

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA ECONÓMICA Y CIENCIAS SOCIALES



**“MODELAMIENTO Y PREDICCIÓN DE LA RECAUDACIÓN DE LOS INGRESOS
CORRIENTES DE LA UNI A TRAVÉS DE UN MODELO ARIMA”**

**INFORME DE SUFICIENCIA
PARA OPTAR EL TÍTULO PROFESIONAL DE:**

LICENCIADO EN ESTADÍSTICA

POR LA MODALIDAD DE ACTUALIZACIÓN DE CONOCIMIENTOS

ELABORADO POR:

FERNANDO LUCANA JARAMILLO

LIMA – PERU

2004

Dedico este trabajo a mis padres Justiniano y Encarnación, quienes con cariño y esfuerzo siempre me dieron lo mejor de ellos.

A mi esposa Antonieta y a mi hija Iris Maribel, porque son parte de mi vida.

A mis hermanos Javier y Erika Liz, que han sido siempre un espejo en el que su hermano mayor ha tratado de mirarse.

A la memoria de mi hermana Maribel Jesús

Mis agradecimientos es a Dios por todos mis logros alcanzados.

A todos mis amigos y profesores quienes siempre se preocuparon por el estado del desarrollo de este presente trabajo, en especial a mi amigo Edgard.

I. CURRÍCULUM VITAE

Enero 2003 **UNI – Oficina Central de Planificación** **Lima, Perú**
- Curso de Planeamiento Estratégico, en calidad de Organizador y Asistente.

Febrero 2003 **UNI – Oficina Central de Planificación** **Lima, Perú**
- Primer, Segundo y Tercer Taller de Planeamiento Estratégico Institucional UNI 2003 - 2007, en calidad de Organizador y Asistente.

Setiembre 1997 y Marzo 2001 **UNI – CENCA – Centro de Computo FIIS** **Lima, Perú**
- Curso “Ensamblaje de Computadoras”
- Curso Taller “ Microsoft Front Page 2000”

Otras Referencias:

- Conocimiento del Sistema Operativo DOS, Word Perfect 6.0, Qpro 5.0, FoxPro 2.6
- Entorno Windows: Microsoft Office (Excel, Word, Power Point y Front Page) y Microsoft Internet Explorer
- Conocimientos del Lenguaje de Programación Turbo Pascal.
- Manejo de Software Estadísticos: Statgraphics 7.0, SPSS 11.0, TSP 7.0, Stackpack, Microstad, EViews 3.1, MINITAB 12.

Seminarios:

Agosto 1995 **UNI – Facultad de Ingeniería Económica y CC.SS.** **Lima, Perú**
- Seminario “Modelización de Series Económicas”, Universidad Nacional de Ingeniería – Universidad de Valladolid – España, Departamento de Economía Aplicada

Agosto 1995 **UNI – Facultad de Ingeniería Económica y CC.SS.** **Lima, Perú**
- Seminario Multidisciplinario “Análisis Económico Espacial y su Aplicación al Desarrollo Municipal y Regional”

Setiembre 1995 **UNI – Facultad de Ingeniería Económica y CC.SS.** **Lima, Perú**
- Seminario “Técnicas Multivariantes en la Investigación Comercial”

Enero 1997 **Instituto Peruano de Ciencias Económicas** **Lima, Perú**
- Seminario “El Proceso Presupuestario y las Operaciones de Tesorería para el Ejercicio fiscal 1997”

Marzo 2002 **UNI – Centro de Capacitación Administrativa CENCA** **Lima, Perú**
- Seminario “Gestión Administrativa en la Universidad y en la Administración Pública”

Noviembre 2003 **UNI – Oficina Central de Planificación** **Lima, Perú**
- Seminario: Proyectos de Inversión en el marco del Sistema Nacional de Inversión Pública

Reconocimiento:

Reconocimiento De " Honor Al Mérito En La Especialidad De Estadística Por La Universidad Nacional De Ingeniería (1995) "

II. INFORME DE SUFICIENCIA

ÍNDICE GENERAL

	Pág.
INTRODUCCIÓN	xiii
ANTECEDENTES	xv
OBJETIVOS	xvii
CAPÍTULO 1	
MARCO TEÓRICO	1
1.1 Introducción al Análisis Univariante de Series Temporales	1
1.2 Proceso Estocástico Univariante	2
1.2.1 Estacionariedad. (Proceso Estacionario)	4
1.2.2 Ergodicidad	4
1.3 Autocorrelación	4
1.3.1 Función de Autocorrelación (FAC). (Correlograma)	5
1.3.2 Función de Autocorrelación Parcial (FACP). (Correlograma)	5
1.4 Procesos Estacionarios y Ergódicos	6
1.4.1 Proceso Autorregresivo AR(p)	6
1.4.2 Proceso de Media Móvil MA(q)	7
1.4.3 Proceso Autorregresivo de Medias Móviles ARMA(p,q)	8

1.4.4	Procesos Autorregresivos Integrados de Medias Móviles	
	ARIMA(p,d,q)	9
1.4.5	Proceso Estacional	9
1.4.6	Proceso Autorregresivo Integrado de Medias Móviles	
	Estacional ARIMA(p,d,q)x(P,D,Q)s	10
1.5	Esquema General de la Metodología Box-Jenkins	11
1.6	Identificación	12
1.6.1	Análisis de la Estabilidad en Varianzas y Transformaciones para conseguir Homocedasticidad	13
1.6.2	Análisis de la Estabilidad en Medias y de la Estacionalidad	15
1.6.2.1	Identificación de d	16
1.6.2.2	Instrumentos que permiten identificar d	16
1.6.2.3	Procedimiento para determinar d	16
1.6.2.4	Identificación de D	18
1.6.3	Identificación de los Procesos Estocásticos Estacionarios subyacentes a Series	18
1.6.3.1	Identificación de Modelos Estacionarios	19
1.6.3.2	Identificación del Modelo Estacional Multiplicativo General	21
1.6.4	Identificación del Término Independiente	21
1.7	Estimación de Parámetros	23
1.8	Validación de los Supuestos del Modelo	25
1.8.1	Análisis de Coeficientes	26

1.8.2	Análisis de Residuales	28
1.8.2.1	Media Cero	28
1.8.2.2	Varianza Constante	29
1.8.2.3	Incorrelación	29
1.8.2.4	Distribución Normal de los Residuales	31
1.8.3	Sobreajuste	31
1.8.4	Bondad de Ajuste	32
1.8.4.1	Línea Verificacionista	32
1.8.4.2	Línea Preferencialista	32
1.8.5	Análisis de Estabilidad	34
1.9	Predicción	35
1.9.1	Predicción por Intervalos	36
1.9.2	Indicadores no Paramétricos del Grado de Exactitud	37
CAPÍTULO 2		
METODOLOGÍA Y CARACTERIZACIÓN		38
2.1	Metodología usada para el Modelamiento y Predicción de la Recaudación de Ingresos Corrientes de la UNI	38
2.2	Caracterización de la Variable	41
CAPÍTULO 3		
RESULTADOS		43
3.1	Modelación y Predicción de la Recaudación de los Ingresos Corrientes de la UNI a través de un Modelo ARIMA por la Fuente	

de Financiamiento de Recursos Directamente Recaudados	43
3.2 Fase de Identificación	45
3.2.1 Estabilización de la Varianza y generación de la serie transformada para conseguir Homocedasticidad	46
3.2.2 Estabilidad en Medias y de la Estacionalidad	50
3.2.3 Identificación del Modelo Generador	56
3.3 Fase de Estimación y Validación de los modelos	59
3.3.1 Sobreajuste del Modelo	60
3.3.2 Análisis de los Residuos del Modelo	63
3.4 Fase de Predicción	67
CONCLUSIONES	72
RECOMENDACIONES	74
APÉNDICE 1	76
AP.1: Deflación de Series en el Tiempo	76
AP.2 Fórmula de Cálculo	78
ANEXOS	
Anexo 1: Desagregado de la Variable Recaudación de Ingresos Corrientes	79
Anexo 2: Índice de Precios Promedio Mensual al Consumidor de Lima Metropolitana: 1993 – 2003 y de Base Diciembre Año 2001=100.00)	80

Anexo 3:	Recaudación de Ingresos Corrientes de la UNI (En Nuevos Soles Corrientes)	81
Anexo 4:	Recaudación de Ingresos Corrientes de la UNI (En Nuevos Soles Constantes)	82
Anexo 5:	Recaudación de Ingresos Corrientes de las Principales Dependencias de la UNI	83
Anexo 6:	Recaudación de Ingresos Corrientes de las Facultades de la UNI	84
	BIBLIOGRAFÍA	85
	GLOSARIO DE TÉRMINOS	87

ÍNDICE DE TABLAS

	Pág.
Tabla 1.1: Características de los Modelos Estacionarios vs. FAC y FACP	20
Tabla 1.2: Características de los Modelos Estacionales vs. FAC y FACP	21
Tabla 1.3: Enfoque Condicional y no Condicional en la estimación de un modelo ARMA(p, q)	24
Tabla 3.1: Estadísticas básicas de la Serie Histórica	44
Tabla 3.2: Prueba de Homogeneidad de Varianza de Z_t	46
Tabla 3.3: Prueba de Homogeneidad de Varianza de W_t	48
Tabla 3.4: Estadística Descriptiva de las series transformadas con diferencia estacionaria, doble diferencia estacionaria, diferencia estacional, diferencia estacionaria y estacional	52
Tabla 3.5: Prueba de Homogeneidad de Varianza de la diferencia estacional W_t	54
Tabla 3.6: Orden de diferenciación de la serie W_t	55
Tabla 3.7: Coeficientes del Modelo ARIMA(2,0,2) x ARIMA(1,1,1) ₁₂	60
Tabla 3.8: Coeficientes del Modelo ARIMA(2,0,3) x ARIMA(1,1,1) ₁₂	62
Tabla 3.9: Matriz de Correlación de los Coeficientes Estimados	63
Tabla 3.10: Estadística Descriptiva de la Serie Residual	64

Tabla 3.11: Prueba de la Homogeneidad de la Varianza de la Serie Residual	65
Tabla 3.12: Error de Predicción de la diferencia estacional de la serie W_t	69
Tabla 3.13: Pronóstico, Actual e Intervalos de Confianza de enero a junio del 2003	69
Tabla 3.14: Pronóstico, Serie Histórica e Intervalos de Confianza de enero a junio del 2003	69
Tabla 3.15: Error de Pronóstico de la serie Z_t	70
Tabla 3.16: Pronóstico, Serie Histórica e Intervalos de Confianza de enero a diciembre del 2003	71
Tabla A.5.1: Principales Dependencias Generadoras de Ingresos Corrientes	83
Tabla A.6.1: Facultades Generadoras de Ingresos Corrientes	84

ÍNDICE DE FIGURAS

	Pág.
Figura 3.1: Serie Histórica de la Recaudación de Ingresos Corrientes (En soles Constantes)	45
Figura 3.2: Gráfico de Cajas por Años de la Serie Histórica	47
Figura 3.3: Diagrama de Nivel y Dispersión de la Serie Histórica por Año	48
Figura 3.4: Diagrama de Nivel y Dispersión de la Serie W_t por Año	49
Figura 3.5: Serie Transformada de la Recaudación de Ingresos Corrientes	50
Figura 3.6: Correlograma de la serie transformada W_t	51
Figura 3.7: Diferencia Estacional de la serie transformada W_t	53
Figura 3.8: Diagrama de Nivel y Dispersión de la diferencia estacional de W_t	55
Figura 3.9: Función de Autocorrelación de la diferencia estacional de W_t	56
Figura 3.10: Función de Autocorrelación Parcial de la diferencia estacional de W_t	57
Figura 3.11: Función de Autocorrelación con retardos periódicos de la diferencia estacional de W_t	57
Figura 3.12: Función de Autocorrelación Parcial con retardos periódicos de la diferencia estacional de W_t	58

Figura 3.13: Función de Autocorrelación de la Serie Residual	66
Figura 3.14: Función de Autocorrelación Parcial de la Serie Residual	66
Figura 3.15: Serie Histórica, Pronóstico e Intervalos de Confianza de enero a diciembre 2003	71
Figura A.5.1: Principales Dependencias generadoras de Ingresos Corrientes	83
Figura A.6.1: Facultades Generadoras de Ingresos Corrientes	84

INTRODUCCIÓN

Toda institución, ya sea la familia, la empresa o el gobierno, tiene que hacer planes para el futuro si ha de sobrevivir y progresar. Hoy en día diversas instituciones requieren conocer el comportamiento futuro de ciertos fenómenos con el fin de planificar, prever y prevenir. En este contexto, la Universidad Nacional de Ingeniería en estos últimos años, como corresponde a una universidad moderna, es la primera productora de bienes y servicios entre las universidades públicas del país. A través de su política de generación de recursos propios, se ha involucrado con los requerimientos de la comunidad nacional.

El tema que se desarrollará es de interés práctico y partirá con los supuestos de un modelo estocástico ARIMA. Se podrá determinar un modelo explicativo del comportamiento de la Recaudación de los Ingresos Corrientes de la UNI y luego la predicción a través del modelo. Se utilizarán los datos de la **Recaudación de Ingresos Corrientes¹ por la Fuente de Financiamiento de Recursos Directamente Recaudados**, correspondientes a los años 1993 – 2003, de la Universidad Nacional de Ingeniería; los cuales se debe remarcar que son datos

¹ Véase el Anexo 1 para observar el desagregado de la variable Recaudación de Ingresos Corrientes.

históricos; primero porque son documentos en los que se encuentra registrada, en forma resumida la evolución financiera de la UNI y segundo porque son documentos conciliados en la Contaduría General de la República y la Dirección Nacional de Presupuesto Público del Ministerio de Economía y Finanzas, cumpliendo con las leyes vigentes de la Nación.

El primer capítulo contiene el marco teórico donde se presenta los conceptos acerca de procesos estocásticos univariantes, autocorrelación, procesos estacionarios y ergódicos, esquema general de la metodología Box-Jenkins donde se identifica, estima, valida el modelo y luego se utiliza para predecir.

Para el segundo capítulo se desarrolla la metodología para el modelamiento de la serie histórica y su respectiva caracterización. Por último el tercer capítulo contiene la modelación y predicción de la serie a través de un modelo ARIMA.

El modelo hallado, como resultado principal de este trabajo, permite obtener un pronóstico estadísticamente significativo, en base a información de una sola variable sin necesidad de generar relaciones de causalidad.

ANTECEDENTE

El análisis clásico de series temporales se basa en un modelo de descomposición que considera a toda serie temporal integrada por cuatro componentes básicas: tendencia, componente estacional, componente cíclica y componente irregular. Cada una de estas componentes se estudia por separado mediante el uso de métodos matemáticos.

El análisis cuantitativo de predicción de series temporales se puede realizar mediante análisis determinístico (medias móviles, alisados exponenciales, ajustes de tendencias) o análisis estocástico (ARIMA). En principio, ambas técnicas pueden ser aplicables, la elección de una u otra depende de las características particulares de cada serie (tendencia y estacionalidad), del tamaño muestral disponible de la serie, del horizonte temporal de predicción, de la complejidad del método y del grado de precisión con que se desea llevar a cabo dicho análisis. Bajo el criterio estrictamente matemático, interesan los métodos que reporten los errores de predicción más bajos.

Una serie temporal se caracteriza por poseer tanto tendencia como estacionalidad. Desde el punto de vista de la información estadística, la variable manifiesta sólo

tendencia cuando la serie histórica se contabiliza anualmente, mientras que, muestra tendencia y estacionalidad cuando dicha serie es contabilizada mensualmente. Para determinar el tipo de metodología predictiva a utilizar, fundamentalmente, ha de tenerse en cuenta el tamaño de la serie histórica.

Los resultados predictivos serán tanto mejores, cuanto más numerosos y mejor indicativos del carácter estacional y tendencia sean los datos de la serie histórica. En este sentido, las técnicas predictivas que mejor se ajustan a este tipo de series son los modelos estocásticos ARIMA, que será tema de este trabajo de investigación.

OBJETIVOS

Objetivo General

Obtener un modelo explicativo del comportamiento de la Recaudación de Ingresos Corrientes, por la Fuente de Financiamiento de Recursos Directamente Recaudados generados por la Universidad Nacional de Ingeniería durante los años correspondientes 1993 - 2002

Objetivos Específicos

Mediante los modelos estocásticos ARIMA:

Identificar, Estimar y Validar el modelo seleccionado.

Predecir mediante el modelo seleccionado la Recaudación de Ingresos Corrientes, por la Fuente de Financiamiento de Recursos Directamente Recaudados, para el año 2003 de la Universidad Nacional de Ingeniería.

CAPÍTULO 1

MARCO TEÓRICO

1.1 Introducción al Análisis Univariante de Series Temporales

El enfoque clásico del análisis de series temporales, basado en la descomposición, tiene a su favor el carácter intuitivo de un razonamiento original y la sencillez relativa de las técnicas e instrumentos que requiere.

Desde una perspectiva estocástica, una serie temporal observada se considera como realización (o muestra) de cierto proceso teórico, integrado por variables aleatorias referidas a momentos o periodos de tiempo.

Los objetivos centrales del análisis univariante de series temporales consiste en **describir** la evolución temporal observada de algunas características de interés de una unidad observable y **prever** su evolución futura (desconocida) a corto plazo, utilizando tan sólo información (conocida) sobre su pasado; dicha información conforma lo que se conoce como una serie temporal (univariante).

La Serie Temporal Univariante es una secuencia de N observaciones (datos numéricos) ordenadas y equidistantes cronológicamente sobre una característica de una única unidad observable en diferentes momentos o fechas. La serie Temporal univariante suele representarse como: z_1, z_2, \dots, z_N ; como un vector columna $z \equiv (z_1, z_2, \dots, z_N)'$; o bien como $(z_t)_{t=1}^N$, donde z_t es la observación número t ($1 \leq t \leq N$) de la serie y N es el número de observaciones de que consta la serie completa (el tamaño o la longitud de la serie).

Como hipótesis de trabajo se considera que una serie temporal es una realización particular de una muestra procedente de un proceso estocástico (univariante):

1.2 Proceso Estocástico Univariante

Un Proceso Estocástico Univariante es la secuencia de variables aleatorias ordenadas y equidistantes cronológicamente referidas a una característica de una única unidad observable en diferentes momentos o fechas. Se representa como: $(Z_t)_{t=0, \pm 1, \pm 2, \dots}$, donde Z_t es una variable aleatoria referida a la característica de la unidad observable considerada en el momento t .

Una **muestra** de tamaño o longitud N (finito) procedente de un proceso estocástico (Z_t) es un subconjunto de N componentes consecutivos de (Z_t)

$$\begin{array}{ll}
 \text{Proceso estocástico:} & \dots, Z_{-1}, Z_0, Z_1, Z_2, \dots, Z_N, Z_{N+1}, \dots \\
 \text{Muestra:} & Z_1, Z_2, \dots, Z_N \qquad (1-1)
 \end{array}$$

Serie temporal:

$$z_1, z_2, \dots, z_N$$

Entonces el análisis estocástico de la serie temporal consiste en realizar una inferencia estadística sobre las propiedades del **proceso teórico** a partir de la información contenida en la serie observada. Por tanto, hay un elevado grado de paralelismo entre el análisis estocástico de series temporales y el análisis estadístico general.

Tras el análisis inicial y la descripción de los datos, en que intervienen las técnicas habituales en estadística y algunas específicas de series temporales, **el objetivo** del análisis se centra en la construcción de un modelo que reproduzca adecuadamente las propiedades del proceso que se supone ha originado los datos.

Conocer el **proceso teórico** implica; conocer la función de distribución conjunta del vector de variables aunque, bajo normalidad, basta con su vector de medias y su matriz de varianzas-covarianzas.

Función de medias $\mu_t = E[z_t]$

Función de Varianzas $\sigma_t = Var[z_t]$

Función de autocovarianzas $Cov(t, t + j) = E[(z_t - \mu_t) \cdot (z_{t+j} - \mu_{t+j})]$ (1-2)

Función de autocorrelación $\rho_{(t,t+j)} = Cov(t, t + j) / \sigma_t \sigma_{t+j}$

Estos elementos pueden ser inferidos a partir de las observaciones pero solo cuando se cumple una serie de condiciones.

Las condiciones que deben verificarse para que la inferencia a partir de una única realización sea posible son dos:

1.2.1 Estacionariedad. (Proceso Estacionario)

Implica que las variables integrantes del proceso tienen media y varianza constantes y finitas, y que la covarianza entre pares de ellas solo depende de su separación temporal.

$$\mu_t = E[z_t] = Cte.$$

$$\sigma_t = Var[z_t] = Cte. \quad (1-3)$$

$$Cov(t, t + k) = Cov(t, t - k) = \gamma_k$$

$$\rho_k = \gamma_k / \gamma_0 \quad \text{Donde } \gamma_0 = \sigma^2$$

1.2.2 Ergodicidad.

Implica que la covarianza entre pares de variables del proceso tiende a reducirse cuanto mayor es su separación temporal.

