

67  
254



**UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO**

**FACULTAD DE CIENCIAS**

**UN MODELO PARA PRONOSTICAR LA  
DEMANDA DE PETROLIFEROS EN MEXICO**

**T E S I S**

QUE PARA OBTENER EL TITULO DE:

**A C T U A R I A**

**P R E S E N T A :**

**MARIA DEL CARMEN MORENO RAMIREZ**



**TESIS CON  
FALLA DE ORIGEN**

DIRECTOR DE TESIS: **MAT. GAUDENCIO HERNANDEZ HERNANDEZ**



**1996**

**FACULTAD DE CIENCIAS  
SECCION ESCOLAR**

**TESIS CON  
FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

M. en C. Virginia Abrín Batule  
Jefe de la División de Estudios Profesionales de la  
Facultad de Ciencias  
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:  
UN MODELO PARA PRONOSTICAR LA DEMANDA DE PETROLIFEROS EN MEXICO

realizado por MARIA DEL CARMEN MORENO RAMIREZ

con número de cuenta 8116860-8 , pasante de la carrera de ACTUARIA

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis MAT. GAUDENCIO HERNANDEZ HERNANDEZ  
Propietario  
Propietario MAT. MARGARTA CHAVEZ CANO  
Propietario DR. JOSE R. MENDOZA BLANCO  
Suplente ACT. DAVID GABRIEL LOPEZ SERVIN  
Suplente MAT. LAURA PASTRANA RAMIREZ

*[Handwritten signature]*  
*[Handwritten signature]*

*[Handwritten signature]*

Laura Pastrana R.

*[Handwritten signature]*  
Consejo Departamental de Matemáticas  
ACT. CLAUDIA CARRILLO QUITOZ  
FACULTAD DE CIENCIAS  
CONSEJO DEPARTAMENTAL  
DE  
MATEMÁTICAS

**UN MODELO PARA PRONOSTICAR  
LA DEMANDA DE PETROLÍFEROS  
EN MÉXICO**

**María del Carmen Moreno Ramírez**

A mi

**Mamá**

# Contenido

<b>Introducción</b> .....	<b>1</b>
<b>Capítulo 1</b>	
<b>El modelo de regresión lineal</b> .....	<b>6</b>
1.1 Diferentes formas de relaciones funcionales .....	6
1.2 El modelo de regresión lineal simple .....	9
1.2.1 Naturaleza probabilística del modelo de regresión .....	10
1.2.2 Estimación de parámetros .....	11
1.2.3 Coeficiente de correlación .....	13
1.2.4 Precisión de la regresión estimada .....	15
1.2.5 Pruebas de hipótesis e intervalos de confianza .....	17
1.3 Pronósticos con el modelo de regresión simple .....	19
1.4 El modelo de regresión lineal múltiple .....	20
1.4.1 Estimación de los parámetros .....	23
1.4.2 Matriz de varianza-covarianza de $\beta$ .....	24
1.4.3 Coeficiente de determinación .....	26
1.4.4 Pruebas de hipótesis e intervalos de confianza .	27
1.5 Pronóstico con el modelo de regresión lineal múltiple .	28
1.6 Violación de los supuestos del modelo de regresión ...	29
1.6.1 Heterocedasticidad .....	30

1.6.2	Autocorrelación .....	31
1.6.3	Multicolinealidad .....	32
1.6.4	Especificación del modelo .....	33

## **C a p í t u l o 2**

<b>Modelos de series de tiempo .....</b>	<b>35</b>
2.1 Definición .....	35
2.2 Naturaleza probabilística de las series de tiempo ....	36
2.3 Técnicas para pronosticar series de tiempo .....	37
2.4 Algunos elementos para realizar el análisis de series de tiempo .....	40
2.4.1 Estabilización de la varianza .....	41
2.4.2 Operadores de retraso y de diferencia .....	43
2.4.3 Autocorrelación .....	44
2.4.3.1 Estacionariedad .....	45
2.4.3.2 Estacionalidad .....	46
2.4.3.3 La prueba $\chi^2$ .....	47
2.4.4 La función de autocorrelación parcial .....	48
2.5 El método de Box-Jenkins .....	49
2.5.1 Modelos autorregresivos .....	50
2.5.2 Modelos de promedios móviles .....	52
2.5.3 Modelos autorregresivos y de promedios móviles, ARMA .....	54
2.5.4 Modelos autorregresivos e integrados de promedios móviles, ARIMA .....	55
2.5.5 La construcción de los modelos .....	56
2.5.5.1 Identificación .....	56

2.5.5.2	Estimación .....	57
2.5.5.3	Verificación .....	58
2.5.6	Pronóstico .....	61
2.5.7	Modelos estacionales .....	64
2.6	Modelos de función de transferencia .....	66
2.6.1	Elementos para la construcción de los modelos de función de transferencia .....	68
2.6.1.1	Función de transferencia .....	68
2.6.1.2	Término estocástico de error .....	69
2.6.1.3	Función de correlación cruzada .....	70
2.6.1.4	Función de impulso-respuesta .....	72
2.6.2	Construcción de modelos de función de transferencia .....	75
2.6.2.1	Identificación .....	76
2.6.2.2	Estimación .....	77
2.6.2.3	Verificación .....	78
2.6.2.4	Uso del modelo .....	78
2.7	Análisis de intervención .....	78
2.7.1	Identificación .....	83
2.7.2	Estimación .....	85
2.7.3	Verificación .....	85
2.7.4	Uso del modelo .....	86

### **Capítulo 3**

<b>Pronósticos ARIMA condicionales .....</b>	<b>87</b>
3.1 La teoría de la combinación de pronósticos .....	87
3.2 El método de pronósticos condicionales .....	89

3.3	Derivación de los pronósticos condicionales óptimos ..	90
3.4	Una prueba estadística para la hipótesis $Y = CZ_f$ .....	93
3.4.1	Intervalos de confianza para los pronósticos condicionales .....	95
3.4.2	Diferencia entre los pronósticos ARIMA y los pronósticos condicionales .....	96
3.5	Pronósticos condicionales cuando la información adicional es una conjetura .....	97

#### **C a p í t u l o 4**

	<b>Pronóstico de la demanda de petrolíferos .....</b>	<b>99</b>
4.1	La importancia de los petrolíferos .....	99
4.2	Pronósticos anuales .....	102
4.2.1	Gasolinas automotrices .....	103
4.2.2	Diesel .....	109
4.2.3	Gas licuado .....	111
4.2.4	Combustóleo .....	114
4.2.5	Gas natural .....	116
4.2.6	Los pronósticos anuales .....	118
4.3	Pronósticos mensuales .....	120
4.3.1	Gasolinas .....	120
4.3.2	Diesel .....	132
4.3.3	Gas licuado .....	134
4.3.4	Combustóleo .....	136
4.3.5	Gas natural .....	138
4.3.6	Los pronósticos mensuales .....	140
4.4	Pronósticos condicionales para 1995 .....	141

4.4.1	Resultados .....	142
4.4.2	Evaluación de los pronósticos .....	144
4.5	Pronósticos para 1996 .....	151
4.5.1	Pronósticos anuales .....	151
4.5.2	Pronósticos mensuales .....	152
4.5.3	Pronósticos condicionales .....	154
4.5.4	Modelo ARIMA con intervenciones para el consumo de las gasolinas .....	159
4.5.5	Resultados .....	166
<b>Conclusiones .....</b>		<b>169</b>
<b>Bibliografía .....</b>		<b>171</b>

## Introducción

El presente trabajo surge de la inquietud por mostrar a los estudiantes de la carrera de Actuaría un caso práctico del manejo de algunas herramientas estadísticas en la obtención de pronósticos de variables económicas, específicamente, la demanda de petrolíferos en México. Dichas herramientas son el análisis de regresión, tema que se plantea comúnmente en las asignaturas de estadística que se imparten en la Facultad de Ciencias, y el análisis de series de tiempo, de gran importancia por su aplicación en la economía y en las finanzas, pero que no se ha incluido en los programas de estadística de licenciatura, salvo esporádicamente en algún seminario. Este material generalmente se expone en cursos de posgrado; sin embargo, ello no implica que un estudiante de licenciatura no pueda abordar el tema con los conocimientos estadísticos que posee.

El objetivo de esta tesis es generar pronósticos de la demanda interna de petrolíferos en México empleando tanto estrategias, como conjuntos de información distintos, que al combinarse de cierta manera producen pronósticos más precisos. Las estrategias mencionadas se refieren a tres métodos estadísticos, dos de ellos que utilizan grupos de información con diferentes características: el análisis de regresión y el análisis de series de tiempo; y uno más que, dentro de la teoría de combinación de pronósticos, explota los dos conjuntos de información que utilizaron los métodos anteriores: este método es el de pronósticos ARIMA condicionales que propuso el Dr. Víctor M. Guerrero en 1989.

La importancia de generar pronósticos de la demanda de petrolíferos se justifica a continuación:

La industria petrolera en México aporta elevados rendimientos a la economía logrando una estrecha vinculación con un gran número de cadenas productivas industriales al suministrar una amplia variedad de productos que abastecen de energía y materias primas. De hecho, los hidrocarburos representan el principal energético al participar con el 90.3% en la producción de energía primaria de acuerdo con lo reportado en el Balance Nacional de Energía de 1994. La estructura por energético del consumo final nacional indica la preponderancia de los petrolíferos y el gas: el consumo de gasolina representa el 27.1%, sigue el gas natural con 17.3%, el diesel con 13.4%, el gas licuado con 10.8%, la electricidad con 10.4%, el combustóleo con 7.7%, la leña con 6.5% y el 6.8% restante proviene de otros combustibles. Lo anterior sin considerar los petrolíferos que se utilizan en la generación de electricidad: combustóleo, gas natural y diesel.

Por sectores, el industrial demandó 313.4 petacalorías de energía en 1994, de las que el 25.8% provino directamente de los petrolíferos, 47.0% del gas, 16.2 de la electricidad y el resto, 11.0%, de otros combustibles como el carbón, la leña y el bagazo de caña. La demanda del sector transporte ascendió a 367.2 petacalorías, de las que el 99.8% se cubrió con petrolíferos. Y el sector residencial, comercial y público requirió 204.0 petacalorías, cubriéndose el 48.33% con petrolíferos, 4.7% con gas natural, 18.1% con electricidad y 28.9% con leña.

Así, la importancia de los energéticos convierten a la industria energética, en particular a la petrolera, en prioritaria y estratégica para el Estado y de ahí que, hasta la fecha, éste mantenga la propiedad de la entidad encargada de su producción que es Petróleos Mexicanos (Pemex). Consecuentemente, los hidrocarburos representan un renglón muy importante dentro de los bienes que produce el sector público y su contribución a las finanzas públicas es fundamental, lo que a su vez convierte a la política de precios de los petrolíferos en un factor determinante dentro del conjunto de medidas disponibles para regular los ingresos públicos.

De esta manera, para el Gobierno Federal resulta imprescindible contar con pronósticos de la demanda de petrolíferos y gas natural que permitan, por un lado, conocer la sensibilidad de los ingresos provenientes de las ventas internas de petrolíferos ante diversas políticas, tanto en el corto, como en el mediano plazo; y por otro, hacer un uso óptimo de la información disponible.<sup>1/</sup>

En este contexto, con objeto de contar con todas las herramientas estadísticas que permitan generar los pronósticos del consumo de petrolíferos, en esta tesis se presenta la teoría básica que sustenta al análisis de regresión y al de series de tiempo. Una característica de esta herramienta es que con el análisis de regresión múltiple los datos disponibles se ajustan a un modelo establecido de antemano, que en economía está respaldado por alguna teoría, mientras que en el análisis de series de tiempo los datos indican la elección del modelo que represente adecuadamente su comportamiento. Asimismo, se presenta el método recursivo, pronósticos ARIMA condicionales, para generar pronósticos óptimos de series de tiempo que incorporan información sobre la probable evolución futura de la variable en estudio.

Mediante el análisis de regresión lineal múltiple, se estiman las ecuaciones de demanda de los petrolíferos utilizando el promedio anual de su consumo de 1970 a 1994 para generar pronósticos anuales que reflejen, principalmente, el comportamiento de la actividad económica, la sustitución entre productos y el impacto de la política de precios. En los ejercicios se consideran ecuaciones sencillas que dejan a un lado factores estructurales como pueden ser el ritmo y patrón de urbanización, las opciones de transporte, el tipo de generación de electricidad y los requerimientos específicos de los

---

<sup>1/</sup> Los productos petrolíferos se obtienen del procesamiento de crudo, condensados y líquidos del gas natural y comprenden gas licuado, pentanos, gasaviones, gasolinas, turbotina, querosenos, diesel, combustóleo, lubricantes, grasas, parafinas, asfalto, coque de petróleo, solventes y otros productos. Por su parte, el gas natural es la mezcla gaseosa de hidrocarburos e impurezas que se extrae de los yacimientos. Teniendo en cuenta estas diferencias, y aun cuando el gas natural no es un petrolífero, en este trabajo las referencias a petrolíferos incluyen también al gas natural.

sectores de la economía. No obstante, los ejercicios consideran las variables más significativas y muestran la aplicación del análisis de regresión lineal múltiple.

Con el análisis de series de tiempo se tienen elementos para identificar la estructura estocástica que genera las series mensuales del consumo de petrolíferos y se construyen los modelos ARIMA univariados con la estrategia que proponen Box y Jenkins (1978), para obtener los pronósticos mensuales basados en el comportamiento histórico de la serie. Debido a las características de algunas series que se utilizaron en este trabajo, y para generar pronósticos más precisos, se consideró conveniente presentar el análisis de intervención, como un caso particular de la teoría de funciones de transferencia.

La presentación del método de pronósticos ARIMA condicionales se enmarca en la combinación de pronósticos, misma que, como se ha probado en diversos estudios, es una estrategia que mejora la calidad de los pronósticos. La combinación de pronósticos surge cuando se obtienen pronósticos de una misma variable con métodos distintos, y entonces la aplicación de criterios para escoger el mejor implica un desperdicio de información útil que se refleja en los pronósticos rechazados. El método de pronósticos ARIMA condicionales utiliza combinaciones lineales para generar pronósticos óptimos basados en información histórica y adicional sobre el futuro, siendo esta última una restricción a la que el pronóstico debe ceñirse exactamente, o bien, es sólo una conjetura acerca de los valores futuros de la serie o el pronóstico de un modelo alternativo. En el ejercicio, los pronósticos mensuales ARIMA que se obtienen para cada producto, se modifican para alcanzar el pronóstico del volumen anual que se obtiene a través de las ecuaciones de demanda estimadas.

El ejercicio de aplicación que se realiza se refiere a la obtención de pronósticos mensuales del consumo de cinco productos: gasolinas, diesel, gas licuado, combustóleo y gas natural. La información disponible es la siguiente: series anuales de 1970 a 1995 del promedio de consumo de cada producto, así como de las variables

macroeconómicas relativas a la actividad económica (Producto Interno Bruto), a la inflación (Índice Nacional de Precios al Consumidor), y al sector externo (tipo de cambio y precio del crudo mexicano de exportación), y de la variable de política representada en este caso por los precios aplicados a la venta al consumidor final de los productos; y series mensuales de enero de 1986 a diciembre de 1995 del promedio mensual de consumo de los petrolíferos. Los datos con periodicidad anual se explotan a través del análisis de regresión lineal múltiple al estimar ecuaciones de demanda de cada producto; por su parte, los datos mensuales se procesan para elegir los modelos de series de tiempo ARIMA univariados más adecuados para representar el comportamiento de las series, y en el caso de las gasolinas fue necesario el empleo del análisis de intervención. Finalmente, se obtienen los pronósticos ARIMA condicionales, los cuales conjuntan toda la información disponible. Es importante señalar que, en general, en este trabajo se han utilizado estadísticas oficiales de carácter público.

El trabajo se estructuró de la siguiente manera: en el primer capítulo se presentan los fundamentos del análisis de regresión simple y múltiple; el segundo capítulo muestra el análisis de series de tiempo univariado, la teoría de las funciones de transferencia y el análisis de intervención; en el tercer capítulo se presenta la teoría que sustenta el método de pronósticos ARIMA condicionales; en el cuarto capítulo se expone el ejercicio integral para generar los pronósticos mensuales de la demanda de petrolíferos, comparando los resultados obtenidos con lo observado en 1995 y mostrando las proyecciones para 1996 de acuerdo con el programa económico oficial para este año y, finalmente, se presentan las conclusiones.

# Capítulo 1

## El modelo de regresión lineal

Las técnicas de regresión lineal en general se refieren tanto a los aspectos de causalidad, como de explicación para poder realizar pronósticos. Con estas técnicas se intenta predecir el comportamiento de alguna variable midiendo algunos factores independientes que la impactan. El procedimiento para pronosticar por medio de regresión difiere del de series de tiempo en que una vez que se han establecido y medido las relaciones de interés, éstas se utilizan para obtener los pronósticos.

En esta sección se presentan los fundamentos teóricos del modelo de regresión lineal, en donde la variable que se desea pronosticar depende de una o más variables exógenas. Esta técnica se utilizará para generar los pronósticos anuales de la demanda de los petrolíferos, en donde las variables explicativas serán variables macroeconómicas y/o de política.

### 1.1 Diferentes formas de relaciones funcionales.

El fenómeno de que una variable cambie a través del tiempo puede expresarse de la siguiente manera:

$$Y = f(\text{tiempo})$$

Esta expresión indica que el nivel de  $Y$  cambia en el tiempo y que los valores futuros de esta variable pueden pronosticarse al tener identificada esta relación.

Una manera funcional de relacionar a dos variables no necesariamente debe de incluir al tiempo. Ésta puede generalizarse para incluir otras variables en una relación explicatoria causal:

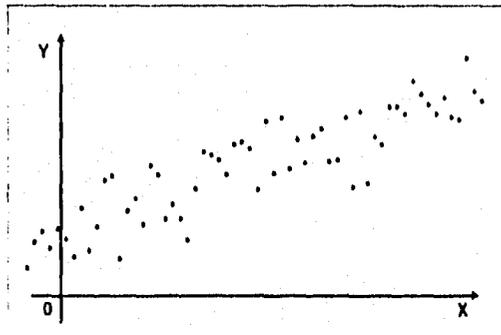
$$Y = f(X_1, X_2, \dots, X_k)$$

en donde  $Y$  es la variable dependiente y  $X_1, X_2, \dots, X_k$  son las variables independientes o exógenas.

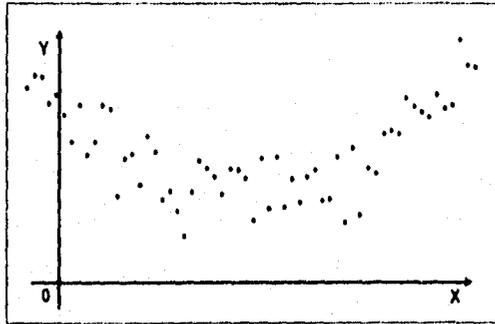
El objetivo del pronóstico causal es poder predecir los valores de la variable dependiente si se descubre la manera en que ésta depende de una o más variables independientes. Cuando se tiene sólo una variable independiente se usa el método de regresión lineal simple, y si se involucra más de una variable independiente se usa regresión lineal múltiple.

Desde que se supone que existe una relación funcional entre dos variables, una dependiente y otra independiente, es necesario determinar si esta relación es lineal, exponencial, cuadrática, cúbica, etc., y estimar los parámetros de la relación. Por ejemplo, la inspección visual de las gráficas puede dar una idea aproximada de la manera en que se relacionan  $X$  y  $Y$ , como pueden ser:

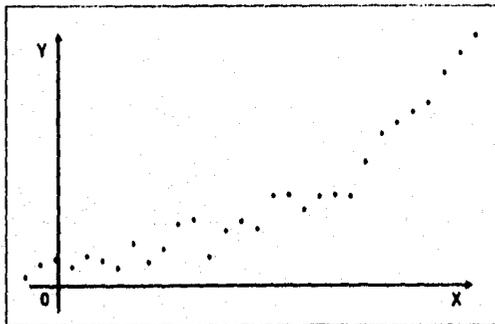
a) relación lineal,



## b) relación cuadrática



## c) relación exponencial

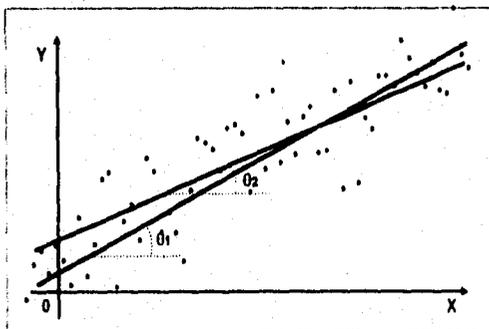


El hecho de que la relación entre las variables  $X$  y  $Y$  no sea lineal no es una limitante ya que muchas relaciones no lineales pueden convertirse en funciones lineales mediante la transformación de los datos. Entonces se puede concluir que basta con encontrar una transformación de las variables que lleve la función no-lineal a una función lineal, para poder trabajar con el modelo de regresión lineal que es de la firma  $Y = \beta_0 + \beta_1 X$ .

Por otra parte, es necesario que la variable independiente  $Y$  sea una función lineal de los parámetros  $\beta_0$  y  $\beta_1$ , y este concepto es el relevante para el análisis de regresión lineal.

### 1.2 El modelo de regresión lineal simple.

Es posible que los primeros modelos de regresión se hayan construido a partir de la simple observación de dos series de datos; por ejemplo, la evolución de las series  $\{X_i\}$  y  $\{Y_i\}$  de la Gráfica 1.2.1 puede describirse con un modelo de regresión por medio de una regla y un transportador, es decir, se puede trazar una línea que pase lo más cerca posible de todos los puntos con un ángulo de inclinación  $\theta_1$ , pero también se puede sugerir otra línea con ángulo de inclinación  $\theta_2$ , que aparentemente pase "igual de cerca de todas las observaciones", por lo que, a simple vista, no se puede escoger la mejor línea.



El método gráfico para obtener regresiones tiene muchas limitaciones, ya que en general se pueden tener más de una variable explicativa en el modelo, y entonces se complica la construcción de los diagramas. Sin embargo, siempre es conveniente elaborar gráficas con el objeto de detectar posibles tendencias, relaciones funcionales o irregularidades en la información para incorporarlas en la especificación del modelo.

### 1.2.1 Naturaleza probabilística del modelo de regresión.

Partiendo del hecho de que para un valor observado de  $X$  se pueden observar muchos valores posibles de  $Y$ , a menos que se tenga conocimiento adicional, se supone que para cada observación de  $X$  las observaciones de  $Y$  diferirán de una manera aleatoria. Esta situación se describe formalmente al sumar un componente estocástico de error al modelo de regresión:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad i = 1, 2, \dots, n$$

en donde para cada observación,  $Y_i$  es una variable aleatoria,  $X_i$  es fijo y  $u_i$  es el término estocástico de error cuyos valores están basados en una distribución probabilística.

El término estocástico de error puede crecer con el juego de diversos factores. Los errores aparecen a causa de que el modelo es una simplificación de la realidad, de la omisión de otras variables explicativas o de la recolección misma de los datos.

Por lo anterior, el modelo de regresión lineal simple se postula bajo los siguientes supuestos:

- 1) La relación entre  $Y$  y  $X$  es lineal en los parámetros.

- 2) La variable  $X$  es no estocástica, siendo sus valores fijos y con varianza distinta de cero para cualquier tamaño de muestra ( $n$ ).
- 3) Las  $u_i$  son variables aleatorias independientes idénticamente distribuidas como una normal con media cero y varianza constante, es decir:

$$a) E[u_i] = 0 \quad i = 1, 2, \dots, n$$

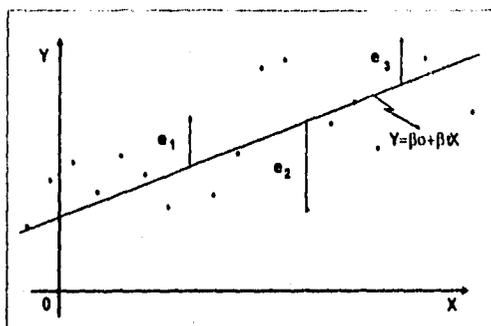
$$b) \text{Var}[u_i] = E[u_i^2] = \sigma^2$$

$$c) E[u_i, u_j] = 0 \quad \forall i \neq j$$

- 4) El modelo de regresión está especificado correctamente, es decir que no existen sesgos ni errores en la especificación (selección de las variables y formas funcionales).

### 1.2.2 Estimación de parámetros

Una vez propuesto el modelo de regresión lineal simple, el paso siguiente es estimar los parámetros del modelo  $Y = \beta_0 + \beta_1 X + u$  y utilizarlos para explicar la posible relación lineal entre las variables  $X$  y  $Y$  para pronosticar los valores futuros de la variable dependiente  $Y$ . Un método que se usa para determinar el valor de los parámetros  $\beta_0$  y  $\beta_1$  es el de mínimos cuadrados ordinarios; éste se basa en minimizar la suma de los cuadrados de la distancia vertical  $e_i$  entre las observaciones reales  $Y_i$  y los valores que están sobre la recta de regresión  $\hat{Y}_i$ .



Cada desviación o residual se calcula como  $e_i = Y_i - \hat{Y}_i$  y cada valor estimado, que se encuentra en la línea de regresión, se calcula como  $\hat{Y}_i = \beta_0 + \beta_1 X_i$ . Entonces la función a minimizar es:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1.2.2.1)$$

para lo cual se sustituye el valor  $\hat{Y}_i$  y se tiene:<sup>1/</sup>

$$\sum e_i^2 = \sum (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (1.2.2.2)$$

Se calculan las derivadas parciales con respecto a  $\beta_0$  y  $\beta_1$ :

$$\frac{\partial \sum e_i^2}{\partial \beta_0} = -2 \sum (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial \sum e_i^2}{\partial \beta_1} = -2 \sum X_i (Y_i - \beta_0 - \beta_1 X_i)$$

<sup>1/</sup> En adelante en este capítulo, las sumas que se indiquen son de  $i=1$  hasta  $n$ , a menos que se especifique otro rango.

e igualando a cero se tienen las ecuaciones normales

$$-\sum Y_i + n\beta_0 + \beta_1 \sum X_i = 0$$

$$-\sum X_i Y_i + \beta_0 \sum X_i + \beta_1 \sum X_i^2 = 0 \quad (1.2.2.3)$$

que resolviendo simultáneamente llevan a obtener los parámetros  $\beta_0$  y  $\beta_1$  dados por:

$$\beta_0 = \sum \frac{Y_i}{n} - \beta_1 \sum \frac{X_i}{n} \quad (1.2.2.4)$$

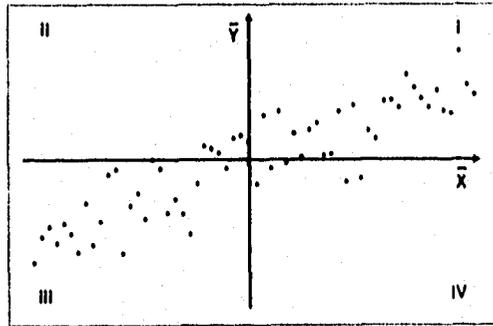
$$\beta_1 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} \quad (1.2.2.5)$$

Los parámetros estimados resultan de una muestra específica de observaciones de las variables dependiente e independiente. Como la muestra puede variar, las estimaciones también, por lo que pueden asociarse con una variable aleatoria. Ya que el modelo es estocástico, se denota a los parámetros estimados por  $\hat{\beta}_0$  y  $\hat{\beta}_1$  para referirse a ellos como una estimación que resulta de una muestra aleatoria y también como el estimador que sigue una distribución de probabilidad, y se puede demostrar que éstos son los mejores estimadores lineales, insesgados (MELI, Gujarati, 1992).

### 1.2.3 Coeficiente de correlación

Al intentar explicar una variable en función de otra es conveniente medir el grado en que ambas variables se encuentran relacionadas linealmente. Una visión más clara se puede tener usando el diagrama de dispersión, en donde se tengan identificadas las rectas  $y = \bar{Y}$  y  $x = \bar{X}$ , que dividen el plano en cuatro cuadrantes con centro en el punto  $(\bar{X}, \bar{Y})$ . Entonces la

cantidad  $\sum(X_i - \bar{X})(Y_i - \bar{Y})$  sirve como una medida de asociación entre  $X$  y  $Y$ . Si la asociación es positiva, la mayoría de los puntos caerán en los cuadrantes I y III y  $\sum(X_i - \bar{X})(Y_i - \bar{Y})$  es positiva. Si la asociación es negativa, la mayoría de los puntos estarán en los cuadrantes II y IV, siendo la suma anterior negativa. Si no existe relación lineal entre  $X$  y  $Y$ , la suma será muy pequeña.



La medida que se utilice para obtener el grado de asociación lineal debe superar dos inconvenientes:

- i) ser una relación independiente de la elección del origen de las variables, lo cual se puede resolver al definir  $x_i = X_i - \bar{X}$  y  $y_i = Y_i - \bar{Y}$ ; y
- ii) una medida independiente de la escala de medición empleada para  $X$  y  $Y$ , es decir, que ambas estén expresadas en las mismas unidades.

De esta manera, el coeficiente de correlación que supera los inconvenientes mencionados está dado por:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{N\sigma_X\sigma_Y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}} \quad (1.2.3.1)$$

El valor de  $r$  puede variar entre 1 y -1. Cuando el valor absoluto es 1, significa que la correlación entre las dos variables es perfecta, y cuando su valor es cero, implica que no existe asociación lineal.

Por otra parte, se puede demostrar que el cuadrado del coeficiente de correlación es la razón entre la variación explicada por la recta de regresión y la variación total de los datos:

$$r^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad (1.2.3.2)$$

es decir, indica el porcentaje de variación total que está siendo explicada por el modelo.

#### 1.2.4 Precisión de la regresión estimada

La precisión con que la línea de regresión estimada se ajusta a los datos se mide de la siguiente manera:

Considérese la identidad:

$$Y_i - \hat{Y}_i = Y_i - \bar{Y} - (Y_i - \bar{Y}) \quad (1.2.4.1)$$

elevando al cuadrado ambos lados y sumando sobre todas las  $i$ 's se tiene:

$$\begin{aligned}\sum(Y_i - \hat{Y})^2 &= \sum((Y_i - \bar{Y}) - (\hat{Y}_i - \bar{Y}))^2 \\ &= \sum(Y_i - \bar{Y})^2 + \sum(\hat{Y}_i - \bar{Y})^2 - 2\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})\end{aligned}$$

y haciendo operaciones se llega a que:

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 - \sum(Y_i - \hat{Y})^2 \quad (1.2.4.2)$$

en donde  $Y_i - \bar{Y}$  es la desviación de la  $i$ -ésima observación con respecto a la media,  $\hat{Y}_i - \bar{Y}$  es la desviación del valor pronosticado respecto a la media y  $Y_i - \hat{Y}$  es la desviación de la  $i$ -ésima observación con relación al valor ajustado (residual). Entonces se puede decir que:

$$\begin{array}{ccc} \text{Suma de cuadrados} & = & \text{Suma de cuadrados} \\ \text{total} & & \text{debida a la regresión} \\ & & \text{Suma de cuadrados} \\ & & \text{residual} \end{array}$$

Esto significa que de la variación de las  $Y$ 's alrededor de su media, una parte se atribuye a la línea de regresión y otra al hecho de que no todas las observaciones están sobre la línea de regresión.

De aquí se puede definir el coeficiente de determinación que mide la proporción de la variación total de  $Y$  que se explica por el modelo de regresión:

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \quad (1.2.4.3)$$

el cual corresponde exactamente al cuadrado del coeficiente de correlación de la expresión 1.2.3.2 (Gujarati 1992); y también, se puede expresar como:

$$r^2 = \beta_1^2 \left( \frac{\sum x_i^2}{\sum y_i^2} \right) = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2} \quad (1.2.4.4)$$

en donde  $x_i = X_i - \bar{X}$  y  $y_i = Y_i - \bar{Y}$

Este coeficiente será siempre positivo y su valor estará en el intervalo  $[0,1]$ . Cuando el valor de  $r^2$  es 1, se interpreta como un ajuste perfecto, y si es cero, entonces no hay relación lineal alguna entre las variables exógena y endógena.

### 1.2.5 Pruebas de hipótesis e intervalos de confianza

La ecuación de regresión estimada es  $\hat{Y} = \beta_0 + \beta_1 X$  en donde  $\beta_0$  y  $\beta_1$  son los parámetros estimados a partir de los datos de una muestra. Si se utilizara una muestra diferente de la misma población, se obtendrían valores de  $\beta_0$  y  $\beta_1$  diferentes.

Para el caso de  $\beta_1$  puede plantearse la hipótesis de que  $\beta_1 = 0$ . Si ésta es verdadera, se tendría que  $\hat{Y} = \beta_0 + 0X$ , lo que implica que la variable  $X$  no tiene ninguna influencia sobre la variable  $Y$  (es decir, que la variable  $X$  no explica a  $Y$ ).

La prueba estadística para no rechazar la hipótesis nula asociada con un coeficiente de regresión está basada en la distribución  $t$ -student. Para utilizar esta distribución y construir intervalos al  $100(1-\alpha)\%$  de confianza,<sup>2/</sup> es necesario estandarizar el parámetro de regresión, restándole su valor real y dividiendo entre su desviación estándar:

$$t = \frac{\hat{\beta}_1 - \beta_0}{\sigma_{\beta_1}} = \frac{\hat{\beta}_1 - \beta_0}{\hat{\sigma} \sqrt{\frac{1}{\sum(X-X)^2}}} \sim t_{n-2} \quad (1.2.5.1)$$

en donde  $\hat{\sigma}$  es la varianza del modelo y está dada por:

$$\hat{\sigma} = \sqrt{\frac{\sum(Y_i - \hat{Y})^2}{n-2}} \quad (1.2.5.2)$$

<sup>2/</sup>  $\alpha$  es el nivel de significancia.

El valor  $t$  se compara con un valor de tablas  $t$ -student con  $n-2$  grados de libertad al nivel de significancia  $\alpha$ . Así, la hipótesis a probar es  $H_0: \beta_1 = 0$  vs  $H_1: \beta_1 \neq 0$ ; y si  $|t| < t_{\alpha}$ , entonces no se rechaza la hipótesis nula  $H_0$ , es decir, se puede considerar que  $\beta_1 = 0$ .

De aquí se puede construir un intervalo de confianza al  $100(1-\alpha)\%$  de confianza para  $\beta_1$ :

$$\hat{\beta}_1 \pm t_{n-2} \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}} \quad (1.2.5.3)$$

De la misma manera, para  $\beta_0$  se tendrá que:

$$t = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}} \frac{\sqrt{N \sum (X - \bar{X})^2}}{\sqrt{\sum X^2}} \sim t_{n-2} \quad (1.2.5.4)$$

y un intervalo de confianza es:

$$\hat{\beta}_0 \pm t_{n-2} \frac{\hat{\sigma} \sqrt{\sum X_i^2}}{\sqrt{N \sum (X_i - \bar{X})^2}} \quad (1.2.5.5)$$

Otra manera alterna para probar que  $\hat{\beta}_1$  es significativamente diferente de cero se asocia a las sumas de cuadrados de los residuales con sus grados de libertad. Para ello se construye una tabla de análisis de varianza de la siguiente manera.

Fuente de la variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Estadística F
Debido a la regresión	$\sum y_i^2 - \beta_1^2 \sum x_i^2$	1	$\beta_1^2 \sum x_i^2$	$\frac{\beta_1 \sum x_i^2}{\frac{\sum e_i^2}{n-2}}$
Debido a los errores	$\sum e_i^2$	n-2	$\frac{\sum e_i^2}{n-2} = \hat{\sigma}^2$	
Total	$\sum y_i^2$	n-1		

La estadística F proporciona una prueba de la hipótesis  $H_0: \beta_1 = 0$  (vs  $H_1: \beta_1 \neq 0$ ), es decir se calcula  $\hat{F}$  y se compara con una  $F_c$  de tablas a un nivel de significancia  $\alpha$  y con 1 y  $n-2$  grados de libertad. Si  $\hat{F} > F_{(1, n-2)}^{1-\alpha}$  entonces se rechaza la hipótesis  $H_0$ .

