



**ESCUELA SUPERIOR POLITÉCNICA DEL  
LITORAL**

**Instituto de Ciencias Matemáticas**

**Ingeniería en Estadística Informática**

**“Control de Calidad de Datos en Series Temporales.  
Una Aplicación a las Series de las Cuentas  
Nacionales”**

**TESIS DE GRADO**

**Previa a la obtención del Título de:**

**INGENIERO EN ESTADÍSTICA INFORMÁTICA**

**Presentada por:**

**María de los Ángeles Rodríguez Aroca**

**GUAYAQUIL - ECUADOR**

**AÑO**

**Junio del 2004**

## **AGRADECIMIENTO**

A Dios por permitirme realizar este trabajo, a mis padres y hermano por darme todo su apoyo y a mi director Mat. Fernando Sandoya por guiarme y brindarme sus conocimientos.

# DEDICATORIA

A Dios  
A mis padres Walter y Ángela  
A mi hermano Watico

## **TRIBUNAL DE GRADUACIÓN**

---

**MAT. JORGE MEDINA**  
**DIRECTOR DEL ICM**

---

**MAT. FERNANDO SANDOYA**  
**DIRECTOR DE TESIS**

---

**MAT. JOHN RAMÍREZ**  
**VOCAL PRINCIPAL**

---

**ING. JORGE FERNÁNDEZ**  
**VOCAL PRINCIPAL**

## **DECLARACIÓN EXPRESA**

**“La responsabilidad del contenido de esta tesis de grado, me corresponde exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL”**

(Reglamento de graduación de la ESPOL)

---

María de los Ángeles  
Rodríguez Aroca



## RESUMEN

El presente trabajo plantea una forma automática de llevar el Control de Calidad para datos en Series Temporales, utilizando el programa TERROR (TRAMO for errors); el cual identifica los datos atípicos, que se producen por errores de diversa índole, tales como errores de medición, de digitación, de procesamiento, etc., corrige estos datos mejorando la calidad de las series, el nombre del proyecto es: **“Control de Calidad de Datos en Series Temporales. Una Aplicación a las Series de las Cuentas Nacionales”**.

En la primera parte se revisa de forma general el marco teórico de las Series Temporales. Luego, en la segunda parte, se analiza la metodología TRAMO. Seguidamente, en el tercer capítulo, se desarrolla la descripción del control de calidad utilizando la aplicación TERROR del programa TRAMO.

En el cuarto capítulo se realiza la aplicación del control de calidad a las series de las Cuentas Nacionales del Banco Central del Ecuador. Y por último, en la quinta parte se presenta las conclusiones y recomendaciones alcanzadas en la elaboración de este trabajo.

# INDICE GENERAL

	<b>Pag.</b>
RESUMEN	I
ÍNDICE GENERAL	II
ABREVIATURAS	VII
SIMBOLOGÍA	IX
ÍNDICE DE GRÁFICOS	XI
INDICE DE TABLAS	XIII
INTRODUCCIÓN	1

## I. INTRODUCCIÓN A LAS SERIES TEMPORALES

1.1. Serie de Datos.....	4
1.1.1. Series Temporales.....	4
1.1.2. Series Transversales.....	4
1.1.3. Datos de Panel.....	4
1.2. Series Temporales.....	5
1.3. Frecuencia de una Serie Temporal de Datos.....	6
1.4. Medición de las Series de Datos.....	7
1.4.1. Series Medidas en Niveles.....	7
1.4.1.1. Variable de flujo.....	7



1.4.1.2.	Variable acumulada o de stock.....	8
1.4.1.3.	Saldo .....	9
1.4.1.4.	Ratio.....	10
1.4.2.	Serie Medidas con números índices.....	10
1.4.2.	Serie Medidas en tasas de crecimiento.....	10
1.5.	Otras Definiciones Básicas en la Teoría de Series de Tiempo.....	11
1.5.1.	Datos Atípicos.....	11
1.5.2.	Ruido Blanco.....	11
1.5.3.	Operador de Diferencias.....	11
1.5.4.	Paseo Aleatorio.....	12
1.5.5.	Función de Autocovarianza de un Proceso Estocástico (FAC).....	12
1.5.6.	Función de Autocorrelación Simple de un Proceso Estocástico (FAS).....	12
1.5.7.	Función de Autocorrelación Parcial de un Proceso Estocástico (FAP).....	13
1.5.8.	Procesos Estacionarios y Ergódicos.....	13
1.5.8.1.	Estacionariedad.....	13
1.5.8.2.	Procesos Ergódicos.....	15
1.5.9.	Modelos de una Serie Temporal.....	16

1.5.9.1.	Modelos Autorregresivos (AR).....	16
1.5.9.2.	Modelos de medias Móviles( MA).....	17
1.5.9.3.	Modelos Mixtos Autorregresivos-Medias Móviles (ARMA).....	17
1.5.10.	Procesos no Estacionarios.....	18
1.5.10.1.	Modelos Arima.....	19

## II. METODOLOGÍA TRAMO

2.1.	Componentes de una Serie Temporal.....	22
2.1.1.	La componente tendencia.....	22
2.1.2.	La componente tendencia – ciclo.....	23
2.1.2.1.	Fase de Expansión.....	24
2.1.2.2.	Fase de Recesión.....	25
2.1.2.3.	Punto de Retorno.....	25
2.1.2.5.	Recuperación.....	25
2.1.3.	Componente Estacional.....	26
2.1.3.1.	Estacionalidad Climática.....	27
2.1.3.2.	Estacionalidad Institucional.....	27
2.1.3.3.	Estacionalidad Inducida.....	27
2.1.4.	La componente de días laborables.....	28
2.1.5.	La componente Pascua.....	29
2.1.6.	La componente Irregular.....	29

2.1.6.1.	Determinantes de la irregularidad de las series.....	30
2.2.	Descomposición de las Series de Tiempo.....	32
2.2.1.	Modelos de descomposición .....	34
2.2.1.1.	Modelos Aditivo.....	34
2.2.1.2.	Modelos Multiplicativo.....	35
2.3.	Método de Extracción de señales.....	36
2.3.1.	Metodología TRAMO.....	36

### **III. CONTROL DE CALIDAD DE DATOS CON TRAMO**

3.1.	TERROR (TRAMO for errors): una aplicación del programa TRAMO.....	43
3.2.	Ejecución del comando terror (TRAMO for errors).....	47

### **IV. APLICACIÓN AL CONTROL DE CALIDAD DE DATOS DE SERIES TEMPORALES**

4.1.	Serie de datos a utilizar.....	50
4.2.	Análisis de Resultados.....	54
4.2.1.	Resultados de las Series de Importaciones.....	54
4.2.1.1	Diferencias tomadas (m).....	55
4.2.1.2	Datos Atípicos (m).....	56
4.2.2.	Resultados de las Series de Exportaciones.....	57

4.2.2.1	Diferencias tomadas (x).....	58
4.2.2.2	Datos Atípicos (x).....	59
4.2.3.	Resultados de las Series de Manufactura.....	60
4.2.3.1	Diferencias tomadas (pr).....	61
4.2.3.2	Datos Atípicos (pr).....	61
4.2.4.	Resultados de las Series Mensuales.....	62
4.2.4.1	Diferencias tomadas (ip).....	64
4.2.4.2	Datos Atípicos (ip).....	64

## **V. CONCLUSIONES**

ANEXOS

BIBLIOGRAFÍA

## ABREVIATURAS

FAC	Función de Autocovarianza
FAS	Función de Autocorrelación Simple
FAP	Función de Autocorrelación Parcial
AR	Modelo Autorregresivo
MA	Modelo de Medias Móviles
ARMA	Modelo autorregresivo y de medias móviles
ARIMA	Modelo autorregresivo y de medias móviles integrado
$c_t$	Componente tendencia-ciclo (el "Aciclo")
$s_t$	Componente estacional ("estacionalidad")
$r_t$	Componente de días laborales
$h_t$	Componente de pascua
$i_t$	Componente irregular (el "irregular")
$p$	Periodo de una serie
TRAMO	Regresión de Series Temporales con ruidos ARIMA, observaciones perdidas y datos atípicos.
TERROR	TRAMO para errores
ECM	Error Cuadrático Medio

AO	Aditivos
LS	Cambios de nivel
TC	Desplazamientos transitorios
m	Series de Importaciones
x	Series de Exportaciones
pr	Series de Manufactura
ip	Series mensuales
BCE	Banco Central del Ecuador

## SIMBOLOGÍA

$Y_t, W_t$	Series de Datos
$Z_t, Y_t$	Vectores de Series de Datos
$p, q$	Orden de las Series
$L$	Operador de Retardo
$\Delta$	Operador de Diferencias
$\varepsilon_t$	Ruido Blanco
$E[x]$	Esperanza
$VAR[x]$	Varianza
$COV[x_t, x_{t-j}]$	Covarianza
$\rho_k(t)$	Función de autocorrelación simple
$\gamma_k$	Función de autocorrelación parcial
$F(Y_{t1}, Y_{t2}, \dots, Y_{tn},)$	Función de distribución conjunta
$m$	Periodos de tiempo
$\mu_t$	Media
$\phi(L), \theta(L), B(L)$	Función de operador de retardos
$Z_t^{des}$	Serie desestacionalizada

$\ln$	Logaritmo natural
$\beta$	Vector de los coeficientes de la regresión
$\phi(B), \delta(B), \theta(B)$	Polinomios finitos en B
$\delta(B)$	Polinomio de raíces unitarias
$\phi(B)$	Polinomio de raíces – parte autorregresiva
$t_1, t_2$	Parámetros de sensibilidad
$\sigma^2$	Varianza de una función
$\lim_{k \rightarrow \infty} \rho_k = 0$	Condición necesaria de ergodicidad
$\bar{y}$	Promedio de la serie
$E[Y_t]$	Esperanza de la serie
$\mu_t$	Media de una función



## INDICE DE GRÁFICOS

Gráfico 1.1	Serie Porcentual del desempleo con relación a la población económicamente activa.....	8
Gráfico 1.2	Serie Porcentual del desempleo .....	9
Gráfico 1.3	Serie del saldo de la balanza comercial Ecuador-Perú.....	10
Gráfico 2.1	Serie del índice Inflacionario del Ecuador.....	26
Gráfico 2.2	Serie temporal del índice I-p11 de las Cuentas Nacionales.....	28
Gráfico 2.3	Serie temporal del índice I-x03 de las Cuentas Nacionales.....	30
Gráfico 4.1	Serie Temporal del índice I-m16 de las Cuentas Nacionales.....	55
Gráfico 4.2	Serie Temporal del índice I-x18 de las Cuentas Nacionales.....	57
Gráfico 4.3	Serie Temporal del índice I-x20 de las Cuentas Nacionales.....	58
Gráfico 4.4	Serie Temporal del índice I-pr16 de las Cuentas Nacionales.....	60

Gráfico 4.5	Serie Temporal del índice Ip-p01 de las Cuentas Nacionales.....	63
Gráfico 4.6	Serie Temporal del índice Ip-p02 de las Cuenta Nacionales.....	63

## INDICE DE TABLAS

Tabla I	Serie del índice de desempleo.....	7
Tabla II	Serie porcentual de desempleados (1995–2000).....	8
Tabla III	Serie del saldo de la balanza comercial Ecuador– Perú.....	9
Tabla IV	Grados de Sensibilidad.....	51
Tabla V	Diferencias tomadas en las series importaciones.....	56
Tabla VI	Datos atípicos en las series importaciones.....	56
Tabla VII	Diferencias tomadas en las series exportaciones.....	56
Tabla VIII	Datos atípicos en las series exportaciones.....	59
Tabla IX	Diferencias tomadas en las series de manufactura....	61
Tabla X	Datos atípicos en las series de manufactura.....	62
Tabla XI	Diferencias tomadas en las series mensuales.....	64
Tabla XII	Datos atípicos en las series mensuales.....	65
Tabla XIII	Resultados obtenidos.....	67

# INTRODUCCIÓN

Dado que cada mes y cada trimestre se reciben miles de datos nuevos desde diferentes entidades para su agregación en las Cuentas Nacionales, es necesario realizar un esfuerzo para detectar posibles errores en los mismos, las causas de estos errores son variadas; pueden ir desde un error en la colocación de una coma, hasta datos mal calculados.

