

Distr.
RESTRINGIDA

LC/DEM/R.132
Serie A, N°228
octubre de 1991

ORIGINAL: ESPAÑOL

CELADE

Centro Latinoamericano de Demografía

**ESTUDIOS DE SERIES TEMPORALES DEMOGRAFICAS MEDIANTE
MODELOS ARIMA DE BOX-JENKINS.**

(Una aplicación a series de la ciudad de Rosario, Argentina).

Este documento fue preparado para ser presentado a la Reunión Técnica sobre "Nuevas Metodologías Estadísticas Aplicadas a la Demografía", organizada por el Centro Latinoamericano de Demografía, con la colaboración de la Universidad Nacional de Rosario, Argentina y del Centro Interamericano de Enseñanza de Estadística (CIENES), y que tuvo lugar los días 2, 3 y 4 de septiembre de 1991. Las opiniones expresadas en este trabajo son de exclusiva responsabilidad de sus autoras y pueden no coincidir con las de las instituciones mencionadas.



NACIONES UNIDAS

UNITED NATIONS

CENTRO LATINOAMERICANO DE DEMOGRAFIA

REUNION TECNICA SOBRE "NUEVAS METODOLOGIAS ESTADISTICAS
APLICADAS A LA DEMOGRAFIA"

2,3 y 4 de Setiembre, 1991

ESTUDIOS DE SERIES TEMPORALES DEMOGRAFICAS

MEDIANTE MODELOS ARIMA DE BOX-JENKINS

(Una aplicación a Series de la Ciudad de Rosario, Argentina)

María Teresa Blaconá
Cristina Cuesta
Claudia Giovagnoli

Santiago, Chile, 1991

UNIVERSIDAD NACIONAL DE ROSARIO
FACULTAD DE CIENCIAS ECONOMICAS Y ESTADISTICA
ESCUELA DE ESTADISTICA

**ESTUDIO DE SERIES TEMPORALES DEMOGRAFICAS
MEDIANTE MODELOS ARIMA DE BOX-JENKINS:**
Una Aplicación a Series de la ciudad de Rosario (Argentina)

Rosario, 1991

**ESTUDIOS DE SERIES TEMPORALES DEMOGRAFICAS
MEDIANTE MODELOS ARIMA DE BOX-JENJINS**
Una Aplicación a Series de la Ciudad de Rosario (Argentina)

*Estadística María Teresa Blaconá
Lic. Estadística Cristina Cuesta
Lic. Estadística Claudia Giovagnoli*

INDICE

I. INTRODUCCION	3
II. MATERIAL DE ESTUDIO	5
III. METODOLOGIA	7
III.1. Modelo Logístico Modificado	7
III.2. Modelos ARIMA de Box y Jenkins	7
III.3. Desestacionalización de series.Método de Medias Móviles.....	8
IV. ANALISIS DE LOS RESULTADOS	8
IV.1. SERIES ANUALES	8
IV.1.1. Serie de población	8
IV.1.2. Serie de tasas de natalidad	13
IV.1.3. Serie de tasas de mortalidad	15
IV.1.4. Serie de tasas de mortalidad infantil	17
IV.2. SERIES MENSUALES	19
IV.2.1. Serie de tasas de mortalidad	19
IV.2.2. Serie de tasas de natalidad	22
V. CONSIDERACIONES FINALES	28
VI. BIBLIOGRAFIA	30
APENDICES :	32
1. Serie de población anual	I
2. Serie de tasas de natalidad anual	II
3. Serie de tasas de mortalidad anual	III
4. Serie de tasas de mortalidad infantil anual	IV
5. Serie de tasas de natalidad mensual	V
6. Serie de tasas de mortalidad mensual	VIII
7. Modelo Logístico Modificado (MLM).....	XI
8. Modelos ARIMA de Box-Jenkins	XII
9. Modelos de descomposición	XVIII

PALABRAS CLAVES: Modelos ARIMA. Series demográficas. Proyecciones demográficas.

I. INTRODUCCION

En demografía existen tres métodos principales que se usan para realizar previsiones⁽¹⁾; ellos son:

1.- Métodos demográficos que tienen en cuenta componentes de cohortes o datos anticipados, usados preferentemente para previsiones a largo plazo.

2.- Modelos estructurales o causales, que ponen en relación variables demográficas o a éstas con variables socio-económicas y permiten si es posible encontrar una causa principal de cambios sobre todo de política demográfica o socio-económica y tenerla en cuenta para realizar previsiones.

3.- Métodos estadísticos de serie de tiempo, estas técnicas tratan de modelar una serie demográfica teniendo en cuenta la información que ella misma brinda para realizar previsiones, por lo general estas previsiones son a corto plazo. Dentro de estos métodos pueden considerarse dos grandes grupos, por un lado los tradicionales los cuales consideran a la serie como una función del tiempo (logística, polinomios de distinto grado, exponencial, etc.) más una componente aleatoria, y los que consideran a la serie observada como la realización de un proceso estocástico y se la ajusta por los modelos ARIMA de Box-Jenkins.

Los tres enfoques planteados no son métodos competitivos entre sí, sino que tienen distintos objetivos, requieren información diferente y en muchos casos en un estudio pueden complementarse entre ellos.

En el primero de ellos **Método de Componentes** tiene en cuenta distintos tipos de indicadores, por un lado los que dan una medida del nivel general por ejemplo fecundidad, mortalidad, migración y por otro permiten visualizar la estructura por edad, sexo, etc.⁽²⁾. Cuando se realiza una proyección por este método se tendrán en cuenta por lo tanto valores globales de las variables como estructuras desagregadas para cada una de ellas.

El método de los **Modelos Estructurales** o modelos causales se basan en plantear relaciones entre variables demográficas y/o variables socio-económicas de tal modo que se pueda predecir una variable demográfica a través de otras variables demográficas y/o socio-económicas. Estos modelos son de difícil construcción requiriendo mucho esfuerzo, y no siempre obtienen resultados alentadores.

Por último los **Métodos de Series de Tiempo** que usan la propia información de la serie demográfica para realizar previsiones, como se dijo anteriormente se puede subdividir en dos tipos distintos de métodos:

3.1 Modelos de Descomposición en donde se considera a la serie como una cierta función matemática que depende de t , más un error aleatorio, los más conocidos son los que toman la función matemática como un polinomio en t , una exponencial, etc., el modelo más usado para población es la llamada curva logística o logística modificada⁽³⁾. Los coeficientes se estiman por mínimos cuadrados y se pueden realizar previsiones puntuales.

3.2 Modelos ARIMA de Box-Jenkins ⁽⁴⁾ consideran la serie como una realización de un proceso estocástico, por ello se pueden calcular tanto estimaciones puntuales como realizar intervalos de confianza para el valor futuro desconocido de la variable. Las previsiones que se calculan son por lo general sólo para variables globales como población, natalidad, mortalidad, mortalidad infantil, etc, por que para realizar proyecciones de las variables por edad, sexo, etc. se debería contar con registros a través de un período bastante largo de tiempo (como mínimo 50 años para series anuales o 60 meses para series mensuales), y esto por lo general no existe.

En este trabajo se aplicaran estos dos últimos métodos para las series de población, natalidad, mortalidad y mortalidad infantil de la ciudad de Rosario (Argentina). Estos análisis se realizan por un lado en series anuales de principio de siglo hasta el año 1986, con lo cual se podrá encontrar la tendencia de dichas variables demográficas y hacer previsiones a corto plazo; por otro lado se analizan las series mensuales de enero de 1972 a diciembre de 1986 de natalidad y mortalidad para conocer el comportamiento estacional de las mismas y realizar previsiones mensuales.

Cabe hacer notar que cualquiera de estos métodos deben ser aplicados a datos confiables, por que no se puede asegurar que su aplicación a datos deficientes conduzcan a resultados que sean los que realmente reflejan la realidad⁽⁵⁾. Debe, por lo tanto, tenerse en consideración, al aplicar técnicas de análisis, las hipótesis subyacentes a los métodos y la calidad de la información que se maneja. Es importante tener juicio crítico y no caer en un uso mecánico de los métodos expuestos.

II. MATERIAL DE ESTUDIO

El sentido de trabajar con series demográficas es que estas sean comparables en el tiempo, es por esto que es necesario tener en cuenta el tamaño de las poblaciones que son objeto de la comparación. Por este motivo surge el concepto de tasas o índices vitales. En estas medidas el numerador representa el número de sucesos vitales de algún tipo especificado durante un intervalo de tiempo preestablecido. Si el denominador representa el tamaño de la población dentro de la cual aconteció el suceso vital, el cociente resultante es llamado tasa de estadísticas vitales (por ej. tasa de natalidad, tasa de mortalidad, etc.)

A continuación se especifican las series bajo estudio. Debido a la disponibilidad de información se trabaja con series anuales cuyo seguimiento parte de principio de siglo y con series mensuales que se pudieron obtener a partir de 1972. Todas ellas llegan hasta 1986 ya que ese año corresponde al del último anuario publicado por la Dirección de Estadística de la Municipalidad de Rosario.

II.1. SERIES ANUALES:

- **Serie de Población de la Ciudad de Rosario de 1900 a 1986.**
(Apéndice 1)

Los datos se obtienen en su gran mayoría de los anuarios estadísticos de la Ciudad de Rosario. Es necesario recurrir a informes de la Dirección General de Estadística de la Municipalidad por no haberse publicado anuarios o boletines para los años comprendidos entre 1956 y 1969. A pesar de esto y consultando otras fuentes, resulta imposible obtener estimaciones de población para los años 1956/59.

Desde 1900 a 1932 inclusive, a los datos de cada censo se suma el crecimiento vegetativo y el saldo migratorio anual. A partir del año 1933 se considera únicamente el crecimiento vegetativo teniendo en cuenta la recomendación de la Primera Conferencia Nacional de Estadística realizada en la Ciudad de Córdoba en 1925. La población para los años 1900 y 1910 corresponde a censos municipales. En 1926 y 1933 se realizan también censos municipales pero no se consideran oficiales los datos. Para 1914, 1947, 1960, 1970 y 1980 se toma la información que brindan los censos nacionales relevados en dichos años.

Para este trabajo, debido a que los datos de la población corresponden al 31 de diciembre de cada año, los últimos tres censos nacionales se actualizan por el crecimiento vegetativo correspondiente a los meses que faltan para completar cada año, esto no se realiza con los otros censos debido a que no se cuenta con la información necesaria.

- **Serie de tasas de Natalidad de Rosario de 1906 a 1986.** (Apéndice 2)

$$\text{tasa de natalidad} = (N/P) * 1000$$

donde N : nacidos vivos en el año t
 P : población en el año t

■ **Serie de tasas de Mortalidad de Rosario de 1906 a 1986.**(Apéndice 3)

$$\text{tasa de mortalidad} = (D/P)*1000$$

donde D : defunciones en el año t
 P : población en el año t

■ **Serie de tasas de Mortalidad Infantil de Rosario de 1906 a 1986.**
(Apéndice 4)

$$\text{tasa de mortalidad infantil} = (M/N)*1000$$

donde M : defunciones de niños menores de un año en el año t
 N : nacidos vivos en el año t

Los valores de estas tasas anuales se obtienen directamente de las publicaciones de la Dirección General de Estadística de la Municipalidad. A partir de Dic. de 1960 y durante la década, por razones que se ignoran, la Dirección de Estadística no publicó más el Boletín Trimestral, abandonando el seguimiento de las series estadísticas de natalidad, mortalidad y mortalidad infantil.

II.2. SERIES MENSUALES:

Las series de tasas mensuales de natalidad y mortalidad no se publican. Se calculan tomando como numerador los valores de nacidos vivos y defunciones en el mes t , publicados en los anuarios y como denominador la población en el mes t . Para obtener la población mensual se parte de la población existente al 31 de diciembre de 1971 y se suma algebraicamente el crecimiento vegetativo mensual, estando estos datos publicados en los anuarios.

