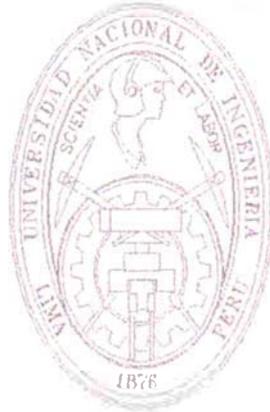


UNIVERSIDAD NACIONAL DE INGENIERÍA

**FACULTAD DE INGENIERÍA
ECONÓMICA Y CIENCIAS SOCIALES**



PROYECCIÓN DE LAS IMPORTACIONES PERUANAS

Una aplicación en la Superintendencia Nacional de Administración
Tributaria SUNAT

PARA OPTAR EL TITULO PROFESIONAL DE:

LICENCIADO EN ESTADÍSTICA

POR LA MODALIDAD DE TESIS

ELABORADO POR:

Edgard Julio Ortega Huayta

**LIMA-PERU
Enero - 2004**

A mis padres Julio y Faustina.

SUMARIO

Las importaciones junto con las exportaciones, son medidas importantes de nuestra economía en el área de comercio exterior. Las primeras están sujetas al pago de impuestos mientras que las últimas no lo están; por esta razón la evolución y proyección de las importaciones es necesaria para la SUNAT como ente recaudador de impuestos; de manera que se puedan tomar las políticas y acciones necesarias para el cumplimiento y alcance de las metas respectivas.

Para entender la evolución de las importaciones y realizar proyecciones a futuro de éstas; se utilizaron metodologías cuantitativas con algunas consideraciones de tipo cualitativo. Se evaluaron inicialmente dos modelos; uno de tipo causal con la técnica del análisis de regresión, y el otro basado en series de tiempo con el uso de modelos ARIMA. Finalmente las proyecciones fueron el resultado de la combinación de las cifras obtenidas con estos modelos.

La combinación de las proyecciones tuvo como resultado una disminución en los niveles de error de pronóstico. Combinar modelos con distintas debilidades y fortalezas, permitió conseguir resultados que integren lo mejor de ambos enfoques.

INDICE

Introducción.....	1
Capitulo I: Marco Teórico.....	3
1.1 Series de tiempo.....	5
1.1.1 Definición.....	5
1.1.2 Estacionariedad.....	7
1.1.3 Procesos lineales.....	9
1.1.4 Condiciones de estacionariedad e invertibilidad.....	11
1.1.5 Identificación y correlograma.....	13
1.1.6 Procesos no estacionarios.....	15
1.1.7 Validación.....	19
1.1.8 Predicción.....	25
1.1.9 Modelos ARIMA con variables exógenas.....	26
1.2 Análisis de Regresión.....	30
1.2.1 Definición.....	30
1.2.2 Suposiciones.....	32
1.2.3 Elección de variables.....	33
1.2.4 Análisis gráfico.....	34

1.2.5	Estimación.....	34
1.2.6	Validación.....	35
1.2.7	Interpretación del modelo.....	41
1.2.8	Predicción.....	42
1.2.9	Regresiones con errores ARMA.....	43
1.3	Evaluación de la capacidad predictiva del modelo.....	45
1.4	Sistemas de pronóstico.....	46
Capitulo II: Objetivos y metodología.....		49
2.1	Objetivo.....	49
2.2	Metodología.....	49
2.3	Metodología de series de tiempo.....	50
2.4	Metodología de análisis de regresión.....	56

Capitulo III: Caracterización de los modelos y proyección.....	61
3.1 Variables.....	61
3.2 Aplicación de la metodología de series de tiempo.....	68
3.3 Aplicación de la metodología de análisis de regresión.....	105
3.4 Comparación de Modelos.....	120
3.5 Pronósticos para el año 2003.....	126
Conclusión.....	129
Bibliografía.....	131
Datos para el ajuste de los modelos.....	133

INTRODUCCIÓN

La Economía Global es el resultado de una serie de cambios económicos y tecnológicos que integran cada vez más a los países. Nuestro país se relaciona con el mundo a través del comercio internacional; por éste se entiende al intercambio de bienes y servicios entre países; el mismo que está compuesto de las exportaciones y las importaciones.

Las *exportaciones* se refieren a la salida de mercancías del territorio nacional para su consumo definitivo en el exterior, mientras que por *importaciones* se entiende al ingreso de mercancías provenientes del exterior para ser destinadas al consumo.

La evolución de las importaciones y exportaciones producto del comercio exterior es de vital importancia para el país pues es uno de los indicadores más importantes del desenvolvimiento de nuestra economía en el sector externo. Este comportamiento se resume en la Balanza Comercial, la cual registra la diferencia entre el valor de las exportaciones y las importaciones de un país. Si el valor de las exportaciones supera al de las importaciones; entonces, se dice que hay superávit. Por el contrario si se tiene la situación inversa, se registra un déficit.

Por otra parte la evolución de las importaciones es de gran importancia para la SUNAT; ya que es ésta la que está afectada a tributos.

De lo expresado en el párrafo precedente es de entenderse que el conocimiento adelantado del comportamiento de las importaciones y exportaciones es de gran importancia. Por ende la proyección de estos valores a futuro se hace necesaria.

El tema de la presente tesis se centra en la proyección de las importaciones; ya que para la SUNAT como ente recaudador de tributos este es de vital importancia. Las exportaciones por otra parte no están afectadas a tributo alguno.

En el desarrollo teórico del presente trabajo no se desarrollan las demostraciones de las afirmaciones estadísticas que se utilizan; pues no es el propósito de este trabajo; este se centra más bien en la aplicación de distintos enfoques para la proyección de las importaciones y de las ventajas y desventajas que estos ofrecen.

Las perspectivas tomadas fueron: modelos causales y modelos de series de tiempo; mientras que las técnicas empleadas: análisis de regresión y metodología ARIMA. Finalmente se optó por un resultado combinado que permitió disminuir el margen de error en los pronósticos.

CAPITULO I: MARCO TEORICO

Para la generación de pronósticos pueden emplearse dos enfoques, el cuantitativo y el cualitativo. En el primero de estos grupos, se tiene a los modelos causales y a las series de tiempo; los modelos causales usan un conjunto de variables denominadas explicativas para obtener el pronóstico de la variable en estudio; por otra parte, las series de tiempo se basan en el análisis del comportamiento pasado de la variable y así determinar su comportamiento futuro. En el enfoque cualitativo, el pronóstico es efectuado por uno o varios expertos que conocen bien la variable en estudio; el carácter de este valor es por lo tanto subjetivo.

Los modelos causales implican tener suficiente información estadística y una previsión precisa de los factores causales. Dado que estas circunstancias no se dan siempre, la metodología usual es proyectar estas variables con series de tiempo, a menos que se tengan escenarios establecidos, u otro tipo de información.

Por otro lado, una serie de tiempo es una secuencia de observaciones de una variable ordenada en el tiempo, el análisis de este tipo de datos usa solamente la

serie histórica de observaciones de la variable para desarrollar un modelo para predecir valores futuros. En este enfoque, no necesitamos conocer ninguna relación de causalidad explicativa para la variable en estudio. La técnica más extendida y rigurosa para la predicción en este contexto son los denominados modelos de Box - Jenkins o modelos ARIMA¹. Esta técnica identifica el modelo generador de las observaciones, para después en un proceso iterativo, estimar y verificar el modelo que una vez aceptado se utiliza para predecir valores futuros de la serie temporal.

En la actualidad los modelos de pronóstico usan una combinación de métodos cualitativos y cuantitativos. Los métodos estadísticos se usan para preparar un pronóstico, luego este se convierte en la información de entrada para una evaluación subjetiva de parte de los expertos que pueden modificar el pronóstico a la vista de otra información relevante y su percepción del futuro.

En este trabajo se evalúa el modelo causal y el modelo de series de tiempo para la generación de pronósticos de importación. El modelo causal, permite el análisis de distintos escenarios en base a las variables explicativas, la técnica aplicada fue el análisis de regresión. En las series de tiempo se usó la metodología ARIMA. El uso del modelo causal o de series de tiempo no excluye la posibilidad de que los errores en un modelo causal puedan modelarse con la metodología ARIMA (si no se comportan como ruido blanco) o que al modelo de

¹ De las palabras inglesas: *autoregressive integrated moving average*, es decir modelos autoregresivos integrados de medias móviles

series de tiempo ARIMA puedan agregársele variables explicativas (llamados a veces modelos ARIMAX o modelos dinámicos).

1.1 Series de tiempo

1.1.1 Definición:

Una serie de tiempo consiste de una secuencia de valores de una variable Y , cada una de las cuales es observada o registrada en un momento específico del tiempo t . De aquí en adelante, se usa la notación Y_t para representar una observación en un momento determinado del tiempo t .

Las series de tiempo pueden ser *continuas* o *discretas*, las primeras ocurren cuando es posible registrar las observaciones en cada instante del tiempo; mientras que en las segundas el conjunto de posibles valores para el tiempo de realización es un conjunto discreto; por ejemplo, cuando las observaciones son realizadas a intervalos fijos de tiempo. Este último tipo de serie es el que se usa en esta tesis y consiste de lecturas tomadas usualmente en forma diaria, mensual, trimestral o anual; son ejemplos en economía la evolución mensual del PBI, las cifras mensuales de importaciones y exportaciones, la cotización Promedio Mensual del Barril de Petróleo (véase Fig. 1, página 6), etc.

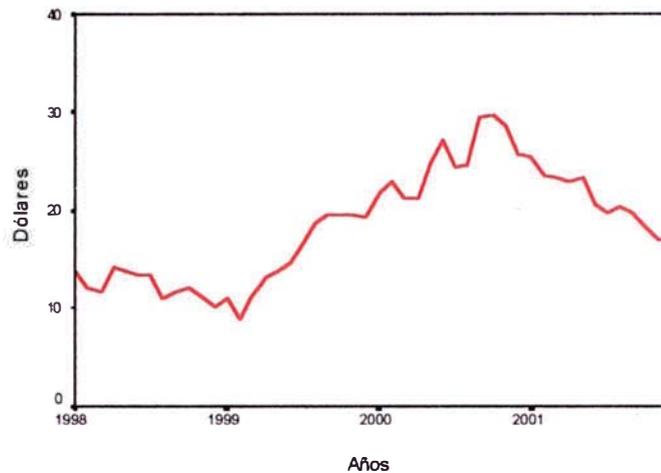


Figura1
Cotización Promedio Mensual del Barril de Petróleo Período: 1998- 2001

Las observaciones de las que se puede disponer para análisis son en general solo una parte de la serie historia que se estudia; esta puede entonces interpretarse como una muestra de la población representada por la serie de tiempo completa. A diferencia de lo que sucede en una muestra aleatoria simple, aquí no se espera que las observaciones sean independientes; más bien, la característica más importante de una serie de tiempo es que las observaciones sean estadísticamente dependientes; es justamente esta dependencia la que se desea investigar, ya que el propósito es que el comportamiento pasado de la serie explique el comportamiento futuro.

Una serie de tiempo puede entenderse en términos más formales como la realización de un proceso estocástico. Un proceso estocástico o aleatorio es una familia de variables aleatorias $\{Y_t, t \in T\}$ definida en un espacio probabilístico (Ω, \mathcal{F}, P) . En cada instante del tiempo se define una variable aleatoria, pero en realidad se dispone solo un valor para cada instante, se puede entonces entender

la serie como una muestra de tamaño uno tomada en periodos sucesivos de tiempo en un proceso estocástico.

1.1.2 Estacionariedad

Para empezar a trabajar con las series de tiempo es necesario determinar sus características y es aquí donde juega un papel importante el concepto de estacionariedad. La estacionariedad es un estado particular de equilibrio estadístico, significa que se tiene un proceso estable en el tiempo, un proceso donde sus parámetros y distribución no cambian. Cuando este es el caso, pueden estimarse la media y covarianzas a partir de una sola realización; lo cual no sería posible de otro modo.

Se tienen dos definiciones de estacionariedad: la estacionariedad débil y la estricta. Un proceso estocástico Y_t es *estacionario en el sentido estricto* si la distribución conjunta F de $Y = (Y_1, Y_2, \dots, Y_n)$ satisface la siguiente condición:

$$F(Y_1, Y_2, \dots, Y_n) = F(Y_{1+k}, Y_{2+k}, \dots, Y_{n+k})$$

Para todo $(Y_1, Y_2, \dots, Y_n) \in \mathbb{R}^n$ y todo escalar k .

En otras palabras la distribución conjunta no se ve afectada al cambiar todos los momentos de observación por una constante hacia atrás o adelante; es decir, el mecanismo generador del proceso es invariante en el tiempo.

La estacionariedad débil tiene requerimientos menos restrictivos, los momentos hasta algún orden dependen solo de las diferencias en los momentos de tiempo; la estacionariedad débil llamada también *estacionariedad de segundo orden* o en *covarianza* requiere que:

- $E(Y_t) = \mu_t = \mu$, Para todo t. (Media del Proceso Estocástico)
- $V(Y_t) = \sigma_t^2 = \sigma^2$, Para todo t.
- $\rho(Y_t, Y_k) = \rho(t-k)$, Para todo t, k

Las condiciones señalan que la media y varianza permanecen constantes y la correlación entre Y_t y Y_k depende solo de la diferencia entre los momentos en el tiempo a los que hacen referencia, esta correlación entre una misma variable es conocida como *autocorrelación*.

La estacionariedad de segundo orden más la suposición de normalidad son suficientes para producir la estacionariedad estricta.

1.1.3 Procesos lineales

Los procesos lineales deben su nombre a que pueden representarse como una combinación lineal de variables aleatorias. Los más importantes son: procesos de ruido blanco, procesos autoregresivos, procesos de medias móviles, y los procesos que resultan como combinación de estos dos últimos.

Antes de presentar los modelos lineales es útil introducir el *operador de retardo* B que es de gran utilidad en la representación de las series. El operador de retardo representado por B tiene la siguiente acción:

$$B Y_t = Y_{t-1}, \quad B^j Y_t = Y_{t-j}$$

Una serie de tiempo muy importante es el llamado *proceso de ruido blanco*. Un proceso $\{\varepsilon_t\}$ se dice que es ruido blanco con media cero y varianza σ^2 , escrito $\{\varepsilon_t\}$ es $WN^2(0, \sigma^2)$; si y solo si, ε_t es no correlacionado en serie; si además la variable aleatoria ε_t es independiente e idénticamente distribuida, se dice que ε_t es ruido blanco independiente e idénticamente distribuido y se escribe $\{\varepsilon_t\}$ es $iid(0, \sigma^2)$. Y si adicionalmente ε_t tiene distribución normal, entonces tenemos un ruido blanco normal Gausiano y se escribirá $\{\varepsilon_t\}$ es $iid N(0, \sigma^2)$.

Un *proceso de medias móviles* de orden q o un proceso MA (q) viene dado por:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} = \mu + (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon_t = \mu + \theta(B) \varepsilon_t$$

² De las siglas en inglés de *White noise* (Ruido blanco)

donde $\{\varepsilon_t\}$ es WN $(0, \sigma^2)$, $(\theta_1, \theta_2, \dots, \theta_q)$ son constantes, μ es la media de Y_t (se usa μ y no μ_t porque el proceso de medias móviles es estacionario), $\theta(B)$ es llamado *polinomio de medias móviles*. El caso más simple es un proceso MA (1): $Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$, media móvil de orden 1.

Un *proceso autoregresivo* AR (p) se expresa de la siguiente forma:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t; \text{ o de manera equivalente}$$

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

donde c es una constante, $\{\varepsilon_t\}$ es WN $(0, \sigma^2)$, y $\phi(B)$ es el *polinomio autoregresivo*. Un proceso autoregresivo de orden 1 o AR (1) satisface la ecuación: $Y_t = c + \phi Y_{t-1} + \varepsilon_t$.

Los *procesos mixtos* incluyen los términos autoregresivos y de medias móviles, su notación es ARMA (p, q): proceso autoregresivo de medias móviles de ordenes p y q:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad \text{ó}$$

$$\phi(B) Y_t = c + \theta(B) \varepsilon_t$$

donde c es una constante, $\{\varepsilon_t\}$ es $WN(0, \sigma^2)$, y $\phi(B)$ con $\theta(B)$ son los definidos anteriormente.

El caso más simple es un ARMA (1,1):

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

1.1.4 Condiciones de estacionariedad e invertibilidad

Por qué usar los procesos estacionarios y lineales, el *Teorema de Wald* ayuda a responder esta pregunta; en términos simples el teorema de Wald establece que *"cualquier proceso de media cero y estacionario en covarianza puede representarse como la suma de de dos procesos; uno de ellos es un proceso de medias móviles de orden infinito y el otro un proceso autoregresivo de orden p ".*

Esto hace recordar a un proceso ARMA; sin embargo, no se puede estimar un número infinito de parámetros para el proceso de medias móviles, el compromiso es entonces buscar el modelo con un número finito de términos que mejor se ajuste a los datos disponibles.

Aparte de la estacionariedad existe otra cualidad llamada invertibilidad que consiste en la propiedad de pasar de una representación a otra. Es decir pasar de un proceso AR a un MA y viceversa.

Las condiciones de estacionariedad e invertibilidad para los procesos presentados son las siguientes:

- Un proceso MA (q) es siempre estacionario; pero es invertible solo si las raíces de su polinomio de medias móviles $\theta(B)$ están fuera del círculo unitario.
- Un Proceso AR (p) es invertible, pero es estacionario solo si las raíces de su polinomio autoregresivo $\phi(B)$ están fuera del círculo unitario.
- Un proceso ARMA (p, q) es estacionario si las raíces del polinomio autoregresivo están fuera del círculo unitario, e invertible si las raíces del polinomio de medias móviles están fuera del círculo unitario.

Por otra parte, el proceso de ruido blanco $\{\varepsilon_t\}$ es estacionario de segundo orden ya que su media y varianza son constantes, y al ser no correlacionado, por definición, su coeficiente de autocorrelación es uno o cero; cumpliendo así las condiciones requeridas por la estacionariedad.

1.1.5 Identificación y correlograma

Este título se centra en los procesos estacionarios, en el siguiente se trata a los no estacionarios. Las herramientas que se utilizan para la identificación de los procesos generadores de las series son las funciones de autocorrelación y autocorrelación parcial.

Sea $\rho_k = \rho(t + k, t) = \gamma_k / \gamma_0$ la *función de autocorrelación o fac*

Una de las características de un proceso estacionario, es una función de autocorrelación que finalmente tiende a cero al crecer k . La representación gráfica de ρ_k para $k = 0, 1, \dots$ recibe el nombre de *Correlograma*. A la *función de autocorrelación parcial fac* la denotaremos por α_k , la autocorrelación parcial en el rezago k puede ser considerada como la correlación entre Y_1 e Y_{k+1} después de controlar los efectos de Y_2, Y_3, \dots, Y_k ; al igual que las autocorrelaciones las autocorrelaciones parciales se suelen graficar para $k = 0, 1, \dots$ este gráfico será llamado *Correlograma Parcial*.

A continuación el comportamiento de los Correlogramas en los distintos procesos estudiados.

Un proceso autoregresivo $AR(p)$ estacionario está caracterizado por una función de autocorrelación parcial que tiene los primeros p coeficientes no nulos y el

resto cero, mientras que la función de autocorrelación presenta en general una mezcla de patrones exponenciales y sinusoidales que decrecen con el tiempo.

Un proceso de medias móviles de orden q se caracteriza porque los primeros q coeficientes de la función de autocorrelación son no nulos y el resto cero; por otra parte la función de autocorrelación parcial presenta en general muchos coeficientes no nulos que decrecen con mezcla de patrones exponenciales y sinusoidales.

En el caso del proceso ARMA los patrones en el correlograma y correlograma parcial son más complejos; la fac y la fap de un proceso ARMA (p, q) es una superposición de las correspondientes en los procesos AR (p) y MA (q) .

Los patrones que se acaban de describir son los teóricos, pero en la práctica estaremos frente a muestras de solo una parte de la serie de tiempo. En este caso los patrones que debemos esperar son aproximados y se recomienda no utilizar el correlograma para un número de coeficientes de autocorrelación superior a un tercio o un cuarto del tamaño de la muestra³.

³ De Introducción al Análisis de Series Temporales de Ezequiel Uriel y Amado Peiró, Editorial AC, año 2000

1.1.6 Procesos no estacionarios

Las series reales no se presentan, en su mayoría, como estacionarias; es común en economía observar series cuya media en el tiempo no es constante, sino más bien que varía con ésta; generalmente con una tendencia creciente o decreciente. Otra fuente de no estacionariedad puede ser la no presencia de una variabilidad constante. Los procesos no estacionarios que se tratarán ahora son aquellos que pueden transformarse en estacionarios mediante la toma de diferencias de algún orden, estos procesos son llamados modelos no estacionarios homogéneos o integrados.

Modelos ARIMA

Son una generalización de los procesos ARMA para incluir el término de diferenciación, también conocido como integración. El nuevo proceso resultante se denomina proceso autoregresivo integrado de medias móviles de ordenes p , d , y q donde d es el orden de diferenciación: ARIMA (p , d , q).

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \nabla^d Y_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon_t + c$$

$$\phi_p(B) W_t = \theta_q(B) \varepsilon_t + c$$

ε_t es ruido blanco, ∇ es el operador de diferencias, definido de la siguiente manera: $\nabla^1 Y_t = Y_t - Y_{t-1}$, $\nabla^d Y_t = Y_t - Y_{t-d}$; y $W_t = \nabla^d Y_t$

Entonces el proceso ARIMA (p, d, q) para Y_t es equivalente al proceso ARMA (p, q) para W_t . Si el proceso $\{ W_t \}$ es estacionario, el proceso ARIMA para Y_t quedará caracterizado por las f_a y f_{ap} de $\{ W_t \}$ y por d el orden de diferenciación del proceso original.

La toma de diferencias esta relacionada con la presencia de una tendencia en la serie. La tendencia ya sea esta estocástica (tiene cambios aleatorios de pendiente y nivel) o determinística, debe ser removida para obtener una serie estacionaria. En el caso de una tendencia determinística, esta será removida restando a la serie la tendencia; si la tendencia es estocástica esta se removerá tomando diferencias; generalmente una o dos diferencias serán suficientes.

Los modelos ARIMA cubren los procesos no estacionarios homogéneos, que en la gran mayoría de los casos son suficientes para describir el comportamiento de las series económicas. No obstante, puede estar presente la falta de homogeneidad para la variabilidad; es usual encontrar series cuya varianza tiene tendencia en el tiempo, y para la cual la toma de diferencias no estabiliza la varianza. Lo usual en estos casos es transformar la serie mediante la toma de logaritmos. Box y Cox definieron una transformación de carácter general para estabilizar la varianza.

$$f_{\lambda}(Y_t) = \begin{cases} \lambda^{-1}(Y_t^{\lambda} - 1) & , Y_t \geq 0, \lambda > 0 \\ \ln(Y_t) & , Y_t > 0, \lambda = 0 \end{cases}$$

en la práctica, si la transformación de Box-Cox es necesaria; es usual que λ sea igual a cero o un medio.

Modelos Arima Estacionales

La estacionalidad es un patrón repetitivo que se presenta cada año y es muy común en las series económicas. Ejemplo de este tipo de comportamiento son las ventas mensuales de gaseosa; en esta serie se observarían picos en las ventas durante los meses de verano de cada año. Se extienden entonces los modelos ARIMA a modelos ARIMA estacionales conocidos también como SARIMA, una característica importante de estos modelos es que permiten la aleatoriedad en el patrón estacional de un ciclo a otro.

Si la autocorrelación entre las observaciones esta presente solo para aquellas observaciones que distan entre si s períodos, siendo s el período estacional, entonces se tendrá un *proceso estacional puro*. La notación general para este proceso será ARMA (P, Q), donde P, Q son los órdenes estacionales autoregresivo y de medias móviles respectivamente:

$$\Phi_P(B^s) Y_t = \Theta_Q(B^s) \varepsilon_t + c, \text{ donde } \Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}$$

Los valores P Y Q son similares a los p y q del modelo no estacional, de hecho las funciones de autocorrelación y autocorrelación parcial tienen el mismo comportamiento, con la diferencia de que solo son distintos de cero para múltiplos de s y cero en otros casos. P, Q serán llamados ordenes estacionales y p, q ordenes regulares; igual tratamiento tendrán los ordenes de diferenciación o integración d y D.

En el modelo estacional puro solo se encuentran presentes las autocorrelaciones en los múltiplos de período s. Sin embargo, esta suposición no es muy realista; pues es muy poco probable que en una serie estacional mensual los diferentes meses estén no correlacionados, podría más bien estar presente algún tipo de tendencia creciente o decreciente entre las observaciones además de la componente estacional.

Para introducir la dependencia entre todas las observaciones tenemos que combinar el modelo estacional puro con el modelo ARMA (p, q). En el caso más general puede resultar que la serie presente algún tipo de orden de integración regular, es decir, se tendría un modelo ARIMA(p, d, q); de igual forma la parte estacional puede no ser estacionaria y necesitar también algún orden de diferenciación estacional y tendríamos así un modelo ARIMA(P, D, Q), un indicio

de que la componente estacional no es estacionaria, es el hecho de que los coeficientes de autocorrelación correspondientes a múltiplos del período s : $s, 2s, 3s, \dots$ disminuyen lentamente. Combinando los modelos que se acaban de mencionar se llega al modelo $ARIMA_s(p, d, q, P, D, Q)$, así el subíndice s indica el período y el hecho de que el modelo $ARIMA$ es estacional)

$$\Phi_p(B^s) \phi_p(B) (1-B)^d (1-B^s)^D Y_t = \theta_q(B) \Theta_Q(B^s) \varepsilon_t + c$$

donde ε_t es ruido blanco, d y D son los ordenes de diferenciación regular y estacional respectivamente, por último los polinomios son los ya definidos anteriormente.

1.1.7 Validación

Al ajustar un modelo $ARMA$ se busca que este sea una representación adecuada del proceso generador de la serie. En otras palabras se busca que este cumpla los siguientes requerimientos:

- La serie debe ser estacionaria e invertible.
- El error deben comportarse aproximadamente como ruido blanco.
- Los coeficientes deben ser estadísticamente significativos.

- El número de coeficientes del modelo debe ser el adecuado para representar la serie
- El grado de ajuste debe ser mayor en comparación al de otros modelos alternativos.

El cumplimiento del primer requisito ya fue tratado en las secciones anteriores.

Error

Los residuos $e_i = y_i - \hat{y}_i$ (no son los errores, pues no se conoce el verdadero valor de los parámetros, solo sus estimaciones) donde $i = 1, 2, \dots, n$; y_i es una observación real e \hat{y}_i es el correspondiente valor ajustado a partir del modelo de series de tiempo; deben tener media cero, varianza constante, y estar no correlacionados; un requerimiento adicional pero no necesario es que tengan distribución normal.

La confirmación de una media nula y varianza constante, puede realizarse utilizando un gráfico de línea de los residuos. De manera más formal puede usarse una prueba de hipótesis para verificar que la media es cero; la homogeneidad de la varianza puede probarse usando la prueba de Levene, aunque esta propiedad debe ser consecuencia de la estacionariedad de la serie; para llevar a cabo esta prueba, se divide la serie en grupos, y la hipótesis nula consiste en que los grupos proceden de poblaciones con varianza común, esta

prueba es menos dependiente de la suposición de normalidad que la mayoría de pruebas. La contraparte gráfica serían los diagramas de caja para cada grupo de la serie.