1.3 Autocorrelación

Cuando el proceso es estacionario y ergódico, la media y la varianza constantes reflejan sus características estáticas (nivel y variabilidad). Las características dinámicas, la manera en que cada variable se ve afectada por las variables previas, aparecen recogidas por las covarianzas que solo dependen de la separación temporal.

El valor de la correlación para sucesivos valores del retardo k proporciona la llamada función de autocorrelación. La función de autocorrelación expresa las características dinámicas del proceso, porque recoge la influencia del pasado en el presente ($k=1,2,3\dots$). Así, un proceso en el cual cada variable dependa sólo de la anterior tendrá nulos todos los coeficientes de autocorrelación excepto el primero. Si cada variable depende de las dos previas, entonces la autocorrelación será no nula para los ordenes uno y dos, y nula para órdenes superiores. En otras palabras, la función de autocorrelación refleja la memoria del proceso: el número de periodos durante los cuales una variable continua teniendo influencia en la evolución del proceso.

La autocorrelación, tanto total como parcial, puede ser estimada a partir de las covarianzas de los datos. La autocorrelación estimada sirve para inferir los ordenes de retardos que son significativos en el estudio de una serie temporal, es decir, los ordenes de retardos implicados en el proceso teórico que subyace a los datos. El modelo que se elija para representar el proceso deberá incluir, precisamente, los órdenes de retardo correspondientes a coeficientes de autocorrelación significativos.

1.3.1 Función de Autocorrelación (FAC). (Correlograma)

Es la representación de los coeficientes de autocorrelación en función del retardo.

1.3.2 Función de Autocorrelación Parcial (FACP). (Correlograma)

Es la representación de los coeficientes de autocorrelación parcial en función del retardo.

1.4 Procesos Estacionarios y Ergódicos

Los procesos lineales estacionarios y ergódicos pueden ser representados mediante un modelo de la clase **ARIMA**. Las siglas corresponden a autorregresivos, integrados y de media móvil.

Un tipo de proceso estacionario particular es el denominado **ruido blanco** (a_t) si:

$$\begin{aligned} E(a_t) &= 0 \quad \forall t \\ V(a_t) &= \sigma^2 \quad \forall t \\ \text{Corr}(a_t, a_{t-k}) &= \rho_k = 0, \quad k = \pm 1, \pm 2, \dots, \quad \forall t \\ a_t, &\text{ sigue una distribución Normal } \forall t \end{aligned} \tag{1-4}$$

En estos procesos conocer los valores pasados no proporciona ninguna información sobre el futuro). El ruido blanco interviene en la formulación de cualquier modelo ARIMA.

Un proceso estocástico se dice ergódico cuando los valores de la serie alejados en el tiempo están poco correlacionados, es decir, cuando el coeficiente de autocorrelación simple decrece al aumentar el retardo k .

1.4.1 Proceso Autorregresivo AR(p)

Un proceso $\{Z_t\}$ es autorregresivo de orden p si:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t + c \tag{1-5}$$

donde $\{a_t\}$ es un proceso de ruido blanco con a_t independiente de Z_{t-h} ,
 $\forall h \geq 1, \quad \forall t$

Si L es el operador de retardo:

$$LZ_t = Z_{t-1}, \dots, L^k Z_t = Z_{t-k} \quad (1-6)$$

La ecuación que resulta es:

$$\phi_p(L)Z_t = a_t + c \quad (1-7)$$

Donde:

$$\phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p \quad (1-8)$$

Un proceso estacionario y autorregresivo AR(p) se caracteriza porque los p primeros coeficientes de la FACP son no nulos y el resto cero, y la FAC, en general, presenta muchos coeficientes no nulos que decrecen con el retardo con mezcla de exponenciales y sinusoidales.

Un proceso autorregresivo AR(p) es estacionario si las raíces del polinomio en $\phi_p(L)$ caen fuera del círculo unitario.

1.4.2 Proceso de Media Móvil MA(q)

Un Proceso $\{Z_t\}$ es una media móvil de orden q si:

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + c \quad (1-9)$$

donde $\{a_t\}$ es un proceso de ruido blanco. Considerando el operador retardo L , la ecuación que resulta es:

$$Z_t = \theta_q(L)a_t + c \quad (1-10)$$

donde:

$$\theta_q(L) = 1 - \theta_1 L - \dots - \theta_q L^q \quad (1-11)$$

Todo proceso de media móvil es estacionaria. Un proceso de media móvil MA(q) se caracteriza porque los q primeros coeficientes de la FAC son no nulos y el resto cero, y la FACP, en general, presentan muchos coeficientes no nulos que decrecen con el retardo con mezcla de exponenciales y sinusoidales.

Un proceso de medias móviles MA(q) es invertible si las raíces del polinomio en L definido por $\theta_q(L)$ caen fuera del círculo unidad.

1.4.3 Proceso Autorregresivo de Medias Móviles ARMA(p,q)

Un proceso $\{Z_t\}$ es autorregresivo de medias móviles de órdenes p y q si:

$$Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + c \quad (1-12)$$

donde $\{a_t\}$ es un proceso de ruido blanco con a_t independiente de las Z_t .

Considerando el operador retardo L , la ecuación que resulta es:

$$\phi_p(L)Z_t = \theta_q(L)a_t + c \quad (1-13)$$

donde:

$$\phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p \quad (1-14)$$

$$\theta_q(L) = 1 - \theta_1 L - \dots - \theta_q L^q \quad (1-15)$$

Dado que los procesos ARMA son una mezcla de procesos AR y procesos MA, la FAC y la FACP de un proceso estacionario ARMA(p,q) será una superposición de las FAC y FACP de los procesos AR(p) y MA(q) correspondientes.

La condición de estacionariedad es que las raíces del polinomio $\phi_p(L)$ estén fuera del círculo unidad.

La condición de invertibilidad es que las raíces del polinomio $\theta_q(L)$ estén fuera del círculo unidad.

1.4.4 Procesos Autorregresivos Integrados de Medias Móviles ARIMA(p,d,q)

Un proceso $\{Z_t\}$ es autorregresivo integrado de medias móviles de órdenes p, d y q si:

$$(1 - \phi_1 L - \dots - \phi_p L^p) \Delta^d Z_t = (1 - \theta_1 L - \dots - \theta_q L^q) a_t + c \quad (1-16)$$

donde $\{a_t\}$ es un proceso de ruido blanco y Δ es el operador de diferencia:

$$\Delta Z_t = (1 - L)Z_t = Z_t - Z_{t-1}, \dots, \Delta^d Z_t = (1 - L)^d Z_t = (1 - L)^{d-1} (Z_t - Z_{t-1}) \quad (1-17)$$

Equivalentemente:

$$\phi_p(L)W_t = \theta_q(L)a_t + c \quad \text{donde: } W_t = \Delta^d Z_t \quad (1-18)$$

Luego el proceso ARIMA(p,d,q) para $\{Z_t\}$ es equivalente al proceso ARMA(p,q) para W_t . Si el proceso $\{W_t\}$ es estacionario el proceso ARIMA para Z_t quedará caracterizada por la FAC del proceso $\{W_t\}$ y por d, el orden de diferenciación del proceso original $\{Z_t\}$.

1.4.5 Proceso Estacional

Un tipo especial de procesos no estacionarios son los estacionales. Un proceso $\{Z_t\}$ será estacional cuando presente una pauta regular de comportamiento periódico. Si la pauta estacional es constante y con periodo s, en general aplicando el operador:

$$(1 - L^s)Z_t = Z_t - Z_{t-s} \quad (1-19)$$

la estacionalidad podrá ser eliminada. En ocasiones será necesaria más de una diferencia estacional para desestacionalizar el proceso:

$$(1-L^s)^D Z_t = (1-L^s)^{D-1} (1-L^s) Z_t = (1-L^s)^{D-1} (Z_t - Z_{t-s}) \quad (1-20)$$

siendo D el orden de diferenciación estacional.

1.4.6 Proceso Autorregresivo Integrado de Medias Móviles Estacional ARIMA(p,d,q)x(P,D,Q)_s

Un proceso estacional $\{Z_t\}$, con periodo s conocido, es un proceso autorregresivo integrado de medias móviles de órdenes p, d, q, P, D y Q si:

$$\Phi_p(L^s)\phi_p(L)(1-L)^d(1-L^s)^D Z_t = \theta_q(L)\Theta_Q(L^s)a_t + c \quad (1-21)$$

donde $\{a_t\}$ es un proceso de ruido blanco, d y D son los órdenes de diferenciación regular y estacional, respectivamente, y

$$\begin{aligned} \Phi_p(L^s) &= 1 - \Phi_1 L^s - \dots - \Phi_p L^{ps} \\ \phi_p(L) &= 1 - \phi_1 L - \dots - \phi_p L^p \\ \theta_q(L) &= 1 - \theta_1 L - \dots - \theta_q L^q \\ \Theta_Q(L^s) &= 1 - \Theta_1 L^s - \dots - \Theta_Q L^{Qs} \end{aligned} \quad (1-22)$$

Equivalentemente:

$$\Phi_p(L^s)\phi_p(L)W_t = \theta_q(L)\Theta_Q(L^s)a_t + c \quad (1-23)$$

donde:

$$W_t = (1-L)^d(1-L^s)^D Z_t \quad (1-24)$$

Luego el proceso ARIMA estacional para Z_t , es equivalente al proceso ARMA estacional para W_t . Si el proceso $\{W_t\}$ es estacionario, el proceso ARIMA estacional para Z_t , quedará caracterizado por las FAC y la FACP del proceso $\{W_t\}$ y por d y D , los órdenes de diferenciación regular y estacional, respectivamente, del proceso original $\{Z_t\}$.

Las FAC y FACP de un proceso ARMA estacional son una mezcla de sus componentes regulares y estacionales, presentando la siguiente estructura:

- 1ª) Los retardos bajos presentan la estructura de la parte regular, mientras que los retardos estacionales ($s, 2s, \dots$) presentan la de la parte estacional.
- 2ª) En la FAC, alrededor de los retardos estacionales se manifiesta la interacción entre la parte regular y la estacional, con la repetición, a ambos lados de cada retardo estacional, de la FAC de la parte regular.
- 3ª) En la FACP, a la derecha de cada coeficiente estacional se repite la estructura de la FACP correspondiente a la parte regular, y a la izquierda de los coeficientes estacionales, la de las FAC correspondiente a la parte regular.

1.5 Esquema General de la Metodología Box-Jenkins

Se trata de una manera de proceder al análisis de una serie temporal utilizando un modelo de la clase ARIMA. El procedimiento consta de cuatro fases o etapas que se aplican de manera **iterativa** hasta alcanzar un resultado satisfactorio.

Consideramos que una serie de tiempo es una realización de un proceso estocástico específico $ARIMA(p, d, q) \times ARIMA(P, D, Q)_s$ bajo estacionalidad o un $ARIMA(p, d, q)$ sin estacionalidad.

En la práctica disponemos de una determinada serie temporal y a partir de ésta averiguar qué modelo $ARIMA(p, d, q) \times ARIMA(P, D, Q)_s$ representa adecuadamente el comportamiento de la misma, con el fin de utilizarla para predecir sus valores futuros.

Para esto se consideran las etapas de:

- 1ª) Identificación.
- 2ª) Estimación de parámetros.
- 3ª) Validación de los supuestos del modelo.
- 4ª) Predicción

1.6 Identificación

En la fase de identificación, en una primera etapa se procede a efectuar un análisis de estacionalidad de la serie. En el caso de que se trate de una serie no estacionaria se aplican las transformaciones adecuadas al objeto de convertirla en estacionaria.

1.6.1 Análisis de la Estabilidad en Varianzas y Transformaciones para conseguir Homocedasticidad

Las series pueden presentar una variabilidad no constante, es decir, la varianza σ_t^2 puede ser dependiente de t. Una forma de detectar si la varianza es constante es agrupar las observaciones por periodos de tiempo con el mismo número de observación (por ejemplo si los datos corresponden a observaciones mensuales, agruparlos en años) y realizar la prueba de Levene para la homogeneidad de varianzas de la serie observada en los K grupos definidos.

Esta prueba permite comprobar la hipótesis de que los K grupos proceden de poblaciones con varianza común.

La prueba de Levene se define como:

$$H_0 : \sigma_1 = \sigma_2 = \dots = \sigma_k \quad (1-25)$$

$$H_a : \sigma_i \neq \sigma_j \quad \text{para por lo menos un par (i, j)}. \quad (1-26)$$

El estadístico de la Prueba de Levene se define como:

$$W = \frac{(N - k) \sum_{i=1}^k N_i (\bar{Z}_i - \bar{Z}_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_i)^2} \quad (1-27)$$

donde Z_{ij} puede tener uno de lo siguiente tres definiciones:

- 1) $Z_{ij} = |Y_{ij} - \bar{Y}_i|$ donde \bar{Y}_i es la media del i-ésimo subgrupo.
- 2) $Z_{ij} = |Y_{ij} - \tilde{Y}_i|$ donde \tilde{Y}_i es el mediana del i-ésimo subgrupo.
- 3) $Z_{ij} = |Y_{ij} - \bar{Y}'_i|$ donde \bar{Y}'_i es un 10% de media recortada del i-ésimo subgrupo.

\bar{Z}_i son el grupo de medias de Z_{ij} y $\bar{Z}_{..}$ es la media global del Z_{ij}

Las tres opciones definidas para Z_{ij} determinan la robustez y poder de la prueba de Levene. Por la **robustez**, tiene la facultad la prueba de descubrir los grupos no homogéneos cuando los datos subyacente no se distribuyen normalmente y los grupos son de hecho homogéneos. Por la **potencia**, la prueba de descubre los grupos no-homogéneos cuando los grupos son de hecho no-homogéneos.

La definición basada en el **mediana** se recomienda como la opción que proporciona la mejor **robustez** contra muchos tipos de datos no-normales pero conserva la mejor potencia.

En la Región crítica la **prueba de Levene** rechaza la hipótesis que las **varianzas son homogéneas** si $W > F_{(1-\alpha, k-1, N-k)}$ donde $F_{(1-\alpha, k-1, N-k)}$ es el valor crítico superior de la distribución de F con $k - 1$ y $N - 1$ grados de libertad a un nivel de significancia de α .

Si la hipótesis de homogeneidad de varianzas fuera rechazada podría ser debido a que la dispersión de la serie cambia con la tendencia central. En concreto, una situación frecuente es que en los periodos en los que la tendencia central es grande también lo sea la dispersión. En dicha situación existe una familia de transformaciones que puede, en general, estabilizar la varianza En la práctica, se utilizan transformaciones del tipo:

$$T(Z_t) = \begin{cases} Z_t^\lambda & \text{si } \lambda \neq 0 \\ \ln(Z_t) & \text{si } \lambda = 0 \end{cases} \quad (1-28)$$

donde λ es el múltiplo de un medio más próximo al poder de transformación proporcionado por el gráfico de nivel y dispersión. El gráfico de nivel y dispersión representa, en cada uno de los K grupos, el logaritmo neperiano del rango intercuartílico de la serie Z_t (en ordenadas) frente al logaritmo neperiano de la mediana (en abscisas). El poder de transformación es igual a **uno** menos la pendiente de la recta de regresión mínima cuadrática ajustada a los K puntos. Esta familia de transformaciones no sólo permite estabilizar la varianza sino que, incluso, puede proporcionar normalidad. Por otro lado, sólo está definida para datos positivos, por lo que, en ocasiones, antes de realizar la transformación será necesaria sumar una misma constante a todos los valores de Z_t ,

1.6.2 Análisis de la Estabilidad en Medias y de la Estacionalidad

Una serie será estacionaria en media cuando pueda mantenerse el supuesto de media constante para toda la serie, es decir, cuando ésta fluctúe en torno a una única media.

Si la tendencia presente en el modelo es creciente o decreciente esto implica no estacionariedad en media o que existen varias medias en la serie.

Adicionalmente si la serie representa **estacionalidad** también debe analizarse la **estacionariedad** en los periodos estacionales.

Dado que la mayoría de las series económicas no son estacionarias (al menos en media) suelen ser homogéneas.

1.6.2.1 Identificación de d

Reglas Básicas

- 1ª) Si la serie original es estacionaria en media, entonces $d = 0$.
- 2ª) Cuando existan segmentos en la serie que difieran en nivel, pero que tengan pendiente constante, basta con diferenciar la serie una única vez, es decir:

$$W_t = (1 - L)Z_t = Z_t - Z_{t-1} \quad (1-29)$$

- 3ª) Cuando diferentes segmentos de la serie difieren tanto en nivel como en pendiente debe diferenciarse dos veces la serie, es decir:

$$W_t = (1 - L)^2 Z_t = (1 - L)(Z_t - Z_{t-1}) = Z_t - 2Z_{t-1} + Z_{t-2} \quad (1-30)$$

- 4ª) Normalmente no es necesario diferenciar más de dos veces.

1.6.2.2 Instrumentos que permiten identificar d

- Gráfico de la serie.
- Correlograma muestral de la serie.

1.6.2.3 Procedimiento para determinar d

- 1ª) Obtener, el gráfico y la FAC de la serie e inspeccionarlas. Si la serie oscila en torno a un valor promedio continuamente y si la FAC presenta rápida tendencia decreciente, entonces $d = 0$.

2ª) En el caso de que los requisitos anteriores no se cumpla, haga $d = 1$ y se repite el paso anterior.

Si el gráfico de la serie y el correlograma presentan las características anteriores, entonces concluimos que la serie transformada es estacionaria.

En caso contrario haga $d = 2$ y así hasta que la serie sea estacionaria.

También se puede utilizar *FACP*, en concreto si existe raíz unitaria.

La *FACP* suele tener el primer rezago muy cercano a la unidad en presencia de una raíz unitaria en la serie. En estos casos se sugiere aplicar contrastes de raíces unitarias: *Dickey - Fuller, DFA*, etc. y verificar las condiciones de Estacionariedad e Invertibilidad en el proceso de estimación.

¿Qué pasa si el orden de d es superior al necesario? Es decir, si sobrediferenciamos la serie.

Respuesta: El proceso será estacionario aunque puede dejar de ser invertible, es decir, si al diferenciar una o algunas de las raíces de la parte MA se aproxima a 1, entonces hay sobrediferenciación.

Otro **método** para determinar d es, diferenciar la serie para distintos ordenes y **analizar** la **varianza** o la **desviación estándar** de las **series diferenciadas**, en la serie que se presente un cambio de **menor a mayor** en la varianza será el **menor orden de diferenciación**.

1.6.2.4 Identificación de D

La expresión de una serie con componente **estacional** es;

$$W_t = (1 - L^s)^D (1 - L)^d Z_t^{(\lambda)} \quad (1-31)$$

Además de identificar λ y d se debe identificar D , el número de diferenciación estacional, lo cual implica, estacionariedad en la serie en la parte estacional.

Los valores habituales de D son; 0 indicando no diferenciación estacional o 1 diferenciación estacional.

La existencia de un componente estacional en la serie puede deducirse además del contraste de *Kruskal - Wallis* por:

- 1^a) El gráfico de la serie (la serie presenta valores superiores o inferiores al valor medio anual, los cuales se repiten frecuentemente para determinar periodos al año).
- 2^a) Correlograma muestral de dicha serie (*FAC* presenta valores elevados en los retardos correspondientes a los períodos estacionales).

1.6.3 Identificación de los Procesos Estocásticos Estacionarios subyacentes a Series Temporales

1.6.3.1 Identificación de Modelos Estacionarios

Cuando la serie Z_t no presenta patrón estacional y una vez identificados los valores de λ y d , efectuada la correspondiente transformación y diferenciación, obteniéndose una serie estacionaria en media y en varianza (W_t).

$$W_t = (1 - L)^d Z_t^{(\lambda)} \quad (1-32)$$

Observe que al efectuar d diferenciaciones regulares, el número de observaciones disponibles en la serie transformada será $T = N - d$, con N número de observaciones en la serie original y T número de observaciones en la serie transformada.

Una vez diferenciada la serie, se compara los correlogramas de la FAC y la FACP. Esto suele ofrecer una orientación para la formulación del modelo tentativo:

- 1ª) Si la FAC está más ajustada que la FACP el modelo suele ser $(0, d, q)$, por tanto se calcula el parámetro MA.
- 2ª) Si la FACP está más ajustada que la FAC el modelo suele ser $(p, d, 0)$, por tanto se calcula el parámetro AR.
- 3ª) Si FACP y FAC están igualmente ajustados el modelo puede ser (p, d, q) , por lo tanto se calculan los parámetros MA y AR (véase la Tabla 1.1).

Tabla 1.1: Características de los Modelos Estacionarios vs. FAC y FACP

FAC Y FACP EN MODELOS AR(p), MA(q) y ARMA(p,q)

MODELOS	FAC	FACP
AR(p)	Decrecimiento rápido de tipo geométrico puro, y geométrico con alternancias de signos, sinusoidal o mezcla de varios tipos	Se anula para retardos superiores a p
MA(q)	Se anula para retardos superiores a q	Decrecimiento rápido de tipo exponencial y/o sinusoidal
ARMA(p,q)	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas

Fuente: Introducción al análisis de series temporales por Ezequiel Uriel

Los procesos autorregresivos presentan la FACP con un número finito de valores distinto de cero. Un proceso AR(p) tiene los primeros p términos de la FACP distintos de cero y los demás son nulos. Esta afirmación es muy fuerte, en la práctica se considera que una muestra dada proviene de un proceso autorregresivo de orden p si los términos de la FACP son casi cero a partir del que ocupa el lugar p.