Para corregir esto se han desarrollado últimamente métodos automáticos que permiten procesar gran cantidad de datos en poco tiempo, utilizando la información disponible en forma eficiente; en particular, la experiencia ha demostrado la eficiencia de una derivación de la metodología TRAMO (Time Series Regression with Arima Noise Missing Observations and Outliers) denominada TERROR (TRAMO for Errors), cuya aplicación es el objetivo de la presente TESIS.

Se usarán las series de las Cuentas Nacionales del Banco Central para la aplicación del trabajo, para ello es necesario introducir brevemente cual es la función de dicha entidad en nuestro país.

El Banco Central del Ecuador (BCE) es la Institución encargada de la elaboración de las estadísticas de síntesis de los principales sectores de la economía, esto es, del sector monetario y financiero, del sector externo, del sector real y del sector fiscal. Adicionalmente también es la entidad encargada de procesar y difundir algunas estadísticas de base, que den cuenta de la marcha económica en el corto plazo de la economía ecuatoriana. En ese contexto y con el objetivo de tener indicadores económicos que sirvan de base para la recomendación y toma de decisiones de política económica, el BCE elabora y difunde las cuentas nacionales anuales y trimestrales; los indicadores monetarios y financieros; la balanza de pagos; las cuentas fiscales del resto del sector público no financiero; las encuestas de coyuntura trimestrales al sector agropecuario, industrial y de la construcción; las encuestas mensuales de opinión empresarial dirigidas a 8 sectores productivos; las estadísticas de comercio exterior; la previsión económica; el índice de actividad económica coyuntural.

De todas estas, nuestro interés se centra en las Cuentas Nacionales, ya que se aplicará un control de calidad a las mismas, valiéndonos del método descrito para detectar posibles errores y corregirlos. Con la información correcta y depurada se podrán tomar decisiones más acertadas en beneficio del país.

El trabajo se ha dividido de la siguiente forma. El primer capítulo tratará del marco teórico de la tesis. Posteriormente, la metodología del programa a utilizar: TRAMO. En tercer lugar se explicará exclusivamente el control de calidad de datos con TRAMO, es decir el comando TERROR. Luego se presenta la aplicación del control de calidad a las series de las Cuentas Nacionales. Finalmente se expondrán las conclusiones y recomendaciones del trabajo.

# CAPITULO 1

## 1. INTRODUCCIÓN A LAS SERIES TEMPORALES

### 1.1. Series de Datos

Se entiende por series de datos cualquier conjunto de datos relativos a una misma variable. Entre las muchas clasificaciones referentes a las series de datos, nos interesan especialmente las siguientes:

**1.1.1. Series Temporales.-** Cuyos datos se refieren a los distintos períodos de un rango de tiempo.

**1.1.2. Series Transversales.-** Cuyos datos hacen referencia a distintos individuos u objetos para un mismo momento del tiempo.

**1.1.3. Datos de panel.-** Cuyos datos cubren, al mismo tiempo un espacio temporal y transversal.

## 1.2. Series Temporales

Una serie temporal es un conjunto de observaciones ordenadas en el tiempo o, también, la evolución de un fenómeno o variable a lo largo de él, ésta variable puede ser: económica, física o social.

El objetivo del análisis de una serie temporal, de la que se dispone de datos en períodos regulares de tiempo, es el conocimiento de su patrón de comportamiento para prever su evolución futura, siempre bajo el supuesto de que las condiciones no cambiarán respecto a las condiciones actuales y pasadas.

Si al conocer la evolución de la serie en el pasado se pudiese predecir su comportamiento futuro sin ningún tipo de error, estaríamos frente a un fenómeno determinista cuyo estudio no tendría ningún interés o complicación especial.

En general, las series de interés llevan asociados fenómenos aleatorios, de forma que el estudio de su comportamiento pasado sólo permite acercarse a la estructura o modelo probabilístico para la predicción del futuro.



Estos modelos se denominan también procesos estocásticos. Así, un proceso estocástico es una sucesión de variables aleatorias  $\{Y_t\}$ , con  $t = 1, 2, \dots, n$ , que evolucionan en el tiempo (representado éste por el subíndice  $t$ ).

Cuando se dispone de  $n$  datos de un proceso estocástico, se está frente a  $n$  muestras, de tamaño unitario, extraídas de la población (variable aleatoria), correspondientes al tiempo en que se realizó la medición, y esto es lo que constituye la serie temporal o cronológica.

### **1.3. Frecuencia de una Serie Temporal de Datos**

Se entiende por frecuencia de una serie temporal, el intervalo de tiempo que separa dos de sus datos.

- Las frecuencias más habituales en las series económicas de las Cuentas Nacionales son la anual (un dato por año), la trimestral (4 datos por año) y la mensual (12 datos por año).
- Cuanto menor es el tiempo transcurrido entre dos datos, decimos que mayor es la frecuencia de la serie o que su nivel de desagregación temporal es mayor.
- La frecuencia de una serie es trascendental para abordar su análisis, condicionando en buena medida la aplicación de técnicas de medición.

#### 1.4. Medición de las Series de Datos

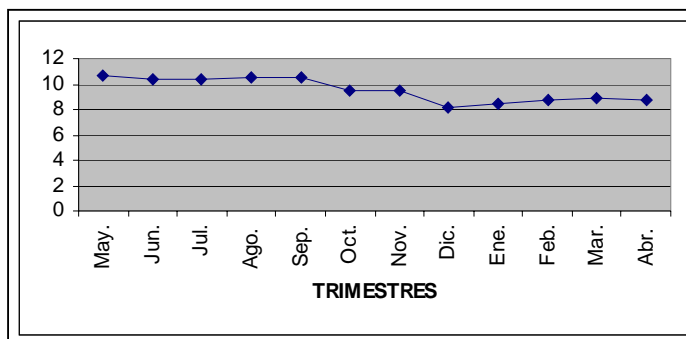
Las series de datos pueden ser medidas de diversas formas según sea el interés del investigador.

**1.4.1. Series medidas en niveles.-** En términos generales decimos que una serie está medida en niveles cuando queda explícita la unidad de medida. A su vez, pueden distinguirse algunos tipos genéricos de medición “en niveles”:

**1.4.1.1. Variable de Flujo.-** Serie que representa en cada observación el valor adquirido por una determinada variable en un determinado período de tiempo, sin relación alguna con otros períodos de tiempo previos o con otras variables.

**Tabla I**

Por ejemplo:											
La serie que se presenta a continuación muestra el índice de desempleo con relación a la P.E.A.											
Mayo 2001 – Abril 2002											
May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.	Ene.	Feb.	Mar.	Abr.
10,6	10,4	10,4	10,5	10,5	9,5	9,5	8,1	8,4	8,8	8,9	8,7
Elaboración propia											
Fuente: INEC (Instituto Nacional de Estadísticas y Censos)											

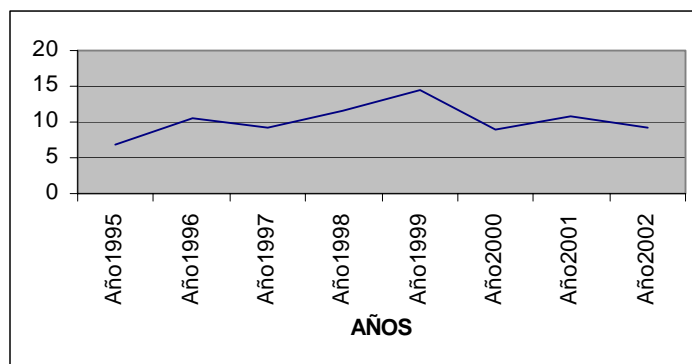


**Gráfico 1.1** Serie porcentual de desempleo (mayo 2001 – abril 2002) con relación a la Población Económicamente Activa

**1.4.1.2. Variable Acumulada o “de stock”.-** Serie cuyas observaciones recogen, de forma acumulada, los valores adquiridos por la variable a lo largo de un determinado período de tiempo y todos los anteriores al menos dentro de un marco temporal más amplio.

**Tabla II**

Por ejemplo:							
La serie que se muestra a continuación recoge el porcentaje De desempleados registrados cada año empezando desde 1995 hasta el 2002							
1995	1996	1997	1998	1999	2000	2001	2002
6,9	10,4	9,2	11,5	14,4	9,0	10,9	9,2
Elaboración propia							
Fuente: INEC (Instituto Nacional de Estadísticas y Censos)							

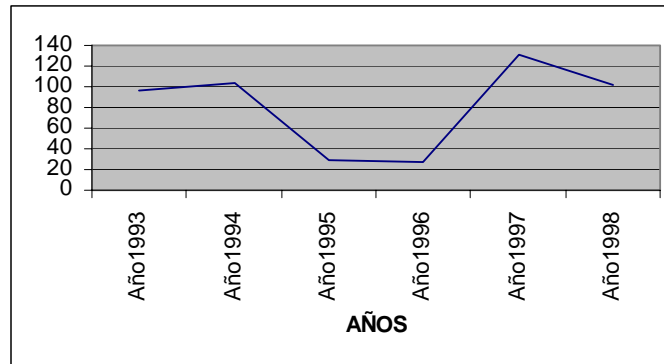


**Gráfico 1.2** Serie porcentual de desempleo

**1.4.1.3. Saldo.-** Variable resultante de la agregación de un conjunto de variables de contenido diverso con cierta utilidad conceptual.

**Tabla III**

Por ejemplo:					
La siguiente serie representa el saldo de la balanza comercial de Ecuador con Perú en miles de dólares desde el año 1993 hasta el año 1998					
1993	1994	1995	1996	1997	1998
96,191	104,389	29,014	26,454	131,154	102,047
Elaboración: Proyecto SICA - BIRF / MAG - Ecuador (www.sica.gov.ec) Fuente: Banco Central del Ecuador					



**Gráfico 1.3** Serie de la balanza comercial de Ecuador con Perú

**1.4.1.4. Ratio.-** Una ratio es una variable resultante de la razón comparada de dos magnitudes medidas en la misma unidad. El porcentaje de exportaciones ecuatorianas a los Estados Unidos, la tasa activa que cobran los bancos, son ejemplos de magnitudes medidas como ratios.

**1.4.2. Series medidas con Números Índices.-** Un número índice es una forma de medida adimensional que compara la magnitud de una variable en cada período con la magnitud de esa misma variable en un momento predeterminado del tiempo que denominamos periodo base.

**1.4.3. Series medidas en Tasas de Crecimiento.-** Se dice que una serie está medida en tasas de crecimiento cuando cada valor expresa

el incremento del valor de la serie respecto a un periodo anterior determinado.

### 1.5. Otras Definiciones Básicas en la Teoría de Series de Tiempo

Es necesario explicar brevemente las siguientes definiciones relacionadas con la teoría de series de tiempo.

**1.5.1 Datos Atípicos.-** Se llamará dato atípico a aquella observación que tiene un comportamiento muy diferente respecto de los demás datos, frente al análisis que se desea realizar sobre las observaciones experimentales.