El cálculo de estas tasas es idéntico al planteado para las tasas anuales, sólo varía el período de referencia ("mes" en lugar de "año")

■ **Serie de tasas de Natalidad de Rosario de enero de 1972 a diciembre de 1986.**(Apéndice 5)

■ **Serie de tasas de Mortalidad de Rosario de enero de 1972 a diciembre de 1986.**(Apéndice 6)

III. METODOLOGIA*

Los procedimientos utilizados para analizar las series antes mencionadas son:

III.1. Series Anuales

En primer lugar se analiza la serie población por dos métodos:

III.1.1. Modelo Logístico Modificado (Apendice 7)

El uso del modelo logístico fue en otros tiempos popular entre los demógrafos. Sin embargo, aproximadamente en 1940 cayó en desuso siendo criticado por que las estimaciones estaban lejos de la realidad. Se consideraba también, que los modelos de población total eran de poco interés práctico, comenzando así el auge del método de componentes como el principal adoptado para proyecciones. Tiempo después resurge el uso de los modelos de tendencia simple (Raeside 1988), como un enfoque sencillo para obtener proyecciones de población global.

III.1.2. Modelos ARIMA de Box-Jenkins (Apendice 8)

Estos modelos aplicados a series económicas a partir de la incorporación de computadoras rápidas en la década del 70, son introducidos luego en la demografía básicamente para estudiar el comportamiento de series demográficas globales y realizar proyecciones a corto plazo.

Los modelos suponen que la serie observada Z_1, Z_2, \dots, Z_T es una realización de un proceso estocástico. En base a la información que brinda dicha serie se construye un modelo probabilístico que describe su comportamiento y mediante el cual se pueden realizar proyecciones.

Esta metodología también fue aplicada a las series anuales de natalidad, mortalidad y mortalidad infantil.

III.2 Series Mensuales

Las series mensuales de natalidad y mortalidad se analizaron por los siguientes métodos:

III.2.1. Modelos SARIMA de Box-Jenkins (Apendice 8)

Estos modelos son del mismo tipo de los modelos ARIMA pero contemplando el comportamiento estacional.

* Los paquetes estadísticos utilizados con los que se aplica esta metodología son: SYSTAT (versión 3.0), STATGRAPHICS (versión 2.6), STATISTIX (versión 2.0) y además se procesan datos en LOTUS-123.

III.2.2. Modelos de Descomposición (Apendice 9)

Son aquellos que piensan la serie como una descomposición de tendencia, estacionalidad y fluctuación aleatoria. Mediante métodos convenientes en este caso el de medias móviles se estima la tendencia y los factores estacionales, obteniendo el comportamiento entre años y el comportamiento estacional intra años.

IV. ANALISIS DE LOS RESULTADOS

Antes de avanzar en el análisis y comprensión de los resultados obtenidos se hace imperiosa la necesidad de recordar que el fin de este trabajo es realizar un estudio de algunas variables demográficas preponderantes para ver cuál fue su comportamiento a través del tiempo. Todas las series son analizadas con los modelos ARIMA, pero algunas de ellas son estudiadas en particular con otra metodología para profundizar en las interpretaciones. El análisis se divide en dos etapas: estudio de las series anuales y estudio de las series mensuales ya que el tratamiento para ambas es algo diferente.

V.1. SERIES ANUALES

IV.1.1. Serie de población

Los cálculos de población a nivel global ya sea para el futuro como para interpolaciones en el pasado resultan útiles para múltiples usos, por ejemplo para el cálculo de tasas demográficas y de indicadores económicos y sociales. Un procedimiento muy utilizado en demografía es el uso de la ecuación compensadora cuyo cálculo generalmente se apoya en los censos de población y en la información de las estadísticas de nacimientos, defunciones y migraciones. Una metodología de mucha aplicación es conocida como el "Método de componentes" que se basa en la ecuación compensadora y que además de considerar las variables demográficas, contempla la edad como variable interviniente. Pero en caso de no disponerse de estas informaciones, o que no se pretenda una información tan detallada lo más sencillo es utilizar procedimientos matemáticos en base a distintos tipos de funciones: lineales, geométricas, exponenciales o logísticas. El problema de todos estos procedimientos es que no se tiene idea del error que se pueda estar cometiendo en la proyección.

En este informe se presenta una modificación de la curva logística (el MLM) y se compara con los modelos ARIMA.

Aplicación del MLM

Al aplicar el MLM a la población de la República Argentina usando los censos nacionales de 1869, 1895, 1914, 1947, 1960 y 1970, Raeside (1988) determinó que a la Argentina le corresponde un MLM con tres parámetros

$$P_t = \frac{P_{\infty}}{1+(P_{\infty}/P_0-1)\exp(-ct^*)} \quad \text{IV.1.1.1.}$$

donde $t^* = \text{año} - 1869$.

Los parámetros estimados con sus desvíos standards son

PARAMETRO	ESTIMACION	ERROR STD.
P_{∞}	37.37	4.01
P_0	1.74	0.08
c	0.0357	0.0021

Suponiendo que el comportamiento de la población de la Ciudad de Rosario es similar al de la República Argentina, se aplica el MLM a la población de esta ciudad. Se consideran los censos municipales de 1900 y 1910 y los nacionales de 1914, 1947, 1960, 1970 y 1980. Se toma como año base el censo de 1900.

Los parámetros se obtienen mediante una estimación no lineal utilizando el paquete estadístico SYSTAT. El algoritmo empleado por este programa no requiere valores iniciales para los parámetros. Las estimaciones obtenidas para la Ciudad de Rosario son

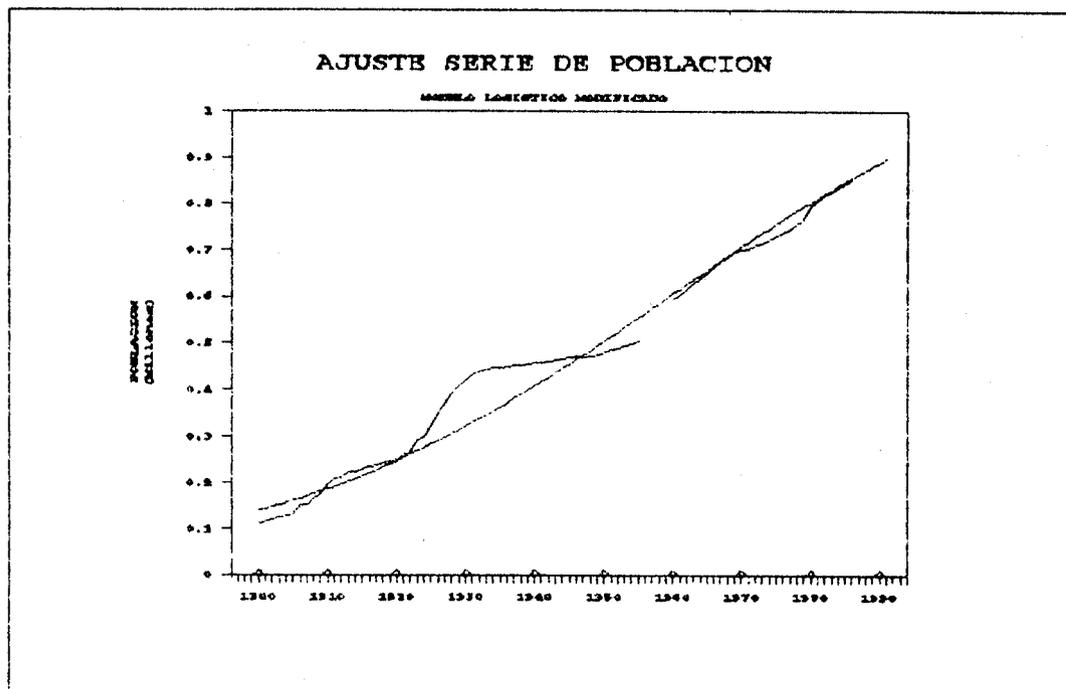
PARAMETRO	ESTIMACION	ERROR STD.
P_{∞}	1.2077	0.2072
P_0	0.1400	0.0107
c	0.0340	0.0030

Vale destacar lo cercano que se encuentra el parámetro $c=0.034$ para Rosario del obtenido para Argentina. Esto avala el supuesto hecho sobre que la población de Rosario sigue un comportamiento similar al de la República Argentina.

Se calculan las proyecciones para los años 1987/91

AÑO	PROYECCION (MLM)	POBLACION OBS.
1987	865 124	858 483
1988	873 406	866 292
1989	881 156	874 384
1990	889 594	
1991	897 496	

GRAFICO 1



FUENTE: Dirección de Estadística de la Municipalidad de Rosario

El GRAFICO 1 muestra la serie original donde se observan los valores faltantes para los años 1956/59 y la serie ajustada a través del MLM.

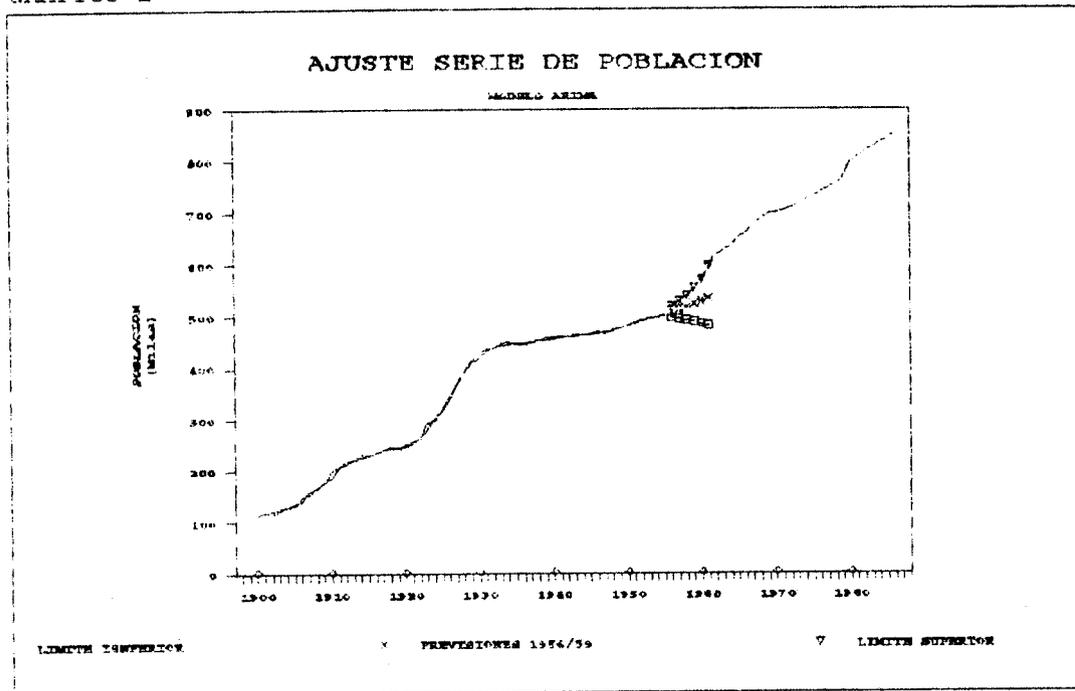
Una experiencia anterior (Cortés, 1951) muestra que para la Ciudad de Rosario la función logística daba buenos resultados aplicándola desde 1881 hasta 1947. Con los datos desde 1900 a 1986 se observa que el MLM no da estimaciones adecuadas para períodos intermedios, es decir, la función no ajusta bien los datos. Es destacable notar que a partir de 1925 y hasta 1930 se da un incremento abrupto en el desarrollo de la población (no hay duda de que este incremento se debe al saldo migratorio) mientras que éste mengua a partir de 1932 al agregarse solamente el crecimiento vegetativo.

Aplicación de los modelos ARIMA

La aplicación de los modelos ARIMA requiere de una serie continua de observaciones, es decir no puede haber valores faltantes para realizar el ajuste.

Dado que no se dispone de las observaciones para los años 1956/59 en una primera etapa se ajusta un modelo para la serie de población 1900/55 y se efectúan las proyecciones para los años faltantes. En una segunda etapa se ajusta la serie completa de 1900 a 1986 y se calculan las proyecciones para los años 1987/91

GRAFICO 2



Se utilizan las estimaciones de los datos faltantes por un modelo ARIMA[0, 2, (1,2,9)], Gráfico 2, para completar la serie. Estas estimaciones pueden considerarse buenas no obstante que el empalme con la serie en el año 1959 presenta un salto. Esto puede deberse a un cambio de metodología o definiciones que provocó un salto inusual. Se busca ajustar la serie original Z_t de 1900 a 1986. Previamente se realizan dos diferencias consecutivas de orden 1 transformándola en estacionaria.

Los correlogramas llevan a postular distintos modelos y luego de ser chequeados se elige al modelo ARIMA [0,2,(1-10)] como el mejor.

$$(1-B)^2 Z_t = (1-\theta_1 B - \theta_{10} B^{10}) u_t \quad \text{IV.1.1.3.}$$

El siguiente cuadro muestra las probabilidades asociadas a cada parámetro, los coeficientes y los errores std.