Los procesos de medias móviles presentan FAC con un número finito de valores distinto de cero. Un proceso MA(q) tiene los primeros q términos de la FAC distintos de cero y los demás son nulos. Estas propiedades son muy importantes con vistas a la identificación de un proceso mediante el análisis de las FAC y FACP.

La **identificación del modelo generador** será tanto más fácil cuanto **mayor** sea el **tamaño de la muestra**. Algunos autores consideran pequeño un tamaño de muestra

inferior a 50 ó 60 observaciones, pero también hay que tener en cuenta la característica de la serie que se está analizando.

1.6.3.2 Identificación del Modelo Estacional Multiplicativo General

Denominaremos por s al periodo estacional, si los datos son mensuales entonces $s = 12$.

El modelo estacional multiplicativo general denotado $ARIMA(p,d,q) \times ARIMA(P,D,Q)_s$.

$$\Phi_p(L^s)\phi_p(L)(1-L^s)^D(1-L)^d Z_t = \Theta_q(L^s)\theta_q(L)a_t + c \quad (1-33)$$

donde $W_t = (1-L^s)^D(1-L)^d Z_t$, puede decirse que la parte de medias móviles, ordinaria y/o estacional, introduce un decaimiento exponencial con posibles oscilaciones sinusoidales. Como resumen de los modelos estacionales puede construirse el siguiente cuadro (véase la Tabla 1.2)

Tabla 1.2: Características de los Modelos Estacionales vs. FAC y FACP

FAC Y FACP EN MODELOS $AR(p) \times AR(P)_s$, $MA(q) \times MA(Q)_s$		
MODELOS	FAC	FACP
$AR(p) \times AR(P)_s$	Decaimiento exponencial y/u oscilaciones sinusoidales atenuadas	Se anula para retardos mayores a $p+sP$
$MA(q) \times MA(Q)_s$	Se anula para retardos mayores a $q+sQ$	Decaimiento exponencial y/u oscilaciones sinusoidales atenuadas

Fuente: Introducción al análisis de series temporales por Ezequiel Uriel

1.6.4 Identificación del Término Independiente

Este término se incluirá si la serie W_t tiene media significativamente distinta de cero.

Para realizar este contraste $H_0 : \mu = 0$ se utiliza el estadístico.

$$t_{N-1} = \frac{\bar{W}}{\sigma_w} = \frac{\text{media muestral de } W}{\sigma_w} \quad (1-34)$$

con

$$\sigma_w = \left[\frac{S_w^2}{N} (1 + 2r_1 + 2r_2 + \Lambda + 2r_k) \right]^{1/2} \quad (1-35)$$

donde S_w^2 es la varianza muestral y N el tamaño muestral de la serie y r_1, r_2, \dots, r_k las primeras k autocorrelaciones muestrales significativas.

O aproximadamente se utiliza

$$\sigma_w = \left\{ \frac{S_N^2}{N} \right\}^{1/2} \quad (1-36)$$

Dado que el estadístico $t = \frac{W}{\sigma_w} \approx t$ con N - 1 grados de libertad, se requiere que el

tamaño muestral N sea lo mas grande posible. Y dado que la distribución t se aproxima a una normal a medida que aumenta el tamaño de la muestra, entonces se acostumbra utilizar como valor crítico para Aceptar o Rechazar.

Para $\alpha = 0.05$, si $|t| > 2$ decidimos incluir el término constante en el modelo.

Otra forma seria, **estimar el modelo incluyendo el término constante** y luego analizar la significancia estadística de éste.

1.7 Estimación de Parámetros

En la fase de estimación, el objetivo es la obtención de estimadores de los parámetros $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \mu_w$ y σ_a^2 . Para designar a los parámetros autorregresivos y de medias móviles se utilizará el vector

$$\beta = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q)' \quad (1-37)$$

Los supuestos que se adoptarán en la exposición del proceso de estimación son los siguientes:

- 1ª) $a \approx N(0, \sigma_a^2 I)$.
- 2ª) El proceso W_t es estacionario.
- 3ª) El proceso W_t es invertible.

Además, el número de observaciones de la serie W_t es $T = N - d$. En el caso de un modelo estacional multiplicativo el número de observaciones de W_t será $T = N - d - sD$, es decir, T es el número de observaciones disponibles para la estimación.

Si se adopta el criterio mínimo cuadrático, los estimadores se obtienen minimizando la expresión $S = \sum a_t^2$

Si se adopta el criterio máxima verosímil, entonces los estimadores se obtendrán mediante maximización de la función de verosimilitud $L = f(a_1, \dots, a_N)$

Tanto en la minimización como en la maximización hay que tener en cuenta el papel que juegan los valores iniciales. A este respecto en el proceso de estimación se

puede adoptar un enfoque condicional o alternativamente un enfoque al que se le puede designar con la denominación de no condicional. En el cuadro siguiente, (véase la Tabla 1.3), se señala para cada enfoque qué valores se determinan en el proceso de estimación y qué valores se consideran como condiciones iniciales.

Los criterios que se utilizan en modelos estacionarios, son también válidos para la estimación de modelos estacionales. Conviene no obstante hacer algunas consideraciones acerca de la inclusión de parámetros estacionales en el proceso de estimación.

Tabla 1.3: Enfoque Condicional y no Condicional en la estimación de un modelo

ARMA(p, q)		
Enfoque	Valores a estimar	Valores dados
Condicional	$\beta = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q)'$ $\sigma_a^2, (\mu_w \text{ en su caso})$	$W = (W_1, W_2, \dots, W_T)'$ $W^0 = (W_0, W_{-1}, \dots, W_{-p+1})'$ $a^0 = (a_0, a_{-1}, \dots, a_{-q+1})'$
No condicional	$\beta = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q)'$ $W^0 = (W_0, W_{-1}, \dots, W_{-p+1})'$ $a^0 = (a_0, a_{-1}, \dots, a_{-q+1})'$ $\sigma_a^2, (\mu_w \text{ en su caso})$	$W = (W_1, W_2, \dots, W_T)'$

Así, un proceso autorregresivo mixto -con componente estacional y no estacional- no se puede estimar por métodos lineales, a diferencia de lo que ocurriría con los procesos autorregresivos no estacionales.

Por otra parte, en la estimación condicionada de los procesos autorregresivos estacionales hay que tener presente que se perderán $D \times s$ observaciones iniciales.

La utilización de métodos de estimación no condicionada requiere el cálculo hacia atrás de muchos más términos que en los procesos no estacionales. Por este motivo, en los modelos estacionales las diferencias entre una estimación condicionada y una estimación no condicionada pueden llegar a ser considerables, siendo preferible, como es lógico, utilizar este último procedimiento.

1.8 Validación de los Supuestos del Modelo

La estrategia a seguir es:

- 1ª) Efectuar el chequeo de los modelos estimados, desechando aquellos que no cumplan los requisitos mínimos.
- 2ª) Simultáneamente en esta etapa, los contrastes pertinentes pueden indicarnos direcciones concretas a seguir respecto a cómo modificar ciertas especificaciones de los modelos desechados.
- 3ª) Todos los modelos que superen esta etapa serán utilizadas posteriormente para realizar predicciones.

En resumen, se tienen dos fases del análisis de validación del modelo.

- La que se realiza previamente a la utilización del modelo.
- La que analizará la adecuación de los modelos al objeto de análisis.

Requisitos que debe satisfacer un modelo para afirmar que es adecuado, es:

- i.) Estimaciones significativas además que cumpla con las condiciones de **Estacionariedad e invertibilidad**
- ii.) Residuos estimados se comporten como un ruido blanco
- iii.) Modelo no omita parámetros relevantes (modelo bien especificado)
- iv.) Que el modelo sea estable.

1.8.1 Análisis de Coeficientes

Para analizar la significancia de los coeficientes estimados, debemos efectuar contrastes de hipótesis respecto a dichos coeficientes del vector.

$$\beta = (\hat{c}, \hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q) \quad (1-38)$$

De donde, bajo el supuesto de normalidad

$$\beta \sim N(\beta, \text{var}(\beta)) \quad (1-39)$$

$$\text{con } \text{var}(\beta) = \sigma_a^2 (X^T X)^{-1} \quad \text{y} \quad \sigma_a^2 = \frac{S_*(\beta)}{T} \quad (1-40)$$

Los elementos de la diagonal principal serán las estimaciones de las varianzas para cada uno de los componentes de β denotados por σ_β^2 con $i = 1, 2, \dots, k$. Por tanto es posible contrastar hipótesis nulas del tipo

$$H_0 : \beta_i = \beta_{i0} \quad (1-41)$$

mediante el estadístico

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{\sigma_{\hat{\beta}}} \quad (1-42)$$

Una cuestión importante en el análisis de coeficientes es comprobar las condiciones de estacionariedad e invertibilidad.

Invertibilidad en los modelos MA y Estacionariedad en los modelos AR.

La existencia de raíces unitarias en la parte autorregresiva implica no estacionariedad, lo que implica la necesidad de diferenciar nuevamente la serie en su parte regular y/o en su parte estacional dependiendo de si la raíz unitaria es estacional o no estacional, de donde se debe hacer una nueva identificación, una nueva estimación de los modelos identificados, y un nuevo chequeo.

La existencia de raíces unitarias en la parte MA puede deberse a un problema contrario de sobrediferenciación.

Por último, **análisis de la matriz de correlación** de los coeficientes estimados. Cuando la correlación entre dos coeficientes estimados es muy alta se tiene un problema grave de multicolinealidad, generando estimaciones muy inestables, además, altas correlaciones implican la existencia de sobreparametrización del modelo, lo cual crea dificultades para en la convergencia de modelos.

Algunos autores recomiendan eliminar algunos de los parámetros implicados en la elevadas correlaciones, especialmente si este no es estadísticamente significativo.

1.8.2 Análisis de Residuales

Si un modelo estacional ha sido correctamente identificado entonces los residuales pueden ser obtenidos a partir de

$$a_t = \frac{(1 - \Phi_1 L^s - \dots - \Phi_p L^{ps})(1 - \phi_1 L - \dots - \phi_p L^p)}{(1 - \Theta_1 L^s - \dots - \Theta_q L^{qs})(1 - \theta_1 L - \dots - \theta_q L^q)} W_t \quad (1-43)$$

Conocidos los β_i y los W_i se calcula a_t que debe ser ruido blanco.

Anderson(1942) demostró que

$$r_j(a) \sim N\left(0, \frac{1}{T}\right) \quad (1-44)$$

Pero dado que a_t es no observable, se tiene en realidad

$$\hat{a}_t = \frac{(1 - \hat{\Phi}_1 L^s - \dots - \hat{\Phi}_p L^{ps})(1 - \hat{\phi}_1 L - \dots - \hat{\phi}_p L^p)}{(1 - \hat{\Theta}_1 L^s - \dots - \hat{\Theta}_q L^{qs})(1 - \hat{\theta}_1 L - \dots - \hat{\theta}_q L^q)} W_t \quad (1-45)$$

Si el modelo es adecuado, puede demostrarse que a medida que T aumenta, \hat{a}_t tiende a parecerse a a_t el cual se distribuye ruido blanco,

$$\lim_{T \rightarrow \infty} \hat{a}_t \rightarrow a_t \quad (1-46)$$

es decir, se trata de averiguar si los residuos de los modelos estimados cumplen con la característica de un **modelo ruido blanco**.

1.8.2.1 Media Cero

Se puede realizar pruebas informales o formales en este caso.

Informal: un análisis gráfico en el cual se analiza el comportamiento de los residuales, es decir, si giran alrededor de un valor constante.

Formal: realizar un análisis o Prueba de significancia estadística de la media de los residuos.

1.8.2.2 Varianza Constante

Realizar un análisis de la dispersión de los residuos a lo largo del tiempo. Si la varianza de la serie es constante esto implica homocedasticidad, si no lo es implica heterocedasticidad, siendo necesario transformar la serie.

1.8.2.3 Incorrelación

Analizar si los residuos son ruido blanco, es decir, que los coeficientes de autocorrelación muestral de los residuos

$$r_j(a) = \frac{\sum_{i=j+1}^T (\hat{a}_i - \bar{a})(\hat{a}_{i-j} - \bar{a})}{\sum_{i=1}^T (\hat{a}_i - \bar{a})^2} \quad (1-47)$$

deben ser pequeños. Interesa probar si la serie es ruido blanco, es decir, si todas las autocorrelaciones son conjuntamente iguales a cero.

Dado que $r_j(\hat{a}) \sim N(0, \frac{1}{T})$ esto es equivalente a tener $\sqrt{Tr_j(\hat{a})} \sim N(0,1)$ luego elevando al cuadrado a ambos lados se tiene que

$$Tr_j^2(\hat{a}) \sim (N(0, \frac{1}{T}))^2 = X_1^2 \quad (1-48)$$

Además se puede demostrar que dado que $r_j(a) \sim N(0, \frac{1}{T})$ éstas en distintos desplazamientos son aproximadamente independientes entre sí.

Dado que la suma de variables independientes X^2 también es X^2 con grados de libertad igual a la suma de los grados de libertad de esas variables, entonces definimos el estadístico Q de Box-Pierce como

$$Q_{BP} = T \sum_{a=1}^m r^2(a) \sim X_m^2 \quad (1-49)$$

con m grados de libertad.

donde la hipótesis nula es H_0 : La serie es ruido blanco, y donde m es el desplazamiento máximo seleccionado, es decir, el numero de rezagos que se desea analizar. En la practica $m = \sqrt{T}$ o $m=1/4T$ o $m=1/3T$.

Como el objetivo es seguir mas de cerca la distribución X^2 en muestras pequeñas, se plantea el estadístico de Ljung-Box.

$$Q_{LB} = T(T+2) \sum_{a=1}^m \left(\frac{1}{T-a}\right) r^2(a) \sim X_m^2 \quad (1-50)$$

Donde $\frac{T+2}{T+a}$ es una ponderación, la cual cuando T es moderadamente grande los pesos son aproximadamente 1.

Por tanto el estadístico Q_{BP} difiere un poco del estadístico de Q_{LB} .

Luego, aceptamos la significancia estadística de cada $r_j(a)$ si su valor no supera el intervalo

$$\pm 1.96 \cdot \left(\frac{1}{\sqrt{T}}\right) \quad (1-51)$$

donde, el conjunto de las $r_j(\hat{a})$ forman la función de autocorrelación muestral de los residuales (FACMR).

Las bandas de significancia de (1-51) además sirven para probar la significancia estadística de la función de autocorrelación parcial muestral de los residuales (FACPMR)

1.8.2.4 Distribución Normal de los Residuales \hat{a}

Para verificar esta, pueden utilizarse bien sea, los coeficientes de Asimetría y de Curtosis definidos como

$$g_1 = \frac{m_3}{m_2^{3/2}} \quad g_2 = \frac{m_4}{m_2^2} - 3 \quad (1-52)$$

donde m_j es el momento muestral central de orden j , donde $j = 1, 2, 3, 4$.

o el estadístico de Jaque-Bera (1987) definido como

$$LM_N = \frac{Tg_1}{6} + \frac{Tg_2^2}{24} = T \left[\frac{m_3^2}{6m_2^3} + \frac{1}{24} \left(\frac{m_4}{m_2^2} - 3 \right)^2 \right] \quad (1-53)$$

donde la hipótesis nula se define que los \hat{a} se distribuyen normal contra no normal.

1.8.3 Sobreajuste

Consiste en incluir parámetros adicionales en el modelo identificado con el fin de comprobar si estos modelos menos parsimoniosos resultan apropiados, es decir, si se identificó el modelo ARIMA(p; d; q), estime lo modelos ARIMA(p + 1, d, q), y ARIMA(p, d, q + 1) y analice la significancia estadística de los parámetros adicionales.

No se estima simultáneamente ambos parámetros adicionales, es decir, el modelo ARIMA(p + 1, d, q + 1), lo cual implica problemas de parámetros redundantes.

1.8.4 Bondad de ajuste

1.8.4.1 Línea Verificacionista

En la metodología Box-Jenkins, el R^2 como medida de bondad de ajuste y el \bar{R}^2 ajustado como medida de comparación de modelos no son aplicables.

$$R^2 = 1 - \frac{\sigma_a^2}{\sigma_w^2} \quad \bar{R}^2 = 1 - \frac{\sigma_a^2 / T - k}{\sigma_w^2 / T - 1} \quad (1-54)$$

La diferencia entre el R^2 y el \bar{R}^2 se debe a que el segundo criterio penaliza la introducción de parámetros adicionales en el modelo. Por ello sería en principio más aconsejable utilizar el \bar{R}^2 ya que, de alguna forma, tiende a evitar la sobreparametrización.

El criterio no paramétrico que habitualmente se maneja para seleccionar entre modelos estocásticos es la varianza estimada del ruido σ_a^2 , según este criterio, de entre todos los modelos que ha superado la etapa de validación se seleccionara aquel que tenga menor varianza estimada del ruido, donde

$$\sigma_a^2 = \frac{\sum_{t=1}^T (\hat{a}_t - \bar{\hat{a}})^2}{T} \quad \sigma_w^2 = \frac{\sum (\hat{W}_t - \bar{\hat{W}})^2}{T} \quad (1-55)$$

Como criterios paramétricos se pueden aplicar directamente los procedimientos de contrastes de hipótesis clásicos, por ejemplo para contraste individuales el

estadístico t y para contrastes globales o restringidos, utilizar los estadísticos LM Wald o LR.

1.8.4.2 Línea Preferencialista

Dadas las limitaciones que presenta el R^2 que son extensibles al \bar{R}^2 , se han propuesto otros criterios de información a partir de un enfoque Bayesiano:

- 1ª) Criterio de información de Akaike (AIC) se considera por aquel modelo para el que se minimice

$$Akaike = \ln(\sigma_a^2) + \frac{2k}{N} \quad (1-56)$$

donde k es el número de parámetros del modelo. Es inmediato ver que al seguir este criterio, se optará por aquel modelo cuyo ajuste sea mejor (una menor varianza residual) de entre aquellos modelos con igual número de parámetros. Por otra parte, se seleccionará aquel modelo que tenga un menor número de parámetros de entre aquellos que presenten un igual grado de ajuste.

- 2ª) Criterio de información de Schwartz (SC) supone seleccionar aquel modelo que minimiza

$$Schwarz = \ln(\sigma_a^2) + \frac{k}{N} \ln(N) \quad (1-57)$$

En este criterio se penaliza la introducción de parámetros (aumentos en k suponen aumentos en el valor del estadístico). La principal diferencia estriba

en que el criterio de Schwarz supone una mayor penalización a la introducción de nuevos parámetros, pues $\ln(N) > 2$, siempre que $N > 7$.

En la interpretación de estos estadísticos conviene tener en cuenta las siguientes observaciones:

- (a) Estos estadísticos pueden aplicarse a modelos sin término independiente, no planteándose ningún problema de interpretación.
- (b) Estos estadísticos no son medidas de carácter relativo como lo es el coeficiente de determinación. Por ello, no puede decirse que un valor obtenido en un modelo sea en sí mismo elevado o bajo. Su utilidad se manifiesta cuando se comparan los valores obtenidos en modelos alternativos.

1.8.5 Análisis de Estabilidad

Para analizar la estabilidad tómense dos periodos muestrales distintos, (por ejemplo dividir el tamaño muestral en dos partes), y estimar sus parámetros para cada una de las partes concluyendo que el modelo es estable cuando las estimaciones de los parámetros no sean muy diferentes.

Esta prueba es equivalente a la aplicación de la prueba de Chow (1960) de cambio estructural.

Si la **F calculada** o experimental es **mayor** que la tabulada o **teórica** a un determinado nivel de significación, se **rechaza** la **hipótesis de estabilidad estructural**.

1.9 Predicción

El objetivo consiste en predecir valores futuros de Z sujetos al menor error posible. es decir, si se tiene información hasta el periodo T , supondremos que se desea obtener predicciones de Z_{T+l} , valor de Z en el periodo $T+l$.

En la obtención del predictor $\hat{Z}_{T+l|T}$ se adoptarán los siguientes supuestos:

- 1ª) Los parámetros ϕ_i y θ_i son conocidos. Para evaluar correctamente la incidencia del supuesto adoptado, conviene tener presente que en series de tamaño superior a 60, que generalmente son las utilizadas en la elaboración de los modelos ARIMA, los errores de estimación de los parámetros son de cuantía reducida.
- 2ª) Se supone que todos los ruidos presentes y pasados $a_T, a_{T-1}, \dots, a_1, a_0, \dots$ son conocidos

Al conjunto de información disponible hasta T lo denotaremos por

$$I_T = (Z_T, Z_{T-1}, Z_{T-2}, \dots, a_T, a_{T-1}, a_{T-2}, \dots) \quad (1-58)$$

$\hat{Z}_{T+l|T}$ será el **predictor óptimo** para el periodo $T+l$ utilizando I_T , luego el predictor óptimo será aquel que minimice el **error cuadrático medio** de predicción que se define como

$$ECM(\hat{Z}_{T+l|T}) = E[(Z_{T+l} - \hat{Z}_{T+l|T})^2] \quad (1-59)$$

es decir, se busca que (1-59) sea mínimo.