**1.5.2. Ruido Blanco.-** Es una sucesión de variables aleatorias con esperanza cero, igual varianza, e independientes en el tiempo. Se denota con  $\{\varepsilon_t\}$

Es decir,

$$E[\varepsilon_t] = 0$$

$$\text{VAR}[\varepsilon_t] = \sigma^2$$

$$\text{COV}[\varepsilon_t, \varepsilon_{t-j}] = 0, \quad \forall j \neq 0$$

**1.5.3. Operador de Diferencias.-** Para una serie de tiempo  $\{y_t\}$  se denomina operador de diferencias, representado con  $\Delta$  o  $B$  al operador que actúa de la siguiente manera:

$$\Delta Y_t = BY_t = Y_t - Y_{t-1}$$

**1.5.4. Paseo Aleatorio.-** Es un proceso estocástico  $\{y_t\}$  cuya primera diferencia es un ruido blanco.

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_t \Rightarrow Y_t = Y_{t-1} + \varepsilon_t$$

donde  $\varepsilon_t$  es ruido blanco

Así entonces:  $E[y_t] = 0$

$$\text{VAR}[y_t] = t\sigma^2$$

$$\text{COV}[y_t, y_{t-j}] = (t-j) \sigma^2$$

**1.5.5. Función de Autocovarianza de un Proceso Estocástico (FAC).-** Es una función que para cada instante  $t$  y cada número entero  $k$  toma un valor, denotado  $\gamma_k(t)$ .

$$\gamma_k(t) = \text{COV}[y_t, y_{t-k}]$$

**1.5.6. Función de Autocorrelación Simple de un Proceso Estocástico (FAS).-** Es una función que para cada instante  $t$  y cada número entero  $k$  toma un valor, denotado  $\rho_k(t)$ .

$$\rho_k(t) = \frac{\text{COV}[y_t, y_{t-k}]}{\sqrt{\text{VAR}(y_t)}\sqrt{\text{VAR}(y_{t-k})}} = \frac{\gamma_k(t)}{\sqrt{\text{VAR}(y_t)}\sqrt{\text{VAR}(y_{t-k})}}$$

**1.5.7. Función de Autocorrelación Parcial de un Proceso Estocástico (FAP).**- Es una función que para cada instante  $t$  y cada número entero  $k$  toma un valor igual a la correlación  $y_t$  e  $y_{t-k}$ , ajustado por el efecto de los retardos intermedios  $y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$ .

$$\gamma_k \approx \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

### 1.5.8. Procesos Estacionarios y Ergódicos

**1.5.8.1. Estacionariedad.**- Para definir la estacionariedad de un proceso se pueden utilizar las funciones de distribución o alternativamente los momentos. Se dice que un proceso estocástico es estacionario en sentido estricto cuando al realizar un mismo desplazamiento en el tiempo de todas las variables de cualquier distribución conjunta finita, resulta que esta distribución no varía.

Considerando la función de distribución conjunta

$$F(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n})$$



Desplazándola  $m$  periodos de tiempo

$$F(Y_{t_1+m}, Y_{t_2+m}, \dots, Y_{t_n+m})$$

Se debe de verificar que:

$$F(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}) = F(Y_{t_1+m}, Y_{t_2+m}, \dots, Y_{t_n+m})$$

El análisis de estacionariedad es más complejo a partir de funciones de distribución que si se efectúa a partir de los momentos. En contrapartida, el concepto de estacionariedad es más limitado.

- Un proceso es estacionario de primer orden o en media si se verifica que:

$$E[Y_t] = \mu_t \quad \forall t$$

Por lo que si un proceso es estacionario en media, la esperanza matemática permanece constante a lo largo del tiempo.

- Un proceso es estacionario de segundo orden (o en sentido amplio) si se verifica:

$$\text{var } Y_t = E(Y_t - \mu)^2 = \sigma^2 < \infty \quad \forall t$$

$$E(Y_{t+k} - \mu_t)(Y_t - \mu) = \gamma_k \quad \forall t$$

Al definir un proceso estacionario en sentido amplio, se tiene en cuenta implícitamente que el proceso es también estacionario en media. Además si un proceso es estacionario en sentido estricto también será estacionario en sentido amplio, lo contrario no necesariamente ocurre. Igualmente se verifica que si un proceso es estacionario en sentido amplio y además normal, el proceso será estacionario en sentido estricto.

**1.5.8.2. Procesos Ergódicos.-** Cuando los valores de la serie temporal alejados en el tiempo están muy correlacionados, es decir, cuando  $\rho_k$  se mantiene en unas cotas elevadas para un  $k$  grande, sucederá que al aumentar el tamaño de la muestra se añade poca información nueva. La consecuencia de este hecho en el plano estadístico será que los estimadores no serán consistentes.

Una condición necesaria, aunque no suficiente, de la ergodicidad es:

$$\lim_{k \rightarrow \infty} \rho_k = 0$$

Cuando se verifica la ergodicidad se pueden obtener estimadores consistentes que caractericen al proceso en cuestión.

En lo que sigue se considerará que las series de tiempo analizadas satisfacen la propiedad de ergodicidad.

**1.5.9. Modelos de una Serie Temporal.-** A continuación se describe de manera general los modelos más comunes que se pueden encontrar en una serie temporal

**1.5.9.1. Modelos Autorregresivos (AR).-** Un modelo autorregresivo de orden  $p$ , o abreviadamente AR( $p$ ), se define de la siguiente manera:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Donde  $\varepsilon_t$  es una variable “ruido blanco” y si se utiliza el operador de retardos:

$$\phi(L) = 1 - \phi_1 L + \phi_2 L^2 + \dots + \phi_p L^p$$

El modelo autorregresivo se puede expresar de manera compacta:

$$\phi(L)Y_t = \varepsilon_t$$

**1.5.9.2. Modelos de Medias Móviles (MA).**- Un modelo de medias móviles de orden  $q$ , o abreviadamente MA( $q$ ), se define de la siguiente forma:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}$$

Utilizando el operador de retardos,

$$\theta(L) = 1 - \theta_1L - \theta_2L^2 - \dots - \theta_qL^q$$

Así, el modelo de medias móviles se puede expresar de forma compacta:

$$Y_t = \theta(L)\varepsilon_t$$

**1.5.9.3. Modelos Mixtos Autorregresivos-Medias Móviles (ARMA).**- Un modelo ARMA( $p,q$ ) está definido de la siguiente forma:

$$Y_t = \phi_1Y_{t-1} + \dots + \phi_pY_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}$$

Utilizando los operadores polinomiales de retardo, el modelo queda expresado en forma compacta del siguiente modo:

$$\phi(L)Y_t = \theta(L)\varepsilon_t$$

**1.5.10. Procesos No Estacionarios.-** En múltiples ocasiones las series económicas no cumplen con las restricciones de estacionariedad y/o invertibilidad. Por lo tanto para obtener un tratamiento de estas series basado en la teoría de los procesos estocásticos es necesario ampliar el campo para incluir los procesos no estacionarios.

El objeto es considerar cierto tipo de procesos no estacionarios que sean adecuados para describir el comportamiento de las series económicas, y al mismo tiempo, sean fácilmente transformables en procesos estacionarios con motivo de utilizar las ventajas que ofrecen estos últimos.

Si se tiene un proceso no estacionario  $Y_t$ , muchas veces, al tomar diferencias de orden  $k$  se puede llegar a un proceso  $w_t$  que es estacionario, entonces se realizan los análisis pertinentes en base a la nueva serie obtenida. Luego integrando el proceso  $w_t$  se regresa al proceso original  $Y_t$ , con el cual se emiten las conclusiones finales.

En general, se denominan modelos no estacionarios homogéneos a todos aquellos modelos que se pueden transformar en estacionarios mediante la toma de diferencias de un determinado orden, o dicho de otra forma, modelos integrados son aquellos que se pueden obtener mediante suma o integración de un proceso estacionario.

**1.5.10.1. Modelos Arima.-** A un proceso integrado  $Y_t$  se le denomina ARIMA(p,d,q) si tomando diferencias de orden d se obtiene un proceso estacionario  $w_t$  del tipo ARMA(p,q). La I central del término ARIMA indica integrado.

Así pues se tiene que:

$$w_t = \Delta^d Y_t = (1-L)^d Y_t$$

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) w_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t$$

En forma más compacta con los operadores de retardo  $\theta$  y  $\phi$ , el modelo ARIMA(p,d,q) se puede representar así:

$$\phi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t$$

Los modelos ARIMA constituyen una clase particular de procesos no estacionarios. Sin embargo, en muchos casos son

suficientes para representar el comportamiento de las series económicas.

Además debe considerarse que cuando una serie económica se observa a lo largo de un periodo dilatado de tiempo ocurre con frecuencia que también la varianza viene afectada por una tendencia, y esta tendencia no desaparece al tomar diferencias. Cuando se presenta esta situación la transformación adecuada puede consistir en tomar logaritmos. Los estadísticos Box y Cox definieron una transformación instantánea de carácter general,

$$Y_t^{(\lambda)} = \begin{cases} \frac{(Y_t^\lambda - 1)}{\lambda} & \lambda \neq 0 \\ \ln Y_t & \lambda = 0 \end{cases}$$

Cuando el parámetro es  $\lambda = 1$ , la transformación consiste prácticamente en tomar valores originales.

# CAPITULO 2

## 2. METODOLOGÍA TRAMO

Las Series Temporales de las Cuentas Nacionales son afectadas por algunos agentes o factores externos que perturban el normal comportamiento de éstas. Estos factores pueden ser: avances o cambios tecnológicos; cambios o nuevas medidas políticas o económicas; huelgas; expedición de nuevas leyes, incremento, disminución o creación de nuevos impuestos, cambios metodológicos en la medición de las estadísticas, etc. Por ello es necesario analizar y corregir estas variaciones, es decir realizar un control de calidad de la serie, para de esta forma trabajar con datos depurados y presentar resultados con los cuales se puede hacer un análisis económico más fino. Para esto existen varios métodos para corregir estas desviaciones, entre ellos se encuentra la metodología TRAMO la constituye una herramienta poderosa, flexible y de fácil uso en el análisis de series de tiempo.



## 2.1. Las Componentes de una Serie Temporal

Partimos de la suposición básica que la serie original  $z_t$  contiene cinco componentes básicas:

- $c_t$ : la componente tendencia-ciclo (el "Aciclo")
- $s_t$ : la componente estacional ("estacionalidad")
- $r_t$ : la componente de días laborales
- $h_t$ : la componente de feriados móviles (o feriados)
- $i_t$ : la componente irregular (el "irregular")

Estas componentes no son observables y por lo tanto no se pueden prever.

La serie desestacionalizada  $ci_t$  está formada por dos de las componentes básicas: la tendencia-ciclo  $c_t$  y la irregular  $i_t$ , es decir, es la serie original a la que se le ha quitado las componentes  $s_t$ ,  $r_t$  y  $h_t$ .

A continuación se describe cada una de las componentes de las series temporales.

**2.1.1. La componente tendencia.-** Representa una variación regular y gradual durante un largo período de años; es una

evolución a largo plazo de la variable sobre varias décadas. Está asociada con las bajas frecuencias. La frecuencia de una serie temporal es igual a:  $\frac{1}{p}$ , donde p es el periodo. Por frecuencia de una serie temporal se entiende el lapso que separa dos de sus datos. Cuanto menor es el tiempo transcurrido entre dos datos, se dice que mayor es la frecuencia de la serie.

La tendencia a largo plazo indica la dirección del movimiento de la serie sobre un período de muchos años, esta dirección puede ser creciente o decreciente.

**2.1.2. La Componente Tendencia-Ciclo.-** El ciclo es un movimiento regular; es decir sin discontinuidades ni oscilaciones erráticas, de la serie alrededor de la tendencia a largo plazo.