### 1.3 Pronósticos con el modelo de regresión lineal simple

Una aplicación del modelo de regresión lineal es la predicción o pronóstico del valor futuro de  $Y$  correspondiente a un determinado valor de  $X$ , denominado  $X_0$ , es decir, se trata de calcular

$$\begin{aligned} \hat{Y}_0 &= E(Y | X_0) \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_0 \end{aligned} \quad (1.3.1)$$

$\hat{Y}_0$  es un estimador de su verdadero valor  $Y$  y entonces se debe estimar el error, el cual puede medirse a través de su varianza:

$$\text{Var}(\hat{Y}_0) = \sigma^2 \left( \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right) \quad (1.3.2)$$

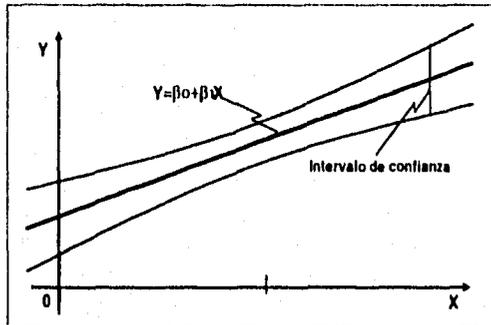
al reemplazar  $\sigma^2$  por su estimador  $\hat{\sigma}^2$  se tiene que la variable

$$t = \frac{\hat{Y}_0 - (\hat{\beta}_0 + \hat{\beta}_1 X_0)}{\hat{\sigma}} - t_{n-2} \quad (1.3.3)$$

Por lo tanto, los límites al  $100(1-\alpha)\%$  de confianza para el verdadero valor medio de  $Y$  dado un valor de  $X_0$ , están dados por

$$\hat{Y}_0 \pm t_{n-2} \hat{\sigma}_{\hat{Y}_0} \quad (1.3.4)$$

La banda de confianza se va haciendo más ancha a medida que se aleje del último punto observado



#### 1.4 El modelo de regresión lineal múltiple

El modelo de regresión lineal múltiple, para  $k-1$  variables explicatorias o independientes y  $n$  observaciones, en términos

generales está dado por  $Y = X\beta + U$ , en donde:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \quad (1.4.1)$$

es un vector de  $n$  observaciones de la variable  $YY$ ;

$$X = \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \dots & X_{1,k-1} \\ 1 & X_{2,1} & X_{2,2} & \dots & X_{2,k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \dots & X_{n,k-1} \end{pmatrix} \quad (1.4.2)$$

es una matriz de cantidades fijas y conocidas y sus columnas corresponden a los datos observados de cada una de las  $k-1$  variables independientes;

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{k-1} \end{pmatrix} \quad (1.4.3)$$

es un vector  $k \times 1$  de parámetros desconocidos, que se asocian a cada una de las variables independientes; y

$$U = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{pmatrix} \quad (1.4.4)$$

es el vector aleatorio que representa a los términos estocásticos de error.

Los supuestos bajo los que se sustenta el modelo, son equivalentes a los del modelo de regresión simple:

- 1) La relación entre  $Y$  y cada una de las variables  $X_1, X_2, \dots, X_{k-1}$  es lineal en los parámetros.
- 2) Los valores de las variables exógenas son no estocásticos con varianza distinta de cero y finita para cualquier tamaño de muestra, es decir, la matriz  $X$  consta de un conjunto de números fijos.
- 3) Los residuales **siguen una distribución normal** con  $E[U]=0$  y  $E[UU'] = \sigma^2 I_n$ ; es decir, que para el primero se tiene

$$E \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (1.4.5)$$

y para el segundo:

$$E(UU') = E \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} (u_1 \ u_2 \ \dots \ u_n) = E \begin{pmatrix} u_1^2 & u_1 u_2 & \dots & u_1 u_n \\ u_2 u_1 & u_2^2 & \dots & u_2 u_n \\ \dots & \dots & \dots & \dots \\ u_n u_1 & u_n u_2 & \dots & u_n^2 \end{pmatrix} \\ = \begin{pmatrix} E(u_1^2) & E(u_1 u_2) & \dots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \dots & E(u_2 u_n) \\ \dots & \dots & \dots & \dots \\ E(u_n u_1) & E(u_n u_2) & \dots & E(u_n^2) \end{pmatrix} \quad (1.4.6)$$

Esta última es la matriz de varianza-covarianza de los errores; sin embargo, haciendo uso de los otros supuestos,  $E[u_i, u_j] = 0$  y  $E[u_i^2] = \text{Var}[u_i] = \sigma^2$ , entonces se reduce a:

$$E(UU^T) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} \quad (1.4.7)$$

$$= \sigma^2 I$$

- 4) Las variables exógenas son linealmente independientes entre sí, lo que implica que el rango de  $X$  es  $k < n$ . Esto significa que no existe multicolinealidad en el modelo.
- 5) El modelo de regresión está especificado correctamente, es decir que no existen sesgos ni errores en la especificación (selección de las variables y formas funcionales).

#### 1.4.1 Estimación de los parámetros

Aplicando el método de mínimos cuadrados y utilizando los supuestos anteriores, se obtiene un estimador lineal insesgado del vector  $\beta$ . El modelo está dado por:

$$Y = X\beta + U$$

de donde

$$U = Y - X\beta \quad (1.4.1.1)$$

Siguiendo el mismo criterio que para el modelo de regresión lineal simple, es necesario minimizar la suma de los residuales al cuadrado:

$$\sum e_i^2 = e_1^2 + e_2^2 + \dots + e_n^2 = (e_1 \ e_2 \ \dots \ e_n) \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = e^T e$$

Por lo tanto,

$$\begin{aligned} e^T e &= (Y - X\beta)^T (Y - X\beta) \\ &= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta \end{aligned} \quad (1.4.1.2)$$

Esta expresión se minimiza cuando:

$$(X^T X)\beta = X^T Y$$

Por tanto, si  $(X^T X)^{-1}$  existe, entonces el estimador de  $\beta$  está dado por:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (1.4.1.3)$$

#### 1.4.2 Matriz de varianza-covarianza de $\hat{\beta}$

En el caso de más de una variable exógena, al igual que en el caso de una sola variable exógena, se calcula la varianza de cada uno de los parámetros involucrados en el modelo y, además, se hace necesario conocer la covarianza entre cualesquiera de los parámetros estimados, mismos que se agrupan en el vector  $\hat{\beta}$ .

La matriz de varianza-covarianza proporciona todas estas relaciones. Por definición, la matriz de varianza-covarianza para  $\hat{\beta}$  es:

$$\begin{aligned} \text{Var} - \text{Cov}(\hat{\beta}) &= E\{(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])^T\} \\ &= \begin{pmatrix} \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{Var}(\hat{\beta}_k) \end{pmatrix} \end{aligned} \quad (1.4.2.1)$$

Para estimar la matriz de varianza-covarianza de  $\beta$  se sustituye  $Y = X\beta + e$  en la expresión que se encontró para estimar el vector de parámetros (1.4.1.3):

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X)^{-1} X^T (X\beta + e) \\ &= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T e \\ &= \beta + (X^T X)^{-1} X^T e\end{aligned}\tag{1.4.2.2}$$

entonces

$$\hat{\beta} - \beta = (X^T X)^{-1} X^T e$$

y sustituyendo en 1.4.2.1 se tiene

$$\begin{aligned}\text{Var-Cov}(\hat{\beta}) &= E\left[\left((X^T X)^{-1} X^T e\right)\left((X^T X)^{-1} X^T e\right)^T\right] \\ &= E\left[\left(X^T X\right)^{-1} X^T e e^T X\left(X^T X\right)^{-1}\right]\end{aligned}\tag{1.4.2.3}$$

Dado que  $X$  es una matriz de valores fijos conocidos, su esperanza es ella misma, por tanto

$$\text{Var-Cov}(\hat{\beta}) = (X^T X)^{-1} X^T E[ee^T] X (X^T X)^{-1}\tag{1.4.2.4}$$

Pero  $E[ee^T] = \sigma^2 I$  y entonces se tiene:

$$\begin{aligned}\text{Var-Cov}(\hat{\beta}) &= (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1}\end{aligned}\tag{1.4.2.5}$$

y un estimador para la varianza es:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k}\tag{1.4.2.7}$$

que matricialmente se expresa como:

$$\hat{\sigma}^2 = \frac{e e^T}{n-k} \quad (1.4.2.7)$$

es decir, la suma de los residuales al cuadrado se calcula como:

$$e e^T = Y^T Y - \beta^T X^T Y \quad (1.4.2.8)$$

### 1.4.3 Coeficiente de determinación

En el caso de regresión lineal simple, el coeficiente de determinación se calcula como

$$r^2 = \frac{\beta^2 \sum x_i^2}{\sum y_i^2} \quad (1.4.3.1)$$

=  $\frac{\text{suma de los errores al cuadrado}}{\text{suma de cuadrados total}}$

y generalizando para  $k$  variables se expresa de la siguiente manera:

$$r^2 = \frac{\beta^T X^T Y - n \bar{Y}^2}{y^T y - n \bar{Y}^2} \quad (1.4.3.2)$$

Sin embargo, para comparar la  $r^2$  de dos modelos alternativos, es necesario considerar el número de variables explicativas que se involucraron en cada modelo. Para ello se calcula un coeficiente de determinación ajustado por los grados de libertad asociados con los  $k$  parámetros que se estiman:

$$\bar{r}^2 = 1 - \frac{\sum e_i^2 / (n-k)}{\sum (Y_i - \bar{Y})^2 / (N-1)} \quad (1.4.3.3)$$

y la relación entre el coeficiente de determinación normal y el ajustado está dada por:

$$\bar{r}^2 = 1 - (1 - r^2) \frac{n-1}{n-k} \quad (1.4.3.4)$$

De aquí se tiene que cuando el número de variables explicativas involucradas en el modelo es grande, con relación al número de observaciones, entonces el coeficiente de determinación ajustado proporciona una medida menos optimista que el no ajustado.

#### 1.4.4 Pruebas de hipótesis e intervalos de confianza

La realización de pruebas de hipótesis y la construcción de intervalos de confianza se efectúa de manera análoga al caso univariado, solo que ahora haciendo operaciones a través de matrices y considerando los efectos cruzados que implica la inclusión de más de una variable explicatoria.

Generalmente las  $e$ 's siguen una distribución normal con media cero y varianza constante  $\sigma^2$ . Matricialmente se tiene:

$$e \sim N(0, \sigma^2 I)$$

con  $e$  y  $0$  vectores columna de  $n \times 1$

Siguiendo las propiedades de los estimadores de mínimos cuadrados ordinarios que se calcularon para el modelo univariado,  $\beta$ , en el sentido de que éstos son iguales a los de máxima verosimilitud, se puede demostrar que cada elemento del vector  $\beta$  se distribuye normalmente con media igual al valor del elemento correspondiente del valor verdadero de  $\beta$  y varianza  $\sigma^2$  multiplicado por  $(X^T X)^{-1}$ :

$$\beta \sim N(0, \sigma^2 (X^T X)^{-1})$$

Utilizando la distribución *t*-student, entonces cada elemento de  $\beta$  se distribuye como una *t*-student con  $n-k$  grados de libertad:

$$t = \frac{\hat{\beta}_i - \beta_i}{\sigma_{\beta_i}} \quad (1.4.4.1)$$

y para probar la hipótesis  $H_0: \beta_i = 0$  vs  $H_1: \beta_i \neq 0$ , se siguen los criterios vistos anteriormente

Para probar la significancia global de la regresión, se utiliza la técnica de análisis de varianza, es decir, la suma total de cuadrados se descompone en la suma de los errores al cuadrado y la suma de los residuales al cuadrado. Entonces, dado que los errores  $e_i$  se distribuyen normalmente, y que se tiene que probar la hipótesis nula  $H_0: \beta_1 = \beta_2 = \dots = \beta_{k-1} = 0$  vs  $H_1$ , se puede demostrar que

$$F = \frac{(\beta X^T Y - n \bar{Y}^2) / (k-1)}{(Y^T Y - \beta X^T Y) / (n-k)} \quad (1.4.4.2)$$

se distribuye como una *F* con  $k-1$  y  $n-k$  grados de libertad, y que su relación con  $r^2$  está dada por:

$$F = \frac{r^2 / (k-1)}{(1-r^2) / (n-k)} \quad (1.4.4.3)$$

### 1.5 Pronóstico con el modelo de regresión lineal múltiple

Una vez que se ha estimado el modelo de regresión, se pueden calcular los valores de la variable  $Y$  si se tiene un vector,  $X_0^T = (1, X_{01}, X_{02}, \dots, X_{0k})$ , de valores conocidos o dados. Entonces el

pronóstico condicional de  $y$  está dado por:

$$E[\hat{y}_0 | X_0^T] = X_0^T \beta \quad (1.5.1)$$

y la varianza del pronóstico está dada por:

$$\text{Var}[\hat{y}_0 | X_0^T] = \hat{\sigma}^2 X_0^T (X^T X)^{-1} X_0 \quad (1.5.2)$$

en donde  $\hat{\sigma}^2$  es el estimador de la varianza de los residuales y el intervalo de confianza para el pronóstico se calcula como:

$$\hat{y}_0 \pm t_{n-1} \sqrt{\hat{\sigma}^2 X_0^T (X^T X)^{-1} X_0} \quad (1.5.3)$$

### 1.6 Violación de los supuestos del modelo de regresión

Los supuestos con los que se construyó el modelo de regresión lineal fueron los siguientes:

- 1) La relación entre la variable dependiente y cada una de las variables independientes es lineal en los parámetros.
- 2) Los valores de las variables exógenas son no estocásticos -valores fijos conocidos- con varianza distinta de cero y finita para cualquier tamaño de muestra
- 3) Los errores son variables aleatorias independientes que siguen una distribución normal con  $E[U] = 0$  y  $E[UU'] = \sigma^2 I_n$  (constante).
- 4) Las variables exógenas son linealmente independientes entre sí.
- 5) El modelo de regresión está especificado correctamente, es decir, que no existen sesgos ni errores de especificación.

En la práctica, es común que uno o varios de estos supuestos no se cumplan. Los dos primeros, desde el momento en que se escoge este método para relacionar las variables se dan por cumplidos, pero los siguientes es preciso verificar que se cumplan y tratar de corregir las violaciones.

Cuando el supuesto 3 no se cumple en lo que se refiere a que la varianza de los residuales sea constante, entonces se dice que se tiene un problema de heterocedasticidad, y cuando se tiene que no son independientes entre sí, entonces se está en un caso de autocorrelación de los residuales. La violación del supuesto 4, es decir que al menos un par de variables exógenas no sea independiente entre sí se denomina multicolinealidad. Finalmente, cuando no se cumple con el supuesto 5, es necesario revisar la especificación del modelo.

### 1.6.1 Heterocedasticidad

La heterocedasticidad origina que los estimadores de mínimos cuadrados ordinarios de la ecuación de regresión no tengan varianza mínima y por tanto que no sean eficientes.

Debido a que en la práctica no es fácil detectar la heterocedasticidad, los métodos que se han desarrollado para advertir su presencia examinan los residuales obtenidos por el método de mínimos cuadrados para encontrar en ellos patrones sistemáticos que pueden sugerir maneras de transformar el modelo original, de tal manera que en la ecuación transformada los errores tengan varianza constante.

### 1.6.2 Autocorrelación

Alguna de las razones por las que se presenta la autocorrelación son: la inercia que muestran la mayoría de las series de tiempo económicas, el sesgo que resulta de excluir algunas variables relevantes para el modelo, la utilización de una forma funcional incorrecta, la exclusión de variables rezagadas y la manipulación de los datos.

Cuando existe autocorrelación, los estimadores del modelo de regresión lineal dejan de ser eficientes y entonces las pruebas  $t$  o  $F$  no se pueden aplicar.

Para detectar la autocorrelación se pueden utilizar varios métodos. Intuitivamente, se pueden graficar los residuales contra el tiempo o contra los valores rezagados de los mismos y tratar de identificar algún patrón sistemático. Sin embargo el más conocido es el de la prueba con el estadístico  $DW$  de Durbin-Watson el cual se define como:

$$Dh' = \frac{\sum_{i=2}^N (e_i - e_{i-1})^2}{\sum_{i=1}^N e_i^2} \quad (1.6.2.1)$$

el cual, como regla general, si es igual a 2 ó cercano, se puede suponer que no existe correlación de primer orden.

Esta prueba es válida sólo en aquellos modelos de regresión lineal en los que no se ha incluido como variable explicativa a la variable dependiente rezagada, y además en donde las perturbaciones estén generadas por un proceso autorregresivo de primer orden.

### 1.6.3 Multicolinealidad

La multicolinealidad se refiere a la existencia de una relación lineal exacta o aproximada entre las variables independientes. Cuando existe colinealidad perfecta entre dos o más variables independientes, entonces la inversa de la matriz  $(X^T X)$  no existe y por tanto los coeficientes de la ecuación de regresión son indeterminados. Cuando la colinealidad es alta pero no perfecta, entonces los coeficientes sí se pueden calcular pero sus errores estándar tienden a ser grandes. Algunos indicadores de la existencia de multicolinealidad son:

- Se tiene una  $r^2$  alta, pero ninguno de los coeficientes de regresión es significativo.
- En el modelo con dos variables independientes, el coeficiente de correlación simple es elevado. Si se tienen más variables, entonces hay que examinar los coeficientes de correlación parcial (véase Gujarati 1992).
- Cuando la  $r^2$  es alta, pero las correlaciones parciales son bajas, es posible que haya multicolinealidad. Sin embargo, si las correlaciones parciales también son elevadas, la multicolinealidad no se puede detectar directamente; entonces, se puede ensayar construir un modelo de regresión para cada una de las variables independientes en función de las restantes en el modelo original y calcular los correspondientes coeficientes de determinación; la variable asociada a un  $r^2$  elevado puede eliminarse del modelo.

Para remediar el problema de la multicolinealidad se puede utilizar información obtenida en forma previa o externa al modelo, combinar información de series de tiempo y de corte transversal, omitir una variable altamente colineal,

transformar la información u obtener información nueva o adicional. De la naturaleza de los datos y la severidad el problema, depende la elección del proceso para solucionarlo.

#### 1.6.4 Especificación del modelo

La especificación correcta de un modelo se puede juzgar a través de algunos criterios prácticos:

- 1) simplicidad,
- 2) identificabilidad,
- 3) bondad de ajuste,
- 4) consistencia teórica, y
- 5) capacidad predictiva.

De éstos, se pueden distinguir errores de especificación de un modelo y errores de especificación equivocada del mismo.

En el análisis de error de especificación se supone que el modelo es correcto y se intenta saber que ocurre con los estimadores de MCO cuando se cometen uno o más de los siguientes errores:

- 1) Omisión de una o más variables relevantes.- En este caso los estimadores de MCO de los coeficientes de las variables que se encuentran en el modelo son sesgados e inconsistentes; además, sus varianzas y errores estándar están estimados en forma incorrecta, lo que vicia los procedimientos usuales de prueba de hipótesis.
- 2) Inclusión de una o más variables innecesarias.- Las varianzas que se estiman tienden a ser más grandes que lo necesario, lo que hace menos precisa la estimación de los parámetros y los intervalos de confianza son más amplios.

- 3) Utilización de una forma funcional incorrecta,
- 4) Especificación incorrecta del término de perturbación  $u_i$ ,
- 5) Errores de medición.- Cuando la medición de la variable dependiente contiene errores y la de las independientes no, los estimadores de MCO continúan siendo insesgados y consistentes pero son menos eficientes. Cuando los errores también se presentan en la medición de las variables explicativas, los estimadores de MCO son inconsistentes y sesgados. El remedio de esta situación es difícil en la práctica ya que requiere la utilización de variables aproximadas o instrumentales que habría que justificar teóricamente para poder incluirlas.

Para analizar los errores de una especificación equivocada, las pruebas no son fáciles de instrumentar ya que se tienen dos o más teorías (modelos) que compiten entre sí. Un procedimiento para decidir entre los modelos consiste en unir los modelos y entonces se escoge aquel cuyos coeficientes sean significativamente distintos de cero; pero surge entonces el problema de colinealidad entre las variables de ambos modelos. Obviamente, teóricamente es muy difícil justificar el llevar a cabo una "mezcla" de esta naturaleza.

## Capítulo 2

### Modelos de series de tiempo

La exposición de las principales herramientas para realizar el análisis de series de tiempo que se presentan en este capítulo, está basada en el trabajo que realizó Víctor Guerrero para proporcionar, de una manera más elemental, la metodología que sugirieron Box y Jenkins para construir modelos de series de tiempo en su libro *Análisis Estadístico de Series de Tiempo Económicas* (1991).

Al contrario del análisis de regresión, donde los datos con que se cuenta para explicar cierto fenómeno se ajustan a un modelo establecido de antemano, la estrategia de Box y Jenkins es dejar que los datos indiquen la elección del modelo que represente adecuadamente su comportamiento.

La exposición de la estrategia para construir modelos de series de tiempo univariados mencionada, se complementa con la presentación de los modelos de función de transferencia, que en series de tiempo es el equivalente al análisis de regresión lineal múltiple, como introducción al análisis de intervención.

#### 2.1 Definición

Una serie de tiempo es un conjunto de observaciones ordenadas en el tiempo. Sólo se consideran series discretas con observaciones  $Z(t)$  hechas a intervalos iguales de tiempo  $\{t, t=1, 2, \dots, N\}$ , en donde  $N$  es el número de observaciones. Aún cuando el fenómeno

que se estudie pueda ser continuo, es posible obtener una serie discreta tomando lecturas en determinados intervalos de tiempo para obtener una serie muestra durante un lapso de tiempo; y por lo general, el intervalo de tiempo  $\tau_i = t_{i+1} - t_i$  entre las observaciones sucesivas es igual (minutos, horas, días, semanas, meses).

Si se puede predecir exactamente el comportamiento futuro de una serie a partir del conocimiento que se tiene de su pasado y su presente, entonces se trata de una serie determinista por lo que no es necesario ningún análisis; sin embargo, en la práctica se tienen series de tiempo estocásticas en las que el conocimiento previo sólo puede indicar la estructura probabilística del comportamiento futuro.

## 2.2 Naturaleza probabilística de las series de tiempo

Una serie de tiempo se puede considerar como una realización de un proceso estocástico. Pero en la práctica es casi imposible obtener otra realización del proceso estocástico que genera a la serie, por ejemplo, no se pueden repetir fenómenos económicos para investigar qué otros patrones pudieron haberse seguido.

Entonces, se tiene que cualquier  $(z_i)$  es una realización de algún proceso estocástico  $\{Z_i\}$  con una función de densidad de probabilidad asociada (fdp),  $P(z_i)$ ; y cualquier conjunto de  $(z_i): (z_{j_1}, z_{j_2}, \dots, z_{j_r})$  tiene una función de densidad de probabilidad conjunta  $P(z_{j_1}, z_{j_2}, \dots, z_{j_r})$ . Si un proceso estocástico es tal que  $P(z_{j_1}, z_{j_2}, \dots, z_{j_r})$  es independiente de  $j$  para algún entero positivo  $m$  y cualquier  $n_1, n_2, \dots, n_m$ , significa que la estructura probabilística no cambia con el tiempo y el proceso es estrictamente estacionario; de otra manera, es no estacionario.

Si se considera un proceso gaussiano cuya función de densidad de probabilidad asociada con cualquier conjunto de intervalos de tiempo, es una normal multivariada, entonces el proceso va a ser estrictamente estacionario con estacionariedad de orden 2, ya que los momentos de orden más alto son fijos. Lo anterior se denomina estacionariedad débil, e implica que  $E[z_i] = \mu$  y  $Cov[z_i, z_{i-k}] = \gamma_k$  en donde  $\mu$  y  $\gamma_k$  son constantes y  $k$  es un entero,  $\mu$  es la media o nivel del proceso y  $\gamma_k$  es la autocovarianza en el rezago  $k$ . En particular,  $z_i$  tiene varianza constante  $\sigma_i^2 = \gamma_0$ .

También para todo entero  $k$ ,  $\gamma_{-k} = \gamma_k$  ya que

$$Cov\{Z_i, Z_{i+k}\} = Cov\{Z_{i+k}, Z_i\} = Cov\{Z_i, Z_{i-k}\} \quad (2.2.1)$$

y esto es necesario para determinar  $\gamma_k$ , con  $k \geq 0$ . Al conjunto  $\{\gamma_k : k = 0, 1, \dots\}$  se le conoce como la función de autocovarianza, y la autocorrelación al rezago  $k$  se define como:

$$\rho_k = \frac{Cov(Z_i, Z_{i-k})}{\sqrt{Var(Z_i)Var(Z_{i-k})}}, \quad k = 0, 1, \dots \quad (2.2.2)$$

o sea,

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad k = 0, 1, \dots \quad (2.2.3)$$

y es independiente de la escala de medición. Así, la función de autocorrelación (FAC) se forma por el conjunto  $\{\rho_k : k = 0, 1, \dots\}$  en donde  $\rho_0 = 1$ .

### 2.3 Técnicas para pronosticar series de tiempo

Existen una gran variedad de técnicas para estimar los valores futuros de una serie de tiempo. La selección de alguna de ellas depende en gran medida de las características de la serie que se

está analizando; para unas es suficiente utilizar modelos sencillos, y para otras se requiere de un análisis más complejo. También intervienen en la selección del método de pronóstico factores tales como las herramientas disponibles para el proceso de los datos y el tipo de información para poder aplicar o construir los modelos (Makridakis, 1980).

Algunas de las técnicas más utilizadas son las siguientes:

A) Métodos de suavizamiento.

La base de los pronósticos es el promedio de los errores pasados ya que se agrega un porcentaje del error al pronóstico previo:

$$F_{t+1} = F_t + \alpha(X_t - F_t) \quad (2.3.1)$$

y sustituyendo la misma definición para  $F_t$  y  $F_{t-1}$  se llega a:

$$F_{t+1} = F_{t-2} + \alpha(X_{t-2} - F_{t-2}) + \alpha(X_{t-1} - F_{t-1}) + \alpha(X_t - F_t) \quad (2.3.2)$$

es decir, dado un pronóstico inicial  $F_t$ , el siguiente se obtiene al agregar un porcentaje del error entre este pronóstico y su valor real, al pronóstico inicial. Algunos de estos errores pueden ser positivos o negativos por lo que el pronóstico final será cercano, en promedio, al patrón que marque el valor real de los datos. Si el porcentaje del error que se agrega, surge del promedio de los últimos  $k$  pronósticos ( $k$  fijo), y éste se va actualizando a medida que se conocen más observaciones. El modelo se llama de **promedios móviles** y se denomina con las siglas **MA** (Moving Average).

B) Métodos de descomposición

Se basa en el principio de separar la serie en cada uno de los componentes: estacionalidad, tendencia, ciclo y

aleatoriedad; y entonces se pronostica separadamente cada uno de estos elementos, excepto la aleatoriedad, y se recombinan para formar el pronóstico de la serie original.

**C) Métodos causales o de regresión**

Como se vió en el capítulo anterior, los métodos de regresión pronostican el valor de la variable dependiente con base en el comportamiento de otros factores para tener un modelo causal

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + u_t \quad (2.3.3)$$

en donde  $X_1, X_2, \dots, X_k$  son variables explicatorias.

**D) Métodos autorregresivos.**

Si en la ecuación (2.3.3) las variables independientes se definen como  $X_1 = Y_{t-1}, X_2 = Y_{t-2}, \dots, X_k = Y_{t-k}$  entonces se convierte en:

$$Y_t = a + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_k Y_{t-k} + u_t \quad (2.3.4)$$

la cual es todavía una ecuación de regresión pero difiere de la original en que las variables independientes son ahora los valores previos de la propia variable dependiente  $Y_t$ , y este tipo de modelos se conocen como **autorregresivos (AR)**. Sin embargo, esta ecuación no se puede tratar de la misma manera que una ecuación de regresión ordinaria porque se viola el supuesto de independencia de las variables independientes, y por otra parte, porque no es posible determinar exactamente el número de rezagos que se deben incluir.

**E) Modelos integrados.**

Los modelos autorregresivos se combinan con términos de promedios móviles en una clase general de modelos de series de tiempo y se denominan procesos autorregresivos y de

promedios móviles (ARMA). Cuando además, para estacionarizar la serie, se transforma a través de diferencias, los modelos se denominan **autorregresivos y de promedios móviles integrados (ARIMA)**.

**F) Modelos de transferencia.**

Estos modelos representan la versión multivariada de los modelos ARIMA y en ellos intervienen conceptos tales como causalidad y correlación cruzada al relacionar la variable dependiente con los valores rezagados de las variables independientes y de ella misma.

En esta trabajo se presenta la metodología que sugieren Box y Jenkins para construir los modelos univariados de series de tiempo y los modelos de función de transferencia, estos últimos como una introducción al análisis de intervención de las series.

#### **2.4 Algunos elementos para realizar el análisis de series de tiempo**

Para aplicar el método de Box-Jenkins en la construcción de los modelos de series de tiempo es necesario identificar de antemano algunas características de la serie de tiempo que deberán incorporarse en la especificación del modelo; por ejemplo, la existencia de patrones estacionales, varianza no constante, media diferente de cero y tendencia.

En principio, se tratará de que la serie en estudio ( $Z_t$ ) tenga varianza constante, para lo cual muchas veces es necesario aplicar una transformación estabilizadora de varianza a los datos originales:  $T(Z_t)$ . Después, para que la serie sea estacionaria se aplica el operador diferencia tantas veces como sea necesario.

Posteriormente, para saber si la serie sobre la que se va a diseñar el modelo cumple con los requisitos de estacionariedad, se utiliza el análisis de autocorrelación.

En esta sección se presentan algunos elementos necesarios para realizar el análisis de series de tiempo, que son la elección de la transformación para estabilizar la varianza, el uso de los operadores de retraso y el análisis de autocorrelación.

#### 2.4.1 Estabilización de la varianza

El método que se sugiere para estabilizar la varianza (Guerrero 1991), contempla la elección de una potencia  $\lambda$  tal que para  $t = 1, 2, \dots, N$ , con  $N$  el número de observaciones que se tengan de la serie  $\{Z_t\}$ , se satisfaga la relación

$$\frac{\sigma_t}{\mu_t^{1-\lambda}} = \text{constante} \quad (2.4.1.1)$$

en donde  $\sigma_t$  y  $\mu_t$  representan la desviación estándar y la media de la variable  $\{Z_t\}$ , respectivamente. Los pasos a seguir son:

- Se dividen las  $N$  observaciones de la serie en  $H$  grupos homogéneos que contengan el mismo número de observaciones  $R$ . Para ésto se dejarán fuera  $n$  observaciones ( $0 \leq n \leq R$ ) ya sea del principio o del final de la serie.
- Se calculan la media y la desviación estándar de cada grupo:

$$\bar{Z}_h = \sum_{r=1}^R \frac{Z_{h,r}}{R} \quad (2.4.1.2)$$

$$y$$

$$S_h = \sqrt{\sum_{r=1}^R \frac{(z_{h,r} - \bar{z}_h)^2}{R-1}} \quad (2.4.1.3)$$

- Para cada valor propuesto de  $\lambda$  se calcula el coeficiente de variación como

$$CV(\lambda) = \frac{de(\lambda)}{M(\lambda)} \quad (2.4.1.4)$$

con

$$de(\lambda) = \sqrt{\frac{\sum_{h=1}^H \left( \frac{S_h}{\bar{z}_h^{1-\lambda}} - M(\lambda) \right)^2}{H-1}} \quad (2.4.1.5)$$

y

$$M(\lambda) = \sum_{h=1}^H \frac{S_h}{\bar{z}_h^{1-\lambda}} \frac{1}{H} \quad (2.4.1.6)$$

y se construye el siguiente cuadro:

Grupo	Potencia $\lambda$				
	-1	-0.5	0	0.5	1
1	$\frac{S_1}{\bar{z}_1^{-1}}$	$\frac{S_1}{\bar{z}_1^{-0.5}}$	$\frac{S_1}{\bar{z}_1^0}$	$\frac{S_1}{\bar{z}_1^{0.5}}$	$S_1$
2	$\frac{S_2}{\bar{z}_2^{-1}}$	$\frac{S_2}{\bar{z}_2^{-0.5}}$	$\frac{S_2}{\bar{z}_2^0}$	$\frac{S_2}{\bar{z}_2^{0.5}}$	$S_2$
...					
h	$\frac{S_h}{\bar{z}_h^{-1}}$	$\frac{S_h}{\bar{z}_h^{-0.5}}$	$\frac{S_h}{\bar{z}_h^0}$	$\frac{S_h}{\bar{z}_h^{0.5}}$	$S_h$
...					
H	$\frac{S_H}{\bar{z}_H^{-1}}$	$\frac{S_H}{\bar{z}_H^{-0.5}}$	$\frac{S_H}{\bar{z}_H^0}$	$\frac{S_H}{\bar{z}_H^{0.5}}$	$S_H$
<b>Coefficiente de variación</b>	<b>CV(-1)</b>	<b>CV(-0.5)</b>	<b>CV(0)</b>	<b>CV(0.5)</b>	<b>CV(1)</b>

- Se elige la potencia de  $\lambda$  que proporcione el mínimo coeficiente de variación para satisfacer de manera aproximada la relación propuesta. En el cuadro los valores de  $\lambda$  son sugeridos, pero se pueden ensayar otro espaciamiento para obtener un valor más exacto.
- Entonces la serie transformada que resulta es  $T(Z_t) = Z_t^\lambda$  si  $\lambda \neq 0$ , y  $T(Z_t) = \log(Z_t)$  si  $\lambda = 0$ .

Este método proporciona únicamente un valor aproximado para el valor de la potencia; sin embargo, puede utilizarse una transformación que surja del conocimiento que se tenga del fenómeno que genera a la serie. Para series económicas la transformación que generalmente se utiliza es la logarítmica debido a la interpretación que se puede dar a los coeficientes que se calculan.

#### 2.4.2 Operadores de retraso y de diferencia

En el método de Box-Jenkins para analizar series de tiempo de acuerdo con su comportamiento pasado, se utilizan el operador de retraso  $B$  (backshift), que se define como:

$$BZ_t = Z_{t-1} \quad (2.4.2.1)$$

y en general,  $B^k Z_t = Z_{t-k}$  para  $k = 1, 2, \dots, N$

El operador diferencia,  $\nabla$ , está definido por:

$$\nabla Z_t = Z_t - Z_{t-1} \quad (2.4.2.2)$$

y también generalizando se tiene que  $\nabla^k Z_t = Z_t - Z_{t-k}$ , con  $k = 1, 2, \dots, N$

La relación que existe entre estos dos operadores es

$$\gamma = 1 - B \quad (2.4.2.3)$$

y obedece las reglas del álgebra elemental.

### 2.4.3 Autocorrelación

El análisis de series de tiempo es un proceso sistemático a través del cual se identifican características de las series tales como estacionariedad y estacionalidad utilizando los coeficientes de autocorrelación para diferentes rezagos de la variable que se pronostica.

Las autocorrelaciones proporcionan información adicional sobre la forma en que los mismos valores de la serie están relacionados. Además de indicar si los datos de la serie se generan por un proceso estacionario y/o estacional, pueden indicar las transformaciones necesarias para obtener series que cumplan, aproximadamente, con estas características.

Para una muestra, la estimación de la función de autocorrelación ( $\rho_k$ ) se calcula como:

$$r_k = \frac{c_k}{c_0} = \frac{\sum_{i=1}^{N-k} (Z_i - \bar{Z})(Z_{i+k} - \bar{Z})}{\sum_{i=1}^{N-k} (Z_i - \bar{Z})^2}, \quad k = 0, 1, 2, \dots \quad (2.4.3.1)$$

en donde

$$\bar{Z} = \frac{1}{N} \sum_{i=1}^N Z_i$$

y  $c_0 = 1$

y cada  $r_k$  se distribuye como una normal con media  $\bar{z}$  y varianza  $\sigma_i^2$ .

