la demanda mensual	19
3.10. Parámetros para el modelo $ARIMA(2, 1, 1) \times (0, 1, 1)_{12}$, la constante resulta ser no significativa	19
3.11. Correlograma simple y parcial de los residuos del modelo ARIMA para la serie demanda de gasolina	20
3.12. Valores reales y predicción para la serie demanda de gasolina para los años 2011 y 2012 a través del modelo $ARIMA(2, 1, 1) \times (0, 1, 1)_{12}$ ajustado con valores de enero de 1992 a diciembre de 2010	20
3.13. Error cometido por la predicción a través del modelo ARIMA	21
3.14. Ajuste de la red neuronal para los años 1993 a 2010	23
3.15. Predicciones para 2011 y 2012 para la red neuronal con 3 nodos en la capa oculta.	24
3.16. Error de la predicción de la red neuronal para los años 2011 y 2012.	24
4.1. Comparación de los errores de predicción de los métodos ARIMA y redes neuronales para los años 2011 y 2012	25

1

INTRODUCCIÓN

En este capítulo se establece el enfoque del estudio, que busca obtener un modelo de demanda de gasolina a través de, por un lado, la técnica de regresión lineal - aprovechando los datos anuales tanto de la demanda como de las variables explicativas - y por otro lado de la metodología de Box y Jenkins y de redes neuronales, usando datos mensualizados. También se pretende dar a conocer el conjunto de características necesarias e imprescindibles de este trabajo: sus objetivos, justificación, sus alcances y limitaciones.

1.1 Origen del Tema

La Empresa Nacional del Petróleo (Enap) es una empresa de energía, 100% del Estado de Chile, líder en hidrocarburos, integrada, que provee productos y servicios que satisfacen las necesidades de los clientes y contribuyen al desarrollo sustentable de los países y de las comunidades en que opera.

En la estadía en esta empresa durante el verano 2013, surgió la necesidad de modelar y hacer predicciones de la demanda de gasolina para ser usada en la Dirección de Planeamiento Estratégico, perteneciente a la Gerencia Planeamiento y Gestión, con el objetivo de proyectar la demanda y tomar decisiones derivadas de esta proyección.

La toma de decisiones se puede clasificar en corto y largo plazo:

- **Corto Plazo:** Conocer por anticipado la demanda de los principales productos que se deben abastecer (gasolinas, diesel, etc.), permite hacer una adecuada compra de crudos, cuyo rendimiento se ajuste a la canasta de demanda proyectada.

Por ejemplo si se vé un incremento de gasolina, será necesario comprar más crudo liviano, que es el que contiene más gasolina ó alternativamente, importar gasolina terminada. Esto se puede hacer en la medida que se tiene una buena estimación de la demanda mensual con una anticipación mínima de 2 meses, que es el tiempo medio que transcurre entre que se compra el crudo, llega a Chile, se refina, se almacena y finalmente se venden los productos obtenidos de él.

Un segundo uso para el modelo es en la preparación del Presupuesto Anual del «próxi-

mo año», donde se establecen la metas de producción, utilidad, etc. En este caso se necesita la canasta de venta mensual y la demanda de gasolina es parte de ésta.

- **Largo Plazo:** Las proyecciones de consumo de largo plazo se utilizan en la evaluación de proyectos de refinería, definiendo el mercado al cual se tendrá acceso, por ejemplo, si el mercado de gasolina se proyecta creciente, surgirán proyectos nuevos, que se harán cargo de dicha demanda.

1.2 Justificación

Si bien es cierto actualmente existe un modelo usado por Enap en función de los precios, consumo rezagado, pib y población, este debe actualizarse considerando otras variables como parque automotriz.

Principalmente esta actualización tiene como beneficio que el modelo tenga credibilidad y que éste no quede obsoleto o falta de información, lo que provocaría ante algunos escenarios, malas predicciones que evidentemente llevan a malas desiciones.

Con el fin de que las predicciones obtenidas sean consideradas en la práctica, es que estas deben tener un error de predicción mínimo o al menos estimado, esto es, tener estimaciones de error a priori.

Considerando lo anterior es muy importante conocer las variables que explican el consumo de la gasolina y sus magnitudes para analizar el comportamiento de éstas. Así por ejemplo para añadir el parque vehicular como variable explicativa se deben tener buenas proyecciones o estimaciones de ésta, lo que implica una labor adicional. Para otras variables como pib o dólar ya se tienen proyecciones bastante aceptables dadas por el Banco Central¹ e INE².

1.3 Objetivos del Estudio

1.3.1. Objetivo General

Elaborar un modelo de predicción de demanda de gasolina.

1.3.2. Objetivos Específicos

Se enumeran los objetivos especificos en forma secuencial.

- Conocer la forma actual en que Enap hace predicción de la demanda de gasolina.
- Hacer una revisión general de las materias que nos ayudarán a elaborar el modelo.

¹www.bcentral.cl

²www.ine.cl

-
- Pensar, obtener y adecuar las posibles variables explicativas.
 - Proponer modelos posibles para predecir la demanda.
 - Estimar los modelos, validando los supuestos y evaluar la capacidad predictiva de ellos para elegir el mejor modelo posible.

2

MODELO DE REGRESIÓN LINEAL MÚLTIPLE

Las series con que trabajaremos (Demanda de gasolina y sus explicativas) se almacenan considerando un dato por año. Es de gran importancia para Enap poder contar con pronósticos para la planificación y toma de decisiones.

Un método usado en forma usual es el de la regresión lineal múltiple que usa el valor de una o más series para predecir otra u otras.

Una vez ajustado los valores por mínimos cuadrados - acá usaremos mínimos cuadrados en dos etapas debido a que el rezago de la demanda es también utilizado como predictora - se tiene el trabajo de predecir la serie en estudio usando la proyección de valores futuros de otras series por información adicional.

2.1 Modelo en Uso

Actualmente Enap usa el modelo de regresión lineal múltiple 2.1 para la predicción de la demanda de gasolina

$$\begin{aligned} dem_t = & \exp(-2,63 - 0,25\ln(pr_t) + 0,33\ln(pib_t/pob_t) + \\ & + 0,51\ln(dem_{t-1}/pob_t) + 0,01\ln(vn_t(1 - td_t)))pob_t + u_t \end{aligned} \quad (2.1)$$

Este predice la demanda por medio del precio pr_t , el pib per cápita pib_t/pob_t , un rezago de la demanda per cápita dem_{t-1}/pob_t y una variable que incorpora el parque de vehículos nuevos que usan gasolina a través de la relación $vn_t(1 - td_t)$, donde vn_t son los vehículos que ingresan al país cada año, registrados por la aduana y td_t es la tasa de diselización, que se define como el cociente entre los vehículos livianos nuevos con motor diesel y el total éstos vehículos. Esta última variable descrita no representa de buena forma el parque de vehículos que consumen gasolina y que circulan en el país, además el ajuste de este modelo se hizo hace ya unos años, por lo que nueva información puede hacer variar en forma significativa los coeficientes del modelo. También estudiaremos si estas son las únicas variables que inciden en la demanda de gasolina o si pueden haber otras que no se han considerado hasta ahora.