Puesto que generalmente las series disponibles son relativamente cortas, a menudo no es posible estimar tendencias que por definición son fenómenos a largo plazo. Por esta razón, la mayor parte de los métodos de descomposición

de las series cronológicas originan la tendencia-ciclo, que es una combinación de la tendencia y el ciclo.

Cuando se expresa tendencia, ciclo o ciclo económico representan siempre la tendencia-ciclo, a menos que se indique lo contrario.

Desde un punto de vista económico se puede decir que la tendencia-ciclo es una secuencia de fluctuaciones regulares de la serie, caracterizada por períodos alternados de expansión y contracción que habitualmente duran entre tres y siete años. Estos períodos se los denomina fases.

La duración y la intensidad de varias fases varían característicamente en cada ocasión. Es necesario recalcar que cada fase dura más de un año.

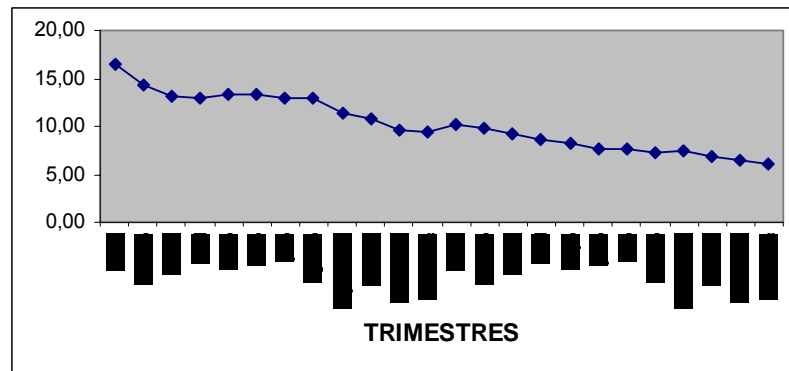
**2.1.2.1. Fase de expansión.-** Es un período de prosperidad o mejoramiento de la situación económica, en el caso de variables como la producción, el empleo; lo opuesto en el caso de variables como el desempleo, las quiebras, etc.

**2.1.2.2. Fase de recesión.-** Es un período de deterioro o crisis de la situación económica.

**2.1.2.3. Punto de retorno.-** Es el pasaje de un período de mejoramiento a un período de deterioro de las condiciones económicas.

**2.1.2.4. Recuperación.-** Es la transición de un período de deterioro a uno de mejoramiento.

La tendencia-ciclo, representa el movimiento fundamental de la serie, por ejemplo, las condiciones económicas dominantes, y no los fenómenos pasajeros como la estacionalidad, los días laborales o los movimientos irregulares y por tanto es esencial para el análisis económico de coyuntura.



**GRÁFICO 2.1** Índice inflacionario del Ecuador desde Ene. 2002 Dic. 2003  
Fuente: Banco Central del Ecuador ([www.bce.fin.ec](http://www.bce.fin.ec))

La representación del índice inflacionario del Ecuador ha sido claramente decreciente a raíz de enero del 2002 hasta diciembre del 2003, como se muestra en el gráfico, es decir muestra una tendencia decreciente, con leves alzas, lo que demuestra estabilidad a lo largo de este periodo.

**2.1.3 Componente Estacional.-** Se refiere a fluctuaciones subanuales, es decir mensuales, trimestrales, semestrales, etc. que se repiten con regularidad año tras año. Por convención, la estacionalidad se anula cada año. Como resultado de ello, las series anuales no pueden contener estacionalidad. Existen varias causas que originan estacionalidad:

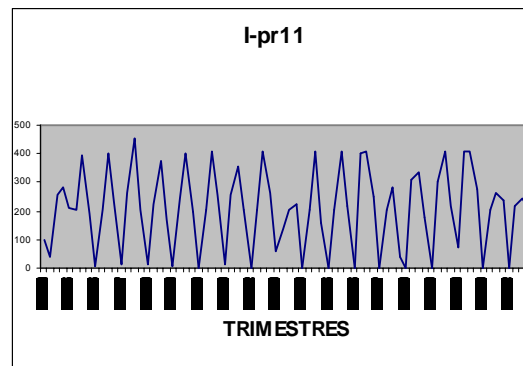
**2.1.3.1. Estacionalidad Climática.-** Atribuible a variaciones climáticas estacionales, por ejemplo: en la agricultura, en la pesca, tiempo de veda, etc.

**2.1.3.2. Estacionalidad institucional.-** Atribuible a convenciones sociales y reglas administrativas, por ejemplo: el efecto de las fiestas navideñas, el inicio del año escolar, etc.

**2.1.3.3. Estacionalidad inducida.-** Atribuible a la estacionalidad en otros sectores, por ejemplo: en la elaboración de alimentos, fabricación de vestuario, etc.

La estacionalidad evoluciona por diferentes motivos; los cambios tecnológicos, se refieren a los avances de la tecnología, como por ejemplo el desarrollo de materiales y técnicas de construcción adaptadas al invierno; el desarrollo biotecnológico de cultivos que resisten mejor a un determinado tipo de clima; los cambios institucionales se refieren a variaciones de las políticas en una determinada organización, como por ejemplo la extensión del año académico a los meses de verano en las universidades, la difusión de las vacaciones a lo largo de todo el año en lugar de verano. También se tienen los cambios de la composición de la serie, como por ejemplo la

sociedad se torna más industrializada y menos dependiente de las industrias primarias (como la pesca y la agricultura), que son esencialmente más estacionales.



**GRÁFICO 2.2** Serie temporal del índice I-pr11 de las Cuentas Nacionales  
Fuente: Banco Central del Ecuador

Observando la serie trimestral del índice de manufactura (I-pr11) se puede notar una marcada estacionalidad, es decir que se repite un patrón más o menos invariable, en momentos similares.

#### **2.1.4. La Componente de días laborables**

**Definición:** Es el resultado de la diferente distribución de los días de la semana que cambia de acuerdo al mes y además está presente cuando algún día de la semana tiene una importancia específica.

**Tasas diarias de Actividad:** Es la proporción de la actividad semanal que ocurre en cada día de la semana.

### 2.1.5. La Componente Pascua

**Definición:** El efecto de feriados de fecha móvil se refiere a las diferentes variaciones debidas a que algunos feriados cambian de fecha dependiendo del año, lo cual origina un desplazamiento en la actividad entre los meses; el proceso más incuestionable es la Pascua (o Semana Santa).

#### **Tipos de efecto Pascua:**

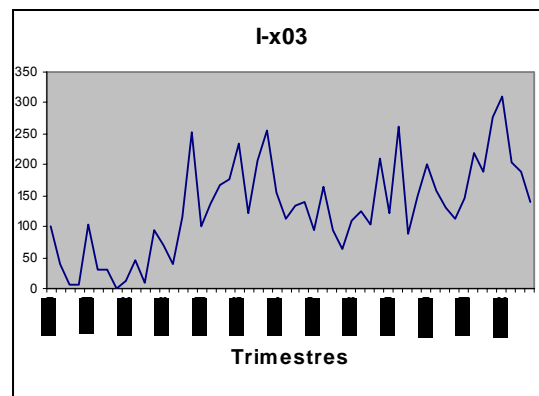
- ❖ *Efecto inmediato:* El efecto que produce la Pascua en el ámbito comercial, específicamente en los días feriados; por ejemplo las ventas realizadas en éstos días exclusivos.
- ❖ *Efecto gradual:* El efecto que produce la Pascua en los días previos a la misma, por ejemplo el número de viajeros y las ventas en las proximidades al feriado.

### 2.1.6. La Componente Irregular

**Definición:** La componente irregular está formada por las fluctuaciones residuales erráticas de la serie, que no se pueden atribuir a las otras componentes ya mencionadas. Estas variaciones se pueden deber a diferentes motivos, entre ellos: Acontecimientos



inusuales como huelgas, clima más frío que lo normal, sequías; errores de medición de la serie tales como tratamiento de datos, errores de muestreo; errores en la toma de decisiones: pánico en el mercado bursátil originado en rumores, etc.



**GRÁFICO 2.3** Serie temporal del índice I-x03 de las Cuentas Nacionales  
Fuente: Banco Central del Ecuador

Como se puede observar en el gráfico de la serie trimestral del índice (I-x03), éste no presenta ninguna de las componentes mencionadas anteriormente, a ello se le atribuye el nombre de irregular.

#### **2.1.6.1 Determinantes de la irregularidad de las series:**

- ❖ *Naturaleza de la variable:* Algunas variables son esencialmente volátiles, como las exportaciones; otras se comportan de manera regular, como la población.

- ❖ *Nivel de agregación de la serie:* Para una variable dada, las series nacionales tienden a ser menos irregulares que las provinciales.
- ❖ *Frecuencia de la serie:* Las series anuales son generalmente menos irregulares que las series trimestrales y mensuales.
- ❖ *Fiabilidad de la fuente de datos:* Las series provenientes de una encuesta son más erráticas que las series que se originan en censos o en fuentes administrativas; las series basadas en modelos son por lo general más regulares, por construcción.

El componente  $i_t$  juega un papel fundamental en el planteamiento estadístico: es una clave para la validación del modelo. En efecto, este término de ajuste no debe contener información puesto que eso invalidaría completamente la interpretación de los otros componentes.

En términos de ingeniería, puede imaginarse que se trata de extraer de la serie original una señal, interpretable y de descuidar el ruido que acompaña la observación. Si después de la descomposición, permanece parte de la señal en el ruido es que el modelo elegido es inadecuado.

En términos estadísticos, se considerará que el componente irregular resultante de la aplicación del modelo debe presentar todas las

características de ruido blanco. La comprobación de esto es imprescindible para la validación del modelo de descomposición.

Las componentes de días laborables y de Pascua son eventos que se pueden prever, a diferencia del resto de las componentes mencionadas.

Después de explicar todos los componentes de una Serie Temporal, se presentan los modelos básicos que exponen de manera explícita la relación que guardan los componentes de una serie.

## **2.2 DESCOMPOSICIÓN DE LAS SERIES DE TIEMPO**

Para efectos del análisis económico, la estimación de los componentes no observables de una serie de tiempo cobra relevancia. Por ejemplo, el conocimiento de los movimientos estacionales contribuye a explicar si los cambios que se están observando en una variable, en determinado momento, obedecen efectivamente a aumentos o disminuciones en su nivel medio o bien a fenómenos estacionales. Adicionalmente, la posibilidad de aislar los factores estacionales permite el estudio de su comportamiento e identificar si son o no estables a lo largo del tiempo. Por otro lado, contar con una

estimación de la tendencia de la serie permite efectuar consideraciones acerca del crecimiento subyacente de la misma.

Antes de presentar los modelos de descomposición, se introducirá el concepto de desestacionalización de una serie.

### **DESESTACIONALIZACIÓN DE UNA SERIE**

**Definición:** Se puede definir la serie desestacionalizada como la combinación de la tendencia-ciclo con la componente irregular, o también como la serie original a la que se le ha quitado la componente estacional, las variaciones por días laborales y los feriados de fecha móvil.

#### **Objetivo de la desestacionalización:**

- ❖ Suprimir las fluctuaciones intra-anales sistemáticas para revelar los movimientos subyacentes de la tendencia-ciclo.
- ❖ La desestacionalización se obtiene estimando las componentes de la serie de tiempo y eliminando la estacionalidad, las variaciones por días laborales y el efecto Pascua.

Las series desestacionalizadas son necesarias para realizar comparaciones mes a mes, pero éstas comparaciones no son válidas en series con estacionalidad o con variaciones de días laborables significativas. También son necesarias para realizar comparaciones con el mismo mes del año anterior, de igual forma éstas comparaciones no son válidas en caso de que la estacionalidad evolucione rápidamente o en presencia de variaciones de días laborables. No se puede hacer ninguna inferencia de la evolución de la serie último año que es generalmente el período más importante para el análisis de coyuntura.