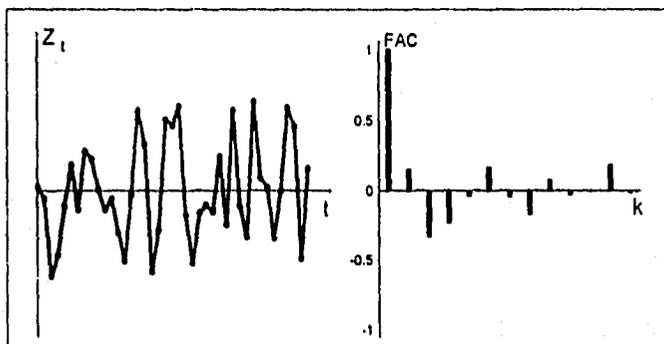
Para determinar las autocorrelaciones que son diferentes de cero, deben compararse los valores los valores de  $r_k$  con sus correspondientes desviaciones estándar. Por tanto, se considerará que las autocorrelaciones son distintos de cero a un nivel de significancia de 5% cuando:

$$|\rho_k| > 1.96 \sqrt{\frac{1}{N} \left( 1 + 2 \sum_{j=1}^k \sigma_j^2 \right)} \quad (2.4.3.2)$$

De esta manera, la función de autocorrelación se puede utilizar para determinar si una serie de tiempo fue generada por un proceso estocástico. Los coeficientes de autocorrelación de varios rezagos en el tiempo se examinan para ver si alguno de ellos es significativamente diferente de cero. También, una vez que se tienen los datos ajustados por el modelo estimado, se calculan las autocorrelaciones de los errores para determinar si éstos son efectivamente aleatorios.

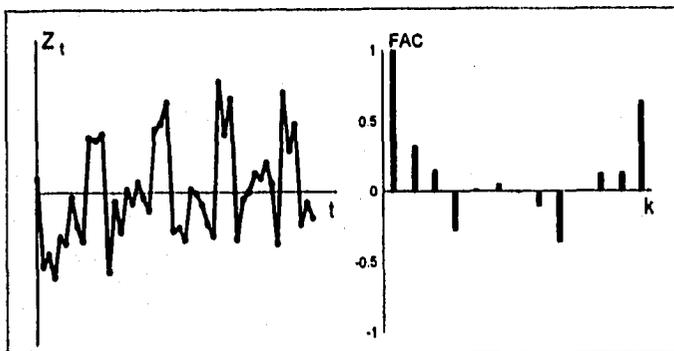
#### 2.4.3.1 Estacionariedad

Una serie es estacionaria si su media es cero y tiene varianza constante en cualquier periodo de tiempo. Las autocorrelaciones de una serie estacionaria decaen a cero después del segundo o tercer rezago, mientras que las no estacionarias son significativamente diferentes de cero para varios periodos.



### 2.4.3.2 Estacionalidad

La estacionalidad se define como un comportamiento que se repite en intervalos de tiempo fijos. Si el patrón es consistente, el coeficiente de autocorrelación correspondiente será elevado y positivo. Esto es posible de identificar en series estacionarias, y la tarea se dificulta si la serie de tiempo tiene otros patrones tales como la tendencia. Por tanto, es recomendable transformar la serie original a una serie aproximadamente estacionaria para identificar los patrones estacionales.



### 2.4.3.3 La prueba $\chi^2$

Guerrero (1991) presenta una prueba estadística conjunta para determinar si los coeficientes de autocorrelación son significativamente diferentes de cero; ésta es la que propusieron Box y Pierce en 1970, a través del cálculo del estadístico Q, el cual se basa en la distribución  $\chi^2$  del coeficiente de autocorrelación.

Si el valor calculado de la prueba es menor que los de una tabla  $\chi^2$ , las autocorrelaciones que se usan para calcular la prueba, no son significativamente diferentes de cero e implica que los datos se generan por un proceso de ruido blanco aleatorio. Si la  $\chi^2$  observada es mayor al valor de tablas, las autocorrelaciones pueden considerarse diferentes de cero e indican la existencia de algún patrón en la serie. El estadístico Q se calcula como:

$$Q = n \sum_{m=1}^{k-1} r_k^2 \quad (2.4.3.3.1)$$

en donde  $m$  es el rezago más antiguo que se utiliza en el cálculo de las autocorrelaciones. De aquí se desprende también que, entre más pequeñas sean las autocorrelaciones, el valor de  $\chi^2$  será menor.

Esta prueba es la primera aproximación en la determinación de la aleatoriedad de los datos.

#### 2.4.4 La función de autocorrelación parcial

Otra herramienta necesaria es la función de autocorrelación parcial (FACP), que se denota por  $\{\phi_{kk}; K=1,2,\dots\}$ , y es el conjunto de autocorrelaciones parciales en varios rezagos  $k$ , y se define como:

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|} \quad (2.4.4.1)$$

en donde  $P_k$  es la matriz de autocorrelación de dimensión  $(k \times k)$  y  $P_k^*$  es  $P_k$  con la última columna reemplazada por el vector  $(\rho_1, \rho_2, \dots, \rho_k)'$ , y así:

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

y los estimadores  $\hat{\phi}_{kk}$  se obtienen reemplazando  $\rho_i$  por  $r_i$ .

Para rezagos suficientemente grandes, se tiene que

$$\text{Var}[\hat{\phi}_{kk}] = \frac{1}{N} \quad (2.4.4.2)$$

y por lo tanto la desviación estándar es:

$$\text{E.S.}[\hat{\phi}_{kk}] = \frac{1}{\sqrt{N}} \quad (2.4.4.3)$$

y para  $N$  grande,  $\hat{\phi}_{kk}$  se distribuye aproximadamente como una normal.

## 2.5 El método de Box-Jenkins

El tipo de modelo que proponen Box y Jenkins es de la forma:

$$\phi(B)\nabla T(Z_t) = \theta(B)a_t \quad (2.5.1)$$

en donde  $\phi$  y  $\theta$  son polinomios de retraso que representan a los procesos autorregresivos y de promedios móviles, respectivamente,  $T(Z_t)$  es una serie estacionaria y  $\{a_t\}$  es una serie de choques generados por un proceso de ruido blanco, es decir, que cada  $a_t$  es independiente y se distribuye con media cero y varianza constante  $\sigma_a^2$ .

Con el método de Box-Jenkins se espera reconocer patrones que puedan explicarse por un modelo de este tipo a partir de las FAC y FACP estimadas para una serie observada. Después de esta identificación, se obtienen los estimadores eficientes de los parámetros y se procede a verificar si el modelo es adecuado. Si no lo es, se cuenta con diversas pruebas que podrían indicar la manera en que el modelo propuesto puede modificarse, volviéndose a identificar, estimar y verificar.

### 2.5.1 Modelos autorregresivos

La forma general de un procesos autorregresivo de orden  $p$ ,  $AR(p)$ , es

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (2.5.1.1)$$

en donde el valor actual del proceso se expresa como una suma ponderada de los valores pasados, más un choque aleatorio. Así, se tiene una ecuación de regresión en la que las variables independientes son la misma variable pero rezagada; de ahí el nombre de autorregresivo.

En términos del operador de retraso, la ecuación anterior se puede escribir como

$$c(B)Z_t = a_t \quad (2.5.1.2)$$

en donde:

$$c(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

En el caso de un proceso autorregresivo de orden 1,  $AR(1)$ :

$$Z_t - \phi Z_{t-1} = a_t \quad (2.5.1.3)$$

en donde  $a_t \sim IN(0, \sigma_a^2)$  y se supone que el proceso es estacionario.

Como las  $a_t$  son independientes de  $Z_{t-1}$ , la varianza es

$$\sigma_z^2 = \phi^2 \sigma_z^2 + \sigma_a^2$$

y de aquí que

$$\sigma_z^2(1 - \phi^2) = \sigma_a^2$$

y para  $\sigma_z^2$  finita y no negativa se requiere que  $-1 < \phi < 1$

Esta es la condición de estacionariedad, y en general, la condición necesaria y suficiente para que un proceso AR(p) sea estacionario es que las raíces de la ecuación característica  $\phi(B)=0$  estén fuera del círculo unitario.

Tomando la esperanza de (2.5.1.3) se tiene:

$$\mu = \phi\mu + 0$$

y como  $\phi \neq 1$ , entonces  $\mu = 0$ .

Ahora bien, si se multiplica la misma ecuación por  $Z_{t-k}$ , y se toma la esperanza matemática,

$$E[Z_t Z_{t-k}] = \phi E[Z_{t-1} Z_{t-k}] + E[a_t Z_{t-k}]$$

para  $k \geq 1$ , y como  $a_t$  y  $Z_{t-k}$  son independientes, se obtiene  $\gamma_k = \phi \gamma_{k-1}$  que es una ecuación en diferencias de primer grado y cuya solución es:

$$\gamma_k = \phi^k \gamma_{k-1}$$

o bien,

$$\rho_k = \phi^k, \quad k \geq 1$$

Para un valor grande y positivo de  $\phi$ , los términos adyacentes de la serie están altamente correlacionados y la FAC decae geoméricamente. Para una  $\phi$  negativa, la FAC alterna en signo decayendo a cero.

Por su parte, la FACP teórica está dada por el término

$$\phi_{11} = \rho_1 = \phi$$

y

$$\phi_{kk} = 0, \quad k \geq 1$$

Además,

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0$$

ya que  $\phi^2 \neq 1$

### 2.5.2 Modelos de promedios móviles

El modelo de promedios móviles de orden  $q$ , MA( $q$ ) está dado por

$$Z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad (2.5.2.1)$$

en donde  $a_t \sim IN(0, \sigma_a^2)$ , es decir, el valor actual de la serie se expresa como la suma ponderada de los choques aleatorios que generan el proceso. Esta última ecuación se puede escribir como:

$$Z_t = \theta(B)a_t \quad (2.5.2.2)$$

en donde  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  es el operador de orden  $q$  y la varianza del proceso está dada por:

$$\sigma_z^2 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2 \quad (2.5.2.3)$$

y para una  $q$  finita este proceso es siempre estacionario.

La ecuación 2.5.2.2 se puede reescribir como:

$$\theta^{-1}(B)Z_t = a_t$$

que es de la forma

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots = \phi(B)Z_t = a_t \quad (2.5.2.4)$$

Se dice que el proceso MA(q) es invertible si las ponderaciones  $\phi$  forman una serie convergente, es decir, si las raíces del polinomio  $\theta(B)$  están fuera del círculo unitario. Esta es una condición análoga a la de estacionariedad de un proceso AR(p).

La ecuación 2.5.2.4 muestra que un proceso MA(q) es equivalente a un proceso AR de orden infinito. De la misma manera, un proceso AR(p) se puede escribir como un proceso MA( $\infty$ ).

Considérese ahora el proceso MA(1)

$$Z_t = a_t + \theta a_{t-1}$$

Por la condición de invertibilidad se tiene que  $-1 > \theta > -1$ , y tomando esperanza,  $\mu = 0$  para toda  $k$  y:

$$E[Z_t, Z_{t-k}] = E[(a_t + \theta a_{t-1})(a_{t-k} + \theta a_{t-k-1})]$$

$$Y \quad \sigma_z^2 = \gamma_0 = (1 + \theta^2)\sigma_a^2$$

$$\gamma_1 = \theta\sigma_a^2$$

$$Y \quad \gamma_k = 0, \quad k > 1$$

Así, la FAC está dada por

$$\rho_1 = \frac{\theta}{1 + \theta^2}$$

$$\rho_k = 0, \quad k > 1$$

La FAC desaparece después de un rezago; pero la FACP decae a cero geométricamente:

$$\phi_{kk} = \frac{(-1)^{k-1} \theta^k (1 - \theta^2)}{1 + \theta^{2(k+1)}}$$

### 2.5.3 Modelos autorregresivos y de promedios móviles integrados, ARMA

Una extensión razonable de los modelos AR y MA es el modelo ARMA de la forma:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (2.5.3.1)$$

o bien

$$\phi(B)Z_t = \theta(B)a_t$$

Para que este proceso cumpla con las condiciones de estacionariedad e invertibilidad se requiere que las raíces de los polinomios característicos  $\phi(B)$  y  $\theta(B)$  estén fuera del círculo unitario. Tomando la esperanza de la ecuación 2.5.3.1, se tiene que  $E[Z_t] = 0$  ya que  $\phi(1) \neq 0$ . Multiplicando la misma ecuación por  $Z_{t-k}$ , para cualquier  $k$  entero, y tomando la esperanzas

$$Y_k = \phi_1 Y_{k-1} + \dots + \phi_p Y_{k-p} + \theta_1 \gamma_{az}(k-1) + \dots + \theta_q \gamma_{az}(k-q) \quad (2.5.3.2)$$

en donde  $\gamma_{az}(j)$  es la covarianza cruzada entre  $a$  y  $Z$  en el rezago  $j$ :

$$\gamma_{az}(j) = \text{Cov}[a_t, Z_{t-j}]$$

Como  $E[a_t]$  y  $E[Z_{t-j}]$  son cero,

$$\gamma_{az}(j) = E[a_t Z_{t-j}]$$

y se tiene la siguiente propiedad

$$\gamma_{az}(j) = 0, \quad j > 0$$

Ya que  $a_t$  es independiente de cualquier  $Z$  anterior. Así, para  $k > q$ , se reduce a una ecuación en diferencias

$$Y_k = \phi_1 Y_{k-1} + \dots + \phi_p Y_{k-p}$$

que no involucra a los parámetros de promedios móviles. Así, después del rezago  $q$  las autocovarianzas, y consecuentemente las autocorrelaciones, se comportan como un proceso AR  $\phi(B)Z_t = \alpha_t$ . Sin embargo, para el proceso ARMA las  $\gamma_k$  más cercanas dependerán de los parámetros MA a través de la ecuación (2.5.3.2).

El modelo se puede escribir como

$$Z_t = \Psi(B)\alpha_t \quad (2.5.3.3)$$

ó

$$\phi(B)Z_t = \alpha_t \quad (2.5.3.4)$$

en donde  $\Psi(B) = \phi^{-1}(B)\theta(B)$  y  $\phi(B) = \theta(B)^{-1}\phi(B)$  son series infinitas en  $B$ . Por lo tanto, el modelo expresado ya sea en la forma de promedios móviles o autorregresivo, conduce a que el comportamiento de la FACP sea decreciente cuando  $t$  tiende a infinito.

#### 2.5.4 Modelos autorregresivos e integrados de promedios móviles, ARIMA

Estos modelos representan la generalización de los modelos ARMA:

$$\phi(B)\nabla^d T(Z_t) = \theta(B)\alpha_t, \quad d \geq 1 \quad (2.5.4.1)$$

en donde se hace explícita la aplicación de  $d$  diferencias para lograr que la serie  $\{T(Z_t)\}$  sea estacionaria.

La referencia a estos modelos se hace como  $ARIMA(p,d,q)$ , siendo  $p$  el grado del polinomio asociado al proceso autorregresivo,  $q$  el de promedios móviles y  $d$  el grado de la diferencia.

### 2.5.5 La construcción de los modelos

La metodología de Box-Jenkins es un proceso iterativo cuyas etapas de construcción son:

- 1a. **Identificación** del modelo dentro de la clase ARIMA.
- 2a. **Estimación** de los parámetros usando técnicas de estimación no lineal.
- 3a. **Verificación** de que el modelo proporcione el ajuste adecuado y de que los supuestos básicos que lo sustentan se cumplan. Si el modelo no cumple con los supuestos, es necesario modificarlo y repetir las primeras tres etapas hasta que se tengan resultados satisfactorios.
- 4a. **Uso** del modelo con fines de pronóstico, control y/o simulación del fenómeno que se está estudiando.

#### 2.5.5.1 Identificación

El primer paso en esta etapa es la elaboración de una gráfica de la serie con el objeto de hacer una inspección visual y determinar indicios de varianza no constante, factores estacionales y/o tendencia. Lo anterior con el propósito de aplicar una transformación estabilizadora de varianza para obtener una serie estacionaria  $\{T(Z_t)\}$ .

El siguiente paso consiste en identificar un modelo tentativo para la serie estacionaria. Se dice que un modelo ARIMA se identifica cuando se propone el valor de  $p$ ,  $d$  y  $q$  una vez que se ha analizado el comportamiento de la FAC y de la FACP calculadas a partir de la serie transformada  $T(Z_1), T(Z_2), \dots, T(Z_N)$ . Es decir, se calcula la media y la varianza, con lo que se espera que  $\{r_t\}$  y  $\{r_{tk}\}$  sigan algún patrón teórico que permita identificar los valores de  $p$ ,  $d$  y  $q$  (véase Guerrero, 1991).

### 2.5.5.2 Estimación

Una vez que se ha identificado un modelo para la serie, se procede a estimar los parámetros. La estimación de los parámetros no resulta fácil pues para obtener la representación más adecuada de la serie  $\{\omega_t = \nabla^d T(Z_t)\}$ , donde  $T(Z_t)$  es la serie transformada y  $Z_t$  es la serie no estacionaria, es necesario maximizar una función de verosimilitud que surge del supuesto de normalidad de los errores y es de la forma:

$$L(\phi, \theta_0, \theta, \sigma_a^2 | \omega) = 2\pi^{-\frac{(N-d)}{2}} \sigma_a^{-N-d} \exp\left\{\frac{-S(\phi, \theta_0, \theta)}{2\sigma_a^2}\right\}$$

en donde

$$\begin{aligned} S(\phi, \theta_0, \theta) &= \sum_{t=d+1}^N \alpha_t^2 & (2.5.5.2.1) \\ &= \sum_{t=d+1}^N (\omega_t - \phi_{1,t-1} - \dots - \phi_{p,t-p} - \theta_0 + \theta_1 \alpha_{t-1} + \dots + \theta_q \alpha_{t-q}) \end{aligned}$$

Para maximizar esta función es necesario minimizar la suma de cuadrados anterior después elegir un estimador de  $\sigma_a^2$ . Box y Jenkins (1978) sugieren un método de estimación no lineal de  $\phi$ ,  $\theta_0$  y  $\theta$  basado en el algoritmo de Marquard que no sólo proporciona una estimación puntual de los parámetros, sino que también proporciona intervalos de confianza. Este método se basa en el desarrollo de series de Taylor para linealizar  $\alpha$ , condicionada en el conocimiento de  $\omega$ , y de valores iniciales para todos los parámetros. Debe tenerse en cuenta que este proceso es iterativo y por lo tanto, es muy sensible a los valores iniciales de los parámetros, por lo que en la etapa de identificación se debe tener cuidado en elegir los valores correctos para alcanzar la convergencia (Guerrero, 1991).

### 2.5.5.3 Verificación

El modelo que se estima se fundamenta en una serie de supuestos, y la verificación es la etapa en la que se detectan las posibles violaciones de ellos, a través del análisis de residuales.

Los supuestos que se deben verificar son:

1.  $\{\alpha_t\}$  tiene media cero.

Si esto no sucede puede deberse a que:

- i) existe una tendencia determinista la cual puede considerarse explícitamente y se denota por  $\theta_0$ , o bien,
- ii) es necesario considerar una diferencia más; sin embargo hay que tener cuidado en no sobrediferenciar la serie.

La verificación de este supuesto se puede hacer construyendo un cociente de la media aritmética entre la desviación estándar de los residuales:

$$\frac{\sqrt{N-d}m(\hat{a})}{\hat{\sigma}_a} \quad (2.5.5.3.1)$$

el cual, si su valor absoluto es menor que dos, no hay evidencia de que la media sea distinta de cero; lo contrario indica que este supuesto no se cumple.

2.  $\{a_i\}$  tiene varianza constante.

En la gráfica de los residuales puede detectarse algún patrón de crecimiento o decrecimiento de la varianza.

3. Las  $\{a_i\}$  son variables aleatorias independientes entre sí.

Para comprobar que este supuesto se cumple se calcula el estadístico  $Q$  de Box y Pierce, haciendo una prueba conjunta de  $k$  autocorrelaciones:

$$Q = (N-d) \sum_{i=1}^k f_i^2(\hat{a}) \quad (2.5.5.3.2)$$

el cual, para  $k$  suficientemente grande se distribuye como una  $\chi^2$  con  $k-p-q$  grados de libertad, en donde  $k$  es el número de autocorrelaciones,  $p$  es el número de parámetros autorregresivos y  $q$  es el número de parámetros de promédios móviles.

Por lo tanto, si  $Q$  es mayor que una  $\chi^2_{(k-p-q)}$  de tablas, se rechaza la hipótesis nula de que las variables aleatorias  $\{a_i\}$  son mutuamente independientes.

4.  $\{a_i\}$  se distribuye normalmente para toda  $t$ .

En la gráfica de los residuales se trazan unas bandas de confianza de  $(-2\hat{\sigma}_a, 2\hat{\sigma}_a)$ , y se espera que aproximadamente el 95% de los residuales queden dentro de la banda de confianza.

5. No existen observaciones aberrantes.

En la gráfica de los residuales se traza una banda de  $(-3\hat{\sigma}_a, 3\hat{\sigma}_a)$ , y los residuales que queden fuera pueden corresponder a una observación que no fue generada por el proceso.

Otros supuestos que se verifican sin utilizar el análisis de residuales son:

6. Parsimonia del modelo

Se construyen intervalos de confianza para cada uno de los parámetros y si el intervalo escogido contiene al cero, este parámetro deberá cancelarse para, de esta manera, considerar el mínimo necesario de ellos.

7. Admisibilidad del modelo.

Se verifican las condiciones de estacionariedad e invertibilidad correspondientes a los polinomios de retraso que intervengan.

8. Estabilidad del modelo

Se calculan las correlaciones entre los parámetros y si son altas, alguno de ellos tendrá que suprimirse.

Si se viola alguno de estos supuestos, es necesario proponer un nuevo modelo y regresar a la etapa de identificación. Este proceso se repite hasta que se encuentre un modelo que no viole los supuestos, o al menos no tan notoriamente, y éste será el que pase a la siguiente etapa.

### 2.5.6 Pronóstico

El criterio que se utiliza para el cálculo de los pronósticos es el de minimizar el error cuadrático medio

$$E[T(Z_{t,t}) - \hat{T}(Z_t)(l)]^2 = \min_{\hat{T}(Z_t)(l)} E[T(Z_{t,t}) - \hat{T}(Z_t)(l)]^2 \quad (2.5.6.1)$$

en donde  $\hat{T}(Z_t)(l)$  es el pronóstico del valor  $T(Z_{t,t})$  calculado a partir del origen en  $t$  y representa a cualquier otro pronóstico lineal de  $T(Z_{t,t})$ .

Se puede demostrar que  $E[T(Z_{t,t})] = \hat{T}(Z_t)(l)$  y que éste es el pronóstico con error cuadrático medio mínimo. El error de este pronóstico está dado por

$$e_t(l) = T(Z_{t,t}) - \hat{T}(Z_t)(l) = \sum_{j=0}^{l-1} a_{t,t-j} \quad (2.5.6.2)$$

ya que

$$E[e_t(l)] = 0$$

por lo que el pronóstico es insesgado, y además se tiene que

$$Var[e_t(l)] = \sum_{j=0}^{l-1} \psi_j^2 \sigma_a^2 \quad (2.5.6.3)$$

en donde las ponderaciones  $\hat{\psi}_1, \dots, \hat{\psi}_{l-1}$  surgen de la representación  $\hat{T}(Z_t) = \hat{\psi}(B)a_t$  y se obtienen al igualar los coeficientes del operador  $B$  en ambos lados de la ecuación:

$$(1 - \hat{\theta}_1 B - \dots - \hat{\theta}_{p,d} B^{p+d})(1 - \hat{\psi}_1 B - \hat{\psi}_2 B^2 - \dots) = 1 - \hat{\theta}_1 B - \dots - \hat{\theta}_q B^q \quad (2.5.6.4)$$

y se utiliza la forma recursiva:

$$\hat{\psi}_j = \begin{cases} \hat{\theta}_j + \hat{\phi}_1 \hat{\psi}_{j-1} + \hat{\phi}_2 \hat{\psi}_{j-2} + \dots + \hat{\phi}_{p-d} \hat{\psi}_{j-p-d} & \text{para } j = 1, 2, \dots, q \\ \hat{\phi}_1 \hat{\psi}_{j-1} + \hat{\phi}_2 \hat{\psi}_{j-2} + \dots + \hat{\phi}_{p-d} \hat{\psi}_{j-p-d} & \text{para } j > q \end{cases} \quad (2.5.6.5)$$

con  $\hat{\psi}_0 = -1$  y  $\hat{\psi}_{j-1} = 0$  si  $j < 1$ .

Con estos resultados se puede decir que  $\{e_t(l) | Z_t, Z_{t-1}, \dots\}$  se distribuye como  $N(0, \text{Var}[e_t(l)])$  y que:

$$\{T(Z_{t,i}) | Z_t, Z_{t-1}, \dots\} \sim N(\hat{T}(Z_t)(l), \text{Var}[e_t(l)])$$

entonces se pueden construir intervalos con límites al  $100(1-\alpha)\%$  de confianza para  $T(Z_{t,i})$  y condicionados al conocimiento del comportamiento histórico de la serie. El intervalo esta dado por:

$$\hat{T}(Z_t)(l) \pm Z_{\frac{\alpha}{2}} \sqrt{\sum_{j=0}^{l-1} \psi_j^2 \sigma_a} \quad (2.5.6.6)$$

en donde  $Z_{\frac{\alpha}{2}}$  es el punto porcentual tal que

$$P[Z > Z_{\frac{\alpha}{2}}] = \frac{\alpha}{2}$$

cuando  $Z$  se distribuye como  $N(0,1)$ .

Para actualizar los pronósticos con la información observada que se va incorporando a la serie, se usa la expresión:

$$\hat{T}(Z_{t,i}(l)) = \hat{T}(Z_t)(l+1) \psi_l a_{t,i} \quad l \geq 1 \quad (2.5.6.7)$$

la cual surge del siguiente procedimiento:

Supóngase que se han calculado los  $l$  pronósticos  $\hat{T}(Z_N)(1), \hat{T}(Z_N)(2), \dots, \hat{T}(Z_N)(l)$  y que el dato  $T(Z_{N+1})$  ya se observó, entonces el residual es:

$$a_{N+1} = T(Z_{N+1}) - \hat{T}(Z_N)(1) \quad (2.5.6.8)$$

y los siguientes pronósticos tendrán su origen en  $N+1$  y serán:

$$\begin{aligned} \hat{T}(Z_{N+1})(1) &= \hat{T}(Z_N)(2) - \psi_1 a_{N+1} \\ \hat{T}(Z_{N+1})(2) &= \hat{T}(Z_N)(3) - \psi_2 a_{N+1} \\ \hat{T}(Z_{N+1})(l-1) &= \hat{T}(Z_N)(l) - \psi_{l-1} a_{N+1} \end{aligned} \quad (2.5.6.9)$$

Hasta aquí se ha visto la manera en que se calculan los pronósticos para la serie transformada  $\{T(Z_t)\}$ ; pero como los datos que finalmente serán los requeridos son los pronósticos de la serie original  $\{Z_t\}$ , es preciso tener en cuenta que las propiedades óptimas de pronóstico  $\hat{T}(Z_t)(l)$  no necesariamente se preservan si la transformación de los datos originales no es lineal y se utiliza como pronóstico de  $Z_{t+l}$  a  $T^{-1}[\hat{T}(Z_t)(l)]$ . Debido a que en la práctica es muy común utilizar la transformación potencia, se calcula el siguiente factor que permite reducir el sesgo que se genera cuando se aplica la transformación inversa para regresar a la escala original:

$$\hat{e}_{t,\lambda}(l) = \left\{ \frac{1}{2} \left( 1 + \sqrt{1 + \frac{2\lambda(\lambda-1) \text{Var}[e_t(l)]}{(1+\lambda \hat{T}(Z_t)(l))^2}} \right) \right\}^{\frac{1}{\lambda}} \quad (2.5.6.10)$$

si  $\lambda \leq 1, \neq 0$ , y

$$\hat{e}_{t,\lambda}(l) = \exp \left\{ \frac{\text{Var}[e_t(l)]}{2} \right\}$$

si  $\lambda = 0$ .

Asimismo, los intervalos de confianza que ya incluyen la corrección por sesgo se calculan como:

$$\hat{c}_{t,\lambda}(t) = T^{-1} \left\{ T(Z_t)(t) \pm Z_{\frac{\lambda}{2}} \sqrt{\text{Var}[e_t(t)]} \right\} \quad (2.5.6.11)$$

De esta manera ya se tienen los elementos necesarios para obtener los pronósticos de la serie de tiempo.

### 2.5.7 Modelos estacionales

Una serie estacional es aquella que muestra fluctuaciones que se repiten anualmente, aunque con cambios graduales a través de los años. Debido a que un gran número de series de tiempo económicas muestran este comportamiento, en esta sección se mencionan los operadores necesarios para obtener series estacionarias a partir de series estacionales.

Se define al operador de diferencia estacional,  $\nabla_E^k$  como:

$$\begin{aligned} \nabla_E^D Z_t &= (1 - B^E)^D Z_t \\ &= \sum_{j=0}^D \frac{D!}{j!(D-j)!} (-1)^j Z_{t-jE} \end{aligned} \quad (2.5.7.1)$$

para  $D=1, 2, \dots$  y  $E=1, 2, \dots$ . No debe perderse de vista el hecho de que al aplicar este operador, se pierden  $DE$  observaciones de la serie original.

Por su parte, el polinomio de retraso estacional de orden  $k$  está dado por:

$$\begin{aligned} G(B^E) &= 1 - g_1 B^E - g_2 B^{2E} - \dots - g_k B^{kE} \\ &= 1 - \sum_{j=1}^k g_j B^{jE} \end{aligned}$$

Con esta notación se puede obtener la representación puramente estacional del tipo ARIMA(P,D,Q)<sub>E</sub>:

$$\phi(B^E)\nabla_E^D(Z_t - \mu) = \theta(B^E)\alpha_t \quad (2.5.7.2)$$

en donde  $\mu$  es la media de  $\{Z_t\}$ ,  $\phi(B^E)$  representa a un polinomio autorregresivo estacional de orden P,  $\theta(B^E)$  es un polinomio de promedios móviles estacional de orden Q y la serie  $\{\alpha_t\}$  es ruido blanco.

Finalmente, el modelo general que incorpora efectos estacionales y no estacionales, es decir en donde se tienen en cuenta posibles relaciones entre los datos dentro de los años y entre los años para los mismos periodos, es:

$$\phi(B^E)\nabla_E^D(Z_t - \mu) = \theta(B^E)\alpha_t$$

en donde la serie  $\{\alpha_t\}$  se genera a través de un proceso ARIMA (p,d,q)

$$\phi(B)\nabla^d\alpha_t = \theta(B)\alpha_t$$

con  $\{\alpha_t\}$  ruido blanco. De las dos expresiones anteriores, se obtiene el modelo multiplicativo estacional:

$$\phi(B)\phi(B^E)\nabla^d\nabla_E^D(Z_t - \mu) = \theta(B)\theta(B^E)\alpha_t$$

y que se denota ARIMA (p,d,q)×(P,D,Q)<sub>E</sub>. Obviamente el análisis de autocorrelación se complica y los posibles modelos a identificar se multiplican. Los pasos que se siguen para la construcción de estos modelos y la obtención de los pronósticos son los mismos que para los modelos no estacionales; sin embargo más detalles se pueden consultar en el libro de Guerrero (1991).

## 2.6 Modelos de función de transferencia

Los modelos de función de transferencia han probado ser una herramienta útil, tanto como métodos de pronóstico, como de explicación de modelos y teorías económicas.

Con estos modelos se pretende tener en cuenta explícitamente los efectos que una o más variables explicativas -exógenas- producen sobre la variable respuesta -endógena- en estudio.

Supóngase que  $\{Y_t\}$  y  $\{X_t\}$  son las series endógena y exógena, respectivamente, las cuales se pretende relacionar para explicar el comportamiento de la variable  $Y$  en respuesta a la variable  $X$ . El modelo a considerar es:

$$y_t = v(B)x_t + \eta_t \quad (2.6.1)$$

en donde:

$y_t$  y  $x_t$  son las series estacionarias con media cero, correspondientes a  $\{Y_t\}$  y  $\{X_t\}$ , respectivamente;

$v(B)$  es la función de transferencia; y

$\eta_t$  es el componente de error aleatorio.

A este modelo se le conoce en Econometría como modelo de rezagos distribuidos si  $\{\eta_t\}$  es un serie de ruido blanco, y podría pensarse en trabajarlo como tal con las técnicas de regresión; sin embargo, es muy posible que se presenten los siguientes problemas:

- i) Dependencia de las  $\eta$ 's, que ocasiona subestimación de la varianza de los coeficientes de regresión y sesga las pruebas  $t$  y  $F$  propiciando que muchos estimadores sean significativos.
- ii) Multicolinealidad, que afecta la estimación de los coeficientes de regresión, ya que se debe a la autocorrelación de la serie exógena  $\{X_t\}$

iii) Identificación del rezago, cuando se desconoce el número de rezagos que deben aparecer en el modelo y ocasiona que no todos los efectos de  $\langle x_t \rangle$  se reflejen en  $\langle y_t \rangle$ .

Con el fin de atacar el primer problema el método que se ha propuesto es combinar el análisis de regresión con el de series de tiempo (Pyndick y Rubinfeld, 1976) de la siguiente manera:

10. estimar el modelo de regresión mediante mínimos cuadrados ordinarios;
20. identificar un modelo ARIMA para el término de error del modelo; y
30. estimar simultáneamente tanto los parámetros de regresión como los del modelo ARIMA.

Para resolver el segundo problema se pueden ensayar las siguientes alternativas (Gujarati, 1992), sin olvidar que el éxito depende de la severidad del problema de multicolinealidad y de la naturaleza de los datos:

10. obtener información a priori de las series altamente colineales;
20. obtener datos nuevos o adicionales de las series altamente colineales;
30. hacer una transformación adecuada de las variables del modelo;
40. replantear el modelo si existe sesgo de especificación; y
50. utilizar las técnicas del análisis factorial y de componentes principales.

Por último, para resolver el tercer problema existen varios métodos, pero el que ha resultado más apropiado y no se enfrenta a los problemas anteriores es el enfoque de función de transferencia.

En la siguiente sección se proporcionan los elementos necesarios para comprender la naturaleza de los modelos de función de transferencia, y posteriormente se presenta el proceso que se sigue para la construcción de dichos modelos.

### 2.6.1 Elementos para la construcción de los modelos de función de transferencia

Para poder utilizar los modelos de función de transferencia es necesario definir lo que es una función de transferencia. Asimismo, hay que derivar los elementos que intervienen en la identificación de ésta y que surgen de los errores que se generan de la aplicación de modelos ARIMA univariados a cada una de las series endógena y exógena y de su interrelación: la función de correlación cruzada y la función de impulso-respuesta.

#### 2.6.1.1 Función de transferencia

Al polinomio de retraso

$$v(B) = v_0 + v_1 B + v_2 B^2 + \dots \quad (2.6.1.1.1)$$

se le conoce como función de transferencia del filtro lineal que transforma a  $\{\tilde{x}_t\}$  en  $\{\tilde{y}_t\}$ , mientras que los valores específicos  $v_0, v_1, v_2, \dots$ , forman la función de impulso-respuesta del proceso.

El modelo de transferencia es un polinomio con un número infinito de términos, y se puede reparametrizar para obtener una representación más parsimoniosa lo cual se logra si se supone que  $v(B)$  es un polinomio racional del tipo:

$$v(B) = \frac{\omega(B)}{\delta(B)} B^b = \frac{\omega_0 - \omega_1 B - \dots - \omega_r B^r}{1 - \delta_1 B - \dots - \delta_r B^r} B^b \quad (2.6.1.1.2)$$

en donde  $B^b$  aparece para indicar un posible retraso de  $b$  periodos antes de que el efecto de  $X$  pueda apreciarse en  $Y$ . Entonces, el modelo de función de transferencia se convierte en:

$$y = \frac{\omega(B)}{\delta(B)} x_{t-b} + \eta_t \quad (2.6.1.1.3)$$

en donde, para que  $\{y_t\}$  sea estacionaria, se requiere que las raíces del polinomio  $\rho(B) = 1 - \delta_1 B - \dots - \delta_r B^r = 0$  estén fuera del círculo unitario.