La propiedad de no-correlación puede verificarse, también, usando el correlograma con límites de confianza al 95%. De igual forma se pueden usar pruebas estadísticas globales como la de Box-Pierce (1970), reformulada más tarde y conocida como prueba de Ljung-Box (1978); en esta prueba se usa un estadístico Q para probar que los coeficientes de correlación hasta un cierto retardo son iguales a cero. Algunos analistas recomiendan examinar esta prueba en el retardo igual a $\frac{1}{4}$ de la longitud de la serie, y si el Valor P es menor que 0,05 concluir que la serie presenta autocorrelación⁴.

La normalidad de los residuos puede comprobarse a primera vista usando la gráfica de distribución de la serie junto con gráficas de probabilidad normal, pero más objetivamente usando una prueba de bondad de ajuste como la de *Kolmogorov Smirnov* o la *Prueba K-S de Lilliefors* (Prueba de Kolmogorov Smirnov con corrección de Lilliefors), o la *Prueba de normalidad de Jarque Vera*.

⁴ De Time Series Analysis and Forecasting with SPSS Trends 9.0. SPSS Inc. 1999

Coeficientes significativos

Una vez que se ha determinado los órdenes de un modelo ARIMA, se deben estimar los parámetros que la caracterizan. Las técnicas usuales de estimación que se usan para los Modelos ARIMA son las de Mínimos Cuadrados y la de Máxima Verosimilitud, dentro de estos métodos pueden usarse enfoques condicionados o no condicionados. Las suposiciones usuales para estas técnicas son que el error es ruido blanco (la normalidad es necesaria si se usa máxima verosimilitud) y que la serie es estacionaria e invertible. Muchos de los programas de computadora usan la técnica de Máxima Verosimilitud para la estimación, por ejemplo SPSS usa una subrutina de máxima verosimilitud que permite estimar series de tiempo con observaciones faltantes; EViews por otra parte utiliza mínimos cuadrados no lineales.

Una vez estimados los parámetros del modelo, debe de realizarse pruebas de significancia de estos; los resultados de los programas de computadora presentan los parámetros estimados junto con los estadísticos T y los Valores P, estos nos permiten probar si los parámetros contribuyen o no de manera significativa al modelo.

Sobreajustes

Luego de identificar un modelo o los modelos candidatos, esto no excluye la posibilidad de un modelo más complejo; la técnica de sobreajuste permite explorar estos modelos, consiste en aumentar en una unidad los ordenes del modelo elegido p, q, P, Q ; este aumento no se da de manera simultánea entre las partes autoregresiva y de medias móviles; por ejemplo, si el modelo elegido es $ARIMA_s(p, d, q, P, D, Q)$ entonces los sobreajustes son:

$$ARIMA_s(p+1, d, q, P, D, Q)$$

$$ARIMA_s(p, d, q+1, P, D, Q)$$

$$ARIMA_s(p, d, q, P+1, D, Q)$$

$$ARIMA_s(p, d, q, P, D, Q+1)$$

Una vez ajustados estos modelos se evalúa la significancia de sus coeficientes; si estos son significativos, entonces se comparan estos modelos en función de los estadísticos de Bondad de Ajuste (ver más adelante). Si uno de los sobreajustes resulta ser mejor que el modelo original, este modelo es elegido y se vuelve a aplicar la técnica de sobreajuste, este procedimiento continua hasta no hallar otro modelo que mejore el modelo original; sin embargo, se recomienda que los ordenes del modelo no pasen de las dos unidades.

Medidas de Bondad de Ajuste

Los estadísticos usuales R^2 y R^2 ajustado usados para medir la bondad de ajuste de modelos; no son los más utilizados en series de tiempo, una de las razones es que no permiten la comparación de modelos con número distinto de diferencias regulares o estacionales. Los estadísticos de mayor utilización son los *criterios de información de Akaike AIC y el Schwarz SIC*; estos no tienen el problema antes mencionado y además permiten evaluar modelos con diferentes transformaciones de Box-Cox. Estos estadísticos castigan la inclusión de parámetros adicionales mediante una penalización. En la práctica se usan ambos y usualmente conducen al mismo modelo, pero si este no es el caso, con frecuencia se prefiere el modelo elegido mediante el criterio de Schwarz puesto que el AIC tiende a seleccionar modelos que son demasiado grandes (sobreparametrización) debido a que tiene una menor penalización que el SIC. El criterio en ambos casos es el de elegir el modelo con el menor valor de AIC o SIC.

También se debe tener en cuenta el principio de parsimonia, éste consiste en lo siguiente: si se tienen dos modelos en el que uno tiene más coeficientes que el otro, pero los criterios de ajuste nos dicen que son similares, debe entonces preferirse el modelo más simple.

1.1.8 Predicción

Uno de los objetivos al construir un modelo, es el de predicción; se eligió los modelos ARIMA porque son óptimos en el sentido de que no otro pronóstico univariante tiene menor error cuadrado medio de pronóstico; siempre que se halla encontrado el modelo apropiado y dentro de los pronósticos de modelos univariantes lineales y de coeficientes fijos.

Estos modelos son usados para pronósticos de corto plazo; es decir, con antelación igual o inferior a un año. Para períodos mayores es preferible el uso de modelos causales; pues el futuro puede significar un comportamiento distinto del pasado reciente.

Para la generación de pronósticos, se asume que los parámetros del modelo son conocidos; en la práctica esta presente el error de estimación de los parámetros pero este es de poca importancia en muestras grandes. El software SPSS genera pronósticos usando mínimos cuadrados exactos y asumiendo que los parámetros estimados son los verdaderos, también se tiene la opción de mínimos cuadrados condicionales. Para una serie estacionaria los pronósticos tienden en el largo plazo a la media de la serie; la rapidez de la convergencia depende del modelo. En general en modelos MA puros la convergencia puede ser rápida mientras que en un modelo con término AR esta puede tomar más tiempo.

Las estimaciones puntuales pueden acompañarse de intervalos de confianza, en un modelo AR los límites irán creciendo continuamente junto con el horizonte de predicción. En el caso de modelos MA, estos permanecen constantes a partir del período q hacia el futuro. El carácter creciente de estos intervalos hace que su implementación no sea adecuada.

1.1.9 Modelos ARIMA con variables exógenas

Hasta el momento no se ha tratado el efecto de otras variables de serie de tiempo en la predicción de la serie que esta bajo estudio; puede ganarse precisión si se incluye otra información que ayude en la construcción del modelo generador de la serie que se busca predecir.

El modelo ARIMA que capta la dinámica propia de la variable Y_t

$$\phi_p(B) W_t = \theta_q(B) \varepsilon_t + c, \quad W_t = \nabla^d Y_t$$

puede ser ampliado para incluir el efecto estacional que presenta esta variable y se obtiene así un modelo SARIMA ó ARIMA estacional

$$\Phi_P(B^s) \phi_p(B) (1-B)^d (1-B^s)^D Y_t = \theta_q(B) \Theta_Q(B^s) \varepsilon_t + c$$

Aún más este último modelo puede ser extendido para captar la dinámica de otra serie de tiempo X_t , esta variante del modelo ARIMA se denomina ARIMAX y en el caso de un modelo estacional tiene el nombre SARIMAX.

$$\Phi_p(B^s) \phi_p(B) (1-B)^d (1-B^s)^D Y_t = w(B) X_t + \theta_q(B) \Theta_Q(B^s) \varepsilon_t + c$$

$$\text{donde } w(B) = w_0 + w_1B + w_2B^2 + \dots$$

Así el nuevo modelo resultante incluye la información de otra variable de serie de tiempo y sus rezagos respectivos. Este nuevo modelo combina la metodología de Box - Jenkins con la del análisis causal.

Para saber cuántos rezagos de la variable exógena incluir, puede usarse la función de correlación cruzada; esta función es útil para analizar si dos series distintas están relacionadas, permite determinar si una serie puede ser predicha a partir de otra y que ordenes de rezago pueden ser los más adecuados. Esta función calcula la correlación entre el valor de una variable en el tiempo t con los valores de otras variables en los tiempos $t-1$, $t-2$, etc:

$$\rho_{YX}(k) = \frac{\gamma_{YX}(k)}{\sigma_Y \sigma_X}, \quad \text{donde } \sigma_Y \text{ y } \sigma_X \text{ son las desviaciones estándar de } Y \text{ y } X$$

respectivamente,

$$\gamma_{YX}(k) = E (Y_{t+k} - \mu_Y) (X_t - \mu_X), \text{ y } k = 0, \pm 1, \pm 2, \pm 3, \dots$$

Antes de examinar la relación entre dos series de tiempo; se debe estar seguro de que ambas series son estacionarias. Las series de tiempo son consideradas como realizaciones de un proceso estocástico y si estas no son estacionarias sus parámetros no son estables en el tiempo.

Una gráfica de autocorrelación cruzada esta usualmente formada por barras que indican el valor de los coeficientes para los rezagos y límites de confianza que señalan si estos son significativos o no; en este tipo de gráfico se muestran los rezagos tanto negativos como positivos. La lógica es la siguiente; un rezago negativo indica que la segunda serie es la que debe ser usada como predictora, por otra parte un rezago positivo es señal de que la primera serie es la que debe usarse como predictora. Los órdenes de rezago de las variables predictoras serán las que se encuentre respectivamente en la parte positiva o negativa. Un rezago cero indica que la correlación es para las dos series diferenciadas cuando ninguna esta rezagada.

Por ejemplo en el gráfico de correlación cruzada que se presenta (véase la Fig. 2, página 29), se muestra una posible relación entre el Producto Bruto Interno (GDP) y el Consumo en valor real.

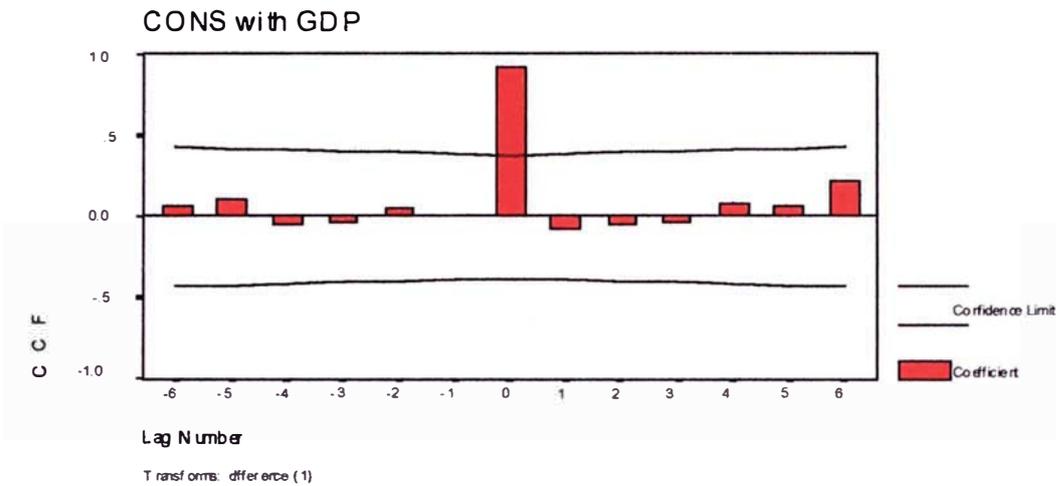


Figura2
Correlación Cruzada entre las variables CONS y GDP

Aquí puede observarse que la única correlación cruzada significativa es en el rezago 0, es decir cuando no hay rezago.

Más de una variable de serie de tiempo puede ser incluida en el modelo ARIMAX con los que se tendría la expresión $\sum w_i(B) X_{ti}$ en vez de $w(B) X_t$.

En la aplicación de la función de autocorrelación cruzada que se desarrollara más adelante; el sentido de si una u otra variable es la predictora estará ya determinado por un marco teórico. El interés será entonces el de conocer los rezagos que deben tomarse en consideración.

1.2 Análisis de Regresión

1.2.1 Definición

El análisis de regresión es una técnica estadística que permite explorar la relación entre un variable (llamada dependiente o explicada) y otras variables (llamadas independientes, explicativas, o regresoras). Se asume en este análisis que la variable independiente denotada por Y es aleatoria y que el valor de las variables explicativas X_s es conocido. Los usos del análisis de regresión son múltiples y variados, abarcan campos como la economía, los negocios, el marketing, etc. Por ejemplo en economía este constituye la base de muchos modelos de previsión económica.

El análisis de regresión puede ser simple o múltiple, en el primer caso solo se maneja una variable explicativa; mientras que en el segundo se pueden tener dos o más de estas variables.

El modelo usual con el que se trabaja es el modelo lineal, recibiendo la técnica el nombre de análisis de regresión lineal; sin embargo, esto no impide la utilización de otros modelos (muchos de los cuales pueden usualmente adaptarse a un

modelo lineal), la ecuación de regresión lineal múltiple tiene la siguiente expresión:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

Este modelo describe un hiperplano en el espacio de dimensión k formado por las variables de regresión X_j ; el término lineal se aplica porque esta ecuación es una función lineal de los parámetros desconocidos β_j , Y es la variable explicada, X_j son las variables explicativas y ϵ es el término que representa al error aleatorio llamado también ruido blanco; se asume que esta última variable tiene una distribución con media igual a cero, varianza constante, y que no presenta autocorrelación, la condición de normalidad también es usualmente exigida para las pruebas estadísticas. En consecuencia la media y varianza de la variable independiente Y es:

$E(Y/X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ la media es condicional porque los valores de X_s son conocidos

$$V(Y_t) = \sigma^2$$

Generalmente los modelos de regresión lineal múltiple se emplean como funciones de aproximación. Es decir, la verdadera relación funcional entre Y y las X_s se desconoce pero para el rango de estudio el modelo de regresión lineal constituye una aproximación adecuada. El diagrama de dispersión (véase la Fig.3, página 32) muestra la relación entre el valor de las importaciones y el

índice del producto bruto interno PBI, puede observarse que la relación lineal puede ser una buena aproximación inicial.

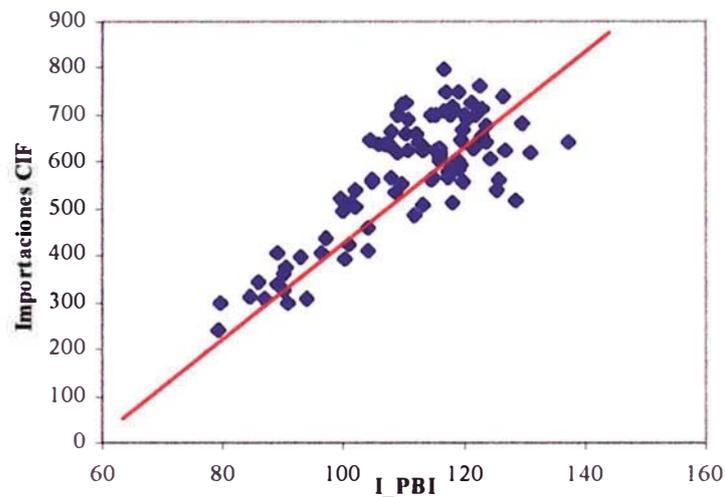


Figura 3
Diagrama de dispersión entre las Importaciones y el Índice del PBI

1.2.2 Suposiciones

Para la correcta aplicación del análisis de regresión deben de cumplirse una serie de condiciones, las cuales se pasa a enumerar:

- El error debe tener media cero
- El error debe tener variabilidad constante
- Los errores no deben estar correlacionados
- Todas las variables predictoras deben ser no estocásticas

- El número de observaciones debe exceder el número de coeficientes ha ser estimado
- No debe haber relaciones lineales exactas entre cualesquiera de las variables explicativas (multicolinealidad).

1.2.3 Elección de variables

La elección de las variables explicativas debe basarse en fundamentos teóricos y conceptuales; los trabajos previos de otros investigadores deben también tomarse en cuenta. Es peligrosa la inclusión de variables no relevantes ya que estas pueden tener correlación con la variable explicada; pero esta puede ser solo porque ambas comparten igual tendencia, lo que ocasionaría relaciones estadísticas entre variables que no están realmente relacionadas.

Cuando se está investigando varias variables como candidatas a explicativas pueden usarse técnicas como "todas las regresiones", o la "regresión por pasos". En la primera de estas técnicas se ajustan todas las ecuaciones de regresión tomando una variable candidata, luego todas las ecuaciones de regresión tomando dos variables candidatas y así sucesivamente; el siguiente paso es evaluar estas ecuaciones de acuerdo a algún criterio adecuado para seleccionar el mejor modelo de regresión. En la regresión por pasos, que es probablemente la técnica más utilizada, el procedimiento construye de manera iterativa una secuencia de modelos de regresión mediante la adición o eliminación de variables en cada paso.

1.2.4 Análisis Gráfico

El paso siguiente a la elección de las variables es realizar un análisis gráfico exploratorio de la información que esta contenida en estas y la variable dependiente; el propósito es identificar las posibles relaciones que están presentes entre las variables, así como de conocer aquellas que podrían ser omitidas del análisis, y detectar la presencia de valores atípicos que puedan distorsionar el estudio.

Para este propósito se hace uso de graficas de dispersión y del coeficiente de correlación lineal de Pearson; esta última estadística mide el grado de relación lineal entre dos variables.

En el caso de regresiones lineales múltiples, los gráficos de dispersión y medidas de correlación solo reflejan la interacción entre dos variables, cuando en realidad en el modelo lineal múltiple se toman todas las variables de manera conjunta. Debe entonces tomarse los resultados como referenciales ya que no representan la interacción simultanea entre las variables.

1.2.5 Estimación

Para la estimación de los coeficientes del modelo de regresión lineal se utiliza la técnica de *mínimos cuadrados ordinarios*; este es un procedimiento mediante el cual se obtienen estimaciones que minimizan la suma de cuadrados de los

errores de estimación. Este método es puramente matemático no siendo necesaria ninguna suposición para su aplicación.

El estimador de mínimos cuadrados $\hat{\beta}$ es el mejor estimador lineal insesgado de β ; si adicionalmente asumimos que los errores están normalmente distribuidos, entonces $\hat{\beta}$ es también el estimador de máxima verosimilitud de β . El estimador de máxima verosimilitud es el estimador insesgado de mínima varianza de β .

Si la muestra no es muy pequeña esta puede dividirse en dos partes, la primera servirá para la estimación propiamente dicha y la otra para medir la capacidad predictiva del modelo.

La muestra está constituida por n observaciones, si x_{ij} es la i -ésima observación de la variable x_j ; entonces las observaciones son:

$$(x_{i1}, x_{i2}, \dots, x_{ik}, y_i) \quad i=1,2,\dots,n$$

1.2.6 Validación

Debe de verificarse varias condiciones necesarias para la aceptación del modelo de regresión lineal.

Errores

El error en el modelo de regresión lineal debe de tener media cero, varianza constante y no estar correlacionado; en la verificación de estas suposiciones se usan los residuos $e_i = y_i - \hat{y}_i$ (no son los errores, ya que no se conoce el verdadero valor de los parámetros, solo sus estimaciones) donde $i = 1, 2, \dots, n$; y_i es una observación real e \hat{y}_i es el correspondiente valor ajustado a partir del modelo de regresión. Para la verificación de las condiciones que deben cumplir la media, varianza y correlación de los residuos pueden usarse las mismas técnicas mencionadas para la validación de residuos en series de tiempo. Si no se cumple la condición de no correlación, como es usual el caso para datos de series de tiempo; el término de error del modelo de regresión lineal puede ser modelado usando procesos ARIMA y así obtener un nuevo error que cumpla las condiciones necesarias.

Adicionalmente a los requerimientos mencionados en el primer párrafo, la normalidad es necesaria para llevar a cabo pruebas de significancia de los coeficientes de regresión. Para verificar que se cumpla este requerimiento pueden también usarse las pruebas mencionadas para los residuos en las series de tiempo.

Coefficientes significativos

Una vez estimados los coeficientes de regresión pueden usarse pruebas de hipótesis para verificar que estos sean significativos. Las pruebas usuales son: la prueba T para cada coeficiente y la prueba F, prueba conjunta de los coeficientes de regresión menos el intercepto, llamada también prueba de significancia de la regresión, esta prueba determina si existe una relación lineal entre la variable Y y las variables independientes X_1, X_2, \dots, X_k . Estas pruebas necesitan de la suposición de normalidad de los residuos.

Medidas de adecuación del modelo

La adecuación de un modelo de regresión lineal múltiple puede medirse empleando las siguientes estadísticas:

- Coeficiente de determinación múltiple ajustado
- Análisis residual y
- Observaciones influyentes

El coeficiente de determinación múltiple ajustado denotado simplemente por R^2 , es una estadística cuyo valor se encuentra entre 0 y 1 y es usualmente presentado en forma porcentual. Su valor expresa cuanto de la variabilidad en los datos es explicada o tomada en cuenta por el modelo de regresión.

El análisis de residuos es útil para verificar que los errores tienen una distribución que es aproximadamente normal con varianza constante, así como para determinar la utilidad que tiene la adición de más términos (variables explicativas) al modelo.

Para verificar la normalidad de los residuos puede usarse un histograma de frecuencias de los residuos o una gráfica de probabilidad normal. A menudo es útil graficar los residuos contra \hat{y}_i , las gráficas resultantes tienen usualmente un aspecto similar a los cuatro patrones generales que aparecen a continuación:

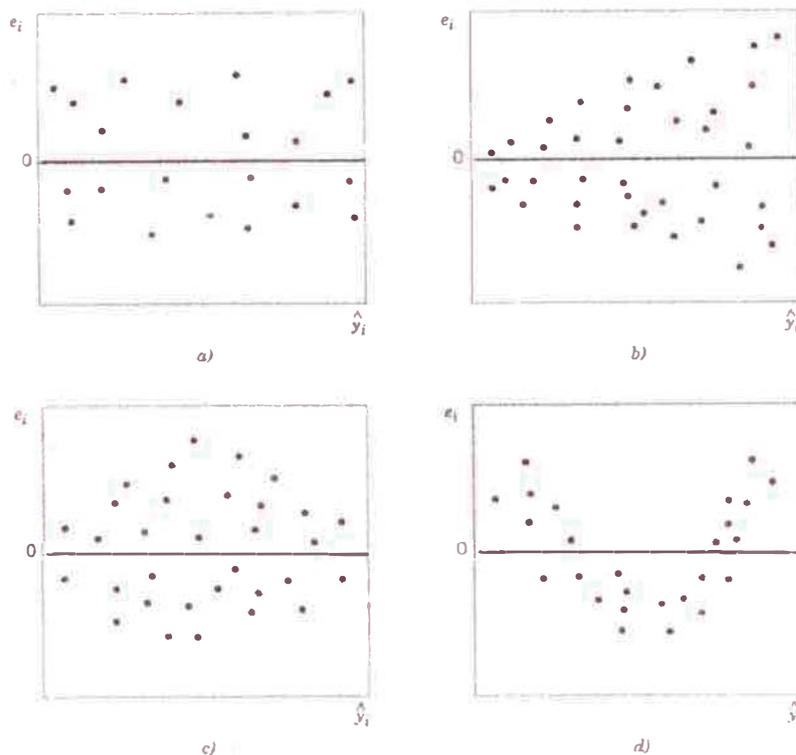


Figura 9-9 Patrones para gráficas de residuos, a) Satisfactorio, b) en embudo, c) doble arco, d) lineal. [Adaptado de Montgomery y Peck (1992).]

Figura 4
Patrones para gráficas de Residuos

La situación ideal es el patrón (a), los patrones (b), (c) y (d) representan anomalías; en (b) la varianza de las observaciones puede aumentar con el

tiempo o la magnitud de y . El patrón (c) también indica una desigualdad en la varianza; el patrón observado en (d) indica que el modelo es inadecuado; es decir, se necesitan más variables de orden superior, transformar las variables, o considerar otros regresores. Otra gráfica de residuos que puede emplearse es la gráfica de serie de tiempo para los residuos, obviamente esto es posible solo si las variables han sido registradas en el tiempo.

Las observaciones influyentes hacen referencia a un subconjunto de estas que tienen una influencia poco usual, estas observaciones están con frecuencia lejos de la vecindad donde se encuentran los demás datos recopilados. Estas observaciones pueden tener mucha influencia en la determinación del R^2 ajustado, en la estimación de los coeficientes de regresión y en la magnitud del error cuadrático medio. Por tanto es importante examinar estos puntos para determinar si ellos controlan muchas de las propiedades del modelo. Una estadística usada para este propósito es la medida de distancia de Cook D_i , un valor de $D_i > 1$ indica que el punto es influyente, también puede recurrirse a la exploración gráfica.

Multicolinealidad

La idea en el modelo de regresión es que la variable dependiente este relacionada con las variables independientes, pero también puede suceder que exista correlación entre las variables independientes; cuando estas dependencias son muy fuertes se dice que existe multicolinealidad. La multicolinealidad tiene efectos serios sobre los coeficientes estimados; ocasionando que las estimaciones de los coeficientes de regresión sean muy imprecisas; sin embargo, la ecuación del modelo ajustado es aún útil en el caso de predicciones por interpolación pero no para extrapolaciones.

Los siguientes son algunos indicadores de posible multicolinealidad.

- La prueba F es significativa, pero ninguna de las pruebas T individuales lo es.
- Los coeficientes estandarizados (ver más adelante) alcanzan magnitudes mayores a 1.
- Valores de tolerancia pequeños, cercanos a 0,01
- Número de condición mayor a 30
- Eigenvalores cercanos a cero.
- Factores de inflación de varianza mayores a 10.
- Coeficientes de correlación altos entre las variables independientes.

Para remediar el problema de multicolinealidad se han propuesto varias medidas:

- Incrementar los datos con nuevas observaciones diseñadas de manera específica para romper las dependencias entre las variables.
- Eliminar ciertas variables del modelo
- Usar otros métodos de estimación como la regresión arista.

1.2.7 Interpretación del modelo

Supongamos que tenemos el siguiente modelo de regresión lineal con k variables independientes o regresoras.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

el parámetro β_0 es llamado intercepto y si el rango de los datos incluye a $X_1 = \dots = X_k = 0$ entonces β_0 es la media de Y cuando $X_1 = \dots = X_k = 0$; de otro modo β_0 no tiene una interpretación física. Los otros coeficientes β_j representan el cambio esperado en la respuesta Y por cambio unitario en X_j cuando el resto de variables se mantienen constantes. Por esta razón los parámetros β son con frecuencia llamados coeficientes parciales de regresión.

La importancia de cada variable en el modelo de regresión puede medirse con la magnitud de los coeficientes respectivos. Sin embargo, en el modelo de regresión lineal las unidades del coeficiente de regresión β son (unidades de

$Y)/(unidades\ de\ X_j)$; por tanto la magnitud de β esta determinada por la selección de la escala para la variable de regresión X_j . Por tanto el trabajo con coeficientes de regresión adimensionales puede facilitar la determinación de la importancia relativa de cada variable, estos son los llamados coeficientes de regresión estandarizados.

No obstante al interpretarse los coeficientes estandarizados debe recordarse que estos siguen siendo coeficientes de regresión parciales; además, el espaciamento en los niveles de las variables X_j también tiene efecto. En consecuencia no debe emplearse estos coeficientes como la única medida de la importancia de las variables de regresión.