PARAMETRO	COEFICIENTE	ERROR STD.	PROB. ASOC.
MA 1	0.84597	0.048931	0.0000
MA 10	0.17614	0.52618	0.0008

La SC = 7 097 300 000 obtenida es menor que la de otros posibles modelos. Realizadas las pruebas de bondad de ajuste de este modelo resultaron satisfactorias.

Este modelo explica el valor de la población de un año en función de fluctuaciones aleatorias del año anterior y de 10 años antes. La dependencia con fluctuaciones aleatorias de 10 años antes puede resultar

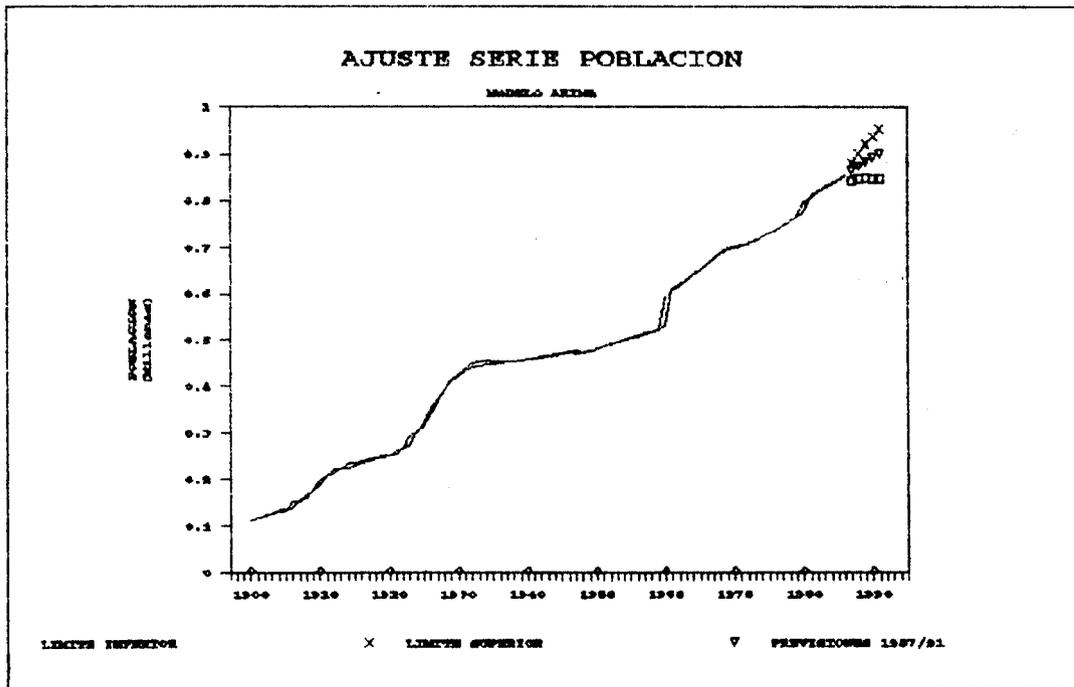
un poco sorprendente, pero si se tiene en cuenta que la mayoría de los censos se realizaron cada 10 años luego de los cuales se corregían los datos quiere decir que la serie fue modificada varias veces cada 10 años.

Es interesante comparar las proyecciones obtenidas con los valores de la población observados en Rosario para los mismos años. Los intervalos construidos mediante el modelo ARIMA cubren los valores observados. Tanto el método MLM como el modelo ARIMA sobreestiman los valores de la población observada en 1987/89. Si se tiene en cuenta el valor provisional de Censo realizado en mayo de 1991 que dió un valor de 899897 habitantes, cae dentro del intervalo de confianza y es muy cercano a las dos proyecciones puntuales.

AÑO	PROYECCION	INTERV. CONF. DEL 95%		POBLACION OBS.
1987	863 300	845 200	881 500	858 483
1988	874 200	846 500	901 900	866 292
1989	885 200	848 700	921 600	874 384
1990	892 300	847 200	937 400	
1991	900 300	846 500	954 000	

En el GRAFICO 3 se observa la serie de población original y la serie ajustada 1900/86 con las proyecciones e intervalos de confianza para los años 1987/91.

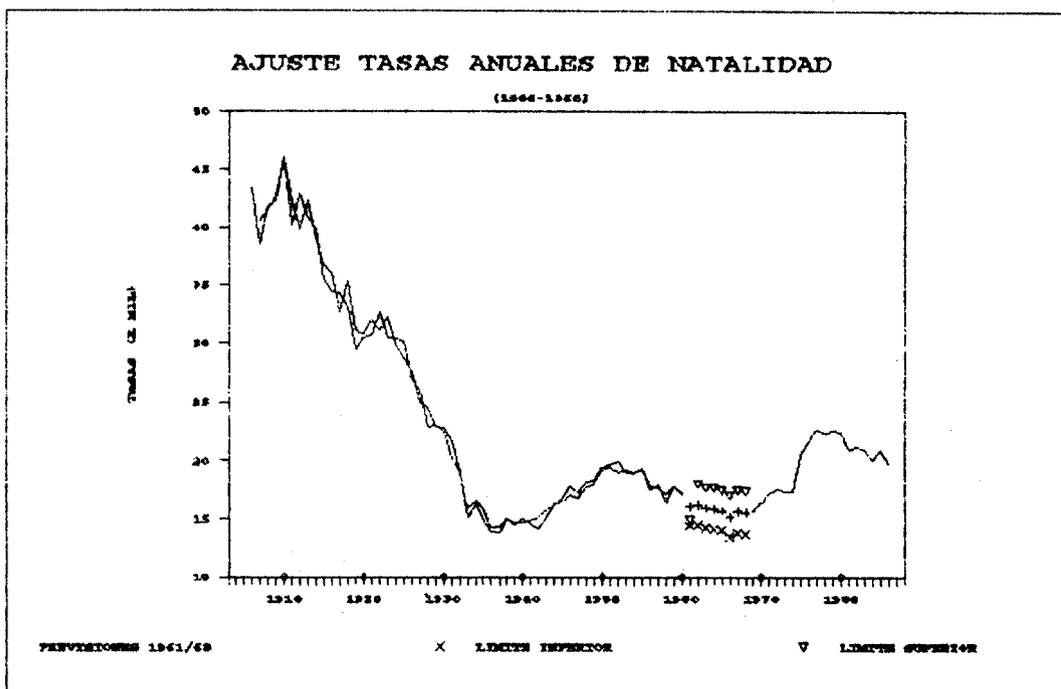
GRAFICO 3



IV.1.2. Serie de tasas de Natalidad

Debido a que no se disponen de las tasas de natalidad para la década del '60 se realiza el análisis de la serie en dos etapas. En la primera se considera la serie de 1906 a 1960, esto es posible ya que se cuenta con suficientes observaciones (55) para poder aplicar esta metodología. Se ajusta un modelo y se utilizan las previsiones para completar los valores faltantes. En la segunda etapa se estudia la serie completa 1906/86, se elige el mejor modelo y se calculan las previsiones para los años 1987/91.

GRAFICO 4



FUENTE: Dirección de Estadística de la Municipalidad de Rosario

La serie completa con las estimaciones de los datos faltantes según un modelo ARIMA[(3,5,11), 1, 0] dieron valores los cuales empalmaron con la serie muy bien. Gráfico 4, requiere una transformación logarítmica y una diferenciación de orden 1 para lograr la estacionariedad. El modelo postulado es un ARIMA [(3-5-11),1,0]

$$(1-\phi_3B^3-\phi_5B^5-\phi_{11}B^{11})(1-B) \ln Z_t = u_t \quad \text{IV.1.2.1.}$$

Sus parámetros son altamente significativos con una SC=0.19981 menor a la de los demás posibles modelos.

PARAMETRO	COEFICIENTE	ERROR STD.	PROB. ASOC.
AR 3	0.25452	0.089722	0.0046
AR 5	0.33972	0.087724	0.0001
AR 11	-0.33937	0.086402	0.0001

Realizadas las pruebas de bondad de ajuste del modelo resultaron satisfactorias. Este modelo indica que la observación de un año esta en función de valores observados tres, cinco y once años antes.

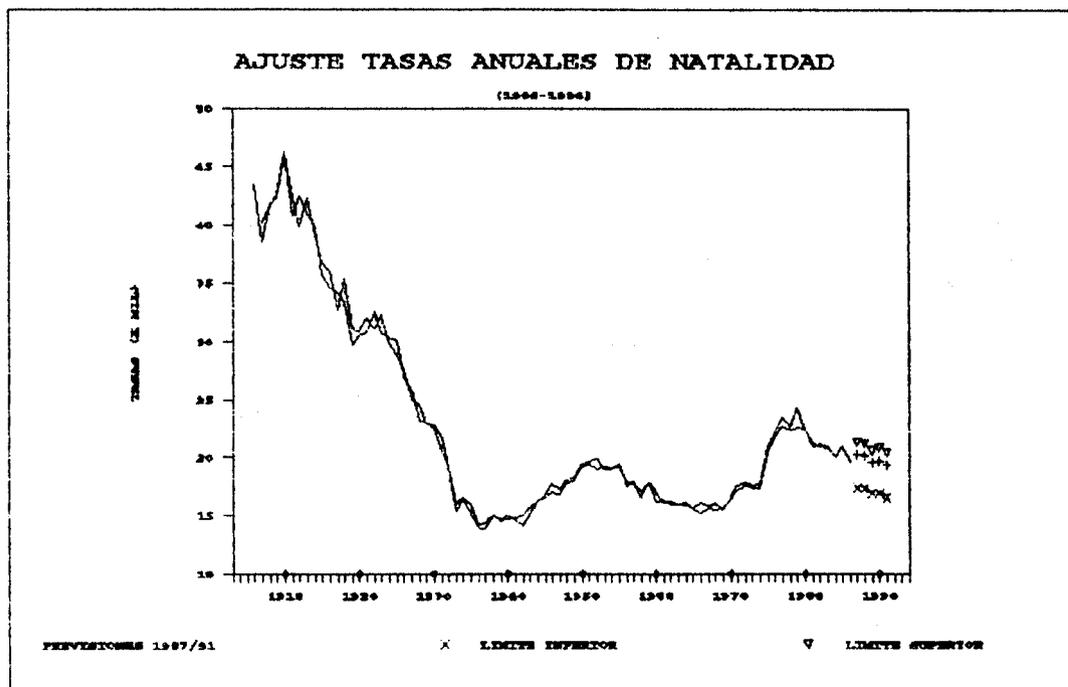
Se calculan las previsiones para los años 1987/91 y sus respectivos intervalos de confianza.

AÑO	PROYECCION	INTERV. CONF. DEL 95%	
1987	20.21	17.41	21.26
1988	20.09	17.30	21.30
1989	19.60	16.88	20.59
1990	19.71	16.91	20.80
1991	19.32	16.59	20.39

En el GRAFICO 5 se observa la serie de tasas de Natalidad original y la serie ajustada 1906/86 con las previsiones e intervalos de confianza para los años 1987/91.

Como se puede observar esta tasa tuvo un decrecimiento muy notorio desde principios de siglo, éste fenómeno que se verifica en muchos países del mundo se debe a una reducción voluntaria de los nacimientos. El descenso de la natalidad está muy vinculado con la industrialización y urbanización; es evidente que este hecho está relacionado con condiciones socio-culturales. La familia tipo fue cambiando desde principios de siglo hasta la fecha.