Reestimando este modelo con información de los años 1986 a 2007 obtenemos un modelo con parámetros estadísticamente iguales a los del modelo 2.1:

Dependent Variable: LOG(DEMANDA/POBL)
 Method: Least Squares
 Date: 10/27/13 Time: 22:17
 Sample (adjusted): 1987 2007
 Included observations: 21 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.468240	0.452655	-5.452811	0.0001
LOG(PRECIO)	-0.217975	0.041642	-5.234512	0.0001
LOG(PIB/POBL)	0.311858	0.080783	3.860429	0.0014
LOG(DEMANDA(-1)/POBL)	0.554308	0.076326	7.262414	0.0000
LOG(NUEVOS*(1-TD))	0.027023	0.015728	1.718137	0.1051

R-squared	0.990235	Mean dependent var	-8.642276
Adjusted R-squared	0.987794	S.D. dependent var	0.170741
S.E. of regression	0.018864	Akaike info criterion	-4.898883
Sum squared resid	0.005694	Schwarz criterion	-4.650188
Log likelihood	56.43828	Hannan-Quinn criter.	-4.844910
F-statistic	405.6258	Durbin-Watson stat	1.826905
Prob(F-statistic)	0.000000		

Figura 2.1: Modelo Enap reestimado con datos de 1986 a 2007

Del resumen del modelo 2.1, vemos que el ajuste es bastante bueno. Sin embargo al hacer la proyección los valores tienden a alejarse del valor real.

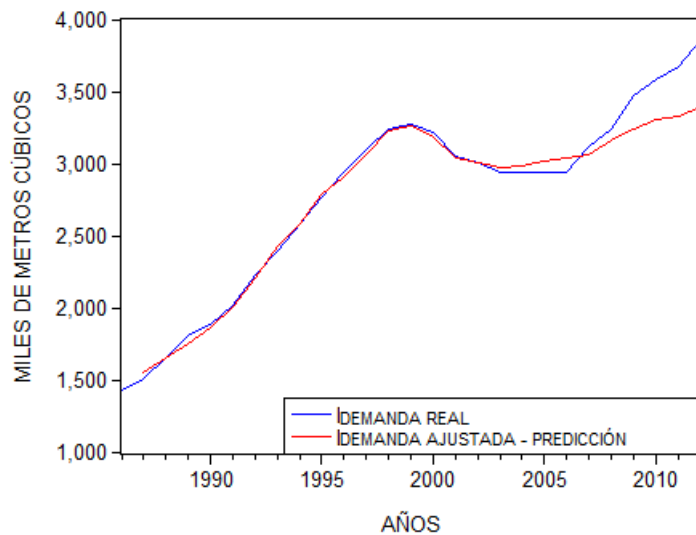


Figura 2.2: Ajuste y estimación modelo 2.1

El Cuadro 2.1 muestra el pronóstico de la demanda para los años 2008 a 2012, en él vemos que, a medida que pasan los años, el ajuste empeora. Esto puede deberse a la falta de información del modelo, ya que hasta 1998 la curva de demanda siempre experimentó alzas superiores al 5%, luego de 2000 a 2006 la demanda bajó o se mantuvo casi constante, 2007 fue el primer año en el que se observó un repunte considerable de la demanda. Así el periodo 2000-2006 hace que la predicción no pueda seguir a la curva demanda en su recuperación.

Lo expuesto hace pensar en que aunque el modelo 2.1 ajuste bien a la curva de demanda, las variables involucradas no son suficientes para que el modelo interprete de buena forma las

Año	2008	2009	2010	2011	2012
Real (Mm3)	3.240	3.475	3.588	3.667	3.856
Predicción (Mm3)	3.160	3.242	3.308	3.323	3.393
Error (%)	2,46	6,71	7,80	9,37	12,02

Cuadro 2.1: Pronósticos para 2008 a 2012 usando modelo 2.1

alzas o bajas que tenga la demanda de gasolina en periodos cortos sucesivos.

2.2 Modelo propuesto

En el proyecto de título I, que antecede a este documento, analizamos las variables de interés para el modelo que propondremos, solo haremos algunas observaciones de la variable parque de vehículos a gasolina.

2.2.1. Parque de vehículos a gasolina

En la actualidad existe un registro anual, proporcionado por el INE, del parque de vehículos a gasolina; sin embargo, esta información es escasa, pues solo se tiene el dato anual desde el año 2001. Unas series más largas son las de vehículos de pasajeros y comerciales, que son una buena aproximación al parque a gasolina, puesto que los vehículos comerciales y de pasajeros son prácticamente los únicos que usan este combustible, ya que camiones y buses, que son el resto del parque, usan diesel.

Es natural pensar, entonces, en usar como medida del parque a gasolina la suma o una ponderación del parque de vehículos comerciales y de pasajeros, teniendo, eso si, como consideración que hay una fracción - que varía año a año y que es difícil de determinar - del parque de pasajeros y comerciales que usan diesel.

Gráficamente (ver figura 2.3) vemos que la suma comerciales y pasajeros se asemeja bastante a la curva de parque de gasolina, al calcular la correlación para el periodo 2001 a 2012 es 0.9996, lo que ratifica lo pensado.

2.2.2. Obtención del modelo

En un principio el modelo propuesto tomará como variables explicativas el PIB, el precio por litro de la gasolina (Pr_t), un rezago de la demanda de gasolina (dem_{t-1}) y la suma del parque automotriz de pasajeros y comerciales (gas_t), además se modela el error por un modelo ARIMA. Dispuestos los datos de nuestras variables para los años 1986 a 2007 y poniendonos en varias situaciones -sacando y poniendo variables- obtenemos el mejor modelo, este es:

$$\begin{aligned}
 dem_t &= \exp(-1,90 - 0,26\ln(pr_t) + 0,51\ln(dem_{t-1}/pob_t) \\
 &\quad + 0,37\ln(gas_t/pob_t))pob_t + u_t \\
 u_t &\sim MA(1)
 \end{aligned}
 \tag{2.2}$$

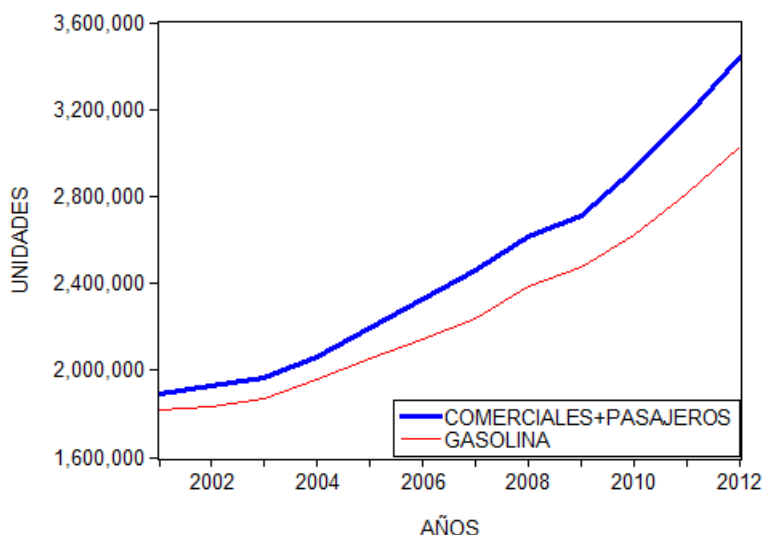


Figura 2.3: Parque de gasolina y suma de parque de comerciales y pasajeros entre los años 2001 y 2012

Entonces, la demanda queda explicada por la variables precio pr_t , un rezago dem_{t-1} , la suma del parque automotriz gas_t , además del error que se modela por un proceso de medias móviles $MA(1)$.