1.2.8 Predicción

Para la proyección de la variable dependiente, es necesario conocer los valores que tomaran las variables independientes en el futuro; sin embargo, no siempre se tiene esta información. En este caso la metodología usual es proyectar estas variables con series de tiempo, a menos que se tengan escenarios establecidos, u otro tipo de información.

En la proyección es necesario diferenciar la interpolación de la extrapolación, en el primero de los casos la proyección esta confinada a valores de la variable independiente dentro del rango de la información estadística usada para estimar

el modelos; en el último, la proyección es para valores de Xs que están fuera del rango mencionado antes, para este caso debe asumirse que la relación observada en la muestra se mantiene para fuera de esta ya que realmente las relaciones de regresión son válidas solo para los valores del regresor que están dentro del rango de los datos considerados, esta relación no es necesariamente válida en la extrapolación. Finalmente las proyecciones pueden ser puntuales o de intervalo.

1.2.9 Regresiones con errores ARMA

Los residuos en un modelo de regresión no deben estar correlacionados; si se viola este supuesto los estimadores obtenidos por mínimos cuadrados ordinarios continúan siendo lineales insesgados al igual que consistentes, pero dejan de ser eficientes (es decir, no tienen mínima varianza).

La causa de la autocorrelación en los residuos puede deberse a que se han omitido variables regresoras; la incorporación de estas puede solucionar el problema, pero si este no es el caso debe trabajarse con modelos que incorporen la estructura de autocorrelación de los residuos. Inicialmente puede suponerse, y es lo que generalmente se hace, que los residuos tienen un comportamiento autoregresivo de primer orden; sin embargo, esto no siempre es cierto. Puede entonces trabajarse con modelos más generales y considerar que los residuos se comportan de acuerdo a un modelo ARMA.

El modelo general sería entonces:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \epsilon_t, \text{ donde } \phi_p(B) \epsilon_t = \theta_q(B) \epsilon_t + c,$$

y ϵ_t es WN $(0, \sigma^2)$

Las regresiones que tienen esta componente ARMA ponen de manifiesto que la regresión (técnica estadística y econométrica de gran aplicación y tradición) y el modelo ARMA (de aparición más reciente no son en realidad una competencia sino complementarios cuando se usan adecuadamente.

1.3 Evaluación de la capacidad predictiva del modelo

Si la muestra es lo suficientemente grande es usual dividirla en dos partes; una será usada para el ajuste o estimación de coeficientes del modelo, y la otra será una muestra de validación; es decir, destinada a evaluar la capacidad predictiva del modelo. Una vez estimado el modelo con la muestra para el período de ajuste se efectúan las predicciones de manera que cubran el período de validación, es aquí donde se evalúa el grado de realismo del modelo, en otras palabras su capacidad para predecir hechos del mundo real. Para la evaluación pueden utilizarse los estadísticos de ajuste MAE y MAPE.

MAE es la media de los valores absolutos del error, y MAPE es el error medio porcentual de los errores absolutos. Con la primera estadística se puede conocer la magnitud promedio del error; con la segunda se tiene el error de predicción porcentual, es decir, una indicación del probable error porcentual de pronosticar.

Otro estadístico usual para medir la capacidad predictiva es la estadística U de Thail, si este es mayor que 1, la predicción es mala, la situación será mejor cuanto más se acerque su valor a cero.

La ecuación que expresa el promedio de los errores de pronóstico elevados al cuadrado puede descomponerse en tres factores de manera que juntos sumen la unidad; estos factores serían proporciones debidas al sesgo, a la varianza y a la covarianza. La primera de estas proporciones nos dice cuán lejos la media del

pronóstico esta de la media de la serie, la segunda proporción nos dice cuan lejos la variación del pronóstico esta de la variación de la serie, la última proporción mide el restante error de pronóstico no sistemático. Si el pronóstico es *bueno* las proporciones de sesgo y varianza deberían ser pequeñas; así la mayor parte de las desviaciones estarían concentradas en la proporción de covarianza.

1.4 Sistemas de pronóstico

Los pronósticos se usan en una gran variedad de situaciones; sin embargo, el objetivo final es casi siempre la toma de decisiones. Es importante definir ciertos elementos relacionados con el pronóstico:

- El *Período de pronóstico*: es la unidad básica de tiempo para la cual se realiza el pronóstico.
- El *horizonte de pronóstico*: es el número de períodos en el futuro cubiertos por el pronóstico; y
- El *intervalo de pronóstico*: es la frecuencia con la cual se preparan nuevos pronósticos.

En el caso del presente trabajo, los datos están constituidos por series históricas mensuales, generándose también las proyecciones de manera mensual; así el período de pronóstico es un mes. El horizonte de pronóstico sería de 12 meses (este horizonte puede ser mayor el caso de modelos causales). La frecuencia de

generación de los pronósticos será trimestral de manera que se pueda captar parte del comportamiento estacional en los datos y evitar la influencia de datos atípicos, esto significa que cada tres meses los pronósticos son revisados. También es importante indicar con que frecuencia se revisará el modelo, en este caso la revisión se efectuará de manera anual.

En la actualidad los pronósticos estadístico matemáticos se usan en combinación con los métodos cualitativos. Es cierto que el modelo es solo una aproximación de la realidad y que el comportamiento de las variables económicas depende de determinadas políticas, coyunturas y relaciones con otras variables; efectos estos que quizá no pueden ser asimilados por la serie histórica, el modelo causal, ó no considerados a tiempo. Por estas razones un sistema de pronóstico incluye el juicio de uno o varios expertos que pueden modificar el pronóstico en razón de otra información relevante y su percepción del futuro.

En general los pronósticos pueden ser de los siguientes tipos:

- Dinámicos : Son pronósticos donde el horizonte de pronóstico abarca a más de un período de pronóstico.
- Estáticos : Calcula una secuencia de períodos que se componen cada uno de un período de pronóstico.

Desde otro punto de vista las proyecciones pueden ser puntuales o de tipo intervalo; ya se menciono el problema con las proyecciones por intervalo en

series de tiempo ARIMA, crecen con el horizonte; en el caso del análisis de regresión se llegará a utilizar como se vera más adelante un modelo ARIMA para modelar el comportamiento de los residuos lo cual complica el modelo. Se decide entonces construir pronósticos de tipo intervalo referenciales, promedio o esperado si quiere llamarse en base al estadístico MAPE.

En la siguiente figura se puede apreciar de manera general las etapas en la generación de pronósticos.

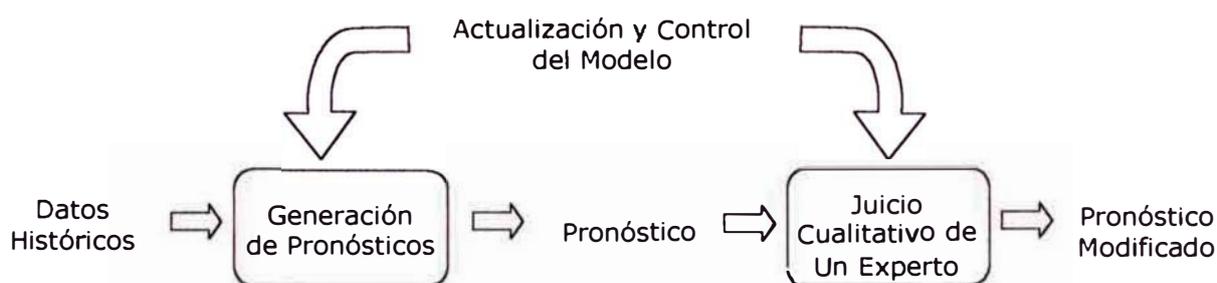


Figura 5
Sistema de Pronóstico

CAPITULO II: OBJETIVOS Y METODOLOGÍA

2.1 Objetivo

El objetivo del presente trabajo de Tesis es la *proyección del valor de las importaciones*. Para este propósito se utilizaron técnicas matemático estadísticas; específicamente se evaluaron dos enfoques: series de tiempo con modelos ARIMA y modelos causales (análisis de regresión), se decidió luego si existía una representación adecuada por parte de estas y se compararon según sus ventajas relativas con el fin de elegir la mejor representación del proceso en estudio.

2.2 Metodología

A continuación se presenta una serie de pasos o etapas a seguirse en la identificación y pronóstico de Modelos ARIMA, y en el ajuste del modelo de regresión.

2.3 Metodología de Series de Tiempo

Análisis Gráfico

El primer paso consiste en la exploración visual de la serie de tiempo; un gráfico lineal de esta permite saber a grandes rasgos si la serie es o no estacionaria, si presenta tendencia, estacionalidad, y varianza constante; para este último punto también es útil el diagrama de cajas.

Otro gráfico muy importante es el correlograma de la serie, si la serie es estacionaria los coeficientes de autocorrelación decrecen rápidamente a cero; caso contrario, un decrecimiento lento indicaría que la serie no es estacionaria. Similar es el caso de la estacionariedad de la componente estacional de la serie, solo que en este caso los coeficientes a analizar son aquellos situados en s , $2s$, $3s$, ... donde s es el período estacional.

Pruebas Estadísticas

El análisis gráfico es un primer paso en el análisis de la serie. Las pruebas estadísticas ayudan a tener más fundamento en las conclusiones del análisis gráfico. Para la prueba de homogeneidad se usará la Prueba de Levene.

Estacionarizar la Serie

Si en base a lo encontrado en los pasos previos se concluye que la serie no es estacionaria; entonces debe previamente estacionarizarse la serie antes de seguir adelante.

Si además, la serie presenta varianza no constante ésta debe estabilizarse mediante transformaciones para este punto pueden usarse las transformaciones sugeridas por el programa SPSS. Si por otra parte, la serie presenta tendencia estocástica esta puede eliminarse tomando diferencias, siendo los ordenes usuales uno o dos; esta operación puede además estabilizar la varianza. En el caso de que la componente estacional no sea estacionaria pueden aplicarse también diferencias estacionales. Si la serie no es estable en medias y en estacionalidad, debe probarse si diferenciando estacionalmente se estabiliza tanto la componente estacional como la media.

Identificación del Modelo

Una vez que la serie ha sido estabilizada puede empezar la etapa de identificación del proceso generador de la serie.

Primero debe dividirse la muestra en dos partes, una será usada para la identificación del modelo ARIMA, y la otra para la evaluación de la capacidad predictiva del modelo. Las herramientas en esta etapa son el Correlograma y el Correlograma Parcial, ambos ayudarán a identificar los ordenes regulares y estacionales del modelo. Así se llegará a uno o varios modelo candidatos, que serán evaluados en la siguiente etapa.

Cuando se este modelando una serie estacional, por ejemplo con orden $P=1$ ó $Q=1$ ó ambos, la notación para los procesos será SAR(12) para la parte autoregresiva y SMA(12) para la parte de medias móviles.

Esta etapa no representa una tarea sencilla y la experiencia del investigador tiene un papel importante.

Elección y Estimación del Modelo

Se empieza estimando los probables modelos generadores de la serie, estos fueron identificados en el punto anterior. La primera evaluación de estos modelos consiste en la significación de sus coeficientes, luego se utilizan criterios de bondad de ajuste como el de Akaike, Schwarz, y el coeficiente de determinación

R^2 . Al final se tendrán uno o varios modelos como probables procesos generadores de la serie.

Sobreajustes

El Modelo elegido es sometido a Sobreajustes; aumentos no simultáneos en una unidad de los ordenes el modelo. Esto tiene su razón en que el modelo elegido no excluye la posibilidad de que otros modelos más complejos puedan ajustar mejor la serie observada. Si un sobreajuste es significativo también se sobreajusta este modelo; este proceso continuara hasta que ningún modelo mejore el ajuste del modelo de partida o hasta que ninguno de los sobreajustes sea significativo.

Validación de Supuestos de los Residuos

Los residuos del modelo deben comportarse como ruido blanco, es decir deben tener media cero, varianza constante, y estar no correlacionados.

El gráfico de línea brinda la primera visión general de los residuos y si éstos estan cumpliendo los supuestos considerados. De manera más formal puede llevarse a cabo pruebas estadísticas de medias y varianzas para apoyar las

conclusiones; de igual forma el Correlograma junto con sus intervalos de confianza permiten saber si la serie de los residuos presenta algún tipo de correlación. Adicionalmente puede probarse la normalidad de los residuos usando una prueba de bondad de ajuste como la de Kolmogorov Smirnov o Jarque Vera.

Si el modelo elegido no cumple las suposiciones respectivas, debe elegirse otro modelo de entre los candidatos o replantearse éste.

Adición de variables

En esta etapa se evalúa la inclusión de variables que puedan ayudar a explicar el comportamiento de la serie. Primero se estacionariza las variables que van a ingresar al modelo, luego se usa la correlación cruzada para determinar con que rezagos entran estas variables en el proceso ARIMA identificado hasta ese momento.

Generación de Pronósticos y Evaluación de la Capacidad Predictiva

Ahora deben generarse los pronósticos para la parte de la muestra que fue asignada a la evaluación de la capacidad predictiva del modelo. El horizonte de pronóstico debe cubrir la longitud de la muestra de evaluación. Una vez generados los pronósticos, estos deben ser evaluados con estadísticas tales

como las U de Thail, MAPE, MAE, y las proporciones de sesgo, varianza y covarianza; si estas medidas son razonables aceptamos el modelo.

Con el modelo aceptado puede efectuarse el pronóstico a futuro fuera de la muestra, suponiendo que el proceso generador de la serie se mantiene; la longitud máxima del horizonte es de 12 meses.

2.4 Metodología de análisis de regresión

Elección de variables a ser usadas como variables explicativas

La primera acción consiste en determinar que variables serán usadas como independientes o predictivas; esta elección, como ya se mencionó con anterioridad, debe apoyarse en fundamentos teóricos, trabajos previos, o la utilización de técnicas como la regresión por pasos para escoger las variables dentro de un conjunto de candidatas.

Análisis Gráfico

En esta etapa se analizan a las variables elegidas en el punto anterior y su relación con la variable dependiente. Se utiliza la gráfica de correlación para investigar la relación entre las variables; y también se utiliza el diagrama de cajas para conocer su distribución.

Ajuste del Modelo

Una vez determinadas las variables a ser usadas como regresoras, se procede a la estimación de modelo de regresión lineal múltiple. Según el resultado de las pruebas de significancia de los coeficientes y el valor del coeficiente de determinación múltiple se puede proceder a la elección de un nuevo modelo, estimarlo y repetir este mismo análisis.

Validación de Supuestos

Primero se verifica que los errores se comporten como ruido blanco; es decir, deben tener media cero, varianza constante, y estar no correlacionados. De manera similar al Análisis de Series de Tiempo la herramienta inicial es el gráfico de línea, el cual brinda una primera visión general de los residuos y si estos están cumpliendo o no los supuestos considerados. Luego puede elegirse un método más objetivo como pruebas de hipótesis para sustentar las conclusiones alcanzadas en la inspección gráfica. Adicionalmente, puede probarse el requisito de normalidad de los residuos usando una prueba de bondad de ajuste como la de Kolmogorov Smirnov o Jarque Vera. En esta etapa puede usarse también el análisis residual para verificar la normalidad y la varianza constante, además de estudiar la posible utilidad que tendría la adición de más términos al modelo.

Si los residuos no se comportan como ruido blanco, es posible que se necesiten más variables explicativas, efectuar alguna transformación en las variables, o que la forma funcional no sea la correcta; si el problema es la correlación pueden modelarse los residuos usando series de tiempo ARIMA.

Otro punto a evaluarse es la presencia de multicolinealidad; como se menciono antes existen varios indicios que pueden indicar la presencia de este problema, los que se presentan a continuación son los que se usan en el presente trabajo.

- La prueba F es significativa, pero ninguna de las pruebas T individuales lo es.
- Coeficientes de correlación altos entre las variables independientes.

Este problema puede corregirse aplicando alguna de las siguientes acciones:

- Incrementar los datos con nuevas observaciones diseñadas de manera específica para romper las dependencias entre las variables.
- Eliminar ciertas variables del modelo
- Usar otros métodos de estimación como la regresión arista.

Si el modelo elegido no cumple las suposiciones respectivas, debe elegirse otro modelo de entre los candidatos o replantearse éste.

Para detectar la presencia de observaciones atípicas influyentes puede usarse la medida de distancia de Cook D_i , un valor de $D_i > 1$ indica que el punto es influyente, o la detección de estos puntos mediante la *exploración gráfica*. Existen dos caminos, en estos casos: no retirar la observación o no tomarla en el análisis, la decisión depende del investigador.

Interpretación del Modelo

Una vez que el modelo final ya ha sido aceptado; puede interpretarse este en función de los coeficientes del modelo y si es necesario utilizar coeficientes estandarizados.

Generación de Pronósticos y Evaluación de la Capacidad Predictiva

Para la proyección a futuro es necesario conocer el comportamiento que tendrán las variables explicativas en ese período. Pueden suceder los siguientes casos, primero conocemos el comportamiento de estas variables o más aún las controlamos; segundo se eligen escenarios para el comportamiento de estas variables; tercero no se conoce el comportamiento de estas variables ni posibles escenarios por lo que se procede a su estimación usualmente con series de

tiempo; finalmente puede elegirse alguna combinación de las situaciones presentadas.

Una vez solucionado el problema de conocer el valor de las variables explicativas, se procede a generar los pronósticos para la parte de la muestra que se eligió para la evaluación de la capacidad predictiva del modelo. Luego de generados los pronósticos se procede a la evaluación de estos en base a las estadísticas ya antes mencionadas: U de Thail, MAPE, MAE, y las proporciones de sesgo, varianza y covarianza.

CAPITULO III: CARACTERIZACIÓN DE LOS MODELOS Y PROYECCIÓN

3.1 Variables

En el año 1992, en el marco del proceso de automatización de la Superintendencia Nacional de ADUANAS (SUNAT a partir de setiembre del 2002), se crea el sistema integrado de gestión aduanera SIGAD, hecho que contribuyo a sistematizar, simplificar y facilitar la operatividad de los regímenes y operaciones aduaneras. De esta forma se cuenta con la fuente para producir y elaborar información estadística de Comercio Exterior.

La información que se utilizó en el Análisis de Series de Tiempo corresponde a las cifras mensuales de importación. Los datos disponibles para este propósito comprenden el período 1985 - Marzo 2003. Sin embargo, el período utilizado para el presente trabajo es el de 1993 - Marzo 2003, no se analizan datos más tempranos por diversos problemas, entre los cuales pueden mencionarse:

- Baja cobertura de las estadísticas
- No se disponía de información oportuna
- Nivel de confiabilidad reducido
- Depuración en forma mecánica y limitada

- Escasos procedimientos de validación

Las importaciones se expresan en valor CIF (véase la Tabla I) (Cost Insurance and Freight), este corresponde al FOB (Free on Board) más el costo del seguro y el flete ; FOB es el valor del producto o mercancía hasta que este es puesto en las bodegas del buque o avión que lo llevará a su destino; este es el valor que se utiliza en el caso de las exportaciones.

Tabla I

PERÚ : IMPORTACIONES MENSUALES											
Período:1993 - 2002											
CIF en Millones de Dólares											
Meses	AÑO										
	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003*
TOTAL	4 024,1	5 434,8	7 583,8	7 774,0	8 341,3	8 042,1	6 717,9	7 344,3	7 252,3	7 486,0	2 072,0
Enero	240,3	374,0	557,0	635,9	697,8	697,3	507,8	584,6	589,5	580,1	717,8
Febrero	297,2	343,4	522,2	540,4	533,6	661,2	484,6	565,7	638,1	545,8	630,4
Marzo	311,1	407,3	662,4	632,6	624,0	709,7	556,6	624,7	605,9	514,4	723,8
Abril	308,3	392,3	564,7	622,8	739,7	760,8	555,1	561,4	583,7	651,5	
Mayo	297,7	457,5	669,2	726,5	680,0	677,4	517,1	640,6	657,1	661,6	
Junio	307,2	408,9	627,9	595,8	640,8	711,3	537,9	621,5	527,7	552,2	
Julio	325,8	423,4	660,9	697,8	719,4	689,0	512,3	604,8	636,8	678,9	
Agosto	359,4	510,5	719,8	690,1	796,4	698,4	567,6	629,8	637,1	624,6	
Setiembre	407,6	496,2	645,5	638,6	750,4	629,6	600,7	561,4	596,1	682,1	
Octubre	339,9	505,0	637,4	727,3	748,1	610,1	591,3	699,4	620,7	661,1	
Noviembre	394,4	563,1	699,6	643,5	710,4	619,5	632,6	647,4	589,0	644,9	
Diciembre	435,2	553,2	617,2	622,9	700,7	577,8	654,3	603,1	570,5	688,7	
NOTA 1: No se incluyen donaciones, servicios diplomáticos y muestras sin valor											
NOTA 2: Las cifras de los años 1993-2001 corresponden a las registradas en los anuarios.											
(*) Cifras Preliminares al 08/04/2003											
Cobertura	: Nacional										
Fuente	: Declaraciones Únicas de Aduanas registradas por el SIGAD										

Para el caso del análisis de regresión se utilizan las siguientes variables predictoras:

Producto bruto interno

Tipo de cambio real multilateral y

Arancel promedio

Las razones de esta elección son las siguientes:

Una de las características más importantes de la economía peruana es la significativa vinculación entre las importaciones y el nivel de actividad económica. Lo anterior se refleja en un alto componente importado del consumo, la producción y la inversión.

De esta manera, cuando se expande el gasto agregado los agentes económicos demandan más bienes y servicios, lo que significa mayores adquisiciones del exterior. Asimismo, las importaciones se relacionan estrechamente con la producción y la acumulación de capital.

Otro aspecto que hay que considerar es la vinculación entre las importaciones y el tipo de cambio real. Este último mide el poder adquisitivo de nuestra moneda y se calcula por medio del índice de paridad del poder de compra que descuenta a la variación nominal del tipo de cambio, la relación entre la tasa de inflación del país y la de nuestros socios comerciales.

Una disminución en el tipo de cambio real implica que nuestros productos resulten menos competitivos en el mercado internacional y que los productos

extranjeros, medidos en moneda nacional, sean más baratos en el mercado interno.

Desde abril 2001 la política arancelaria ha jugado un rol más activo en la economía nacional. La estructura arancelaria que afecta las importaciones fue modificada a tres niveles de 4, 12 y 20 por ciento, registrando un arancel promedio de 11,8 por ciento. Posteriormente, en marzo 2002 el número de tasas aumentó a 4, 7, 12 y 20 por ciento. En la actualidad el arancel promedio simple disminuyó a 11.0 por ciento.

De otro lado, con miras a lograr una participación competitiva dentro de la economía internacional en los próximos años, en el Marco Macroeconómico Multianual (MMM) 2003 – 2005 se contempla efectuar una reducción progresiva del promedio arancelario desde el 11 por ciento actual hasta llegar a 8 por ciento en el 2005, de acuerdo a los compromisos internacionales asumidos. Se esperaría un aumento en las importaciones ante la reducción del nivel arancelario.

En lo que sigue se presentan los datos que se utilizara para las variables mencionadas.

Comportamiento del PBI

Para representar el Producto Bruto Interno se ha elegido a su índice con base 1994; la información que se presenta (véase la Fig. 6) corresponde a los últimos tres años, la información desde el año 1993 puede encontrarse en el Apéndice de datos. El gráfico refleja una tendencia creciente y también un comportamiento estacional. Las cifras del año 2003 a partir de marzo son proyecciones basadas en el comportamiento estacional y un crecimiento esperado de 4%⁵ del PBI para el año 2003.

	2001	2002	PROYECCIÓN
			2003*
Ene	117,4	123,2	128,9
Feb	114,1	117,9	124,2
Mar	122,7	122,2	126,8
Abr	125,9	139,9	145,2
May	136,7	144,0	149,5
Jun	128,2	132,9	138,0
Jul	124,0	131,7	136,7
Ago	123,0	127,8	132,7
Sep	117,8	127,3	132,1
Oct	123,5	129,9	134,8
Nov	122,0	131,4	136,4
Dic	124,1	134,1	139,2
Promedio	123,3	130,2	135,4
Var %		5,6	4,0

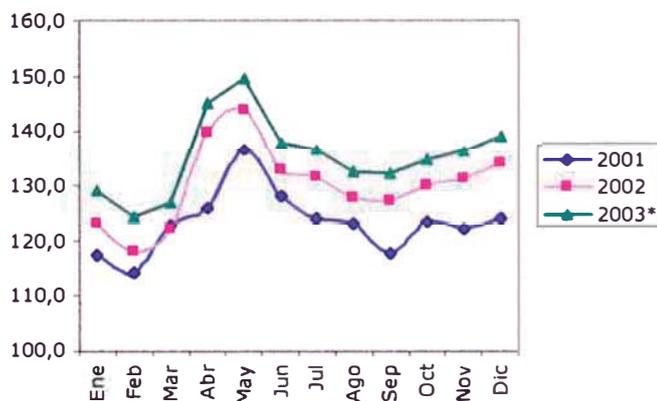


Figura 6
Comportamiento del Índice del PBI, 2001-2003(proyección)

Comportamiento del Arancel

En lo que respecta al arancel promedio las cifras representan el promedio mensual del impuesto de Ad-Valorem (véase Tabla II, página 66) , los números

⁵ Según el reporte de Inflación: Evolución y Perspectivas de Marzo del 2003 del BCRP

en el cuadro son los porcentajes promedio a los que estarían afectos las mercancías. Las cifras de los años 2001 y 2002 son cifras ya dadas mientras que para el 2003 no se esperan, en este momento, modificaciones importantes. Como puede observarse el arancel promedio anual esta disminuyendo.

Tabla II
Comportamiento del Arancel 2001-2003 (Proyección)

	2001	2002	PROYECCIÓN 2003
Ene	11,8	11,8	10,8
Feb	11,8	11,9	10,8
Mar	11,8	11,9	10,8
Abr	11,8	11,0	10,8
May	11,8	11,0	10,8
Jun	11,8	11,0	10,8
Jul	11,8	11,0	10,8
Ago	11,8	11,0	10,8
Sep	11,8	10,9	10,8
Oct	11,8	10,8	10,8
Nov	11,8	10,8	10,8
Dic	11,8	10,8	10,8
Promedio	11,8	11,1	10,8
Var %		-5,5	-3,0

Las cifras de años anteriores pueden encontrarse en el apéndice de datos.

Comportamiento del TCRM

Para el tipo de cambio real multilateral se usa como en el caso del PBI, el índice del tipo de cambio real multilateral (véase Fig. 7, página 67). El comportamiento que ha presentado este índice en los últimos años es de decrecimiento. No se han encontrado mayor información para el escenario que se tendría en el año 2003; pero, según la tendencia se esperaría un decrecimiento; sin embargo, se plantean dos escenarios:

- Decrecimiento, en este caso se utiliza la media geométrica para obtener la tasa de decrecimiento promedio anual y realizar la proyección ó
- Estancamiento no hay crecimiento ni decrecimiento

Las cifras de los tres primeros meses del año 2003 son cifras reales, es con las restantes que se hacen los ajustes.