La cantidad $(Z_{T+l} - \hat{Z}_{T+l|T})$ se conoce como error de predicción,

$$e_t(l) = (Z_{t+l} - \hat{Z}_{t+l|T}) \quad \text{Donde } \hat{Z}_{T+l|T} = E(Z_{T+l}) \quad (1-60)$$

1.9.1 Predicción por Intervalos

El objetivo es obtener predicciones con el menor error de predicción posible

$$e_T(l) = Z_{T+l} - \hat{Z}_{T+l|T} \quad (1-61)$$

Una forma alternativa de escribir (1-61) es

$$e_T(l) = a_{T+l} + \psi_1 a_{T+l-1} + \psi_2 a_{T+l-2} + \dots + \psi_{l-1} a_{T+1} \quad (1-62)$$

donde

$$E(e_T(l)) = E(a_{T+l} + \psi_1 a_{T+l-1} + \psi_2 a_{T+l-2} + \dots + \psi_{l-1} a_{T+1}) = 0 \quad (1-63)$$

$$Var(e_T(l)) = E(a_{T+l} + \psi_1 a_{T+l-1} + \psi_2 a_{T+l-2} + \dots + \psi_{l-1} a_{T+1})^2 = \sigma_a^2 (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{l-1}^2) \quad (1-64)$$

$$\therefore e_T(l) \sim N \left[0, \sigma_a^2 \left(\sum_{i=0}^{l-1} \psi_i^2 \right) \right] \quad (1-65)$$

con $\psi_0 = 1$

Por lo tanto, un intervalo de confianza para $\hat{Z}_{T+l|T}$ será igual a

$$\hat{Z}_T(l) \pm N_{\epsilon/2} \cdot d.e.(e_T(l)) \quad (1-66)$$

De (1-62), (1-64) y (1-65) el error cuadrado medio es

$$ECM(\hat{Y}_{T+l|T}) = Var(e_T(l)) = \sigma_a^2 (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{l-1}^2) \quad (1-67)$$

entonces se verifica que el ECM coincide exactamente con la varianza de error de predicción.

Teniendo en cuenta que las a_t se distribuyen normalmente se pueden establecer intervalos de confianza del siguiente tipo

$$\text{Pr ob}\{\hat{Z}_{T+H|T} - 1.96\sqrt{ECM(\hat{Z}_{T+H|T})} < Z_{T+H} < \hat{Z}_{T+H|T} + 1.96\sqrt{ECM(\hat{Z}_{T+H|T})} = 0.95 \quad (1-68)$$

1.9.2 Indicadores no Paramétricos del Grado de Exactitud

Cuando es necesario comparar entre la mejor de varias predicciones respecto a una serie $\{Z_T\}$, se utiliza las medidas no paramétricas, **el error cuadrático medio (ECM)**, **el error absoluto medio (EAM)** y **el error absoluto porcentual medio (EAPM)** las cuales no tienen en consideración la estructura estocástica del modelo.

Definidas como:

$$ECM = \frac{1}{H} \sum_{l=1}^H e_{T-H}^2(l) = \frac{1}{H} \sum (Z_{T+l-H} - \hat{Z}_{T-H}(l))^2 \quad (1-69)$$

$$EAM = \frac{1}{H} \sum_{l=1}^H |e_{T-H}(l)| = \frac{1}{H} \sum_{l=1}^H |Z_{T+l-H} - \hat{Z}_{T-H}(l)| \quad (1-70)$$

$$EAPM = \frac{1}{H} \sum_{l=1}^H \frac{|e_{T-H}(l)|}{Z_{T+l-H}} * 100 = \frac{1}{H} \sum_{l=1}^H \frac{|Z_{T+l-H} - \hat{Z}_{T-H}(l)|}{Z_{T+l-H}} * 100 \quad (1-71)$$

CAPÍTULO 2

METODOLOGIA Y CARACTERIZACION

2.1 Metodología usada para el Modelamiento y Predicción de la Recaudación de Ingresos Corrientes de la UNI

El presente trabajo de investigación utiliza datos reales mensuales de la **Recaudación de Ingresos Corrientes en soles corrientes**, (véase el Anexo 3), desde enero de 1993 hasta junio del 2003, de la Universidad Nacional de Ingeniería.

Luego se procederá a una **deflactación**¹ utilizando los **índices de precios al consumidor** (véase el Anexo 2) para eliminar el efecto de los precios; obteniéndose así una serie histórica en soles constantes (véase el Anexo 4).

La **serie histórica en soles constantes** partirá con los supuestos de un modelo estocástico ARIMA. Pero la suposición básica fundamental en el análisis de las series de tiempo es que los factores que han influido en el pasado y en el presente

¹ En el Apéndice 1 se desarrolla la fórmula de cálculo para la deflactación

en los patrones de la actividad económica de la institución continuaran haciéndolo mas o menos en la misma forma en el futuro.

A partir de la serie histórica, se tiene que averiguar qué modelo ARIMA es susceptible de haber generado dicha serie (es decir, que modelo representa adecuadamente el comportamiento de la misma), de lo cual se siguen cuatro etapas: identificación, estimación, validación y predicción.

En la etapa de identificación consiste en inducir - a partir de los datos, la función de autocorrelación muestral y la función de autocorrelación parcial muestral- qué modelos ARIMA se adaptarían mejor a las características de la serie. Cuando se trata de una serie no estacionaria, en primer lugar se procede a estacionalizar la serie, tanto en varianza, esto es identificando el valor de λ , como en media, es decir, identificando el valor de d y D (estacionalidad). Una vez que esta serie transformada es estacionaria (en media y en varianza) se deben de averiguar los posibles valores tanto de la parte regular del modelo (p y q) como de la parte estacional (P y Q).

En la etapa de estimación una vez identificados los posibles modelos que han podido generar la serie temporal, es preciso cuantificar los parámetros de los mismos. Los dos problemas fundamentales a los que se enfrenta la estimación de los modelos ARIMA son el de los valores iniciales y el de no linealidad.

La etapa de validación, consiste en analizar si los residuos del modelo tienen un comportamiento similar a las perturbaciones del mismo (a_t), esto es, si puede afirmarse que son semejantes a un ruido blanco.

Adicionalmente, se trata de comprobar la calidad de las estimaciones.

En la etapa de la predicción, tras la validación, viene el fin básico de esta metodología, esto es, la obtención de predicciones de valores futuros de la serie temporal. Una vez obtenidas las predicciones del modelo se trata de volver a chequear la adecuación del mismo, pudiendo utilizarse para ello tanto métodos no paramétricos (como el error cuadrático medio) como paramétricos (estadísticos de contenido informativo).

Para la elaboración de los distintos análisis se han utilizado herramientas informáticas disponibles en el mercado. Para la Identificación del modelo y para las gráficas de los correlogramas de las series se ha empleado el Statistical Package for Social Sciences -SPSS. En lo que se refiere a la modelización de las series temporales se ha utilizado el Econometric Views –Eviews, para la contrastación de hipótesis, para la estimación, validación y predicción de modelos ARIMA. Se ha escogido este programa por las posibilidades de elaborar programas para el tratamiento de los datos.

2.2 Caracterización de la Variable

Los **Ingresos Corrientes** son recursos financieros que se obtienen de modo regular o periódico, y que no alteran de manera inmediata la situación patrimonial del Estado.

Los **Recursos Directamente Recaudados**; comprende los **Ingresos Corrientes**, generados por las **entidades** de la UNI, que no constituyen recursos ordinarios. Tienen su origen en el esfuerzo propio de recaudación y provienen de los servicios que brindan; entre los cuales se pueden mencionar las **tasas**, la **venta de bienes**, la **prestación de servicios** y las **rentas de la propiedad** ².

Entre las principales **entidades** generadoras de estos **Ingresos Corrientes** han sido la Oficina Central de Admisión (**OCAD**), la Oficina Central de Infraestructura (**OCI**), el Centro de Extensión y Proyección Social (**CEPS**), Centro de Estudios Pre-Universitarios de la UNI (**CEPREUNI**), Centro Peruano Japonés de Investigaciones Sísmicas y Mitigación de Desastres (**CISMID**), Instituto de Ciencias de la Información (**ICI**), Laboratorio Nacional de Hidráulica (**LNH**), (véase el Anexo 5).

También las **Facultades** de la **Universidad Nacional de Ingeniería**, (véase el Anexo 6), forman parte de la generación de **Ingresos Corrientes** y esto a contribuido a la calidad profesional de nuestros académicos, la confianza de la

² Con más detalle se puede ver el desagregado de la variable Recaudación de Ingresos Corrientes en el Anexo 1

sociedad y el Estado en la UNI, una apropiada enseñanza en relación a una practica real y la Magnitud de la inversión que ha permitido desarrollar íntegramente a la UNI.

La suma de la **Recaudación de Ingresos Corrientes** de todas las **Dependencias y Facultades** que conforman la **Universidad Nacional de Ingeniería** determinó la variable de estudio de este presente trabajo.

Z_t : Recaudación de Ingresos Corrientes de la UNI (En Soles Constantes)
(Base Dic. 2002 = 100) (véase el Anexo 4)

Por la Fuente de Financiamiento : Recursos Directamente Recaudados

t: Mes (desde enero del año 1993 a junio del año 2003) ($t = 1, 2, \dots, 126$)

CAPÍTULO 3

RESULTADOS

3.1 Modelación y Predicción de la Recaudación de los Ingresos Corrientes de la UNI a través de un Modelo ARIMA por la Fuente de Financiamiento de Recursos Directamente Recaudados.

En este capítulo los objetivos que se desarrollan es el modelamiento y la predicción de la Recaudación de los Ingresos Corrientes de la UNI a través de un Modelo ARIMA, para lo cual contamos con una serie de tiempo que empieza en el mes de enero de 1993 y termina en junio del 2003, es decir 126 observaciones (véase el Anexo 4).

Los últimos datos de la serie como periodo de validación, es decir el último semestre quedará reservada para efectos de comprobación de la capacidad predictiva del modelo.

A continuación se desarrolla un cuadro con las estadísticas básicas correspondientes a 120 observaciones.

Tabla 3.1: Estadísticas básicas de la Serie Histórica

Zt: Serie Histórica	Estadísticos
Media	3,946,729
Intervalo Confidencial de la Media al 95%	
Límite Inferior	3,545,375
Límite Superior	4,348,084
Mediana	3,803,863
Desviación Estándar	2,220,405
Mínimo	329,621
Máximo	9,546,444
Rango	9,216,823

En el Tabla anterior, se observa que el promedio mensual de Recaudación de Ingresos Corrientes es de 3,946,729 Soles. La cantidad mínima fue de 329,621 en el mes de marzo de 1993 y la máxima de 9,546,444 en el mes de agosto del 2000. Además el 50% de las observaciones encuentran sobre los 3,803,863 Soles.

Para desarrollar el análisis de la serie temporal Z_t , de la Recaudación de los Ingresos Corrientes de la UNI en soles constantes, se utilizará la metodología Box-Jenkins que es un procedimiento de análisis estadístico para ajustar a una serie un tipo especial de modelos, denominados ARIMA (Autorregresive Integrated Moving Average). A continuación se desarrollará las siguientes fases o etapas:

3.2 Fase de Identificación

La primera etapa en el **proceso de modelamiento o identificación del modelo univariante de una serie temporal, consiste en comprobar si esta es estacionaria en media y en varianza.** Si además los datos de la serie están recogidos con una periodicidad inferior a un año (en el caso de la Recaudación de Ingresos Corrientes de la UNI, se trata de datos mensuales) tendremos que **comprobar también si es estacionaria estacionalmente.** En el caso que la serie no sea estacionaria, ya sea en varianza o en media, habrá que aplicarle las transformaciones necesarias para que lo sea.

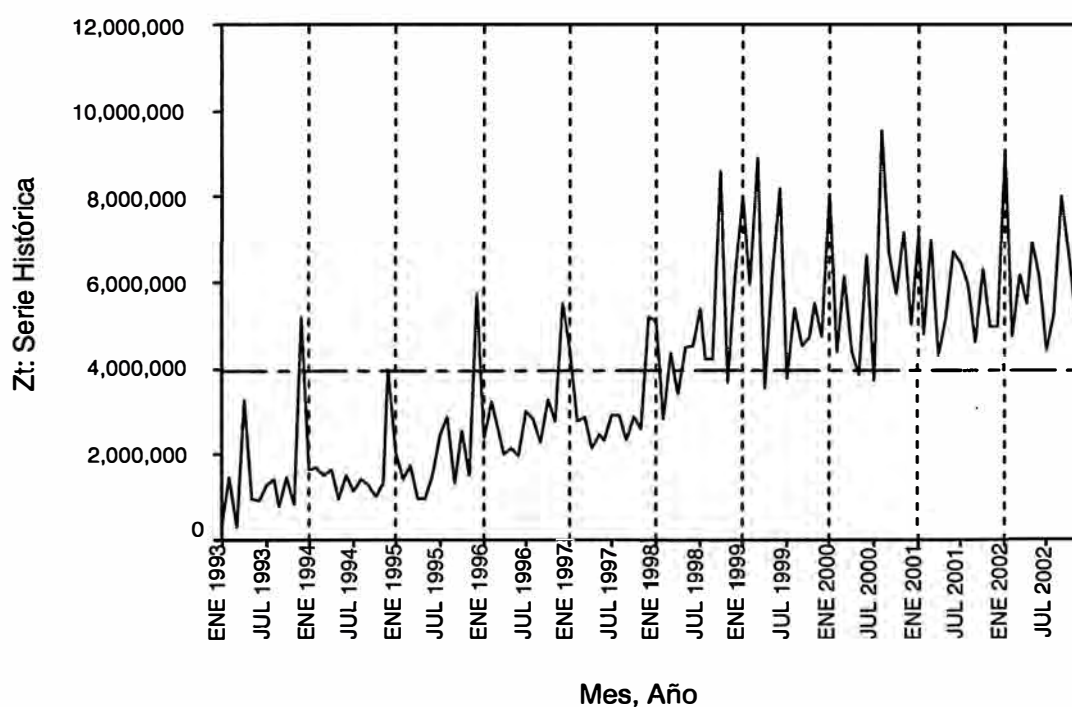


Figura 3.1: Serie Histórica de la Recaudación de Ingresos Corrientes (En soles Constantes)

3.2.1 Estabilización de la Varianza y generación de la serie transformada para conseguir Homocedasticidad.

En el **gráfico de secuencia**, (véase la Figura 3.1) se puede observar que al parecer la serie **no es estacionaria en varianza**. Se puede ver que en los inicios de la serie hay un tipo de dispersión, en cambio a medida que se avanza en el tiempo, más o menos en la mitad y sobre todo al final de la serie, la dispersión de los datos es de otro tipo, lo que muestra indicios de no estacionariedad en varianza. Para detectar si la **varianza** de la serie es **constante**, es decir si no cambia con el tiempo, se agrupa las observaciones por **años** y se aplica la prueba de **Levene** para la **homogeneidad de varianzas**. Dado que el periodo de ajuste comprende desde 1993 hasta 2002, ambos años inclusive, ya que se dispone de 10 grupos con 12 observaciones cada uno. La prueba de Levene permitirá contrastar la hipótesis nula de que no existen diferencias significativas entre las varianzas de la serie en los 10 grupos definidos. A continuación se muestra una tabla con los resultados de la prueba y después un Box Plot o gráfico de cajas.

Tabla 3.2: Prueba de Homogeneidad de Varianza de Z_t

Zt: Serie Histórica	Estadístico de Levene	k-1	N-k	Nivel de Significancia
Basado en la Media	1.796	9	110	0.077
Basado en la Mediana	1.548	9	110	0.140
Basado en la Mediana y con ajuste en N-k	1.548	9	90	0.144
Basado en la Media ajustada	1.779	9	110	0.080

k=10 (Número de grupos: años)

N=120 (Tamaño de la Serie Histórica)

Utilizando la Prueba de Homogeneidad de Varianza de Levene, (véase la Tabla 3.2), se llega a la conclusión que **se puede aceptar al 5% la hipótesis nula de que las varianzas son las mismas a lo largo de los años.**

Observando el gráfico de caja, (véase la Figura 3.2), sólo nos da una referencia de lo que puede estar pasando, pero la respuesta definitiva la da la Prueba de Levene.

Para determinar una potencia apropiada de transformación de los datos, podemos graficar, para cada grupo, el logaritmo de la mediana versus el logaritmo del rango intercuartílico.

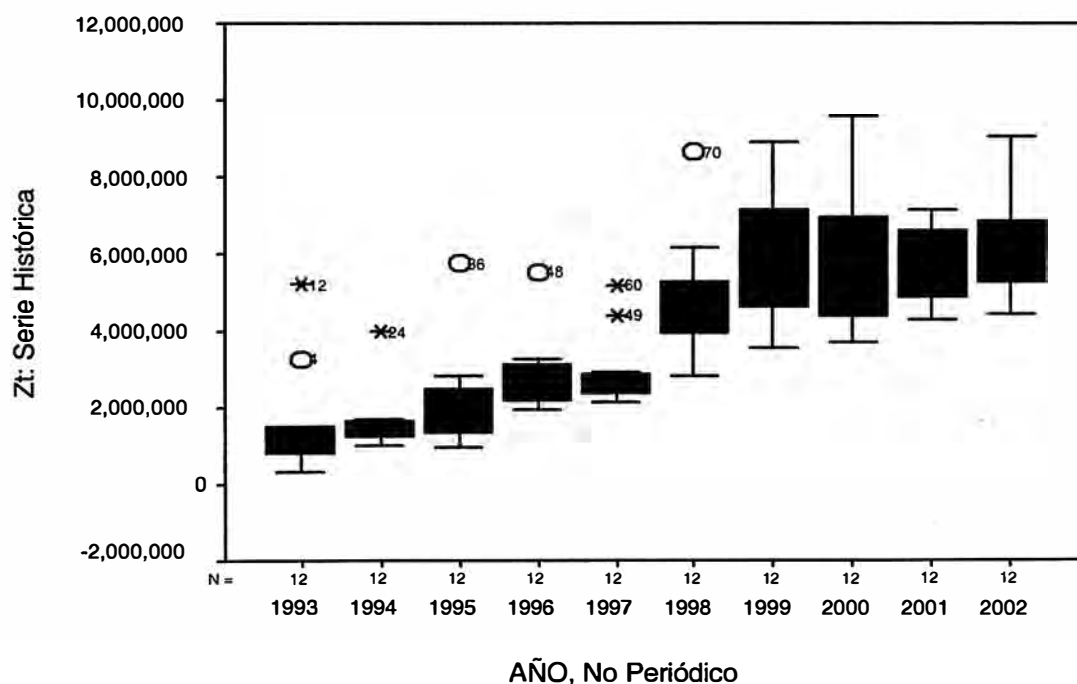
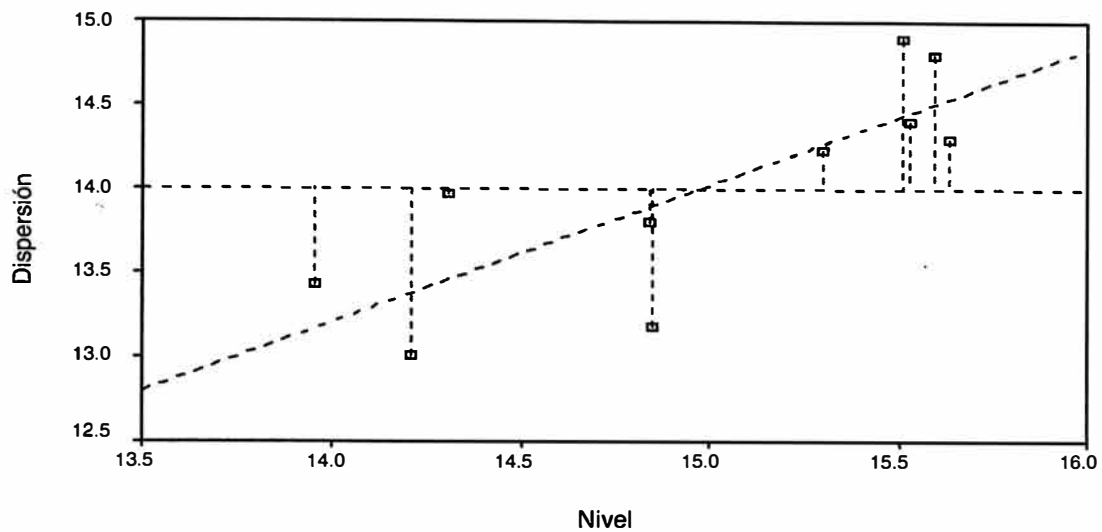


Figura 3.2: Gráfico de Cajas por Años de la Serie Histórica



* Plano de LN(Dispersion) vs LN(Nivel)