### 2.6.1.2 Término estocástico de error

En el modelo de transferencia se supone que el componente estocástico de error  $\eta_t$  es independiente del nivel de la serie  $\{x_t\}$  y aditivo con respecto a la influencia de  $x$ . Además, se supone que  $\eta_t$  admite la representación ARIMA  $(p, d, q)$ :

$$\phi(B) \nabla^d \eta_t = \theta(B) a_t \quad (2.6.1.2.1)$$

en donde  $\{a_t\}$  es un proceso de ruido blanco con media cero.

Por tanto, el modelo de función de transferencia se puede expresar como

$$y = \frac{\omega(B)}{\delta(B)} X_{t-b} + \frac{\theta(B)}{\phi(B)} \nabla^d a_t \quad (2.6.1.2.2)$$

que en Econometría se conoce como un modelo de regresión dinámica (Pindick, Rubinfeld 1978).

### 2.6.1.3 Función de correlación cruzada

Generalmente, la relación entre dos series de tiempo no se puede hacer directamente. Se ha encontrado que pueden tenerse correlaciones altas entre dos series independientes y estacionarias; este fenómeno se conoce como correlación espuria ya que se origina por variaciones aleatorias en las series (Granger y Newbold, 1974).

Para evitar lo anterior se recomienda trabajar con series preblanqueadas (Jenkins, 1979), es decir, se construye el modelo ARIMA para la serie  $\{\tilde{x}_t\}$

$$\phi_x(B)\tilde{x}_t = \theta_x(B)\alpha_t$$

del que se obtiene

$$\alpha_t = \phi_x(B)\theta_x^{-1}(B)\tilde{x}_t \quad (2.6.1.3.1)$$

el cual es un proceso que transforma a la serie autocorrelacionada  $\{\tilde{x}_t\}$  en una serie no correlacionada de ruido blanco  $\{\alpha_t\}$ . Cuando este modelo se aplica a la serie  $\{\tilde{y}_t\}$  se tiene:

$$\phi_x(B)\tilde{y}_t = \theta_x(B)\beta_t \quad (2.6.1.3.2)$$

es decir:

$$\beta_t = \phi_x(B)\theta_x^{-1}(B)\tilde{y}_t \quad (2.6.1.3.3)$$

que transforma a la serie autocorrelacionada  $\{\tilde{y}_t\}$  en una serie  $\{\beta_t\}$  que no necesariamente es ruido blanco. Al sustituir  $\{\alpha_t\}$  y  $\{\beta_t\}$  en el modelo de transferencia se correlacionan las series estacionarias y preblanqueadas  $\{\tilde{x}_t\}$  y  $\{\tilde{y}_t\}$  a través de la función de correlación cruzada.

La función de correlación cruzada (FCC) se calcula sustituyendo  $\{\alpha_t\}$  y  $\{\beta_t\}$  en el modelo de transferencia, y se tiene:

$$\beta_t = v(B)\alpha_t + \epsilon_t, \quad (2.6.1.3.4)$$

con  $\epsilon_t = \psi_x(B)O_x^{-1}(B)\eta_t$ . Por lo tanto:

$$\begin{aligned} E[\beta_t \alpha_{t-k}] &= E[v_0 \alpha_t \alpha_{t-k} + \dots + v_k \alpha_{t-k} \alpha_{t-k} + \dots] + E[\epsilon_t \alpha_{t-k}] \\ &= v_k E[\alpha_{t-k}^2] \end{aligned}$$

o sea que

$$v_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2}, \quad k = 0, 1, 2, \dots \quad (2.6.1.3.5)$$

es decir, la  $k$ -ésima ponderación es el cociente de la covarianza cruzada en el retraso  $k$  entre  $\alpha$  y  $\beta$ . Por consiguiente, si se hace uso de correlaciones cruzadas más que de covarianzas cruzadas, se tiene

$$v_k = \rho_{\alpha\beta}(k) \frac{\sigma_\beta}{\sigma_\alpha} \quad k = 0, 1, 2, \dots \quad (2.6.1.3.6)$$

con  $\rho_{\alpha\beta}(k) = \gamma_{\alpha\beta}(k) / \sigma_\alpha \sigma_\beta$

La estimación de  $\{v_k\}$  se convierte en un problema de estimación de  $\sigma_\alpha$ ,  $\sigma_\beta$  y de  $\gamma_{\alpha\beta}(k)$ ,  $h = 0, 1, 2, \dots$

Estimadores de  $\rho_{\alpha\beta}(k)$  y de  $v_k$  son:

$$\hat{\rho}_{\alpha\beta}(k) = \sqrt{\frac{\sum_{i=1}^N (\hat{\alpha}_i - \bar{\alpha})(\hat{\beta}_i - \bar{\beta})}{\sum_{i=1}^N (\hat{\alpha}_i - \bar{\alpha})^2 (\hat{\beta}_i - \bar{\beta})^2}} \quad k = 0, 1, 2, \dots \quad (2.6.1.3.7)$$

$$\hat{v}_k = \hat{\rho}_{\alpha\beta}(k) \sqrt{\frac{\sum_{i=1}^N (\hat{\alpha}_i - \bar{\alpha})^2}{\sum_{i=1}^N (\hat{\beta}_i - \bar{\beta})^2}} \quad k = 0, 1, 2, \dots \quad (2.6.1.3.8)$$

En particular, la estimación de  $v_k$  es ineficiente debido a que se desconoce la forma del modelo; sin embargo, es útil para proporcionar la guía de un modelo que sea razonable de postularse.

Es importante tomar en cuenta que  $\hat{\rho}_{\alpha\beta}(k)$  está sujeta a fluctuaciones meramente muestrales, por lo que es conveniente conocer la varianza del coeficiente de correlación cruzada muestral,  $v_k$ :

$$\text{Var}[\hat{\rho}_{\alpha\beta}(k)] \approx \frac{1}{N-k} \quad (2.6.1.3.9)$$

y los coeficientes de correlación cruzada podrán compararse contra la desviación estándar aproximada, dada por  $1/\sqrt{N-k}$ , es decir, aquellos valores  $\hat{v}_k$  mayores a 2 veces la desviación estándar se pueden considerar significativamente distintos de cero.

#### 2.6.1.4 Función de impulso-respuesta

La identificación de la función de transferencia debe considerar las ponderaciones de impulso respuesta asociadas

con los valores de las  $\delta$ 's y de las  $\omega$ 's a través de la ecuación en diferencias con coeficientes B

$$\delta(B)v(B) = \omega(B)B^b \quad (2.6.1.4.1)$$

o sea

$$(1 - \delta_1 B - \dots - \delta_r B^r)(v_0 + v_1 B + \dots) = (\omega_0 - \omega_1 B - \dots - \omega_s B^s)B^b$$

de donde

$$v_0 + (v_1 - v_0 \delta_1)B + (v_2 - v_1 \delta_1 - v_0 \delta_2)B^2 + \dots - \omega_0 B^b - \omega_1 B^{b+1} - \omega_2 B^{b+2} - \dots - \omega_s B^{b+s}$$

Por lo tanto

- (1)  $v_j = 0$   $j < b$
- (2)  $v_j = v_{j-1} \delta_1 + v_{j-2} \delta_2 + \dots + v_{j-r} \delta_r + \omega_0$   $j = b$
- (3)  $v_j = v_{j-1} \delta_1 + v_{j-2} \delta_2 + \dots + v_{j-r} \delta_r - \omega_{j-b}$   $j = b+1, \dots, b+s$
- (4)  $v_j = v_{j-1} \delta_1 + v_{j-2} \delta_2 + \dots + v_{j-r} \delta_r$   $j > b+s$

De esta manera, el comportamiento de la función de impulso-respuesta consiste de cuatro grupos de ponderaciones:

- i)  $v_0, v_1, \dots, v_{b-1}$ , b valores iguales a cero, es decir, valores que no son significativamente distintos de cero al compararlos con su desviación estándar.
- ii)  $v_b, v_{b+1}, \dots, v_{b+s-r}$ , s-r+1 valores que no siguen un patrón determinado para s>r y que no existen para s<r; el valor  $v_b$  está dado por (2).
- iii)  $v_{b+s-r+1}, \dots, v_{b+s}$ , r valores que proporcionan valores iniciales para la ecuación en diferencia de orden r dada por (3).

- iv)  $v_{b,s+1}, v_{b,s+2}, \dots, b+s+m$  valores que siguen un patrón dictado por la ecuación en diferencia de orden  $r$  dada por (4),  $m \in [1, \infty)$ .

Los modelos más sencillos para la función de transferencia pueden ser adecuados para representar una gran cantidad de comportamientos, así que la parsimonia debe ser siempre un criterio importante al identificar un modelo (Guerrero, 1991). Los patrones más comunes en la identificación de la transferencia son:

- a)  $r=0$

En una gráfica de las  $v_i$  se verá un patrón geométrico proporcional en el que  $s+1$  valores son significativamente distintos de cero. Los valores de las ponderaciones  $v_i$ , así como la función de transferencia se muestran en la siguiente tabla:

Valor de $s$	Ponderación	Función de transferencia
$s=0$	$v_b = \omega_0$	$\hat{Y} = \omega_0 B^s \hat{X}_t$
$s=1$	$v_b = \omega_0$ $v_{b,1} = -\omega_1$	$\hat{Y}_t = (\omega_0 - \omega_1 B) B^s \hat{X}_t$
$s=2$	$v_b = \omega_0$ $v_{b,1} = -\omega_1$ $v_{b,2} = -\omega_2$	$\hat{Y}_t = (\omega_0 - \omega_1 B - \omega_2 B^2) B^s \hat{X}_t$
...		
$s+1$	$v_b = \omega_0$ $v_{b,1} = -\omega_1$ ... $v_{b,s} = -\omega_s$	$\hat{Y}_t = (\omega_0 - \omega_1 B - \dots - \omega_s B^s) B^s \hat{X}_t$

- b)  $r=1$

En la gráfica de las  $v_i$  se verá un patrón que oscila senoidalmente a cero a partir de los valores iniciales

$v_{b-a-r+1}$  y  $v_{b-a-r+2}$  y se tendrán  $s-r+1$  que no siguen un patrón determinado. La función de transferencia para distintos valores de  $s$  se presentan en la próxima tabla.

Para  $s=0$  la flexibilidad del modelo es limitada porque el primer valor inicial de las  $v_i$  es cero; y para  $s=2$   $v_b = \omega_0$  no sigue un patrón fijo.

Cabe señalar que aunque la función de transferencia sea de orden  $r$  o  $s$ , algunos de los parámetros  $\delta$ 's u  $\omega$ 's de orden menor pueden llegar a valer cero, debido a que no son necesarios en el modelo.

Valor de $s$	Las ponderaciones oscilan a partir de:	Función de transferencia
$s=0$	$v_{b,1} = 0$ $v_b = \omega_0$	$(1 - b_1 B \delta_2^2) \hat{y}_t = \omega_0 B^b \hat{x}_t$
$s=1$	$v_b = \omega_0$ $v_{b,1} = \delta_1 \omega_0 - \omega_1$	$(1 - b_1 B \delta_2^2) \hat{y}_t = (\omega_0 - \omega_1 B) B^b \hat{x}_t$
$s=2$	$v_{b,1} = \delta_1 \omega_0 - \omega_1$ $v_{b,2} = (\delta_1^2 - \delta_2) \omega_0 - \delta_1 \omega_1 - \omega_2$	$(1 - b_1 B \delta_2^2) \hat{y}_t = (\omega_0 - \omega_1 B - \omega_2 B^2) B^b \hat{x}_t$

### 2.6.2 Construcción de los modelos de función de transferencia

La construcción de modelos de función de transferencia sigue esencialmente las etapas de la estrategia propuesta por Box y Jenkins (1970) para construir modelos de series de tiempo en general, esto es: identificación, estimación, verificación y uso del modelo.

Para construir un modelo de función de transferencia es necesario haber construido los modelos univariados de las series exógena y endógena, este último como base de comparación,

sobre todo si lo que se pretende es mejorar los pronósticos de una serie incorporando información adicional a su propia historia.

### 2.6.2.1 Identificación

El proceso que se sigue para indentificar un posible modelo de función de transferencia es el siguiente:

10. Estacionarizar las series  $\{X_t\}$  y  $\{Y_t\}$ , es decir, obtener las series estacionarias:

$$\begin{aligned}x_t &= \nabla^{d_0} \nabla_{E_0}^{D_0} T_0(X_t) - \bar{X} \\y_t &= \nabla^{d_1} \nabla_{E_1}^{D_1} T_1(Y_t) - \bar{Y}\end{aligned}\quad (2.6.2.1.1)$$

20. Construir los modelos univariados para  $\{X_t\}$  y  $\{Y_t\}$ :

$$\begin{aligned}\phi_x(B)\hat{x}_t &= \theta_x(B)\alpha_t \\ \phi_y(B)\hat{y}_t &= \theta_y(B)\gamma_t\end{aligned}\quad (2.6.2.1.2)$$

en donde  $\{\alpha_t\}$  y  $\{\gamma_t\}$  son series de ruido blanco.

30. Con el fin de analizar primero la estructura intrínseca de las series antes de analizar la estructura que ligue a las series entre sí, la serie  $y_t$  se preblanquea con el modelo univariado construido para  $x_t$ , y se tiene:

$$\begin{aligned}\phi_x(B)\hat{x}_t &= \theta_x(B)\alpha_t \\ \phi_x(B)\hat{y}_t &= \theta_x(B)\beta_t\end{aligned}\quad (2.6.2.1.3)$$

Las series de residuales  $\{\hat{\alpha}_i\}$  y  $\{\hat{\beta}_i\}$  se utilizan para determinar las ponderaciones  $v_i$  que surgen de la función de correlación cruzada 2.6.1.3.8.

- 4o. Identificación de la función de transferencia a partir de la FCC entre  $\{\hat{\alpha}_i\}$  y  $\{\hat{\beta}_i\}$ , es decir, se identifican los órdenes de los polinomios  $\omega(B)$  y  $\delta(B)$ ,  $r$  y  $s$  respectivamente, así como el retraso  $b$ .
- 5o. Identificación del modelo ARIMA para  $\eta_i$ , o sea, determinar los valores de  $p$  y  $q$  para los polinomios  $\phi(B)$  y  $\theta(B)$  que rigen el comportamiento de  $\{\eta_i\}$ .

En este caso, el modelo univariado que se construyó para la serie  $\{y_i\}$  se puede usar como una primera aproximación para la serie  $\{\eta_i\}$  y se tendría:

$$\phi_y(B)\nabla^d \eta_i = \theta_y(B)\epsilon_i$$

con  $\{\epsilon_i\}$  ruido blanco.

### 2.6.2.2 Estimación

En esta etapa se estiman los parámetros del modelo

$$\hat{y}_i = \frac{\omega(B)}{\delta(B)} \hat{x}_{i-b} + \eta_i$$

o sea:

$$\delta(B)\hat{y}_i = \omega(B)\hat{x}_{i-b} + \frac{\theta(B)}{\phi(B)}\epsilon_i$$

y se calculan los residuales para someterlos a la verificación de los supuestos.

### 2.6.2.3 Verificación

Se verifica que el modelo satisface los supuestos de la función de transferencia y del modelo ARIMA correspondiente al proceso de ruido blanco.

### 2.6.2.4 Uso del modelo

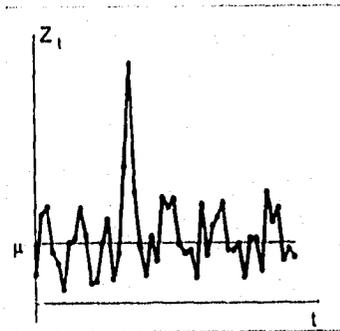
El modelo se utiliza para explicar el comportamiento y/o para pronosticar a la variable endógena. En esta etapa es donde se utiliza el modelo univariado de la variable endógena para comparar los resultados.

## 2.7 Análisis de intervención

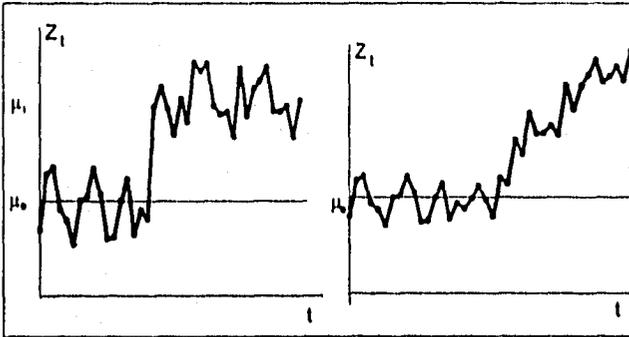
Las intervenciones son los mecanismos o políticas que llegan a afectar el comportamiento de una serie de tiempo, es decir, el término "intervención" se refiere a la ocurrencia de un evento exógeno al comportamiento histórico de la serie en estudio.

Las intervenciones pueden identificarse como:

- i) Una caída o elevación instantánea (pulso) de la serie en el momento en que se dá la intervención; o



ii) un cambio de nivel o tendencia de la serie (escalón).



Los modelos de intervención y los de función de transferencia son muy similares entre sí, de hecho son equivalentes: una función de transferencia puede emplearse para estudiar los efectos de una intervención sobre la variable de interés, suponiendo que fuera posible tener la serie exógena que origina el cambio de la serie en estudio y que su efecto pueda representarse en una forma equivalente a la ecuación 2.6.1.1.1 a través del siguiente un modelo:

$$f(\delta, \omega, X, t) = \sum_{j=1}^k Y_{ij} = \sum_{j=1}^k \frac{\omega_j(B)}{\delta_j(B)} X_{ij} \quad (2.7.1)$$

en donde :

$k$  es el número de intervenciones;

$Y_{ij}$  es la transferencia dinámica;

$\omega_j(B)$  y  $\delta_j(B)$  son polinomios de retraso de grado  $s_j$  y  $r_j$ , respectivamente y  $\omega_j(B)$  tiene sus raíces fuera del círculo unitario; y

**ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA**

$X_t$ , es la variable artificial que sustituye a la variable exógena que origina la intervención.

La forma general del modelo de intervención es:

$$T(Z_t) = Y_t + N_t \quad (2.7.2)$$

en donde  $Y_t$  acepta la representación 2.7.1 y  $N_t$  es la parte aleatoria que se representa como un modelo estocástico de la forma:

$$\phi(B)N_t = \theta(B)\alpha_t$$

Por lo tanto, la ecuación 2.7.2 se puede representar como:

$$T(Z_t) = \frac{\omega(B)}{\delta(B)}X_t + \frac{\theta(B)}{\phi(B)}\alpha_t \quad (2.7.3)$$

Por otro lado,  $Y_t$  también puede representarse como:

$$\delta(B)Y_t = \omega(B)X_t \quad (2.7.4)$$

Para el caso de una intervención de pulso se tendría:

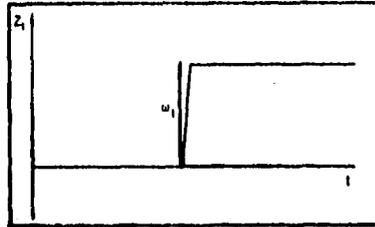
$$X_t = P_{t,1} = \begin{cases} 0 & \text{si } t \neq 1 \\ 1 & \text{si } t = 1 \end{cases} \quad (2.7.5)$$

y para la intervención de escalón:

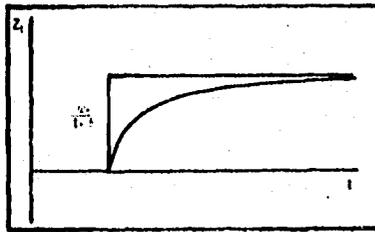
$$X_t = S_{t,1} = \begin{cases} 0 & \text{si } t < 1 \\ 1 & \text{si } t \geq 1 \end{cases} \quad (2.7.6)$$

De lo anterior es claro que pueden proponerse una amplia variedad de modelos de intervención ya que para diferentes órdenes  $r$  y  $s$ , existe todo un modelo dinámico de intervención. Sin embargo, en la práctica sólo se usan formas funcionales muy sencillas. Los modelos de intervención más simples son:

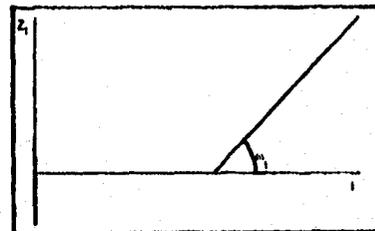
- 1) Cambio de nivel de la serie, que puede cuantificarse al estimar  $\omega_1$ .



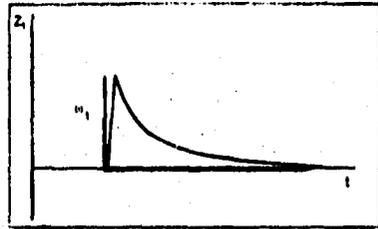
- 2) La intervención origina un crecimiento asintótico de la serie, dado por  $\frac{\omega_1}{1-\delta}$ , hacia el nuevo nivel.



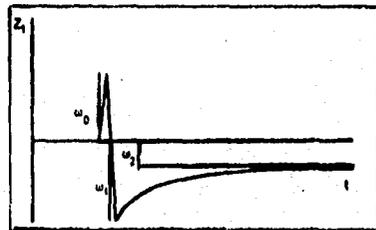
- 3) Caso particular del anterior, en donde la serie crece indefinidamente cuando  $\delta = 1$ .



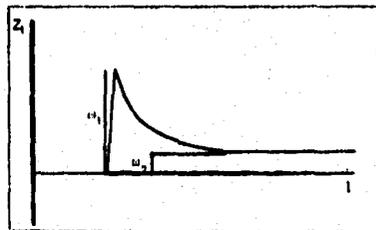
- 4) Después de la intervención, el nivel de la serie se recupera paulatinamente.



- 5) Combinación de los anteriores;  $\omega_1$  es el valor del primer pico y  $\omega_2$  es el nuevo nivel.



- 6)  $\omega_0$  es el tamaño del primer pulso y es a partir de  $\omega_1$  que la serie regresa gradualmente al nuevo nivel  $\omega_2$ .



Considerando únicamente las intervenciones de pulso, en la ecuación 2.7.3 puede sustituirse  $X_t$  por  $P_{t,l}$ :

$$Y_t = \frac{\omega(B)}{\delta(B)\nabla^b} P_{t,l} + \frac{\theta(B)}{\phi(B)\nabla^d} a_t$$

en donde

$$P_{t,l} = \begin{cases} 0, & t < l \\ \omega_0, & t = l \\ \delta^{t-l}\omega_0 - \delta^{t-l-1}\omega_1, & t > l \end{cases}$$

y  $\nabla^b$  es un polinomio de retraso que opera sobre  $Y_{t,l}$  para hacer que el efecto de la intervención no sea inmediato, pero que si se refleje en periodos posteriores.

### 2.7.1 Identificación

Mediante la verificación de los supuestos de un modelo ARIMA, construido con la técnica de Box-Jenkins, es posible identificar la intervención que está asociada a un cambio de la serie. Para ello, se estima el modelo en el periodo comprendido entre el inicio de la serie y una observación previa a la intervención y se calculan los pronósticos para verificar si el modelo propuesto continúa siendo el adecuado. Si existen discrepancias significativas entre los valores observados y los pronosticados puede deberse a que:

- a) en el modelo no se está considerando algún factor exógeno que origina que la serie cambie de nivel;
- b) no se está tomando en cuenta algún efecto estacional;
- c) alguno de los parámetros estimados cambió.

Guerrero (1991) propone una técnica gráfica respaldada con la prueba de Box-Tiao para demostrar que la discrepancia entre los valores observados y los pronósticos es real y para identificar las intervenciones y obtener los valores iniciales de los parámetros.

Para ello se identifican los residuales que son significativamente distintos de cero, construyendo los límites de confianza:

$$\pm 2\hat{\sigma}_e(t) = \pm 2\sqrt{\sum_{j=0}^{t-1} \psi_j^2 \hat{\sigma}_a} \quad (2.7.1.1)$$

con  $\psi_0 = 1$  y  $\psi_j$ , las ponderaciones calculadas con la expresión 2.5.6.5.

La gráfica de los errores puede indicar algún comportamiento que advierta la necesidad de incluir una intervención, y los valores iniciales de los parámetros serán el valor de los errores de pronóstico que resulten significativamente diferentes de cero.

Este proceso se lleva a cabo siempre y cuando se tenga conocimiento de que se introdujeron mecanismos o políticas que afectaron a la serie, y que se tengan observaciones posteriores a la fecha de la intervención.

Para saber si el comportamiento de la serie se vio alterado significativamente, se calcula el estadístico:

$$C = \sum_{h=1}^H e_{t-2+h}^2 / \hat{\sigma}_a^2 \quad (2.7.1.2)$$

con  $\hat{\sigma}_a^2$  la varianza de los residuales del modelo ARIMA que se construyó para el periodo previo a la intervención ( $t=I-1$ ), y  $e_{t-1}(1), e_{t-2}(1), \dots, e_{t-H-2}(1)$ , son los errores del pronóstico un

periodo hacia adelante, calculados a través de la expresión 2.5.6.9. Este estadístico se compara con una  $X^2_{(H)}$  de tablas, y si es mayor, se podrá considerar que la intervención ocurrida sí afectó el comportamiento de la serie

### 2.7.2 Estimación

Una vez identificado el tipo de intervención se estiman simultáneamente los parámetros que aparecen en la función de intervención y los parámetros del modelo ARIMA que se había construido previamente para el periodo desde el origen hasta la observación previa a la siguiente intervención, si la hay.

Al igual que en los modelos ARIMA, la estimación se realiza utilizando alguno de los paquetes que existen el mercado.

### 2.7.3 Verificación

En esta etapa se prueba que el modelo dinámico de intervención estimado es el adecuado para los periodos anterior y posterior a la intervención, es decir, si el modelo es el adecuado hasta antes de la intervención se espera que sea el adecuado después, al agregarle la función de intervención.

El modelo de intervención dinámico será el adecuado si los intervalos de confianza de los parámetros no contienen al cero o al uno, ya que si no, el modelo debe modificarse. Además se siguen básicamente los mismos pasos de verificación de los modelos ARIMA univariados.

Debido a que la función de intervención elegida inicialmente no es necesariamente la correcta, la construcción de modelos de intervención se debe pensar entonces como un proceso iterativo para encontrar el modelo adecuado.

#### 2.7.4 Uso del modelo

Una vez que se encuentra un modelo adecuado, entonces éste puede utilizarse para explicar el efecto de la intervención sobre el comportamiento de la serie de tiempo en estudio, o bien, para pronosticar su trayectoria futura. Los valores que se den a la variable exógena serán acordes con el tipo de función de intervención que se utilizó,  $P_{t,t}$  o  $S_{t,t}$ .

## Capítulo 3

### Pronósticos ARIMA condicionales

En este capítulo se presenta, como caso particular de la teoría de combinación de pronósticos, un método recursivo para generar pronósticos de series de tiempo que incorporan información adicional acerca de su probable evolución en el futuro. Aquí se vierte únicamente el procedimiento que propone Guerreo (1989) para calcular pronósticos ARIMA condicionales óptimos.

La herramienta que se propone permitirá obtener los pronósticos ARIMA de la demanda mensual de petrolíferos que estarán condicionados a los resultados anuales que se obtengan a través de regresión lineal.

#### 3.1 La teoría de la combinación de pronósticos

Frecuentemente se presenta el caso de que se obtienen pronósticos para una misma variable con métodos distintos y entonces es necesario establecer criterios para escoger el mejor. Pero al hacer esto se puede estar desperdiciando información útil que contengan los pronósticos rechazados y surge la pregunta: ¿porqué no combinarlos?. A simple vista, proceder de esta manera llevaría a un pronóstico más preciso basado en un conjunto de información más amplio.

Otra motivación para combinar pronósticos se presenta en situaciones prácticas, sobretodo en economía, cuando los

pronósticos que se obtienen de un modelo cuantitativo son modificados a juicio, empleando información de naturaleza cualitativa.

La teoría de la combinación de pronósticos, que trata de encontrar procedimientos óptimos para combinar diferentes pronósticos individuales del valor futuro de la misma variable, empezó a desarrollarse por Bates y Granger en 1969, quienes mostraron que una combinación lineal de pronósticos individuales, obtenidos cada uno con diferentes métodos, puede producir pronósticos mejores. Estudios subsecuentes se han enfocado en promedios aritméticos de los pronósticos individuales (Makridakis y Winkler, 1983) o promedios ponderados en donde los pesos se restringen a que sumen la unidad (Newbold y Granger, 1974) y posteriormente Granger y Ramanathan (1984) proponen utilizar regresión lineal para expresar el pronóstico combinado en función de los pronósticos individuales. En general, la idea es minimizar los errores del pronóstico asignando ponderaciones que tomen en cuenta las dependencias entre los pronósticos individuales y sus precisiones relativas, suponiendo que los errores se distribuyen normalmente. Las investigaciones siguientes se han enfocado fundamentalmente a revisar el trabajo de Granger y Ramanathan (Diebold 1986, Clemen 1986, Holmen 1987), a mejorar la eficiencia de los pronósticos combinados (Bunn, 1985), a presentar aplicaciones prácticas (Newbold y Zumwalt, 1987) y aplicar programación matemática para considerar otro tipo de aspectos relacionados con el pronóstico, introduciendo más de un objetivo (Gaytán, 1989).

Sin embargo, los métodos anteriores se enfocan a pronósticos individuales con la misma periodicidad, es decir, los pronósticos que se van a combinar, son necesariamente mensuales, trimestrales o anuales. Pero puede presentarse también el caso de que un método de pronóstico arroje mejores resultados si se utilizan series más agregados (por ejemplo un modelo de regresión múltiple para

series económicas que utiliza datos anuales), y otro requiera datos muy desagregados (series de tiempo que emplea datos mensuales o trimestrales). Ambos tienen una temporalidad diferente y los métodos tradicionales para combinar los pronósticos no son aplicables. Una solución a este problema es el método de pronósticos condicionales que propone Víctor M. Guerrero, mediante el cual se aplican ciertas restricciones a los pronósticos que se obtienen de un modelo ARIMA. El método anterior se convierte así en un procedimiento de combinación de pronósticos con diferente periodicidad.

### 3.2 El método de pronósticos condicionales

Esta metodología fue desarrollada por Guerrero en 1989, y tiene como antecedente en México el trabajo realizado por Pérez-Porrúa en 1984, donde se presenta un método recursivo sencillo para generar pronósticos más eficientes que los ARIMA como herramienta para sistematizar la "labor de adecuación de los pronósticos que comúnmente se hace en forma quizá demasiado empírica" y también para las tareas de seguimiento, evaluación y control de programas y presupuestos. De hecho, el trabajo de Guerrero formaliza las propiedades estadísticas de los pronósticos óptimos que, además de Pérez-Porrúa, se habían presentado implícitamente en trabajos publicados por Doan, Litterman y Sims en 1983 y por Cholette en 1982.

La base de este método es que se cuenta con una serie de tiempo -información histórica- para la cual se contruye un modelo ARIMA con el propósito de obtener pronósticos que se modifiquen a través de combinaciones lineales para atender la información adicional que puede estar contenida en pronósticos más agregados -valores, sumas o promedios ya sea trimestrales o anuales- de la misma

variable obtenidos a través de otros métodos; por ejemplo un modelo econométrico, la opinión de los expertos, o también, la información adicional puede tratarse de una meta planteada en algún programa o presupuesto.

Cuando se calcula un pronóstico univariado óptimo basado en información histórica, dada por un modelo ARIMA, y en información adicional sobre el futuro, se pueden tener cualquiera de las dos situaciones siguientes:

- i) la información adicional impone una restricción a la que el pronóstico debe ceñirse estrictamente; y
- ii) la información es sólo una conjetura acerca de los valores futuros de la serie o un pronóstico de un modelo alternativo.

En ambos casos las restricciones sobre el valor futuro de la serie de tiempo se dan generalmente en forma de combinaciones lineales; por ejemplo promedios simples, ponderados, sumas, etc.

En esta sección se presenta el procedimiento que propone Guerrero para combinar tales restricciones con los pronósticos que se obtienen de un modelo univariado ARIMA. La formulación que sugiere atiende a las reglas de la minimización cuadrática, estadísticamente es congruente con las propiedades de los pronósticos, y deriva estadísticos para decidir la eficacia del procedimiento sugerido.

### 3.3 Derivación de los pronósticos condicionales óptimos

Para derivar los pronósticos ARIMA condicionales, una vez construido el modelo ARIMA para la serie  $\{Z_t\}$ , se parte de la

representación de promedios móviles:<sup>1</sup>

$$Z_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad (3.3.1)$$

en donde la serie  $(a_t)$  es una secuencia de choques aleatorios con media cero y varianza  $\sigma_a^2$ , y las ponderaciones  $\psi_j$  son constantes conocidas, que se calculan de acuerdo con la expresión 2.5.6.5.

El pronóstico óptimo de  $Z_{N+h}$  (lineal con error cuadrático medio mínimo), dada la información histórica, que se representa a través del vector de observaciones  $Z_0 = (Z_1, Z_2, \dots, Z_N)$ , es:

$$E[Z_{N+h} | Z_0] = \mu + \sum_{j=h}^{\infty} \psi_j a_{N+h-j} \quad h = 1, \dots, H \quad (3.3.2)$$

y su error de pronóstico está dado por:

$$Z_{N+h} - E[Z_{N+h} | Z_0] = \sum_{j=0}^{h-1} \psi_j a_{N+h-j}, \quad h = 1, 2, \dots, H \quad (3.3.3)$$

Usando notación matricial se tiene que:

$$Z_F - E[Z_F | Z_0] = \psi a_F \quad (3.3.4)$$

con  $Z_F = (Z_{N+1}, \dots, Z_{N+H})^T$ , el vector de pronósticos, y  $a_F = (a_{N+1}, \dots, a_{N+H})^T$ , el vector de los errores que se generan con el modelo de series de tiempo ARIMA,  $\psi$  es la matriz de dimensión  $H \times H$ :

<sup>1</sup> En esta sección, para simplificar, se usa la notación  $\{Z_t\}$  para la serie transformada  $\{T(Z_t)\}$  que es la serie correspondiente a aquella sobre la que se construyó el modelo ARIMA.

$$\Psi = \begin{bmatrix} \psi_0 & 0 & \dots & 0 \\ \psi_1 & \psi_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{H-1} & \psi_{H-2} & \dots & \psi_0 \end{bmatrix} \quad \text{con } \psi_0 = 1$$

y en donde  $E[a_F | Z_0] = 0$ , y  $E[a_F a'_F | Z_0] = \sigma_a^2 I$ .

Para obtener  $H$  pronósticos condicionales, en el vector  $Y$  de  $H \times 1$ , la información adicional,  $m$  restricciones, se integrará al pronóstico del modelo de series de tiempo en la matriz  $C$  ( $m \times H$ ,  $m \leq H$ ) de constantes, a través de la combinación lineal:

$$Y = CZ_F \tag{3.3.5}$$

Sustituyendo (3.3.4) se tiene

$$Y = CE[Z_F | Z_0] + C\psi a_F \tag{3.3.6}$$

Así, el objetivo es minimizar la función objetivo siguiente:

$$\min_{Z_F} (a_F^T a_F) \quad \text{sujeto a } Y = CZ_F$$

Para resolverlo se tiene, de (3.3.4), que:

$$a_F = \psi^{-1}(Z_F - E[Z_F | Z_0])$$

donde la función objetivo es no singular.