## 2.2.1 MODELOS DE DESCOMPOSICIÓN

**2.2.1.1 Modelo Aditivo:** Se utiliza cuando los componentes son independientes entre sí o se presentan valores negativos o ceros.

$$Z_t = C_t + S_t + r_t + h_t + i_t$$

Las componentes están expresadas en unidades. En este caso la serie desestacionalizada se obtiene como:

$$Z_t^{\text{des}} = Z_t - S_t - r_t - h_t = C_t + i_t$$

**2.2.1.2 Modelo Multiplicativo:** Se utiliza cuando los componentes son dependientes entre sí, o cuando el nivel de las series es muy cambiante.

$$z_t = c_t \times s_t \times r_t \times h_t \times i_t$$

En este modelo, la serie desestacionalizada se obtiene como:

$$z_t^{\text{des}} = z_t / (s_t \times r_t \times h_t) = c_t \times i_t$$

Si la amplitud del componente estacional varía en forma proporcional a la tendencia media anual, el modelo más apropiado es el *multiplicativo*. Si en cambio el componente estacional permanece constante ante variaciones en la tendencia, el modelo más adecuado es el *aditivo*, es decir la estacionalidad de la serie es independiente de su tendencia.

Se sabe por estudios realizados que la mayoría de las series económicas siguen un modelo multiplicativo.

En los casos en que la serie presenta valores negativos o ceros, el único modelo que se puede aplicar es el “aditivo”.

## 2.3 MÉTODOS DE EXTRACCIÓN DE SEÑALES

### 2.3.1 METODOLOGÍA TRAMO

**Definición:** TRAMO ("Time Series Regression with Arima Noise, Missing Observations and Outliers") es un programa diseñado para la estimación y predicción de modelos de regresión con observaciones perdidas y errores ARIMA, que generalmente no son estacionarios. TRAMO también realiza interpolaciones para cualquier secuencia posible de observaciones perdidas, e identifica y corrige las series de cuatro tipos de datos atípicos, de algunos tipos de efectos especiales; como los ya mencionados anteriormente: efecto de días de vacación y efecto calendario.

Existen ciertas consideraciones técnicas que deben tomarse en cuenta en la metodología TRAMO:

Dado el siguiente vector de observaciones:

$$Z = (Z_{t1}, \dots, Z_{tM})'$$

Donde  $0 < t_1 < t_M$ ,

El programa ajusta el modelo de regresión:

$$Z_t = \beta y'_t + \gamma_t,$$

Donde  $\beta = (\beta_1, \beta_2, \dots, \beta_n)'$  Es el vector de coeficientes de regresión.

$y'_t = (y_{1t}, \dots, y_{nt})$  Denota las  $n$  variables regresoras, y  $\gamma_t$  sigue el proceso general ARIMA de la forma siguiente:

$$\phi(L)\delta(L)\gamma_t = \theta(L)a_t,$$

Donde:

$L$  Es el operador de retardo.

$\phi(L), \delta(L), \theta(L)$  Denotan los polinomios finitos en  $L$

$a_t$  Se asume como innovación ruido blanco n.i.i.d. parámetros  $(0, \sigma_a^2)$ .



Además:

$\delta(L)$  Es el polinomio de raíces unitarias – parte integrada.

$$\delta(L) = (1-L)^d (1-L^s)$$

$\phi(L)$  Es el polinomio de raíces – parte autorregresiva.

$$\phi(L) = (1 + \phi_1 L + \dots + \phi_p L^p) (1 + \Phi_1 L^s + \dots + \Phi_p L^{s \times p})$$

$\theta(L)$  Es el polinomio de medias móviles.

$$\theta(L) = (1 + \theta_1 L + \dots + \theta_q L^q) (1 + \Theta_1 L^s + \dots + \Theta_q L^{s \times q})$$

El programa:

1. Estima los parámetros de máxima verosimilitud exacta (o incondicional / condicional de mínimos cuadrados).
2. Detecta y corrige las series de tiempo de algunos tipos de datos atípicos.

3. Genera interpolaciones óptimas de las observaciones perdidas y sus ECM (Error Cuadrático Medio) asociados.
4. Presenta una opción para la identificación automática del modelo y para el tratamiento automático de los datos atípicos.

A continuación se presentan algunos ejemplos de datos atípicos:

*Innovativos:* Es una alteración ocasional que tendrá un comportamiento tipo ARIMA.

*Aditivos:* Alteración ocasional de un solo dato de la serie observada.

*Cambio temporal:* Es una perturbación que ocasiona un dato que afecta algunos de los siguientes hasta disiparse.

*Cambio de Nivel:* Es una perturbación ocasional de un dato que afecta a todos los datos futuros.

TRAMO no es el único método utilizado para la descomposición de una serie temporal, conocido también como extracción de señales; existen otros métodos que serán presentados a continuación:

❖ Métodos Empíricos:

1. X11-ARIMA

## 2. Census X11

### ❖ Método basados en modelos:

#### X12-ARIMA

Los métodos empíricos se basan en promedios móviles, si bien son capaces de extraer características comunes de una gran cantidad de series, no tienen un modelo definido y por tanto se limita la capacidad de análisis y del diagnóstico de los resultados.

Los métodos basados en modelos utilizan estimadores que son variables aleatorias con todas sus propiedades e intervalos de confianza. Estos métodos permiten efectuar un diagnóstico amplio de las estimaciones, esta característica es precisamente la que los hace superiores.

El interés del presente trabajo se centra en la metodología TRAMO y más adelante se presentará el comando TERROR incluido en el software TRAMO-SEATS.

# CAPITULO 3

## 3. CONTROL DE CALIDAD DE DATOS CON TRAMO

Los modelos econométricos presentan deficiencias por no contemplar las expectativas racionales y basarse en las expectativas adaptativas de los agentes económicos; además están expuestos a la calidad de la información disponible para realizar las estimaciones, de ahí surge la necesidad de que una dependencia dedicada a la investigación económica se interese por el desarrollo y aplicación de técnicas más avanzadas para detectar algún tipo de error en las cifras. Es decir, previo al tratamiento de estadístico de una serie de datos, es preciso realizar primero un control de calidad de los mismos.

Las técnicas disponibles hasta el presente para tratar conjuntos grandes de series han presentado limitaciones importantes, que hacen que la detección de errores sea fundamentalmente manual y se base en unos

criterios simples que utilizan una cantidad muy pequeña de la información disponible en la serie.

En estos casos, la detección se realiza esencialmente comparando el crecimiento de los dos últimos periodos, ignorando la historia completa de la serie y por ende la existencia de errores del pasado, por otro lado, la comparación se realiza a un nivel de agregación relativamente elevado. La información exógena a priori, es de vital importancia en la detección de errores, pero el disponer de un método automático que muestre indicios de error de una cifra, facilitará la revisión de gran cantidad de series en muy poco tiempo, (según Luna y Maravall 1999), un método de este tipo podría proporcionar una relación básica de partida sobre cuáles de los nuevos datos pueden catalogarse como “sospechosos” de ser erróneos, y por tanto de ser corregidos o eliminados.

Cualquier método automático tiene que basarse en una definición precisa de lo que significa que un dato sea “sospechoso”. Dado que se trata en el caso de series temporales, parece natural plantear el problema en los siguientes términos: dada la evolución que ha seguido la serie, se obtiene una predicción óptima para el próximo período, así como una medida de la desviación típica del error de predicción. Si, al llegar el nuevo, dato, éste resulta “excesivamente” alejado de la predicción, se clasifica como

sospechoso. El concepto de “excesivo” es sencillo de precisar. Por ejemplo, se puede considerar excesivo un error de predicción de un tamaño tal que, dadas las características de la serie, sólo se esperaría encontrarlo 1 de cada 100.000 veces, es decir con una probabilidad de que ocurra bastante insignificante estadísticamente.

A Continuación se presenta el comando TERROR incluido en el paquete integrado TRAMO SEATS, con el fin de ilustrar el potencial que tiene esta aplicación dentro del contexto de la agilización y sistematización en la auditoría de datos, y mejorar así la calidad de la información.

Este procedimiento se aplicará a las series de datos de las Cuentas Nacionales el cual es el motivo del presente trabajo.

### **3.1 TERROR (TRAMO FOR ERRORS): UNA APLICACIÓN DEL PROGRAMA TRAMO**

La idea de detectar errores en la producción de datos por medio de los errores de predicción en series temporales no es nueva, desde hace algunos años se han desarrollado diversos procedimientos con el fin de detectar errores en la producción de datos por medio de los errores de predicción en series temporales, en los cuales se destacan las ventajas de la utilización de modelos ARIMA; sin embargo, existen

problemas previos que deben ser considerados, como *la relación entre error de predicción, observación atípica y error en los datos*, podrán existir también otros fenómenos atípicos que afecten a las observaciones, como por ejemplo cambios en las normas de cálculo del dato o algún tipo de acontecimiento especial; en las bases que no han sido controladas cuidadosamente, los errores no corregidos del pasado tenderán a presentarse como valores atípicos y por ende pueden distorsionar fuertemente la identificación y estimación de los modelos ARIMA, así como sus predicciones, además las series desagregadas de las bases de datos con frecuencia no están completas.

El procedimiento que sigue el programa TRAMO es el siguiente:

1. El programa determina la transformación adecuada de la serie.
2. Identifica el modelo apropiado.
3. Detecta y corrige las observaciones atípicas que puedan existir.
4. Interpola las observaciones ausentes y estima, en su caso, efectos especiales de tipo calendario, pascua y días festivos. (Los cuales fueron explicados en el capítulo 2).

Los valores atípicos que reconoce el programa son de tres tipos:

- a) Aditivos. AO

- b) Cambios de nivel. LS
- c) Desplazamientos transitorios. TC

En el primer caso, son valores atípicos aislados que afectan solo un periodo; el segundo captura un salto en el nivel medio de la serie, que se mantiene en los periodos siguientes y el tercero representa saltos en el nivel de la serie que, de forma amortiguada, eventualmente regresan a su nivel anterior.

La efectividad del método para depurar series mensuales y trimestrales es grande.

Finalmente, basándose en el modelo identificado, se obtienen predicciones óptimas (con error cuadrático medio mínimo) para la serie corregida de las distorsiones producidas por los valores atípicos. TRAMO es un programa muy rápido que permite tratar, en un PC ordinario, más de 500.000 series en un día. El programa está fundamentalmente orientado a series mensuales y trimestrales, con un máximo de 600 observaciones y un mínimo de 36 observaciones si la serie es mensual, y 15 sí trimestral.



Esta facilidad del programa permite incrementar la efectividad del método a la hora de depurar series de tiempo mensuales y trimestrales.

Sin embargo, cuando se utiliza el programa TRAMO para la detección de errores es necesario tomar en consideración los siguientes puntos muy importantes:

- El primero se refiere a la detección de datos atípicos en la última (o penúltima) observación. Supongamos que se cuenta con la serie  $[x(1), \dots, x(T-1)]$  y en base a ella se obtiene la predicción para el período  $T$ . Si el programa detecta un valor atípico para  $T-1$  (el último período de la muestra), es imposible que pueda determinar si ese valor atípico es del tipo aditivo y durará sólo un período, o se trata de un cambio de nivel, que permanecerá en la serie. Obviamente, que la observación atípica sea de un tipo u otro afectará drásticamente a la predicción (el valor atípico aditivo prácticamente se ignora, mientras que el cambio de nivel se incorpora plenamente a la predicción). En el período  $T-1$  es imposible identificar el tipo de valor atípico para ese período, y el programa lo trata como aditivo, es decir, no lo considera. Una identificación mínima razonable requiere un mínimo de 2 o 3 períodos adicionales. En consecuencia, parece juicioso utilizar una opción que

permita detectar, pero no corregir, valores atípicos en la(s) última(s) observaciones. Es una forma de protegerse frente a la incertidumbre de cuál será la corrección final.