Pendiente = 0.815 Potencia de Transformación = 0.185

Figura 3.3: Diagrama de Nivel y Dispersion de la Serie Histórica por Año

En el gráfico anterior, **de nivel y dispersion** se puede observar que a medida que pasa el tiempo, la varianza va aumentando, incluso la **pendiente** de la recta de regresión para la nube de puntos representada es **0,815** y la Potencia de Transformación sugerida para la variable es $W_t = (Z_t)^{0.185}$ (véase la Figura 3.3)

Tabla 3.3: Prueba de Homogeneidad de Varianza de W_t

Wt: Serie Transformada	Estadístico de Levene	k-1	N-k	Nivel de Significancia
Basado en la Media	2.040	9	110	0.041
Basado en la Mediana	1.857	9	110	0.066
Basado en la Mediana y con ajuste en N-k	1.857	9	62	0.076
Basado en la Media ajustada	1.992	9	110	0.047

Se vuelve a hacer la prueba de **homogeneidad de la varianza**, pero ahora la transformación fue de $W_t = (Z_t)^{0.185}$. Luego se **acepta la hipótesis nula** con un nivel de significancia de 5%. En otras palabras, la estabilidad de la varianza cumple basada en la **mediana**. (véase la Tabla 3.3)

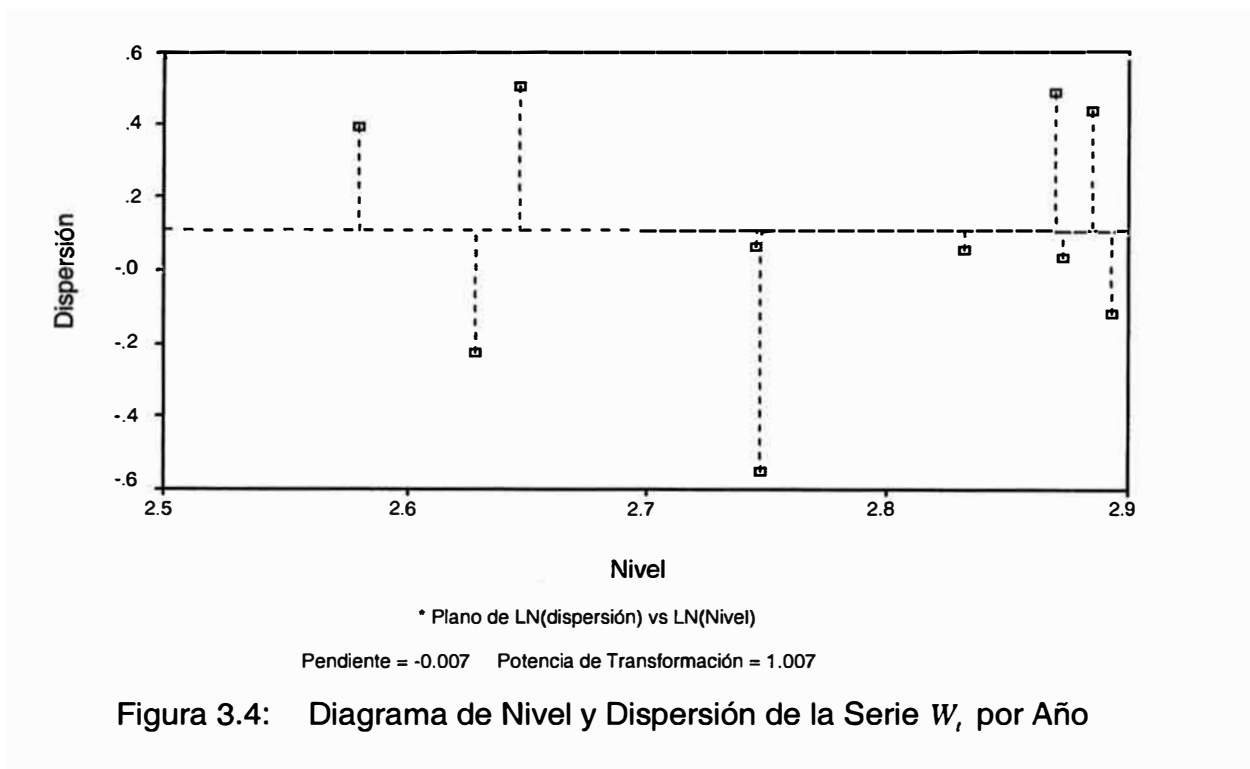


Figura 3.4: Diagrama de Nivel y Dispersión de la Serie W_t por Año

En el diagrama **de nivel y dispersión**, (véase la Figura 3.4), se nota que a medida que pasa el tiempo, el problema de la falta de estabilidad de la varianza desaparece, incluso la pendiente de la recta de regresión para la nube de puntos representada es -0.007 , en otras palabras es horizontal y la transformación sugerida es la 1.007, es decir, es cercana a 1, lo cual implica que no hay cambios en la serie. **Concluimos que la serie W_t es estacionaria en varianza.**

3.2.2 Estabilidad en Medias y de la Estacionalidad

La serie de datos W_t , también tiene que ser estacionaria en media, cuando pueda mantenerse el supuesto de media constante para toda la serie, es decir, cuando ésta fluctúe en torno a una única media.

Si la tendencia presente en el modelo es creciente o decreciente esto implica no **estacionariedad en media** o que existen varias medias en la serie, para comprobar lo anterior veremos algunos gráficos.

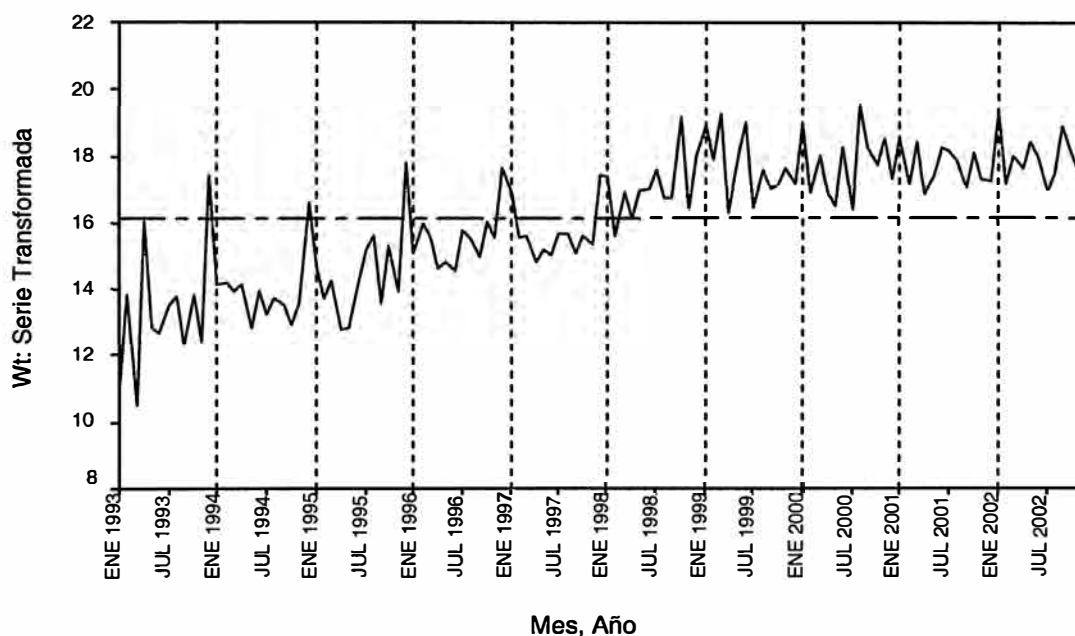


Figura 3.5: Serie Transformada de la Recaudación de Ingresos Corrientes

Se observa una clara tendencia alcista en los datos, lo que es el principal síntoma de **no estacionariedad en media** (véase la Figura 3.5). Por otro lado, hay algunos valores por debajo de la media y otros por sobre la media, lo que también indica **no estacionariedad en media**. La media es la línea horizontal de la serie W_t .

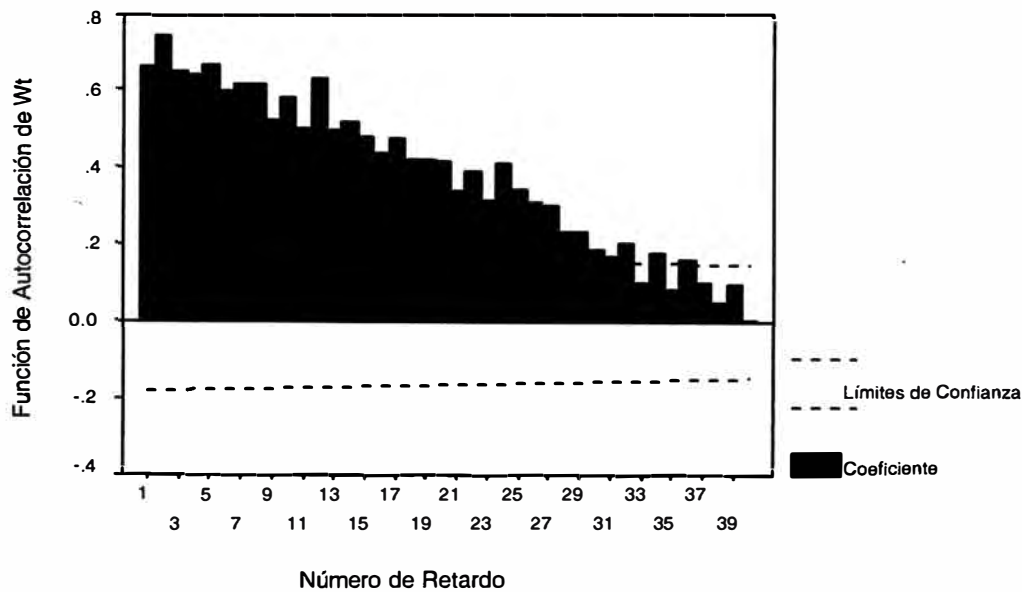


Figura 3.6: Correlograma de la serie transformada W_t ,

El gráfico anterior nos muestra el correlograma de la serie, (véase la Figura 3.6), específicamente la función de autocorrelación. Se nota como los picos van descendiendo paulatinamente a cero, los primeros 36 rezagos de la serie están fuera de los límites de confianza, **otra evidencia** de que **la serie no es estacionaria en media**.

La observación más importante es el hecho de que descienden lentamente a cero y si ya se estudió o vió otros modelos ARIMA, en éste no es tan patente como aquellos, porque tiene una pequeña particularidad que se verá más adelante y que tiende a hacer no tan patente la visualización de la **falta de estacionariedad** y es que presenta una falta de **estacionariedad también estacionalmente**.

Tomando diferencias a la serie de orden **d** y **D**; para la parte **no estacional** y para la parte **estacional** respectivamente se conseguirá en general que la serie sea **estacionaria en media**.

Se usará el calculo de varianzas de la serie original W_t y de la serie sometida a diferentes diferenciaciones, tomándose como diferenciación adecuada aquella para la cual la varianza o la **desviación estándar es mínima**.

Mas adelante se utilizaran gráficos de serie y la función de autocorrelación estimada FACE

Tabla 3.4: Estadística Descriptiva de las series transformadas con diferencia estacionaria, doble diferencia estacionaria, diferencia estacional, diferencia estacionaria y estacional

Serie	N	Mínimo	Máximo	Media	Desviación Estándar
Wt	120	10.49	19.56	16.134	1.974
DIFF(Wt,1)	119	-3.37	5.55	0.062	1.545
DIFF(Wt_1,1)	118	-8.75	8.93	-0.018	2.786
SDIFF(Wt,1,12)	108	-2.03	3.55	0.517	1.118
SDIFF(Wt_1,1,12)	107	-5.34	3.32	-0.021	1.464

Luego se escoge la serie $SDIFF(W_t,1,12)$ que es la diferencia estacional de W_t cuya desviación estándar es 1.118, y es la menor entre todas las demás series (véase la Tabla 3.4)

Donde:

$DIFF(W_t,1)$ es equivalente a:

$$\Delta W_t = (1 - L)W_t = W_t - W_{t-1} \tag{3-1}$$

$DIFF(W_t, 1, 1)$ es equivalente a:

$$\Delta \cdot \Delta W_t = (1 - L) \cdot (1 - L)W_t = (1 - 2L - L^2)W_t = W_t - 2W_{t-1} + W_{t-2} \quad (3-2)$$

$SDIFF(W_t, 1, 12)$ es equivalente a:

$$\Delta_{12}W_t = (1 - L^{12})W_t = W_t - W_{t-12} \quad (3-3)$$

$SDIFF(W_t, 1, 1, 12)$ es equivalente a:

$$\Delta \cdot \Delta_{12}W_t = (1 - L) \cdot (1 - L^{12})W_t = (1 - L - L^{12} + L^{13})W_t \quad (3-4)$$

Luego
$$\Delta \cdot \Delta_{12}W_t = W_t - W_{t-1} - W_{t-12} + W_{t-13} \quad (3-5)$$

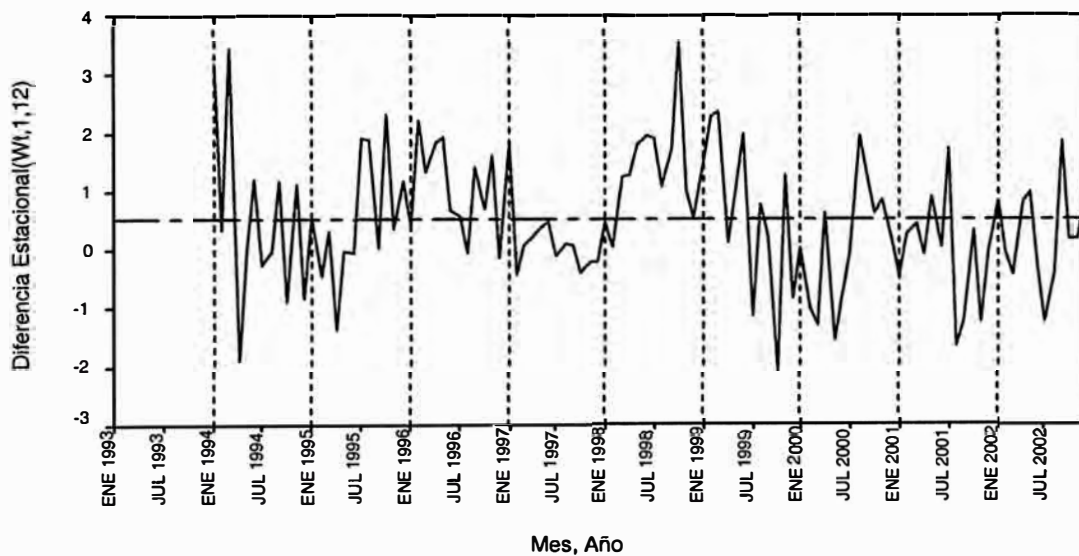


Figura 3.7: Diferencia Estacional de la serie transformada W_t

El grafico esta representando que la serie esta **estacionalizada en media** (véase la Figura 3.7)

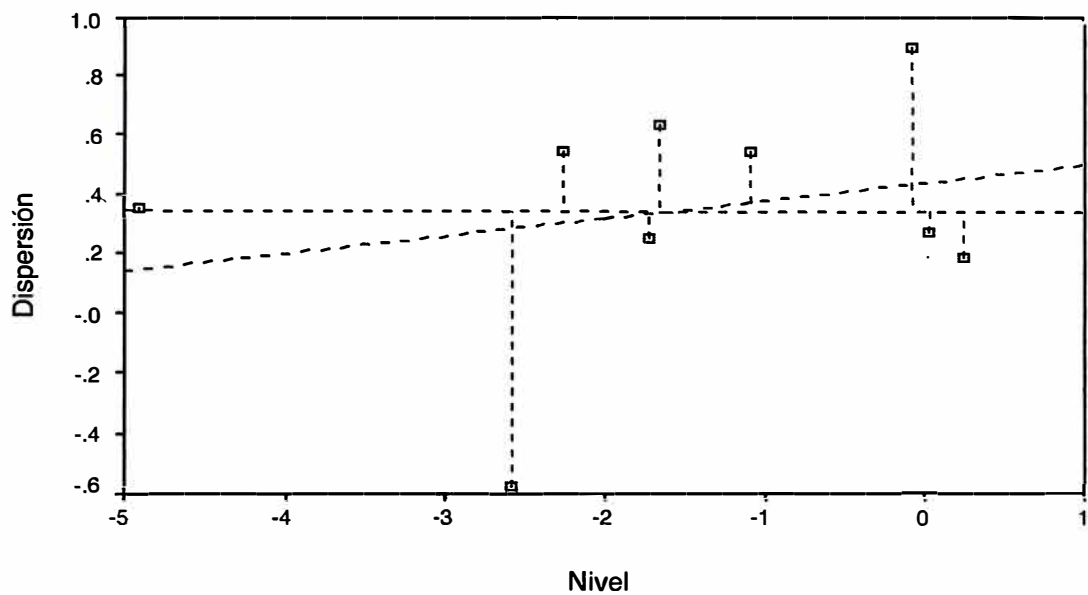
Para detectar si la varianza de la serie $SDIFF(W_t, 1, 12)$, es constante, es decir si no cambia con el tiempo, agruparemos las observaciones por años y aplicaremos la prueba de Levene para la homogeneidad de varianzas. Dado que el periodo de ajuste comprende desde 1994 hasta 2002, ambos años inclusive, dispondremos de 9 grupos con 12 observaciones cada uno. La prueba de Levene permitirá contrastar la hipótesis nula de que no existen diferencias significativas entre las varianzas de la serie en los 9 grupos definidos. A continuación se muestra los resultados.

Tabla 3.5: Prueba de Homogeneidad de Varianza de la diferencia estacional W_t

Diferencia Estacional de la Serie W_t	Estadístico de Levene	k-1	N-k	Nivel de Significancia
Basado en la Media	2.113	8	99	0.041
Basado en la Mediana	1.717	8	99	0.104
Basado en la Mediana y con ajuste en N-k	1.717	8	71	0.109
Basado en la Media ajustada	2.107	8	99	0.042

Luego es posible aceptar la hipótesis nula con un nivel de significancia de 5%, basada en la mediana (véase la Tabla 3.5).

En el gráfico de nivel y dispersión, (véase la Figura 3.8), de la serie $SDIFF(W_t, 1, 12)$, se observa que la estabilidad de la varianza esta basada en la mediana. La pendiente de la recta de regresión para la nube de puntos representada es 0,061 y la transformación sugerida es la 0.939, es decir, es cercana a 1, lo cual implica que no hay cambios en la serie. **Concluimos que la serie es estacionaria en varianza.**



* Plano de LN(Dispersión) vs LN(Nivel)

Pendiente = 0.061 Potencia de Transformación = 0.939

Figura 3.8: Diagrama de Nivel y Dispersión de la diferencia estacional de W_t ,

Tabla 3.6: Orden de diferenciación de la serie W_t ,

SERIE	Diferenciación Estacionaria	Diferenciación Estacional
	d	D
W_t	0	0
$DIFF(W_t,1)$	1	0
$DIFF(W_{t-1},1)$	2	0
$SDIFF(W_{t-1},12)$	0	1
$SDIFF(W_{t-1},1,12)$	1	1

Como la serie $SDIFF(W_t,1,12)$, tiene la **menor desviación estándar** (véase la Tabla 3.4), **estacionaria en media estacionalmente** (véase la Figura 3.7) y **estacionaria en varianza** (véase Tabla 3.5 y la Figura 3.8). Concluimos que la diferenciación estacionaria es **d=0** y la diferenciación estacional es **D=1** (véase la Tabla 3.6)

3.2.3 Identificación del Modelo Generador

Antes de determinar los valores de p y q , del modelo de la parte no estacional (estacionaria) previamente se debe tener en cuenta las características para la elección del modelo apropiado (véase la Tabla 1.1).

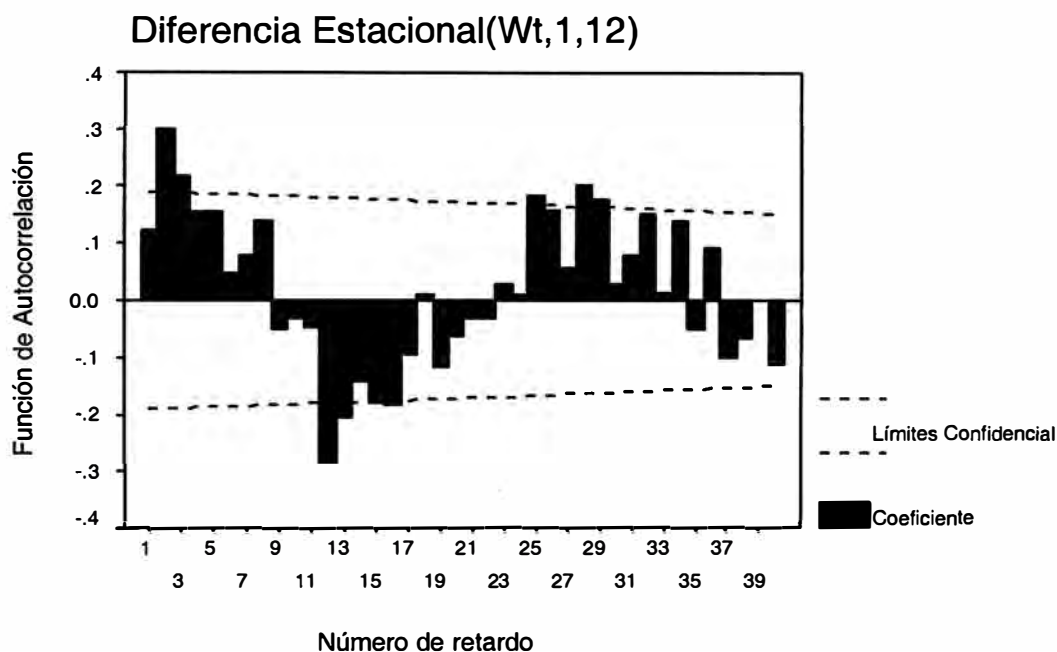


Figura 3.9: Función de Autocorrelación de la diferencia estacional de W_t .