	1999	2000	2001	2002	PROYECCIÓN 2003	
					DECRECE	ESTANCA
Ene	109,0	111,7	105,8	100,3	100,6	100,6
Feb	110,7	109,2	105,3	99,2	100,1	100,1
Mar	108,8	108,6	103,5	98,4	99,8	99,8
Abr	108,3	109,0	104,3	98,1	96,4	100,0
May	106,9	107,5	105,3	98,5	97,0	100,7
Jun	105,9	108,3	102,6	99,4	96,2	99,8
Jul	104,8	107,4	101,0	101,7	96,5	100,1
Ago	107,0	106,0	102,2	101,7	97,1	100,7
Sep	108,6	105,1	102,2	101,9	97,2	100,9
Oct	111,0	104,7	100,6	99,9	95,5	99,0
Nov	110,6	105,6	100,5	101,0	96,0	99,5
Dic	110,9	105,2	100,6	99,9	95,5	99,0
Promedio	108,5	107,3	102,8	100,0	97,3	100,0
Var %		-1,1	-4,2	-2,7	-2,7	0,0

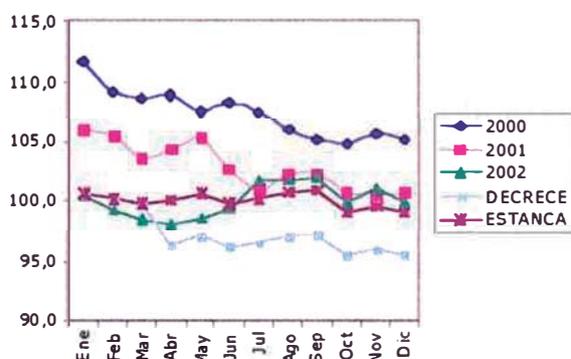


Figura 7
Comportamiento del Índice del Tipo de Cambio Real Multilateral 2001-2003 (Proyección)

Aplicación de la metodología de series de tiempo

Análisis Gráfico

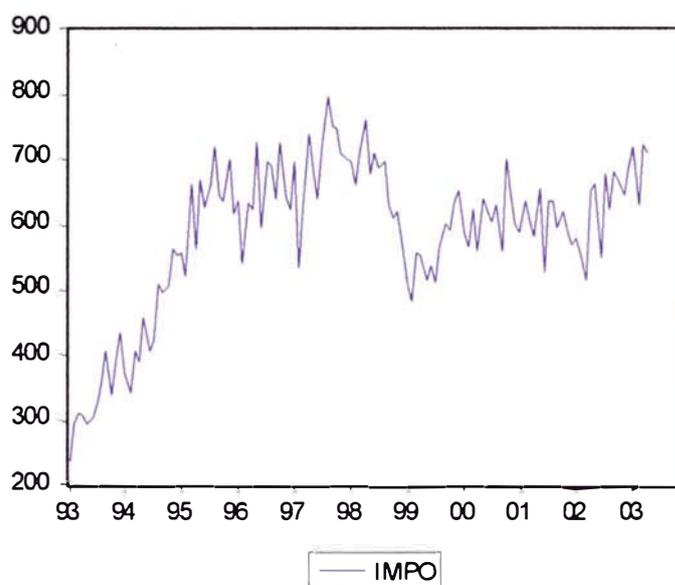


Figura 8
Serie mensual de las Importaciones, 1993 – Marzo 2003 en millones de \$

Este gráfico presenta la evolución mensual de las importaciones del Perú en millones de dólares americanos, desde enero de 1993 hasta marzo del año 2003. Lo primero que se puede observar de la serie es la tendencia creciente al principio para luego caer y tener después un crecimiento lento casi estacionario. Respecto a la homogeneidad de la varianza esta no parece ser problema como se aprecia en el diagrama de cajas (véase Fig. 9, página 69) desde 1993 hasta el año 2002.

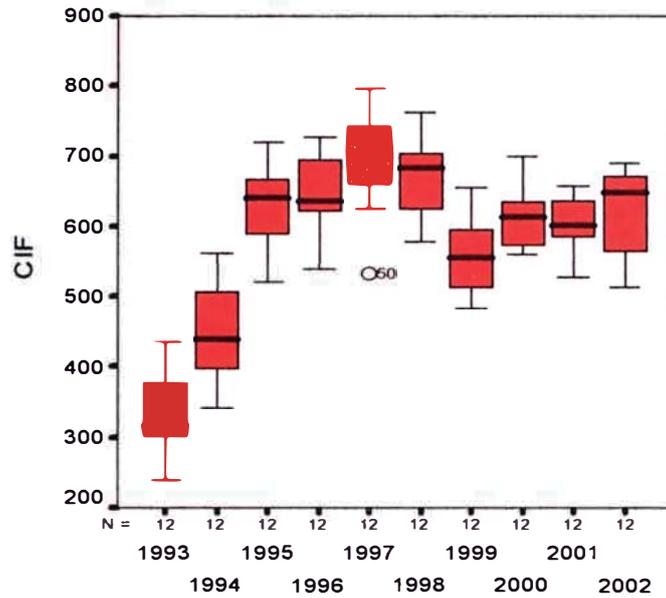
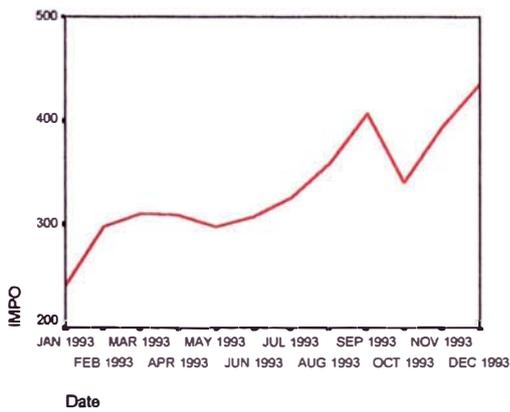
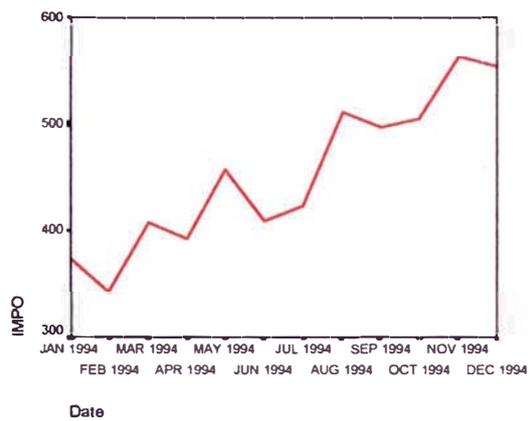


Figura 9
Diagrama de cajas de las importaciones, 1993-2002 en millones de \$

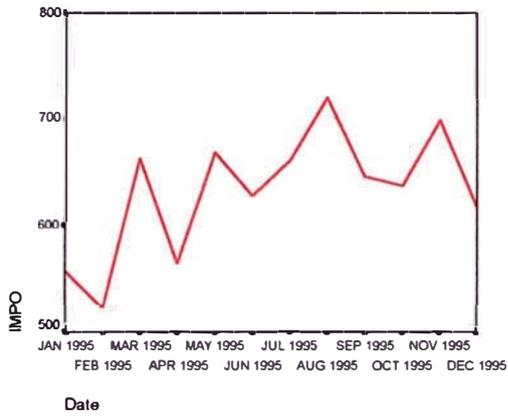
Para detectar la presencia de estacionalidad es útil presentar gráficos individuales para cada año:



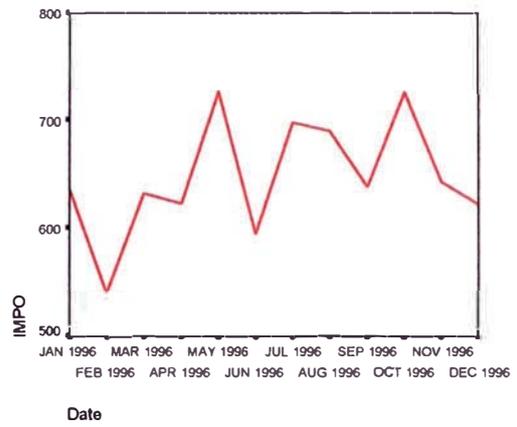
(10a)



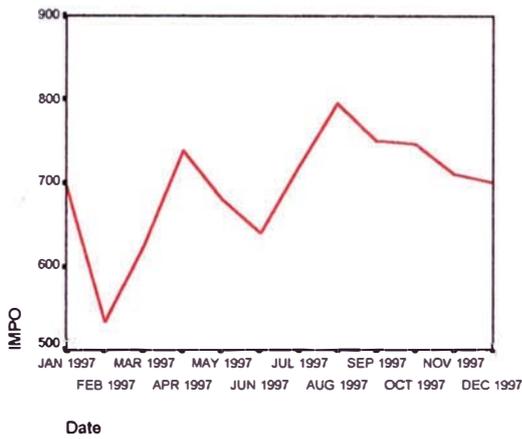
(10b)



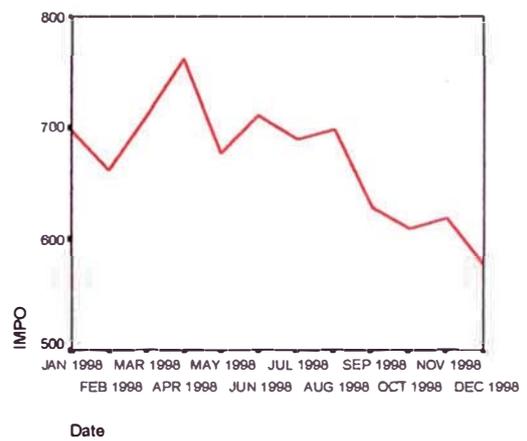
(10c)



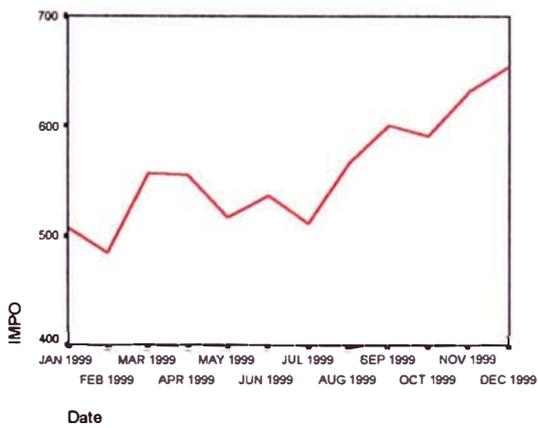
(10d)



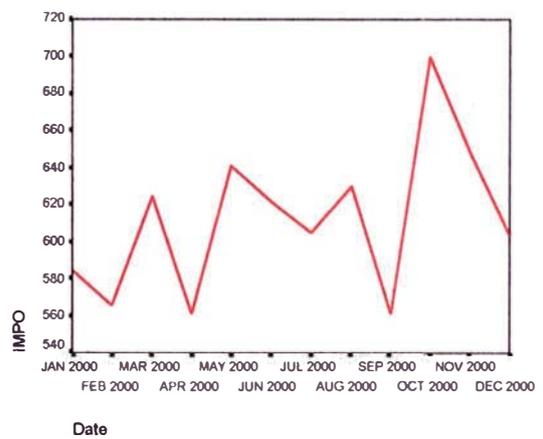
(10e)



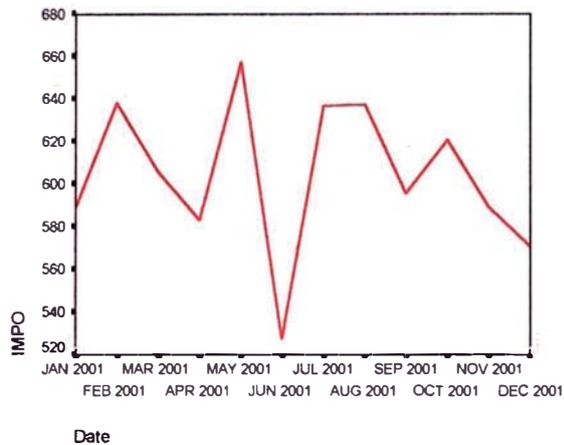
(10f)



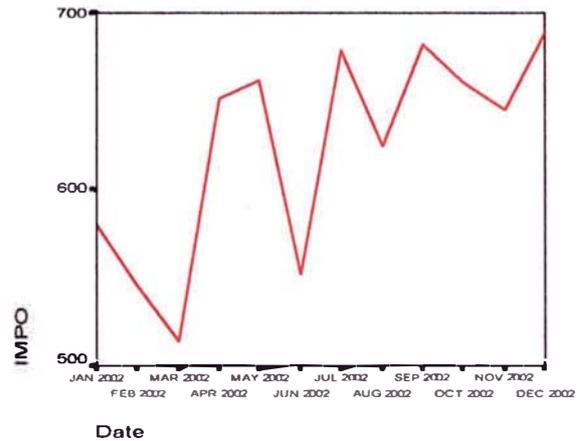
(10g)



(10h)



(10i)



(10j)

Figura 10

Serie mensual de las Importaciones por año, 1993 - 2002 en millones de \$

En la mayor parte de los gráficos anuales puede observarse dos picos en el primer semestre y otros dos en el segundo, estos meses corresponden aproximadamente a los meses de Mayo, Agosto, y Noviembre. Al igual que la tendencia, esta componente estacional detectada no se repite de manera exacta año tras año.

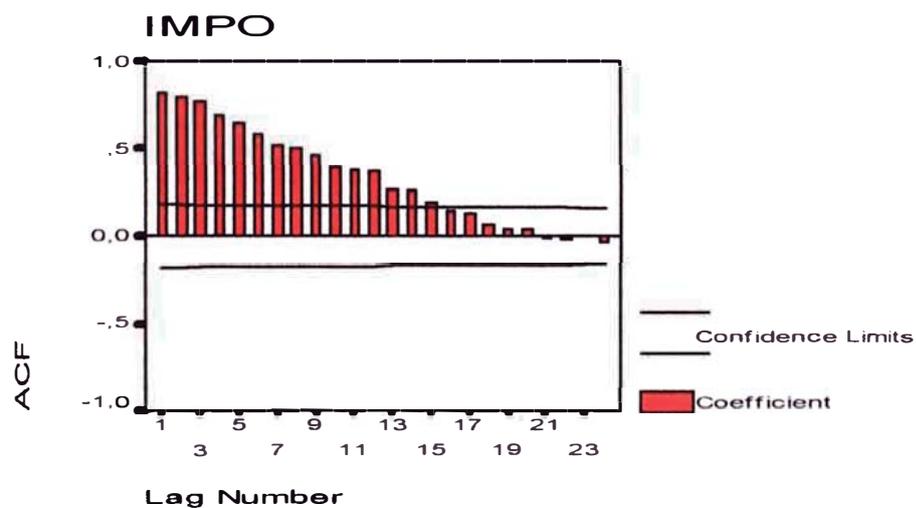


Figura 11
Correlograma de las importaciones

El correlograma (véase Fig. 11, página 71) tiene un decrecimiento lento hacia cero lo que constata que la serie no es estacionaria; por otra parte la componente estacional parece ser estacionaria como lo refleja el correlograma en los múltiplos del período estacional $s = 12$ (véase la Fig. 12).

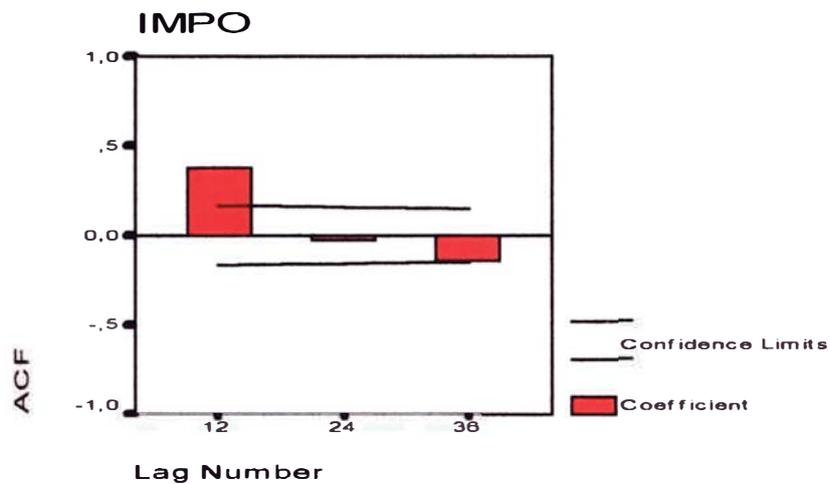


Figura 12
Correlograma de las importaciones, parte estacional

Pruebas Estadísticas

Prueba de Homogeneidad

Se efectuó la Prueba de Levene de Homogeneidad de Varianzas; para esto se agruparon las observaciones por año, no se tomó el año 2003 por tener solo cuatro casos. Los resultados fueron los siguientes:

Tabla III
Prueba de Homogeneidad de Levene

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
IMPO	Based on Mean	1,067	9	110	,393
	Based on Median	,786	9	110	,630
	Based on Median and with adjusted df	,786	9	93,153	,630
	Based on trimmed mean	1,019	9	110	,430

El Valor P (sig. en el cuadro) es mayor que 0,05 por lo que no se rechaza la hipótesis nula de igualdad de varianzas; esto confirma lo que se observó en el diagrama de cajas.

Estacionarizar la serie

Sobre la base de lo encontrado en los puntos anteriores se procedió a tomar una diferencia regular a la serie (véase Fig. 13, página 74).

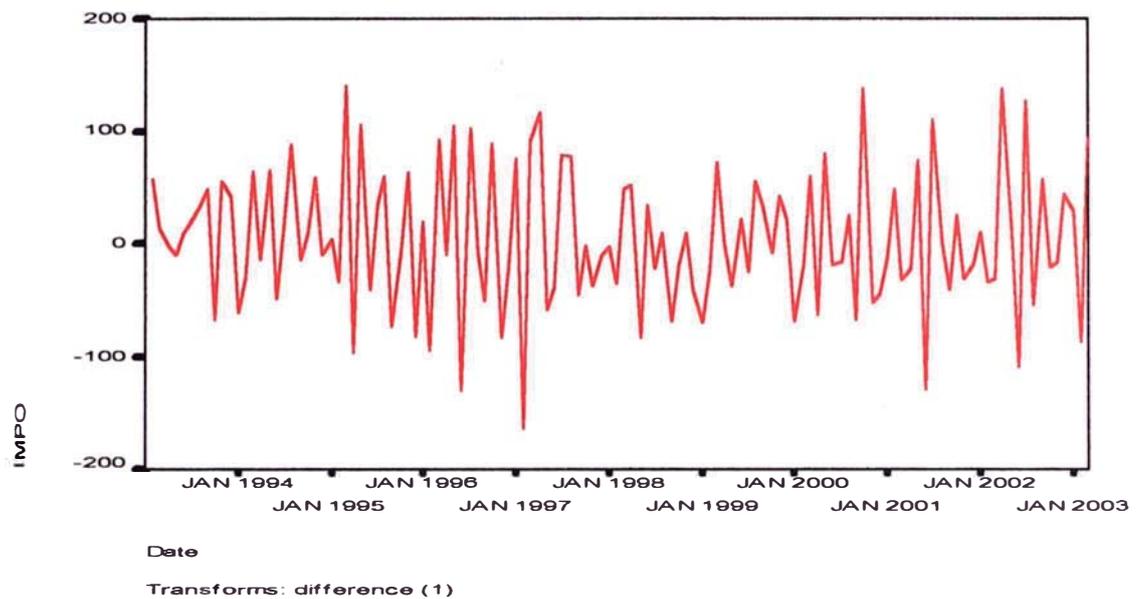


Figura 13
Serie de tiempo de las importaciones con una diferencia

Puede apreciarse ahora que la serie tiene un comportamiento estacionario; respecto a la varianza, esta es constante como ya se ha comprobado, y además no es necesario tomar diferencias estacionales ya que como se vio al principio la serie es estacionaria en esta componente.

Se presenta ahora el correlograma de la serie diferenciada para confirmar que esta se ha estacionarizado (véase Fig. 14, página 75).

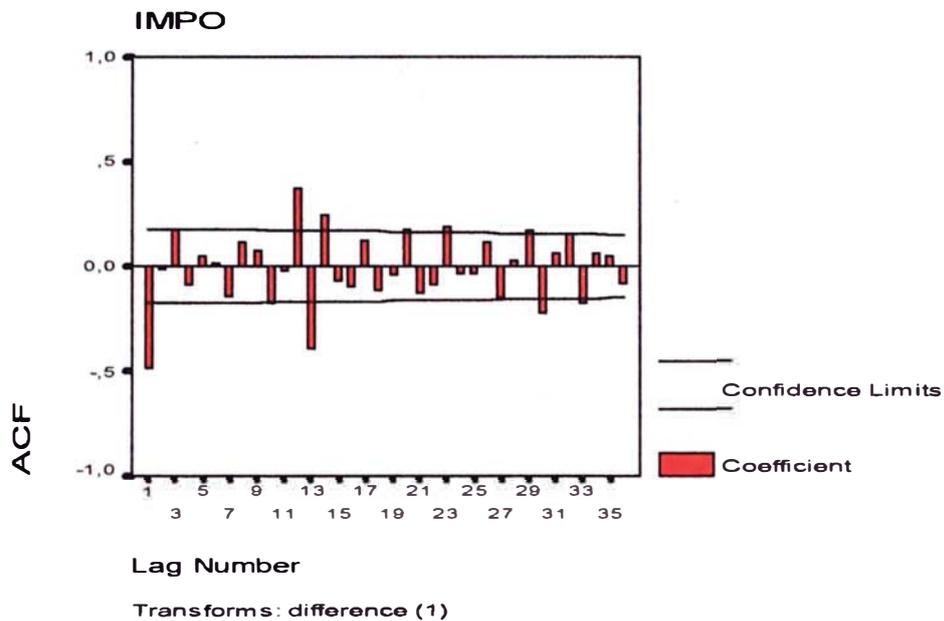


Figura 14
Correlograma de las importaciones con una diferencia

La serie es estacionaria pues tiende rápidamente a cero, el alto coeficiente de correlación en el rezago 12, debe estar reflejando algún orden estacional; mientras que el observado en el punto 13 podría ser más bien consecuencia de los errores muestrales, así como los demás valores subsecuentes que en algunos casos pasan ligeramente los límites de confianza al 95%. Si las autocorrelaciones son significativamente distintas de cero al principio estas pueden tomarse en cuenta; mientras que si estas se presentan más adelante, es muy probable que sean consecuencia de los errores debido a la muestra.

El siguiente gráfico es el correlograma de la componente estacional, la cual como en el correlograma de la parte regular tiene un descenso rápido hacia cero.

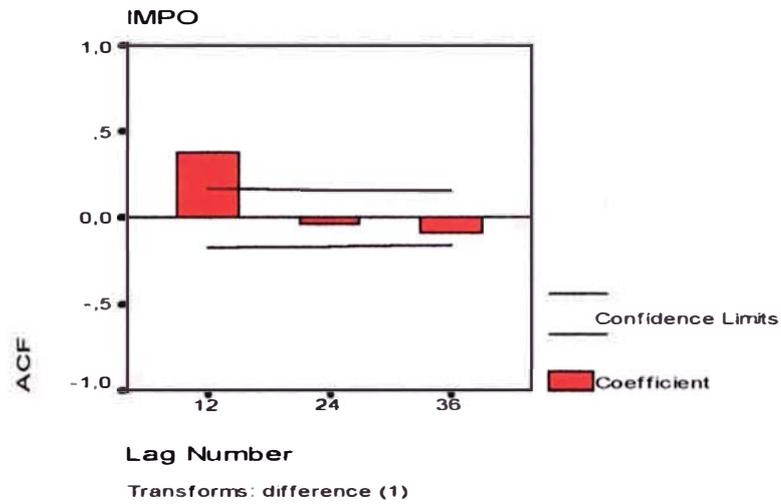


Figura 15
Correlograma de la parte estacional con una diferencia

Podemos concluir por tanto que la serie ha sido estabilizada.

Identificación del Modelo

Para la identificación del Proceso generador de la serie se tomó el siguiente rango muestral: 1993-2001, el período 2002 se usará para medir la capacidad predictiva del modelo.

El correlograma y correlograma parcial son los siguientes:

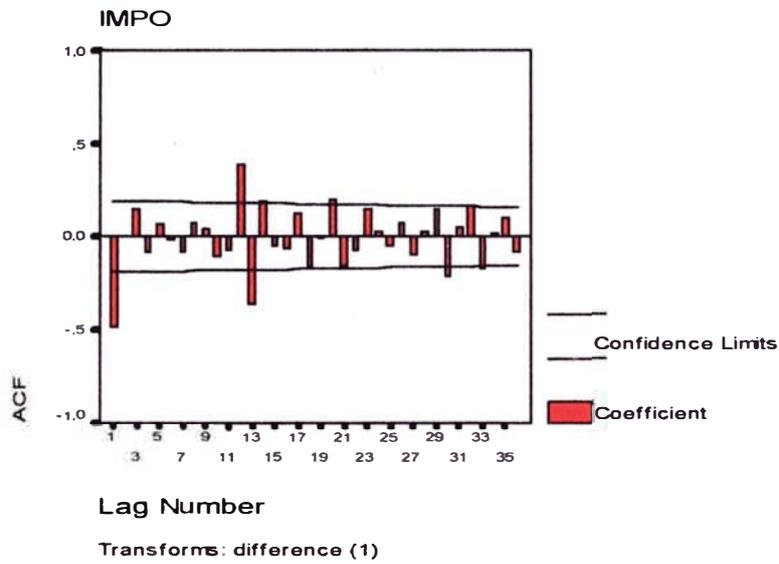


Figura 16
Correlograma de las importaciones con una diferencia

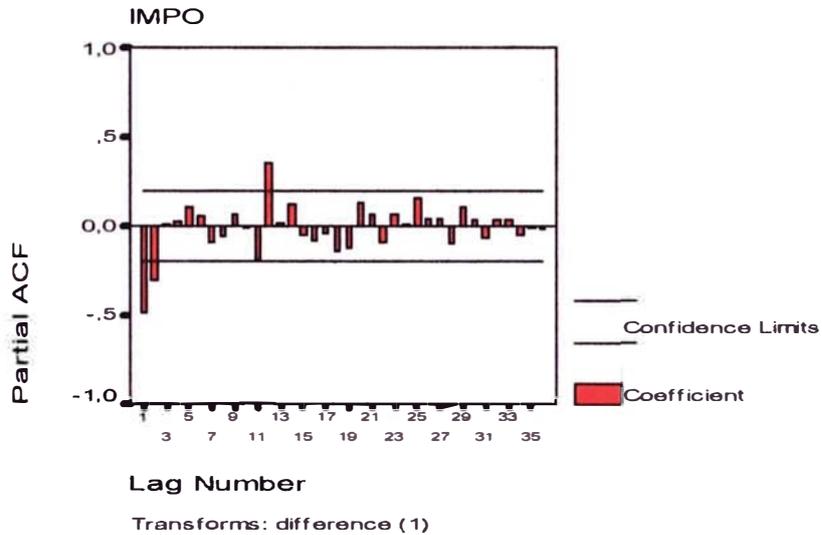


Figura 17
Correlograma parcial de las importaciones con una diferencia

Estos gráficos puede estar sugiriendo tres posibles modelos (sin considerar los rezagos en la parte estacional) ; primero un modelo MA(1) ya que se tiene el primer coeficiente distinto de cero en el Correlograma y un descenso rápido y

progresivo hacia cero en el Correlograma Parcial; también podría decirse que estamos frente a un proceso AR(2) porque los dos primeros coeficientes son distintos de cero en el Correlograma Parcial y se observa un decrecimiento rápido hacia cero en el Correlograma; finalmente se puede optar por un modelo ARMA(2,1) es decir una combinación de los modelos anteriores, esto para agrupar los dos posibles modelos que se han mencionado y porque el comportamiento de los correlogramas en un modelo ARMA son combinación de los respectivos en el modelo AR y el MA.