GRAFICO 5



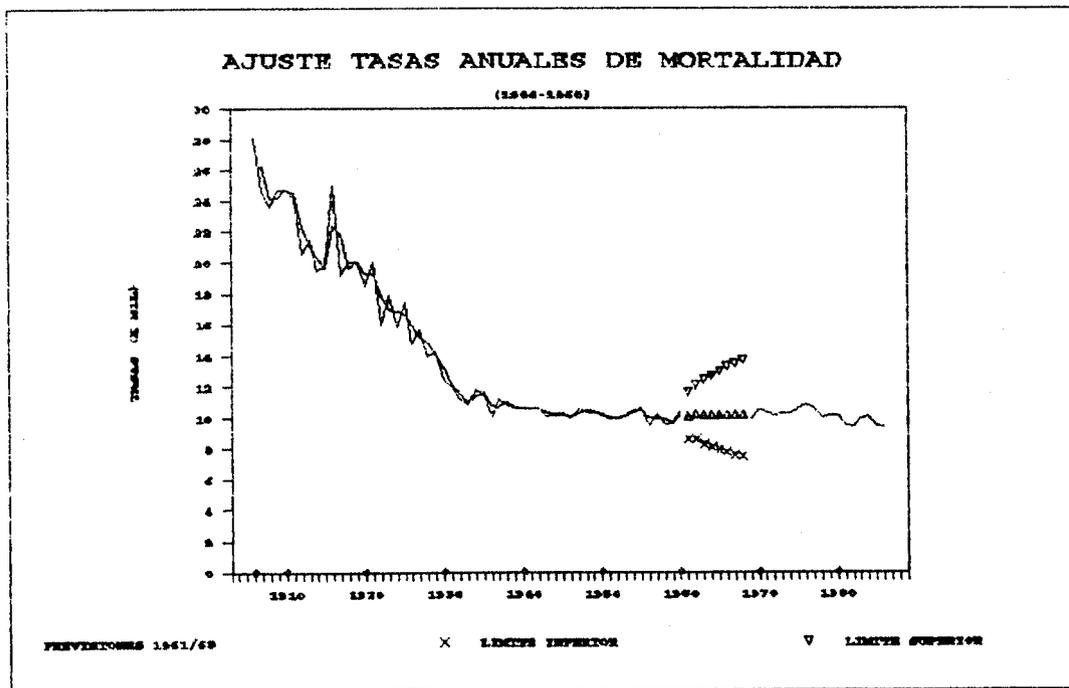
Nótese que para el período en que el incremento de la población fue abrupto (1926/32) la tasa de natalidad decreció a gran velocidad, esto es lógico pues aumentó el denominador en forma explosiva.

Otro hecho de gran notoriedad es que durante los años 1975/77 se produce un ascenso en la tasa, este hecho también se verifica a nivel nacional. Luego de este aumento y a partir de 1980 la tasa tiende lentamente a retomar las cifras que había dejado antes del incremento, aunque las proyecciones, a pesar de ser decrecientes, se mantienen en un nivel mayor que para 1974.

IV.1.3. Serie de tasas de Mortalidad

El análisis se realiza en dos etapas. En la primera se analiza la serie desde 1906 a 1960, se ajusta un modelo ARIMA y se hacen previsiones de manera de completar los valores faltantes. En la segunda se hace un ajuste de toda la serie y se calculan las previsiones para los años 1987/91.

GRAFICO 6



FUENTE: Dirección de Estadística de la Municipalidad de Rosario

La serie completa, estimados los valores faltantes según un modelo ARIMA(1, 1, 0) con un empalme muy satisfactoria con la serie observada, Gráfico 6, tomando los valores faltantes como los predichos con el modelo

anterior, requiere una transformación logarítmica y una diferenciación de orden 1 para que la misma sea estacionaria.

El comportamiento de esta serie es muy similar al comportamiento de la serie hasta 1960, observando la $facm$ y $facpm$ se elige nuevamente como el mejor modelo al siguiente: ARIMA (1,1,0), cuyo coeficiente es

PARAMETRO	COEFICIENTE	ERROR STD.	PROB. ASOC.
AR 1	-0.48792	0.09482	0.0000

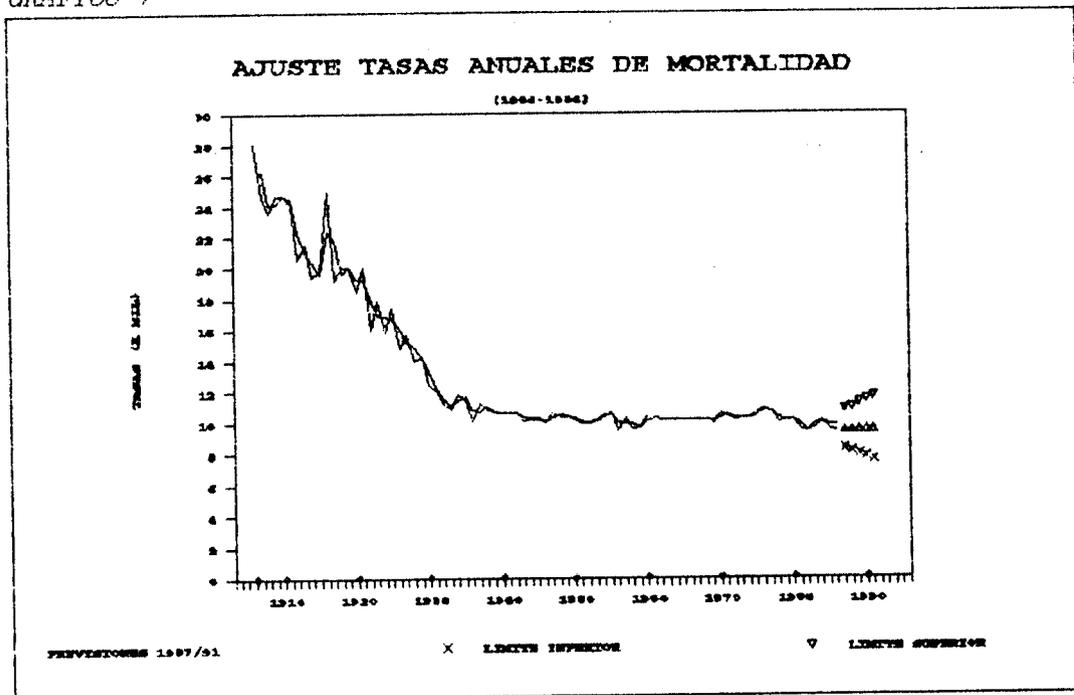
SC=0.34859

Las pruebas de bondad de ajuste del modelo resultaron satisfactorias.

AÑO	PROYECCION	INTERV. CONF. DEL 95%	
1987	9.45	8.30	10.76
1988	9.46	8.14	10.91
1989	9.48	7.92	11.26
1990	9.47	7.77	11.45
1991	9.47	7.61	11.68

La serie original de las tasas de Mortalidad, la serie ajustada 1906/86 y las previsiones con sus intervalos de confianza para los años 1987/91 pueden verse en el GRAFICO 7.

GRAFICO 7



A principios de siglo la mortalidad era elevada. El descenso de la misma, al menos hasta cierto nivel, depende esencialmente de medidas de salud pública: aplicación masiva de la medicina preventiva y curativa, así como de la mejora de condiciones sanitarias de la población. El

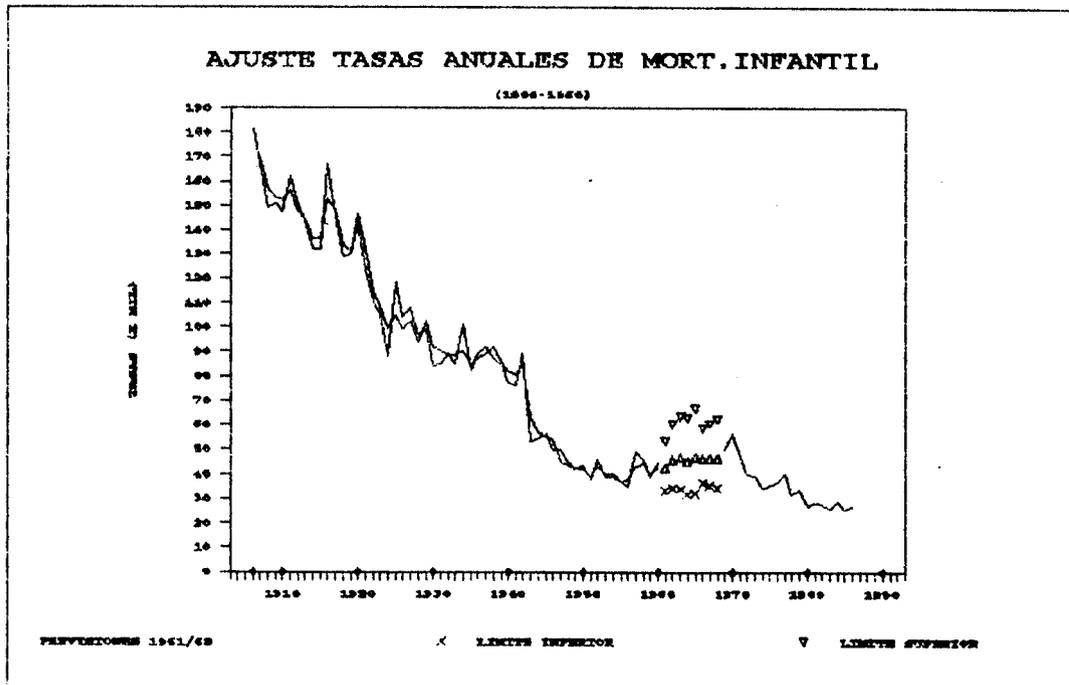
declive de la mortalidad depende entonces, en gran medida, del desarrollo económico y social como de programas médicos y tecnológicos. Si bien como dice O.Arguello⁽⁶⁾, esta tasa tomada aisladamente y en forma global no es un indicador contundente de desarrollo, el que la serie no presente picos tan marcados como al principio de siglo, estaría indicando un avance en las condiciones sociales de la población.

A partir de 1930, esta tasa parece estabilizarse en un nivel constante hasta la fecha y es el mismo nivel en que se mantienen las proyecciones.

IV.1.4. Serie de tasas de Mortalidad Infantil

El estudio de esta serie consta de tres etapas. En la primera, la serie de 1906/60 se ajusta con un modelo que sólo permite realizar previsiones 5 años hacia adelante, esto obliga en la segunda etapa a ajustar un modelo para la serie de 1906/65 que al calcular las previsiones posibilita completar finalmente la serie. En la tercera etapa se analiza la serie completa 1906/86.

GRAFICO 8



FUENTE: Dirección de Estadística de la Municipalidad de Rosario

La serie completa, estimada mediante modelos ARIMA[0, 1, (1,5)] con un buen empalme, Gráfico 8, presenta un decrecimiento brusco por lo que se la debió transformar con la función logaritmo y luego fue diferen-

ciarla para lograr la necesaria estacionariedad. Se prueban varios modelos siendo el mejor un ARIMA [0,1,(1-5)]

$$(1-B) \ln Z_t = (1-\theta_1 B - \theta_5 B^5) u_t \quad \text{IV.1.4.1.}$$

En este modelo la $SC=0.9899$ es menor que la de otros modelos probados.

PARAMETRO	COEFICIENTE	ERROR STD.	PROB. ASOC.
MA 1	0.4286	0.09377	0.0000
MA 5	-0.2985	0.09255	0.0013

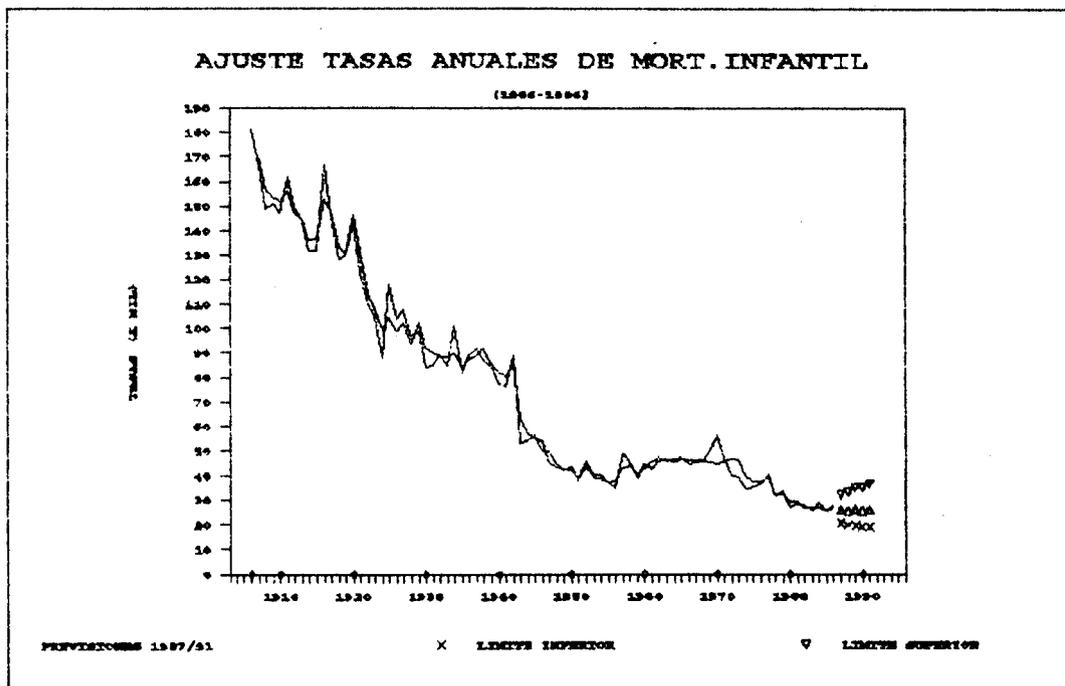
Los tests de bondad de ajuste de este modelo resultaron satisfactorios.