Observando el gráfico siguiente se aprecia que la función de autocorrelación parcial desciende a cero después del **segundo rezago** (véase la Figura 3.10); de acuerdo a las pautas de la tabla, (véase la Tabla 1.1), el valor es $p=2$, entonces el modelo es un **AR(2)**, posteriormente al momento de modelar haciendo sobreajustes se deberá tener en cuenta que el primer coeficiente es menor que la línea del limite superior, lo cual cae la posibilidad que ese primer coeficiente sea cero.

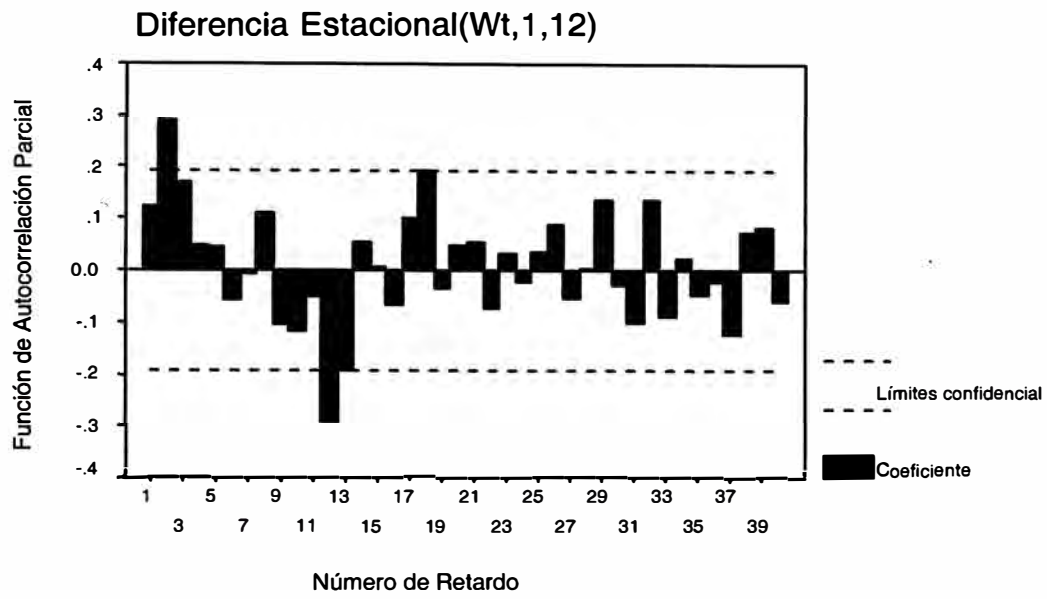


Figura 3.10: Función de Autocorrelación Parcial de la diferencia estacional de W_t

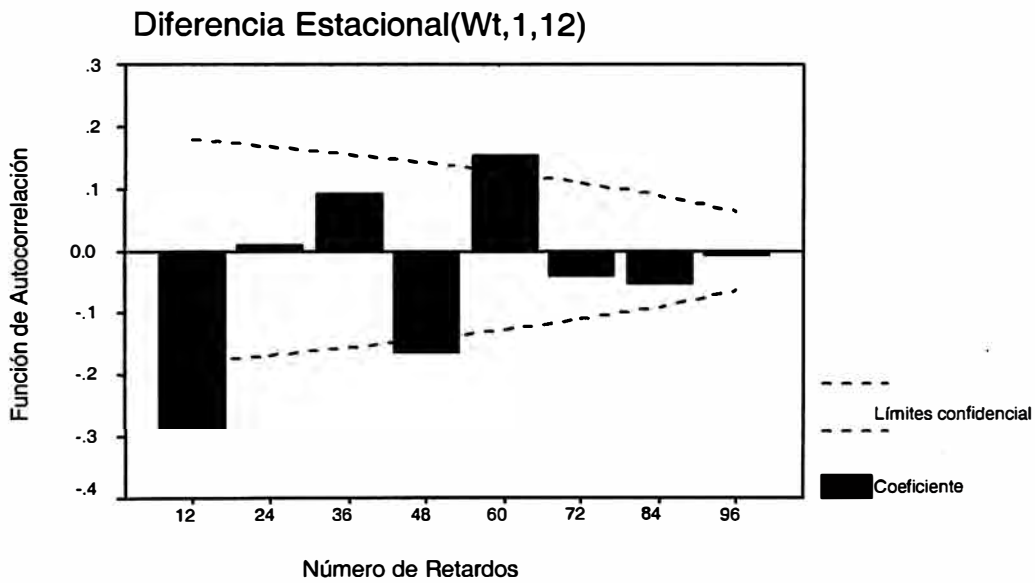


Figura 3.11: Función de Autocorrelación con retardos periódicos de la diferencia estacional de W_t

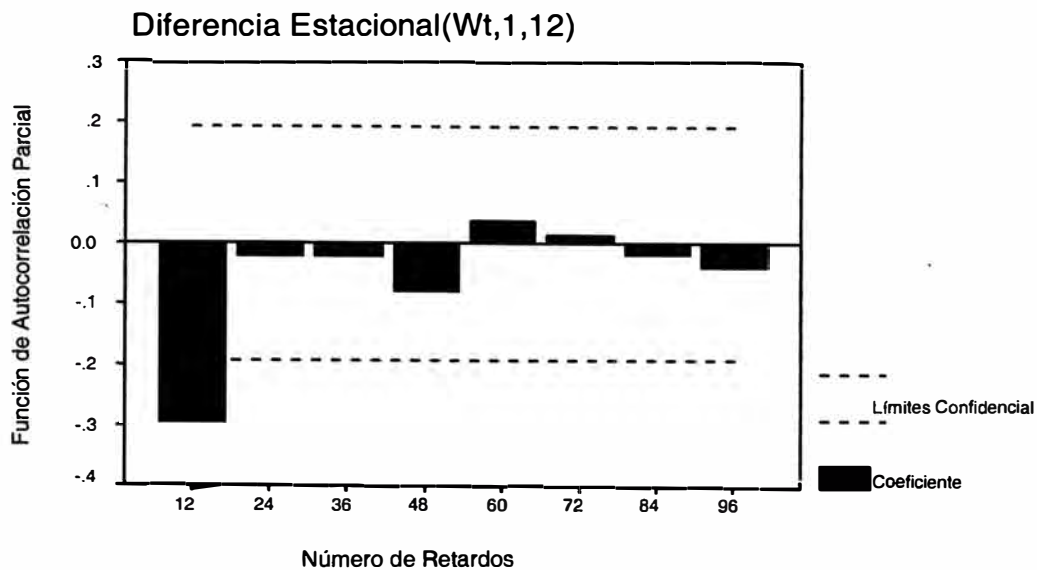


Figura 3.12: Función de Autocorrelación Parcial con retardos periódicos de la diferencia estacional de W_t

Para la función de autocorrelación desciende a cero después del **segundo rezago** (véase la Figura 3.9); de acuerdo a las pautas de la tabla, (véase la Tabla 1.1), el valor es **q=2**, entonces el modelo es un **MA(2)**, posteriormente al momento de modelar haciendo sobreajustes se deberá tener en cuenta que el primer coeficiente es menor que la línea del límite superior, lo cual cae la posibilidad que ese primer coeficiente sea cero.

Para determinar los valores de **P** y **Q**, del modelo de la parte **estacional** previamente se debe tener en cuenta las características para la elección del modelo apropiado (véase la Tabla 1.2).

Observando el gráfico anterior se aprecia que la función de autocorrelación parcial desciende a cero después del primer rezago estacional de $s=12$ meses, (véase la Figura 3.12); de acuerdo a las pautas de la tabla, (véase la Tabla 1.2), el valor es $P=1$ entonces el modelo es un $AR(1)_{12}$.

Para la función de autocorrelación desciende a cero después del primer rezago (véase la Figura 3.11); de acuerdo a las pautas de la tabla, (véase la Tabla 1.2), el valor es $Q=1$ entonces el modelo es un $MA(1)_{12}$.

Una vez identificado el **modelo generador**; concluimos que el modelo hallado es $ARIMA(2,0,2) \times ARIMA(1,1,1)_{12}$.

3.3 Fase de Estimación y Validación de los modelos

En la **fase de identificación** se utilizó básicamente como herramienta de trabajo el software del SPSS. En esta **fase de estimación** se utiliza también el software de Eviews, en el cual nos facilita la estimación de los coeficientes del modelo ARIMA

Previamente se debe tener en cuenta la inclusión del **termino independiente**. Como la **media** de la serie $SDIFF(W_t, 1, 12)$ es 0.517 (véase la Tabla 3.4), lo cual tiene una media significativa distinto de cero, lógicamente **el termino independiente se incluirá** en el modelo $ARIMA(2,0,2) \times ARIMA(1,1,1)_{12}$

Tabla 3.7: Coeficientes del Modelo ARIMA(2,0,2) x ARIMA(1,1,1)₁₂

Variable	Coeficiente	Error Estándar	t-estadístico	Probabilidad
C	0.505485	0.134483	3.758723	0.0003
AR(1)	0.431345	0.162382	2.656358	0.0094
AR(2)	0.403543	0.144708	2.788668	0.0065
SAR(12)	0.176755	0.10794	1.637526	0.1051
MA(1)	-0.294096	0.146448	-2.008195	0.0477
MA(2)	-0.166588	0.110407	-1.508851	0.1350
SMA(12)	-0.849546	0.036569	-23.23104	0.0000
R ²	0.420683	F-estadístico		10.5295
R ² ajustado	0.38073	Probabilidad (F-estadístico)		0.0000

La tabla muestra los estadísticos más importantes del modelo, lo que nos interesa en estos modelos que sus regresores sean estadísticamente significativos, y esto nos lo señala la última columna (véase la Tabla 3.7)

Para los coeficientes de las variables SAR(12) y MA(2), el $\alpha_{prob.}$ es mayor para un nivel de significancia de $\alpha = 0.05$; entonces se **acepta la hipótesis nula** que los coeficiente de las variables es **cero**; ósea **no es estadísticamente significativo**.

El modelo **ARIMA(2,0,2) x ARIMA(1,1,1)₁₂**, sus regresores no son **estadísticamente significativos**.

3.3.1 Sobreajuste del Modelo

El hecho de que el modelo identificado no sea el adecuado, no se debe descartar la posibilidad de que otro modelo algo más complejo pueda ajustarse mejor a la serie

observada. Para comprobar si algún modelo se ajusta mejor conviene utilizar la técnica de sobreajuste.

El sobreajuste consiste en estimar cada uno de los modelos obtenidos al aumentar en una unidad cada uno de los órdenes **p, q, P, Q**

Por lo tanto si el modelo candidato recién calculado es **ARIMA(2,0,2)** x **ARIMA(1,1,1)₁₂**

Los sobreajustes serán:

1^a) **ARIMA(3,0,2)** x **ARIMA(1,1,1)₁₂**

2^a) **ARIMA(2,0,3)** x **ARIMA(1,1,1)₁₂**

3^a) **ARIMA(2,0,2)** x **ARIMA(1,1,2)₁₂**

4^a) **ARIMA(2,0,2)** x **ARIMA(2,1,1)₁₂**

Procediendo a los cálculos de los modelos del sobreajuste, se determinó:

1^a) Los coeficientes de las variables AR(1), MA(1), MA(2) y SAR(12), no son significativos. Se modeló sin considerar AR(1) y MA(1) porque se tuvo en cuenta que el **primer coeficiente es menor** que la línea del límite superior para un **AR** y **MA**, (véase la Figura 3.9 y 3.10), lo cual cae la posibilidad que ese primer coeficiente sea **cero**. Luego modelando se obtuvo que MA(2) y SAR(12), no son significativos.

2^a) En este modelo los coeficientes de las variables AR(1), MA(1) y MA(2), no son significativos, luego se modeló sin considerar AR(1), MA(1) y MA(2), entonces el **modelo obtuvo un buen resultado** (véase la Tabla 3.8). En la última columna de la tabla el $\alpha_{prob.}$ fue menor para el nivel de significancia $\alpha = 0.05$ para todos los coeficientes del modelo.

Tabla 3.8: Coeficientes del Modelo ARIMA(2,0,3) x ARIMA(1,1,1)₁₂

Variable	Coeficiente	Error Estándar	t-estadístico	Probabilidad
C	0.440752	0.058133	7.581765	0.0000
AR(2)	0.252888	0.091217	2.772382	0.0068
SAR(12)	0.215204	0.064332	3.345227	0.0012
MA(3)	0.223729	0.099244	2.254318	0.0266
SMA(12)	-0.885796	0.000205	-4329.778	0.0000
R ²	0.421968	F-estadístico		16.2427
R ² ajustado	0.395989	Probabilidad (F-estadístico)		0.0000
Raíces invertidas AR	0.88	0.76+0.44i	0.76-0.44i	0.50
	0.44-0.76i	0.44+0.76i	0.88i	-0.88i
	-0.44+0.76i	-0.44-0.76i	-0.50	-0.76+0.44i
	-0.76-0.44i	-0.88		
Raíces invertidas MA	0.99	0.86+0.49i	0.86-0.49i	0.49+0.86i
	0.49-0.86i	0.30+0.53i	0.30-0.53i	-0.99i
	0.99i	-0.49-0.86i	-0.49+0.86i	-0.61
	-0.86+0.49i	-0.86-0.49i	-0.99	

3^a) Los coeficientes de las variables AR(1) y SAR(12) no son significativos. Se modeló sin considerar AR(1) y MA(1) entonces SAR(12), no era significativo.

4^a) Los coeficientes de las variables SAR(12) y SAR(24), no son significativos. Se modeló sin considerar SAR(12) entonces SAR(24), no era significativo.

El modelo determinado en los sobreajustes es: **ARIMA(2,0,3) x ARIMA(1,1,1)₁₂**.
 Donde los regresores **si son estadísticamente significativos** (véase la Tabla 3.8).
 Las **raíces invertidas** de la **parte autorregresiva** y la **media móvil** son menores a la **unidad** en valor absoluto. Por lo cual se concluye que el **modelo es estacionario e invertible** respectivamente (véase la Tabla 3.8).

Tabla 3.9: Matriz de Correlación de los Coeficientes Estimados

Variable	C	AR(2)	SAR(12)	MA(3)	SMA(12)
C	1.00E+00	-2.20E-02	-6.02E-03	-5.75E-03	-1.42E-03
AR(2)	-2.20E-02	1.00E+00	-8.81E-04	-1.12E-02	2.42E-04
SAR(12)	-6.02E-03	-8.81E-04	1.00E+00	1.42E-02	-9.33E-04
MA(3)	-5.75E-03	-1.12E-02	1.42E-02	1.00E+00	-1.31E-03
SMA(12)	-1.42E-03	2.42E-04	-9.33E-04	-1.31E-03	1.00E+00

Por último, en el análisis de la matriz de correlación de los coeficientes estimados en la tabla se observa la correlación entre dos coeficientes estimados es **cercano a cero**; esto quiere decir que **no hay problemas de multicolinealidad**, las **estimaciones son estables**, **no hay la existencia de sobreparametrización** del modelo, lo cual no crea dificultades en la convergencia del modelo (véase la Tabla 3.9).

3.3.2 Análisis de los Residuos del Modelo

En primer lugar se ha supuesto que la **serie histórica de Recaudación de ingresos corrientes (en soles constantes)** tiene memoria, se ha tratado de descubrir cuál es

la caja negra que genera la serie bajo análisis. Si el modelo que se ha desarrollado es la caja negra, **los residuos del modelo** cumplen una serie de características

Estas características o la validación consistirá en comprobar:

- 1º Qué la serie de residuales tiene **media cero**.
- 2º Que es estable en **varianzas**
- 3º Que sus observaciones están **incorrelacionadas** y que proceden de una **distribución normal**.

En otras palabras que sea un **ruido blanco**

Para la serie residual del modelo **la media es 0.0216** se aproxima a cero

Por la prueba estadística de Jarque Bera para un nivel de significancia $\alpha = 0.05$ se muestra que la serie residual esta normalmente distribuida con un $\alpha_{prob} = 0.0781$ ya que $\alpha < \alpha_{prob}$ (véase la Tabla 3.10).

Tabla 3.10: Estadística Descriptiva de la Serie Residual

Serie Residual	Estadísticos
Media	0.0216
Intervalo Confidencial de la Media al 95%	
Límite Inferior	-0.1439
Límite Superior	0.1870
Mediana	-0.0777
Varianza	0.6523
Desviación Estándar	0.8077
Mínimo	-1.5146
Máximo	2.2987
Rango	3.8133
Jarque-Bera	5.0990
Probabilidad	0.0781

Tabla 3.11: Prueba de la Homogeneidad de la Varianza de la Serie Residual

Serie Residual(Diferencia Estacional de la Serie W_t)	Estadístico de Levene	k-1	N-k	Nivel de Significancia
Basado en la Media	3.104	7	86	0.006
Basado en la Mediana	2.606	7	86	0.017
Basado en la Mediana y con ajuste en N-k	2.606	7	67	0.019
Basado en la Media ajustada	3.045	7	86	0.007

Analizando la homogeneidad de varianza por la prueba de Levene, se tiene que la serie de residuales basado en la mediana con un 1% de nivel de significancia cumple la hipótesis que la varianza es homogénea, (véase la Tabla 3.11).

Por lo cual se concluye que el modelo, cumple con las condiciones de **ruido blanco**.

Además la **incorrelación de la serie residual del modelo**, en ambos gráficos de FAC y FACP (véase las Figuras 3.13 y 3.14); se puede apreciar que ninguno de los picos queda fuera del intervalo de confianza por lo tanto los errores están incorrelacionados unos con otros. El estadístico de Ljung-Box hasta el rezago 30 tiene $\alpha_{prob} > \alpha$, donde el nivel de significancia es $\alpha = 0.05$ incluso hay α_{prob} cercanos a la unidad con lo que se está señalando que no están correlacionados.

El coeficiente de determinación ajustada \hat{R}^2 es una medida de carácter relativo del modelo y su valor está a un 39.59% bondad de ajuste; aunque no es cercana al cien por ciento pero el modelo ARIMA hallado se comporta como un ruido blanco (véase la Tabla 3.8).

El nivel de significancia de la F calculada del modelo es menor para un nivel de significación $\alpha = 0.05$, entonces se **acepta la hipótesis de estabilidad estructural** del modelo hallado (véase la Tabla 3.8).