Obtenemos pronósticos para este modelo:

Año	2008	2009	2010	2011	2012
Real (Mm3)	3.240	3.475	3.588	3.667	3.856
Predicción (Mm3)	3.179,36	3.472,27	3.540,43	3.571,01	3.679,01
Error (%)	1,87%	0,08%	1,33%	2,62%	4,59%

Cuadro 2.2: Pronósticos para 2008 a 2012 usando modelo 2.2

Mostramos en la figura 2.4 que tanto ajuste como predicción siguen de cerca a la serie de- manda de gasolina.

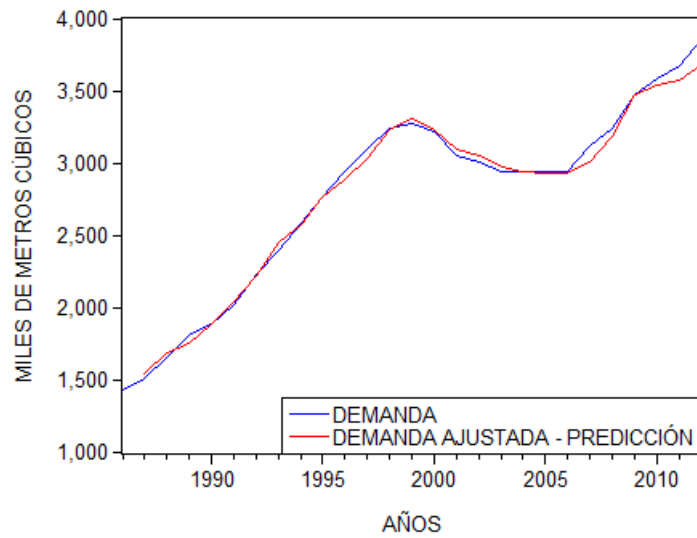


Figura 2.4: Demanda de gasolina y ajuste (1986 a 2007)/pronóstico (2008 a 2012)

Dependent Variable: LOG(DEMANDA/POBL)
 Method: Least Squares
 Date: 10/29/13 Time: 19:12
 Sample (adjusted): 1987 2007
 Included observations: 21 after adjustments
 Convergence achieved after 19 iterations
 MA Backcast: 1986

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.897801	0.303666	-6.249638	0.0000
LOG(PRECIO)	-0.255829	0.026623	-9.609469	0.0000
LOG(DEMANDA(-1)/POBL)	0.506807	0.064505	7.856864	0.0000
LOG(GASOLINA/POBL)	0.373264	0.064057	5.827101	0.0000
MA(1)	-0.901664	0.250305	-3.602265	0.0024
R-squared	0.989457	Mean dependent var	-8.642276	
Adjusted R-squared	0.986821	S.D. dependent var	0.170741	
S.E. of regression	0.019601	Akaike info criterion	-4.822199	
Sum squared resid	0.006147	Schwarz criterion	-4.573503	
Log likelihood	55.63309	Hannan-Quinn criter.	-4.768225	
F-statistic	375.3880	Durbin-Watson stat	0.996025	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.90			

Figura 2.5: Estadísticos para el modelo 2.2

3

MODELOS PARA LA SERIE DEMANDA DE GASOLINA MENSUALIZADA

Se dispone de datos mensualizados de la demanda de gasolina, con esta información -y dado las características de la serie- es posible modelar la demanda con modelos ARIMA y Redes Neuronales Artificiales RNA.

3.1 Metodología de Box y Jenkins, ARIMA

La metodología de Box y Jenkins se aplicará siguiendo las etapas que describimos a continuación:

- Análisis exploratorio de la serie
 - Análisis gráfico de la serie

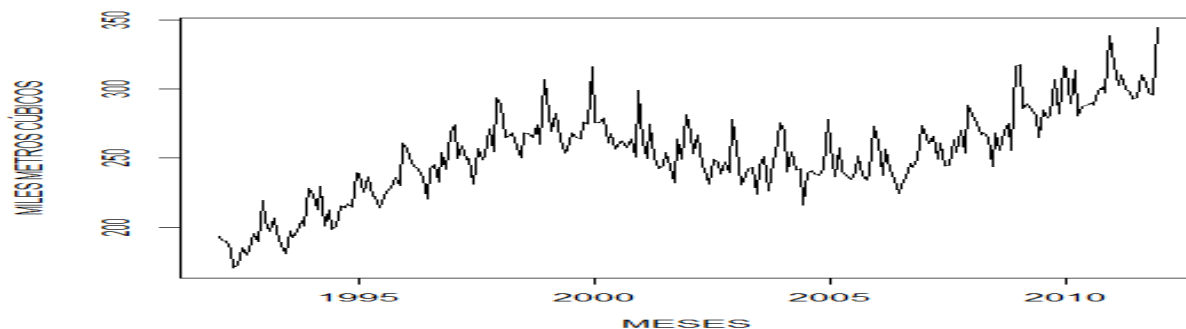


Figura 3.1: Secuencia de la demanda de gasolina mensual desde enero de 1992 hasta diciembre de 2011

Hasta el año 2000 aproximadamente hay una tendencia al alza, luego entre 2001 a 2006 la demanda se presenta a la baja o es casi constante, luego retoma la tendencia al alza. Claramente la media no es constante, por lo que la serie debe diferenciarse. Observamos estacionalidad anual. La serie no es estacionaria.

- Obtenemos una idea del comportamiento de la varianza mediante un diagrama de caja por año:

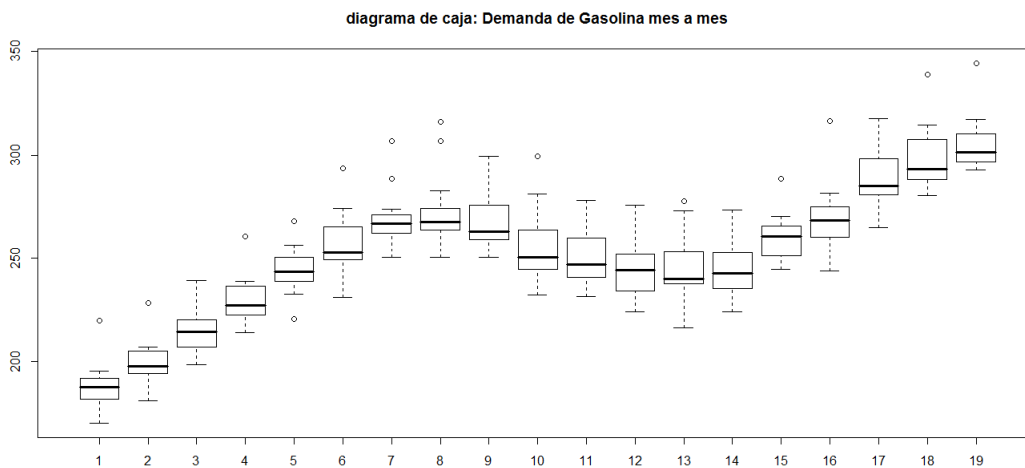


Figura 3.2: Diagrama de caja por año para la demanda mensual entre 1992 y 2011

Del gráfico puede verse que quizás hay homocedasticidad, esto puede verificarse con la Prueba de Homogeneidad de Varianza de Levene:

Prueba de homogeneidad de varianzas			
Demanda mensual de gasolina			
Estadístico de Levene	gl1	gl2	Sig.
,231	19	220	1,000

Figura 3.3: Tabla de resultados de la prueba de homogeneidad de varianzas de Levene para la demanda de gasolina

Dada la significancia obtenida (1,000) no existe evidencia significativa para rechazar la hipótesis nula de homogeneidad de varianzas y verificamos la homocedasticidad.

- Al diferenciar la serie se obtiene la gráfica de la Figura 3.4:

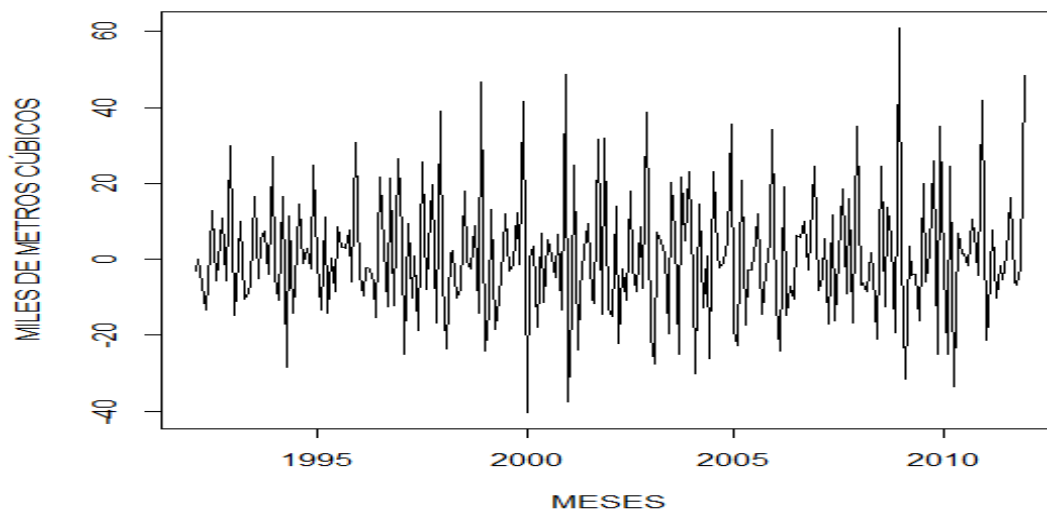


Figura 3.4: Serie demanda de gasolina diferenciada una vez: $Y_t = y_t - y_{t-1}$, enero 1992-diciembre 2011

Ahora la serie no tiene tendencia (media constante) y también se cumple la homogeneidad de la varianza.

- Identificación del modelo: Haciendo uso de los patrones que se observan en los correlogramas simples y parciales podemos identificar el orden del componente autorregresivo y de promedio móvil:

Correlograma Simple	Correlograma Parcial	Modelo
Decae lentamente	Se corta después del retardo p	$AR(p)$
Se corta después del retardo q	Decae lentamente	$MA(q)$
Decae lentamente	Decae lentamente	$ARMA(p, q)$

Para la identificación del modelo se considera el 90% de los datos, esto es, desde enero de 1992 a diciembre de 2009, correspondientes a 216 meses de los 240 meses que corresponden al lapso de tiempo estudiado.

Por simplicidad, llamaremos **demanda1** a la serie demanda una vez diferenciada y extraído el 90% de los datos ya mencionados.

- Correlogramas simple y parcial de la serie demanda1:

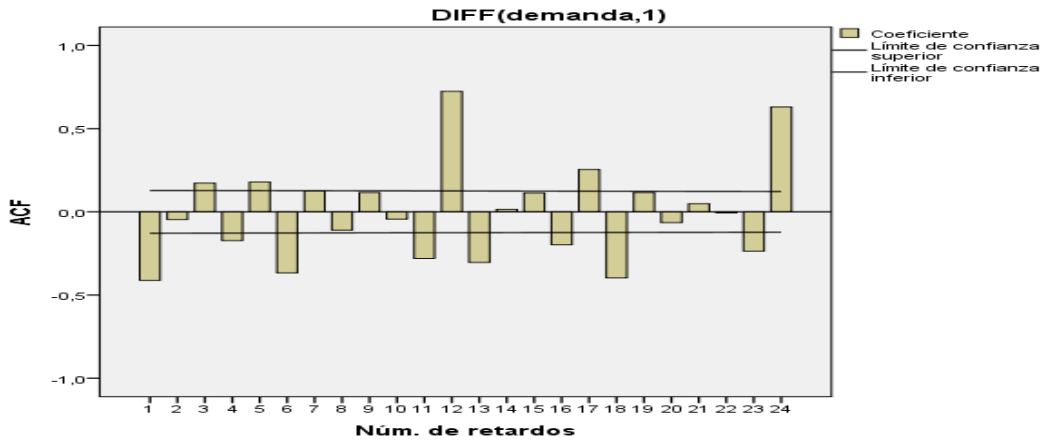


Figura 3.5: Correlograma simple de la serie demanda1

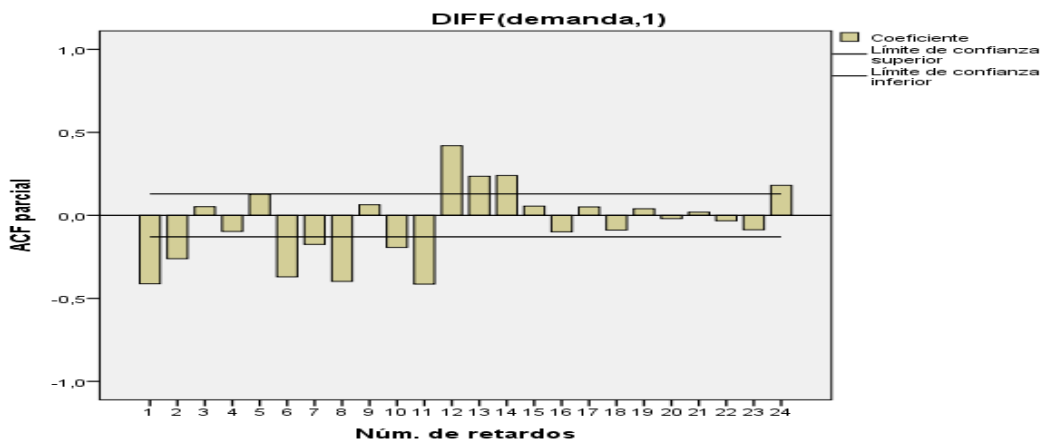


Figura 3.6: Correlograma parcial de la serie demanda1

En el correlograma simple se nota estacionalidad de 12 meses, se debe aplicar primera diferencia en la parte estacional.