Utilizando multiplicadores de Lagrange, con el vector  $\lambda$ , se construye la función a optimizar:

$$Q = (Z_F - E[Z_F | Z_0])^T (\psi \psi^T)^{-1} (Z_F - E[Z_F | Z_0]) + 2\lambda^T (Y - CZ_F) \tag{3.3.7}$$

Las condiciones de primer orden son:

$$\frac{\partial Q}{\partial Z_F} = 0$$

y

$$\frac{\partial Q}{\partial \lambda^T} = 0$$

y resolviéndolas simultáneamente se obtiene la expresión para el pronóstico óptimo condicional:

$$\begin{aligned} Z_F &= E[Z_F | Z_0] + \psi \psi^T C^T \lambda \\ &= E[Z_F | Z_0] + \psi \psi^T C^T (C \psi \psi^T C^T)^{-1} (Y - C E[Z_F | Z_0]) \end{aligned} \quad (3.3.8)$$

### 3.4 Una prueba estadística para la hipótesis $Y = CZ_F$

El procedimiento para obtener pronósticos ARIMA condicionales consiste en construir un modelo ARIMA para la serie de tiempo  $\{Z_t\}$  y utilizar las ponderaciones  $\psi$  como los elementos que aportan la información histórica del comportamiento de la serie. El pronóstico condicional se obtiene al incorporar la información adicional en el vector  $Y$ . El supuesto básico es que  $Y$  es de la forma  $Y = CZ_F$  y que cada elemento de  $Z_F$  se genera a través de un proceso de promedios móviles. Entonces es necesario verificar que la información adicional sobre el comportamiento futuro de la serie sea consistente con el comportamiento histórico.

Partiendo de que  $Z_t$  cumple con las condiciones de estacionariedad de un proceso gaussiano multivariado, el error del pronóstico condicional está dado por:

$$\begin{aligned} e_Y &= Y - CE[Z_F | Z_0] \\ &= C\psi a_F \sim N_m(0, \sigma_a^2 C\psi\psi^T C^T) \end{aligned} \quad (3.4.1)$$

El estadístico que deriva Guerrero (1988) se calcula suponiendo que  $\sigma_a^2$  es conocido:

$$K_1 = \frac{e_Y^T (C\psi\psi^T C^T)^{-1} e_Y}{\sigma_a^2} \sim \chi_m^2 \quad (3.4.2)$$

Debido a que la varianza muestral de un proceso ARIMA está dada por:

$$\hat{\sigma}_a^2 = \frac{a_0^T a_0}{N-k} \quad (3.4.3)$$

en donde  $a_0 = (a_1, a_2, \dots, a_N)^T$  y  $kk$  es el número de parámetros que se estimaron en el modelo ARIMA, entonces se puede calcular el estimador:

$$K_2 = \frac{e_Y^T (C\psi\psi^T C^T)^{-1} e_Y}{m\hat{\sigma}_a^2} \sim F_{m, N-k} \quad (3.4.4)$$

que se compara con los valores de una tabla de distribución  $F$  con  $m$  y  $N-k$  grados de libertad a un nivel de significancia  $\alpha$ , o bien,  $K_1$  se compara con valores de tablas de una  $\chi^2$  con  $m$  grados de libertad.

Si los estadísticos calculados,  $K_1$  o  $K_2$  no muestran evidencia de que la hipótesis nula  $H_0: Y = CZ_F$  se debe rechazar, entonces los pronósticos condicionales se pueden utilizar, y además se pueden construir intervalos de confianza para éstos.

### 3.4.1 Intervalos de confianza para los pronósticos condicionales

El proceso para calcular los intervalos de confianza de los pronósticos condicionales utiliza la propiedad de insesgamiento que se busca en las predicciones estadísticas. Para ello, se plantea la solución del problema de optimización desde el punto de vista de las propiedades de los pronósticos.

Para que el pronóstico condicional sea el óptimo, debe existir una matriz constante  $A$ , de dimensiones  $m \times H$ , tal que

$$\begin{aligned} Z &= AY \\ &= A(E[Z_f | Z_0] + \psi a_f) \end{aligned} \quad (3.4.1.1)$$

el cual, debido a que el error de pronóstico,  $Z_0$  es insesgado, debe satisfacer la siguiente igualdad:

$$0 = E[Z_f - Z_f | Z_0] = (AC - I)E[Z_f | Z_0] \quad (3.4.1.2)$$

y de esta manera el error de pronóstico está dado por:

$$Z_f - Z_f = (AC - I)\psi a_f \quad (3.4.1.3)$$

cuya matriz de varianza-covarianza está dada por

$$\text{Cov}[(Z_f - Z_f | Z_0)] = \sigma_a^2 (C\psi\psi^T C^T A^T - AC\psi\psi^T - \psi\psi^T C^T A^T + \psi\psi^T)$$

y para que el vector de pronósticos estimado tenga el error cuadrático medio mínimo, la matriz  $A$  se debe escoger de tal manera que minimice la varianza generalizada, definida como:

$$\text{var } Z_f = \text{tr}(\text{Cov}[Z_f - Z_f | Z_0])$$

y entonces  $A$  se puede calcular al resolver

$$0 = \left( \frac{\partial \text{var } Z_f}{\partial A} \right)$$

y así:

$$\hat{\lambda} = \psi \psi^T (C \psi \psi^T C^T)^{-1} \quad (3.4.1.4)$$

Esta matriz es completamente consistente con el pronóstico condicional obtenido a través de minimización cuadrática ya que, de 3.4.1.1 y 3.4.1.2, el pronóstico condicional óptimo está dado por

$$\hat{Z}_f = E[Z_f | Z_0] + \hat{\lambda} C a_f \quad (3.4.1.5)$$

y utilizando 3.3.6 se tiene que

$$\begin{aligned} \hat{Z}_f &= E[Z_f | Z_0] + \hat{\lambda} (Y_0 - CE[Z_f | Z_0]) \\ &= E[Z_f | Z_0] + \psi \psi^T C^T (C \psi \psi^T C^T)^{-1} (Y - CE[Z_f | Z_0]) \end{aligned} \quad (3.4.1.6)$$

### 3.4.2 Diferencia entre los pronósticos ARIMA y los pronósticos condicionales.

Cuando se comparan las covarianzas asociadas de los pronósticos del modelo simple de series de tiempo y de los pronósticos condicionales se tiene lo siguiente:

La covarianza de los errores de pronóstico del modelo ARIMA,  $\hat{Z}_f$ , es:

$$Cov(\psi a_f | Z_0) = \sigma_a^2 \psi \psi^T \quad (3.4.2.1)$$

Por su parte, la matriz de covarianzas del pronóstico  $E[Z_f | Z_0]$  excede a la matriz de covarianzas de  $\hat{Z}_f$  por

$$\sigma_a^2 \hat{\lambda} C \psi \psi^T = \sigma_a^2 \psi \psi^T C^T (C \psi \psi^T C^T)^{-1} C \psi \psi^T \quad (3.4.2.2)$$

la cual es positiva semidefinida.

De aquí que el pronóstico óptimo condicional, basado en información  $Z_0$  y  $Y$ , sea más preciso que aquel que se genera utilizando únicamente  $Z_0$ .

### 3.5 Pronósticos condicionales cuando la información adicional es una conjetura

Cuando los estadísticos calculados muestran evidencia de que se rechace la hipótesis  $Y=CZ_f$  es posible considerar otras alternativas. Una de ellas es la que propone Pérez-Porrúa (1984) la cual consiste en resolver el mismo problema que incluye un supuesto acerca de la diferencia entre los valores que se conjeturan y los valores futuros que efectivamente observará la variable, suponiendo que la información adicional es incierta. El desarrollo del pronóstico condicional se realiza suponiendo que

$$Y - CZ_f = u \quad (3.5.1)$$

en donde el vector  $u$  satisface las siguientes condiciones:

$$u \sim N_H(0, U), \quad E[Z_0 u^T] = E[a_f u^T] = 0$$

Expresando los pronósticos como esperanzas condicionales se llega a la expresión:

$$\begin{aligned} \hat{Z}_{f,U} &= E[Z_f | Z_0, Y] \\ &= E[Z_f | Z_0] + \hat{A}_U(Y) - CE[Z | Z_0] \end{aligned} \quad (3.5.2)$$

en donde:

$$\hat{A}_U = \psi \psi^T C^T \left( C \psi \psi^T C^T + \frac{U}{\sigma_a^2} \right)^{-1} \quad (3.5.3)$$

La matriz  $U$  se puede calcular utilizando la misma fuente de información externa; por ejemplo, en el caso en que ésta sea un pronóstico de un modelo alternativo se cuenta con los intervalos de confianza de los pronósticos que en sí constituyen un margen de incertidumbre, o bien, utilizando la evidencia empírica de los mismos datos. Para este último caso, Guerrero propone que la matriz  $U$  se seleccione de la siguiente manera:

Partiendo de los errores del pronóstico condicional

$$e_{Y,U} = Y - CE[Z_F | Z_0] = C\psi a_f + u \quad (3.5.4)$$

entonces

$$E[e_{Y,U} | Z_0] = 0,$$

$$Cov[e_{Y,U} | Z_0] = \sigma_a^2 (C\psi\psi^T C^T + U)$$

y

$$K_U = e_{Y,U}^T \left( C\psi\psi^T C^T + \frac{U}{\sigma_a^2} \right)^{-1} e_{Y,U} / \sigma_a^2 \sim \chi_m^2 \quad (3.5.5)$$

y así, para  $N$  suficientemente grande, se puede calcular el estadístico:

$$K_U = (Y - CE[Z_F | Z_0])^T \left( C\psi\psi^T C^T + \frac{U}{\sigma_a^2} \right)^{-1} (Y - CE[Z_F | Z_0]) / \hat{\sigma}_a^2 \quad (3.5.6)$$

el cual se utiliza para especificar la matriz  $U$ . Si se compara este estadístico con una  $\chi_m^2$  de tablas a un nivel de significancia  $\alpha$ , y se tiene que  $K_U < \chi_m^2(\alpha)$ , significa que los datos no muestran evidencia de que deba rechazarse la hipótesis  $Y - CZ_F = u$  con  $u \sim N(0, U)$ .

## Capítulo 4

### Pronóstico de la demanda de petrolíferos

En este capítulo se aplican las técnicas expuestas en los capítulos anteriores para generar los pronósticos anuales (mediano plazo) y mensuales (corto plazo) de la demanda interna de petrolíferos. Para ello, se seleccionaron los cinco productos más importantes en el consumo de combustibles: las gasolinas de uso automotriz (Nova y Magna Sin), el diesel, el gas licuado, el combustóleo y el gas natural.

El ejercicio en el que se muestra la utilización del método de pronósticos condicionales se realizó de la siguiente manera:

10. Se estimaron las ecuaciones de demanda de cada uno de los cinco productos, utilizando los modelos de regresión lineal múltiple, y con ellas se generaron las proyecciones anuales para 1995.
20. Se construyeron los modelos ARIMA para las series mensuales del volumen de venta de cada uno de los productos.
30. Se aplicó el método propuesto de combinación de pronósticos, y se obtuvieron los pronósticos ARIMA condicionales.

#### 4.1 La importancia de los petrolíferos

La industria petrolera en México juega un papel prioritario en el desarrollo económico del país. Contribuye al producto interno bruto (PIB), a la balanza comercial y a las finanzas públicas como ninguna otra rama industrial; su vinculación productiva con

otros sectores es crítica. Durante muchos años sus aportaciones provocaron que se hablara de una "petrolización" de la economía; sin embargo, la importancia de la industria petrolera en términos de los grandes agregados económicos ha venido declinando como resultado de la caída del precio internacional del petróleo que se inició en 1981, de la restricción presupuestal a la que ha estado sujeta la industria petrolera en México y de una estructura industrial más diversificada.

México es el 6o. país productor de petróleo crudo y el 8o. de gas natural en el mundo. En 1995 su producción alcanzó los 2,617 miles de barriles diarios (mbd) de crudo y 3,759 miles de pies cúbicos diarios (mpcd) de gas natural. Exceptuando la exportación de 1,295 mbd de crudo, el resto de la extracción se destina a la producción de petrolíferos y petroquímicos, en su mayor parte para satisfacer la demanda interna de este tipo de energéticos.

La explotación de crudo, así como la comercialización y distribución de los hidrocarburos es facultad exclusiva del Estado y la empresa encargada para ello es Petróleos Mexicanos (Pemex). La importancia de este organismo es estratégica para la economía y las finanzas públicas. En términos generales, en 1995 aportó 9.5% al PIB y contribuyó con 36.5% de los ingresos fiscales. Durante ese año, Pemex enteró al Gobierno Federal 103,478 millones de nuevos pesos de impuestos y derechos sobre la extracción de hidrocarburos, lo que equivale al 70.2% de los ingresos totales que se derivan de esta actividad. Correspondieron a derechos sobre extracción de hidrocarburos y otros impuestos directos 73,570 millones de nuevos pesos, y como retenedor del impuesto

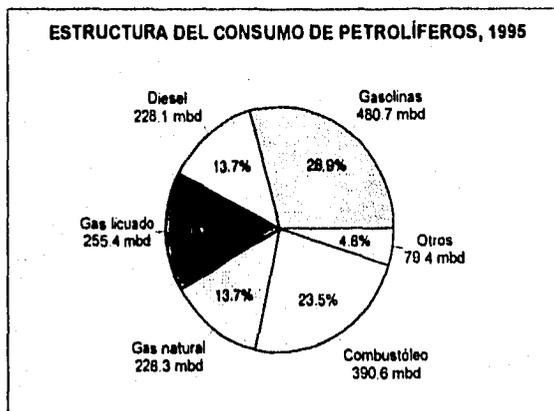
al valor agregado (IVA) y del impuesto especial a la producción y servicios (IEPS), la empresa liquidó 29,908 millones de nuevos pesos a la federación.<sup>1/</sup>

Tanto el IVA como el IEPS se aplican al consumo nacional de los petrolíferos y éstos juegan un papel sustancial en la política de precios y tarifas, misma que constituye un elemento fundamental en la regulación de los ingresos del sector público. Consecuentemente, para el Gobierno Federal resulta imprescindible elaborar pronósticos a mediano plazo de la demanda interna de petrolíferos a través de modelos que consideren el contexto macroeconómico y el manejo de diferentes políticas. Asimismo, las tareas de seguimiento y control de las finanzas públicas, requiere de proyecciones a corto plazo que permitan instrumentar medidas correctivas en el transcurso de los ejercicios presupuestales.

Los petrolíferos que se seleccionaron para pronosticar su demanda, en 1995 representaron el 95.2% del volumen de petrolíferos consumidos en el país y su composición se muestra en la siguiente gráfica.

---

<sup>1/</sup> Datos contenidos en el Informe sobre la Situación Económica y las Finanzas Públicas, correspondiente al cuarto trimestre de 1995.

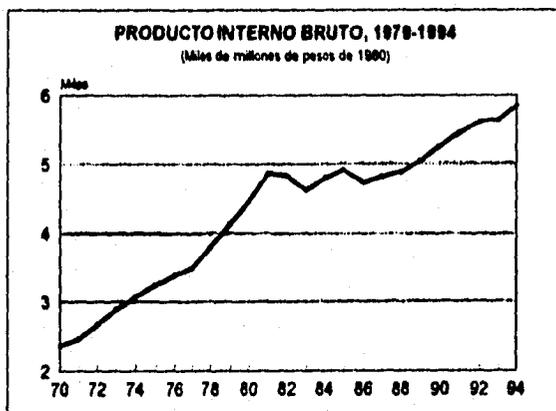


#### 4.2 Pronósticos anuales

Utilizando los modelos de regresión lineal múltiple se estimaron las ecuaciones de demanda para cada uno de los productos señalados.<sup>2/</sup> De acuerdo con la teoría económica, la demanda de un producto de consumo depende generalmente del ingreso del consumidor, del precio de ese producto y del precio de otros productos que compiten con éste o que son complementarios. La dependencia se supone que es lineal y se representa con un modelo de regresión lineal múltiple. Es importante señalar que esta parte del proceso es la que permite ligar los pronósticos del consumo de petrolíferos con las proyecciones del entorno macroeconómico, ya que en frecuentemente la variable que aproxima al ingreso de

<sup>2/</sup> Para realizar los cálculos se utilizó el paquete "Time Series Processor", versión 6.1 (1988).

los consumidores es el Producto Interno Bruto (PIB) general nacional, o alguna de sus ramas, y los precios se pueden establecer de acuerdo con las metas de inflación que se estén considerando.

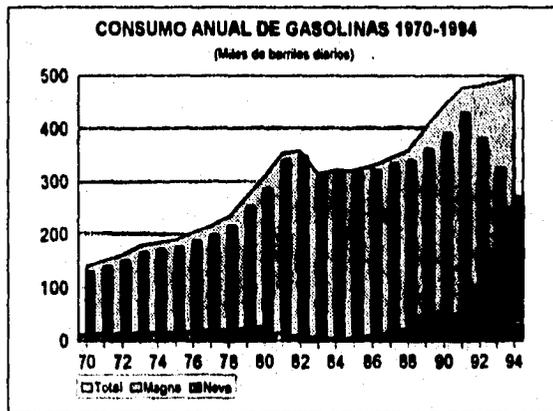


En la estimación de las ecuaciones se utilizaron datos anuales del periodo 1970 a 1994, es decir, se cuenta con 25 observaciones. Asimismo, las variables que intervinieron en la estimación de las ecuaciones de demanda corresponden al logaritmo de las observaciones originales, con el fin de poder interpretar los coeficientes estimados en términos de elasticidades. El proceso para encontrar la ecuación más adecuada para cada uno de los productos se presenta a continuación.

#### 4.2.1 Gasolinas automotrices

El consumo de gasolinas automotrices representa el 28.9% del total de petrolíferos. Actualmente su venta se realiza en dos

variedades: la Nova y la Magna Sin. La venta de ésta última se inició en 1989 en sustitución de la gasolina Extra y su consumo ha ido aumentando paulatinamente de tal manera que en diciembre de 1994 ya se vendía igual cantidad de Magna Sin que de Nova, y para diciembre de 1995 la proporción fue 45.0% de Nova y 55.0% de Magna Sin, pretendiéndose continuar con la sustitución del consumo de Nova por Magna Sin.



Es por ello que, aun cuando la demanda de cada una de las dos variedades de gasolina se comporta de manera diferente, se decidió pronosticar el consumo de las dos gasolinas en conjunto. De hecho se está induciendo la sustitución de una por otra a través de la política de precios; no obstante que la calidad de la gasolina Magna Sin es muy superior a la de la Nova, el diferencial de precios es muy reducido para que el consumidor, en un principio sea indiferente al consumo de cualquiera de las dos, y por tanto prefiera la de más calidad. Adicionalmente,

los vehículos que se incorporan a la circulación a partir de 1990 están equipados con un convertidor catalítico y requieren necesariamente el uso de la Magna Sin.

Las variables que se utilizaron en este caso fueron:

LGASO= logaritmo natural del volumen de ventas de gasolinas, medido en miles de barriles diarios,

LPIB= logaritmo natural del PIB

LPRPO= logaritmo natural del precio ponderado de las gasolinas Nova y Magna Sin, deflactado con el Índice Nacional de Precios al Consumidor (INPC), y

LVEHI= logaritmo natural del número de vehículos en circulación en el país.

Inicialmente se consideraron como variables independientes el PIB y el precio; la ecuación resultante fue la siguiente:

$$\begin{array}{rcccc} \text{LGASO} & = & -6.438 & +1.499 & \text{LPIB} & -0.114 & \text{LPRPO} \\ & & (0.445) & (0.065) & & (0.059) \\ t & = & (-14.47) & (22.92) & & (-1.94) \end{array}$$

$$r^2 = 0.9732 \quad g.l. = 22 \quad Dw = 0.529 \quad F(1, 22) = 399.54$$

$$\bar{r}^2 = 0.9708$$

en donde los números que aparecen abajo de los coeficiente entre paréntesis corresponden a la desviación estándar y al estadístico  $t$  de cada uno de ellos, empleando las expresiones 1.4.2.7 y 1.4.4.1, respectivamente; los coeficientes de determinación simple ( $r^2$ ) y ajustado ( $\bar{r}^2$ ) se obtienen utilizando las expresiones 1.4.3.2 y 1.4.3.3, en ese orden; y el estadístico  $F$  con la ecuación 1.4.4.2.

El estadístico  $Dw$ , calculado de acuerdo con la expresión 1.6.2.1, así como la gráfica de los residuales indicaron la presencia de autocorrelación, y entonces se intentó corregirla

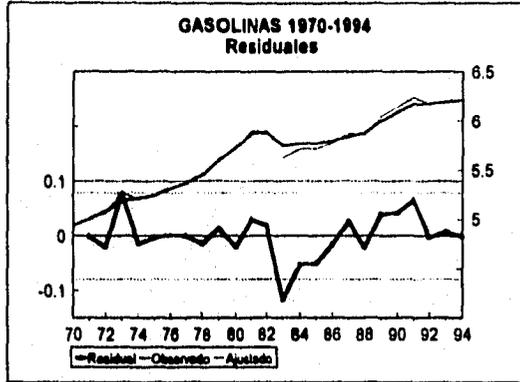
incluyendo la variable de acervo vehicular, que está ligada estrechamente al comportamiento de la demanda de gasolina, y los resultados fueron los siguientes:

$$\begin{array}{rcccc}
 LGASO & = & -4.025 & +0.941 & LPIB & -0.115 & LPRPO & +0.270 & LVEHI \\
 & & (2.652) & (0.607) & & (0.059) & & (0.293) \\
 t & = & (-1.52) & (1.55) & & (-1.95) & & (0.92) \\
 \\ 
 r^2 & = & 0.9743 & & g.l. & = 21 & DW & = 0.583 & F(1,22) = 264.84 \\
 \bar{r}^2 & = & 0.9706 & & & & & & 
 \end{array}$$

Esta variable no contribuyó a elevar el coeficiente de explicación del modelo, no eliminó la presencia de autocorrelación y, sobretodo, al incluirla el valor del coeficiente de la variable de ingreso (LPIB) se reduce significativamente. Ello se puede interpretar como la existencia de correlación entre estas variables, lo cual es razonable ya que la variación del acervo vehicular está dada en mayor medida por el número de unidades nuevas que se incorporan a la circulación, y la industria automotriz está ligada fuertemente con el desarrollo de la economía. Entonces, se incluyó la misma variable dependiente rezagada un periodo:

$$\begin{array}{rcccc}
 LGASO & = & -2.224 & +0.559 & LPIB & -0.121 & LPRPO & +0.649 & LGASO(-1) \\
 & & (0.829) & (0.173) & & (0.037) & & (0.114) \\
 t & = & (-2.68) & (3.22) & & (-3.22) & & (5.70) \\
 \\ 
 r^2 & = & 0.9887 & & g.l. & = 20 & DW & = 1.642 & F(1,22) = 581.11 \\
 \bar{r}^2 & = & 0.9870 & & & & & & 
 \end{array}$$

La ecuación es apropiada, pero al analizar la gráfica de los residuales hay dos de ellos que rebasan por un amplio margen los límites marcados por  $2\hat{\sigma}^2$ . Estos residuales corresponden a 1973 y 1983. En el primero, si bien hubo un incremento en el precio después de 15 años de mantenerlo fijo, la sustitución de Supermexolina, Gasolmex y Pemex-100 por únicamente dos tipos



de gasolina, Pemex-Nova y Pemex-Extra, propició una demanda mayor. En 1983, el incremento de los precios, aunado a una drástica caída de la economía, originó una reducción más severa del consumo de la gasolina.

Por estas razones se decidió incorporar una variable artificial (DUM, de "dummy") que considerara el efecto de estos dos factores. La ecuación resultante fue la siguiente:<sup>3/</sup>

$$\begin{array}{rcccccc}
 LGASO & = & -2.286 & +0.563 & LPIB & -0.132 & LPRPO & +0.660 & LGASO(-1) & +0.089 & DUM \\
 & & (0.513) & (0.107) & & (0.023) & & (0.070) & & (0.15) \\
 t & = & (-4.46) & (5.25) & & (-5.67) & & (9.38) & & (5.78)
 \end{array}$$

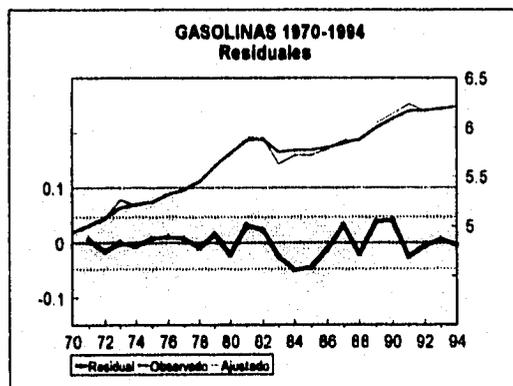
$$R^2 = 0.9959 \quad g. l. = 19 \quad DW = 1.744 \quad F(1,22) = 1150.2$$

$$\bar{R}^2 = 0.9950$$

Para esta ecuación la verificación de los supuestos del modelo de regresión es satisfactoria ya que el valor de  $\bar{r}^2$  es muy elevado; todos los coeficientes estimados son estadísticamente significativos dado que el valor crítico de la distribución t

3/ Véase Pindyk R. y Rubinfeld D. (1981) pp. 424-425.

con 19 grados de libertad al 5% de significancia es 2.093; la gráfica de los residuales no muestra presencia de autocorrelación; y el valor de la  $F$  es muy elevado lo que indica que en conjunto, todos los coeficientes son significativamente diferentes de cero.



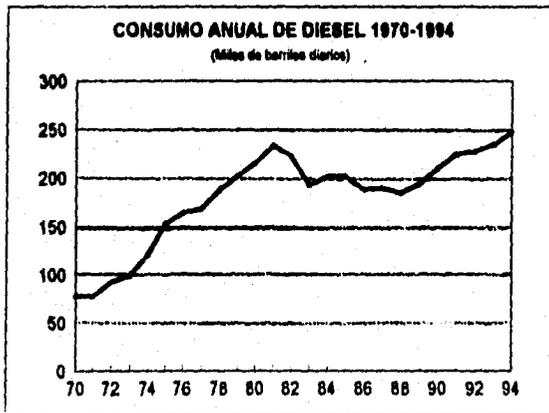
De acuerdo con la teoría económica del consumo, la ecuación de demanda estimada es también adecuada. Los coeficientes estimados de las variables exógenas relativas al ingreso y al precio, se pueden interpretar de la siguiente manera: el consumo de gasolina crece alrededor de 0.56% cuando la actividad económica aumenta 1.0%, manteniendo constante el resto de las variables; de la misma forma, si el precio de las gasolinas aumenta 1% real en el año, se espera que el consumo decline 0.13% por este efecto.<sup>4/</sup>

<sup>4/</sup> El incremento real del precio se refiere a la variación del precio de un año a otro, eliminando el efecto de la inflación.

#### 4.2.2 Diesel

El diesel es el combustible que utilizan generalmente los vehículos de autotransporte y de transporte de carga. El comportamiento de su demanda está ligado al de las gasolinas, y en algún modo puede decirse que son productos sustitutos.

Las variedades de este producto son: el diesel desulfurado, el diesel Sin a partir de 1993, y el diesel automotriz.



Las variables que se utilizaron en este caso fueron:

LDIES= logaritmo natural del volumen de ventas de diesel, medido en miles de barriles diarios,

LPIB= logaritmo natural del PIB,

LPRDI= logaritmo natural del precio del diesel, deflactado con el INPC, y

LPRPO= logaritmo natural del precio ponderado de las gasolinas Nova y Magna Sin, deflactado con el INPC.

Inicialmente se consideraron como variables independientes el PIB y el precio del diesel y de las gasolinas; la ecuación resultante fue la siguiente:

$$\begin{array}{rcccccc}
 LDIES & = & -6.827 & +1.430 & LPIB & -0.242 & LPRDI & +0.209 & LPRPO \\
 & & (0.551) & (0.077) & & (0.033) & & (0.063) \\
 t & = & (-12.40) & (18.48) & & (-7.24) & & (3.33)
 \end{array}$$

$$r^2 = 0.9657 \quad g.l. = 21 \quad DW = 1.478 \quad F_{(1,21)} = 196.93$$

$$\bar{r}^2 = 0.9608$$

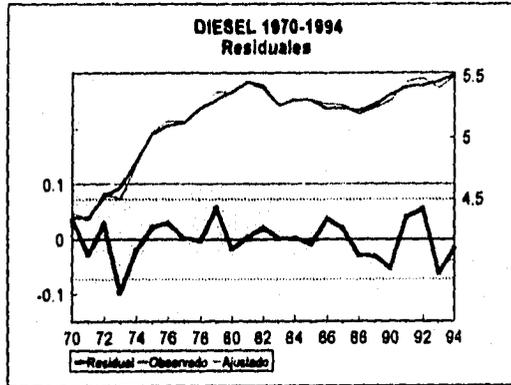
La ecuación es adecuada, pero el residual correspondiente a 1973 es muy grande. En este año la introducción de nuevas gasolinas que reunían especificaciones adecuadas para la operación mecánica de todos los tipos y modelos de automóviles y camiones que circulaban, indujo la sustitución del diesel por estos productos. Por lo tanto, se decidió incorporar una variable artificial para considerar este efecto. Así, la ecuación resultante fue:

$$\begin{array}{rccccccccc}
 LDIES & = & -7.072 & +1.449 & LPIB & -0.262 & LPRDI & +0.257 & LPRPO & -0.161 & DUM \\
 & & (0.330) & (0.046) & & (0.020) & & (0.038) & & (0.026) \\
 t & = & (-21.41) & (31.36) & & (-12.99) & & (6.73) & & (-6.26)
 \end{array}$$

$$r^2 = 0.9884 \quad g.l. = 19 \quad DW = 2.146 \quad F_{(1,19)} = 426.16$$

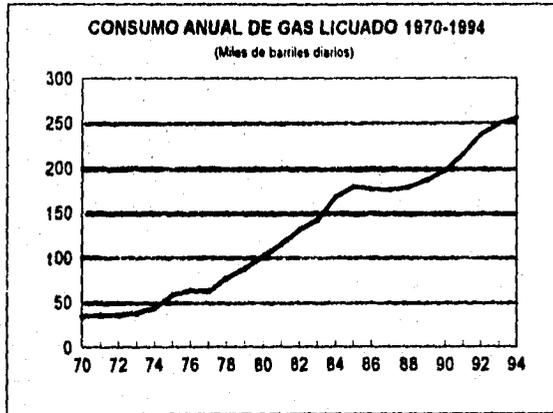
$$\bar{r}^2 = 0.9860$$

En este caso, los parámetros asociados a las variables de ingreso y precio indican lo siguiente: el consumo de este producto es elástico al ingreso, cuando la economía crece 1.0%, la demanda aumenta 1.45%; con relación al precio, un incremento del precio del diesel de 1.0% reduce el consumo en 0.26%, y cuando el precio de la gasolina disminuye 1.0% real, el consumo del diesel aumenta en 0.26%.



#### 4.2.3 Gas licuado

El gas licuado se usa como combustible en los hogares, principalmente para preparar alimentos y calentar agua, y en menor medida, para los sistemas de aire acondicionado. Aunque a partir de 1992 una parte de éste se utilizó como carburante para vehículos, la falta de seguridad en el sistema de suministro, y sobretodo el subsidio que representaba para los usuarios, comparando con el precio de la gasolina y del diesel, ocasionó que actualmente este uso se haya reducido significativamente.



Las variables que se utilizaron en este caso fueron:

LGLP= logaritmo natural del volumen de ventas de gas licuado, medido en miles de barriles diarios,

LPIB= logaritmo natural del PIB, y

LPRGL= logaritmo natural del precio al público del gas licuado, medido en N\$/lt y deflactado con INPC.

Inicialmente se consideraron como variables independientes el PIB y el precio; la ecuación resultante fue la siguiente:

$$\begin{aligned}
 LGLP &= -16.125 + 2.479 LPIB + 0.095 LPRGL \\
 &\quad (1.317) \quad (0.132) \quad (0.115) \\
 t &= (-12.25) \quad (18.80) \quad (0.39)
 \end{aligned}$$

$$r^2 = 0.9570 \quad g.l. = 23 \quad Dw = 0.404 \quad F(1.22) = 244.54$$

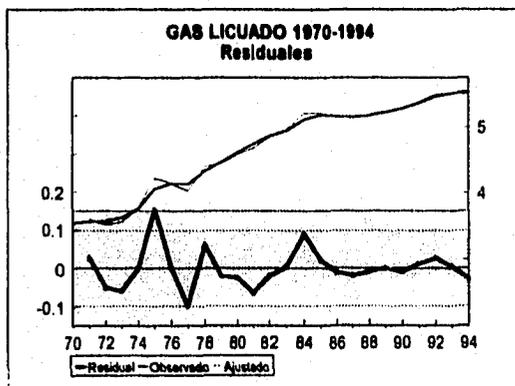
$$\bar{r}^2 = 0.9530$$

En esta ecuación el coeficiente del precio del producto es no significativo y su signo es diferente al que se espera, el estadístico Durbin-Watson indica la existencia de autocorrelación entre los residuales. Debido a que no se encontró una variable que pudiera agregarse a la especificación del modelo se probó incluir la misma variable rezagada un periodo:

$$\begin{array}{rccccccc}
 LGLP & = & -4.836 & +0.760 & LPIB & -0.004 & LPRGL & +0.694 & LGLP(-1) \\
 & & (1.273) & (0.180) & & (0.043) & & (0.065) \\
 t & = & (-3.80) & (4.21) & & (-0.11) & & (10.66)
 \end{array}$$

$$r^2 = 0.9940 \quad g.l. = 19 \quad DW = 1.958 \quad F(1,22) = 1,109.63$$

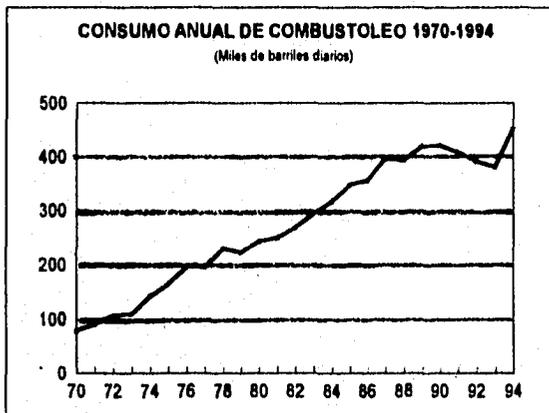
$$\bar{r}^2 = 0.9931$$



En esta ecuación el coeficiente del precio ya tiene el signo esperado, aunque no es significativo, lo que se interpreta como que el consumo del Gas LP es casi indiferente al incremento de su precio; sin embargo, se decidió utilizar esta ecuación.

#### 4.2.4 Combustóleo

El combustóleo y el gas natural son los principales combustibles que utiliza la industria en México. La mayor parte del primero (73.6%), se utiliza en la generación de electricidad. El combustóleo es un producto relativamente barato ya que es un residual del proceso de refinación del crudo; es por ello que su precio es muy cercano al del petróleo crudo. De hecho, cuando se utilizó en la estimación de las ecuaciones de demanda el precio del combustóleo, éste no resultó significativo y mostraba signo contrario al esperado; entonces se utilizó el precio de la mezcla de petróleo crudo de exportación.



De esta manera, las variables que se emplearon fueron:

L<sub>COMB</sub>= logaritmo natural del volumen de ventas del total de combustóleo, medido en miles de barriles diarios,

L<sub>PIB</sub>= logaritmo natural del PIB,

LPRCR= logaritmo natural del precio en pesos de la mezcla de crudo mexicano de exportación, en pesos y deflactado con el INPC, y

LPRGN= logaritmo natural del precio de venta del gas natural, deflactado con el Índice Nacional de Precios al Productor.

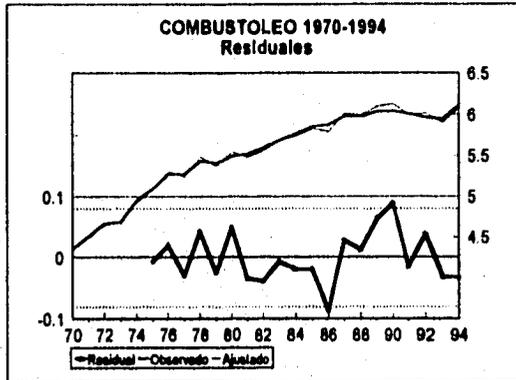
En este caso, el periodo de estimación fue de 1975 a 1994 debido a que la variable del precio del petróleo crudo está disponible a partir de 1975, año en que México inició la exportación de este producto.