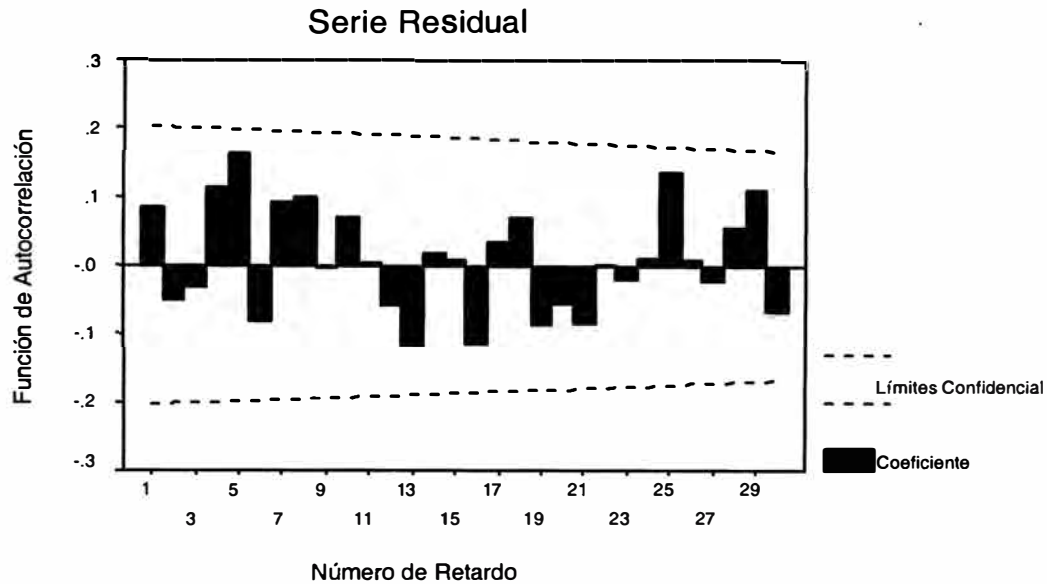


Figura 3.13: Función de Autocorrelación de la Serie Residual

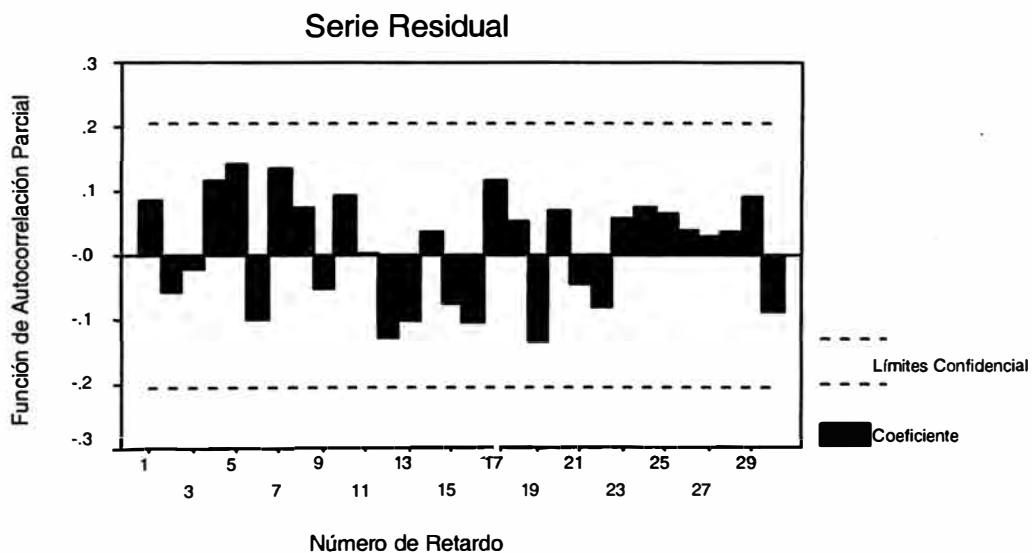


Figura 3.14: Función de Autocorrelación Parcial de la Serie Residual

3.4 Fase de Predicción

Se tiene el modelo definitivo, el mismo tiene residuos que cumplen con todos los supuestos de los Modelos ARIMA. Por lo tanto se procederá a hacer una proyección. Ahora se tomará todos los datos, es decir los datos del ajuste y los datos de validación, en otras palabras la serie comienza en enero de 1993 y termina en junio del 2003. De esta manera se podrán comparar los datos verdaderos y las proyecciones que hace el modelo.

El modelo ARIMA elegido es: **ARIMA(2,0,3) x ARIMA(1,1,1)**₁₂; donde la serie es un **Modelo Estacional Multiplicativo Estacionario** (véase la Tabla 3.8)

Formulando el modelo tenemos:

$$\phi(L)\Phi(L^s)\Delta^d\Delta_s^D(W_t + C) = \theta(L)\Theta_s(L^s)a_t \quad \text{donde } W_t = Z_t^\lambda \quad (3-6)$$

Los valores de los coeficientes del modelo se toma de la tabla (véase la Tabla 3.8).

$$\phi(L) = (1 - \phi_1 L - \phi_2 L^2) = (1 - 0.2528L^2) \quad \Phi(L^s) = (1 - \Phi_{12}L^{12}) = (1 - 0.2152L^{12}) \quad (3-7)$$

$$\Delta^d = (1 - L)^0 = 1 \quad \Delta_s^D = (1 - L^s)^D = (1 - L^{12}) \quad (3-8)$$

$$\theta(L) = (1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3) = (1 - 0.2237L^3) \quad (3-9)$$

$$\Theta_s(L^s) = (1 - 0.885796L^{12}) \quad C = 0.440752 \quad (3-10)$$

Luego se reemplaza (3-7), (3-8), (3-9) y (3-10) en la formula del modelo (3-6), tenemos:

$$(1 - 0.2528L^2)(1 - 0.2152L^{12})(1 - L^{12})(W_t - C) = (1 - 0.2237L^3)(1 + 0.8858L^{12})a_t \quad (3-11)$$

$$W_t - 0.25W_{t-2} - 1.21W_{t-12} + 0.31W_{t-14} + 0.21W_{t-24} - 0.054W_{t-26} = a_t - 0.22a_{t-3} - 0.88a_{t-12} + 0.19a_{t-15} \quad (3-12)$$

Usando la transformaciones de la serie:

$$W_t = (Z_t)^{0.185} \quad (3-13)$$

se obtiene el modelo en función de la serie histórica:

$$Z_t = \{0.25(Z_{t-2})^{0.185} + 1.21(Z_{t-12})^{0.185} - 0.31(Z_{t-14})^{0.185} - 0.21(Z_{t-24})^{0.185} + 0.054(Z_{t-26})^{0.185} + a_t - 0.22a_{t-3} - 0.88a_{t-12} + 0.19a_{t-15}\}^{\frac{1000}{185}} \quad (3-14)$$

donde Z_t es la variable modelada, (en este caso la Recaudación de Ingresos Corrientes por la Fuente de Financiamiento de Recursos Directamente Recaudados), a_t es el termino error y t es el tiempo.

Por tanto, las predicciones serán a partir de $t = T + 1$, siendo T el instante correspondiente a la última observación de la serie histórica. La siguiente ecuación para la predicción es el siguiente:

$$Z_t = \{0.25(Z_{t-2})^{0.185} + 1.21(Z_{t-12})^{0.185} - 0.31(Z_{t-14})^{0.185} - 0.21(Z_{t-24})^{0.185} + 0.054(Z_{t-26})^{0.185} - 0.22a_{t-3} - 0.88a_{t-12} + 0.19a_{t-15}\}^{\frac{1000}{185}} \quad (3-15)$$

Al incluir los datos de los meses de enero a junio del 2003 se pudo comparar los datos verdaderos y las proyecciones que hace el modelo y los resultados que se muestran, (véase la Tabla 3.12 y la Tabla 3.13), por ejemplo el **error absoluto porcentual medio** es de **6.18%**. Además teniendo en cuenta que las a_t se distribuyen normalmente se pueden establecer intervalos de confianza del siguiente tipo, donde el nivel de significancia es $\alpha = 0.05$.

$$Pr ob[W'_t - 1.96\sqrt{ECM(W'_t)} < Actual < W'_t + 1.96\sqrt{ECM(W'_t)}] = 0.95 \quad (3-16)$$

Tabla 3.12: Error de Predicción de la diferencia estacional de la serie W_t

Pronóstico: W_t	
Actual: Diferencia Estacional de la serie W_t	
Pronóstico de la muestra: 2003:01 2003:06	
Inclusión de observaciones: 6	
Raíz del Error Cuadrático Medio	1.14
Error Absoluto Medio	1.08
Error Absoluto Porcentual Medio	6.18

Tabla 3.13: Pronóstico, Actual e Intervalos de Confianza de enero a junio del 2003

Año y Mes	Pronóstico W_t	Actual	Dif. Absoluta	%error absoluto por mes	Intervalo de Confianza al 95%	
					Límite Inferior	Límite Superior
2003.01	19.72	17.92	1.80	10.05	17.48	21.96
2003.02	17.69	16.76	0.92	5.52	15.45	19.93
2003.03	18.76	17.82	0.94	5.28	16.52	21.00
2003.04	17.82	17.22	0.60	3.46	15.58	20.06
2003.05	18.49	17.42	1.07	6.13	16.25	20.73
2003.06	18.66	17.50	1.17	6.66	16.42	20.90

Luego tomando la serie w'_t (Pronóstico) y el Actual, (véase la Tabla 3.13), la ecuación (3-16) y usando las transformaciones correspondientes se obtuvo los siguientes resultados (véase la Tabla 3.14).

Tabla 3.14: Pronóstico, Serie Histórica e Intervalos de Confianza de enero a junio del 2003

Año y Mes	Pronóstico Z_t	Serie Histórica Z_t	Dif. Absoluta	%error absoluto por mes	Intervalo de Confianza al 95%	
					Límite Inferior	Límite Superior
2003.01	6,553,413	5,955,180	598,233	10.05	5,860,206	7,246,620
2003.02	4,377,839	4,148,933	228,906	5.52	3,684,632	5,071,046
2003.03	6,085,860	5,780,843	305,017	5.28	5,392,653	6,779,066
2003.04	4,968,913	4,802,691	166,222	3.46	4,275,707	5,662,120
2003.05	5,424,673	5,111,449	313,224	6.13	4,731,466	6,117,880
2003.06	5,581,835	5,233,308	348,527	6.66	4,888,628	6,275,042

Donde la raíz del error cuadrático medio es de 353,677 y el error absoluto porcentual medio es de 6.18% (véase la Tabla 3.15).

Tabla 3.15: Error de Pronóstico de la serie Z't

Pronóstico: Z't	
Serie Histórica: Zt	
Pronóstico de la muestra: 2003:01 2003:06	
Inclusión de observaciones: 6	
Raíz del Error Cuadrático Medio	353,677
Error Absoluto Medio	326,688
Error Absoluto Porcentual Medio	6.18

Para la pronosticar de julio a diciembre del 2003, se considera las proyecciones que hace el modelo a través del software E-views.

Su respectivo intervalo de confianza por mes se considero como raíz del error cuadrático medio el valor 353,677 y su nivel de significancia de $\alpha = 0.05$.

Donde el pronóstico estimado anual es de S/. 66,380,913, su mínimo calculado es de S/. 58,062,429 y su máximo es S/. 74,699,396 (véase la Tabla 3.16).

Gráficamente se muestra el pronóstico, los intervalos de confianza al 95% de los meses de enero a diciembre del 2003.

La serie histórica de los meses de enero a junio del 2003 que sirvió para comparar con las proyecciones que hace el modelo (véase la figura 3.15).

Tabla 3.16: Pronóstico, Serie Histórica e Intervalos de Confianza de enero a diciembre del 2003

Año y Mes	Pronóstico W't	Pronóstico Z't	Serie Histórica Zt	Intervalo de Confianza al 95%	
				Límite Inferior	Límite Superior
2003.01	19.72	6,553,413	5,955,180	5,860,206	7,246,620
2003.02	17.69	4,377,839	4,148,933	3,684,632	5,071,046
2003.03	18.76	6,085,860	5,780,843	5,392,653	6,779,066
2003.04	17.82	4,968,913	4,802,691	4,275,707	5,662,120
2003.05	18.49	5,424,673	5,111,449	4,731,466	6,117,880
2003.06	18.66	5,581,835	5,233,308	4,888,628	6,275,042
2003.07	18.18	5,396,348		4,703,141	6,089,555
2003.08	18.69	5,549,323		4,856,116	6,242,530
2003.09	18.81	5,583,447		4,890,240	6,276,654
2003.10	19.04	5,652,471		4,959,264	6,345,678
2003.11	18.40	5,461,754		4,768,547	6,154,961
2003.12	19.35	5,745,037		5,051,830	6,438,244
Total	223.61	66,380,913		58,062,429	74,699,396

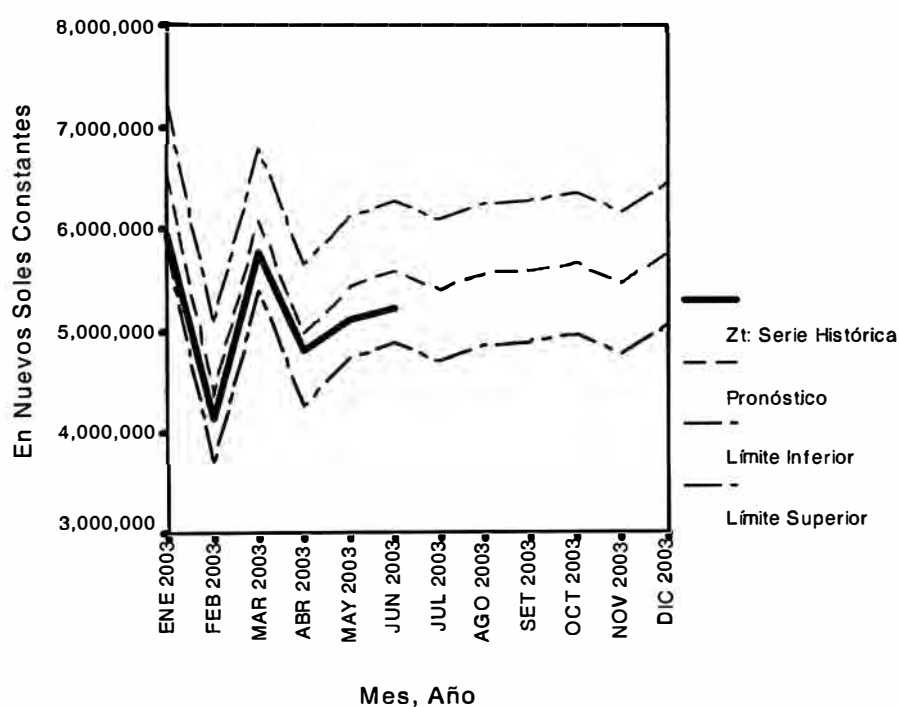


Figura 3.15: Serie Histórica, Pronóstico e Intervalos de Confianza de enero a diciembre 2003.

CONCLUSIONES

1. Esta serie se caracteriza por poseer tendencia creciente y estacionalidad anual. El modelo óptimo calculado resulta ser el siguiente:

$$Z_t = \{0.25(Z_{t-2})^{0.185} + 1.21(Z_{t-12})^{0.185} - 0.31(Z_{t-14})^{0.185} - 0.21(Z_{t-24})^{0.185} + \\ + 0.054(Z_{t-26})^{0.185} + a_t - 0.22a_{t-3} - 0.88a_{t-12} + 0.19a_{t-15}\}^{\frac{1000}{185}}$$

donde Z_t es la variable modelada, (en este caso la Recaudación de Ingresos Corrientes por la Fuente de Financiamiento de Recursos Directamente Recaudados), a_t es el termino error y t es el tiempo.

2. Las predicciones serán a partir de $t = T + 1$, siendo T el instante correspondiente a la última observación de la serie histórica. La ecuación para la predicción es el siguiente:

$$Z_t = \{0.25(Z_{t-2})^{0.185} + 1.21(Z_{t-12})^{0.185} - 0.31(Z_{t-14})^{0.185} - 0.21(Z_{t-24})^{0.185} + \\ + 0.054(Z_{t-26})^{0.185} - 0.22a_{t-3} - 0.88a_{t-12} + 0.19a_{t-15}\}^{\frac{1000}{185}}$$

3. Al comparar la **serie histórica** de los meses de enero a junio del 2003 con las **proyecciones** que hace el **modelo**, se obtuvo un **error absoluto porcentual medio** de **6.18%**, lo cual indica un aceptable grado de precisión.
4. La **proyección anual** para el año **2003** que se estima recaudar es de **S/. 66,380,913**, siendo esta la **menor recaudación de ingresos corrientes** de los **últimos cuatro años**.
5. El coeficiente negativo en el componente **estacional de la media móvil** en el modelo estimado corresponde a dos fenómenos principales. En primer lugar, el componente esta relacionado estacionalmente con los datos mensuales con una **tendencia de menor Recaudación de Ingresos** en el año **2001** respecto al año **2002**. En segundo lugar esta relacionado con la existencia de una **tendencia central** de la serie en los **cuatro últimos años** desde **enero de 1999** hasta **diciembre del 2002** donde se puede apreciar, que la Recaudación de Ingresos Corrientes de la UNI no tuvo una tendencia creciente respecto a los años anteriores.

RECOMENDACIONES

1. Con frecuencia las series económicas que se analizan están sujetas a la influencia de factores externos de carácter excepcional, por ejemplo las **medidas de política económica**, los cambios tecnológicos, las huelgas, **las campañas publicitarias** y **la competencia entre empresas** que prestan el mismo servicio. La UNI debe estar vigilante ante este reto; **modernizando su infraestructura** y **desarrollando nuevas tecnologías** útiles en el mercado para que a través de la **prestación de servicios** su Recaudación de Ingresos Corrientes, sea **creciente** en el tiempo.
2. En la modelización de la serie económica en estudio, se deberá tener en cuenta la técnica del **análisis de intervención**; y esto contribuiría en obtener un modelo ARIMA mas óptimo y con una mejor proyección.
3. Tener actualizado el **modelo** con los **últimos datos** para establecer un mecanismo de apoyo **teórico practico**, y de esa forma lograr óptimas proyecciones, lo cual contribuirá a una eficiente toma de decisiones.

4. Para un futuro estudio, se debe analizar las **genéricas** de la **Recaudación** de los **ingresos corrientes** que conforman la serie histórica como: **Tasas, Venta de bienes, Prestación de servicios y Rentas de la propiedad**. Esto dará como resultado implementar estrategias de competitividad en las dependencias y facultades de la UNI, para el desarrollo y crecimiento institucional.

5. En un proceso de planificación estratégica institucional, los modelos estocásticos ARIMA de series temporales proporcionan una alternativa eficaz para describir y predecir el comportamiento futuro de la Recaudación de Ingresos Corrientes. Por lo cual sugiero su aplicación.

APÉNDICE 1

AP.1 Deflación de series en el tiempo.

En términos matemáticos, éste método de deflación de series en el tiempo es estrictamente aplicable sólo si los números índice cumplen el criterio de inversión de factores, y por esta razón el índice de precios se calcula con la formula de Laspeyres que comprende de un conjunto de “ponderaciones fijas”.

Las ponderaciones son fijas en el sentido de las cantidades consumidas, pero no en el que respecta al valor de cada componente, ya que el mismo cambia su participación con la variación de los precios relativos.

La formula de Laspeyres es comúnmente presentada bajo la formula:

Índice de Precios de Laspeyres

$$w_i = p_{io} \cdot q_{io} \quad L_p = \frac{\sum_{i=1}^N \frac{p_{it}}{p_{io}} p_{io} q_{io}}{\sum_{i=1}^N p_{io} q_{io}} = \frac{\sum_{i=1}^N p_{it} q_{io}}{\sum_{i=1}^N p_{io} q_{io}} \quad (\text{AP-1})$$

De este modo, todos los países del mundo, al aplicar la fórmula de Laspeyres, pueden dar a conocer los resultados de la evolución de los precios al consumidor, con metodologías homogéneas que garantizan la comparabilidad internacional en la medición de los precios.

Los cambios que se producen en los precios de los bienes y servicios a través del tiempo, ocasionan problemas al tratar de comparar valores monetarios en periodos diferentes, debido a que tienen poderes adquisitivos distintos.

Para efectuar un adecuado análisis y alcanzar una conclusión correcta en relación al comportamiento de una variable, es necesario homogenizar los valores de la variable en un mismo poder de compra, de tal forma que se elimine o aisle el efecto precio.

Este proceso de transformación de valores nominales a valores reales (deflactación), requiere la aplicación de un índice deflactor que elimine el efecto de los precios.

En la deflactación de valores se debe de utilizar el índice más apropiado, generalmente se utiliza como deflactor, el Índice de Precios al Consumidor (IPC).

AP.2 Fórmula de Cálculo

$$VR \text{ de "t" a precios de "o"} = \frac{VN_t}{\frac{IPC_t}{IPC_o}} \quad (\text{AP-2})$$

VR de "t" a precios de "o" = valor real de una variable en el período "t" a precios del período "o".

VN_t = Valor Nominal de una variable en el período "t"

IPC_t = Índice de Precios al Consumidor del período "t"

IPC_o = Índice de Precios al Consumidor del período "o"

El subíndice "o" indica que se trata del **período base**, es decir **año 2001**, y el subíndice "t" corresponde al **período actual**, es decir **diciembre del año 2002**. (Se utilizará en la deflactación de la serie original, para luego pronosticar valores reales del año 2003).

El campo de aplicación de la deflactación es muy amplio en el análisis económico, por intermedio de ella, se puede calcular: las remuneraciones reales, la tasa de interés real, la emisión primaria en términos reales, el crédito real del sector bancario, entre otros.

ANEXO 1: Desagregado De La Variable Recaudación De Ingresos Corrientes

(Fuente de Financiamiento de Recursos Directamente Recaudados).