A continuación tenemos los correlogramas para la parte estacional:

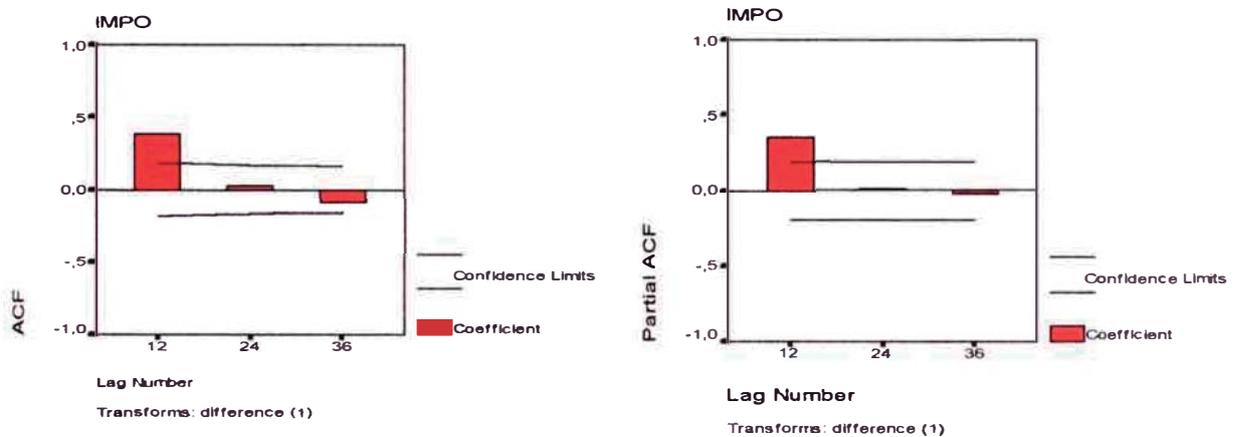


Figura 18
Correlogramas de la parte estacional de la serie de Importaciones con una diferencia

Como en el caso anterior aquí también existen varias posibilidades; puede ser un MAS(1) o un ARS(1) o una combinación de ambos. La componente MAS(1) quiere decir que $Q=1$ en el modelo estacional multiplicativo, y el ARS(1) implica que $P=1$. En ambos casos el período es $s=12$.

Elección y estimación del Modelo

Combinando los modelos para la parte regular y estacional, en la que se considera la combinación de los candidatos, se obtienen:

ARIMA_s (0,1,1,1,0,1)

ARIMA_s (0,1,1,1,0,0)

ARIMA_s (0,1,1,0,0,1)

ARIMA_s (2,1,0,1,0,1)

ARIMA_s (2,1,0,1,0,0)

ARIMA_s (2,1,0,0,0,1)

ARIMA_s (2,1,1,1,0,1)

ARIMA_s (2,1,1,1,0,0)

ARIMA_s (2,1,1,0,0,1)

Se estima primero el modelo ARIMA_s (0,1,1,1,0,1) con constante, al ajustar este modelo resulta que la constante es la menos significativa; luego al volver a estimar el modelo sin la constante sucede que la componente estacional de

medias móviles tampoco es significativa lo que finalmente resulta en el siguiente modelo

Dependent Variable: D(IMPO)

Method: Least Squares

Date: 05/11/03 Time: 20:11

Sample(adjusted): 1994:02 2001:12

Included observations: 95 after adjusting endpoints

Convergence achieved after 11 iterations

Backcast: 1994:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	0,399518	0,096330	4,147402	0,0001
MA(1)	-0,503153	0,089836	-5,600787	0,0000
R-squared	0,389342	Mean dependent var		2,067943
Adjusted R-squared	0,382776	S.D. dependent var		62,56660
S.E. of regression	49,15457	Akaike info criterion		10,64864
Sum squared resid	224704,0	Schwarz criterion		10,70241
Log likelihood	-503,8106	F-statistic		59,29484
Durbin-Watson stat	2,176911	Prob(F-statistic)		0,000000
Inverted AR Roots	,93	,80+,46i	,80 -,46i	,46+,80i
	,46 -,80i	,00+,93i	-,00 -,93i	-,46 -,80i
	-,46+,80i	-,80 -,46i	-,80+,46i	-,93
Inverted MA Roots	,50			

Es el turno del modelo $ARIMA_s(0,1,1,1,0,0)$, el cual por coincidencia acaba de ser estimado luego de retirar la constante y la componente $ma(1)$ en el modelo anterior.

El siguiente modelo a estimarse es $ARIMA_s(0,1,1,0,0,1)$, y como en los anteriores casos la constante no es significativa.

Dependent Variable: D(IMPO)
 Method: Least Squares
 Date: 05/11/03 Time: 19:02
 Sample(adjusted): 1993:02 2001:12
 Included observations: 107 after adjusting endpoints
 Convergence achieved after 8 iterations
 Backcast: 1992:01 1993:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0,485521	0,081570	-5,952191	0,0000
SMA(12)	0,448626	0,083272	5,387449	0,0000
R-squared	0,388692	Mean dependent var	3,085447	
Adjusted R-squared	0,382870	S,D, dependent var	60,48026	
S.E. of regression	47,51185	Akaike info criterion	10,57835	
Sum squared resid	237024,4	Schwarz criterion	10,62831	
Log likelihood	-563,9417	F-statistic	66,76286	
Durbin-Watson stat	2,106310	Prob(F-statistic)	0,000000	
Inverted MA Roots	,90+,24i	,90 -,24i	,66 -,66i	
			,66+,66i	
	,49	,24+,90i	,24 -,90i	-
			,24+,90i	
	-,24 -,90i	-,66+,66i	-,66+,66i	-
			,90+,24i	
	-,90 -,24i			

La estimación para el segundo grupo de candidatos sin considerar la constante que resulta no significativa es la siguiente:

Dependent Variable: D(IMPO)
 Method: Least Squares
 Date: 05/11/03 Time: 19:06
 Sample(adjusted): 1994:04 2001:12

Included observations: 93 after adjusting endpoints

Convergence achieved after 25 iterations

Backcast: 1993:04 1994:03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0,675151	0,078017	-8,653891	0,0000
AR(2)	-0,373085	0,072156	-5,170523	0,0000
SAR(12)	0,807405	0,037152	21,73266	0,0000
MA(12)	-0,885772	0,000161	-5504,568	0,0000
R-squared	0,581446	Mean dependent var		1,754573
Adjusted R-squared	0,567338	S.D. dependent var		62,82054
S,E, of regression	41,32151	Akaike info criterion		10,32270
Sum squared resid	151964,6	Schwarz criterion		10,43163
Log likelihood	-476,0056	F-statistic		41,21232
Durbin-Watson stat	1,933284	Prob(F-statistic)		0,000000
Inverted AR Roots	,98	,85+,49i	,85 -,49i	,49+,85i
	,49 -,85i	-,00 -,98i	-,00+,98i	,34+,51i
	-,34 -,51i	-,49 -,85i	-,49+,85i	,85+,49i
	-,85 -,49i	-,98		
Inverted MA Roots	,99	,86 -,49i	,86+,49i	,49 -
	,49+,86i	,00+,99i	-,00 -,99i	,86i
	-,49 -,86i	-,86+,49i	-,86 -,49i	-,99

Dependent Variable: D(IMPO)

Method: Least Squares

Date: 05/11/03 Time: 19:08

Sample(adjusted): 1994:04 2001:12

Included observations: 93 after adjusting endpoints

Convergence achieved after 6 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0,620633	0,100468	-6,177438	0,0000
AR(2)	-0,306596	0,099573	-3,079105	0,0028
SAR(12)	0,411983	0,097289	4,234611	0,0001
R-squared	0,427483	Mean dependent var		1,754573
Adjusted R-squared	0,414760	S.D. dependent var		62,82054
S.E. of regression	48,05833	Akaike info criterion		10,61443
Sum squared resid	207864,3	Schwarz criterion		10,69613
Log likelihood	-490,5712	F-statistic		33,60023
Durbin-Watson stat	1,958153	Prob(F-statistic)		0,000000
Inverted AR Roots	,93	,80+,46i	,80 -,46i	
				,46+,80i
	,46 -,80i	,00+,93i	,00 -,93i	-
				,31+,46i
	-,31 -,46i	-,46+,80i	-,46 -,80i	-,80 -
				,46i
	-,80+,46i	-,93		

Dependent Variable: D(IMPO)

Method: Least Squares

Date: 05/11/03 Time: 19:09

Sample(adjusted): 1993:04 2001:12

Included observations: 105 after adjusting endpoints

Convergence achieved after 6 iterations

Backcast: 1992:04 1993:03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0,575634	0,094444	-6,094951	0,0000
AR(2)	-0,292297	0,093933	-3,111750	0,0024
MA(12)	0,414085	0,095269	4,346473	0,0000
R-squared	0,410174	Mean dependent var		2,470334
Adjusted R-squared	0,398608	S,D, dependent var		60,81843
S.E. of regresión	47,16435	Akaike info criterion		10,57331
Sum squared resid	226896,6	Schwarz criterion		10,64914

Log likelihood	-552,0987	F-statistic	35,46612
Durbin-Watson stat	1,975473	Prob(F-statistic)	0,000000
Inverted AR Roots	-,29 -,46i	-,29+,46i	
Inverted MA Roots	,90+,24i	,90 -,24i	,66 -,66i
			,66+,66i
	,24+,90i	,24 -,90i	-,24+,90i
			-,24 -,90i
	-,66+,66i	-,66+,66i	-,90 -,24i
			-,90+,24i

La estimación de los últimos tres modelos es la siguiente:

Al estimar el modelo $ARIMA_s(2,1,1,1,0,1)$ resulta que la componente $ma(1)$ no es significativa; al ser retirada y estimarse el modelo nuevamente la constante tiene un valor de P de 0,0428 y es también retirada del modelo; en consecuencia el modelo resultante es $ARIMA_s(2,1,0,1,0,1)$ el cual ya ha sido estimado antes.

Luego al estimar $ARIMA_s(2,1,1,1,0,0)$, la constante no es significativa como en los casos anteriores; al ser esta retirada y estimarse el modelo nuevamente resulta que la componente $ma(1)$ tampoco es significativa; el modelo final es $ARIMAs(2,1,0,1,0,0)$ el cual también ya ha sido estimado con anterioridad.

Finalmente el último candidato es el modelo $ARIMA_s(2,1,1,0,0,1)$, al estimar este y volverse a estimar el modelo la componente $ma(1)$ es no significativa, al ser esta retirada la constante corre igual suerte, lo que finalmente conduce a un modelo ya analizado antes: $ARIMAs(2,1,0,0,0,1)$

La siguiente tabla presenta los modelos que finalmente han quedado como candidatos; puede verse que el primer modelo tiene el mayor R^2 ajustado y los menores valores de Akaike y Schwarz; es por lo tanto el modelo elegido.

Tabla IV
Comparación de Modelos ARIMA

Modelos	R^2 ajustado	Akaike	Schwarz
ARIMA _s (2,1,0,1,0,1)	0,567	10,323	10,432
ARIMA _s (0,1,1,0,0,1)	0,383	10,578	10,628
ARIMA _s (2,1,0,0,0,1)	0,399	10,573	10,649
ARIMA _s (2,1,0,1,0,0)	0,415	10,614	10,696
ARIMA _s (0,1,1,1,0,0)	0,383	10,649	10,702

La expresión matemática del modelo elegido, ARIMA_s(2,1,0,1,0,1), es:

$$(1-0,800B^{12})(1+0,623B+0,371B^2)(1-B) Y_t = (1-0,886B^{12})\varepsilon_t$$

Sobreajustes

Se procede ahora a efectuar los correspondientes sobreajustes al modelo ARIMA_s(2,1,0,1,0,1), estos son:

ARIMA_s(3,1,0,1,0,1)

ARIMA_s(2,1,1,1,0,1)

ARIMA_s(2,1,0,2,0,1)

ARIMA_s(2,1,0,1,0,2)

Se añade un ar(3)

Dependent Variable: D(IMPO)

Method: Least Squares

Date: 05/14/03 Time: 21:27

Sample(adjusted): 1994:05 2001:12

Included observations: 92 after adjusting endpoints

Convergence achieved after 21 iterations

Backcast: 1993:05 1994:04

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0,643091	0,084658	-7,596321	0,0000
AR(2)	-0,304178	0,098407	-3,091036	0,0027
AR(3)	0,092916	0,078454	1,184335	0,2395
SAR(12)	0,809196	0,037043	21,84505	0,0000
MA(12)	-0,885836	0,000162	-5456,139	0,0000
R-squared	0,585288	Mean dependent var	1,936989	
Adjusted R-squared	0,566221	S.D. dependent var	63,13999	
S.E. of regression	41,58520	Akaike info criterion	10,34618	
Sum squared resid	150451,6	Schwarz criterion	10,48323	
Log likelihood	-470,9243	F-statistic	30,69606	
Durban-Watson stat	2,025447	Prob(F-statistic)	0,000000	
Inverted AR Roots	,98	,85+,49i	,85 -,49i	,49 - ,85i
	,49+,85i	,20	,00+,98i	-,00 - ,98i
	-,42 -,54i	-,42+,54i	-,49 -,85i	- ,49+,85i
	-,85+,49i	-,85 -,49i	-,98	
Inverted MA Roots	,99	,86+,49i	,86 -,49i	,49+,86i
	,49 -,86i	-,00 -,99i	-,00+,99i	-,49 - ,86i
	-,49+,86i	-,86+,49i	-,86 -,49i	-,99

El término ar(3) no es significativo.

En el segundo modelo se añade un ma(1), pero esta componente no es significativa como ya se ha visto antes (1er modelo del tercer grupo de modelos evaluados).

Para el tercer modelo se aumenta una componente estacional P =2. (sar=24)

Dependent Variable: D(IMPO)

Method: Least Squares

Date: 05/14/03 Time: 21:29

Sample(adjusted): 1995:04 2001:12

Included observations: 81 after adjusting endpoints

Convergence achieved after 17 iterations

Backcast: 1994:04 1995:03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0,651420	0,085713	-7,599989	0,0000
AR(2)	-0,442648	0,084598	-5,232389	0,0000
SAR(12)	0,815821	0,091367	8,929054	0,0000
SAR(24)	-0,114860	0,090520	-1,268890	0,2084
MA(12)	-0,885804	0,000151	-5863,047	0,0000
R-squared	0,638422	Mean dependent var		-
				1,135249
Adjusted R-squared	0,619392	S.D. dependent var		63,63702
S.E. of regression	39,25986	Akaike info criterion		10,23802
Sum squared resid	117141,6	Schwarz criterion		10,38583
Log likelihood	-409,6399	F-statistic		33,54751
Durbin-Watson stat	2,028567	Prob(F-statistic)		0,000000
Inverted AR Roots	,96	,87	,83 - ,48i	
				,83+ ,48i
	,75+ ,43i	,75 - ,43i	,48 - ,83i	

				,48+,83i
	,43 -,75i	,43+,75i	,00+,96i	,00 -
				,96i
	-,00+,87i	-,00 -,87i	-,33+,58i	-,33 -
				,58i
	-,43+,75i	-,43 -,75i	-,48+,83i	-,48 -
				,83i
	-,75+,43i	-,75 -,43i	-,83+,48i	-,83 -
				,48i
	-,87	-,96		
Inverted MA Roots	,99	,86+,49i	,86 -,49i	,49 -
				,86i
	,49+,86i	-,00 -,99i	-,00+,99i	-,49 -
				,86i
	-,49+,86i	-,86+,49i	-,86 -,49i	-,99

Esta componente tampoco es significativa.

Por último en el cuarto sobreajuste se añade la componente estacional $Q=2$ (sma=24), esta componentes resulta ser significativa pero al comparar las medidas de ajuste de este nuevo modelo con el elegido en la etapa anterior, resulta que éste sigue siendo el mejor.

Dependent Variable: D(IMPO)

Method: Least Squares

Date: 05/14/03 Time: 21:34

Sample(adjusted): 1994:04 2001:12

Included observations: 93 after adjusting endpoints

Convergence achieved after 11 iterations

Backcast: 1992:04 1994:03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0,601694	0,088435	-6,803830	0,0000
AR(2)	-0,309943	0,083031	-3,732886	0,0003

SAR(12)	0,743481	0,080015	9,291746	0,0000
MA(12)	-0,534930	0,119893	-4,461740	0,0000
MA(24)	-0,310777	0,112804	-2,755019	0,0071
R-squared	0,542256	Mean dependent var	1,754573	
Adjusted R-squared	0,521450	S.D. dependent var	62,82054	
S.E. of regression	43,45758	Akaike info criterion	10,43371	
Sum squared resid	166193,4	Schwarz criterion	10,56987	
Log likelihood	-480,1676	F-statistic	26,06181	
Durbin-Watson stat	1,987269	Prob(F-statistic)	0,000000	
Inverted AR Roots	,98	,84+,49i	,84 -,49i	,49+,84i
	,49 -,84i	-,00 -,98i	-,00+,98i	-
				,30+,47i
	-,30 -,47i	-,49 -,84i	-,49+,84i	-
				,84+,49i
	-,84 -,49i	-,98		
Inverted MA Roots	,99	,89+,24i	,89 -,24i	,86 -
				,49i
	,86+,49i	,65+,65i	,65 -,65i	,49 -
				,86i
	,49+,86i	,24+,89i	,24 -,89i	,00 -
				,99i
	,00+,99i	-,24 -,89i	-,24+,89i	-
				,49+,86i
	-,49 -,86i	-,65+,65i	-,65+,65i	-,86 -
				,49i
	-,86+,49i	-,89 -,24i	-,89+,24i	-,99

El modelo elegido hasta este momento es el que se obtuvo en la etapa de estimación.

El gráfico siguiente muestra el ajuste entre la serie diferenciada y el modelo elegido.

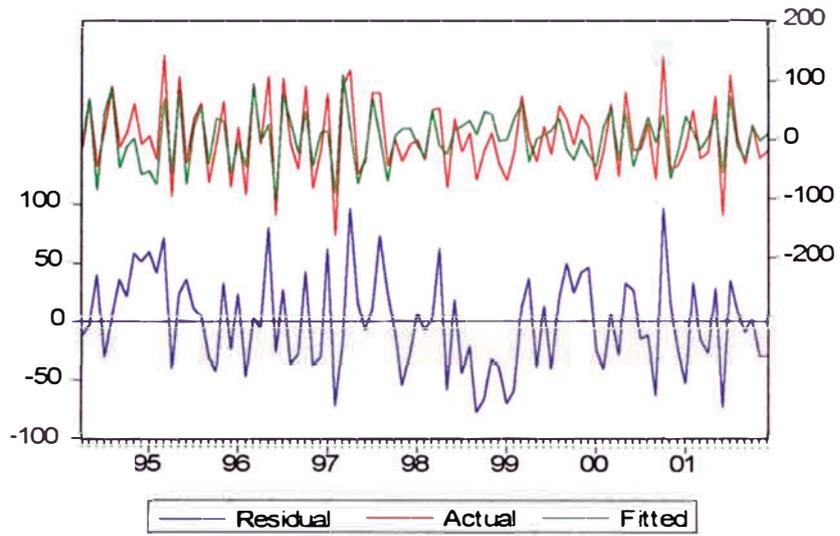


Figura 19

Serie mensual diferenciada de las importaciones, valores Ajustados y Residuos

Actual en el gráfico representa a la serie con una diferencia regular

Validación de supuestos de los residuos:

El siguiente es el gráfico de línea de los residuos :

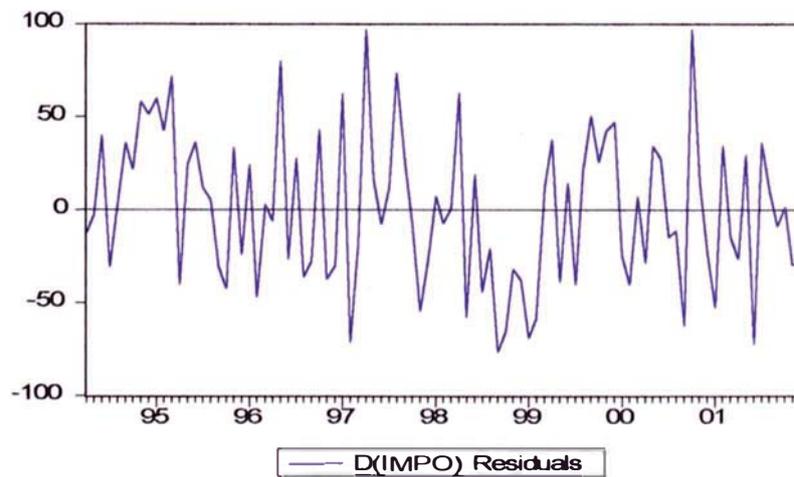


Figura 20 , residuos

El promedio de los residuos en este gráfico es de $-0,29$ cerca del supuesto de media cero para el ruido blanco. La varianza puede tomarse en general como constante según este gráfico

Los gráficos de autocorrelación y auto correlación parcial (véase la Fig. 21) muestran que los residuos no están autocorrelacionados.

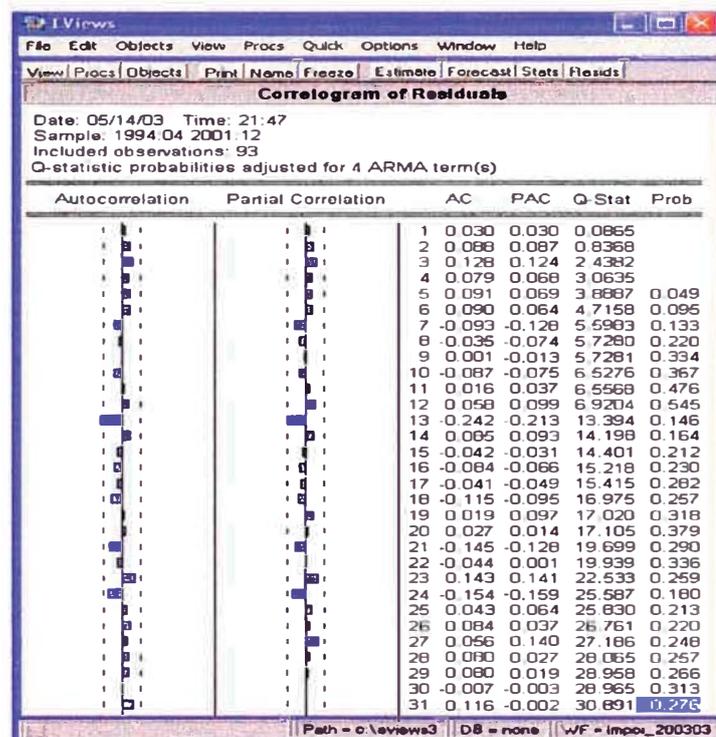


Figura 21
Correlograma de los residuos

Sin embargo, el correlograma muestra la posible presencia de una componente MA(13) O AR(13) en el modelo, sin los términos intermedios; si incluimos la primera componente en el modelo resulta que no es significativa; pero, la componente AR(13) si lo es y mejora el ajuste del modelo.

Dependent Variable: D(IMPO)

Method: Least Squares

Date: 05/14/03 Time: 21:52

Simple(adjusted): 1995:03 2001:12

Included observations: 82 after adjusting endpoints

Convergence achieved after 19 iterations

Backcast: 1994:03 1995:02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0,636501	0,099085	-6,423779	0,0000
AR(2)	-0,363411	0,083544	-4,349932	0,0000
AR(13)	-0,208159	0,064570	-3,223768	0,0019
SAR(12)	0,757649	0,036552	20,72790	0,0000
MA(12)	-0,885788	0,000164	-5387,735	0,0000
R-squared	0,658776	Mean dependent var		0,588404
Adjusted R-squared	0,641050	S.D. dependent var		65,14058
S.E. of regression	39,02733	Akaike info criterion		10,22544
Sum squared resid	117281,2	Schwarz criterion		10,37219
Log likelihood	-414,2430	F-statistic		37,16449
Durbin-Watson stat	2,108434	Prob(F-statistic)		0,000000
Inverted AR Roots	,98	,85+,49i	,85 -,49i	
				,81+,21i
	,81 -,21i	,61+,59i	,61 -,59i	,49 -
				,85i
	,49+,85i	,25 -,83i	,25+,83i	
				,00+,98i
	-,00 -,98i	-,17 -,90i	-,17+,90i	-,49 -
				,85i
	-,49+,85i	-,55+,76i	-,55 -,76i	-
				,81+,43i
	-,81 -,43i	-,85 -,49i	-,85+,49i	-,91
	-,98			
Inverted MA Roots	,99	,86 -,49i	,86+,49i	,49 -
				,86i
	,49+,86i	,00+,99i	-,00 -,99i	-
				,49+,86i
	-,49 -,86i	-,86+,49i	-,86 -,49i	-,99

Sin embargo, es probable que la presencia de esta componente se deba más bien a errores muestrales, por lo que se prefiere no admitirla en el modelo - al probar la inclusión de esta componente, es cierto que mejora el ajuste del modelo; pero, los residuos resultan tener una media de -1,7 alejada del valor de cero esperado.