Se obtienen las previsiones para los años 1987/91.

AÑO	PROYECCION	INTERV. CONF. DEL 95%	
1987	25.89	20.63	32.10
1988	25.66	19.79	32.88
1989	26.20	19.61	34.57
1990	25.84	18.84	35.06
1991	26.20	18.63	36.45

El GRAFICO 9 muestra la serie de tasas de Mortalidad Infantil original y la serie ajustada 1906/86 con las previsiones e intervalos de confianza para los años 1987/91.

GRAFICO 9



La mortalidad infantil es uno de los indicadores socio-económicos que puede ayudar a conocer la realidad socio-económica de un país o región, tiene por lo tanto mucho sentido que sea estudiada aparte de la mortalidad general. La acentuada disminución de la tasa de mortalidad infantil hasta alrededor de 1950 aparentemente se ha detenido. Las primeras mejorías reflejan no sólo los factores que han beneficiado las condiciones de salud en general, sino también, la difusión de los cuidados pre y post natales, así como el aumento de hospitalización.

No obstante esta disminución es una serie que presenta muchos picos, sobre todo hasta la década del 40, luego si bien se estabiliza aun presenta algunos picos de menor cuantía, lo que indicaría una deficiencia socio-económica en este sector. El modelo también muestra esta recurrencia de valores altos cada cinco años, por que el valor de un año esta influenciado por fluctuaciones aleatorias de uno y cinco años anteriores.

IV.2. SERIES MENSUALES

Entre otras cosas, las series anuales estudiadas dan una buena idea de la tendencia que estas series tienen a través del tiempo. Para profundizar el análisis, se quiere ver cual es la estacionalidad que ellas presentan. Es por esta razón que se estudian las series mensuales. Interesa particularmente, saber qué ocurre con la estacionalidad en las series de tasas de natalidad y tasas de mortalidad. Estas son estudiadas bajo dos enfoques: el primero es el enfoque que Box y Jenkins dan al estudio de las series y que ya se justificó ampliamente en párrafos anteriores; el segundo es tomando a la serie como una combinación lineal de tres factores que influyen en la misma: tendencia, estacionalidad y un error aleatorio, en este caso interesa la interpretación de la componente estacional.

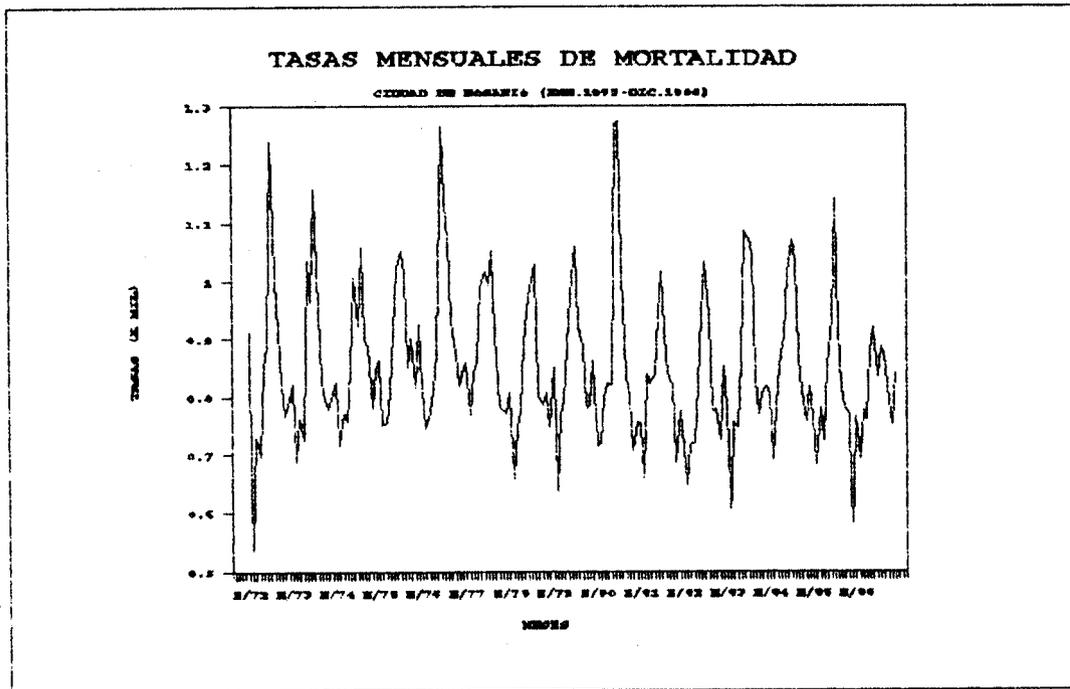
IV.2.1. Serie de tasas de mortalidad (GRAFICO 10)

Modelos ARIMA

Esta serie presenta una fuerte estacionalidad, por lo tanto para destruirla se hace una diferencia de orden 12 y debido a que no presenta tendencia, la serie diferenciada resulta estacionaria. Luego de visualizar la $facm$ y $facpm$ de esta serie se identifica como el mejor modelo al siguiente: SARIMA $[1,0,(1-7)][0,1,1]_{12}$ es decir

$$(1-B^{12})(1-\phi_1 B) Z_t = (1-\alpha_1 B^{12})(1-\theta_1 B - \theta_7 B^7) u_t \quad \text{IV.2.1.1.}$$

GRAFICO 10



FUENTE: Dirección de Estadística de la Municipalidad de Rosario.

En este modelo los parámetros que resultan significativos son

PARAMETRO	COEFICIENTE	ERROR STD.	PROB. ASOC.
AR 1	0.81658	0.08257	0.0000
MA 1	0.71133	0.09373	0.0000
MA 7	-0.18703	0.05704	0.0010
SMA 1	0.88140	0.01926	0.0000

La suma de cuadrados de este modelo es (SC) = 0.64297. Las pruebas de bondad de ajuste de este modelo resultaron satisfactorias.

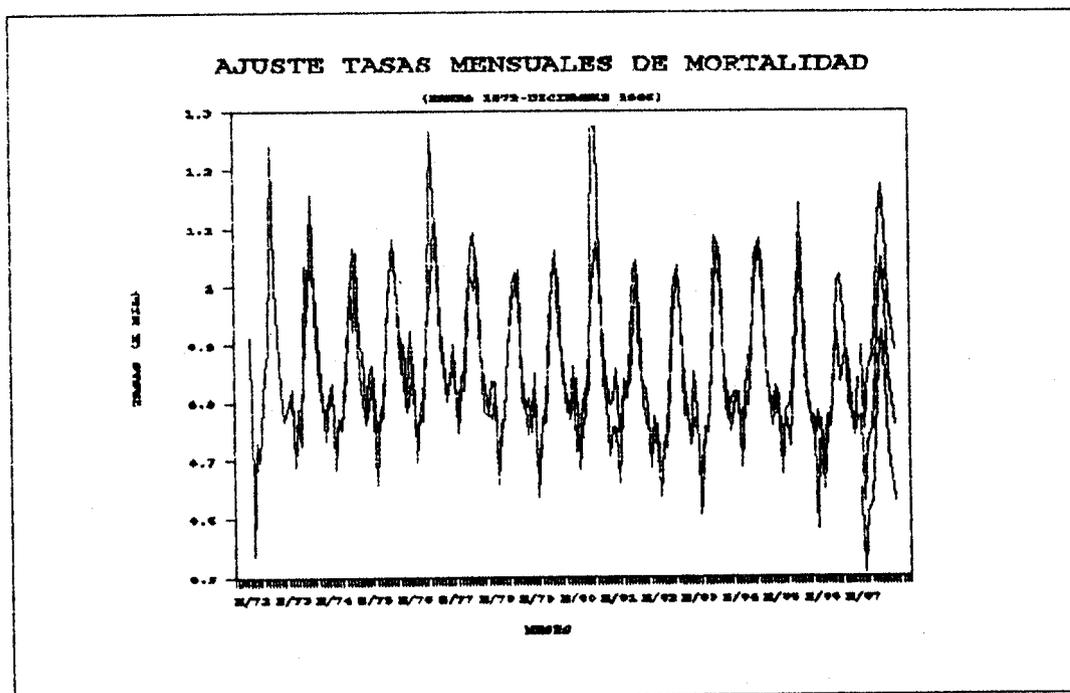
Este modelo indica que la observación de un período esta en relación con la observación del mes anterior y con las fluctuaciones aleatorias de 1, 7 y 12 meses anteriores, siendo el coeficiente más grande el de 12 meses lo que indica un fuerte patron estacional.

Con este modelo se calculan 12 previsiones con un intervalo de confianza del 95%. Estas son para el año 1987

MES	PROYECCION	INTERV. CONF. DEL 95%	
Enero	0.7732	0.6505	0.8959
Febrero	0.6293	0.5059	0.7527
Marzo	0.7310	0.6071	0.8548
Abril	0.7535	0.6293	0.8776
Mayo	0.8604	0.7361	0.9850
Junio	0.9900	0.8657	1.1150
Julio	1.0470	0.9225	1.1720
Agosto	0.9700	0.8422	1.0970
Setiembre	0.8485	0.7193	0.9780
Octubre	0.8006	0.6701	0.9311
Noviembre	0.7590	0.6276	0.8903
Diciembre	0.7885	0.6566	0.9203

El GRAFICO 11 muestra la serie ajustada por este modelo y las previsiones con sus correspondientes intervalos de confianza.

GRAFICO 11

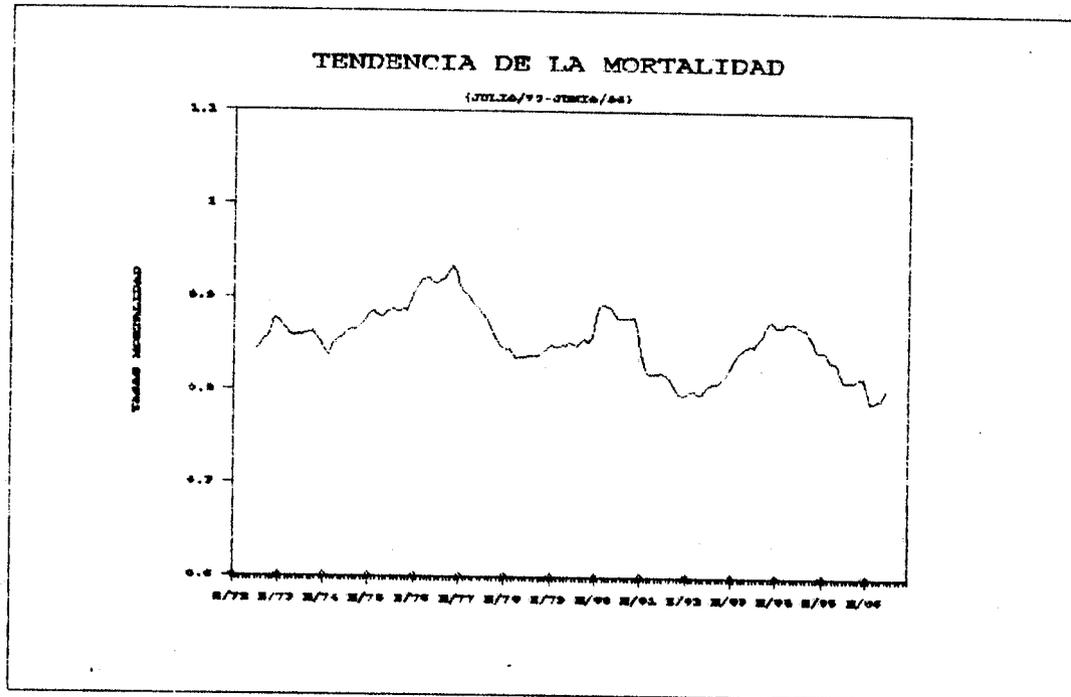


Método de desestacionalización

Este segundo procedimiento estima las componentes de la serie, es decir, estima T_t , S_t y e_t para cada mes t .