CODIGO DE PARTIDA	DENOMINACION
1.0.0	INGRESOS CORRIENTES
1.2.0	TASAS
1.2.3	DE EDUCACION
1.2.3.001	CARNETS
1.2.3.002	DERECHO DE EXAMEN DE ADMISION
1.2.3.003	GRADO, TITULOS, CONSTANCIAS Y CERTIFICADOS
1.2.3.004	DERECHOS UNIVERSITARIOS
1.2.3.006	TESIS
1.2.3.008	PENSION DE ENSEÑANZA
1.2.3.009	MATRICULAS
1.2.3.010	CONVALIDACION
1.2.3.011	TRASLADO
1.2.3.099	OTROS
1.2.4	DE SALUD
1.2.4.001	TASAS DE SALUD
1.2.4.003	EXAMEN MEDICO
1.4.0	VENTA DE BIENES
1.4.4	DE EDUCACION, RECREACION
1.4.4.001	VENTA DE LIBROS, BOLETINES
1.5.0	PRESTACION DE SERVICIOS
1.5.4	DE EDUCACION, RECREACION
1.5.4.001	CENTRO PRE-UNIVERSITARIO
1.5.4.003	SEMINARIOS, FORUMS, CHARLAS
1.5.4.004	CURSOS
1.5.4.007	ACTIVIDADES RECREATIVAS
1.5.4.099	OTROS
1.5.6	OTROS
1.5.6.007	SERVICIOS DE LABORATORIO
1.5.6.010	SERVICIOS DE ASESORIA
1.5.6.014	SERVICIOS DE IMPRENTA
1.5.6.020	PROYECTOS DE INVESTIGACION
1.5.6.023	SERVICIOS DE PROCESAMIENTO AUTOMATICO
1.5.6.030	SERVICIOS A TERCEROS
1.5.6.099	OTROS
1.6.0	RENTAS DE LA PROPIEDAD
1.6.1	DE LA PROPIEDAD REAL
1.6.1.006	CONCESION
1.6.1.099	OTROS
1.6.2	DE LA PROPIEDAD FINANCIERA
1.6.2.004	INTERESES BANCARIOS
1.6.2.099	OTROS
1.6.3	OTROS
1.6.3.001	DIFERENCIAL CAMBIARIO

Fuente: Presupuesto Del Sector Publico 2003 - Compendio De Normas Presupuestales- Dirección Nacional De Presupuesto Publico – MEF.

ANEXO 2: Índice de Precios Promedio Mensual al Consumidor de Lima Metropolitana: 1993 – 2003

En este anexo se presenta los Índice de Precios Promedio Mensual al Consumidor de Lima Metropolitana: 1993 a 2003 y de Base Diciembre Año 2001=100.00), para deflactar la serie original de Soles Corrientes a Soles Constantes.

INDICE DE PRECIOS PROMEDIO MENSUAL AL CONSUMIDOR DE LIMA METROPOLITANA: 1993 - 2003

(Base: Dic 2001=100.00)

MESES	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
ENERO	43.5728	59.0311	67.1325	74.6406	82.8503	88.5782	93.0685	96.5900	100.3154	99.4801	101.7508
FEBRERO	44.8511	60.1059	67.8954	75.7855	82.9240	89.6717	93.3617	97.0538	100.5616	99.4410	102.2278
MARZO	46.7512	61.5030	68.8269	76.8314	83.9804	90.8544	93.9331	97.5791	101.0723	99.9764	103.3705
ABRIL	48.8209	62.4532	69.5045	77.5008	84.3056	91.4096	94.4870	98.0783	100.6515	100.7055	103.3179
MAYO	50.3015	62.9002	70.0847	78.0646	84.9422	91.9521	94.9322	98.0947	100.6760	100.8453	103.2848
JUNIO	51.2155	63.6176	70.6515	78.4280	85.8649	92.4398	95.1027	98.1578	100.6182	100.6161	102.7962
JULIO	52.6184	64.1837	71.0543	79.5056	86.5764	93.0203	95.3531	98.6646	100.7920	100.6506	
AGOSTO	53.9517	65.1676	71.7917	80.2405	86.7727	93.2663	95.5162	99.1264	100.4866	100.7522	
SETIEMBRE	54.8261	65.5033	72.0727	80.5006	87.0265	92.7646	95.9556	99.6776	100.5497	101.2293	
OCTUBRE	55.6529	65.6912	72.4396	81.0919	87.1584	92.4539	95.8397	99.9095	100.5873	101.9561	
NOVIEMBRE	56.5456	66.4931	73.3401	81.4714	87.2261	92.4821	96.1055	99.9733	100.0916	101.5489	
DICIEMBRE	57.9662	66.8836	73.7246	82.4536	87.7826	93.0555	96.5231	100.1275	100.0000	101.5158	
PROMEDIO	51.4228	63.6278	70.7099	78.8762	85.6175	91.8290	95.0149	98.5861	100.5335	100.7264	51.3957

Fuente: INEI: Boletín Mensual, Agosto 2003, N° 08-2003-INEI

ANEXO 3: Recaudación de Ingresos Corrientes de la UNI (En Nuevos Soles Corrientes)

En este anexo se presenta la Recaudación de Ingresos Corrientes (En Nuevos Soles Corrientes), por la Fuente de Financiamiento de Recursos Directamente Recaudados, desde Enero del año 1993 hasta Junio del año 2003.

RECAUDACION DE INGRESOS CORRIENTES

Pliego : 514 Universidad Nacional de Ingeniería

Fuente de Financiamiento : Recursos Directamente Recaudados

AÑOS MESES	AÑO (En Nuevos Soles Corrientes)										
	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
ENERO	178,462.03	966,566.78	1,355,170.88	1,751,102.22	3,600,157.63	4,453,878.11	7,331,121.86	7,686,788.44	7,051,734.11	8,857,640.55	5,968,965.24
FEBRERO	657,551.06	1,005,462.62	957,952.11	2,392,868.90	2,265,526.50	2,491,784.25	5,475,532.39	4,198,082.94	4,755,101.35	4,631,931.19	4,178,032.28
MARZO	151,800.97	930,083.43	1,170,396.23	2,112,311.33	2,366,423.63	3,896,316.45	8,224,880.41	5,912,402.06	6,966,086.51	6,099,205.10	5,886,459.86
ABRIL	1,574,472.07	1,025,566.45	664,898.82	1,525,077.72	1,795,642.56	3,054,427.32	3,308,889.20	4,227,460.44	4,252,898.01	5,481,438.51	4,887,947.93
MAYO	487,225.12	619,712.57	687,363.74	1,639,941.62	2,059,142.20	4,070,142.18	5,799,143.93	3,707,957.31	5,099,941.60	6,898,092.94	5,200,520.35
JUNIO	472,008.92	959,407.88	1,047,656.00	1,513,187.00	1,970,848.06	4,099,943.96	7,684,079.23	6,388,363.91	6,640,067.20	6,094,443.76	5,299,315.08
JULIO	681,499.04	750,032.61	1,715,366.78	2,357,294.37	2,471,915.72	4,935,391.17	3,541,558.98	3,612,334.27	6,389,518.41	4,387,197.54	
AGOSTO	764,196.84	915,766.08	2,014,668.98	2,215,580.78	2,481,825.16	3,854,869.34	5,062,949.95	9,321,747.56	5,875,123.35	5,214,688.16	
SETIEMBRE	431,993.69	848,500.56	946,637.61	1,805,206.06	2,008,782.20	3,835,729.14	4,287,185.08	6,534,765.18	4,559,840.18	7,993,936.52	
OCTUBRE	817,797.83	672,988.91	1,815,734.26	2,600,926.37	2,447,709.61	7,846,989.50	4,444,562.82	5,622,326.97	6,272,691.36	6,703,593.69	
NOVIEMBRE	464,673.30	875,070.34	1,114,661.79	2,235,474.56	2,221,116.14	3,343,208.24	5,228,213.60	7,065,884.80	4,902,391.93	5,262,619.52	
DICIEMBRE	2,968,776.54	2,625,826.00	4,175,044.06	4,484,037.78	4,483,977.66	5,614,512.84	4,510,379.04	4,937,671.90	4,882,075.07	6,611,399.35	
TOTAL	9,650,457.41	12,194,984.23	17,665,551.26	26,633,008.71	30,173,067.07	51,497,192.50	64,898,496.49	69,215,785.78	67,647,469.08	74,236,186.83	31,421,240.74

Fuente: UNI-Oficina Central de Planificación, Unidad de Presupuesto

ANEXO 4: Recaudación de Ingresos Corrientes de la UNI (En Soles Constantes)

En el anexo se presenta la Recaudación de Ingresos Corrientes (En Soles Constantes), por la Fuente de Financiamiento de Recursos Direct. Recaudados, desde Enero de 1993 a Junio de 2003 (Base Dic.2002=100).

RECAUDACION DE INGRESOS CORRIENTES (En Nuevos Soles constantes)

Pliego : 514 Universidad Nacional de Ingeniería

Fuente de Financiamiento : Recursos Directamente Recaudados

MESES \ AÑOS	AÑO (En Nuevos Soles Constantes) (Base: Dic. 2002 = 100.00)										
	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
ENERO	415,780	1,662,205	2,049,250	2,381,607	4,411,244	5,104,405	7,996,526	8,078,792	7,136,117	9,038,898	5,955,180
FEBRERO	1,488,298	1,698,175	1,432,310	3,205,283	2,773,464	2,820,906	5,953,759	4,391,088	4,800,221	4,728,575	4,148,933
MARZO	329,621	1,535,180	1,726,269	2,790,955	2,860,541	4,353,534	8,888,830	6,150,930	6,996,653	6,193,118	5,780,843
ABRIL	3,273,880	1,667,027	971,128	1,997,650	2,162,206	3,392,123	3,555,034	4,375,627	4,289,418	5,525,543	4,802,691
MAYO	983,292	1,000,166	995,628	2,132,593	2,460,914	4,493,467	6,201,318	3,837,274	5,142,483	6,943,957	5,111,449
JUNIO	935,583	1,530,945	1,505,327	1,958,642	2,330,082	4,502,488	8,202,243	6,606,911	6,699,302	6,148,940	5,233,308
JULIO	1,314,805	1,186,285	2,450,757	3,009,884	2,898,463	5,386,138	3,770,451	3,716,723	6,435,402	4,424,910	
AGOSTO	1,437,917	1,426,548	2,848,807	2,803,029	2,903,499	4,195,836	5,380,966	9,546,444	5,935,297	5,254,210	
SETIEMBRE	799,878	1,314,990	1,333,357	2,276,467	2,343,230	4,197,583	4,535,608	6,655,276	4,603,652	8,016,561	
OCTUBRE	1,491,735	1,040,002	2,544,544	3,255,999	2,850,915	8,616,115	4,707,792	5,712,720	6,330,593	6,674,644	
NOVIEMBRE	834,224	1,335,980	1,542,891	2,785,468	2,584,988	3,669,775	5,522,538	7,174,905	4,972,148	5,260,904	
DICIEMBRE	5,199,198	3,985,474	5,748,867	5,520,689	5,185,476	6,124,966	4,743,680	5,006,134	4,956,078	6,611,399	
TOTAL	18,504,211	19,382,977	25,149,135	34,118,266	35,765,022	56,857,336	69,458,745	71,252,824	68,297,364	74,821,659	31,032,404

ANEXO 5: Recaudación de Ingresos Corrientes de las Principales Dependencias de la UNI

En este anexo se describe en tabla y en gráfico la Recaudación de Ingresos Corrientes (En Nuevos Soles Constantes), por la Fuente de Financiamiento de Recursos Directamente Recaudados, de las **principales Dependencias** de la UNI **generadoras de Ingresos Corrientes** (Base: Año 2002=100)

Tabla A.5.1: Principales Dependencias Generadoras de Ingresos Corrientes.

DEPENDENCIA	AÑO	1998	1999	2000	2001	2002	2003 (*)
O.C. Admisión		4,740,678	6,299,228	4,876,652	4,552,531	4,439,325	4,238,842
O.C. Infraestructura		17,928,253	21,261,727	24,130,045	18,963,622	26,391,839	10,920,976
C. E. P. S.		2,658,727	2,808,374	2,645,602	2,530,864	1,523,114	1,447,081
CEPREUNI		4,332,251	6,617,859	5,614,381	5,980,025	5,646,611	5,853,375
CISMID		1,533,332	1,311,453	1,026,727	690,864	792,905	1,718,245
ICI		384,213	545,606	567,849	625,717	558,063	601,668
L. N. Hidráulica		132,972	135,819	458,694	569,766	361,907	98,325

(*) De Enero a Junio (Recaudación de Ingresos Corrientes), De Julio a Diciembre (Proyección)

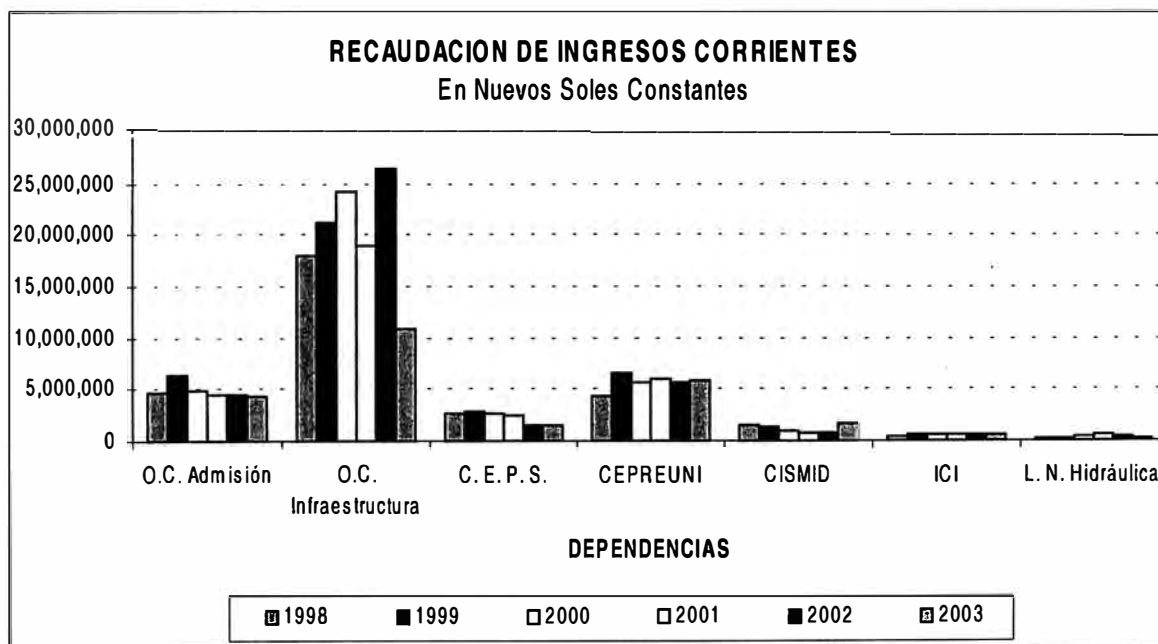


Figura A.5.1: Principales Dependencias generadoras de Ingresos Corrientes.

ANEXO 6: Recaudación de Ingresos Corrientes de las Facultades de la UNI

En este anexo se describe en tabla y en gráfico la Recaudación de Ingresos Corrientes (En Nuevos Soles Constantes), por la Fuente de Financiamiento de Recursos Directamente Recaudados, de las **Facultades** de la UNI generadoras de **Ingresos Corrientes** (Base: Año 2002=100).

Tabla A.6.1: Facultades Generadoras de Ingresos Corrientes.

FACULTADES	AÑO	1998	1999	2000	2001	2002	2003 (*)
FAUA		727,106	1,015,276	847,476	512,785	1,271,696	751,251
FC		381,444	402,206	446,758	770,964	639,969	548,690
FIA		636,999	690,755	870,560	1,039,340	1,044,681	878,194
FIC		2,591,206	3,325,522	2,896,143	2,535,870	4,768,322	3,982,451
FIECS		458,625	456,231	648,972	991,685	1,207,116	1,216,533
FIEE		789,506	847,041	989,659	2,014,105	1,607,055	1,430,552
FIGMM		2,338,897	2,059,079	2,411,929	2,105,583	1,239,180	1,031,788
FIIS		4,368,799	6,114,831	6,909,409	7,054,462	6,614,420	5,901,228
FIM		1,477,103	1,766,566	1,824,026	2,715,133	2,814,652	3,396,635
FIP		265,408	484,192	399,328	735,948	700,258	1,300,613
FIQT		707,152	805,900	707,751	760,006	797,960	786,375

(*) De Enero a Junio (Recaudación de Ingresos Corrientes), De Julio a Diciembre (Proyección)

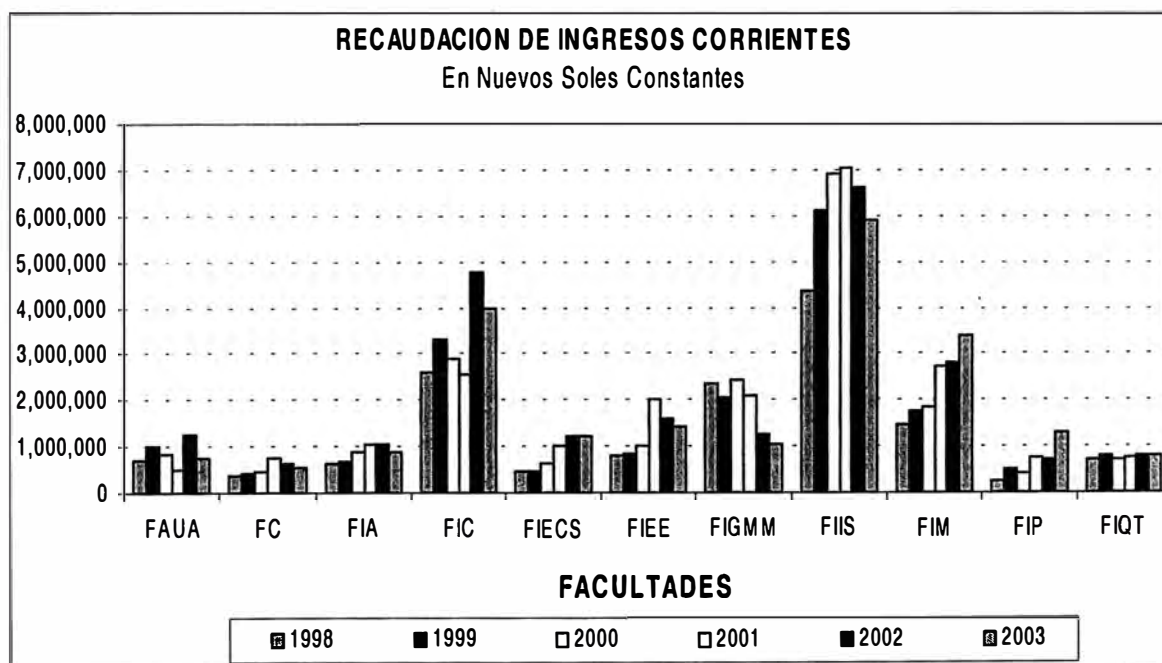


Figura A.6.1: Facultades Generadoras de Ingresos Corrientes.

BIBLIOGRAFÍA

1. **Boletín Mensual:** Indicadores de Precios de la Economía. Instituto Nacional de Estadística e Informática, Agosto 2003, N° 08-2003-INEI
2. **Cornejo, Gabriel:** "Proyección de pasajeros para Lan Chile en la ruta Santiago Isla de Pascua a través de un modelo ARIMA". (disponible en www.monografias.com, documento de trabajo), año 2001.
3. **Ferrán Aranaz, Magdalena:** SPSS para WINDOWS: Programación y Análisis Estadístico. McGraw-Hill, Madrid, año 1996.
4. **González Chávez, S.:** "Modelos Estocásticos ARIMA para Predicción de Variables Energéticas". "TECNIA", Vol 9 N° 1, pp. 81-87, año 1999.
5. **Kendall, Maurice G. y Buckland, William R.:** Diccionario de Estadística. Ediciones Pirámide, S.A., Madrid, año 1976.

6. **Manjares, Mario y Ligia Guevara, Martha:** "Pronóstico de Producción de Aceite mediante Metodologías de Series de Tiempo", CENIPALMA, año 2002.
7. **Parzen, Emanuel:** Procesos Estocásticos. Paraninfo, Madrid, año 1972
8. **Pérez López, César:** Técnicas Estadísticas con SPSS. Prentice Hall, Madrid, año 2001.
9. **Pindyck, Robert S.:** Modelos y Pronósticos, Econometría. McGraw-Hill, México, año 2001.
10. **Presupuesto Del Sector Publico:** Compendio De Normas Presupuestales - Dirección Nacional De Presupuesto Publico - Ministerio De Economía Y Finanzas, año 2003.
11. **Sota Nadal, Javier:** Memoria 1989-1994: la universidad en los años difíciles, año 1994.
12. **Sota Nadal, Javier:** Memoria 1995-1998: Universidad Nacional de Ingeniería, año 1998.
12. **Uriel, Ezequiel y Peiró, Amado:** Introducción al Análisis de Series Temporales. Editorial AC, Madrid, año 2000.

GLOSARIO

Tasas: Son los recursos financieros cuya obligación tiene como hecho generador la prestación efectiva de un servicio individualizado en el contribuyente.

Venta de Bienes: Son los recursos financieros provenientes de la venta de bienes que el Organismo Público produce, independientemente de su función de administración, procurando la cobertura individualizada de sus costos.

Prestación de Servicios: Son recursos financieros provenientes de la prestación de servicios que el Organismo Público brinda, independientemente de su función de administración, procurando la cobertura individualizada de sus costos.

Rentas de la Propiedad: Son los recursos financieros percibidos por concepto de renta de activos reales y financieros:

De la Propiedad Real: Son aquellos que se obtienen por la explotación o concesión de bienes públicos.

De la Propiedad Financiera: Son aquellos que se obtienen por la tenencia de títulos, valores e intereses por depósito en el sistema financiero nacional.

Otros: Son rentas de propiedad que se obtienen por conceptos no considerados dentro de las anteriores partidas subgenéricas.