- Correlograma Simple y Parcial con diferencia en la parte regular y en la parte estacional:

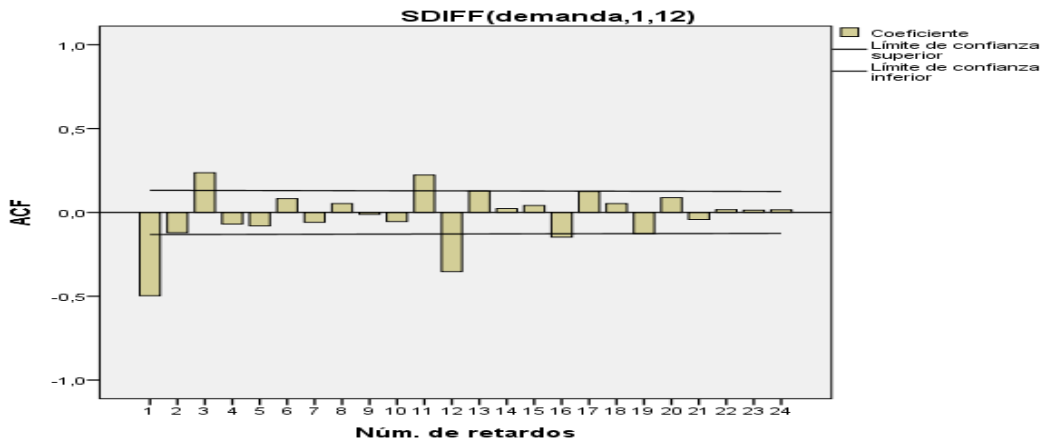


Figura 3.7: Correlograma simple con las transformaciones de primera diferencia en la parte regular y primera diferencia en la parte estacional

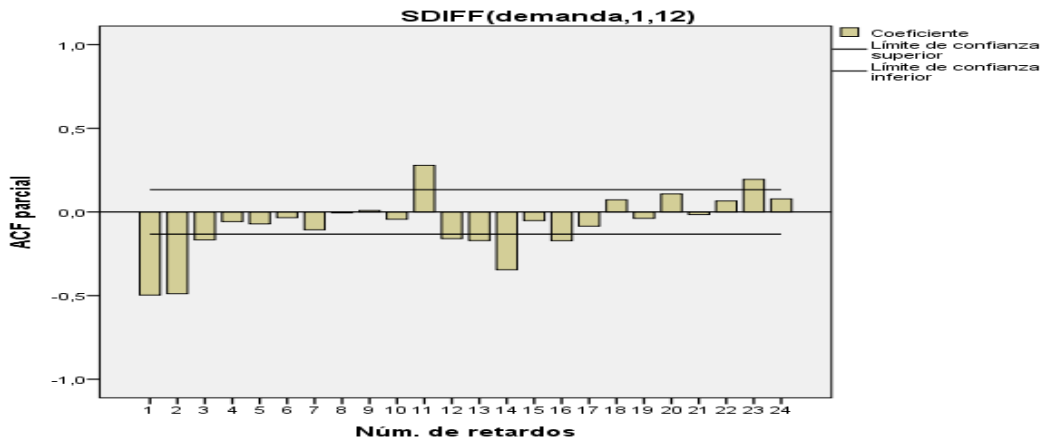


Figura 3.8: Correlograma parcial con las transformaciones primera diferencia en la parte regular y primera diferencia en la parte estacional

- Estimación de Parámetros De los correlogramas expuestos y teniendo en cuenta las variaciones de los coeficientes de esta se identifica el modelo $ARIMA(2, 1, 1) \times (0, 1, 1)_{12}$. Se realizaron numerosas pruebas, obteniéndose resultados satisfactorios con los siguientes modelos:
 - a. $ARIMA(2, 1, 1) \times (0, 1, 0)_{12}$
 - b. $ARIMA(2, 1, 1) \times (0, 1, 1)_{12}$
 - c. $ARIMA(2, 1, 0) \times (0, 1, 2)_{12}$

El modelo de mejor ajuste resultó ser **b.** con un $BIC = 17,98$, un $MAPE=2,270$ y un $R^2=0,942$

En la figura 3.9 vemos la demanda real y sus valores ajustados, al principio el ajuste no es muy bueno y los picos alcanzados en enero de cada año son valores difíciles de alcanzar, pero el ajuste va mejorando paulatinamente. Puede decirse que la serie ajustada representa bastante bien a la serie original.

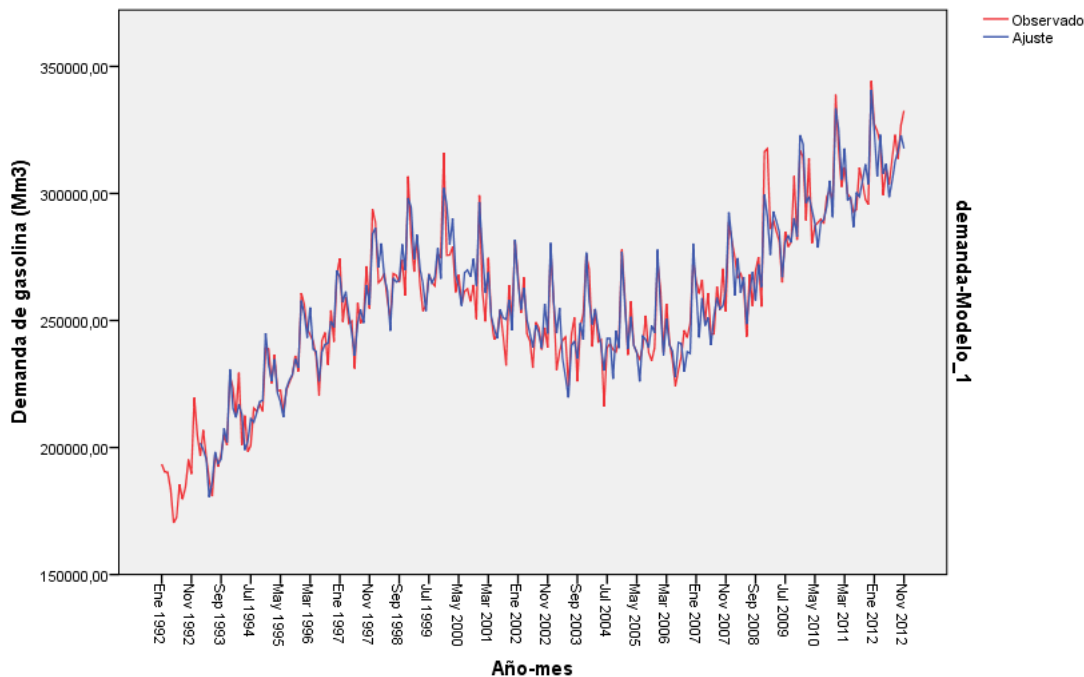


Figura 3.9: La demanda real se muestra en rojo, la demanda ajustada (Enero 1992 a Diciembre 2009) y predicción (Enero 2010 a Noviembre 2012) en azul

Parámetros del modelo ARIMA				Estimación	ET	t	Sig.
demanda de gasolina en Mm3 Sin transformación	Constante			-22,454	49,328	-.455	,649
	AR	Retardo 1		-,503	,104	-4,856	,000
		Retardo 2		-,418	,080	-5,198	,000
	Diferencia			1			
	MA	Retardo 1		,343	,110	3,116	,002
		Diferenciación estacional			1		
	MA, estacional	Retardo 1		,765	,052	14,725	,000

Figura 3.10: Parámetros para el modelo $ARIMA(2, 1, 1) \times (0, 1, 1)_{12}$, la constante resulta ser no significativa

- **Análisis de los residuos** La prueba de Kolmogorov Smirnov permite verificar si los residuos siguen una distribución Normal. Se obtuvo una significancia de 0,855, lo que indica que no puede rechazarse la hipótesis nula de normalidad de los residuos. La media de los errores es 0,0078, aproximadamente cero.