- El segundo problema que es necesario mencionar es el de determinar el valor (en términos absolutos) a partir del cual un error de predicción se juzga excesivo. Si, por ejemplo, se pretende aplicar TERROR a 100 series, un tamaño de error que se presenta 1 de cada 1.000 veces puede ser suficiente. Si se aplica a 100.000 series la proporción es demasiado alta, ya que (en promedio) se detectarían 100 valores sospechosos que serían completamente espúreos; para este caso, una proporción de 1 en 10 millones sería más adecuada. En consecuencia, el criterio para señalar un dato como sospechoso depende, en última instancia, del número de series a analizar. Depende también, por supuesto del interés del analista, es decir, del grado de precisión con que se quieren analizar los posibles errores en el grupo de series de que se trate (lo que el programa llama nivel de sensibilidad).

### **3.2 EJECUCIÓN DEL COMANDO TERROR (TRAMO FOR ERRORS)**

La ejecución del programa se resume en lo siguiente:

El archivo de entrada contiene como última observación (periodo T) el dato recién recibido, el cual se pretende clasificar como aceptable o sospechoso de ser erróneo. Para ello, el programa ignora ese último dato y calcula la predicción para T con la serie de datos hasta el periodo (T-1), así como el error estándar de la predicción. Por otro lado, divide el error de predicción (nueva observación menos su predicción) por dicho error estándar, obteniendo un valor  $t$ , el cual se utilizará para evaluar el dato, de acuerdo a los siguientes criterios:

- Si  $t$  (en valor absoluto) es menor que un cierto valor  $t_1$ , especificado a priori, el dato se clasifica como aceptable.
- Si  $t_1 < |t| < t_2$ , donde  $t_2 (> t_1)$  también ha sido especificado a priori, el dato se clasifica como “posiblemente” erróneo.
- Si  $|t| > t_2$ , el dato se clasifica como “probablemente” erróneo.

A continuación se presenta la aplicación del comando TERROR en las Cuentas Nacionales del Ecuador.

# **CAPITULO 4**

## **4. APLICACIÓN AL CONTROL DE CALIDAD DE DATOS DE SERIES TEMPORALES.**

El objetivo del presente trabajo es ejercer un control de calidad sobre las Series de las Cuentas Nacionales, y de esta manera depurar los posibles datos erróneos, presentando así una información veraz y confiable. Lo cual se convierte en eficiencia y así poder tomar decisiones más acertadas para el desarrollo del país.

La División Económica del Banco Central del Ecuador recibe gran cantidad de información mensual y trimestral de diferentes fuentes, tal información constituye la base para la construcción de indicadores económicos sector externo. Por ello, esta tesis tiene como fin el estudio del comando TERROR del paquete integrado TRAMO SEATS, con el fin de proponer un sistema ágil y sistematizado de auditoría de datos, que busque asegurar la calidad de la información que se utiliza

en el Banco Central para efectuar los diferentes cálculos y estudios de investigación que le competen.

#### **4.1 SERIES DE DATOS A UTILIZAR**

Las Series de las Cuentas Nacionales que serán estudiadas se las ha dividido en cuatro secciones.

- Importaciones
- Exportaciones
- Manufactura
- Mensuales

En total se analizarán 76 series de las cuales, 19 corresponden a Importaciones (m) cada una con un mínimo de 36 datos y un máximo de 64 datos; 21 series correspondientes a Exportaciones (x) cada una con 52 datos; 13 a Manufactura (pr) cada una con 76 datos y 23 series mensuales (ip) con un total de 264 datos cada serie.

Es importante determinar el valor (en términos absolutos) a partir del cual un error de predicción se juzga excesivo, por un lado el criterio para señalar un dato como sospechoso dependerá del interés del analista, o sea del grado de precisión con que se

quieran analizar los posibles errores (lo que el programa denomina niveles de sensibilidad). Conforme el nivel de sensibilidad aumente, los parámetros para  $t_1$  y  $t_2$  tienden a disminuir.

**Tabla IV**

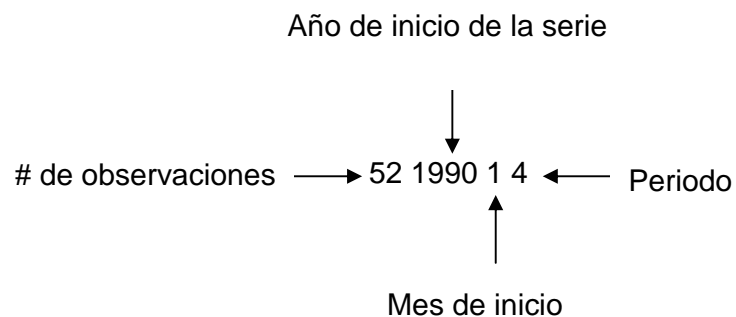
<b>Grados de Sensibilidad</b>		
<b>Nivel de Sensibilidad</b>	<b>Parámetro Inicial</b>	<b>Parámetro Final</b>
S=0: baja	$t_1: 5$	$t_2: 6$
S=1: media	$t_1: 4$	$t_2: 5$
S=2: alta	$t_1: 3$	$t_2: 4$

Elaboración Propia

Dependiendo del grado de sensibilidad, el programa detecta más o menos datos atípicos; si el grado de sensibilidad aumenta, con ello también aumenta la detección de datos atípicos. El grado de sensibilidad depende básicamente de la precisión que le asigne el investigador. La tabla IV muestra el nivel de sensibilidad que se desee utilizar, siendo la de mayor precisión, los parámetros  $t_1: 3$  y  $t_2: 4$ .

Para que el programa TRAMO SEATS lea correctamente el archivo que contiene la serie de la variable de interés, este debe cumplir con ciertas especificaciones:

- En Primer lugar, se debe guardar la serie en una hoja de Microsoft Excel.
- En la celda A1 se debe escribir el nombre o nomenclatura de la variable a utilizar.
- En la celda A2, se establecen las características del archivo en el siguiente orden y con un espacio entre sí: número de observaciones, año de la primera observación, mes de observación y periodicidad. Por ejemplo para el caso de la serie de exportaciones **I-x01** se tiene:



Para efectos prácticos el procedimiento automático para llevar a cabo la aplicación del comando **TERROR** es el siguiente:

- 1.- Establecer el formato de las series de datos a utilizar de acuerdo a las condiciones definidas anteriormente.

2.- Una vez seleccionadas las series de tiempo, en el panel de especificación de los datos se escoge la opción TERROR, que acciona ITER=2, lo que significa que el programa muestra un modelo común para todas las series; posteriormente, con (RSA= 4) se habilita el comando para generar la opción del modelo automático a las diferentes series de datos. Ilustración (ver anexo 1 a).

3) En la pestaña TERROR, se definen los parámetros t1 (K1) y t2 (K2) que determinan el grado de sensibilidad al cual el analista va a someter la veracidad del último dato recibido. Además de la opción Nmatrix = 1 (por defecto) que adiciona otra información en la salida resumen del Outmatrix como el t de los parámetros encontrados, t del efecto calendario, t de los outliers y t de las variables de regresión.

4) Finalmente se da clic a "OK" y otro clic al botón "Run"; en la opción Output se podrá ver la clasificación del último dato para las diversas series como "posiblemente erróneo" y "probablemente erróneo" (ver anexo 1 b).



## **4.2 ANÁLISIS DE RESULTADOS**

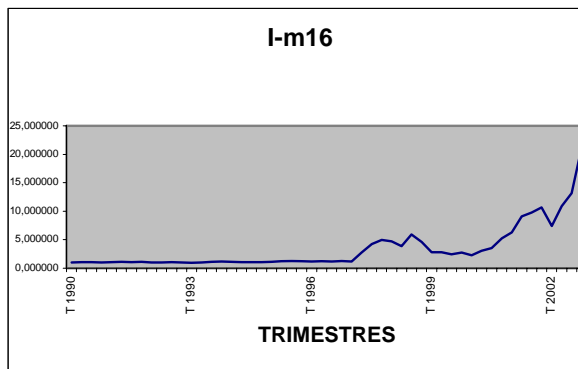
Al aplicar el comando TERROR se pueden obtener dos resultados importantes:

- I. El análisis de calidad de la última cifra recibida, ubicando el dato como probablemente erróneo, posiblemente erróneo o aceptable.
  
- II. Un resumen de las principales características del modelo encontrado por TRAMO accediendo a la opción Out Matrix, en donde, se obtiene información acerca del modelo, como parámetros ARMA, Efecto Determinístico, datos atípicos y Efecto Calendario.

A continuación se muestra los resultados obtenidos:

### **4.2.1 Resultados de las Series de Importaciones**

De las 19 series analizadas 18 pasaron la prueba íntegramente. La serie I-m16, presenta como último valor: 21.16486; mientras que el programa presenta como valor predicho 13.25443, mostrando una diferencia de 7.910430, siendo el valor t: 5.328650 es decir que es probablemente erróneo.



**GRÁFICO 4.1** Serie Temporal del índice I-m16 de las Cuentas Nacionales  
Fuente: BCE

De las 19 series de importaciones analizadas, 15 presentan un comportamiento no estacionario, y tan solo 4 pueden verse como oscilaciones en torno a un nivel relativamente fijo. Por otra parte, aproximadamente el 32% contiene estacionalidad.

**4.2.1.1 Diferencias tomadas (m)-** Se resume en la tabla siguiente, donde BD es el orden de las diferencias estacionales, se tomaron, 7 diferencias con  $D=0$  y 12 diferencias con  $D=1$ .

Tabla V

# de Series con	D=0	D=1	D=2	TOTAL
BD=0	4	9	0	13
BD=1	3	3	0	6
TOTAL	7	12	0	19

Diferencias tomadas en las series de importaciones

**4.2.1.2 Datos Atípicos (m)-** El programa reconoció un total de 69 datos atípicos, de los cuales 22 corresponden a datos atípicos aditivos; 17 son por cambios de nivel y 30 corresponden a desplazamientos transitorios.

La siguiente tabla resume los datos atípicos reconocidos por TRAMO:

Tabla VI

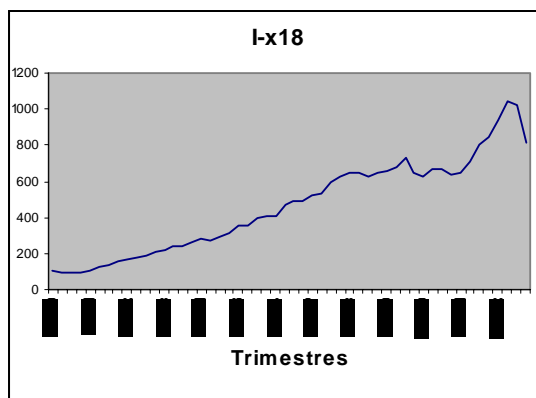
	AO	LS	TC	TOTAL
<b>% de series con</b>	63.2	47.4	73.7	89.5
<b>Media</b>	1.2	0.9	1.6	3.6
<b>Máximo</b>	4	3	5	8
<b>Mínimo</b>	0	0	0	0

Datos atípicos de las series importaciones

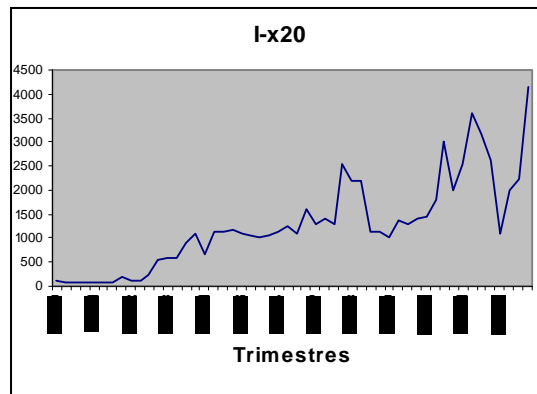
No existieron observaciones perdidas, tampoco el efecto de los días laborables, ni el efecto pascua. Solo dos series no presentaron ningún dato atípico, las otras 17 series presentaron por lo menos un tipo de dato atípico.