El GRAFICO 12 muestra la tendencia estimada, se observa que no hay tendencia evidente, es decir, la estimación de T_t fluctúa alrededor de un valor constante a través del tiempo.

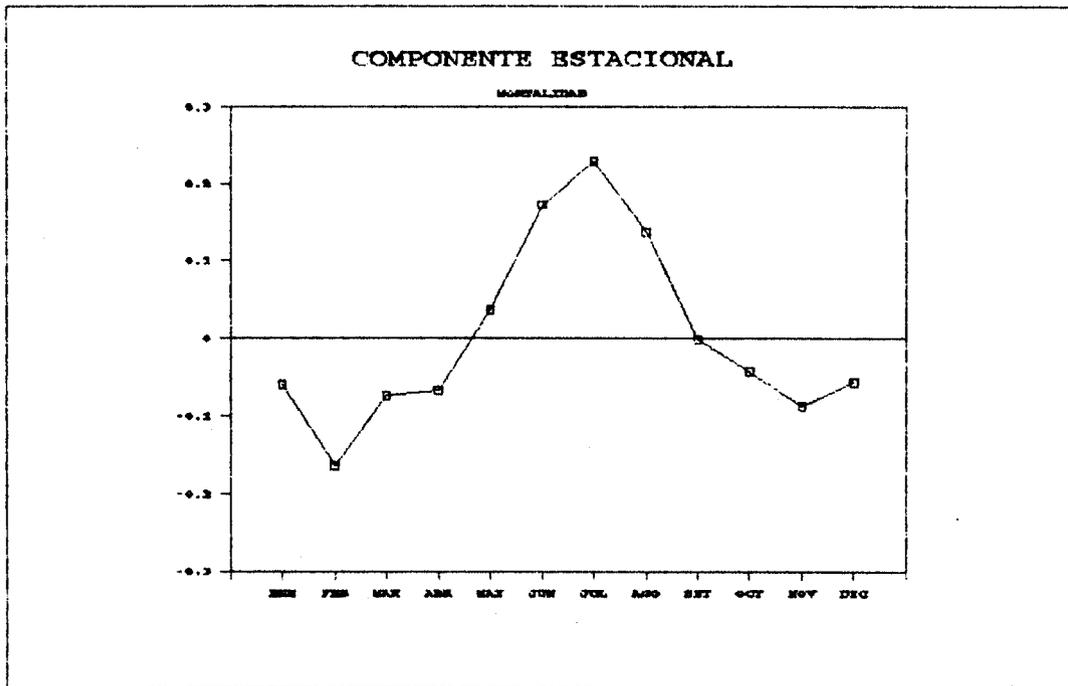
GRAFICO 12



El GRAFICO 13 de la componente estacional muestra que el mayor pico de mortalidad se da en el mes de julio; en los meses de invierno esta componente es siempre positiva, mientras que disminuye resultando negativa en el resto de los meses. Otro pequeño pico en que la componente estacional aumenta (aunque es negativa) se produce en los meses de verano, esto concuerda con los patrones dados para esta tasa. (Rosenberg, 1966).

Es necesario aclarar que la tasa de mortalidad tal como aquí se la estudia, está fuertemente influenciada por la mortalidad infantil. El estudio sería más riguroso si se pudiera extraer la influencia de la mortalidad infantil, lo cual no es posible por no contar con los registros necesarios.

GRAFICO 13



IV.2.2. Serie de tasas de natalidad (Gráfico 14)

Modelos ARIMA

Esta serie presenta un inusual incremento en la tasa para el año 1975 (esto se aprecia también en la serie anual, ver GRAFICO 5) lo cual da idea de una tendencia creciente, por esto se realiza una diferencia de orden 1; además debido a la estacionalidad presente en la serie se realiza una diferencia de orden 12. De esta manera la serie diferenciada resulta estacionaria pudiéndose asignar un modelo ARIMA. El modelo que resulta más adecuado es:

$$SARIMA (0,1,1)(0,1,1)_{12}$$

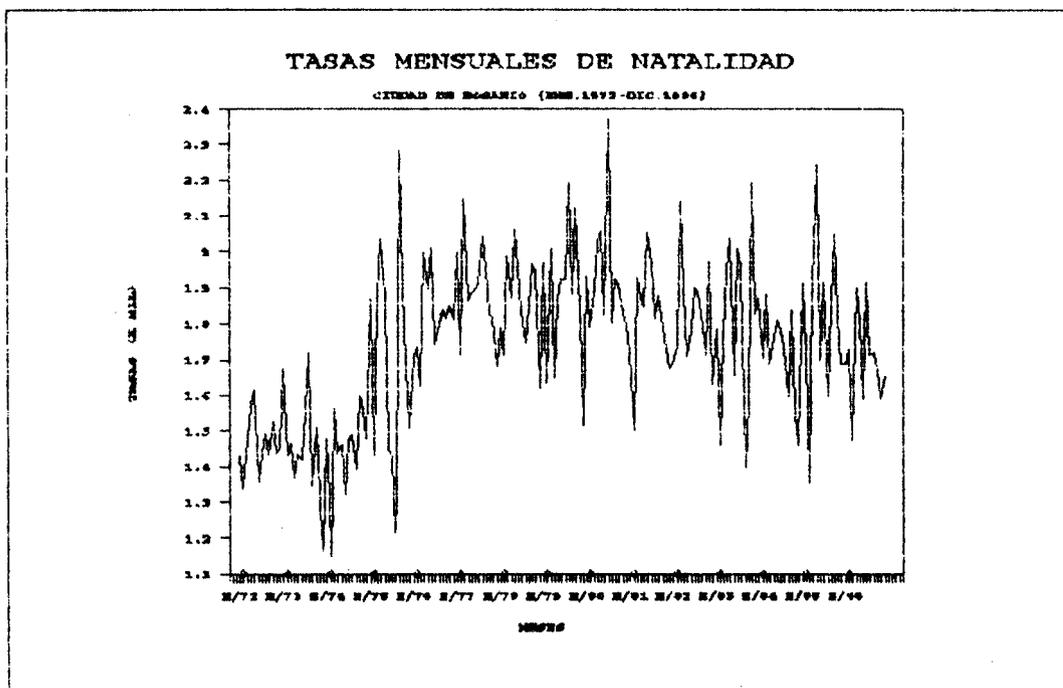
$$(1-B)(1-B^{12}) Z_t = (1-Q_{12}B^{12})(1-\theta_1B) u_t \quad \text{IV.2.2.1.}$$

donde los parámetros tienen los siguientes coeficientes, errores std. y probabilidades asociadas

PARAMETRO	COEFICIENTE	ERROR STD.	PROB. ASOC.
MA 1	0.87256	0.036073	0.0000
SMA 1	0.90359	0.025796	0.0000

La suma de cuadrados de este modelo es SC=5.1607. Las pruebas de bondad de ajuste de este modelo resultaron satisfactorias.

GRAFICO 14



FUENTE: Dirección de Estadística de la Municipalidad de Rosario

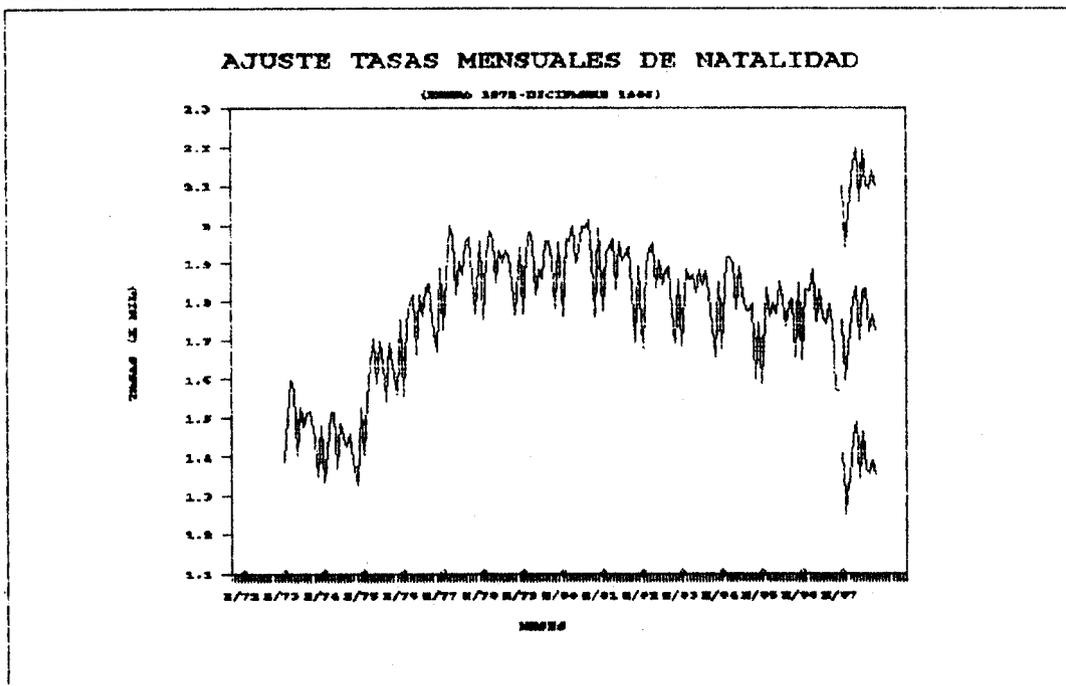
Este modelo también presenta un fuerte patrón estacional.

Con este modelo se calculan 12 previsiones con un intervalo de confianza del 95%, éstas son para el año 1987

MES	PROYECCION	95% INTERV. CONF.	
Enero	1.753	1.407	2.100
Febrero	1.596	1.246	1.945
Marzo	1.726	1.374	2.078
Abril	1.804	1.449	2.159
Mayo	1.840	1.489	2.197
Junio	1.700	1.340	2.061
Julio	1.828	1.465	2.191
Agosto	1.737	1.369	2.100
Setiembre	1.723	1.355	2.092
Octubre	1.769	1.398	2.141
Noviembre	1.726	1.353	2.100
Diciembre	1.595	1.218	1.971

El GRAFICO 15 muestra la serie ajustada a través de este modelo y las previsiones con sus correspondientes intervalos de confianza.

GRAFICO 15



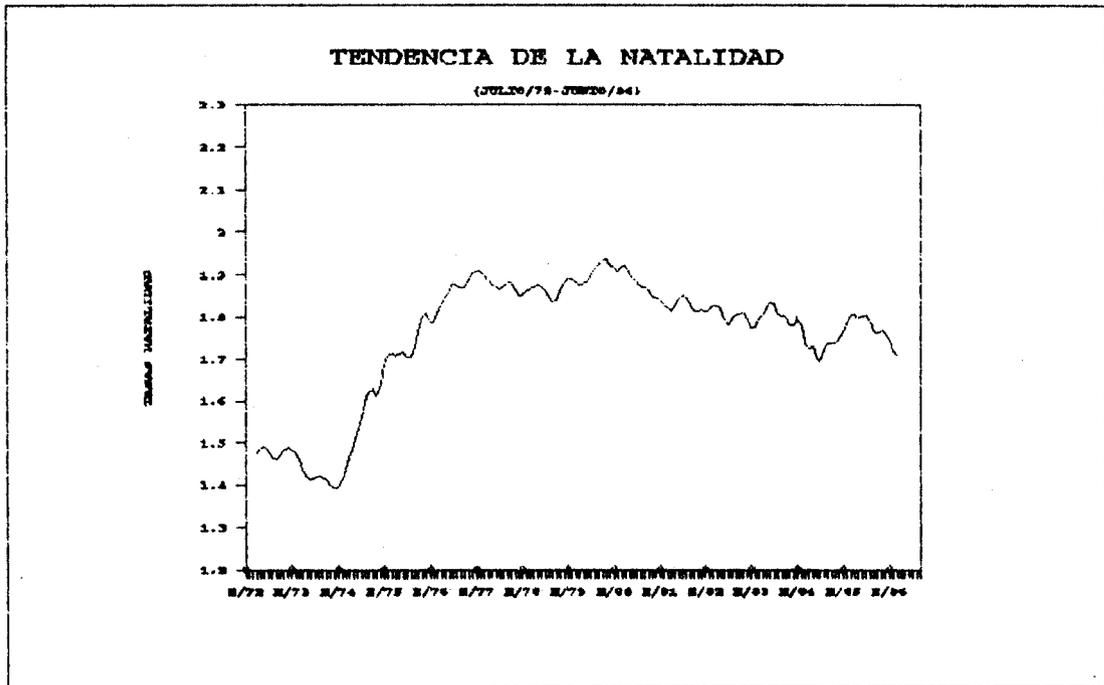
Método de desestacionalización

Se estiman las componentes T_t , S_t y e_t de la serie Z_t para cada mes t .