En los correlogramas de los residuos (figura 3.11) se observan valores significativos, lo que indica que queda cierta información sobre la estructura de la serie.

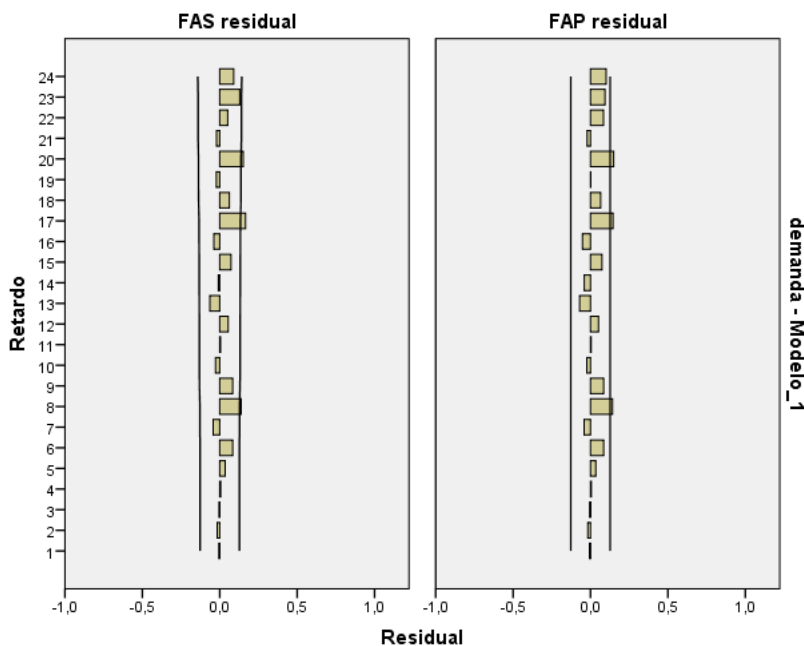


Figura 3.11: Correlograma simple y parcial de los residuos del modelo ARIMA para la serie demanda de gasolina

- **Predicción** En el apéndice se presentan los valores de la serie demanda de gasolina, las predicciones del modelo ARIMA estudiado y el error cometido en el ajuste y predicción. En la figura 3.9 se representan los valores originales de la serie demanda de gasolina y las predicciones para el periodo 2012, se observa que el ajuste es bastante bueno, aun cuando hay algunos errores algo elevados para algunos meses, debido quizás a que algo de información no es proporcionada por los rezagos ni por la parte de medias móviles.

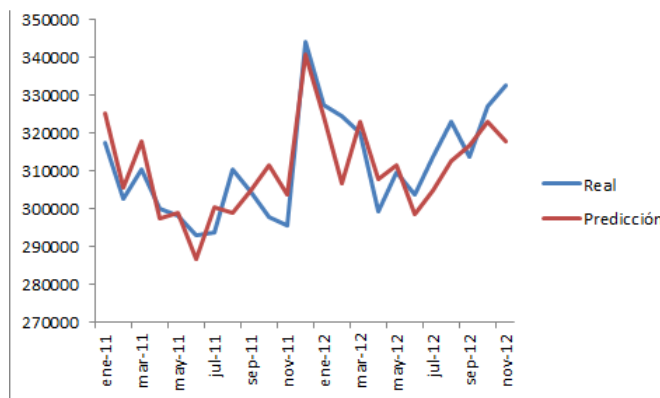


Figura 3.12: Valores reales y predicción para la serie demanda de gasolina para los años 2011 y 2012 a través del modelo $ARIMA(2, 1, 1) \times (0, 1, 1)_{12}$ ajustado con valores de enero de 1992 a diciembre de 2010

El error promedio para los dos años de predicción es de un 2,11%.

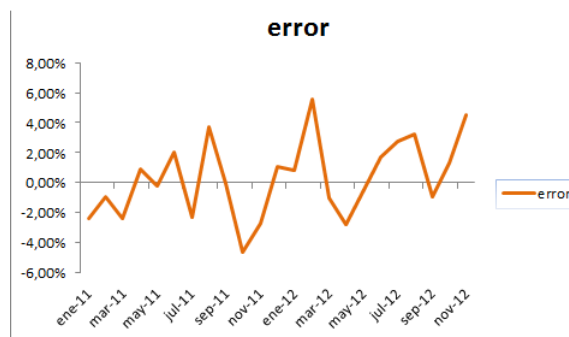


Figura 3.13: Error cometido por la predicción a través del modelo ARIMA

3.2 Experimentos y Resultados para Redes Neuronales Artificiales

Se usará el software MatLab para ajustar la curva de la demanda de gasolina por redes neuronales artificiales. MatLab usa una mejora del algoritmo Backpropagation llamado Algoritmo de Levenberg - Marquardt para hacer el ajuste. Este algoritmo es una modificación del método de Newton, el que fue diseñado para minimizar funciones que sean la suma de los cuadrados de otras funciones no lineales; es por ello que el algoritmo de Levenberg - Marquardt, tiene un excelente desempeño en el entrenamiento de redes neuronales donde el rendimiento de la red esté determinado por el error medio cuadrático.

Presentamos una noción del algoritmo de Levenberg - Marquardt.

3.2.1. Algoritmo de Levenberg - Marquardt

El método de Newton para optimizar el rendimiento $e(x)$ es:

$$X_{k+1} = X_k - A_k^{-1} g_k$$

$$A_k = \nabla^2 e(x)|_{x=x_k} \quad g_k = \nabla e(x)|_{x=x_k}$$

Si asumimos que el error del ajuste, $e(x)$, es una suma de funciones cuadráticas:

$$e(x) = \sum_{i=1}^n v_i^2(x) = v^T(x)v(x)$$

El gradiente puede ser escrito entonces en forma matricial:

$$\nabla e(x) = 2J^T(x)v(x)$$

Donde $J(x)$ es la matriz Jacobiana. Ajustando el método de Newton, obtenemos el algoritmo de Levenberg Marquardt:

$$x_{k+1} = x_k - [J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k)v(x_k)$$

O lo que es lo mismo

$$\Delta x_k = - [J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k)v(x_k)$$

La constante μ determina la tendencia del algoritmo, cuando μ_k se incrementa, este algoritmo se aproxima al algoritmo de pasos descendientes para pasos de aprendizaje muy pequeños; cuando μ_k se decrecienta este algoritmo se convierte en el método de Gauss - Newton. El algoritmo comienza con un valor pequeño para μ_k , por lo general 0,01, si en ese paso no se alcanza el valor para $e(x)$ entonces el paso es repetido con μ_k multiplicado por un factor $\vartheta > 1$. Si se ha escogido un valor pequeño de paso en la dirección de paso descendiente, $e(x)$ debería decrecer. Si un paso produce un pequeño valor para $e(x)$, entonces el algoritmo tiende al método de Gauss - Newton, el que se supone garantiza una rápida convergencia. Este algoritmo genera un compromiso entre la velocidad del método de Gauss-Newton y la garantía de convergencia del método de paso descendiente.