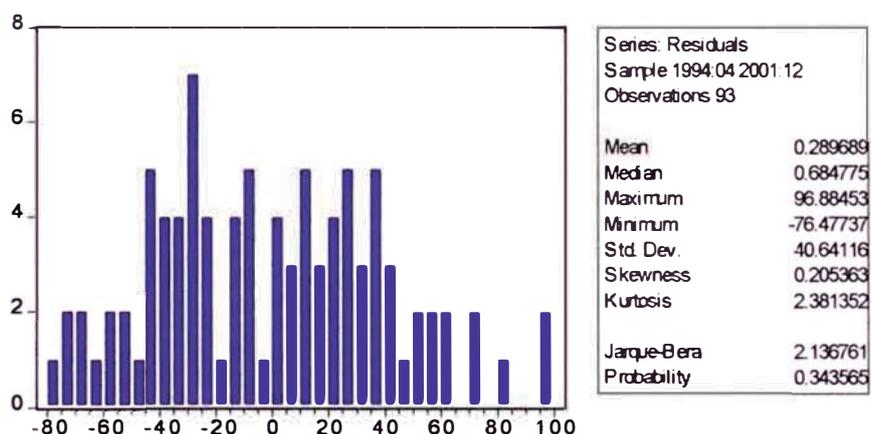


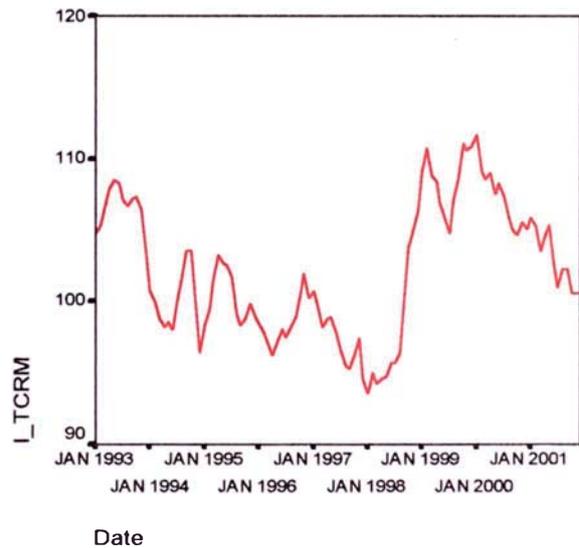
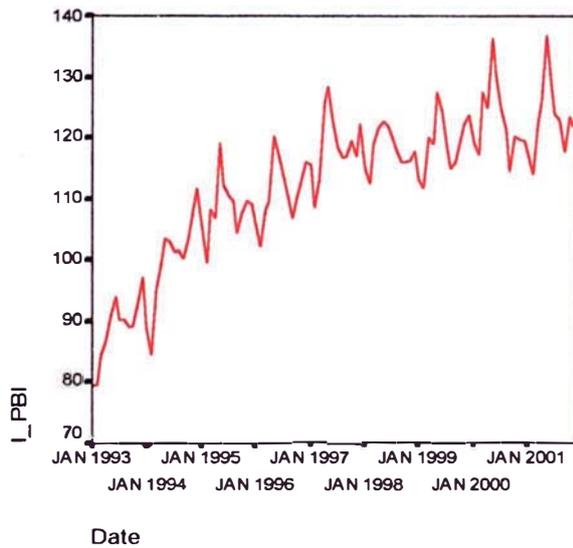
Figura 22
Histograma y estadísticas descriptivas de los residuos

No se rechaza la normalidad, y la media tiene un valor cercano a cero (véase la Fig. 22). El ajuste y el comportamiento de los residuos puede mejorar al incluir variables explicativas adicionales en el modelo, como se verá inmediatamente.

Adición de variables

Las posibles variables que pueden ayudar en la explicación del comportamiento de la serie de tiempo de las importaciones son:

- El Producto Bruto Interno (se usa el índice)
- El Tipo de Cambio Real Multilateral (se usa el índice) y
- El Arancel promedio (Ad-valorem)



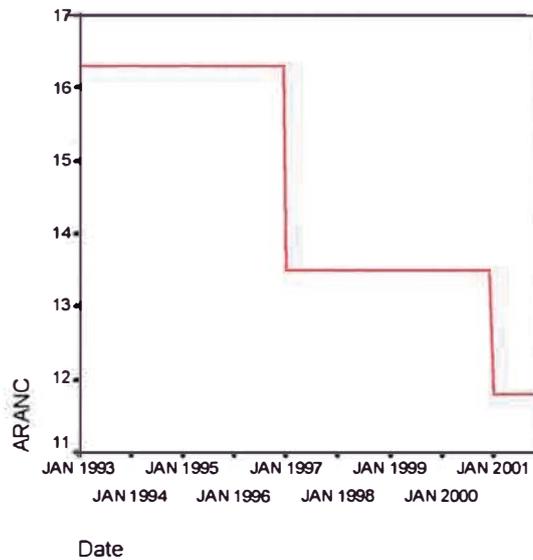
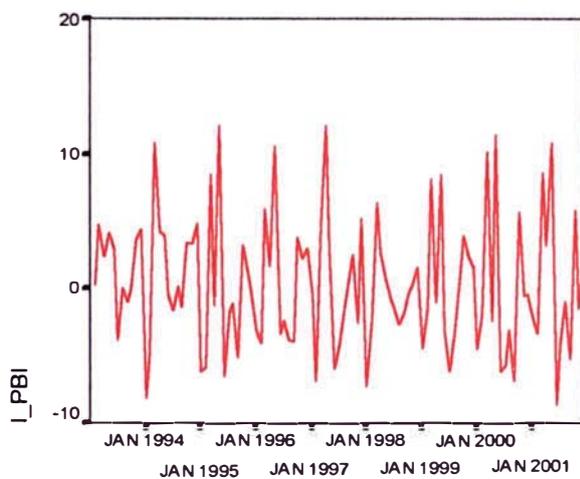


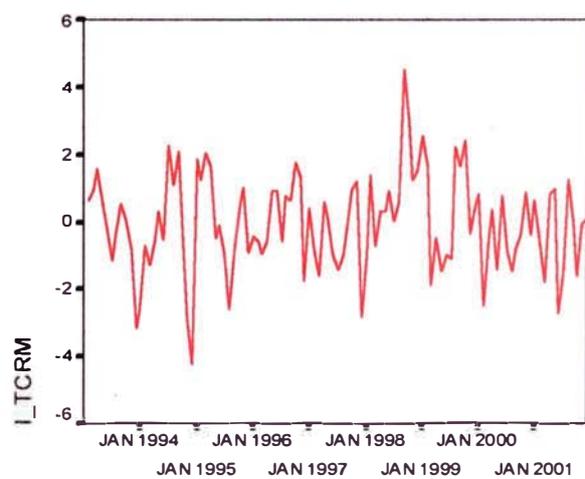
Figura 23
Gráficas de serie del I_PBI, I_TCRM y el ARANCEL

Como puede apreciarse las tres series no son estacionarias (véase Fig. 23, páginas 94 y 95), luego de una diferencia regular parece que logran ser estacionarias:



Date

Transforms: difference (1)



Date

Transforms: difference (1)

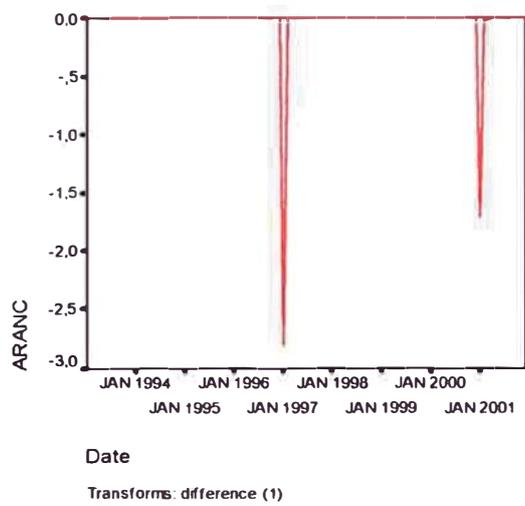
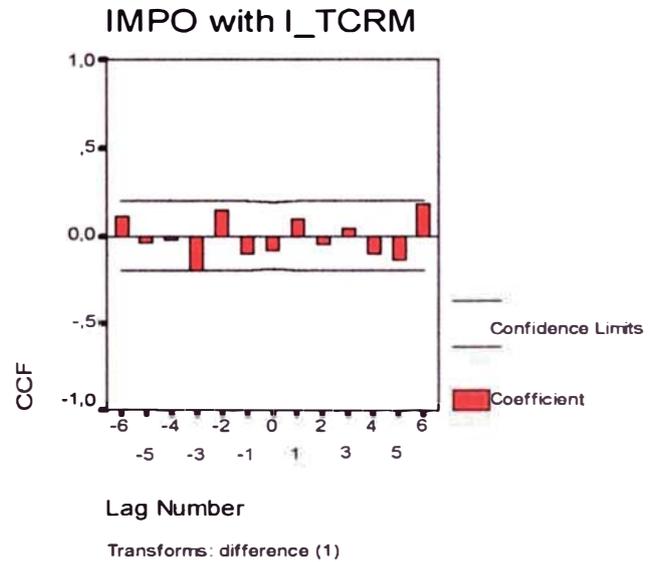
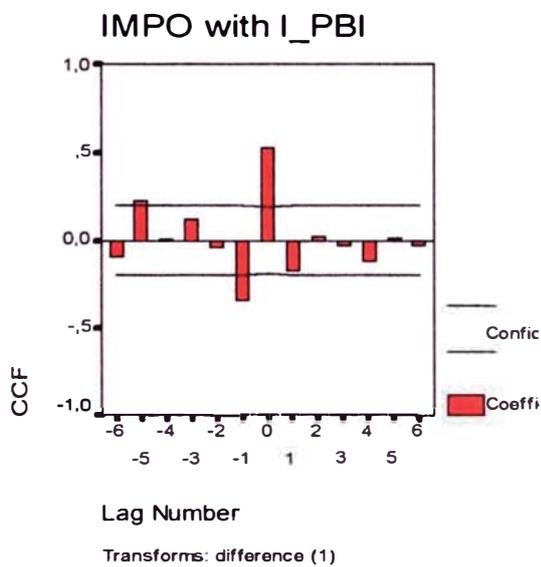


Figura 24
Gráficas diferenciadas del I_PBI, I_TCRM y el ARANCEL

Una vez que se ha conseguido estabilizar las series, puede ahora obtenerse los gráficos de correlación cruzada:



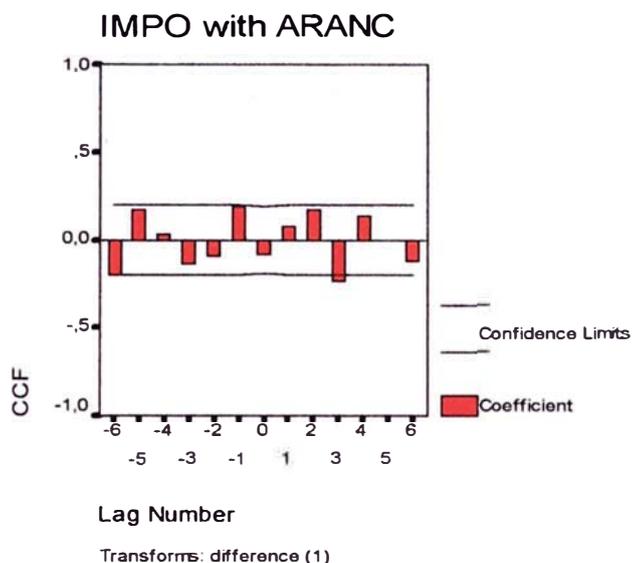


Figura 25
Correlación cruzada de las Importaciones con el I_PBI, I_TCRM y el ARANCEL

El primer gráfico de correlación cruzada, Impo vs. I_PBI indica que el rezago 0 y -1 son significativos; esto quiere decir, que las siguientes variables pueden incluirse en el modelo $(I_PBI_t - I_PBI_{t-1})$ y $(I_PBI_{t-1} - I_PBI_{t-2})$. En lo que respecta a las otras variables, estas parecen no tener algún aporte significativo.

El modelo estimado con las dos variables mencionadas en el párrafo anterior es:

Dependent Variable: D(IMPO)
 Method: Least Squares
 Date: 05/18/03 Time: 13:57
 Sample(adjusted): 1994:05 2001:12
 Included observations: 92 after adjusting endpoints
 Convergence achieved after 17 iterations
 Backcast: 1993:05 1994:04

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I_PBI)	4,485075	1,255748	3,571636	0,0006
D(I_PBI(-1))	-3,825958	1,238725	-3,088625	0,0027

AR(1)	-0,557848	0,107698	-5,179737	0,0000
AR(2)	-0,247220	0,106542	-2,320391	0,0227
SAR(12)	0,780777	0,058063	13,44701	0,0000
MA(12)	-0,882032	0,030704	-28,72648	0,0000
R-squared	0,649421	Mean dependent var	1,936989	
Adjusted R-squared	0,629039	S.D. dependent var	63,13999	
S.E. of regression	38,45640	Akaike info criterion	10,19992	
Sum squared resid	127185,0	Schwarz criterion	10,36439	
Log likelihood	-463,1964	F-statistic	31,86176	
Durbin-Watson stat	1,903250	Prob(F-statistic)	0,000000	
Inverted AR Roots	,98	,85+,49i	,85 -,49i	,49 - ,85i
	,49+,85i	,00+,98i	-,00 -,98i	- ,28+,41i
	-,28 -,41i	-,49 -,85i	-,49+,85i	-,85 - ,49i
	-,85+,49i	-,98		
Inverted MA Roots	,99	,86+,49i	,86 -,49i	,49+,86i
	,49 -,86i	-,00 -,99i	-,00+,99i	-,49 - ,86i
	-,49+,86i	-,86+,49i	-,86 -,49i	-,99

Aquí puede observarse que el R^2 ajustado ha ganado 5 puntos porcentuales y que los criterios de Akaike y Schwarz han disminuido ligeramente.

El correlograma (véase Fig. 26, página 99) muestra que los residuos no están correlacionados.

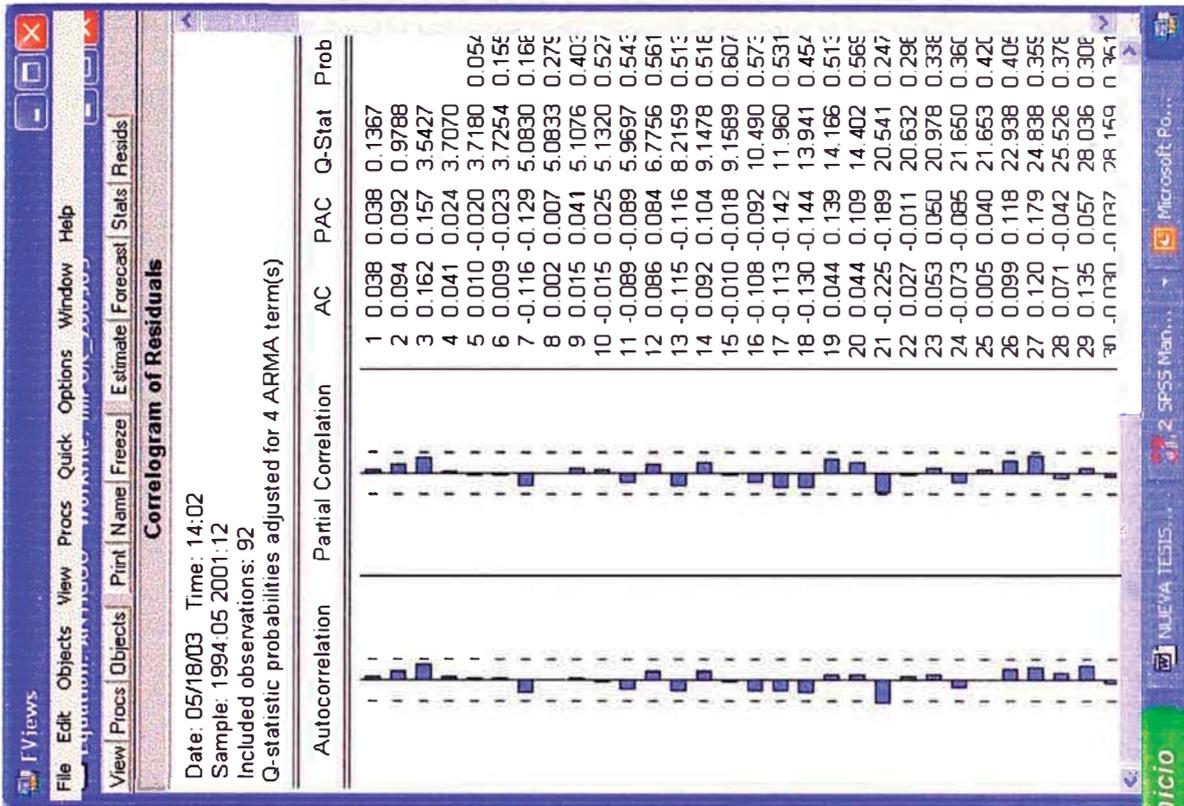


Figura 26
Correlograma de los residuos

En lo concerniente a la normalidad y resultados descriptivos se tiene el siguiente resultado:

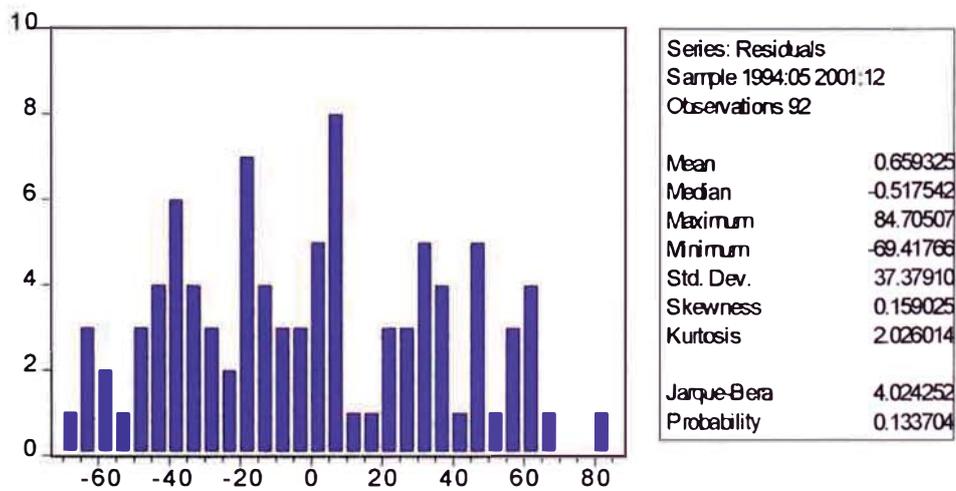


Figura 27
Histograma y estadísticas descriptivas de los residuos

La distribución se acepta como normal de acuerdo a la prueba de Jarque-Vera; sin embargo, la media de los residuos esta algo alejada del valor esperado de cero.

Finalmente con este modelo se generara los pronósticos para el período de prueba y se evaluara la capacidad descriptiva.

$$D(IMPO) = [AR(1)=0,56 ; AR(2)=-0,25 ,SAR(12)=0,78, MA(12)=-0,88] + 4,49*D(I_PBI) - 3,83*D(I_PBI(-1))$$

Generación de Pronósticos y Evaluación de la Capacidad Predictiva

Se procede a generar pronósticos para el período 2002.

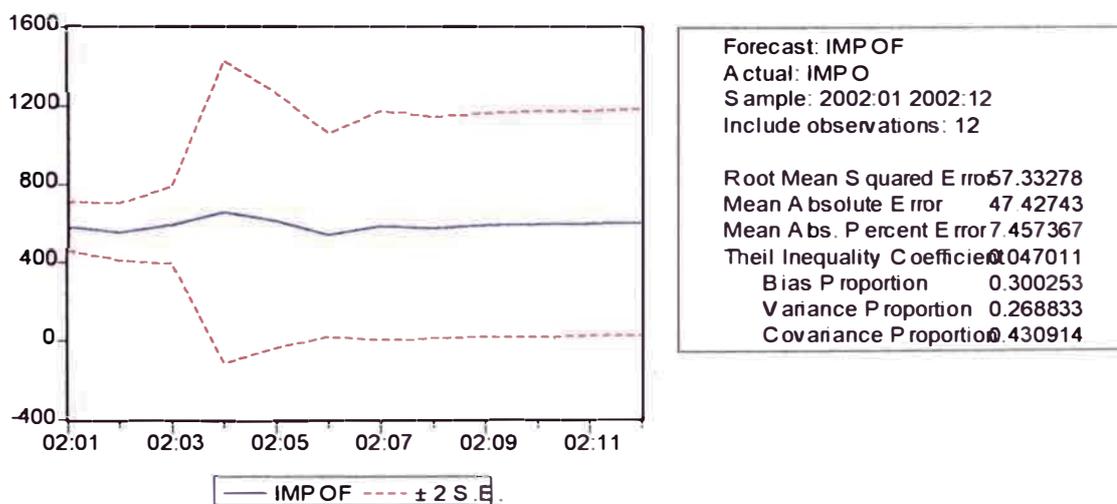


Figura 28
Evaluación de la capacidad predictiva

Las estadísticas MAE, MAPE, y el Coeficiente de Theil (véase Fig. 28, página 100) permiten evaluar la capacidad predictiva del modelo en las predicciones mensuales:

MAE = 47,4 Es el error medio absoluto, la magnitud promedio del error de predicción ya sea este de subestimación o sobrestimación es de 47,4 millones de dólares.

MAPE = 7,5 Es el error medio porcentual absoluto, el modelo pronostica importaciones que en promedio se espera que estén 7,5% por encima o por debajo de las cifras reales de importación.

Coef.

Theil = 0,05 Mientras más cercano a cero se encuentre este valor mejor.

De entre las proporciones la más alta es la correspondiente a la covarianza, por lo que parece que los pronósticos son buenos.

Tabla V
Comparación de Valores Reales y Proyectados en el Período de Validación

Mes	Reales 2002	Pronóstico 2002	Error Absoluto Porcentual
Enero	580,1	583,6	0,6
Febrero	545,8	554,7	1,6
Marzo	514,4	592,9	15,3
Abril	651,5	656,7	0,8
Mayo	661,6	613,3	7,3
Junio	552,2	541,0	2,0
Julio	678,9	587,2	13,5
Agosto	624,6	578,1	7,4
Septiembre	682,1	594,0	12,9
Octubre	661,1	598,7	9,4
Noviembre	644,9	601,1	6,8
Diciembre	688,7	607,7	11,8
Total	7486,0	7109,0	5,0

El error de pronóstico para la proyección anual del año 2002 es -377,0 millones lo que significa que la proyección resulto ser 5,0% (véase la Tabla V) menor que la cifra real. Considerando el valor de MAPE de 7,5% para las proyecciones mensuales puede construirse un intervalo promedio: 6615,6 – 7681,8 estos valores extremos pueden tomarse como los escenarios pesimista y optimista respectivamente; la cifra real del año 2002 de 7486,0 se encuentra en este intervalo.

Luego de la proyección inicial para el 2002 efectuada al cierre del año 2001; se procede a realizar las actualizaciones respectivas cada tres meses, de acuerdo al plan establecido. Esta actualización trimestral permite afinar los pronósticos del

año y acercarse cada vez más a la cifra real gracias al aporte de nueva información (véase la Tabla VI).

Tabla VI
Proyecciones y Actualizaciones Trimestrales

Real	1er trim.	2do trim	3er trim
580,1	580,1	580,1	580,1
545,8	545,8	545,8	545,8
514,4	514,4	514,4	514,4
651,5	619,0	651,5	651,5
661,6	576,7	661,6	661,6
552,2	496,8	552,2	552,2
678,9	548,0	603,6	678,9
624,6	538,6	598,2	624,6
682,1	550,8	601,8	682,1
661,1	556,3	610,5	670,3
644,9	559,2	615,9	673,7
688,7	564,4	619,0	680,5
7486,0	6650,2	7154,8	7515,8
Var. %	-11,2	-4,4	0,4

El cuadro precedente muestra las cifras que se proyectan para lo que resta del año 2002 luego del primer trimestre, después del segundo trimestre y luego del tercer trimestre. A partir del segundo trimestre el error porcentual empieza a disminuir; sin embargo, después al primer trimestre la proyección tiene un margen porcentual de error de mas del doble en comparación con la proyección al cierre del año 2001.

Las proyecciones efectuadas luego del primer trimestre; ponen en evidencia lo importante que es contar con otro tipo de información relevante para ajustar las

proyecciones. Como ya se dijo antes es importante el juicio de un experto que ayude a dar mayor precisión a los resultados.

Con los datos hasta diciembre del año 2002 se procede a proyectar las cifras del año 2003 (véase la Tabla VII). La cifra anual proyectada para ese año resulta ser de 8129,0 millones de dólares; luego del primer trimestre puede observarse que las cifras reales fueron mayores a las que se habían predicho inicialmente para los primeros tres meses del año.

Tabla VII
Proyecciones para el año 2003

Proyección 2003	Valores reales	Proyección 1trim. 2003	Proyección 1trim. 2003
641,3	717,8	717,8	717,8
620,3	630,4	630,4	630,4
657,7	723,8	723,8	723,8
748,2		811,2	748,2
714,0		792,3	714,0
636,5		718,0	636,5
679,2		744,3	679,2
669,3		704,5	669,3
680,4		717,4	680,4
689,2		759,3	689,2
692,7		732,7	692,7
700,3		748,5	700,3
8129		8800,3	8281,7

Una vez conocidas las cifras del primer trimestre del año 2003, se procede a realizar la actualización de las proyecciones y resulta que la cifra anual proyectada es de 8 800,3 millones de dólares; sin embargo según el entorno económico (ver el capítulo *Pronostico para el 2003*) que se presenta y la opinión de los especialistas es difícil que se alcance dicha cifra. Se decide entonces mantener las proyección anteriores para lo que resta del año y así se obtiene la cifra de 8281,7 millones de dólares.

3.3 Aplicación de la metodología de análisis de regresión

Elección de variables a ser usadas como variables explicativas

Las variables que se usan como explicativas o variables independientes en el modelo causal son las siguientes:

- El Índice del Producto Bruto Interno,
- El Índice del Tipo de Cambio Real Multilateral y
- El Arancel Promedio mas la Sobretasa Adicional

Las razones para considerar estas variables ya se mencionaron en el capítulo III.

Análisis Gráfico

Las siguientes son las gráficas de dispersión de las Importaciones en valor CIF con cada una de las variables independientes:

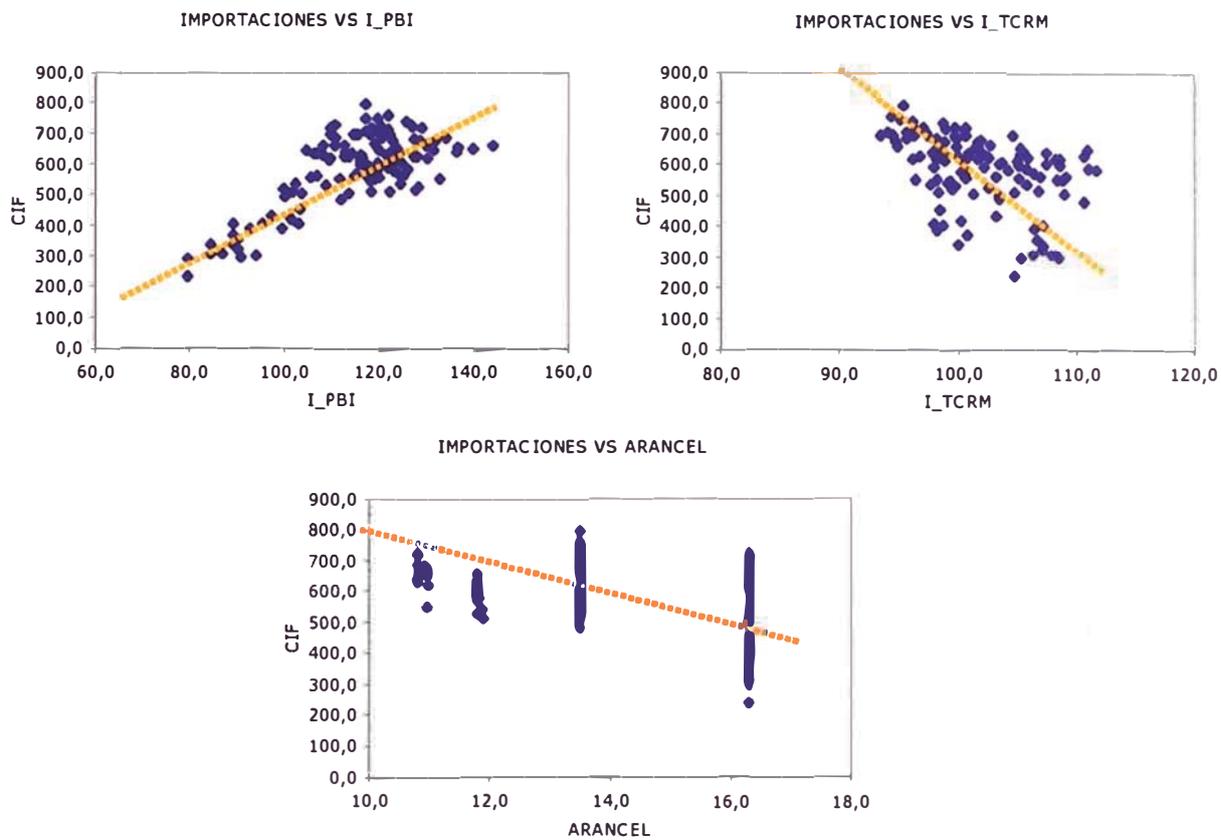


Figura 29
Gráficas de dispersión entre la variable dependiente y las explicativas

Tabla VIII
Coeficientes de correlación

	IMPO	I_PBI	I_TCRM	ARANC
IMPO	1			
I_PBI	0,74	1		
I_TCRM	-0,44	-0,10	1	
ARANC	-0,43	-0,79	-0,02	1

El primero de los gráficos muestra que las importaciones tienen una relación lineal positiva importante con el índice del producto bruto interno, 0,74 en términos cuantitativos (véase la Tabla VIII). En los dos siguientes gráficos la

relación lineal es menos fuerte $-0,44$ y $-0,43$ y es de tipo negativa. Entre los tres pueda quizá lograrse una buena relación lineal múltiple.