Los resultados fueron los siguientes:

$$\begin{array}{r}
 LCOMB = -4.628 + 1.236 LP/B - 0.135 LPRCR + 0.160 LPRGN \\
 \quad \quad (1.075) \quad (0.138) \quad \quad (0.048) \quad \quad (0.032) \\
 t = (-4.31) \quad (8.98) \quad \quad (-2.81) \quad \quad (5.00) \\
 \\
 r^2 = 0.9575 \quad g.l. = 16 \quad DW = 1.381 \quad F(1,16) = 120.29 \\
 \bar{r}^2 = 0.9496
 \end{array}$$

La gráfica de residuales y el estadístico  $DW$  indicaron la presencia de autocorrelación y esta se corrigió utilizando la variable dependiente rezagada un periodo:

$$\begin{array}{r}
 LCOMB = -5.119 - 1.322 LP/B - 0.196 LPRCR - 0.146 LPRGN - 0.141 LCOMB(-1) \\
 \quad \quad (0.743) \quad (0.096) \quad \quad (0.036) \quad \quad (0.022) \quad \quad (1.032) \\
 t = (-6.89) \quad (13.75) \quad \quad (-5.50) \quad \quad (6.43) \quad \quad (-4.38) \\
 \\
 r^2 = 0.9814 \quad g.l. = 15 \quad DW = 1.950 \quad F(1,15) = 197.80 \\
 \bar{r}^2 = 0.9764
 \end{array}$$



#### 4.2.5 Gas natural

El consumo del gas natural, como ya se mencionó en la sección anterior, es básicamente industrial; sólo un 5.0% se utiliza para consumo doméstico. El consumo de este producto se puede explicar también con los requerimientos del sector industrial.

En este caso las variables utilizadas fueron:

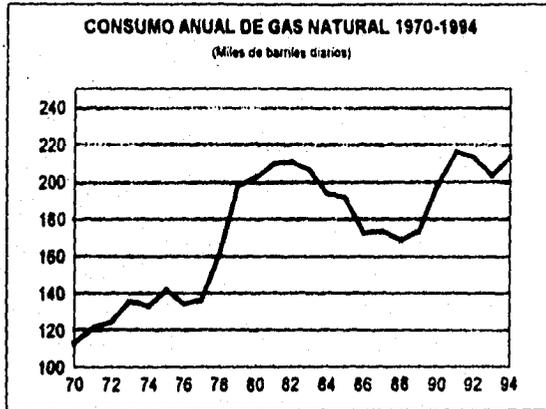
LGNAT =logaritmo natural del volumen de ventas del total de gas natural, medido en miles de barriles diarios de combustóleo,

LPIB = logaritmo natural del PIB total,

LPRGN =logaritmo natural del precio de venta del gas natural, deflactado con el INPC, y

LPRCR =logaritmo natural del precio en pesos de la mezcla de crudo mexicano de exportación, como precio de un producto sustituto, en este caso el combustóleo.

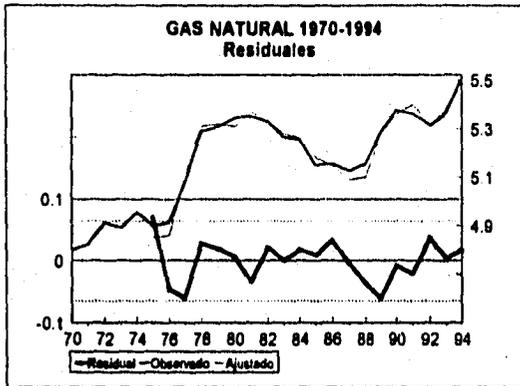
La ecuación resultante fue la siguiente:



<i>LGNAT</i>	=	-2.348	+0.865	<i>LPIB</i>	=	-0.064	<i>LPRGN</i>	+0.127	<i>LPRCR</i>
		(0.594)	(0.076)			(0.017)		(0.026)	
<i>t</i>	=	(-3.95)	(11.37)		=	(-3.63)		(4.80)	

$r^2 = 0.9513$        $g.l. = 16$        $DW = 1.727$        $F(1,16) = 73.28$

$\bar{r}^2 = 0.9383$



#### 4.2.6 Los pronósticos anuales

Para el cálculo de los pronósticos correspondientes a 1995, se consideró que la actividad económica reportaría una caída de 6.5% con relación a 1994 y que la inflación de fin de periodo ascendería a 50%. De esta manera, el valor de las variables macroeconómicas exógenas que se utilizaron se presentan en el siguiente Cuadro.

MARCO MACROECONOMICO, 1995		
Variable	Unidad de medida	Valor
PIB total	Miles de millones pesos de 1980	5,472.8
INPC	Millones de pesos de 1980	58,418.2
Tipo de cambio	Nuevos pesos por dólar	6.29
Precio del petróleo crudo de exportación	Dólares por barril	14.98

Los precios promedio ponderados, que se utilizaron son acordes con la política que se estableció en el Programa de Acción para Reforzar el Acuerdo de Unidad para Superar la Emergencia Económica (PARAUSEE).

PRECIOS INTERNOS DE PETROLIFEROS	
Producto	Precio
Gasolinas	1.96 N\$/lt
Diesel	1.60 N\$/lt
Gas licuado	1.09 N\$/lt
Combustóleo */	15.6 dpb
Gas natural	0.420 N\$/kg

\*/ Precio del crudo de exportación.

De esta manera, los pronósticos anuales de cada producto para 1995, calculados utilizando la expresión 1.3.1, así como sus intervalos de confianza, obtenidos mediante la expresión 1.3.4, se presentan en el siguiente cuadro:

DEMANDA DE PETROLIFEROS PARA 1995			
Producto	Pronóstico (mbd)	Intervalos de confianza	
		Límite inferior	Límite superior
Gasolinas	494.5	472.3	515.3
Diesel	228.8	215.8	242.8
Gas licuado	254.5	232.4	278.6
Combustóleo	382.1	356.9	409.1
Gas Natural	219.2	205.9	233.4
	(mpcd) 1,486.8	1,396.6	1,583.2

Estas proyecciones, que consideran el efecto del comportamiento general de la economía, implican descensos importantes en el consumo del combustóleo (15.7%), del diesel (7.6%), y de las

gasolinas (1.3%); por su parte, para el gas natural se obtiene un crecimiento de sólo 1.9% y el gas licuado queda prácticamente en el mismo nivel.

#### 4.3 Pronósticos mensuales

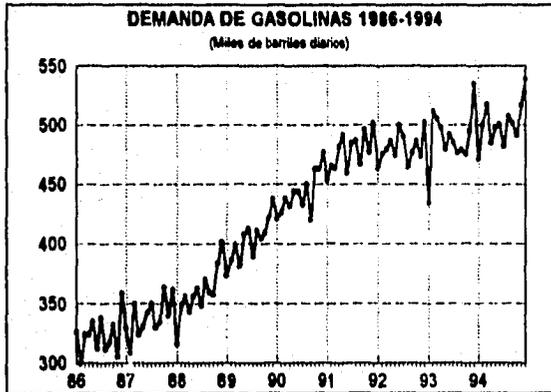
La construcción de los modelos ARIMA se realizó para las series mensuales del volumen de ventas de los petrolíferos de 1986 a 1994, utilizando la versión 2.0 del paquete estadístico "The Autoboxp System".

El proceso para la construcción de estos modelos se presenta para el caso de la gasolina, y para los demás productos se muestran sólo los aspectos relevantes de la estimación y los resultados.

##### 4.3.1 Gasolines

El comportamiento mensual de la demanda de gasolinas en el mercado interno se muestra en la siguiente gráfica:

En la gráfica se aprecia que es necesario estacionarizar la serie ya que su nivel no es constante y es diferente de cero. En lo que se refiere a su varianza, para seleccionar la transformación estabilizadora de varianza, se calculó el coeficiente de determinación para valores de  $\lambda$  de -2 a 2, tomando 9 grupos con 12 observaciones de la serie original

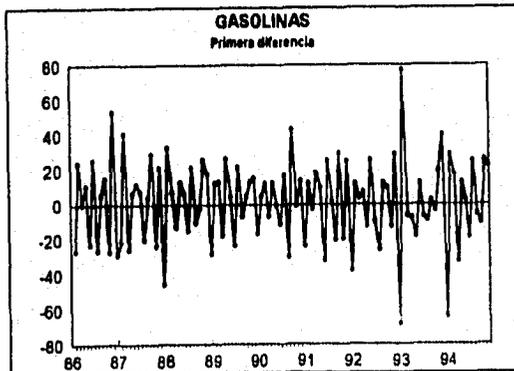


dividida entre 100. El valor mínimo se obtuvo cuando  $\lambda = 1$ , lo que significa que no es necesario aplicar ninguna transformación.<sup>5/</sup>

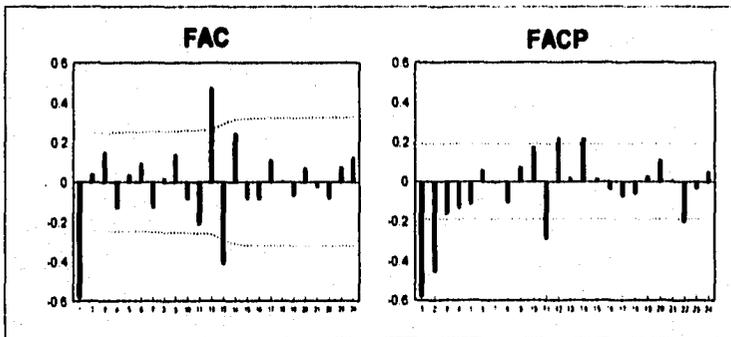
<sup>5/</sup> Es importante señalar que para poder aplicar las restricciones a los pronósticos ARIMA, se deben manejar combinaciones lineales de estos pronósticos, por lo que, en caso de haber obtenido  $\lambda = 0$ , no se hubiese podido utilizar el método de pronósticos condicionales expuesto. Por ello, para el resto de los productos sólo se realizaron las transformaciones de cambio de escala.

DETERMINACION DE $\lambda$ PARA LA SERIE DE CONSUMO DE GASOLINAS								
		Potencia $\lambda$						
Año	Grupo	-2	-1	-0.5	0	0.5	1	2
1986	1	729.575	119.860	48.582	19.691	7.981	3.235	0.531
1887	2	858.467	135.668	53.933	21.440	8.523	3.388	0.535
1888	3	384.860	80.896	37.152	17.041	7.816	3.585	0.754
1889	4	677.024	122.700	52.235	22.237	9.467	4.030	0.730
1990	5	776.324	138.692	58.621	24.777	10.472	4.426	0.790
1991	6	1271.073	197.564	77.889	30.707	12.106	4.772	0.741
1992	7	2562.632	316.160	111.049	39.005	13.700	4.812	0.593
1993	8	334.672	81.818	40.454	20.002	9.889	4.890	1.195
1894	9	832.239	151.399	64.574	27.542	11.747	5.010	0.911
1995	10	933.869	161.045	66.877	27.772	11.533	4.789	0.825
de		668.524	71.822	22.578	6.812	2.012	0.695	0.207
M		936.329	149.428	60.499	24.716	10.189	4.239	0.753
Coef. de variación		0.714	0.480	0.373	0.275	0.197	0.164	0.275

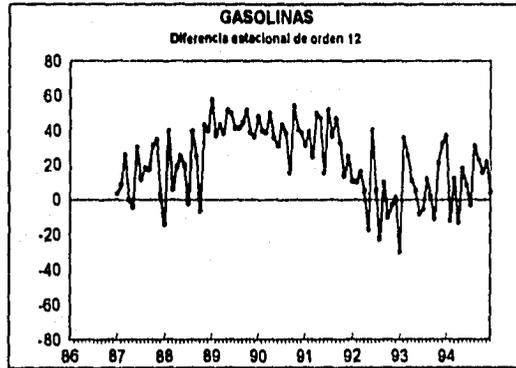
Para obtener una serie estacionaria se calculó la primera diferencia de la serie,  $\nabla T(Z_t)$  en donde  $(T(Z_t))$  es la demanda promedio mensual de gasolina de enero de 1986 a diciembre de 1994, medida en miles de barriles diarios y dividida entre 100.



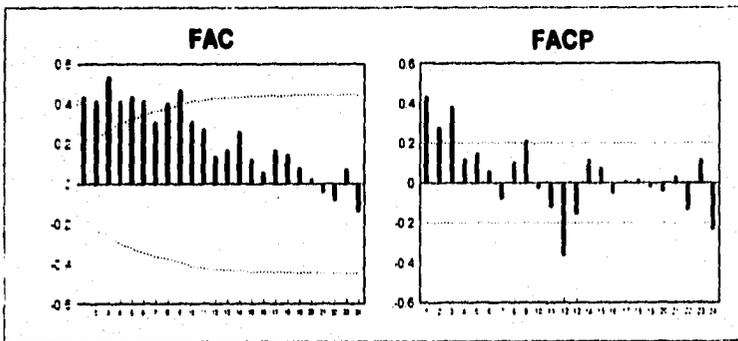
Y se calcularon las autocorrelaciones (FAC) y las autocorrelaciones parciales (FACP) de la nueva serie:



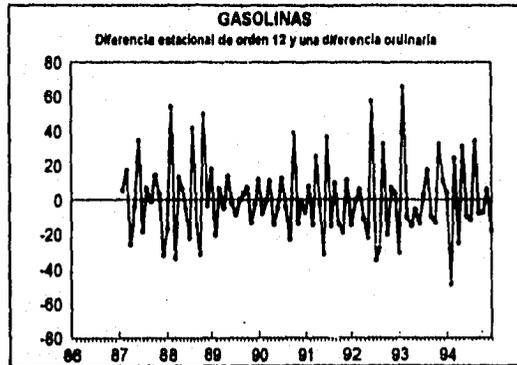
En éstos, los valores  $r_{11}$ ,  $r_{12}$  y  $r_{13}$  resultaron significativamente diferentes de cero y las autocorrelaciones parciales muestran valores diferentes de cero con una tendencia decreciente. El valor alto para  $r_{12}$  indica que la serie contiene un patrón estacional lo cual es congruente con la inspección visual de la gráfica de la serie original,  $T(Z_t)$ , en donde se aprecia un patrón que se repite cada año. Por lo tanto, es necesaria la aplicación de una diferencia estacional de orden 12 a la serie original ( $\nabla_{12}$ ).



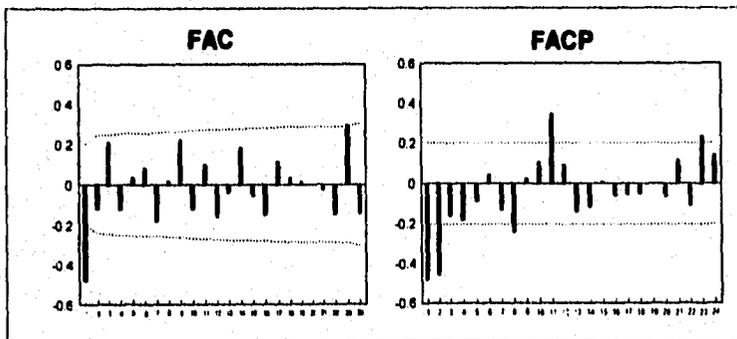
Las gráficas de la FAC y de la FACP para la serie  $(\nabla_{12}T(Z_t))$  son las siguientes:



Esta serie todavía no es estacionaria porque las autocorrelaciones decrecen lentamente, y entonces se aplicó otra diferencia de orden 1.



La nueva serie,  $\nabla\nabla^{12}T(Z_t)$ , es ya aproximadamente estacionaria como lo muestran las gráficas de las FAC y las FACP:



De hecho es la serie en la que el valor de la media es el más cercano a cero:

	$T(Z_t)$	$\nabla T(Z_t)$	$\nabla^{12} T(Z_t)$	$\nabla \nabla^{12} T(Z_t)$
$\mu$	4.2390	0.0199	0.2219	0.0001
$\sigma$	0.6781	0.2318	0.2024	0.2524

De aquí, ya se tienen los elementos para proponer un modelo a estimar:

- las diferencias que se aplicarán son una estacional de orden 12 y una ordinaria de orden 1;
- la gráfica de la FAC señala dos autocorrelaciones grandes,  $r_1$  y  $r_{23}$ , y en la FACP cinco autocorrelaciones parciales elevadas, correspondientes a los rezagos 1, 2, 8, 11 y 23.

Por lo tanto el modelo que se propone para estimar es un  $ARI(2,1,0)(0,1,0)_{12}$ , es decir:

$$(1 - \phi_1 B - \phi_2 B^2) \nabla \nabla_{12} T(Z_t) = a_t$$

Utilizando el paquete Autobox para estimar el modelo, se obtuvo el valor de los coeficientes y sus intervalos de confianza al 95%:

$$\begin{aligned} \hat{\phi}_1 &= -0.7089 & (-0.8905, -0.5273) \\ \hat{\phi}_2 &= -0.4639 & (-0.6454, -0.2824) \end{aligned}$$

Ahora se procede a verificar que no se violen los supuestos establecidos para estos modelos:

- $\{a_t\}$  tiene media cero.

El valor  $\sqrt{N-D-d-p} m(\hat{a})/\hat{\sigma}_a = -0.05220$ , es menor que 2 en valor absoluto por lo que no hay evidencia de que la media de  $\{a_t\}$  sea distinta de cero.

- $\{a_t\}$  tiene varianza constante.

La gráfica de los residuales no indica una violación grave a este supuesto.

3. Las  $\{a_t\}$  son variables aleatorias independientes entre sí.

En este caso el valor calculado de  $Q'$  es 66.0, mientras que en tablas  $X_{0.05}^2(22) = 33.92$ , con lo que se rechaza la hipótesis de que los residuales son ruido blanco; además, hay dos autocorrelaciones de los residuales que son distintas de cero:  $r_9 = 0.394$  con un error estandar de 0.114, y  $r_{12} = -0.316$  con error de 0.130.

Lo anterior indica que es necesario introducir en el modelo los parámetros correspondientes a éstos patrones.

La violación de este supuesto nos indica que este modelo no es adecuado para la serie, y para identificar otro se corregirá el ya estimado con los mismos elementos que llevaron a su rechazo. Así, como  $r_{12}$  es elevado y se asocia al orden de la estacionalidad que presenta la serie, se considera conveniente incluir un término de orden 12, ya sea autorregresivo o de promedios móviles. Un término de grado 9 para abatir el valor significativo de  $r_9$ , en principio se descarta ya que se espera que con la corrección anterior éste ya no aparezca y, sobretodo, porque aparentemente no hay un sustento teórico que justifique la inclusión de este término; no obstante, si la naturaleza de los datos lo requiere entonces se incluirá.

Por lo tanto se realizó la estimación de los modelos:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \phi B^{12}) \nabla \nabla_{12} T(Z_t) = a_t$$

y

$$(1 - \phi_1 B - \phi_2 B^2) \nabla \nabla_{12} T(Z_t) = (1 - \theta_{12} B^{12}) a_t$$

Al incluir el término de orden 12 en ambos modelos prevaleció la autocorrelación  $r_9$  grande, y el estadístico  $Q'$  indicaba que los residuales no era una serie de ruido blanco. Debido a que el error estándar de los residuales del primer modelo resultó más reducida se optó por agregar a éste un término de promédios móviles de orden 9:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \phi B^{12})\nabla \nabla_{12} T(Z_t) = (1 - \theta_9 B^9) a_t$$

y los valores estimados de los parámetros, con sus respectivos límites de confianza fueron:

$$\hat{\phi}_1 = -0.8241 \quad (-0.9960, -0.7541)$$

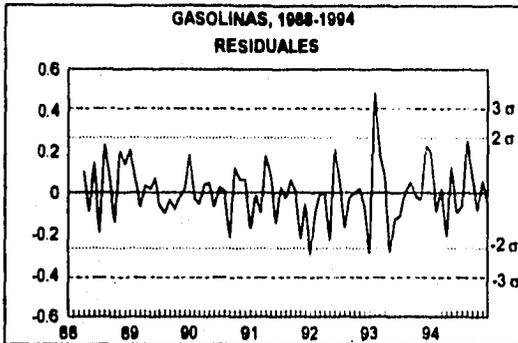
$$\hat{\phi}_2 = -0.5637 \quad (-0.7451, -0.3814)$$

$$\hat{\phi} = -0.4080 \quad (-0.6079, -0.2081)$$

$$\hat{\theta}_2 = -0.4474 \quad (-0.6673, -0.2277)$$

Con las correcciones anteriores, esta vez se concluye con la verificación de todos los supuestos:

1.  $\{a_t\}$  tiene media cero porque el valor  $\frac{\sum_{t=1}^n a_t}{N-D-d-p} = m(\hat{a})/\hat{\sigma}_a = 0.01424$ , es menor que 2 en valor absoluto.
2. La gráfica de los residuales no indica una violación grave al supuesto de que  $\{a_t\}$  tiene varianza constante.



3. Las  $(a_t)$  son variables aleatorias independientes entre sí.

El valor calculado de  $Q'$ , utilizando la expresión 2.5.4.3.2, es 20.0, mientras que en tablas  $N_{0.05}^2(20) = 31.41$ , y no se rechaza la hipótesis de que los residuales son ruido blanco; además, ya no hay autocorrelaciones de los residuales distintas de cero.

4.  $(a_t)$  se distribuye normalmente para toda  $t$ .

En la gráfica de los residuales, sólo 4 de ellos caen fuera de la banda  $(-2\hat{\sigma}_a, 2\hat{\sigma}_a)$ .

5. No existen observaciones aberrantes.

El residual correspondiente a enero de 1993 rebasa la banda de  $(-3\hat{\sigma}_a, 3\hat{\sigma}_a)$ ; se verificó que no fuera un error de captura de los datos y este no fue el caso; la fuente de la información contenía este dato y era congruente con los resultados que se presentan en otros documentos.

6. Parsimonia del modelo.

Ninguno de los intervalos de confianza de los parámetros estimados incluye al cero.

7. Admisibilidad del modelo.

Los resultados que arrojó el programa indican que todos los polinomios involucrados en el modelo tienen sus raíces características fuera del círculo unitario.

8. Estabilidad del modelo.

La correlación entre los parámetros  $\phi_1$  y  $\phi_2$  es 0.5590, por lo que uno de los dos parámetros debería eliminarse; se probó eliminar cada uno de ellos y en ambos casos las autocorrelaciones indicaban la necesidad de incluir ambos parámetros.

En el siguiente cuadro se presenta el resumen de la estimación de los modelos.

RESUMEN DE LA ESTIMACION DE LA DEMANDA DE GASOLINAS

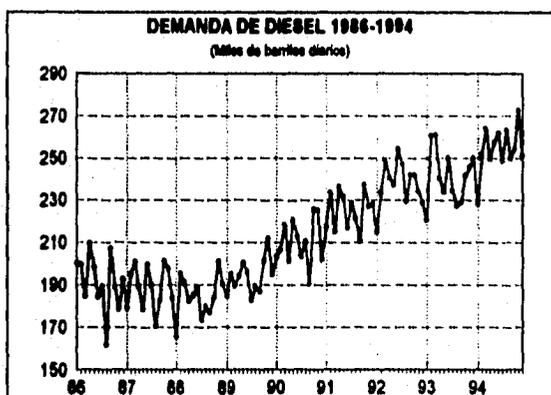
Periodo: Ene 1986-dic 1994

No. de observaciones: 108 Transformación:  $T(Z_t) = Z_t/100$

Modelo	Parámetros estimados	Intervalos del 95% de confianza	Correlación entre parámetros > 0.5 ó < -0.5	Análisis de residuos					
				$m(d)$	$\hat{\sigma}_e$	$\frac{m(d)}{\hat{\sigma}_e}$	$Q^*$ , g.l.	$r_a(d) \neq 0$	Residuales grandes
$(1 - \phi_1 B - \phi_2 B^2) \nabla \nabla_{111} T(Z_t) = \alpha$	$\hat{\phi}_1 = -0.7089$ $\hat{\phi}_2 = -0.4639$	(-0.8905, -0.5273) (-0.6454, -0.2824)		-0.00091	0.16801	-0.05220	66.0, 22	$r_a = 0.394$ (0.140) $r_{11} = -0.316$ (0.130)	$\hat{d}_{23} = 0.379$ $\hat{d}_{27} = -0.383$ $\hat{d}_{71} = -0.365$ $\hat{d}_{88} = -0.457$ $\hat{d}_{98} = 0.405$
$(1 - \phi_1 B - \phi_2 B^2) \nabla \nabla_{11} T(Z_t) = (1 - \theta_{11} B^{12}) \alpha$	$\hat{\phi}_1 = -0.8468$ $\hat{\phi}_2 = -0.5614$ $\hat{\theta}_{11} = 0.6245$	(-0.9977, -0.6957) (-0.7325, -0.3903) (0.4455, 0.8035)	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5345$	0.00043	0.14824	0.02803	35.0, 21	$r_a = 0.337$ (0.111)	$\hat{d}_{23} = 0.420$ $\hat{d}_{95} = 0.363$ $\hat{d}_{88} = -0.393$ $\hat{d}_{98} = 0.382$
$(1 - \phi_1 B - \phi_2 B^2) (1 - \phi B^{11}) \nabla \nabla_{11} T(Z_t) = \alpha$	$\hat{\phi}_1 = -0.8282$ $\hat{\phi}_2 = -0.5171$ $\hat{\phi} = -0.4262$	(0.9991, -0.6573) (-0.7062, -0.3890) (-0.6915, -0.2109)	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5724$	0.00022	0.15161	0.01286	49.0, 21	$r_a = 0.438$ (0.118)	$\hat{d}_{95} = 0.301$ $\hat{d}_{88} = -0.395$ $\hat{d}_{91} = 0.354$ $\hat{d}_{98} = -0.308$
$(1 - \phi_1 B - \phi_2 B^2) (1 - \phi B^{12}) \nabla \nabla_{11} T(Z_t) = (1 - \theta_{11} B^{11}) \alpha$	$\hat{\phi}_1 = -0.8768$ $\hat{\phi}_2 = -0.5637$ $\hat{\phi} = -0.4080$ $\hat{\theta}_{11} = -0.4474$	(-0.9960, -0.7400) (-0.7451, -0.3814) (-0.6679, -0.2611) (-0.6673, -0.2277)	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5590$	0.00022	0.13697	0.01424	20.1, 21		$\hat{d}_{73} = 0.296$ $\hat{d}_{95} = 0.293$ $\hat{d}_{88} = -0.485$ $\hat{d}_{98} = 0.287$

### 4.3.2 Diesel

La serie del diesel muestra de 1990 a 1992 cierto patrón estacional; sin embargo, éste no se encuentra bien definido en todos los años, lo que se refleja en el hecho de que la serie no admite diferencias estacionales de orden 12.



En la construcción de este modelo destaca que la inclusión de un término autorregresivo de orden 10, que se derivó del análisis de la autocorrelación de los residuales del primer modelo identificado; este término fue no significativo y, sobretodo, no abatió la necesidad de incluir un término constante para que la media de los residuales fuera cero; por tal motivo se decidió eliminar este parámetro. El modelo más adecuado fué el siguiente:

$$(1 + .693B - 0.217B^{12})7(Z, / 100) - 0.010 + (1 - 0.7650B^2)a_t$$

El proceso para encontrar el modelo más adecuado se muestra en el siguiente cuadro:

RESUMEN DE LA ESTIMACION DE LA DEMANDA DE DIESEL

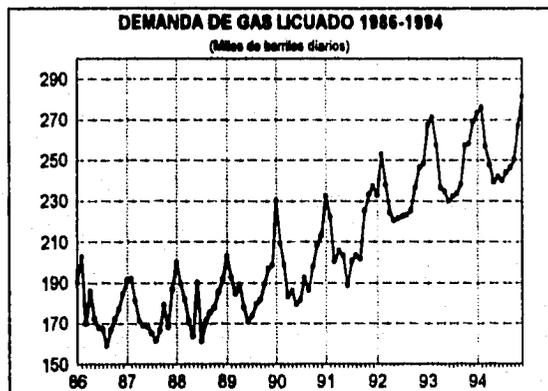
Periodo: Ene 1986-dic 1994

No. de observaciones: 108 Transformación:  $T(Z_t) = Z_t/100$

Modelo	Parámetros estimados	Intervalos del 95% de confianza	Correlación entre parámetros > 0.5 & < -0.5	Análisis de residuos					
				$m(\hat{\alpha})$	$\hat{\sigma}_e$	$\frac{n(n+1)}{6}$	$Q^*$ g.l.	$r_e(\hat{\alpha}) = 0$	Residuales grandes
$VT(Z_t) = (1 - \theta_1 B - \theta_2 B^2)\alpha_t$	$\hat{\theta}_1 = 0.7692$ $\hat{\theta}_2 = 0.0022$	(-0.5781, 0.9603) (-0.2134, 0.2178)	$r(\hat{\theta}_1, \hat{\theta}_2) = -0.6995$	0.02017	0.11542	1.80750	54.0, 22	$r_{22} = -0.213$ (0.097) $r_{14} = -0.236$ (0.109) $r_{13} = 0.323$ (0.114)	$d_{22} = 0.327$ $d_{23} = 0.239$ $d_{24} = -0.292$
$(1 - \theta_1 B - \theta_{11} B^{11})VT(Z_t) = (1 - \theta_2 B^2 - \theta_{12} B^{12})\alpha_t$	$\hat{\theta}_1 = -0.5413$ $\hat{\theta}_{12} = 0.2717$ $\hat{\theta}_2 = 0.5247$ $\hat{\theta}_{18} = 0.0911$	(-0.7100, -0.3226) (-0.1141, 0.4293) (0.1192, 0.3690) (0.3472, 0.7726)		0.02384	0.10186	2.28100	29.0, 21		$d_{22} = -0.254$ $d_{23} = -0.204$ $d_{24} = -0.251$ $d_{25} = 0.2531$ $d_{26} = -0.251$ $d_{27} = -0.284$
$(1 - \theta_1 B - \theta_{11} B^{11} - \theta_{12} B^{12})VT(Z_t) = (1 - \theta_2 B^2)\alpha_t$	$\hat{\theta}_1 = -0.5524$ $\hat{\theta}_{12} = -0.1627$ $\hat{\theta}_{18} = -0.2441$ $\hat{\theta}_2 = 0.5599$	(-0.7323, -0.3725) (-0.3146, -0.0108) (0.1192, 0.3690) (0.3472, 0.7726)		0.025926	0.09970	2.33890	23.0, 21		$d_{22} = 0.2032$ $d_{23} = -0.199$ $d_{24} = -0.237$ $d_{25} = 0.273$ $d_{26} = -0.243$
$(1 - \theta_1 B - \theta_{11} B^{11} - \theta_{12} B^{12})VT(Z_t) = \alpha_t - (1 - \theta_2 B^2)\alpha_t$	$\hat{\theta}_1 = -0.6766$ $\hat{\theta}_{12} = -0.1474$ $\hat{\theta}_{18} = 0.2849$ $\hat{\theta}_2 = 0.0113$ $\hat{\theta}_3 = 0.7100$	(-0.8632, -0.5170) (-0.2956, 0.0008) (0.1383, 0.4315) (0.0051, 0.0175) (0.5217, 0.8983)	$r(\hat{\theta}_1, \hat{\theta}_2) = -0.5500$	0.00068	0.09897	0.06697	23.0, 21		$d_{22} = -0.198$ $d_{23} = 0.2407$ $d_{24} = -0.214$ $d_{25} = 0.291$
$(1 - \theta_1 B - \theta_{11} B^{11})VT(Z_t) = \alpha_t - (1 - \theta_2 B^2)\alpha_t$	$\hat{\theta}_1 = -0.6931$ $\hat{\theta}_{12} = 0.2168$ $\hat{\theta}_2 = 0.0108$ $\hat{\theta}_3 = 0.7650$	(-0.8449, -0.5413) (0.0865, 0.3471) (0.0053, 0.0163) (0.6013, 0.9287)	$r(\hat{\theta}_1, \hat{\theta}_2) = -0.5320$	0.00025	0.10108	0.02172	32.0, 21.0		$d_{22} = 0.252$ $d_{23} = -0.238$ $d_{25} = 0.268$

### 4.3.3 Gas licuado

La gráfica del consumo mensual de gas LP muestra un comportamiento estacional bien definido, lo que se comprobó al determinar el grado de diferenciación que requería la serie para ser aproximadamente estacionaria.



En este caso, al igual que las gasolinas, se requirió de la aplicación de una diferencia estacionaria de orden 12 y una diferencia ordinaria de grado 1.

Inicialmente se identificó un modelo con un parámetro de promedios móviles de grado 1 y otro de grado 12; sin embargo, la correlación alta entre los parámetros y las autocorrelaciones de los rezagos 1 y 12 significativamente distintas de cero llevaron a ensayar el modelo:

$$(1 - \phi_{12}B^{12})\nabla_{12}T(Z_t) = (1 - \theta_1B)a_t$$

el cual resultó adecuado y el valor de los parámetros se muestra al final del siguiente cuadro:

RESUMEN DE LA ESTIMACION DE LA DEMANDA DE GAS LICUADO

Periodo: Ene 1986-dic 1994

No. de observaciones: 108 Transformación:  $T(Z_t) = Z_t/100$

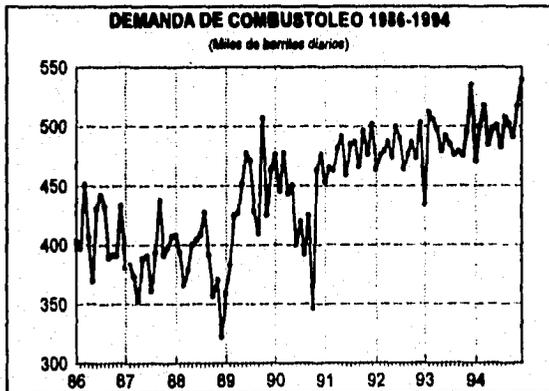
Modelo	Parámetros estimados	Intervalos del 95% de confianza	Correlación entre parámetros > 0.5 ó < -0.5	Análisis de residuos					
				$m(d)$	$\hat{\sigma}_e$	$\frac{m(d)}{\hat{\sigma}_e}$	$\sigma^2$ s.l.	$r_1(d) \neq 0$	Residuales grandes
$\forall \forall_{12} T(Z_t) = (1 - \theta_1 R - \theta_{12} R^{12}) a_t$	$\hat{\theta}_1 = 0.7575$ $\hat{\theta}_{12} = 0.1613$	(0.9170, 0.5980) (0.0149, 0.3077)	$r(\hat{\theta}_1, \hat{\theta}_{12}) = -0.7643$	0.01247	0.08376	1.45100	33.0, 22	$r_1 = -0.225$ (0.103) $r_{12} = -0.277$ (0.114)	$\hat{d}_{20} = -0.216$ $\hat{d}_{42} = 0.260$ $\hat{d}_{46} = -0.201$ $\hat{d}_{52} = 0.171$ $\hat{d}_{72} = 0.188$ $\hat{d}_{76} = -0.177$ $\hat{d}_{78} = -0.191$ $\hat{d}_{85} = -0.181$
$(1 - \theta_{12} R^{12}) \forall \forall_{12} T(Z_t) = (1 - \theta_1 R) a_t$	$\hat{\theta}_{12} = -0.3804$ $\hat{\theta}_1 = 0.7797$	(-0.5847, -0.1761) (0.6366, 0.9227)		0.00253	0.08158	0.28227	19.0, 22		$\hat{d}_{20} = -0.208$ $\hat{d}_{42} = 0.197$ $\hat{d}_{46} = -0.181$ $\hat{d}_{72} = 0.229$ $\hat{d}_{76} = -0.174$

#### 4.3.4 Combustóleo

La serie del combustóleo requirió solamente una diferencia ordinaria de grado 1. Se identificó un modelo AR(2), pero la FAC de los residuales sugirió la inclusión de un término de grado 11, y se decidió que éste fuera de promédios móviles. De esta manera el modelo para esta serie fue:

$$(1 + 0.605B + 0.278B^2)\nabla(Z_t/100) = (1 + 0.230B^{11})a_t$$

Los resultados se muestran en el siguiente cuadro:



RESUMEN DE LA ESTIMACION DE LA DEMANDA DE COMBUSTOLEO

Periodo: Ene 1986-dic 1994

No. de observaciones: 108 Transformación:  $T(Z_t) = Z_t/100$

Modelo	Parámetros estimados	Intervalos del 95% de confianza	Correlación entre parámetros > 0.5 ó < -0.5	Análisis de residuales					
				$m(d)$	$\hat{\sigma}_e$	$\frac{n(d)}{n_e}$	$\theta^*$ g.l.	$r_e(d) \neq 0$	Residuales grandes
$(1 - \phi_1 B - \phi_2 B^2)VT(Z_t) = a_t$	$\hat{\phi}_1 = -0.6334$ $\hat{\phi}_2 = -0.3201$	(-0.7903, -0.4193) (-0.4972, -0.1429)	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5025$	0.00371	0.07410	0.51254	30.0, 22	$r_{11} = 0.231$ (0.104)	$\hat{a}_{12} = -0.720$ $\hat{a}_{22} = 0.683$ $\hat{a}_{32} = -0.670$ $\hat{a}_{42} = -0.507$ $\hat{a}_{52} = -0.696$ $\hat{a}_{124} = -0.724$
$(1 - \phi_1 B - \phi_2 B^2)VT(Z_t) = (1 - \theta_1 B)u_t$	$\hat{\phi}_1 = -0.6018$ $\hat{\phi}_2 = -0.2779$ $\hat{\theta}_1 = -0.2296$	(-0.7903, -0.4193) (-0.4601, -0.0957) (-0.4332, 0.0260)	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5003$	0.00333	0.07210	0.47221	24.0, 21		$\hat{a}_{12} = -0.731$ $\hat{a}_{22} = 0.655$ $\hat{a}_{32} = -0.659$ $\hat{a}_{42} = -0.586$ $\hat{a}_{52} = 0.526$ $\hat{a}_{124} = -0.794$

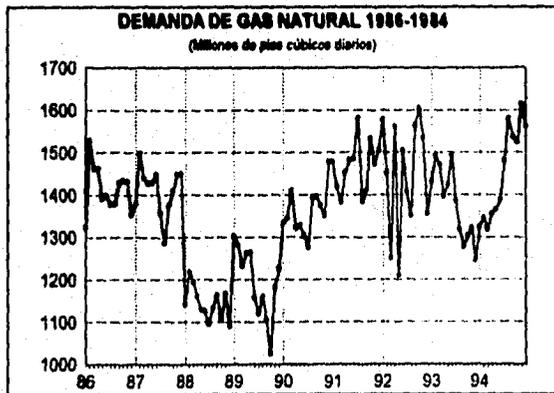
#### 4.3.5 Gas natural

La serie del gas natural requirió solamente una diferencia ordinaria de grado 1. En este caso el análisis de FAC y de la FCAP llevó a la identificación de dos modelos:

$$(1 - \phi_1 B - \phi_2 B^2) \nabla T(Z_t) = a_t$$

$$Y \quad (1 - \phi_1 B) \nabla T(Z_t) = (1 - \theta_2 B^2) a_t$$

Ambos modelos fueron adecuados y se optó por utilizar el segundo para calcular los pronósticos porque presentaba la menor desviación estándar de los residuales (véase cuadro siguiente).