3.2.2. Ajuste

Se considera la serie demanda de gasolina mensual a nivel país medida en miles de metros cúbicos (Mm³) desde enero de 1992 a diciembre de 2010.

Tomaremos como nodos de entrada los mismos rezagos que se usaron en la sección anterior y una capa oculta. Los distintos experimentos consisten en ir variando el número de nodos en la capa oculta. A medida que los nodos varían en cantidad las medidas de bondad del ajuste cambian también.

En la teoría no se muestra una forma del como obtener el número óptimo de nodos de la capa oculta, una regla a priori, que en experimentos habitualmente resultado de utilidad, es asumir que el valor inicial del número de nodos de la capa oculta sea igual al promedio entre el número de entradas y salidas, es decir: $(\# \text{ Entradas} + \# \text{ Salidas})/2$ (si el valor obtenido es decimal se redondea), en nuestro caso seis.

Las medidas de bondad de ajuste que usaremos serán R^2 , medida que se usa corrientemente en el ajuste de algún modelo, Error cuadrático medio (ECM), Error absoluto medio (EAM), Desviación estándar de error absoluto (DEEA).

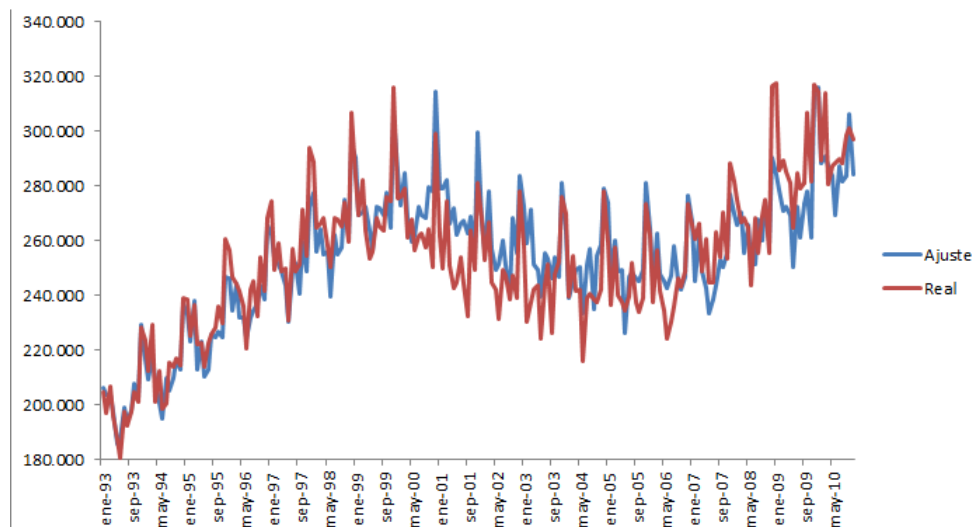


Figura 3.14: Ajuste de la red neuronal para los años 1993 a 2010

La tabla siguiente muestra los coeficientes mencionados para las experiencias con 6 a 2 nodos en la capa oculta.

Nº Nodos	R^2	ECM	EAM	DEEA
6	0,8383	12117,95	9493,67	7530,93
5	0,8427	10286,17	7897,97	6589,94
4	0,8257	11626,72	9228,49	7072,17
3	0,8565	10997,86	8701,05	6726,42
2	0,8369	11511,35	9177,02	6949,36

Medidas de bondad de ajuste para las redes neuronales con dos a seis nodos en la capa oculta.

El mejor ajuste se encuentra para la capa oculta con tres nodos.

3.2.3. Predicción

Una vez encontrada la red con el mejor ajuste procedemos a utilizarla para encontrar valores futuros. Haremos predicciones para los años 2011 y 2012.

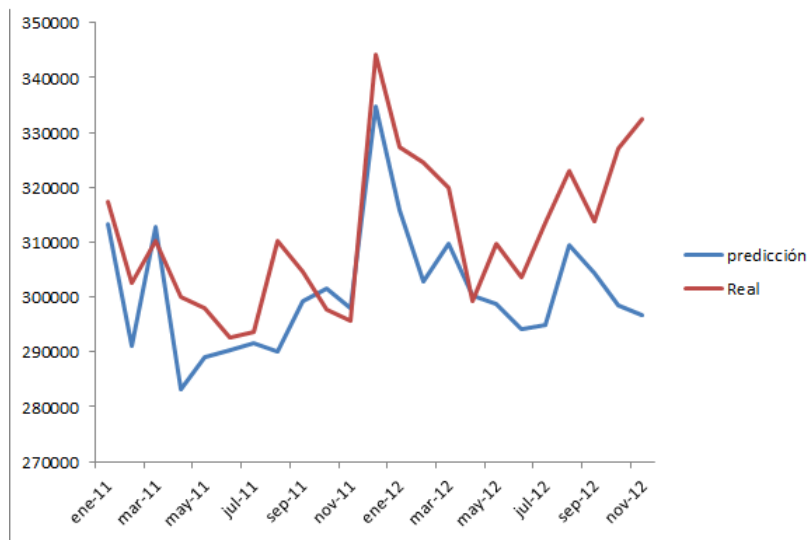


Figura 3.15: Predicciones para 2011 y 2012 para la red neuronal con 3 nodos en la capa oculta.

El error promedio para los dos años de predicción es de un 3,58%.

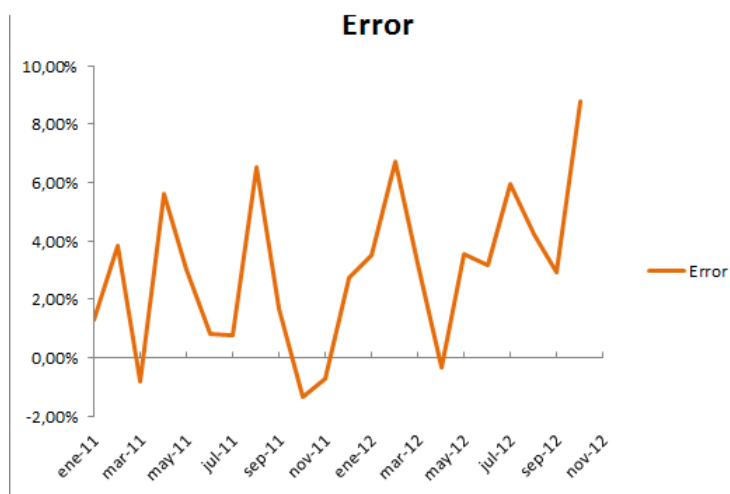


Figura 3.16: Error de la predicción de la red neuronal para los años 2011 y 2012.