Los diagramas de caja (véase la Fig. 30), menos para la variable arancel que casi tiene un comportamiento similar al de una variable discreta, revelan que los montos CIF de importación y el Índice del Producto Bruto Interno tienen ambos distribuciones con sesgo negativo, y algunas observaciones atípicas (los círculos fuera de los bigotes de la caja) mientras que el Índice del Tipo de Cambio Real Multilateral tiene un comportamiento con un ligero sesgo positivo.

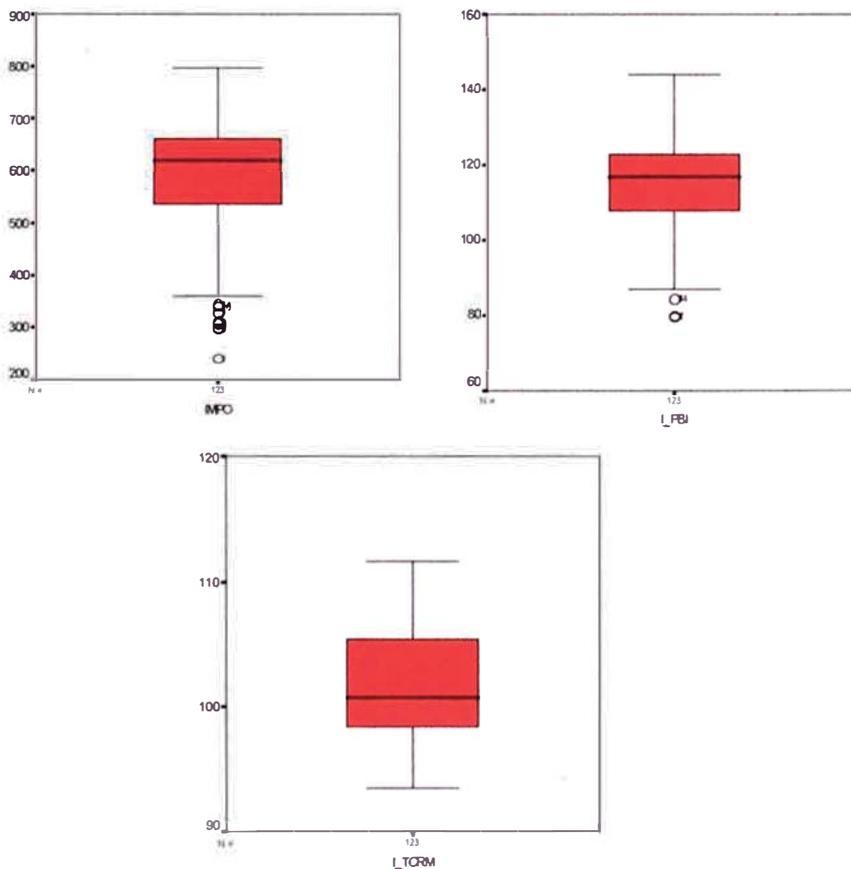


Figura 30
Diagrama de Cajas de las Importaciones, I_PBI y el I_TCRM

Ajuste del Modelo

Se ajustó un modelo de regresión lineal múltiple con todas las variables independientes y que corresponden al periodo 1993-2001.

Dependent Variable: IMPO
Method: Least Squares
Date: 05/21/03 Time: 22:21
Simple: 1993:01 2001:12
Included observations: 108

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I_PBI	8,911364	0,719129	12,39188	0,0000
I_TCRM	-9,934367	1,271444	-7,813454	0,0000
ARANC	14,34251	5,379171	2,666304	0,0089
C	384,0761	221,1145	1,737001	0,0853
R-squared	0,768221	Mean dependent var	578,8381	
Adjusted R-squared	0,761535	S.D. dependent var	121,7993	
S.E. of regression	59,47809	Akaike info criterion	11,04543	
Sum squared resid	367914,9	Schwarz criterion	11,14477	
Log likelihood	-592,4530	F-statistic	114,9010	
Durbin-Watson stat	0,867445	Prob(F-statistic)	0,000000	

La constante no parece ser muy significativa por lo que es retirada del modelo, así resulta:

Dependent Variable: IMPO
Method: Least Squares
Date: 05/21/03 Time: 22:23
Simple: 1993:01 2001:12
Included observations: 108

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I_PBI	9,860027	0,472304	20,87643	0,0000
I_TCRM	-8,261164	0,837799	-9,860550	0,0000
ARANC	21,64280	3,389451	6,385342	0,0000

R-squared	0,761497	Mean dependent var	578,8381
Adjusted R-squared	0,756954	S,D, dependent var	121,7993
S.E. of regression	60,04670	Akaike info criterion	11,05551
Sum squared resid	378588,6	Schwarz criterion	11,13001
Log likelihood	-593,9974	F-statistic	167,6227
Durbin-Watson stat	0,882878	Prob(F-statistic)	0,000000

No hay cambios muy importantes en las estadísticas de este modelo respecto al anterior; según la prueba F conjunta y las pruebas T individuales todos los coeficientes de este modelo son significativos. El coeficiente de determinación múltiple ajustado de este modelo es 75,7%.

Validación de supuestos

De acuerdo a la gráfica de línea de los residuos (véase la Fig. 31), estos parecen comportarse como ruido blanco; es decir con media cero y varianza constante, respecto a la condición de no correlación esta no puede ser claramente discernida a partir de este gráfico.

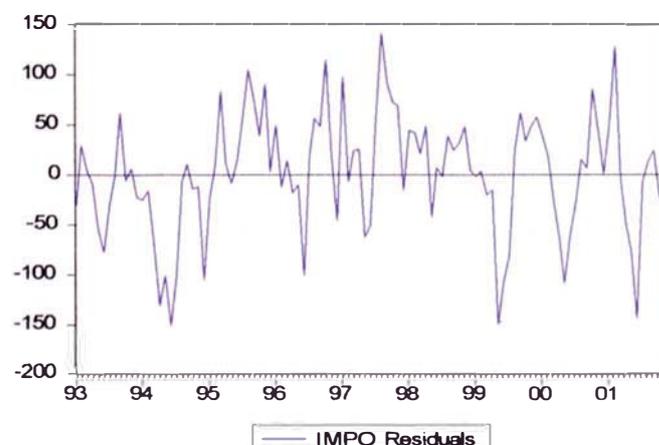


Figura 31
Gráfica lineal de los residuos

En el correlograma y correlograma parcial (véase la Fig. 32), se observa que esta presente alguna estructura de autocorrelación.

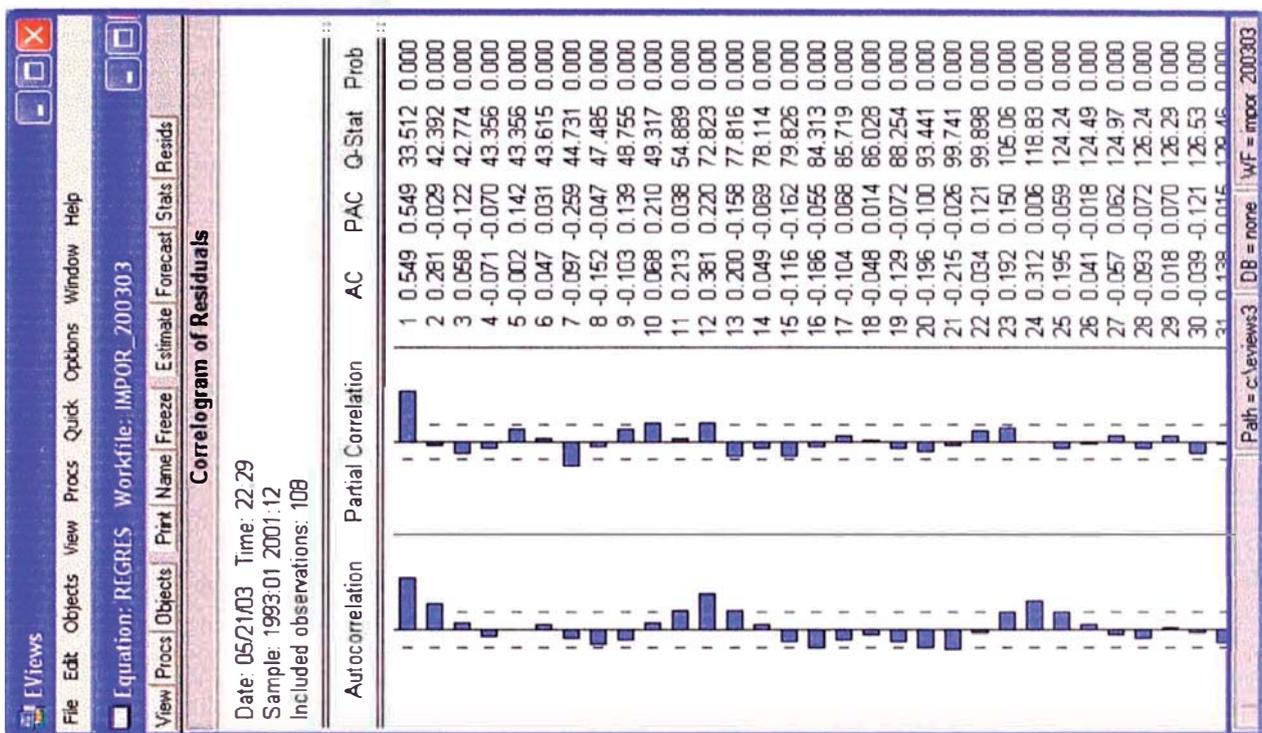


Figura 32
Correlograma y correlograma parcial

Específicamente los coeficientes significativos son $ma(1)$, $ma(2)$, $ar(1)$, $sma(12)$, y $sma(24)$; esta estructura de correlación en los residuos puede modelarse con modelos ARIMA de series de tiempo. Al añadir esta estructura al modelo resulta:

Dependent Variable: IMPO
 Method: Least Squares
 Date: 05/21/03 Time: 23:05
 Sample(adjusted): 1993:02 2001:12
 Included observations: 107 after adjusting endpoints
 Convergence achieved after 100 iterations
 Backcast: 1990:12 1993:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I_PBI	8,912671	0,826823	10,77942	0,0000
I_TCRM	-6,472843	1,481763	-4,368339	0,0000
ARANC	17,13683	6,226647	2,752177	0,0070
AR(1)	0,460194	0,114222	4,028953	0,0001
MA(1)	-0,030975	0,081186	-0,381533	0,7036
MA(2)	0,038257	0,065990	0,579738	0,5634
SMA(12)	0,271904	0,056395	4,821442	0,0000
SMA(24)	0,784640	0,000156	5041,900	0,0000

R-squared	0,886799	Mean dependent var	582,0016
Adjusted R-squared	0,878795	S.D. dependent var	117,8302
S.E. of regression	41,02205	Akaike info criterion	10,33792
Sum squared resid	166598,0	Schwarz criterion	10,53776
Log likelihood	-545,0787	F-statistic	110,7928
Durbin-Watson stat	1,994043	Prob(F-statistic)	0,000000
Inverted AR Roots	,46		
Inverted MA Roots	,98+,14i	,98 -,14i	,92 -,37i
	,78+,61i	,78 -,61i	,61+,78i
	,37 -,92i	,37+,92i	,14 -,98i
	,02+,19i	,02 -,19i	-,14 -,98i
	-,37 -,92i	-,37+,92i	-,61 -,78i
	-,78 -,61i	-,78+,61i	-,92 -,37i
	-,98+,14i	-,98 -,14i	

Las componentes ma(1) y ma(2) no son significativas, y al retirarlas del modelo se obtiene:

Dependent Variable: IMPO

Method: Least Squares

Date: 05/21/03 Time: 23:04

Simple(adjusted): 1993:02 2001:12

Included observations: 107 after adjusting endpoints

Convergence achieved after 31 iterations

Backcast: 1991:02 1993:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I_PBI	9,193513	0,762285	12,06046	0,0000
I_TCRM	-7,133621	1,369366	-5,209432	0,0000
ARANC	19,57123	5,812524	3,367079	0,0011
AR(1)	0,435570	0,091212	4,775346	0,0000
MA(12)	0,261069	0,056521	4,618996	0,0000
MA(24)	0,784613	0,000135	5798,322	0,0000
R-squared	0,886645	Mean dependent var	582,0016	
Adjusted R-squared	0,881033	S.D. dependent var	117,8302	
S.E. of regression	40,64151	Akaike info criterion	10,30190	
Sum squared resid	166825,0	Schwarz criterion	10,45178	
Log likelihood	-545,1515	F-statistic	158,0008	

Durbin-Watson stat	2,006840	Prob(F-statistic)	0,000000
Inverted AR Roots	,44		
Inverted MA Roots	,98+,14i	,98 -,14i	,92 -,37i
			,92+,37i
	,78+,61i	,78 -,61i	,61+,78i
			,61 -
			,78i
	,37 -,92i	,37+,92i	,14 -,98i
			,14+,98i
	-,14 -,98i	-,14+,98i	-,37 -,92i
			-,37+,92i
	-,61 -,78i	-,61+,78i	-,78 -,61i
			-,78+,61i
	-,92 -,37i	-,92+,37i	-,98 -
			-,98 +
			,14i

Este último modelo ya no tiene el problema de autocorrelación (véase la Fig. 33).

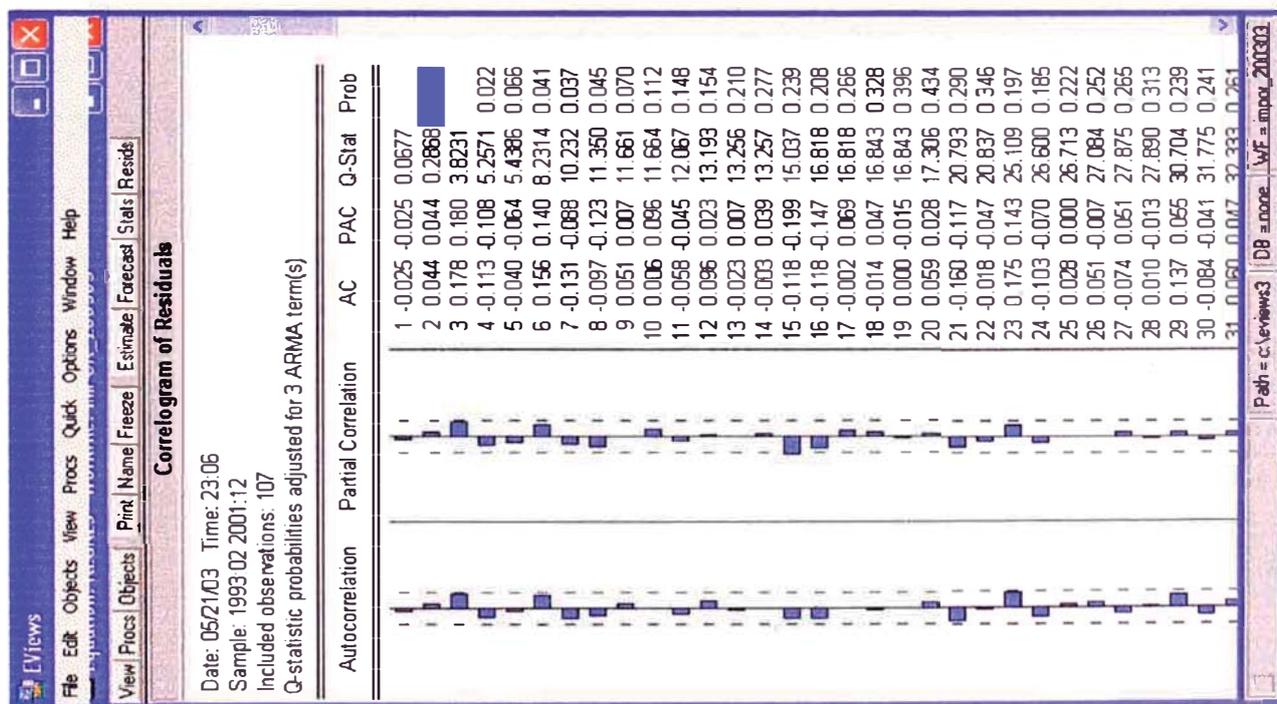


Figura 33, Correlograma y correlograma parcial

Además este último modelo representa una mejora respecto del anterior en lo referente al ajuste pues ahora el coeficiente de determinación ajustado es de 88,1%. La media de estos nuevos residuos es 0,28, la desviación estándar es 39,67, y no se rechaza la hipótesis de distribución normal (véase la Fig. 34).

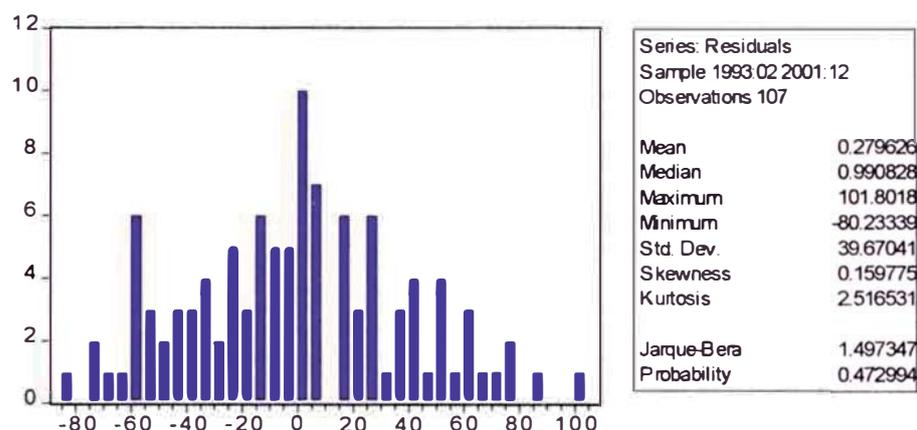


Figura 34
Histograma y estadísticas descriptivas de los residuos

Respecto a la presencia de multicolinealidad, se tiene lo siguiente:

- Las prueba F es significativa, así como todas las pruebas T individuales.
- Se observa también que los coeficientes de correlación (véase la Tabla IX) entre las variables independientes no son muy altos salvo la relación entre las variables I_PBI y ARANC que es de -0,79 pero aún así no está muy cerca de -1.

Tabla IX
Coeficientes de correlación

	IMPO	I_PBI	I_TCRM	ARANC
IMPO	1			
I_PBI	0,74	1		
I_TCRM	-0,44	-0,10	1	
ARANC	-0,43	-0,79	-0,02	1

Uno podría pensar quizá entonces en retirar la variable ARANC y de hecho el modelo no sufriría cambios importantes. Es cierto que esta variable no era considerada al principio entre las determinantes de las importaciones; sin embargo como ya se explicó esta variable esta adquiriendo importancia en el contexto actual, por lo que su inclusión se hace necesaria.

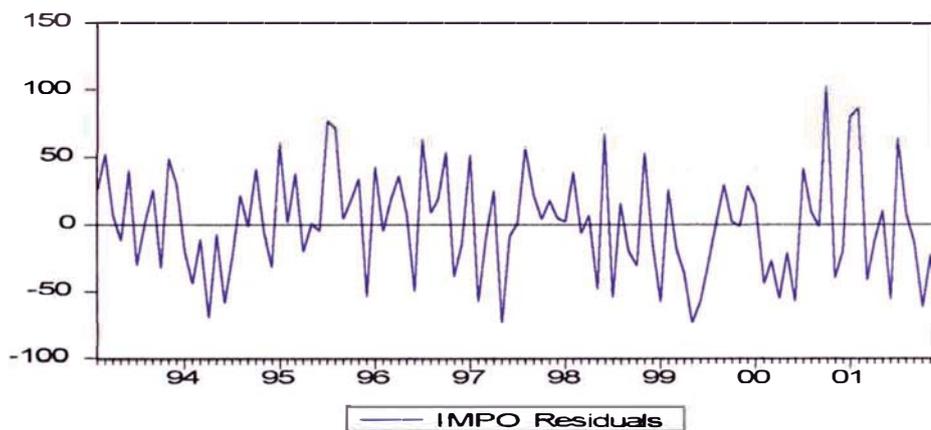


Figura 35
Gráfica de serie de los residuos

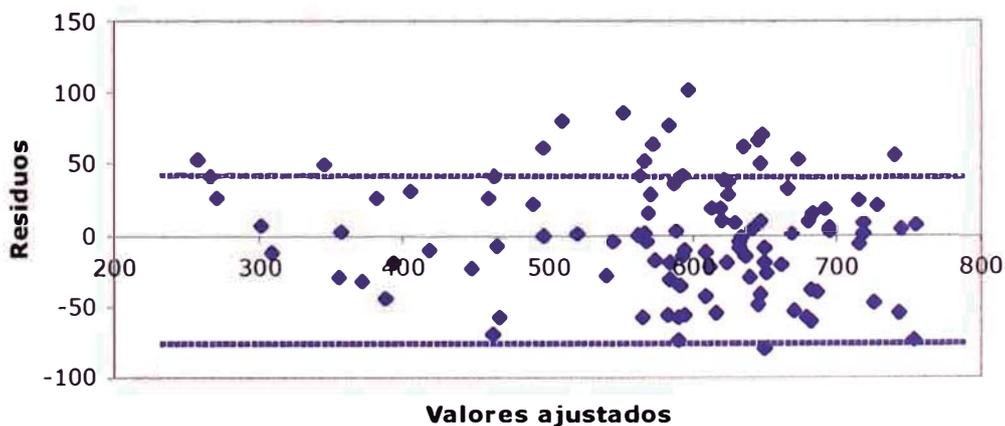


Figura 36
Gráfica de los residuos vs. los valores ajustados \hat{y}

Tabla X
Prueba de homogeneidad de los residuos

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
RESIDUOS	Based on Mean	,855	8	98	,557
	Based on Median	,625	8	98	,755
	Based on Median and with adjusted df	,625	8	79,156	,754
	Based on trimmed mean	,825	8	98	,582

Según los gráficos de residuos anteriores (véase la Fig. 35 y 36, página 114) y la exploración gráfica de estos no parece estar presente la falta de homogeneidad en la variabilidad lo que se comprueba al aplicar la prueba de homogeneidad de Levene (véase la Tabla X), la cual no rechaza la hipótesis nula de igualdad de varianzas; se presentan algunos valores altos cerca del extremo derecho de la serie de tiempo de los residuos, el mayor de éstos alcanza la cifra de 101,8 pero no pasa los $3\sigma = 119,1$. Parece así que la presencia de estos puntos no tiene mayor influencia en el análisis, lo cual se refleja en el valor alto del coeficiente de determinación ajustado 88,1%, y la conclusión de la prueba de normalidad. La siguiente gráfica presenta los valores reales con los ajustados.

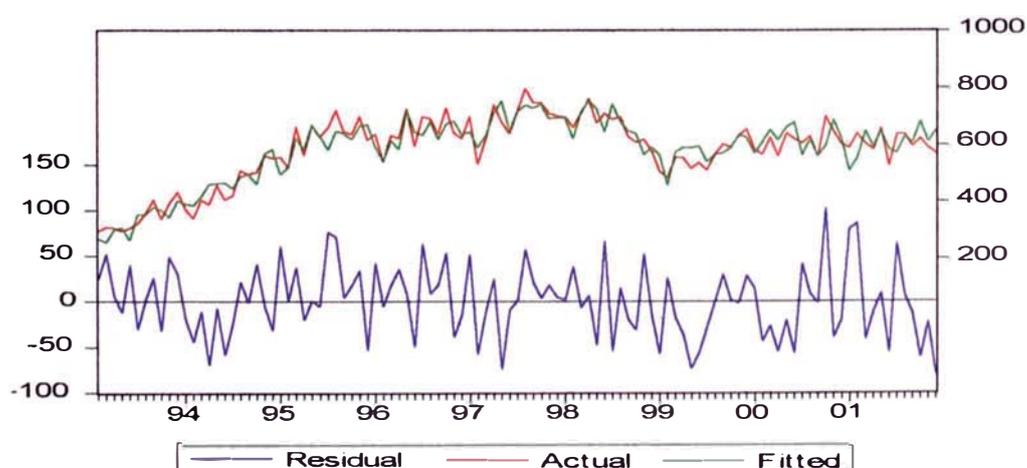


Figura 37
Valores Observados y Valores Ajustados en el Modelo Causal

Interpretación del Modelo

El modelo es finalmente el siguiente:

$$\text{IMPO} = 9,2 * \text{I_PBI} + -7,1 * \text{I_TCRM} + 19,6 * \text{ARANC} + [\text{AR}(1) = 0,4, \text{MA}(12) = 0,3, \text{MA}(24) = 0,8] + e$$

El programa E-views no tiene la opción para obtener coeficientes estandarizados; que si lo tiene el programa SPSS, sin embargo este tampoco tiene dicha opción en el caso de modelos más complicados como el presente (ARIMA en los residuos). No obstante podrían compararse los coeficientes del I_PBI y I_TCRM pues ambas son variables de tipo índice más no así para la variable ARANC que esta expresada en porcentaje. Entonces el PBI es más importante que el tipo de cambio real múltilateral y además los efectos de estas variables son opuestas, hay que recordar siempre que la interpretación de estos coeficientes es relativa.

Generación de Pronósticos y Evaluación de la Capacidad Predictiva

Se generan los pronósticos para el año 2002 y se obtienen las estadísticas MAE, MAPE, y el Coeficiente de Theil que permiten evaluar la capacidad predictiva del modelo en las predicciones mensuales (véase la Fig. 38, página 117).