#### 4.2.2 Resultados de las Series de Exportaciones

De las 21 series analizadas 19 pasaron la prueba íntegramente. Las series I-x18 y la serie I-x20 presentan como nuevos valores 816.5859 y 4163.109 respectivamente siendo los valores predichos por el programa: 1068.338 para la primera y 2293.058 para la segunda, teniendo así una diferencia en valor absoluto de 251.7519 y 1870.051; y encontrándose el valor  $t$  entre  $t_1$  y  $t_2$ ; por lo que los datos son posiblemente erróneos.



**GRÁFICO 4.2** Serie Temporal I-x 18 de las Cuentas Nacionales  
Fuente: BCE



**GRÁFICO 4.3** Serie Temporal I-x 20 de las Cuentas Nacionales  
Fuente: BCE

De las 21 series de exportaciones, 14 presentan un comportamiento no estacionario y las otras 7 tienen un comportamiento estacionario. Aproximadamente el 29% de las series contiene estacionalidad.

**4.2.2.1 Diferencias tomadas (x)-** A continuación en la tabla se resume las diferencias tomadas:

Donde BD es el orden de las diferencias estacionales, se tomaron, 9 diferencias con  $D=0$  y 12 diferencias con  $D=1$ .

Tabla VII

# de Series con	D=0	D=1	D=2	TOTAL
BD=0	7	8	0	15
BD=1	2	4	0	6
TOTAL	9	12	0	21

Diferencias tomadas en las series de exportaciones.

**4.2.2.2 Datos Atípicos (x)-** El programa reconoció un total de 45 datos atípicos, de los cuales 15 corresponden a datos atípicos aditivos; 14 son por cambios de nivel y los 16 restantes corresponden a desplazamientos transitorios.

La siguiente tabla resume los datos atípicos reconocidos por TRAMO:

Tabla VIII

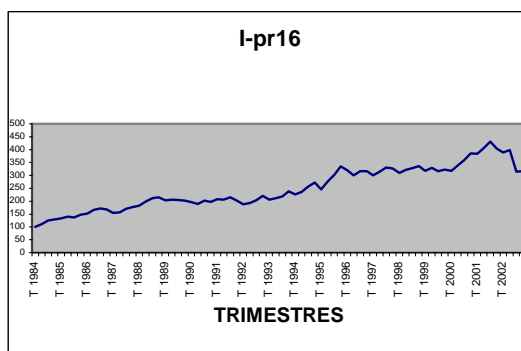
	AO	LS	TC	TOTAL
<b>% de series con</b>	42.9	28.6	33.3	57.1
<b>Media</b>	0.7	0.7	0.8	2.2
<b>Máximo</b>	4	6	3	11
<b>Mínimo</b>	0	0	0	0

Datos atípicos de las series exportaciones

No existieron observaciones perdidas, tampoco el efecto de los días laborables, ni el efecto pascua. De las 21 series 9 de ellas no presentaron ningún dato atípico, las otras 12 series presentaron por lo menos un tipo de dato atípico, siendo la serie I- x14 la que presenta mayor cantidad de atípicos de los tres tipos, siendo 11 en total.

#### 4.2.3 Resultados de las Series de Manufactura

De las 13 series analizadas, 1 serie no pasó la prueba del programa. La serie I- pr16 presenta como nuevo valor 316.6371 y el valor predicho por el programa es: 418.4113, es decir que se presenta una diferencia de 101.7742; siendo el valor t 4.057254 por lo que TRAMO lo califica como posible de ser erróneo.



**GRÁFICO 4.4** Serie Temporal del índice I-pr 16 de las Cuentas Nacionales  
Fuente: BCE

Todas la series de manufactura presentan un comportamiento no estacionario. Diez de las series contienen estacionalidad.

**4.2.3.1 Diferencias tomadas (pr)-** A continuación en la tabla se resume las diferencias tomadas:

Como ya se describió anteriormente BD es el orden de las diferencias estacionales, se tomaron, 2 diferencias con D=0 y 11 diferencias con D=1.

**Tabla IX**

# de Series con	D=0	D=1	D=2	TOTAL
BD=0	0	3	0	3
BD=1	2	8	0	10
TOTAL	2	11	0	13

Diferencias tomadas en las series de manufactura.

**4.2.3.2 Datos Atípicos (pr)-** El programa reconoció un total de 25 datos atípicos, de los cuales 7 corresponden a datos atípicos aditivos; 8 corresponden a cambios de nivel y los 10 restantes corresponden a desplazamientos transitorios.

La siguiente tabla resume los datos atípicos reconocidos por TRAMO:



Tabla X

	<b>AO</b>	<b>LS</b>	<b>TC</b>	<b>TOTAL</b>
<b>% de series con</b>	61.5	23.1	46.2	76.9
<b>Media</b>	0.8	0.3	0.8	1.9
<b>Máximo</b>	3	2	3	6
<b>Mínimo</b>	0	0	0	0

Datos atípicos de las series de manufactura

No existieron observaciones perdidas, tampoco el efecto de los días laborables, ni el efecto pascua. De las 13 series 3 de ellas no presentaron ningún dato atípico, las otras 10 series presentaron por lo menos un tipo de dato atípico, siendo la serie I- pr21 la que presenta mayor cantidad de atípicos de los tres tipos, siendo 6 en total.

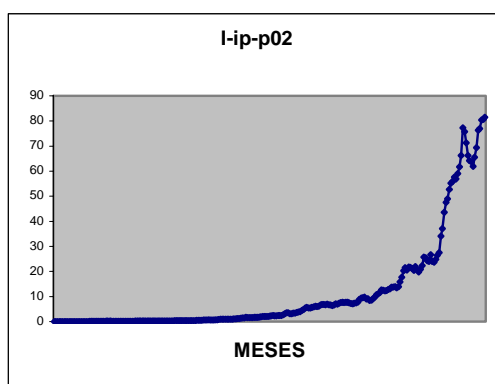
#### 4.2.4 Resultados de las Series Mensuales

De las 23 series analizadas 21 pasaron la prueba íntegramente. El último dato de las series: I-ip-p02 y I-ip-p01 es clasificado como posible y probablemente erróneo respectivamente.

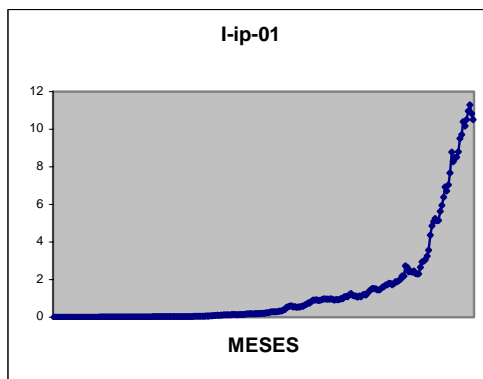
Con la primera serie I-ip-p02, el último dato es 82.73577 mientras que el dato que predice el programa es: 84.76686, existiendo una

diferencia de 2.031090, y siendo el valor  $t$  4.144580 por lo que es clasificado como posiblemente erróneo.

Por otro lado con la serie I-ip-p01 tenemos que el último dato es 11.09519 y el dato predicho por TRAMO es 12.15853, por lo que se tiene una diferencia de 1.063341, con un valor  $t$  9.430237, lo cual significa que el dato es probablemente erróneo.



**GRÁFICO 4.5** Serie Temporal I-ip-p02 de las Cuentas Nacionales  
Fuente: BCE



**GRÁFICO 4.6** Serie Temporal I-ip-p01 de las Cuentas Nacionales  
Fuente: BCE

Todas la series mensuales presentan un comportamiento no estacionario. Alrededor del 35% de las series contienen estacionalidad.

**4.2.4.1 Diferencias tomadas (ip)-** En la tabla siguiente se resume las diferencias tomadas.

Siendo BD el orden de las diferencias estacionales, se tomaron, 9 diferencias con D=1 y 14 diferencias con D=2.

**Tabla XI**

# de Series con	D=0	D=1	D=2	TOTAL
BD=0	0	3	12	15
BD=1	0	6	2	8
TOTAL	0	9	14	23

Diferencias tomadas en las series mensuales

**4.2.4.2 Datos Atípicos (ip)-** TRAMO reconoció un total de 484 datos atípicos, a diferencia de las series analizadas anteriormente, éstas contienen 264 datos cada una de las 23 series, por lo tanto la detección de datos atípicos es proporcional a la cantidad de datos que se analizan. De los 484 atípicos: 141 corresponden a datos atípicos aditivos; 147

corresponden a cambios de nivel y por último los 196 restantes corresponden a desplazamientos transitorios.

La siguiente tabla resume los datos atípicos reconocidos por TRAMO:

**Tabla XII**

	<b>AO</b>	<b>LS</b>	<b>TC</b>	<b>TOTAL</b>
<b>% de series con</b>	100	95.7	95.7	100
<b>Media</b>	6.1	6.4	8.5	21
<b>Máximo</b>	13	13	19	30
<b>Mínimo</b>	1	0	0	5

Datos atípicos de las series mensuales

De igual forma que en las series analizadas anteriormente no existieron observaciones perdidas, tampoco el efecto de los días laborables, ni el efecto pascua. Todas las series presentaron por lo menos un tipo de dato atípico, siendo las series I-ip-p19 y I-ip-p20 las que presentaron mayor cantidad de ellos con total 30 de los tres tipos reconocidos de los 264 datos de cada serie.

Es necesario recalcar que para el análisis de éstas series se aplicó el nivel de sensibilidad media es decir con  $t_1$ : 4 y  $t_2$ : 5, aunque también se realizaron pruebas con sensibilidad alta donde el programa reconoció unos cuantos

más datos atípicos. Pero como ya se mencionó anteriormente esto depende básicamente de la precisión que le desee dar el analista.

En general se puede resumir lo siguiente:

De las 19 series de importaciones , la serie del índice I-m16 presenta el último dato como probable de ser erróneo. En total se obtuvieron 69 datos atípicos. En el caso de las 21 series de exportaciones, se tiene que en las series I-x18 y I-x20 los últimos datos de éstas son clasificados como posiblemente erróneos; el programa reconoció un total de 45 atípicos. Para las series de manufactura, de las 13 las serie del índice I-pr16 presenta el último dato como posiblemente erróneo y un total de 25 datos atípicos. Por último, del total de las 23 series mensuales, la serie del índice Ip-p02 presenta el último dato como posible erróneo y la serie del índice IP-p01 como probablemente erróneo, aquí el programa reconoció en total 484 datos atípicos.

Tabla XIII

	<b>M</b>	<b>X</b>	<b>pr</b>	<b>lp</b>
No. de series analizadas	19	21	13	23
Total de datos atípicos	69	45	25	484
Dato probablemente erróneo	1	0	0	1
Dato posiblemente erróneo	0	2	1	1

## Resultados obtenidos

En total se detectaron 623 datos atípicos, es necesario recalcar que: en la serie de importaciones se analizaron un total 1116 datos de los cuales 69 son datos atípicos lo que representa un 6% del total de los datos analizados. En cuanto a las exportaciones se analizaron 1092 datos de los cuales el 4% corresponde a datos atípicos. En las series de manufactura se analizaron 932 datos con un 3% de datos atípicos y por último en las series mensuales se analizaron en total 6072 datos con un 8% de datos atípicos.