4

CONCLUSIONES

En el capítulo 2 se intentó mejorar el modelo usado por Enap para la predicción de la demanda anual de gasolina. Al menos para el período de predicciones planteado el modelo propuesto funciona bastante bien y tiene mejores predicciones que el usado en la actualidad. Acá debemos considerar dos aspectos muy importantes, primero el hecho de que los coeficientes del modelo existente se reestimaron para el período 1986 a 2007, mismo período con el que se ajustó el modelo propuesto, esto pudo haber influenciado (y muy probablemente lo hizo) en las predicciones. El otro hecho que debemos considerar, y claramente el que más influye en la diferencia de la predicción de los dos modelos, es el del cambio de algunas variables por otras, que nutrió de más información al modelo, lo que lo hace mucho más preciso ante eventuales cambios por ejemplo en la tendencia.

En el capítulo 3 se presentan dos técnicas para modelar/ajustar la serie mensualizada de demanda de gasolina con el fin hacer predicciones.

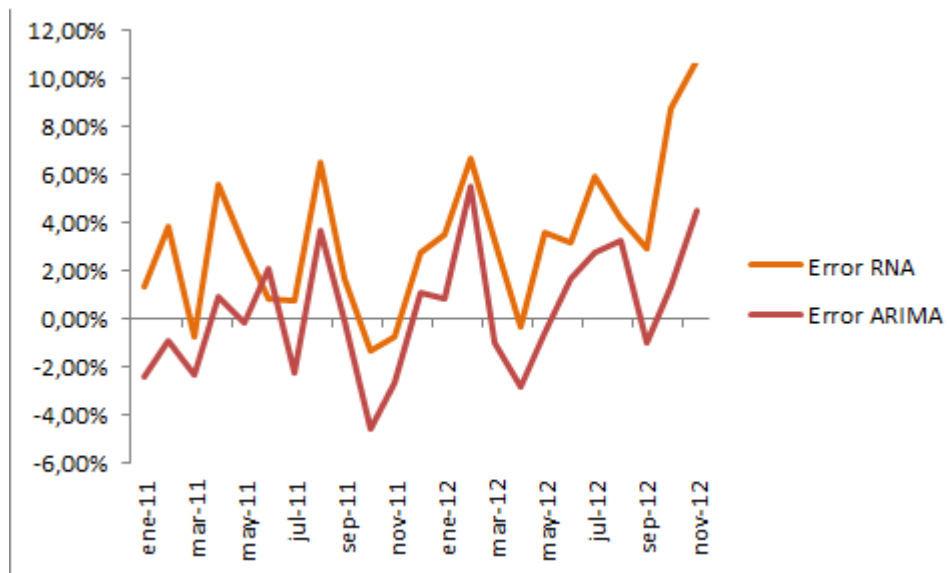


Figura 4.1: Comparación de los errores de predicción de los métodos ARIMA y redes neuronales para los años 2011 y 2012

Las redes neuronales tiene como una de sus limitaciones para alcanzar un uso generalizado el establecimiento de la arquitectura de la red. En este trabajo se uso una herramienta del software MatLab para realizar el ajuste sin inmiscuirnos demasiado en la topología de la red. El punto más álgido en el diseño del modelo de la red son las entradas o retrasos de la serie. En forma empírica, se llegó a la conclusión de que puede utilizarse la metodología ARIMA como una herramienta de preprocesamiento de datos, considerando como entradas los retrasos involucrados en el modelo proporcionado. En la serie estudiada, el modelo RNA proporcionó un mejor ajuste, mientras que la ARIMA con un ajuste similar, aunque algo inferior, logra mejores predicciones.

Este es un hecho que sorprende pues el modelo de RNA no pide linealidad y tiene restricciones menores al del modelo ARIMA, es decir, el modelo ARIMA es mucho más restrictivo que el de redes neuronales.

En ningún caso podemos decir que una técnica es mejor que otra, todo dependerá del tipo de dato y período estudiado, para este caso particular resultó ser una mejor herramienta el modelo ARIMA por sobre las RNA.

Algunas últimas consideraciones que tienen que ver con la experiencia:

- No siempre se tiene la libertad de implementar el modelo que uno estime, por ejemplo por desconocimiento de la contraparte en este sentido es importante poder persuadir/convencer.
- El manejo y utilización de las variables de buena forma es fundamental, por ejemplo a variable precio fue deflactada en pesos de 2000 utilizando IPC, usarla con valores corrientes hubiese sido un error ya que los valores así usados no son comparable.
- La forma de obtener las variables debe ser simple, dentro de lo posible, o al menos los valores comprobable de forma práctica, la variable vehículos nuevos es obtenida con información de las bases de datos que entrega la aduana, teniendo que filtrar la información, lo que conlleva mucha trabajo, la variable propuesta (parque de vehículos a pasajeros más comerciales) se obtiene directamente de información que tiene el INE por lo que es fácil de obtener, además de confiable.
- La forma de elegir un modelo SARIMA sobre otro se basó en la comparación de predicciones para los modelos propuestos, más que en lo que indica la teoría.
- En ningun caso podemos decir que una técnica es mejor que otra, a mi juicio, todo depende del tipo de dato y período estudiado, para este caso en particular resultó ser una mejor herramienta el modelo SARIMA por sobre las RNA.

Bibliografía

- [1] Bishop Christopher, Neural Networks for Patter Recognition, Oxford University Press, 1997.
- [2] Chumacero Rómulo, Modelo de Proyección del Parque Vehicular, 2007.
- [3] Collantes Joanna, Predicción con redes neuronales: Comparación con la metodología de Box y Jenkins.
- [4] Durbin James - Koopman Siem Jan, Time Series Analysis by State Space Methods, T.J. International, 2001.
- [5] Esteban M^a Victoria, Análisis Económico, Departamento de Econometría Aplicada III, Universidad del País Vasco/Euskal Arriko Unibertsitatea, [<http://www.sarriko-online.com/cas/fichas/2008/04-08.pdf>], 2008.
- [6] Faraway Julian, Time series forecasting with neural networks: a comparative study using airline data, Applied Statistic 47(2), 1998.
- [7] Freeman James A. - Skapura David M., Redes Neuronales, Addison-Wesley Iberoamericana, 1993.
- [8] Gujarati Damodar N., Econometría, cuarta edición española, McGraw-Hill, 2004.
- [9] O’Ryan Raúl , Diseño de un Modelo de Proyección de Demanda Energética Global de Largo Plazo, Ministerio de Energía, Informe preparado para la Comisión Nacional de Energía, 2008.
- [10] Pérez-Chavarría Miguel, Predicción de series de tiempo aplicando redes neuronales artificiales, Ciencias Marinas, 2002.
- [11] Rumelhart D.E., Hinton G.E. y Williams R.J., Learning internal representation by error propagation, MIT Press, Cambridge, 1986.