El GRAFICO 16 revela la tendencia creciente estimada de la que se habló anteriormente con un salto abrupto durante el año 1975 y por la cual es necesaria una diferencia de orden uno en la parte no estacional del modelo ARIMA.

La componente estacional reflejada en el GRAFICO 17 no presenta una interpretación muy clara, aunque se puede ver que existen dos picos negativos en los meses de diciembre y febrero donde las tasas de natalidad sufren un fuerte decrecimiento. El máximo se da en el mes de mayo. Esto no concuerda con el patrón observado a nivel mundial donde el pico máximo se exhibe durante los meses de primavera y otro pico menos importante en los meses de otoño, los mínimos son en verano e invierno (Sauvy, 1960 y Rosenberg, 1966).

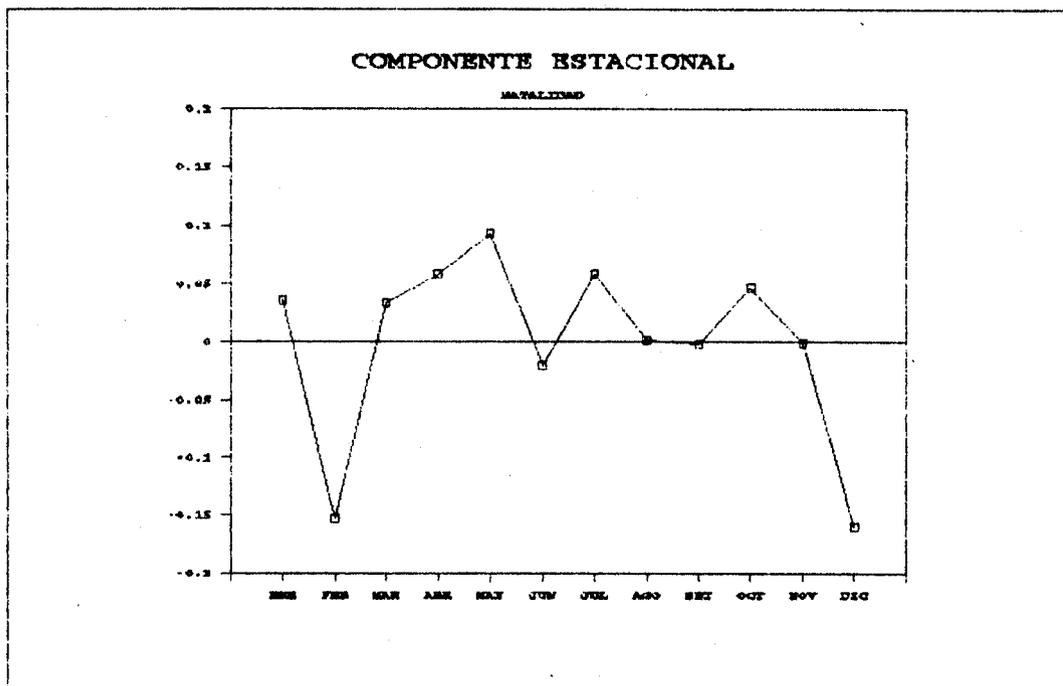
GRAFICO 16



Es importante notar que en nuestra ciudad, como en el resto del país, es muy frecuente que los recién nacidos no sean anotados dentro de las fechas establecidas en los registros civiles (lo cual depende en gran medida, de la situación socio-económica por la que estén atravesando los padres). Este hecho provoca una distorsión en las tasas de natalidad que influye en el patron estacional debido a los malos registros. Este inconveniente no se presenta en las tasas de mortalidad ya que las defunciones deben ser declaradas para poder realizar las inhumaciones correspondientes.

Resultaría interesante comprobar la bondad de los registros analizando el comportamiento del índice de masculinidad. Esto no tiene sentido hacerlo a través de la metodología de series de tiempo ya que es irrelevante ajustar una serie a dicho índice y menos aún hacer previsiones del mismo que se sabe debe variar alrededor de 105. Este analisis no es posible realizarlo debido a que la Dirección de Estadística de la Municipalidad de Rosario no lo publica y además no cuenta con los registros completos para calcularlo.

GRAFICO 17



Tanto en el estudio de esta tasa como en la de mortalidad no se realizaron las correcciones por efecto calendario⁽⁷⁾, por lo cual ellas estan debilmente influenciadas por la extensión de cada mes.

V. CONSIDERACIONES FINALES

De los análisis realizados podemos concluir las siguientes consideraciones generales para cada variable estudiada:

- **Población:** el ajuste realizado a través de los modelos ARIMA dió resultados más alentadores que la logística modificada en cuanto a la bondad de ajuste. Con respecto a las proyecciones si bien ambos métodos brindaron buenos resultados, el modelo ARIMA[0, 2, (1,10)] proporciona intervalos de confianza muy aceptables. Si se considera la cifra provisional brindada del censo de mayo de 1991 de 899897 fue cubierta por el intervalo de confianza del modelo ARIMA (846500-954000) y los valores de las estimaciones puntuales de este método 900300 y del MLM 897496 resultaron muy aproximados.

- **Tasa anual de natalidad:** Esta tasa muestra un marcado decrecimiento desde principio de siglo hasta aproximadamente la década del 30, con un pequeño repunte en los 40, y un ascenso muy notorio en los 70 y luego en los 80 un crecimiento leve, este comportamiento es muy parecido al de la tasa de natalidad del total del país. En cuanto al ajuste del modelo ARIMA[(3,5,11), 1, 0] resultó muy satisfactorio, con buenas estimaciones del período faltante (1961/68), las previsiones se acompañan con intervalos de confianza aceptablemente precisos.

- **Tasa anual de mortalidad:** Esta tasa también muestra un marcado decrecimiento con muchos picos que evidencian una inestabilidad en la tasa desde principio de siglo hasta la década del 30, a partir de la cual se estabiliza, la inestabilidad del primer período puede deberse a deficientes condiciones socio-sanitarias en dicho período. El ajuste del modelo ARIMA(1, 1, 0) resultó ampliamente satisfactorio, pudiéndose realizar buenas estimaciones para el período faltante (1961/68), con un empalme muy bueno a la serie observada. Las previsiones puntuales son estables con intervalos de confianza que pueden considerarse un poco amplios, esto puede deberse a las fluctuaciones sufridas por la serie en el primer período. Para mejorar esta situación podrían usarse métodos robustos.

- **Tasa anual de mortalidad infantil:** Presenta un comportamiento decreciente con picos y altibajos que se presentan hasta la década del 70 lo que estaría indicando una deficiencia en las condiciones socio-económicas y atención sanitaria a este sector de la población. El ajuste del modelo ARIMA[0, 1, (1,5)] brindó buenos resultados con buena estimación de datos faltantes que empalman bien la serie observada. Las previsiones puntuales son estables con intervalos de confianza un poco amplios, lo cual puede deberse como en la tasa anterior a la gran fluctuación de la serie y podría tener la misma solución.

- **Tasa mensual de mortalidad:** Como era de esperar esta serie presenta un fuerte comportamiento estacional sin evidenciar tendencia y puede explicarse a través de un modelo SARIMA[1, 0, (1,7)][0, 1, 1]₁₂, con proyecciones e intervalos de confianza bastante precisos. Al estudiar el patrón estacional por el método de descomposición, este mostró un pico

en el mes de julio, presentandose las componenets positivas en los meses de invierno.

- *Tasa mensual de natalidad:* Esta serie presenta un inusual crecimiento a partir de 1975 (hecho observado tambien en los datos a nivel nacional), esta tendencia se visualiza mejor al aplicar el método de descomposición y aislar la tendencia. El modelo que ajustó bien la serie fue un SARIMA(0, 1, 1)(0, 1, 1)₁₂, con proyecciones e intervalos de confianza aceptables. El patron estacional no presenta un comportamiento muy acorde con los patrones mundiales, esto como se dijo podria deberse a que los registros de nacimientos sobre todo en las clases de menores ingresos suele dilatarse por más de un mes.

Como se dijo anteriormente las técnicas aplicadas en este trabajo sólo brindan, el comportamiento a través del tiempo de variable demográficas globales y permiten realizar proyecciones a corto plazo. Si se dispone de registros confiables de variables demográficas en el tiempo de por lo menos 50 años de variables anuales o 60 meses de datos mensuales, estas metodologías permiten una descripción de su comportamiento y realizar proyecciones para las cuales se puede medir la probabilidad del error que se comete en la estimación.

Este tipo de estudio puede ser completado por otros donde se puedan desagregar las variables demográficas por ejemplo por sexo, edad, causa de muerte, etc. . Las proyecciones encontradas por este método pueden ser empleadas como datos en otras investigaciones en las que se apliquen técnicas distintas.

Para finalizar cabria agregar lo importante que es para una comunidad (pueblo, ciudad, provincia, país), contar con buenos registros en este caso demográficos, por que ellos permiten conocer la dinámica de su comportamiento histórico, permitiendo a su vez la realización de previsiones sustentadas por métodos que permiten medir la probabilidad de error de dichas proyecciones.

Trabajos de este tipo tratan de analizar, interpretar y comprender algunos aspectos demográficos de una zona. Uno de sus objetivos es que sean una pequeña contribución para conocer mejor una región para ayudar a sus gobernantes a elaborar y ajustar programas de desarrollo económico y social.

VII. BIBLIOGRAFIA

- (1) Land K., "Methods for National Population Forecasts: A review", *JASA* 1986 Vol. 81 N°396.
- (2) Pujol José M. "Proyecciones de población a nivel nacional. Método de los componentes" en "Métodos para Proyecciones Subnacionales de Población", DANE-CELADA 1989.
- (3) Raeside R. (1988), "The Use of Sigmoids in Modelling and Forecasting Human Populations," *Journal Royal Statistical Society*, 151. Part 3, 499-513.
- (4) Box G.E.P., Jenkins G.M. (1976), *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day, Inc.
- (5) CELADE, "Métodos para proyecciones demográficas", 1984, en presentación.
- (6) Arguello Omar, "Estilos de desarrollo y dinámica demográfica", CELADE 1989.
- (7) Bell W.R. and Hillmer S.C., "Modelling Time Series with Calendar Variation", *JASA*, Vol 78 n°383, 1983.
- (8) Cortés F. (1951), "Experiencia Estadística sobre el Desarrollo de la Población de la Ciudad de Rosario en su Aspecto Cuantitativo". Ateneo de Economía. Rosario.
- (9) Gérard H., Wunsch G. (1973), *Demografía*. Madrid: Editorial Pirámide.
- (10) Leach D. (1981), "Re-evaluation of the Logistic Curve for Human Population," *Journal Royal Statistical Society*, 144, 94-103.
- (11) Meade N. (1988), "A Modified Logistic Model Applied to Human Populations," *Journal Royal Statistical Society*, 151, Part 3, 491-498.
- (12) Nelson C.R. (1973), *Applied Time Series Analysis for Managerial Forecasting*. San Francisco: Holden-Day, Inc.
- (13) Pressat R. (1983), *El Análisis Demográfico. Métodos, resultados, aplicaciones*. México: La Impresora Azteca.

- (14) Rosenberg H.M. (1966), "Recent Developments in Seasonally Adjusting in Vital Statistics," *Demography*, 3. 305-318.
- (15) Sauvy A. (1960), *La Población. Sus movimientos, sus leyes.* Buenos Aires: UDEBA.
- (16) Spiegelman M. (1972), *Introducción a la Demografía.* México: Gráfica Panamericana.
- (17) Vandaele W. (1983), *Applied Time Series and Box-Jenkins Models.* Orlando: Academic Press, Inc.
- (18) Uriel E. (1985), *Análisis de Series Temporales, Modelos ARIMA.* Paraninfo.