RESUMEN DE LA ESTIMACION DE LA DEMANDA DE GAS NATURAL

Periodo: Ene 1986-dic 1994

No. de observaciones: 108 Transformación:  $T(Z_t) = Z_t / 1000$

Modelo	Parámetros estimados	Intervalos del 95% de confianza	Correlación entre parámetros > 0.5 ó < -0.5	Análisis de residuales					
				$m(\hat{a})$	$\hat{\sigma}_a$	$\frac{\sigma(\hat{a})}{\hat{\sigma}_a}$	$Q^2$ , g.l.	$r_a(\hat{a}) = 0$	Residuales grandes
$(1 - \phi_1 B - \phi_2 B^2) \nabla T(Z_t) = \alpha_t$	$\hat{\phi}_1 = -0.4695$ $\hat{\phi}_2 = -0.1990$	(-0.6585, -0.2805) (-0.3847, -0.0133)		0.00171	0.08621	0.20400	12.0, 22		$\hat{a}_{25} = 0.300$ $\hat{a}_{75} = 0.245$ $\hat{a}_{77} = 0.246$ $\hat{a}_{78} = -0.193$ $\hat{a}_{95} = 0.204$
$(1 - \phi_1 B) \nabla T(Z_t) = (1 - \theta_1 B^2) \alpha_t$	$\hat{\phi}_1 = -0.4670$ $\hat{\theta}_1 = 0.2440$	(-0.6501, -0.2839) (0.0352, 0.4128)		0.0187	0.08571	0.22505	11.0, 22		$\hat{a}_{25} = 0.293$ $\hat{a}_{77} = -0.187$ $\hat{a}_{78} = 0.239$ $\hat{a}_{79} = 0.265$ $\hat{a}_{81} = -0.180$ $\hat{a}_{95} = 0.181$

#### 4.3.6 Los pronósticos mensuales

Con los modelos ARIMA univariados estimados se procedió a calcular los pronósticos mensuales para 1995 de la demanda de los cinco productos petrolíferos seleccionados.

Los pronósticos obtenidos por este método se determinan únicamente por el proceso estocástico que se identificó del propio comportamiento de la serie durante nueve años. Con el crecimiento del consumo de petrolíferos registrado en los años anteriores, acorde al desarrollo de la economía en general, era fácil prever que la demanda durante 1995 mostraría una tendencia creciente también, y ello se refleja en los pronósticos que se obtuvieron a través de los modelos ARIMA: el crecimiento anual promedio calculado para los pronósticos es de alrededor de 3% para las gasolinas y el diesel, de 4% para el combustóleo y el gas licuado, y de 8% para el gas natural.

Los resultados obtenidos se muestran a continuación:

PRONOSTICOS ARIMA DE LA DEMANDA DE PETROLIFEROS, 1995										
	Gasolinas		Diesel		Gas licuado		Combustóleo		Gas natural	
	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mpcd	$\bar{\sigma}$
Ene	485.2	13.70	244.2	10.11	284.8	8.16	488.2	28.80	1,564.8	85.71
Feb	534.1	13.80	258.9	10.57	287.9	8.35	471.0	30.85	1,566.1	97.13
Mar	524.7	14.39	252.4	17.37	270.9	8.54	472.5	31.27	1,565.5	103.79
Abr	503.1	16.84	254.9	18.51	257.2	8.73	474.3	31.40	1,565.8	111.54
May	523.4	16.86	255.9	20.01	251.3	8.91	464.6	31.97	1,565.6	111.83
Jun	515.5	17.57	257.3	21.12	251.1	9.09	461.5	31.97	1,565.7	112.85
Jul	499.6	18.50	254.6	21.36	250.9	9.27	485.0	31.99	1,565.7	113.39
Ago	523.6	18.50	280.6	22.10	253.8	9.44	488.0	32.02	1,565.7	113.41
Sep	509.7	19.06	254.7	22.10	257.0	9.61	474.8	32.03	1,565.7	113.58
Oct	505.3	19.23	260.8	22.47	266.7	9.78	476.2	32.03	1,565.7	113.80
Nov	532.3	19.33	261.6	22.49	277.3	9.94	466.5	32.03	1,565.7	113.81
Dic	556.6	19.45	257.5	22.63	290.3	10.10	472.1	32.86	1,565.7	113.83
Prom.	516.1		256.1		266.6		471.2		1,565.6	

#### 4.4 Pronósticos condicionales para 1995

El escenario que se desarrolló para calcular los pronósticos ARIMA condicionales se basa en los pronósticos ARIMA para 1995 que se presentaron en la sección 4.2.6 y la restricción a la que se sujetarán está dada por la proyección anual promedio de la demanda de los petrolíferos que se obtuvo a través de los modelos de regresión lineal múltiple y que se presentaron en la sección 4.1.6. De esta manera, se van a calcular los pronósticos mensuales de la demanda de petrolíferos para 1995, por lo que  $H=12$ , y la restricción que se aplicará es el pronóstico anual por lo que  $m=1$ . Este pronóstico anual corresponde al promedio, por lo que la combinación lineal que relaciona a los pronósticos ARIMA con la restricción está dada por la matriz  $C=(1/12, 1/12, \dots, 1/12)$ . Por su parte,  $E(Z_F | Z_0)$  es el vector formado por los 12 pronósticos ARIMA y  $Y$  es la demanda anual promedio estimada. Las ponderaciones  $\psi$  se calcularon con los valores de los parámetros estimados para cada uno de los modelos, mediante la expresión 2.5.6.5, y se presentan en el siguiente cuadro:

VALOR DE LAS PONDERACIONES $\psi$					
t	Gasolinas	Diesel	Gas licuado	Combustóleo	Gas natural
1	-0.1230	-0.3070	-0.2200	-0.3840	-0.5330
2	0.2979	1.3839	-0.2200	0.1781	0.4269
3	0.6390	0.6313	-0.2200	-0.1010	0.4765
4	0.0549	-0.7820	-0.2200	-0.2082	0.0546
5	-0.3610	-0.9890	-0.2200	0.0042	-0.1930
6	-0.4220	0.3159	-0.2200	0.0395	-0.1290
7	0.0300	0.5601	-0.2200	-0.0470	0.0218
8	0.3384	-0.0470	-0.2200	-0.0300	0.0717
9	-0.1770	-0.4030	-0.2200	0.0154	0.0280
10	-0.1450	-0.0010	-0.2200	0.0017	-0.0190
11	-0.1640	0.2511	-0.2200	-0.2640	-0.0230

## 4.4.1 Resultados

Para calcular los pronósticos condicionales se utilizó la versión 4.02 (1992) del paquete estadístico PC-RATS (Regression Analysis and Time Series). En primer lugar se calcularon los pronósticos suponiendo que las restricciones dadas por las proyecciones del modelo de regresión eran conjeturas exactas sobre el comportamiento que seguiría el consumo de los petrolíferos; en este caso, para calcular los pronósticos ARIMA condicionales se empleó la expresión 3.3.8, es decir, el procedimiento en donde únicamente se calcula la matriz  $A$ , al cual se hará referencia como Procedimiento A. En el siguiente cuadro se muestran los pronósticos condicionales obtenidos, así como sus desviaciones estándar y el valor del estadístico  $K_1$ , obtenido mediante la expresión 3.4.2:1/

PRONOSTICOS CONDICIONALES PARA LA DEMANDA MENSUAL DE PETROLIFEROS, 1995 Procedimiento A										
	Gasolinas		Diesel		Gas licuado		Combustóleo		Gas natural	
	Mbd	$\hat{\sigma}$	Mbd	$\hat{\sigma}$	Mbd	$\hat{\sigma}$	Mbd	$\hat{\sigma}$	Mpcd	$\hat{\sigma}$
Ene	452.9	13.38	228.5	9.65	312.5	6.97	413.3	28.65	1,493.9	81.81
Feb	521.4	13.46	250.1	10.44	305.2	7.82	372.6	30.37	1,531.5	96.33
Mar	506.7	13.73	220.8	16.28	278.7	8.48	389.4	30.93	1,500.3	101.09
Abr	474.6	15.40	213.1	18.68	256.5	8.73	388.7	31.05	1,488.4	105.88
May	497.1	15.63	226.9	19.22	243.1	8.82	387.8	31.69	1,466.7	108.79
Jun	491.6	16.60	236.9	20.78	236.3	8.81	389.6	31.73	1,478.5	108.66
Jul	475.9	17.62	226.9	20.88	230.5	8.72	423.7	31.81	1,483.1	109.42
Ago	499.9	17.60	227.6	21.17	228.7	8.62	401.7	31.81	1,477.2	109.83
Sep	483.4	18.01	228.0	21.50	228.1	8.52	363.8	31.45	1,478.9	109.19
Oct	483.1	18.48	236.8	21.99	235.6	8.48	362.6	31.42	1,494.0	110.63
Nov	508.4	18.54	228.6	21.58	243.7	8.48	385.1	31.79	1,500.3	111.14
Dic	537.9	18.92	221.7	21.57	255.8	8.59	296.3	31.45	1,471.0	108.38
Prom.	494.5		228.8		254.5		382.1		1,487.1	
$K_1$	17.38		27.16		42.46		332.68		7.70	

1 /Para estos ejercicios los estadísticos  $K_1$  y  $K_2$  tienen el mismo valor porque se tiene una sola restricción ( $m = 1$ ).

En todos los casos, el valor del estadístico  $\hat{K}_1$ , es muy elevado, comparado con el valor de tablas  $\chi^2(0.05) = 3.84$ . Ello, de acuerdo con los criterios establecidos, muestra evidencia de que el valor de las restricciones no es compatible con el proceso que generó la serie de tiempo para la que se construyó el modelo ARIMA. Por ello, fue necesario incluir la matriz  $U$ , dada en la ecuación 3.5.2, cuya dimensión es  $m \times 1$ , es decir  $1 \times 1$  y cuyo valor se decidió que fuera la desviación estándar del pronóstico obtenido con el modelo de regresión lineal. Los resultados derivados de la expresión 3.5.2, que se denominará Procedimiento AU, se muestran enseguida:

PRONOSTICOS CONDICIONALES PARA LA DEMANDA MENSUAL DE PETROLIFEROS, 1995										
Procedimiento AU										
	Gasolinas		Diesel		Gas licuado		Combustóleo		Gas natural	
	Mbd	$\hat{\sigma}$	Mbd	$\hat{\sigma}$	Mbd	$\hat{\sigma}$	Mbd	$\hat{\sigma}$	Mpcd	$\hat{\sigma}$
Ene	484.0	13.67	243.2	10.06	284.9	8.15	485.6	28.90	1,563.4	85.63
Feb	533.0	13.77	258.3	10.57	288.0	8.35	486.2	30.83	1,565.4	87.11
Mar	523.0	14.33	250.4	17.31	270.9	8.54	488.4	31.25	1,564.2	103.73
Abr	500.4	16.72	252.3	18.40	257.2	8.73	470.2	31.38	1,563.8	111.42
May	521.0	16.75	254.1	19.98	251.2	8.91	460.9	31.98	1,563.6	111.52
Jun	513.3	17.48	256.0	21.10	251.0	9.09	458.0	31.98	1,563.9	112.76
Jul	497.4	18.42	252.8	21.32	250.8	9.26	482.1	31.98	1,564.0	113.31
Ago	521.4	18.42	258.5	22.04	253.7	9.44	484.8	32.01	1,563.8	113.32
Sep	507.3	18.98	253.0	22.08	256.8	9.60	468.5	32.00	1,563.9	113.49
Oct	503.2	19.18	259.3	22.44	266.6	9.77	470.7	32.00	1,564.2	113.54
Nov	530.2	19.28	259.5	22.44	277.1	9.93	463.1	32.02	1,564.3	113.56
Dic	554.9	19.40	255.2	22.57	290.1	10.09	463.7	32.78	1,563.7	113.52
Prom.	514.1		254.4		266.5		466.9		1,564.0	
$K_U$	1.61		1.72		0.24		15.97		0.16	

El empleo de este procedimiento, si bien lleva a valores del estadístico  $K_U$ , calculado mediante la expresión 3.5.6, con los que no se rechaza la hipótesis de que las restricciones con su margen de error son consistentes con los datos históricos

-con excepción del combustóleo- en este caso, en que la combinación lineal es un promedio, genera pronósticos casi iguales a los pronósticos ARIMA originales; incluso calculando por ensayo y error el valor de  $U$  para el cual se satisfaga el criterio de consistencia, los pronósticos resultantes son cada vez más cercanos a los ARIMA.

Lo anterior justifica el planteamiento que hace Pérez-Porrúa (1984) de que, cuando la precisión que se asigna a las restricciones es amplia, entonces la opción es mantener los pronósticos ARIMA; y si se asigna un alto valor de confianza a las conjeturas entonces los pronósticos satisfarán exactamente la conjetura.

Por lo tanto, los pronósticos que se tomaron fueron los que se obtuvieron con el Procedimiento A.

#### **4.4.2 Evaluación de los pronósticos**

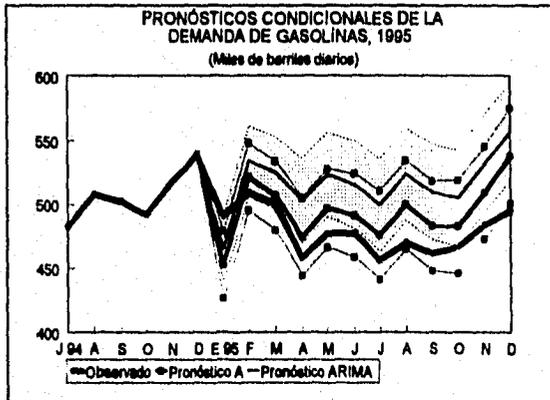
La evaluación de los pronósticos calculados se hizo respecto a los datos ya observados. Lo sucedido en 1995 resultó un buen ejemplo para mostrar la ventaja de combinar pronósticos que provienen de modelos que utilizan información distinta. Así, con el comportamiento que había registrado el consumo de petrolíferos hasta 1994 era obvio esperar un crecimiento moderado en la demanda de 1995; la propia historia de las series, considerada en los modelos ARIMA, justifica también este planteamiento.

Por otro lado, en los primeros meses de 1995 se tuvieron indicadores que permitían prever cambios sustanciales en muchas variables económicas. Los factores que a primera vista impactarían directamente el consumo de energéticos eran la

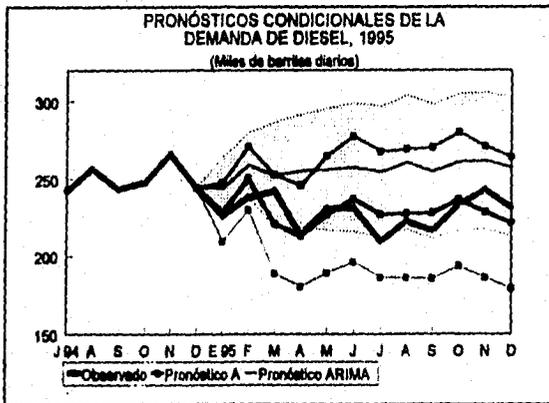
reducción de la actividad económica y el incremento de los precios de los petrolíferos, tanto por la política aplicada, como por el efecto de la devaluación del peso. Esta información anticipaba un descenso de la demanda de los petrolíferos, contrario a lo que indicaba el sólo comportamiento histórico de las series.

La integración de toda la información, a través del método de pronósticos condicionales, permitió obtener pronósticos más precisos. En los siguientes cuadros se presenta para cada producto la comparación entre los pronósticos obtenidos con los modelos ARIMA y los condicionales, tanto con el Procedimiento A (A), como con el AU (AU), y las observaciones reales en 1995. Asimismo, en las gráficas se muestra: la banda de confianza de los los pronósticos ARIMA univariados (sombreada), los pronósticos ARIMA condicionales utilizando el Procedimiento A (Pronóstico A), y los datos observados.

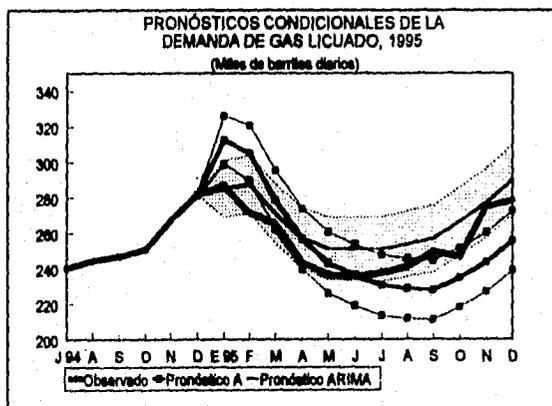
EVALUACION DE LOS PRONOSTICOS DE LA DEMANDA MENSUAL DE GASOLINAS, 1995 (Miles de barriles diarios)							
	Obs.	Pronósticos			Diferencia % con observado		
		ARIMA	A	AU	ARIMA	A	AU
Ene	491.0	485.17	457.0	481.54	-5.3	-8.9	-6.0
Feb	508.8	534.14	524.7	529.96	5.0	3.1	4.2
Mar	500.6	524.69	512.5	519.32	4.8	2.4	3.8
Abr	458.3	503.08	487.0	495.95	9.8	6.3	8.2
May	477.0	523.40	503.5	514.60	9.7	5.8	7.9
Jun	477.7	516.84	492.1	505.18	7.9	3.0	5.7
Jul	486.1	499.56	473.0	487.82	9.5	3.7	7.0
Ago	489.9	523.65	484.5	510.77	11.4	5.2	8.7
Sep	481.8	509.71	478.5	495.93	10.4	3.8	7.4
Oct	486.6	505.26	473.5	491.22	8.3	1.5	5.3
Nov	484.2	532.33	504.2	518.91	9.9	4.1	7.4
Dic	495.2	556.60	533.2	546.23	12.4	7.7	10.3
Prom.	478.9	518.09	494.5	506.53	7.8	3.2	5.8



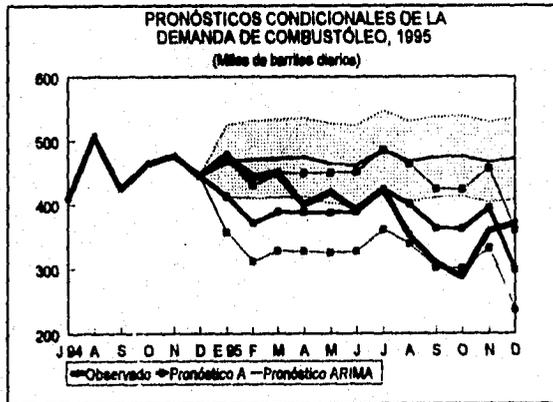
EVALUACION DE LOS PRONOSTICOS DE LA DEMANDA MENSUAL DEL DIESEL, 1995 (Miles de barriles diarios)							
	Obs.	Pronósticos			Diferencia % con observado		
		ARIMA	A	AU	ARIMA	A	AU
Ene	226.2	241.65	227.1	240.65	6.8	0.4	6.4
Feb	237.8	255.78	235.2	254.37	7.6	-1.1	7.0
Mar	242.0	250.66	230.8	249.30	3.6	-4.6	3.0
Abr	213.3	250.10	227.1	248.52	17.3	6.4	16.5
May	230.3	252.35	230.3	250.84	9.6	-0.0	8.9
Jun	231.3	252.18	228.8	250.57	9.0	-1.1	8.3
Jul	209.8	249.03	224.8	247.37	18.7	7.1	17.9
Ago	222.6	254.35	230.4	252.71	14.3	3.5	13.5
Sep	216.4	248.08	223.4	246.39	14.6	3.2	13.9
Oct	233.6	252.74	229.2	251.13	8.2	-1.9	7.5
Nov	242.9	254.61	231.4	253.01	4.8	-4.7	4.2
Dic	231.4	248.23	227.5	246.81	7.3	-1.7	6.7
Prom.	228.1	250.81	228.8	249.31	9.9	0.3	9.3



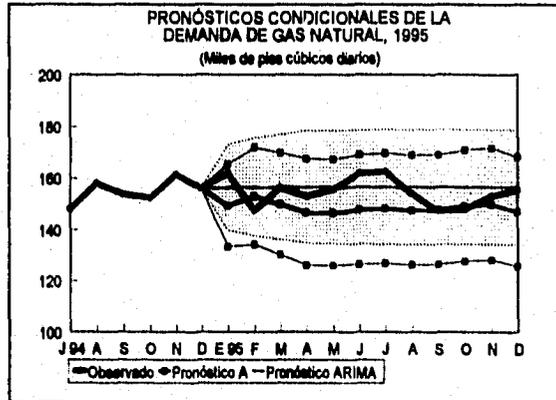
EVALUACION DE LOS PRONOSTICOS DE LA DEMANDA MENSUAL DEL GAS LICUADO, 1995 (Miles de barriles diarios)							
	Obs.	Pronósticos			Diferencia % con observado		
		ARIMA	A	AU	ARIMA	A	AU
Ene	288.6	284.80	276.4	284.59	-0.6	-3.6	-0.7
Feb	271.4	287.88	277.7	287.63	6.1	2.3	6.0
Mar	285.3	270.87	259.3	270.59	2.1	-2.3	2.0
Abr	243.5	257.18	244.7	256.85	6.6	0.5	5.5
May	236.1	251.27	238.1	250.85	6.4	0.9	6.3
Jun	235.4	251.09	237.5	250.76	6.7	0.9	6.5
Jul	237.4	250.87	237.2	250.54	5.7	-0.1	5.5
Ago	240.9	253.85	240.2	253.52	5.4	-0.3	5.2
Sep	249.4	256.98	243.7	256.64	3.0	-2.3	2.9
Oct	246.7	266.73	254.0	266.42	8.1	3.0	8.0
Nov	275.1	277.33	265.5	277.04	0.8	-3.5	0.7
Dic	278.3	290.32	278.7	290.07	4.3	0.5	4.2
Prom.	255.5	266.59	254.5	266.30	4.3	-0.4	4.2



EVALUACION DE LOS PRONOSTICOS DE LA DEMANDA MENSUAL DEL COMBUSTOLEO, 1995 (Miles de barriles diarios)							
	Obs.	Pronósticos			Diferencia % con observado		
		ARIMA	A	AU	ARIMA	A	AU
Ene	477.4	455.62	409.8	439.98	-4.6	-14.2	-7.8
Feb	443.2	455.62	393.4	434.40	2.8	-11.2	-2.0
Mar	450.1	455.62	383.8	431.10	1.2	-14.7	-4.2
Abr	400.5	455.62	377.2	428.86	13.8	-5.8	7.1
May	418.6	455.62	373.6	427.62	8.6	-11.0	1.9
Jun	392.6	455.62	371.5	426.92	16.1	-5.3	8.8
Jul	424.7	455.62	371.0	426.75	7.3	-12.6	0.5
Ago	353.3	455.62	371.8	427.02	29.0	5.2	20.9
Sep	309.1	455.62	374.0	427.78	47.4	21.0	38.4
Oct	287.8	455.62	378.3	428.24	59.3	31.5	49.1
Nov	358.0	455.62	384.9	431.49	26.9	7.2	20.2
Dic	371.4	455.62	395.8	435.21	22.7	6.6	17.2
Prom.	390.7	455.62	382.1	430.53	16.6	-2.2	10.2



EVALUACION DE LOS PRONOSTICOS DE LA DEMANDA MENSUAL DEL GAS NATURAL, 1995 (Miles de pies cúbicos diarios)							
	Obs.	Pronósticos			Diferencia % con observado		
		ARIMA	A	AU	ARIMA	A	AU
Ene	1,618.6	1,564.8	1,518.9	1,562.4	-3.4	-6.2	-3.5
Feb	1,477.4	1,566.1	1,495.9	1,562.3	6.0	1.3	5.7
Mar	1,562.9	1,565.5	1,485.6	1,561.2	0.2	-4.9	-0.1
Abr	1,531.5	1,565.7	1,480.4	1,561.2	2.2	-3.3	1.9
May	1,567.4	1,565.6	1,478.3	1,561.0	0.5	-5.1	0.2
Jun	1,618.5	1,565.7	1,477.5	1,561.0	-3.3	-8.9	-3.6
Jul	1,626.8	1,565.7	1,477.4	1,561.0	-3.8	-9.2	-4.0
Ago	1,540.3	1,565.7	1,477.8	1,561.0	1.6	-4.1	1.3
Sep	1,474.9	1,565.7	1,478.0	1,561.0	6.2	0.3	5.8
Oct	1,479.9	1,565.7	1,482.2	1,561.2	5.8	0.2	5.5
Nov	1,530.3	1,565.7	1,488.5	1,561.6	2.3	-2.7	2.0
Dic	1,555.2	1,565.7	1,503.4	1,562.3	0.7	-3.3	0.5
Prom.	1,548.0	1,565.6	1,487.1	1,561.4	1.1	-3.9	0.9



#### 4.5 Pronósticos para 1996

Con el programa económico propuesto por el Gobierno Federal para 1996 y la información observada hasta diciembre de 1995 se calcularon las proyecciones anuales de la demanda de los productos y se actualizó la estimación de los modelos ARIMA para las series mensuales.

##### 4.5.1 Pronósticos anuales

Para 1996 se consideró el escenario que sirvió de base para establecer el presupuesto del sector público y que se encuentra en el documento Criterios Generales de Política Económica para 1996. En éste se plantea, para las principales variables involucradas, un crecimiento de la economía de 3%, un tipo de cambio promedio de 7.7 pesos por dólar, una inflación de 20% y un precio promedio del petróleo crudo mexicano de exportación de 13.25 dpb.

En cuanto a la política de precios y tarifas a instrumentarse en este año, se contemplan aumentos para gasolinas, diesel y gas licuado de 7% en diciembre de 1995, 6% en abril y un incremento de 2.1% en cada uno del resto de los meses; para el combustóleo y gas natural, su precio se fijará conforme a la referencia internacional.

Con los datos anteriores se obtuvieron los pronósticos anuales de la demanda de petrolíferos:

DEMANDA DE PETROLIFEROS PARA 1996			
Producto	Pronóstico mbd	Intervalos de confianza	
		Límite inferior	Límite superior
Gesolinas	484.5	464.7	505.1
Diesel	236.8	225.2	253.2
Gas licuado	260.2	237.7	284.9
Combustóleo	404.6	377.8	433.2
Gas Natural	221.1	207.7	235.4
(mpcd)	1,499.9	1,408.7	1,597.0

#### 4.5.2 Pronósticos mensuales

La construcción de los modelos ARIMA se inició a partir los modelos estimados para el periodo 1986-1994, incluyendo ahora los datos observados de 1995, teniendo de esta manera 120 observaciones desponibles para cada producto. En la reestimación, los modelos resultaron adecuados y sólo cambió el valor de los parámetros; únicamente en el caso del diesel fue necesario identificar otro modelo diferente. Los resultados de la estimación de los modelos se presentan en el siguiente cuadro.

MODELOS ESTIMADOS PARA LA DEMANDA DE PETROLIFEROS

Periodo: Ene 1986-dic 1995

No. de observaciones: 108

Variable	Transformación	Parámetros estimados	Intervalo del 95% de confianza	Correlación entre parámetros > 0.5 ó < -0.5	Análisis de residuos					
					m(d)	d <sub>0</sub>	$\frac{m(d)}{d_0}$	g.l.	r <sub>1</sub> (d) ≠ 0	Residuales grandes
Gasolina	$T(Z_t) = Z_t / 100$ $(1 - \phi_1 B - \phi_2 B^2)(1 - \delta B^{12}) \nabla \nabla_{12} T(Z_t) = (1 - \theta_1 B^6) a_t$	$\hat{\phi}_1 = -0.7707$ $\hat{\phi}_2 = -0.4399$ $\hat{\delta} = -0.2894$ $\hat{\theta}_1 = -0.3726$	(-0.9577, -0.5837) (-0.6222, -0.2576) (-0.4850, -0.0938) (-0.5843, -0.1609)	$P(\hat{\phi}_{11}, \hat{\phi}_2) = 0.5373$	-0.01098	0.14512	-0.72636	27.0, 21	$r_{11} = 0.232$ (0.116)	$d_{26} = -0.509$ $d_{111} = 0.234$ $d_{112} = 0.312$
Diesel	$(Z_t) = Z_t / 100$ $(1 - \phi_1 B - \phi_2 B^{12}) \nabla \nabla_{12} T(Z_t) = (1 - \theta_1 B^2 - \theta_{12} B^{12}) a_t$	$\hat{\phi}_1 = -0.4925$ $\hat{\phi}_2 = -0.3041$ $\hat{\theta}_1 = -0.4782$ $\hat{\theta}_{12} = 0.1780$	(-0.6549, -0.3281) (-0.1564, -0.4516) (-0.2889, -0.6675) (0.0090, 0.3550)		+0.01704	0.10623	1.69130	17.0, 21		$d_{22} = -0.259$ $d_{28} = -0.211$ $d_{29} = -0.249$ $d_{72} = 0.223$ $d_{88} = -0.260$ $d_{112} = 0.261$
Gas Licuado	$T(Z_t) = Z_t / 100$ $(1 - \phi_1 B) \nabla T(Z_t) = (1 - \theta_1 B) a_t$	$\hat{\phi}_1 = -0.3437$ $\hat{\theta}_1 = 0.7617$	(-0.5362, -0.1512) (0.4223, 0.9011)		0.00278	0.08124	-0.33400	17.0, 22		$d_{28} = -0.211$ $d_{29} = 0.207$ $d_{77} = -0.201$ $d_{72} = 0.223$ $d_{74} = -0.172$ $d_{112} = 0.167$
Gas Natural	$T(Z_t) = Z_t / 1000$ $(1 - \phi_1 B) \nabla T(Z_t) = (1 - \theta_1 B) a_t$	$\hat{\phi}_1 = -0.4735$ $\hat{\theta}_1 = 0.2528$	(-0.6457, -0.3013) (0.0585, 0.4471)		0.00122	0.08345	0.15893	11.0, 22		$d_{25} = 0.292$ $d_{22} = -0.187$ $d_{74} = -0.193$ $d_{77} = 0.265$ $d_{78} = -0.175$ $d_{88} = 0.180$
Combustible	$T(Z_{subt}) = Z_t / 100$ $(1 - \phi_1 B - \phi_2 B^2) \nabla T(Z_t) = (1 - \theta_1 B) a_t$	$\hat{\phi}_1 = -0.5241$ $\hat{\phi}_2 = -0.2118$ $\hat{\theta}_1 = -0.2141$	(-0.7031, -0.3451) (-0.3923, -0.0313) (-0.4168, -0.0141)		-0.805213	0.30801	-0.15309	20.0, 22		$d_{12} = -0.724$ $d_{22} = 0.665$ $d_{38} = -0.658$ $d_{104} = -0.814$ $d_{117} = 0.775$

Pronóstico de la demanda de petrolíferos

Los pronósticos que se obtuvieron con los modelos estimados fueron los siguientes:

PRONOSTICOS ARIMA DE LA DEMANDA DE PETROLIFEROS, 1996										
	Gasolinas		Diesel		Gas licuado		Combustóleo		Gas natural	
	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mpcd	$\bar{\sigma}$
Ene	442.7	14.67	226.9	10.69	284.7	8.12	343.4	30.80	1,536.2	83.45
Feb	464.7	16.14	236.0	11.99	275.5	8.35	352.7	34.11	1,535.8	84.32
Mar	461.9	16.19	232.7	14.21	265.1	8.57	341.8	34.22	1,535.9	101.50
Abr	423.4	17.22	226.0	15.64	247.6	8.79	346.0	34.22	1,535.7	109.29
May	436.0	17.66	237.7	16.68	239.8	9.00	341.7	34.48	1,535.8	109.36
Jun	438.0	17.66	232.7	16.02	240.3	9.20	344.1	34.51	1,535.8	110.70
Jul	418.1	17.86	229.0	16.09	241.0	9.40	331.9	34.51	1,535.8	111.23
Ago	435.5	17.86	232.9	16.12	244.7	9.60	321.2	34.51	1,535.8	111.26
Sep	427.0	17.96	227.5	16.17	261.0	9.79	316.3	34.52	1,535.8	111.45
Oct	429.6	18.76	235.0	16.17	260.7	9.98	333.3	34.52	1,535.8	111.47
Nov	448.6	18.87	234.1	16.23	275.0	10.17	334.5	34.52	1,535.8	111.49
Dic	462.7	18.88	231.0	16.24	282.0	10.35	330.3	35.16	1,535.8	111.51
Prom.	441.1		231.8		258.1		336.4		1,535.8	

#### 4.5.3 Pronósticos condicionales

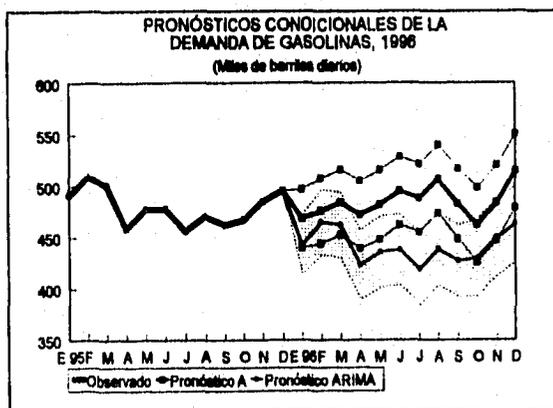
Para el cálculo de los pronósticos condicionales, se utilizaron los Procedimientos A y AU. Los pronósticos ARIMA originales y los condicionales con ambos procedimientos, así como los errores estándar de los pronósticos y el valor de los estadísticos  $K_1$  y  $K_2$ , se muestran a continuación:

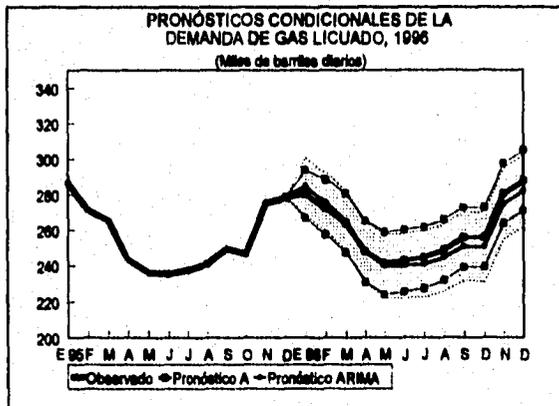
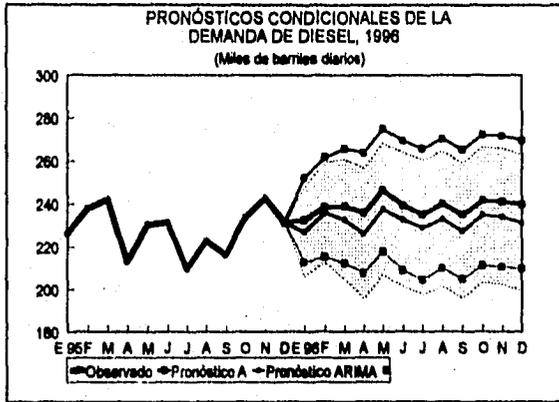
PRONOSTICOS CONDICIONALES DE LA DEMANDA DE PETROLIFEROS, 1996										
Procedimiento A										
	Gasolinas		Diesel		Gas licuado		Combustóleo		Gas natural	
	mbd	σ	mbd	σ	mbd	σ	mbd	σ	mpcd	σ
Ene	488.5	14.40	232.6	10.05	280.3	6.88	390.7	30.63	1,504.4	79.68
Feb	476.7	16.11	238.6	11.87	272.8	7.92	432.0	33.69	1,519.9	93.53
Mar	484.1	16.07	239.0	13.62	263.9	8.49	398.5	34.00	1,505.7	98.71
Abr	472.1	16.70	235.8	14.31	247.8	8.78	407.2	33.97	1,490.8	103.49
May	482.0	17.20	246.5	14.58	241.3	8.88	399.3	34.28	1,490.8	103.48
Jun	495.0	16.95	239.3	15.45	242.8	8.86	391.9	34.35	1,496.9	106.44
Jul	487.9	16.82	235.2	15.60	244.5	8.77	373.3	34.40	1,498.5	107.35
Ago	506.5	16.98	240.2	15.44	248.9	8.66	375.9	34.31	1,495.5	106.70
Sep	482.5	17.30	235.2	15.38	255.9	8.56	398.5	34.06	1,495.9	107.00
Oct	481.9	18.54	241.8	15.56	256.0	8.51	404.1	34.18	1,502.8	108.42
Nov	483.6	18.63	241.2	15.58	280.6	8.54	384.5	34.35	1,505.7	108.98
Dic	514.3	18.34	239.7	15.25	287.8	8.66	498.9	33.24	1,493.1	106.39
Prom.	484.5		238.8		260.2		404.6		1,500.0	
K1	132.81		2.43		1.05		217.18		1.64	

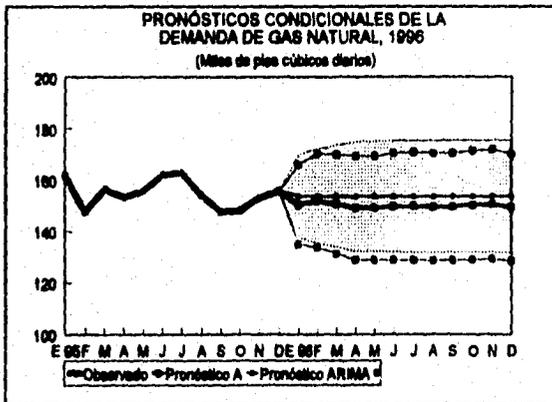
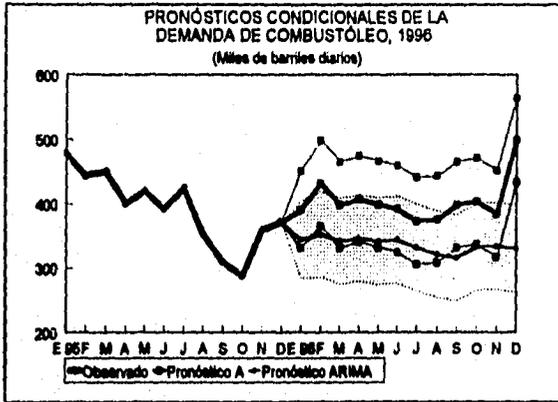
En los casos de las gasolinas y el combustóleo, el valor del estadístico K1 indicó que la conjetura dada por la demanda anual estimada no era consistente con la información histórica. Mientras que con el modelo ARIMA se estima una caída anual promedio con relación a lo observado en 1995 de 7.9 y 13.9%, respectivamente, la demanda estimada con los modelos de regresión crece 1.2 y 3.6%. Sin embargo, debido a que el procedimiento AU generó pronósticos prácticamente iguales a los ARIMA originales, se decidió tomar los pronósticos obtenidos con el procedimiento A en el caso del combustóleo. Para las gasolinas, se ensayó la utilización de un modelo ARIMA con efectos de intervención para considerar el comportamiento atípico que se observó en 1995. El procedimiento para la construcción de este modelo se presenta en la siguiente sección.

PRONOSTICOS CONDICIONALES DE LA DEMANDA DE PETROLIFEROS, 1996 Procedimiento AU										
	Gasolinas		Diesel		Gas licuado		Combustóleo		Gas natural	
	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mbd	$\bar{\sigma}$	mpcd	$\bar{\sigma}$
Ene	443.4	14.57	227.2	10.66	284.7	8.12	344.2	30.80	1,535.6	83.37
Feb	465.0	16.14	236.2	11.96	275.5	8.35	354.1	34.11	1,535.3	94.31
Mar	462.5	16.18	233.0	14.18	265.1	8.57	342.8	34.21	1,535.3	101.44
Abr	424.8	17.21	226.5	15.58	247.6	8.79	347.0	34.22	1,534.8	109.18
May	437.3	17.64	238.1	15.60	239.8	9.00	342.7	34.48	1,534.9	109.23
Jun	439.7	17.64	233.0	15.95	240.3	9.20	344.9	34.51	1,535.0	110.92
Jul	421.1	17.82	229.3	16.07	241.0	9.40	332.6	34.51	1,535.0	111.16
Ago	440.4	17.93	233.3	16.09	244.7	9.69	322.1	34.51	1,535.0	111.17
Sep	428.6	17.94	227.8	16.13	251.0	9.79	317.7	34.51	1,535.0	111.36
Oct	430.5	18.74	235.3	16.14	250.7	9.87	334.5	34.51	1,535.1	111.41
Nov	450.6	18.86	234.4	16.20	275.0	10.16	335.4	34.52	1,535.2	111.44
Dic	464.2	18.86	231.4	16.18	282.1	10.34	333.1	35.12	1,534.9	111.40
Prom.	442.3		232.1		258.1		337.6		1,535.1	
KU	3.81		0.11		0.01		3.65		0.03	

Las gráficas que se muestran a continuación muestran los pronósticos derivados de los modelos ARIMA originales con sus intervalos de confianza (área sombreada), y los pronósticos ARIMA condicionales generados con el Procedimiento A.







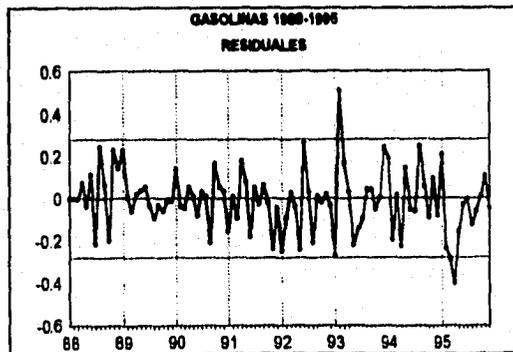
#### 4.5.4 Modelo ARIMA con intervenciones para el consumo de las gasolinas

Para incorporar el efecto de las intervenciones al modelo ARIMA de la serie mensual del consumo de las gasolinas se partió del modelo estimado para el periodo enero de 1986 a diciembre de 1995:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \phi B^{12}) \nabla \nabla_{12} T(Z_t) = (1 - \theta_0 B^9) a_t$$

En la verificación de los supuestos de este modelo, los residuales correspondientes a las observaciones 86, 111 y 112 resultaron mayores a dos veces la desviación estándar de la serie de residuales. En el caso del residual 86, asociado a la observación de enero de 1993, se trata de una sola observación que sale del comportamiento regular de la serie; sin embargo, no se encontró ninguna explicación que pudiese indicar que este comportamiento haya sido resultado de un efecto de intervención.

Los otros dos residuales se asocian a los efectos de la crisis económica que se desencadenó con la devaluación del peso ocurrida en diciembre de 1994; por lo tanto, el análisis de intervención se avoca a medir el impacto que tuvo la devaluación sobre el consumo de gasolinas.



La inspección visual de la gráfica del consumo de gasolinas indica claramente un cambio de nivel en la serie a partir de 1995. Aún así, para verificar que la intervención tuvo un efecto significativo sobre el comportamiento de la serie, se calculó el estadístico  $C$  (expresión 2.7.1.2), a partir de los pronósticos obtenidos con el modelo estimado de enero de 1986 a diciembre de 1994 (sección 4.2.1):

ERRORES DE PRONOSTICO Y ESTADISTICO $C$			
$h$	$\hat{T}(Z_{t-1})(h)$	$\hat{T}(Z_{t-2,h})(1)$	$e_{t-2,h}(1)$
109	4.852	4.852	0.000
110	5.341	5.373	0.032
111	5.247	5.297	0.050
112	5.031	5.067	0.036
113	5.234	4.960	-0.274
114	5.155	4.822	-0.334
115	4.988	4.551	-0.444
116	5.237	4.803	-0.434
117	5.007	4.802	-0.418
118	5.053	4.688	-0.367
119	5.323	4.688	-0.628
120	5.586	4.888	-0.580
$C = 28.998$		$\chi^2_{(12)0.05} =$	21.030

El valor de  $C$  es mayor que la  $\chi^2_{(12)}$  de tablas, por lo que no se rechaza la hipótesis de que la intervención afectó el comportamiento de la serie.

La serie que capta el efecto de la intervención es del tipo escalón:

$$S_{t,1} = \begin{cases} 0, & (t \leq 108) \\ 1, & (t > 108) \end{cases}$$

y por el comportamiento de los residuales a partir de enero de 1995, se propuso que el modelo a estimar para la intervención fuese:

$$T(Z_t) = \left\{ \frac{\omega_0 - \omega_1 B - \omega_2 B^2}{(1 - \delta B) \nabla} \right\} S_{t,1} + N_t$$

en donde el modelo ARIMA que se propone para  $N_t$  es el mismo que se construyó para obtener los pronósticos de 1995:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \phi B^{12}) \nabla \nabla_{12} N_t = (1 - \theta_9) a_t$$

Cabe señalar que, como se muestra en el cuadro de resumen de la estimación, no fue necesario modificar este modelo en el proceso para identificar la función de intervención.

De esta manera, el primer modelo estimado no era adecuado porque el parámetro  $\hat{\omega}_0$  fue no significativo, y el intervalo de confianza de  $\hat{\delta}_1$  y  $\hat{\omega}_1$  contenía al 1, teniéndose correlaciones muy elevadas entre todos los parámetros del numerador de la función de intervención; además, no disminuyó el valor de los residuales que habían resultado significativamente distintos de cero.

El siguiente modelo que se ensayó fue eliminando los parámetros mencionados y ahora incluyendo uno en el numerador de orden 3,  $\omega_3$ , considerando que el cambio en el nivel de la serie se reflejó 2 ó 3 periodos más adelante; de esta manera, el modelo estimado fue:

$$T(Z_t) = \left\{ \frac{\omega_2 B^2 - \omega_3 B^3}{\nabla} \right\} S_{t,1} + N_t$$

En este modelo, se tuvo correlación muy elevada entre los parámetros  $\omega_2$  y  $\omega_3$ , y la autocorrelación de los residuales  $r_{24}$  fue distinta de cero.

La gráfica de la serie indicaba que el efecto de la intervención se prolongó durante varios meses, por ello se ensayó incluir términos de orden más elevado; así, de una gran cantidad de modelos estimados los que resultaban más adecuados fueron:

$$T(Z_t) = \left\{ \frac{\omega_2 B^2 - \omega_3 B^3 - \omega_4 B^4}{(1 - \delta_1 B^1) \nabla} \right\} S_{t,t} + N_t$$

$$T(Z_t) = \left\{ \frac{\omega_2 B^2 - \omega_5 B^5}{(1 - \delta_2 B^2) \nabla} \right\} S_{t,t} + N_t$$

y

$$T(Z_t) = \left\{ \frac{\omega_0}{(1 - \delta_3 B^3) \nabla} \right\} S_{t,t} + N_t$$

pero aún presentaban violaciones a los supuestos muy significativas (véase cuadro resumen). Inclusive, para el último modelo, los resultados indicaron que las raíces características del polinomio  $1 - \delta_3 B^3 = 0$  quedaban fuera del círculo unitario.

Finalmente, el modelo que resultó más adecuado fue el siguiente:

$$T(Z_t) = \left\{ \frac{-0.0693 B^2 + 0.0165 B^4}{(1 + 0.5706 B^3) \nabla} \right\} S_{t,t} + \left\{ \frac{1 + 0.4187 B^2}{(1 + 0.8718 B + 0.5204 B^2)(1 + 0.3278 B^{11}) \nabla \nabla_{11}} \right\} a_t$$

El resumen de los resultados de la estimación se presenta en los siguientes cuadros.

**RESUMEN DE LA ESTIMACION DE LOS MODELOS ARIMA CON INTERVENCION PARA LA DEMANDA DE GASOLINAS**

Periodo: Ene 1986-dic 1995

No. de observaciones: 108

Transformación:  $T(Z_t) = Z_t / 100$

Intervención:  $S_{i,t} = \begin{cases} 0, & t < 108 \\ 1, & t \geq 108 \end{cases}$

Modelo	Parámetros estimados	Intervalos del 95% de confianza	Correlación entre parámetros > 0.5 ó < -0.5	Análisis de residuos					
				m(Δ)	σ <sub>a</sub>	$\frac{\sigma(\hat{\epsilon}_t)}{\sigma_a}$	Q, g.l.	r <sub>ε</sub> (Δ)=0	Residuales grandes
$(1-\phi_1 B - \phi_2 B^2)(1-\Phi B^{12})VV, T(Z_t) = (1-\theta_1 B^6)n$ Periodo de estimación: 1986-1996	$\hat{\phi}_1 = -0.8768$ $\hat{\phi}_2 = -0.5637$ $\hat{\Phi} = -0.4080$ $\hat{\theta}_1 = -0.4474$	(-0.9960, -0.7400) (-0.7451, -0.3814) (-0.6079, -0.2081) (-0.6673, -0.2277)	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5590$	0.00022	0.13697	0.01424	21.0, 20		$\hat{d}_{75} = 0.296$ $\hat{d}_{95} = 0.293$ $\hat{d}_{98} = -0.485$ $\hat{d}_{99} = 0.287$
$T(Z_t) = \left\{ \frac{\omega_1 + \phi_1 B + \omega_2 B^2}{(1-\phi_1 B - \phi_2 B^2)} \right\} S_{i,t} + N_t$ $(1-\phi_1 B - \phi_2 B^2)(1-\Phi B^{12})VV, N_t = (1-\theta_1 B^6)n$	$\hat{\phi}_1 = -0.8536$ $\hat{\phi}_2 = -0.5438$ $\hat{\Phi} = -0.31641$ $\hat{\theta}_1 = -0.4432$ $\hat{\theta}_2 = -0.7064$ $\hat{\omega}_2 = 0.1787$ $\hat{\omega}_1 = 0.5939$ $\hat{\omega}_2 = -0.4449$	(-1.0337, -0.6735) (-0.7208, -0.3667) (-0.5519, -0.1763) (-0.6447, -0.2416) (-1.0847, -0.3281) (-0.0697, 0.4270) (0.0788, 1.090) (-0.7336, -0.1562)	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5545$ $r(\hat{\omega}_1, \hat{\omega}_2) = 0.9764$ $r(\hat{\omega}_1, \hat{\omega}_2) = -0.8992$ $r(\hat{\omega}_1, \hat{\omega}_2) = -0.9722$	-0.00014	0.13292	-0.00999	29.0, 16		$\hat{d}_{75} = 0.286$ $\hat{d}_{95} = 0.284$ $\hat{d}_{98} = -0.495$ $\hat{d}_{99} = 0.269$
$T(Z_t) = \left\{ \frac{\omega_1 + \phi_1 B + \omega_2 B^2}{(1-\phi_1 B - \phi_2 B^2)} \right\} S_{i,t} + N_t$ $(1-\phi_1 B - \phi_2 B^2)(1-\Phi B^{12})VV, N_t = (1-\theta_1 B^6)n$	$\hat{\phi}_1 = -0.8551$ $\hat{\phi}_2 = -0.5099$ $\hat{\Phi} = -0.3112$ $\hat{\theta}_1 = -0.4166$ $\hat{\omega}_2 = -0.0532$ $\hat{\omega}_1 = -0.0144$	(-1.0376, -0.6725) (-0.6874, -0.3324) (-0.4966, -0.1258) (-0.6202, -0.2130) (-0.0821, -0.0243) (-0.0261, -0.0027)	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5677$ $r(\hat{\omega}_1, \hat{\omega}_2) = 0.7368$	0.00154	0.13690	0.10836	35.0, 18	$r_{24} = -0.252$ (0.126)	$\hat{d}_{95} = 0.273$ $\hat{d}_{98} = -0.494$ $\hat{d}_{99} = -0.271$

Ponderado de la demanda de petrolíferos

**RESUMEN DE LA ESTIMACION DE LOS MODELOS ARIMA CON INTERVENCION PARA LA DEMANDA DE GASOLINAS**

Periodo: Ene 1986-dic 1995

No. de observaciones: 108

Transformación:  $T(Z_t) = Z_t/100$

Intervención:  $S_{t,t} = \begin{cases} 0, & t < 108 \\ 1, & t \geq 108 \end{cases}$

Modelo	Parámetros estimados	Intervalos del 95% de confianza	Correlación entre parámetros > 0.5 ó < -0.5	Análisis de residuos					
				$m(d)$	$\hat{a}_s$	$\frac{\sum \hat{e}_t^2}{n}$	$Q^*$ , g.l.	$r_s(d) \neq 0$	Residuales grandes
$T(Z_t) = \left( \frac{1 - \theta_1 B - \theta_2 B^2}{1 - \phi_1 B - \phi_2 B^2} \right) S_{t,t} + N_t$ $(1 - \theta_1 B - \theta_2 B^2) \nabla \nabla_{12} N_t = (1 - \phi_1 B - \phi_2 B^2) a_t$	$\hat{\phi}_1 = -0.8780$ $\hat{\phi}_2 = -0.5210$ $\hat{\theta}_1 = -0.3469$ $\hat{\theta}_2 = -0.4138$ $\hat{\delta}_1 = 0.5902$ $\hat{\omega}_1 = -0.0566$ $\hat{\omega}_2 = -0.1382$ $\hat{\omega}_3 = 0.1455$	$(-1.0603, -0.6957)$ $(-0.6995, -0.3434)$ $(-0.5337, -0.1606)$ $(-0.6166, -0.2111)$ $(-0.2996, 1.4799)$ $(-0.1092, -0.0040)$ $(-0.2415, -0.0348)$ $(0.0281, 0.2629)$	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5775$ $r(\hat{\delta}_1, \hat{\omega}_2) = 0.8216$	0.00252	0.13401	0.18154	34.0, 16		$\hat{d}_{77} = 0.273$ $\hat{d}_{85} = 0.280$ $\hat{d}_{88} = -0.483$ $\hat{d}_{95} = 0.267$ $\hat{d}_{108} = -0.307$
$T(Z_t) = \left( \frac{1 - \theta_1 B - \theta_2 B^2}{1 - \phi_1 B - \phi_2 B^2} \right) S_{t,t} + N_t$ $(1 - \theta_1 B - \theta_2 B^2) \nabla \nabla_{12} N_t = (1 - \phi_1 B - \phi_2 B^2) a_t$	$\hat{\phi}_1 = -0.8546$ $\hat{\phi}_2 = -0.5113$ $\hat{\theta}_1 = -0.3099$ $\hat{\theta}_2 = -0.4184$ $\hat{\delta}_1 = -0.1795$ $\hat{\omega}_1 = -0.0531$ $\hat{\omega}_2 = -0.0174$	$(-1.0403, -0.6689)$ $(-0.6921, -0.3304)$ $(-0.4998, -0.1201)$ $(-0.6244, -0.2123)$ $(-2.1344, 1.7754)$ $(-0.1206, 0.0145)$ $(-0.0381, 0.0033)$	$r(\hat{\phi}_1, \hat{\phi}_2) = 0.5751$ $r(\hat{\delta}_1, \hat{\omega}_2) = 0.8583$ $r(\hat{\omega}_1, \hat{\omega}_2) = 0.6363$	0.00257	0.13755	0.17998	35.0, 17	$r_{26} = 0.251$ $(0.126)$	$\hat{d}_{85} = 0.275$ $\hat{d}_{88} = -0.489$ $\hat{d}_{108} = -0.289$

**RESUMEN DE LA ESTIMACION DE LOS MODELOS ARIMA CON INTERVENCION PARA LA DEMANDA DE GASOLINAS**

Periodo: Ene 1986-dic 1995

No. de observaciones: 108

Intervención:  $S_{i,t} = \begin{cases} 0, & t < 108 \\ 1, & t \geq 108 \end{cases}$

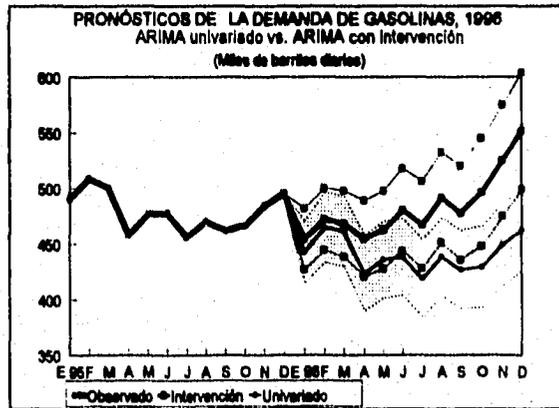
Transformación:  $T(Z_t) = Z_t/100$

Modelo	Parámetros estimados	Intervalos del 95% de confianza	Correlación entre parámetros > 0.5 ó < -0.5	Análisis de residuos					
				m(d)	$\hat{\sigma}_e$	$\frac{m(d)}{\hat{\sigma}_e}$	0', g.l.	$r_e(d)=0$	Residuales grandes
$T(Z_t) = \left\{ \frac{1 - \theta_1 B^2 - \theta_2 B^4}{(1 - \phi_1 B - \phi_2 B^2)} \right\} S_{i,t} + N_t$ $(1 - \theta_1 B - \theta_2 B^2)(1 - \phi_1 B^2) VV_{12} N_t = (1 - \theta_1 B^2) \alpha$	$\hat{\theta}_1 = -0.8718$ $\hat{\theta}_2 = -0.5204$ $\hat{\phi}_1 = -0.3278$ $\hat{\phi}_2 = -0.4187$ $\hat{\delta}_1 = -0.5706$ $\hat{\omega}_1 = -0.0693$ $\hat{\omega}_2 = -0.0165$	$(-1.0547, -0.6889)$ $(-0.6981, -0.3427)$ $(-0.5124, -0.1432)$ $(-0.6204, -0.2171)$ $(-0.1382, -0.0029)$ $(-0.1047, -0.0339)$ $(-0.0323, -0.0007)$	$r(\hat{\theta}_1, \hat{\theta}_2) = 0.5676$ $r(\hat{\omega}_1, \hat{\omega}_2) = 0.5150$	0.00244	0.13522	0.17429	34.0, 17		$d_{77} = 0.266$ $d_{81} = 0.276$ $d_{85} = -0.488$ $d_{108} = -0.291$
$T(Z_t) = \left\{ \frac{1 - \theta_1 B - \theta_2 B^2}{(1 - \phi_1 B - \phi_2 B^2)} \right\} S_{i,t} + N_t$ $(1 - \theta_1 B - \theta_2 B^2)(1 - \phi_1 B^2) VV_{12} N_t = (1 - \theta_1 B^2) \alpha$ m	$\hat{\theta}_1 = -0.7905$ $\hat{\theta}_2 = -0.4569$ $\hat{\phi}_1 = -0.2894$ $\hat{\phi}_2 = -0.3808$ $\hat{\delta}_1 = -1.6540$ $\hat{\omega}_1 = -0.0102$	$(-0.9790, -0.6020)$ $(-0.8401, -0.2738)$ $(-0.6844, -0.0948)$ $(-0.5978, -0.1638)$ $(-2.9105, -0.3975)$ $(-0.0224, 0.0019)$	$r(\hat{\theta}_1, \hat{\theta}_2) = 0.5468$	-0.00323	0.14385	-0.21666	30.0, 18	$r_{23} = 0.244$ $(0.118)$	$d_{98} = -0.506$ $d_{112} = 0.353$

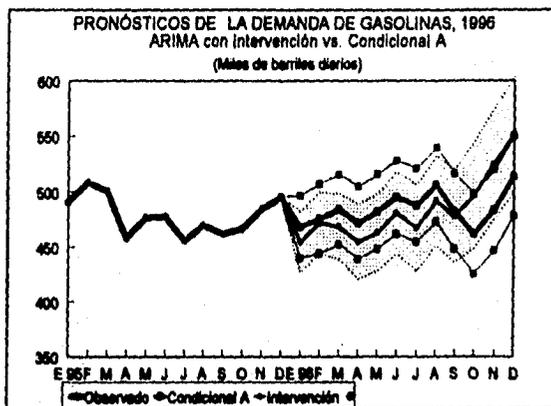
#### 4.5.5 Resultados

De esta manera, para 1996 los pronósticos mensuales del consumo de petrolíferos se conformaron de la siguiente manera:

Para las gasolinas se tomaron los pronósticos obtenidos mediante el modelo ARIMA con intervenciones, y si bien el promedio anual de éstos es similar a los obtenidos a través del modelo de regresión, se decidió utilizar los primeros y no los condicionales generados con el Procedimiento A por el comportamiento mensual. Se supuso que en este año la recuperación económica será paulatina, y por tanto el comportamiento de la demanda de gasolinas seguiría el patrón que marcan los pronósticos del modelo ARIMA con intervenciones.



Para el resto de los productos, el estadístico  $K_1$ , con excepción de combustóleo, indicó que los pronósticos calculados a través



del Procedimiento A eran los adecuados. Así, los pronósticos mensuales para la demanda de petrolíferos en 1996 se presenta a continuación:

PRONÓSTICOS PARA LA DEMANDA MENSUAL DE PETROLÍFEROS, 1996 (Miles de barriles diarios)						
	Gasolinas	Diesel	Gas licuado	Combustibles	Gas natural	Total
Ene	454.6	232.6	290.3	390.7	221.6	1,590.0
Feb	472.3	236.6	272.6	432.0	224.1	1,639.6
Mar	466.4	239.0	263.9	396.5	222.0	1,587.8
Abr	456.0	235.8	247.8	407.2	219.8	1,566.6
May	463.0	246.5	241.3	389.3	219.8	1,569.9
Jun	480.7	239.3	242.8	381.8	220.7	1,575.4
Jul	487.3	235.2	244.5	373.3	220.8	1,541.1
Ago	491.9	240.2	248.9	375.9	220.6	1,577.4
Sep	478.0	235.2	255.9	396.5	220.6	1,586.2
Oct	486.8	241.8	256.0	404.1	221.6	1,620.1
Nov	525.3	241.2	280.6	384.5	222.0	1,653.6
Dic	551.6	239.7	287.8	488.9	220.1	1,788.2
Prom.	483.7	236.8	260.2	404.6	221.1	1,606.4
Crec.	1.0	4.7	1.8	3.5	-3.1	1.7

Los resultados anteriores muestran que el consumo de las gasolinas podría presentar un crecimiento en 1996 de 1.0%, y el resto de los productos, si se cumplen los supuestos del programa macroeconómico podrían alcanzar crecimientos de 4.7% el diesel, 1.8% el gas licuado, 3.5% el combustóleo, mientras que para el gas natural se esperaría una caída de 3.1%.

## Conclusiones

En este trabajo se ha presentado un procedimiento integral que combina información histórica de una serie con algunas conjeturas acerca de su probable evolución futura para obtener pronósticos que resultan más eficientes que los pronósticos ARIMA convencionales.

La aplicación al pronóstico de la demanda de petrolíferos en México resultó muy ilustrativo para mostrar el uso de dos herramientas estadísticas en la generación de pronósticos, éstas son el análisis de regresión lineal múltiple y el análisis de series de tiempo. Además, con el método de pronósticos ARIMA condicionales que propuso Víctor M. Guerrero en 1989 se expuso la manera en que se puede explotar un conjunto de información más amplio que el que se utilizaría mediante el empleo de un sólo método.

En particular, el comportamiento en 1995 de las series utilizadas permitió realizar una evaluación, que en este caso fue sólo intuitiva, de la eficacia del método propuesto y mostrar el empleo de una herramienta adicional en el análisis de series de tiempo, es decir, el análisis de intervención.

Sin embargo, aun cuando Pérez-Porrúa (1984) sugiere la extensión de los procedimientos de pronósticos condicionales a modelos ARIMA donde se utilizan variables explicatorias, por ejemplo los modelos de funciones de transferencia, el método no se aplicó a los pronósticos generados con modelos ARIMA con intervención debido a que no se revisaron los fundamentos teóricos que justificaran tal aplicación; así, esto constituye un tema de futuras investigaciones.

Por lo que hace al ejercicio de aplicación, la herramienta estadística que se expuso proporciona otras alternativas para generar diferentes pronósticos mensuales de la demanda de petrolíferos; por ejemplo, se puede utilizar el análisis de regresión con variables explicativas que tengan una periodicidad mensual, o bien, construir los modelos de función de transferencia utilizando la historia de la propia serie de consumo y de otras series económicas exógenas que estén disponibles con la misma periodicidad.

De esta manera, el procedimiento expuesto puede ser muy útil no sólo para obtener pronósticos más precisos, sino para medir el impacto de la aplicación de ciertas políticas sobre la variable de interés, y como auxiliar en el seguimiento, evaluación y control de programas o presupuestos. Una ventaja adicional del procedimiento aquí expuesto, y que señala Pérez-Porrúa, radica en la posibilidad de sistematizar las labores de "adecuación" de los pronósticos, que comúnmente se realizan en forma muy empírica.

## Bibliografía

- [1] ANDERSON, O. (1979) "Time Series Analysis and Forecasting, The Box-Jenkins Approach". 1a. Edición. Londres Butterworht & Co. Ltd.
- [2] ANUARIO Estadístico 1989-1995, Petróleos Mexicanos.
- [3] BARRON, M.O. (1986) "Aplicación del Análisis de Intervención de Box y Tiao a la Serie de Tiempo del Índice nacional de Precios al Consumidor en México". Tesis de licenciatura, UNAM.
- [4] BORDLEY, R.F. (1982) "The Combination of Forecasts: A Bayesian Approach" Journal Operation Research Society 33, pp 171-174.
- [5] BOX, G.E.P., Jenkins G.M. (1976) "Time Series Analysis: Forecasting and Control". Edición revisada. San Francisco, Holden-Day.
- [6] BUNN, D.W. (1985) "Statistical Efficiency in the Linear Combination of Forecasts" International Journal of Forecasting 1, pp 151-163.
- [7] CLEMEN, R.T. y Winkler R.L. (1986) "Combining Economic Forecasts" Journal of Business and Economic Statistics 4, pp 39-46.
- [8] CLEMEN, R.T. (1986) "Linear Constraints and the Efficiency of Combined Forecasts" Journal of Forecasting 5, pp 31-38.
- [9] CRITERIOS Generales de Política Económica para 1996. Secretaría de Hacienda y Crédito Público.
- [10] DIEBOLD, F.X. (1988) "Serial Correlation and the Combination of Forecasts" Journal of Business and Economic Statistics 6, pp 105-111.
- [11] DORNBUSCH, R. y Fisher, S. (1989) "Macroeconomía" 4a. Edición. Mc. Graw-Hill.

- [12] GRANGER C.W.J. y Newbold, P. (1974) "Experience with Forecasting Univariate Time Series and the Combination of Forecasts". Journal of the Royal Statistical Society, Serie A 139, pp 246-257.
- [13] GRANGER C.W.J. y Newbold, P. (1977) "Forecasting Economic Time Series". Nueva York, Academic Press, Inc.
- [14] GRANGER, C.W.J. y Ramanathan, R. (1984) "Improved Methods of Combining Forecasts". Journal of Forecasting 3, pp 197-204.
- [15] GUERRERO, V.M. (1991) "Análisis Estadístico de Series de Tiempo Económicas". México, Universidad Autónoma Metropolitana.
- [16] GUERRERO, V.M. (1989) "Desagregación de Series de Tiempo". Documento Interno, Banco de México.
- [17] GUERRERO, V.M. (1989) "Optimal Conditional ARIMA Forecasts". Journal of Forecasting 8, pp 215-229.
- [18] GUJARATI, D. (1992) "Econometría". 2a. Edición. Mc Graw-Hill, México.
- [19] INDICADORES Económicos del Banco de México, Agosto 1992-Diciembre 1995.
- [20] INDICADORES Petroleros, Informe Mensual sobre Producción y Comercio de Hidrocarburos, Julio 1992-Diciembre 1995. Petróleos Mexicanos.
- [21] LEY del Impuesto Especial sobre Producción y Servicios (1995). Secretaría de Hacienda y Crédito Público.
- [22] LEY del Impuesto al Valor Agregado (1995). Secretaría de Hacienda y Crédito Público.
- [23] LEY Federal de Derechos (1995). Secretaría de Hacienda y Crédito Público.
- [24] MADDALA, G.S. (1976) "Econometrics". Nueva York. Mc. Graw-Hill.
- [25] MAKRIDAKIS, S. y Wheelwright, S. (1978) "Forecasting. Methods and Applications". Nueva York, John Wiley & Sons Inc.
- [26] MEMORIA de Labores, Petróleos Mexicanos (1990).

- [27] PINDYCK, R. S. Y Rubinfeld D.L. (1981) "Econometric Models and Economic Forecasts, 2a. Edición, Nueva York, Mc Graw-Hill, Inc.
- [28] PÉREZ-PORRUA, J.M. (1984), "Pronósticos Condicionales". Documento interno. Banco de México.
- [29] SALAS, F. (1991) "Incentivos en una Empresa de Gobierno: El Caso de Pemex". Trabajo de Investigación, Stanford University.
- [30] SALAS, J. (1990) "Econometría Aplicada a los Países en Desarrollo. El Caso Mexicano". México. Fondo de Cultura Económica.
- [31] TRONCOSO, F.J. (1973) "Aplicación de Modelos para Pronosticar el Abastecimiento de Gasolina en la República Mexicana". Tesis de licenciatura, Universidad Anáhuac.
- [32] WINKLER, R.L. y Makridakis, S. (1983) "The Combination of Forecasts". Journal of the Real Statistical Society. Serie A 146, pp 150-147.
- [33] WINKLER, R.L. y Makridakis, S. (1983) "Average of Forecasts: Some Empirical Results". Management Science 29, pp 987-996.