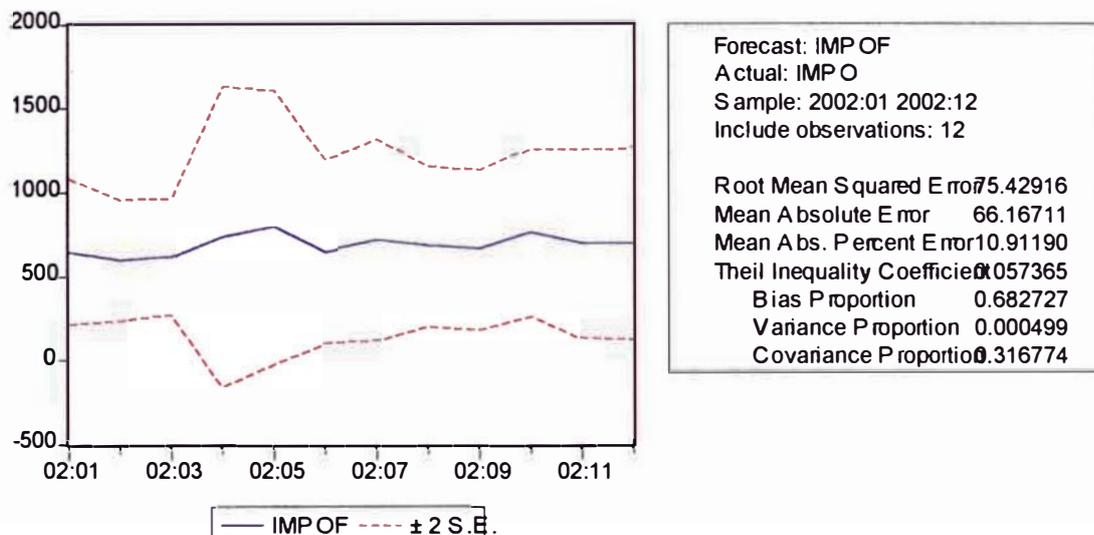


Figura 38
Evaluación de la capacidad predictiva

MAE = 66,2 Es el error medio absoluto, la magnitud promedio del error de predicción mensual ya sea este de subestimación o sobrestimación es de 66,2 millones de dólares.

MAPE = 10,9 Es el error medio porcentual absoluto, el modelo pronostica importaciones que en promedio se espera que estén 10,9% por encima o por debajo de las cifras reales de importación.

Coef.

Theil = 0,06 Mientras más cercano a cero se encuentre este valor mejor.

De entre las proporciones la más alta, es la que corresponde a la componente del sesgo y no a la covarianza como debía esperarse; las proyecciones de este modelo tienden entonces a dar cifras algo mayores de las reales.

Tabla XI
Comparación de valores reales y proyectados en el período de validación
y actualizaciones trimestrales

Real	2002	1er. Trim 2002	2do. Trim 2002	3er. Trim 2002
580,1	644,7	580,1	580,1	580,1
545,8	595,3	545,8	545,8	545,8
514,4	616,4	514,4	514,4	514,4
651,5	738,6	681,9	651,5	651,5
661,6	794,5	763,1	661,6	661,6
552,2	650,1	632,7	552,2	552,2
678,9	715,2	706,1	658,8	678,9
624,6	680,2	675,7	648,1	624,6
682,1	659,1	658,4	640,4	682,1
661,1	758,1	760,9	748,5	769,1
644,9	690,6	691,0	681,7	698,8
688,7	691,2	693,1	685,0	694,5
7486,0	8233,9	7903,2	7568,3	7653,6
Var %	10,0	5,6	1,1	2,2

El error de pronóstico para el año 2002, al cierre de las cifras del año 2001 fue de 748,1 millones de dólares, cifra que representa un error de 10% respecto de la cifra real para ese año que alcanzo los 7 486,0 millones (véase la Tabla XI). Considerando el valor de MAPE de 10,9 % para las proyecciones mensuales puede construirse un intervalo promedio para la proyección anual: 7424,6 – 9241,2 estos valores extremos pueden tomarse como los escenarios pesimista y optimista respectivamente, el rango de variación de este intervalo es muy amplio.

Luego del primer, segundo y tercer trimestre del año 2002; se procede a efectuar la actualización de las proyecciones según el plan establecido. Y como

puede observarse del cuadro precedente los porcentajes de error fueron disminuyendo.

Finalmente con los datos hasta diciembre del 2002 se procede a efectuar la proyección de las cifras del año 2003 (véase la Tabla XII) y resulta una proyección de 8 761,3 millones para ese año.

Tabla XII
Proyecciones para el año 2003, I_TCRM en decrecimiento

Proyección 2003	Valores reales	Proyección 1trim. 2003	Proyección 1trim. 2003
689,2	717,8	717,8	717,8
711,3	630,4	630,4	630,4
642,1	723,8	723,8	723,8
816,6		852,0	816,6
839,4		845,2	839,4
716,3		724,3	716,3
785,2		782,1	785,2
729,5		723,6	729,5
728,9		753,9	728,9
666,8		723,1	666,8
716,3		729,1	716,3
719,6		768,4	719,6
8761,3		8973,7	8790,6

Luego de transcurrido el primer trimestre y conocidas las cifras de este período se procede a la actualización de los pronósticos y resulta una cifra de 8 973,7 millones de dólares, cifra que es poco probable de alcanzar; similar al caso del pronóstico con series de tiempo. Si se acepta la solución que se aplicó en el caso del modelo ARIMA se llega en este caso a una cifra de 8 790,6 millones, cifra que aún es alta; hay que recordar que de acuerdo a la descomposición de proporciones del error, este modelo tiende a dar proyecciones sesgadas.

Los pronósticos que acaban de obtenerse se basan en la suposición de que el índice del tipo de cambio real multilateral seguirá la tendencia decreciente observada en los últimos años. Si se toma la otra opción, la de estancamiento para este índice los pronósticos son los siguientes:

Tabla XIII
Proyecciones para el año 2003, I_TCRM estancado

Proyección 2003	Valores reales	Proyección 1trim. 2003	Proyección 1trim. 2003
689,2	717,8	717,8	717,8
711,3	630,4	630,4	630,4
642,1	723,8	723,8	723,8
795,5		842,7	795,5
818,2		835,8	818,2
695,3		715,0	695,3
764,1		772,7	764,1
708,2		714,2	708,2
707,6		744,5	707,6
645,9		713,8	645,9
695,3		719,8	695,3
698,7		759,1	698,7
8571,4		8889,5	8600,8

Estos pronósticos son todavía altos aunque menores que los obtenidos con la suposición de decrecimiento del índice del tipo de cambio real multilateral.

3.4 Comparación de Modelos

Finalmente debe elegirse el modelo que será usado para efectuar las proyecciones. Si observamos las estadísticas de ajuste de ambos modelos resulta el cuadro siguiente (véase la Tabla XIV, página 121):

Tabla XIV
Comparación de modelos 1

	R² ajustado	Akaike	Schwarz
Causal	0,881	10,302	10,452
Serie ARIMA	0,629	10,200	10,364

A primera vista al comparar el valor del R² ajustado pareciera que el modelo causal tiene un mejor ajuste; debe recordarse, sin embargo, que el coeficiente de determinación ajustado no puede usarse para comparar estos modelos, ya que en el modelo causal la serie no esta diferenciada mientras que si lo esta en el modelo ARIMA. Por otra parte si observamos las estadísticas de Akaike y Schwarz, que si son aplicables, el mejor modelo es el ARIMA estacional, ya que este tiene los menores valores para estas estadísticas

A continuación se comparan los modelos según las estadísticas que miden la capacidad predictiva para las proyecciones del año 2002 en base a las cifras hasta diciembre del 2001.

Tabla XV
Comparación de modelos 2

	MAE	MAPE	THEIL	Sesgo	Var.	Cov.
Serie ARIMA	43,4	7,5	0,05	0,30	0,27	0,43
Causal	66,2	10,9	0,06	0,68	0,00	0,32

Claramente el modelo de series de tiempo parece ser el de mejor desempeño: tiene menor error medio absoluto, menor error medio porcentual absoluto, y menor valor para el coeficiente de Theil. En las proporciones del error, el mayor

valor le corresponde a la covarianza, lo que indica mejores pronósticos que el modelo causal que tiene la mayor proporción en el sesgo, hecho que corrobora los altos observados con este modelo.

En lo que respecta al horizonte de proyección, los modelos de series de tiempo solo pueden usarse para horizontes cortos, a lo más de un año. Ya que el pasado reciente puede tener un comportamiento distinto del futuro. Los modelos causales pueden desempeñarse mejor en las proyecciones de horizonte mayor a un año si se tienen las previsiones o escenarios posibles del comportamiento futuro de sus variables independientes. En el caso particular de los modelos estudiados el horizonte de proyección fue de un año y se tuvieron los siguientes pronósticos para el año 2002, con las actualizaciones trimestrales.

Tabla XVI
Comparación de modelos 3

	Real	Series ARIMA	Error %	Causal	Error %
2001	7486,0	7109,0	-5,0	8233,9	10,0
1 trim.	7486,0	6650,2	-11,2	7903,2	5,6
2 trim.	7486,0	7154,8	-4,4	7568,3	1,1
3 trim.	7486,0	7515,8	0,4	7653,6	2,2

Al comparar las proyecciones iniciales al término del año 2001, resulta que el modelo causal tienen un error de 10% (véase la Tabla XVI) mientras que el modelo de series de tiempo ARIMA solo de 5%; luego en las actualizaciones trimestrales, el error porcentual varía entre 0,4% y 11,2% en términos absolutos para el modelo ARIMA; y entre 1,1% y 5,6% en el modelo causal.

En conclusión parece ser que el modelo de series de tiempos es el más adecuado. No obstante, el modelo causal es muy importante por que permite la elección de diversos escenarios económicos que no son posibles al usar series de tiempo. Hay entonces tres posibles caminos: usar solo series de tiempo, lo que no ayudaría en la construcción de escenarios macroeconómicos (este modelo incluye al PBI pero no a las otras variables); usar el modelo de regresión, el cual tienen un menor desempeño que el modelo de series por lo que deberían hacérsele mayores ajustes de juicio; y la tercera opción en la que no se descarta el aporte de ambos modelos : el buen desempeño del modelos de series y la posibilidad de trabajar con escenarios del modelo causal.

Se tendría entonces un nuevo pronóstico que combina las cifras de los dos modelos; esta combinación puede darse en la forma de un promedio ponderado donde los pesos para cada pronóstico son las inversas de las desviaciones estándar (las inversas permiten asignar un mayor peso a los pronósticos con menor error) de los residuos de cada modelo. Las desviaciones estándar de los residuos de los modelos de series de tiempo y el modelo de regresión para la proyección de las cifras del año 2002 en base al período 1993-2001 son: 38,45640 y 40,64151 como puede apreciarse en las siguientes salidas de Eviews:

ARIMA

Dependent Variable: D(IMPO)

Method: Least Squares

Date: 08/14/03 Time: 23:01

Sample(adjusted): 1994:05 2001:12
 Included observations: 92 after adjusting endpoints
 Convergence achieved after 17 iterations
 Backcast: 1993:05 1994:04

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I_PBI)	4,485075	1,255748	3,571636	0,0006
D(I_PBI(-1))	-3,825958	1,238725	-3,088625	0,0027
AR(1)	-0,557848	0,107698	-5,179737	0,0000
AR(2)	-0,247220	0,106542	-2,320391	0,0227
SAR(12)	0,780777	0,058063	13,44701	0,0000
MA(12)	-0,882032	0,030704	-28,72648	0,0000
R-squared	0,649421	Mean dependent var	1,936989	
Adjusted R-squared	0,629039	S,D, dependent var	63,13999	
S.E. of regression	38,45640	Akaike info criterion	10,19992	
Sum squared resid	127185,0	Schwarz criterion	10,36439	
Log likelihood	-463,1964	F-statistic	31,86176	
Durbin-Watson stat	1,903250	Prob(F-statistic)	0,000000	

REGRESION

Dependent Variable: IMPO
 Method: Least Squares
 Date: 08/14/03 Time: 23:03
 Sample(adjusted): 1993:02 2001:12
 Included observations: 107 after adjusting endpoints
 Convergence achieved after 31 iterations
 Backcast: 1991:02 1993:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I_PBI	9,193513	0,762285	12,06046	0,0000
I_TCRM	-7,133621	1,369366	-5,209432	0,0000
ARANC	19,57123	5,812524	3,367079	0,0011
AR(1)	0,435570	0,091212	4,775346	0,0000

MA(12)	0,261069	0,056521	4,618996	0.0000
MA(24)	0,784613	0,000135	5798,322	0.0000
R-squared	0,886645	Mean dependent var	582,0016	
Adjusted R-squared	0,881033	S.D. dependent var	117,8302	
S.E. of regression	40,64151	Akaike info criterion	10,30190	
Sum squared resid	166825,0	Schwarz criterion	10,45178	
Log likelihood	-545,1515	F-statistic	158,0008	
Durbin-Watson stat	2,006840	Prob(F-statistic)	0,000000	

Las inversas correspondientes son: $1/38,45640 = 0,026003$ y $1/40,64151 = 0,024605$. Los resultados comparativos para las proyecciones del año 2002 realizadas al cierre del año 2001 y en los trimestre siguientes son, para cada uno de los modelos:

Tabla XVII
Comparación de modelos 4

	Real	Series ARIMA	Error %	Causal	Error %	Combinación	Error %
2001	7486,0	7109,0	-5,0	8233,9	10,0	7655,9	2,3
1 trim.	7486,0	6650,2	-11,2	7903,2	5,6	7259,4	-3,0
2 trim.	7486,0	7154,8	-4,4	7568,3	1,1	7355,8	-1,7
3 trim.	7486,0	7515,8	0,4	7653,6	2,2	7582,8	1,3

Los pronósticos del modelo combinado se obtuvieron de la siguiente manera:

$$Combinación = \frac{pronóstico_ARIMA * (1/\sigma_{ARIMA}) + pronóstico_causal * (1/\sigma_{CAUSAL})}{(1/\sigma_{ARIMA} + 1/\sigma_{CAUSAL})}$$

Puede observarse que el error porcentual producto de la combinación de pronósticos, tiene un menor intervalo de variación que el de los modelos

individuales: 1,3% a 2,3%. Los recorridos de las variables error porcentual en términos absolutos son: 10,8 en el modelo ARIMA, 8,9 en el modelo Causal, y sólo 1,0 en el modelo combinado. Respecto a los valores de MAE y MAPE, sus valores son los siguientes:

Tabla XVIII
Comparación de modelos 5

	MAE	MAPE
Serie ARIMA	43,4	7,5
Causal	66,2	10,9
Combinado	35,3	5,9

El modelo combinado, tiene también aquí un mejor desempeño que los modelos tomados de manera individual 35,3 millones de dólares de error absoluto promedio y 5,9% de error medio porcentual de los errores absolutos. En vista de estas ventajas observadas para el modelo combinado, se decide optar por este para la generación de los futuros pronósticos de la variable importaciones.

3.5 Pronósticos para el año 2003

Para el año 2003 los pronósticos resultantes de la aplicación del modelo de series de tiempo ARIMA y el modelo causal, son los que se presentan a continuación (véase la Tabla XIX, página 127):

Tabla XIX
Proyección del año 2003, Modelo ARIMA y Causal

Proyección 1trim. 2003	Proyección 1trim. 2003
717,8	717,8
630,4	630,4
723,8	723,8
748,2	795,5
714,0	818,2
636,5	695,3
679,2	764,1
669,3	708,2
680,4	707,6
689,2	645,9
692,7	695,3
700,3	698,7
8281,7	8600,8

Estas proyecciones para el año 2003, tienen ya las cifras ejecutadas del primer trimestre de ese año; además, corresponden las obtenidas al cierre del año 2002 y no a las que se obtuvieron en la actualización del primer trimestre; la razón es que las cifras de importación registradas para este periodo fueron mayores a las usuales debido al conflicto entre Estados Unidos e Irak lo que condujo a una alta importación de combustibles y que afecto a las actualizaciones del primer trimestre, ocasionando cifras altas para el resto del año.

Las proyecciones correspondientes al modelo causal son aquellas que se basan en un comportamiento estable para el índice del tipo de cambio real multilateral, ya que esta es la cifra que parece estar más de acuerdo a escenario económico en lo que resta del año. Pues según los especialistas se espera un menor ritmo

de crecimiento para las importaciones a partir del segundo trimestre, acentuándose la desaceleración en los dos últimos trimestres del año.

Luego con las proyecciones de ambos modelos se procede finalmente a obtener la proyección que combina ambos modelos (véase la Tabla XX)

Tabla XX
Proyección del año 2003, Modelo Combinado

Proyección 1 trim 2003	Proyección Trimestral 2003	
717,8	Trim1	2072,0
630,4	Trim2	2200,9
723,8	Trim3	2102,3
771,2	Trim4	2061,6
764,6	Total	8436,8
665,1		
720,4		
688,2		
693,6		
668,2		
693,9		
699,5		
8436,8		

Conclusiones

- La proyección más adecuada de las importaciones, se alcanzo con el uso combinado de la metodología causal y de series de tiempo; análisis de regresión y modelos ARIMA respectivamente.
- La combinación de las proyecciones tuvo como resultado menores niveles de error en los periodos de evaluación. Se reproduce aquí la tabla XVII que corresponde a las proyecciones para el año 2002 al cierre del año 2001 y las actualizaciones trimestrales respectivas; y la tabla XVIII que se corresponde con los pronósticos del año 2002 en base a cifras del 2001.

Tabla XVII

Comparación de modelos 4

	Real	Serie ARIMA	Error %	Causal	Error %	Combinación	Error %
2001	7486,0	7109,0	-5,0	8233,9	10,0	7655,9	2,3
1 trim.	7486,0	6650,2	-11,2	7903,2	5,6	7259,4	-3,0
2 trim.	7486,0	7154,8	-4,4	7568,3	1,1	7355,8	-1,7
3 trim.	7486,0	7515,8	0,4	7653,6	2,2	7582,8	1,3

Tabla XVIII

Comparación de modelos 5

	MAE	MAPE
Serie ARIMA	43,4	7,5
Causal	66,2	10,9
Combinado	35,3	5,9

- Las proyecciones para el año 2003 son las siguientes:

Proyección Trimestral 2003	
Trim1	2072,0
Trim2	2200,9
Trim3	2102,3
Trim4	2061,6
Total	8436,8

y si se toma en cuenta que el máximo porcentaje de error encontrado en las proyecciones trimestrales para el modelo combinado en el período de evaluación es de 3%, entonces se esperaría que las proyecciones estén en el intervalo 8191,1-8697,7 millones de dólares.

- Como punto final algunas recomendaciones de carácter general
 - Tenga en cuenta que un modelo es una aproximación de la realidad.
 - De disponerse de información relevante no considerada en el modelo debe de evaluarse el impacto de ésta en los pronósticos.
 - El modelo puede cambiar en el futuro, por lo que es necesario el monitoreo y evaluación periódica.

Bibliografía

BOX, GEORGE E. P. ; JENKINS, GWILYN M. (1976)

Time Series Analysis, Forecasting and Control, Revised Edition

Estados Unidos de Norteamérica; Ed. Holden Day, San Francisco.

BROCKWELL, PETER J.; DAVIS, RICHARD A. (1991)

Time Series: Theory and Methods, Second Edition

Estados Unidos de Norteamérica; Ed. Springer-Verlag New York.

DIEBOLD, FRANCIS X. (1999)

Elementos de Pronósticos

México, International Thomson Editores

FERRÁN ARANAZ, MAGDALENA (1996)

SPSS para Windows, Programación y Análisis Estadístico

España, McGraw – Hill / Interamericana de España.

GREENE, WILLIAM H. (1999)

Análisis Econométrico, Tercera edición

España; Prentice Hall, Madrid.

GUJARATI, DAMODAR N. (1997)

Econometría, Tercera edición

Colombia; McGraw-Hill, Santafé de Bogotá.

MONTGOMERY, DOUGLAS C; JOHNSON, LYNWOOD A. (1976)

Forecasting and Time Series Analysis

Estados Unidos de Norteamérica; McGraw-Hill, New York.

NORUSIS, MARIJA J. (1993)

SPSS for Windows, Base System User´s Guide. Release 6.0

Estados Unidos de Norteamérica; SPSS INC.

SIGAD, Sistema Integrado de Gestión Aduanera (1992).

Superintendencia Nacional de Aduanas, ADUANAS

SPSS INC. TRAINING DEPARTMENT (1999)

Time Series Analysis and Forecasting with SPSS Trends 9.0

Estados Unidos de Norteamérica; SPSS INC. (1999).

URIEL, EZEQUIEL; PEIRO , AMADO (2000)

Introducción al Análisis de Series Temporales

España; Editorial AC, Madrid

Datos para el ajuste de los modelos

CIF : en millones de US \$

I_PBI : Índice del PBI

I_TCRM : Índice del tipo de cambio real multilateral (cifras de abril-diciembre del 2003 bajo suposición de estancamiento)

ARANC : Arancel promedio

Año	Meses	CIF	I_PBI	I_TCRM	ARANC
1 993	Ene	240,3	79,4	104,8	16,3
	Feb	297,2	79,7	105,4	16,3
	Mar	311,1	84,4	106,3	16,3
	Abr	308,3	86,8	107,9	16,3
	May	297,7	90,9	108,5	16,3
	Jun	307,2	93,9	108,3	16,3
	Jul	325,8	90,1	107,1	16,3
	Ago	359,4	90,1	106,7	16,3
	Sep	407,6	89,1	107,2	16,3
	Oct	339,9	89,1	107,3	16,3
	Nov	394,4	92,8	106,5	16,3
	Dic	435,2	97,2	103,3	16,3
1 994	Ene	374,0	89,1	100,8	16,3
	Feb	343,4	84,5	100,1	16,3
	Mar	407,3	95,3	98,8	16,3
	Abr	392,3	99,5	98,2	16,3
	May	457,5	103,4	98,5	16,3
	Jun	408,9	103,0	98,0	16,3
	Jul	423,4	101,4	100,3	16,3
	Ago	510,5	101,6	101,4	16,3
	Sep	496,2	100,2	103,5	16,3
	Oct	505,0	103,6	103,5	16,3
	Nov	563,1	106,9	100,7	16,3
	Dic	553,2	111,7	96,5	16,3
1 995	Ene	557,0	105,5	98,3	16,3
	Feb	522,2	99,7	99,5	16,3
	Mar	662,4	108,1	101,6	16,3
	Abr	564,7	106,9	103,2	16,3
	May	669,2	119,0	102,7	16,3
	Jun	627,9	112,5	102,6	16,3
	Jul	660,9	110,7	101,7	16,3
	Ago	719,8	109,6	99,1	16,3
	Sep	645,5	104,5	98,4	16,3
	Oct	637,4	107,7	98,8	16,3

	Nov	699,6	109,6	99,8	16,3
	Dic	617,2	109,3	98,9	16,3
1 1996	Ene	635,9	106,2	98,4	16,3
	Feb	540,4	102,2	97,8	16,3
	Mar	632,6	108,1	96,8	16,3
	Abr	622,7	109,8	96,2	16,3
	May	726,5	120,4	97,2	16,3
	Jun	595,8	117,0	98,1	16,3
	Jul	697,8	114,6	97,5	16,3
	Ago	690,1	110,8	98,3	16,3
	Sep	638,6	106,9	98,9	16,3
	Oct	727,3	110,7	100,6	16,3
	Nov	643,5	113,0	102,0	16,3
	Dic	622,8	116,0	100,3	16,3
1 1997	Ene	697,8	115,6	100,6	13,5
	Feb	533,6	108,8	99,8	13,5
	Mar	624,0	113,5	98,2	13,5
	Abr	739,7	125,6	98,8	13,5
	May	680,0	128,5	98,9	13,5
	Jun	640,8	122,6	97,9	13,5
	Jul	719,4	118,4	96,5	13,5
	Ago	796,4	116,9	95,5	13,5
	Sep	750,4	117,0	95,3	13,5
	Oct	748,1	119,5	96,2	13,5
	Nov	710,4	117,0	97,4	13,5
	Dic	700,7	122,2	94,6	13,5
1 1998	Ene	697,3	115,0	93,6	13,5
	Feb	661,2	112,7	95,0	13,5
	Mar	709,7	119,1	94,2	13,5
	Abr	760,8	121,8	94,5	13,5
	May	677,4	122,7	94,8	13,5
	Jun	711,3	122,0	95,7	13,5
	Jul	689,0	120,6	95,7	13,5
	Ago	698,4	118,0	96,3	13,5
	Sep	629,6	116,2	100,8	13,5
	Oct	610,1	116,0	103,8	13,5
	Nov	619,5	116,3	105,0	13,5
	Dic	577,8	117,8	106,5	13,5
1 1999	Ene	507,8	113,4	109,0	13,5
	Feb	484,6	111,9	110,7	13,5
	Mar	556,6	120,0	108,8	13,5
	Abr	555,1	119,0	108,3	13,5
	May	517,0	127,5	106,9	13,5
	Jun	537,9	124,4	105,9	13,5
	Jul	512,3	118,3	104,8	13,5
	Ago	567,6	115,0	107,0	13,5
	Sep	600,7	116,0	108,6	13,5
	Oct	591,3	119,9	111,0	13,5
	Nov	632,6	122,3	110,6	13,5
	Dic	654,3	123,8	110,9	13,5

2 000	Ene	584,6	119,3	111,7	13,5
	Feb	565,7	117,3	109,2	13,5
	Mar	624,7	127,4	108,6	13,5
	Abr	561,4	125,0	109,0	13,5
	May	640,6	136,4	107,5	13,5
	Jun	621,5	130,3	108,3	13,5
	Jul	604,8	124,6	107,4	13,5
	Ago	629,8	121,5	106,0	13,5
	Sep	561,4	114,7	105,1	13,5
	Oct	699,4	120,4	104,7	13,5
	Nov	647,4	119,9	105,6	13,5
	Dic	603,1	119,5	105,2	13,5
2 001	Ene	589,5	117,4	105,8	11,8
	Feb	638,1	114,1	105,3	11,8
	Mar	605,9	122,7	103,5	11,8
	Abr	583,7	125,9	104,3	11,8
	May	657,1	136,7	105,3	11,8
	Jun	527,7	128,2	102,6	11,8
	Jul	636,8	124,0	101,0	11,8
	Ago	637,1	123,0	102,2	11,8
	Sep	596,1	117,8	102,2	11,8
	Oct	620,7	123,5	100,6	11,8
	Nov	589,0	122,0	100,5	11,8
	Dic	570,5	124,1	100,6	11,8
2 002	Ene	580,1	123,2	100,3	11,8
	Feb	545,8	117,9	99,2	11,9
	Mar	514,4	122,2	98,4	11,9
	Abr	651,5	139,9	98,1	11,0
	May	661,6	144,0	98,5	11,0
	Jun	552,2	132,9	99,4	11,0
	Jul	678,9	131,7	101,7	11,0
	Ago	624,6	127,8	101,7	11,0
	Sep	682,1	127,3	101,9	10,9
	Oct	661,1	129,9	99,9	10,8
	Nov	644,9	131,4	101,0	10,8
	Dic	688,7	134,1	99,9	10,8
2 003	Ene	717,8	128,9	100,6	10,8
	Feb	630,4	124,2	100,1	10,8
	Mar	723,8	126,8	99,8	10,8
	Abr		145,2	100,0	10,8
	May		149,5	100,7	10,8
	Jun		138,0	99,8	10,8
	Jul		136,7	100,1	10,8
	Ago		132,7	100,7	10,8
	Sep		132,1	100,9	10,8
	Oct		134,8	99,0	10,8
	Nov		136,4	99,5	10,8
	Dic		139,2	99,0	10,8