# CAPITULO 5

## 5.1 CONCLUSIONES

Dados los resultados obtenidos en el presente estudio se pueden resaltar los siguientes puntos:

- I. En este trabajo se estudió el comando TERROR incluido en el paquete integrado TRAMO SEATS, con el fin de ilustrar el potencial que tiene este comando dentro del contexto de la agilización y sistematización en la auditoría de datos. Lo anterior ayudará a mejorar la calidad de la información que se utiliza en el Banco Central del Ecuador para efectuar los diferentes cálculos y estudios de investigación que le competen. Desde este punto de vista se ha cumplido con este objetivo, ya que se ha evaluado el comando

TERROR como alternativa para realizar una auditoría formal a los datos recibidos de diferentes fuentes.

- II. Un elemento a resaltar es el hecho de que los programas son gratuitos y pueden ser obtenidos mediante Internet o ser solicitados a sus autores; sin embargo, dado que estos programas constantemente se encuentran en continuo mejoramiento, se hace necesario contar con la última versión o posibles modificaciones a la versión.
- III. En cuanto a la detección de valores sospechosos se puede agregar que del conocimiento que posea el analista de la serie dependerán los valores  $t_1$  y  $t_2$  que se utilicen como criterio para evaluar si un dato es sospechoso de error. Sin embargo, de acuerdo a los resultados cuanto más elevados sean los valores de  $t_1$  y  $t_2$ , menos observaciones sospechosas serán detectadas, pero la probabilidad de que un dato sea sospechoso de un error aumenta.
- IV. El programa TRAMO da la posibilidad de personalizar un modelo con base en información observada, con el fin de obtener un mejor ajuste y cumplir con los criterios de decisión de mejor forma que el modelo automático.



- V. Los resultados del este trabajo revelan que ante casos como “atípicos” la herramienta funciona de forma eficiente para detectar valores sospechosos; sin embargo, es importante considerar que las series están siendo adecuadas para ser detectadas por TERROR. Adicionalmente, la detección de datos sospechosos puede indicar la ocurrencia de alguna situación de gran importancia, ya sea un hecho particular como algún desastre natural, una huelga, etc.

## **5.2 RECOMENDACIONES**

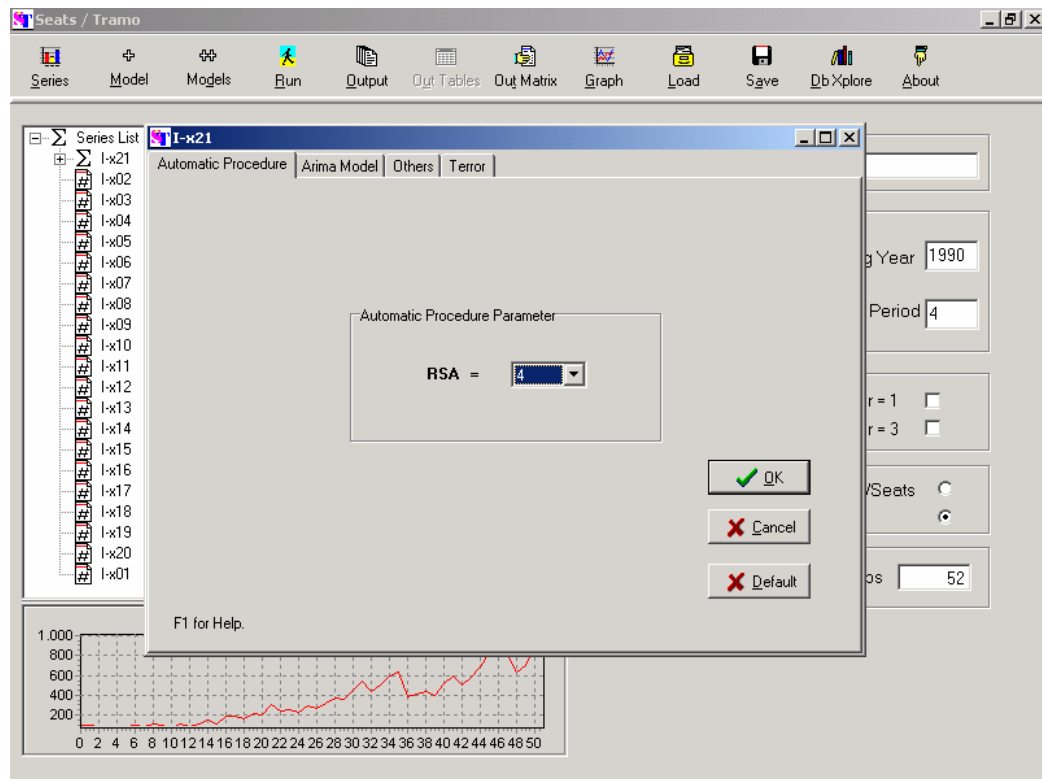
- I. En este estudio se han utilizado series de diversas áreas del Banco Central del Ecuador mensuales y trimestrales de exportaciones, importaciones y referente al área de manufactura, para mostrar y generalizar el uso de esta herramienta en el control y calidad de datos, se han utilizado series de datos consideradas como “producto final” con el fin de asimilar el uso de la herramienta, sin embargo, se recomienda aplicar este procedimiento a las series que son las fuentes de datos con los cuales se generan los “datos agregados en general del BCE”, esto permitirá una verificación de las cifras

consideradas como sospechosas antes de ser utilizadas para la elaboración de estadísticas.

- II. Como se mencionó anteriormente el grado de sensibilidad que se desee asignar depende básicamente del grado de precisión que el investigador asigne al análisis de datos. En el presente trabajo se escogió el nivel de sensibilidad medio ( $t_1: 4$ ,  $t_2: 5$ ), aunque también se realizaron pruebas con sensibilidad alta obteniéndose más cantidad de datos atípicos.
  
- III. Finalmente, esta tesis constituye un aporte a la auditoría de datos que maneja el Banco Central de una forma más rápida, eficiente y automática, por lo cual, se recomienda ampliar el uso de esta herramienta en el Control y Calidad de los datos en todas las áreas del BCE y considerarlo una opción viable para ser planteado como proyecto.

## ANEXO 1 : Ilustraciones del programa TRAMO – SEATS , comando TERROR

### a) Pantalla principal y elección del modelo automático (RSA = 4)



## b) Parámetros de Sensibilidad (Media)

The screenshot displays the 'Seats / Tramo' software interface. The main window shows a 'Series List' on the left with series I-x01 through I-x21. A 'Time Series Plot' is visible at the bottom left, showing a red line graph of data points over time (0 to 50). The 'Error Parameters' dialog box is open, showing the following settings:

- Sens = 1
- Nmatrix = 1
- K1 = 4
- K2 = 5
- Minabs = 1

The dialog box also includes 'OK', 'Cancel', and 'Default' buttons. The background interface includes a menu bar with options like 'Series', 'Model', 'Models', 'Run', 'Output', 'Out Tables', 'Out Matrix', 'Graph', 'Load', 'Save', 'Db Xplore', and 'About'. The 'Automatic Procedure' tab is selected in the dialog box.

ANEXO 2 : Hoja de resultados del programa TRAMO – SEATS , comando  
TERROR

a) Resultados Series Importaciones

Input Parameters :

- int2=-3
- rsa= 4
- terror= 1
- modelsumm= 1
- k1= 4.000
- k2= 5.000
- Minabs= 1.000

SERIES TITLE	New Value	Forecast	Diff.	StdDev	T-Value	Results
I-m16	21.16486	13.25443	7.910430	0.0885651	5.328650	Likely

Summary Statistics

- 19 Series were tested.
- 0 Releases were suspicious (possibly wrong).
- 1 Releases were very suspicious (likely wrong).
- 0 Series produced a Run-Time EXCEPTION.
- 0 Series did not match TERROR memory constraints.
- 18 Series passed the plausibility tests.

b) Resultados Series Exportaciones

Input Parameters :

- int2=-3
- rsa= 4
- terror= 1
- modelsumm= 1
- k1= 4.000
- k2= 5.000
- Minabs= 1.000

SERIES TITLE	New Value	Forecast	Diff.	StdDev	T-Value	Results
I-x18	816.5859	1068.338	-251.7519	0.0580134	-4.603149	Possible
I-x20	4163.109	2293.058	1870.051	0.1285479	4.703602	Possible

#### Summary Statistics

- 21 Series were tested.
- 2 Releases were suspicious (possibly wrong).
- 0 Releases were very suspicious (likely wrong).
- 0 Series produced a Run-Time EXCEPTION.
- 0 Series did not match TERROR memory constraints.
- 19 Series passed the plausibility tests.

#### c) Resultados Series Manufactura

##### Input Parameters :

- int2=-3
- rsa= 4
- terror= 1
- modelsumm= 1
- k1= 4.000
- k2= 5.000
- Minabs= 1.000

SERIES TITLE	New Value	Forecast	Diff.	StdDev	T-Value	Results
I-pr16	316.6371	418.4113	-101.7742	0.0681220	-4.057254	Possible

#### Summary Statistics

- 13 Series were tested.
- 1 Releases were suspicious (possibly wrong).
- 0 Releases were very suspicious (likely wrong).
- 0 Series produced a Run-Time EXCEPTION.
- 0 Series did not match TERROR memory constraints.
- 12 Series passed the plausibility tests.

## d) Resultados Series Mensuales

## Input Parameters :

- int2=-3
- rsa= 4
- terror= 1
- modelsumm= 1
- k1= 4.000
- k2= 5.000
- Minabs= 1.000

SERIES TITLE	New Value	Forecast	Diff.	StdDev	T-Value	Results
l-ip-p02	82.73577	84.76686	-2.031090	0.4900594	-4.144580	Possible
l-ip-p01	11.09519	12.15853	-1.063341	0.1127587	-9.430237	Likely

## Summary Statistics

23 Series were tested.

1 Releases were suspicious (possibly wrong).

1 Releases were very suspicious (likely wrong).

0 Series produced a Run-Time EXCEPTION.

0 Series did not match TERROR memory constraints.

21 Series passed the plausibility tests.

## BIBLIOGRAFÍA

- I. Hamilton, James D. (1994), Time Series Analysis, Princeton University Press.
- II. Mahía, Ramón. (Abril, 2001). Conceptos Básicos de Matemática Económica. Departamento de Investigaciones Económicas, Banco de España .
- III. Luna, Cristina y Maravall, Agustín. 1999. Un nuevo método para el control de calidad de los datos en series temporales. Boletín Económico, Servicio de Estudios, Banco de España.
- IV. Gómez, Víctor. y Maravall, Agustín. (1997), "Programs TRAMO and SEATS: Instructions for the User", Servicio de Estudios, Banco de España.
- V. Maravall, Agustín. (Abril, 2002). Brief description of the programs, Servicio de Estudios, Banco de España.



- VI. Quiroz S., Juan Carlos. (Junio, 2002). Departamento de Investigaciones Económicas, División Económica, Die/01-2002-Nt. Banco Central de Costa Rica,
- VII. Caporello, Gianluca y Maravall, Agustín. (Abril, 2002) A tool for Quality Control of Time series data Program TERROR. Banco de España.
- VIII. Caporello, G.; Maravall, A. y Sánchez, F. (Marzo, 2002) Program TRAMO – SEATS Refence Manual. Banco de España (<http://www.bde.es>)
- IX. Maravall Agustín, “Consultas realizadas sobre el comando TERROR”, [maravall@bde.es], septiembre 2003, enero 2004.