APENDICES

Apendice 1. SERIE DE POBLACION DE LA CIUDAD DE ROSARIO (1900-1986)

<i>Año</i>	<i>Población</i>	<i>Año</i>	<i>Población</i>	<i>Año</i>	<i>Población</i>
1900	113168	1929	409022	1958	517100*
1901	117410	1930	420631	1959	522600*
1902	121139	1931	432295	1960	594063
1903	127040	1932	441305	1961	604853
1904	129117	1933	443526	1962	615814
1905	131412	1934	445638	1963	626971
1906	150845	1935	447607	1964	638343
1907	152923	1936	449427	1965	649900
1908	167516	1937	450830	1966	661683
1909	176071	1938	452695	1967	673671
1910	198781	1939	454436	1968	685883
1911	208980	1940	456491	1969	698317
1912	211607	1941	458338	1970	699943
1913	223611	1942	459949	1971	702600
1914	225101	1943	462226	1972	709859
1915	230335	1944	465007	1973	714881
1916	236851	1945	467945	1974	723238
1917	240041	1946	471249	1975	730419
1918	243709	1947	467937	1976	738382
1919	246641	1948	473087	1977	747342
1920	250730	1949	476672	1978	756575
1921	260568	1950	480980	1979	765861
1922	265015	1951	485569	1980	797755
1923	294307	1952	489945	1981	807038
1924	298872	1953	494391	1982	816664
1925	320842	1954	498612	1983	825788
1926	348261	1955	502942	1984	834019
1927	371506	1956	507200*	1985	843724
1928	391957	1957	512000*	1986	852535

FUENTE: Dirección de Estadística de la Municipalidad de Rosario.

* Valores estimados a través del modelo ARIMA [0,2,(1-2-9)].

Apendice 2. TASAS DE NATALIDAD ANUALES DE LA CIUDAD DE ROSARIO (1906-1986)

Año	Tasas	Año	Tasas	Año	Tasas
1906	43.4	1933	15.9	1960	17.1
1907	38.5	1934	16.5	1961	16.11*
1908	41.8	1935	15.9	1962	16.19*
1909	42.3	1936	14.2	1963	15.92*
1910	46.0	1937	14.4	1964	15.83*
1911	42.2	1938	15.0	1965	15.67*
1912	39.8	1939	14.5	1966	15.14*
1913	42.3	1940	15.1	1967	15.57*
1914	39.0	1941	14.6	1968	15.45*
1915	36.7	1942	14.1	1969	15.6
1916	36.0	1943	15.1	1970	16.4
1917	32.7	1944	16.3	1971	17.2
1918	35.3	1945	16.5	1972	17.6
1919	31.1	1946	17.0	1973	17.4
1920	30.9	1947	16.8	1974	17.4
1921	32.0	1948	17.7	1975	20.5
1922	31.1	1949	18.0	1976	21.7
1923	32.3	1950	19.2	1977	22.7
1924	29.8	1951	19.4	1978	22.3
1925	28.8	1952	19.0	1979	22.5
1926	27.6	1953	19.2	1980	22.4
1927	25.0	1954	19.0	1981	21.0
1928	24.5	1955	19.3	1982	21.2
1929	22.9	1956	17.8	1983	21.0
1930	22.8	1957	17.7	1984	20.0
1931	21.7	1958	17.1	1985	21.0
1932	19.1	1959	17.8	1986	19.8

FUENTE: Dirección de Estadística de la Municipalidad de Rosario.

* Valores estimados a través del modelo ARIMA [(3-5-11),1,0].

Apendice 3. TASAS DE MORTALIDAD ANUALES DE LA CIUDAD DE ROSARIO (1906-1986)

Año	Tasa	Año	Tasa	Año	Tasa
1906	28.1	1933	10.9	1960	10.4
1907	24.6	1934	11.8	1961	10.08*
1908	23.6	1935	11.5	1962	10.27*
1909	24.7	1936	10.1	1963	10.18*
1910	24.7	1937	11.2	1964	10.23*
1911	24.2	1938	10.9	1965	10.20*
1912	20.6	1939	10.6	1966	10.22*
1913	21.5	1940	10.6	1967	10.21*
1914	19.4	1941	10.6	1968	10.21*
1915	19.8	1942	10.6	1969	9.9
1916	25.0	1943	10.1	1970	10.6
1917	19.2	1944	10.3	1971	10.4
1918	20.0	1945	10.2	1972	10.2
1919	20.0	1946	10.0	1973	10.3
1920	18.5	1947	10.6	1974	10.3
1921	20.0	1948	10.4	1975	10.6
1922	16.0	1949	10.4	1976	10.9
1923	17.9	1950	10.2	1977	10.6
1924	15.8	1951	9.9	1978	10.0
1925	17.5	1952	10.0	1979	10.2
1926	14.7	1953	10.2	1980	10.2
1927	15.7	1954	10.5	1981	9.5
1928	14.0	1955	10.6	1982	9.4
1929	14.3	1956	9.5	1983	9.9
1930	12.4	1957	10.3	1984	10.1
1931	12.0	1958	9.6	1985	9.5
1932	11.2	1959	9.7	1986	9.4

FUENTE: Dirección de Estadística de la Municipalidad de Rosario.

* Valores estimados a través del modelo ARIMA (0,1,1).

**Apéndice 4. TASAS DE MORTALIDAD INFANTIL ANUALES DE LA CIUDAD DE ROSARIO
(1906-1986)**

<i>Año</i>	<i>Tasas</i>	<i>Año</i>	<i>Tasas</i>	<i>Año</i>	<i>Tasas</i>
1906	181	1933	85	1960	45
1907	167	1934	101	1961	42.72*
1908	149	1935	82	1962	46.14*
1909	151	1936	89	1963	47.17*
1910	147	1937	92	1964	45.50*
1911	162	1938	87	1965	47.03*
1912	150	1939	84	1966	46.34*
1913	144	1940	77	1967	46.30*
1914	132	1941	76	1968	46.34*
1915	132	1942	89	1969	50
1916	167	1943	53	1970	56
1917	146	1944	55	1971	47
1918	128	1945	56	1972	39.8
1919	130	1946	50	1973	39.4
1920	143	1947	50	1974	34.3
1921	123	1948	45	1975	35.5
1922	110	1949	42	1976	36.8
1923	105	1950	44	1977	40.5
1924	98	1951	38	1978	31.8
1925	118	1952	46	1979	33.7
1926	104	1953	39	1980	27.3
1927	108	1954	39	1981	28.4
1928	97	1955	37	1982	27.5
1929	99	1956	35	1983	25.7
1930	84	1957	49	1984	28.6
1931	85	1958	46	1985	25.3
1932	89	1959	39	1986	26.9

FUENTE: Dirección de Estadística de la Municipalidad de Rosario.

** Valores estimados a través del modelo ARIMA [0,1,(1-5)].*

Apendice 5. TASAS DE NATALIDAD MENSUALES DE LA CIUDAD DE ROSARIO
(ENE.1972-DIC.1986)

Año	Mes	Tasas	Año	Mes	Tasas
1972	enero	1.43	1975	enero	1.87
	febrero	1.34		febrero	1.43
	marzo	1.48		marzo	1.87
	abril	1.57		abril	2.03
	mayo	1.62		mayo	1.88
	junio	1.36		junio	1.45
	julio	1.44		julio	1.43
	agosto	1.49		agosto	1.22
	setiembre	1.43		setiembre	2.29
	octubre	1.53		octubre	1.90
	noviembre	1.44		noviembre	1.69
	diciembre	1.45		diciembre	1.51
1973	enero	1.67	1976	enero	1.71
	febrero	1.42		febrero	1.74
	marzo	1.46		marzo	1.62
	abril	1.36		abril	2.00
	mayo	1.43		mayo	1.90
	junio	1.42		junio	2.01
	julio	1.56		julio	1.75
	agosto	1.72		agosto	1.80
	setiembre	1.35		setiembre	1.84
	octubre	1.51		octubre	1.82
	noviembre	1.28		noviembre	1.85
	diciembre	1.16		diciembre	1.81
1974	enero	1.48			
	febrero	1.15			
	marzo	1.56			
	abril	1.44			
	mayo	1.46			
	junio	1.32			

FUENTE: Dirección de Estadística de la Municipalidad de Rosario.

(continuación apéndice 5)

<i>Año</i>	<i>Mes</i>	<i>Tasas</i>	<i>Año</i>	<i>Mes</i>	<i>Tasas</i>
1977	enero	2.00		julio	1.92
	febrero	1.71		agosto	2.19
	marzo	2.15		setiembre	1.88
	abril	1.87		octubre	2.12
	mayo	1.89		noviembre	1.83
	junio	1.89		diciembre	1.52
	julio	1.91	1980	enero	1.93
	agosto	2.05		febrero	1.79
	setiembre	1.96		marzo	1.89
	octubre	1.83		abril	2.03
	noviembre	1.82		mayo	2.06
	diciembre	1.68		junio	1.83
1978	enero	1.79		julio	2.37
	febrero	1.71		agosto	1.80
	marzo	1.99		setiembre	1.92
	abril	1.87		octubre	1.91
	mayo	2.06		noviembre	1.83
	junio	1.96		diciembre	1.80
	julio	1.83	1981	enero	1.72
	agosto	1.75		febrero	1.50
	setiembre	1.85		marzo	1.93
	octubre	1.96		abril	1.87
	noviembre	1.95		mayo	1.84
	diciembre	1.62		junio	2.06
1979	enero	1.97		julio	1.97
	febrero	1.64		agosto	1.81
	marzo	2.01		setiembre	1.88
	abril	1.65		octubre	1.79
	mayo	1.87		noviembre	1.73
	junio	1.92		diciembre	1.68

(continuación apéndice 5)

<i>Año</i>	<i>Mes</i>	<i>Tasas</i>	<i>Año</i>	<i>Mes</i>	<i>Tasas</i>
1982	enero	1.69		julio	1.78
	febrero	1.74		agosto	1.74
	marzo	2.14		setiembre	1.60
	abril	1.89		octubre	1.84
	mayo	1.71		noviembre	1.55
	junio	1.78		diciembre	1.46
	julio	1.90	1985	enero	1.91
	agosto	1.89		febrero	1.73
	setiembre	1.83		marzo	1.36
	octubre	1.71		abril	1.90
	noviembre	1.97		mayo	2.24
	diciembre	1.63		junio	1.70
1983	enero	1.79		julio	1.92
	febrero	1.46		agosto	1.60
	marzo	1.77		setiembre	1.86
	abril	1.95		octubre	2.05
	mayo	2.04		noviembre	1.82
	junio	1.66		diciembre	1.69
	julio	2.01	1986	enero	1.69
	agosto	1.96		febrero	1.73
	setiembre	1.40		marzo	1.48
	octubre	1.58		abril	1.90
	noviembre	2.19		mayo	1.81
	diciembre	1.83		junio	1.59
1984	enero	1.87		julio	1.92
	febrero	1.71		agosto	1.71
	marzo	1.88		setiembre	1.72
	abril	1.69		octubre	1.67
	mayo	1.76		noviembre	1.59
	junio	1.81		diciembre	1.66

Apendice 6. TASAS DE MORTALIDAD MENSUALES DE LA CIUDAD DE ROSARIO
(ENE. 1972-DIC. 1986)

Año	Mes	Tasas	Año	Mes	Tasas	
1972	enero	0.91		julio	0.92	
	febrero	0.53		agosto	1.05	
	marzo	0.72		setiembre	0.89	
	abril	0.69		octubre	0.88	
	mayo	0.84		noviembre	0.78	
	junio	0.89		diciembre	0.84	
	julio	1.24		1975	enero	0.86
	agosto	1.00			febrero	0.75
	setiembre	0.91			marzo	0.75
	octubre	0.82			abril	0.80
	noviembre	0.76			mayo	0.91
	diciembre	0.79			junio	1.02
1973	enero	0.82	julio		1.05	
	febrero	0.68	agosto		1.01	
	marzo	0.76	setiembre		0.85	
	abril	0.72	octubre		0.90	
	mayo	1.03	noviembre		0.81	
	junio	0.96	diciembre		0.92	
	julio	1.15	1976	enero	0.81	
	agosto	0.94		febrero	0.74	
	setiembre	0.82		marzo	0.77	
	octubre	0.79		abril	0.81	
	noviembre	0.77		mayo	0.87	
	diciembre	0.80		junio	1.26	
1974	enero	0.82		julio	1.13	
	febrero	0.71		agosto	1.07	
	marzo	0.77		setiembre	0.93	
	abril	0.75		octubre	0.87	
	mayo	0.84		noviembre	0.81	
	junio	1.00		diciembre	0.84	

FUENTE: Dirección de Estadística de la Municipalidad de Rosario.

(continuación apéndice 6)

<i>Año</i>	<i>Mes</i>	<i>Tasas</i>	<i>Año</i>	<i>Mes</i>	<i>Tasas</i>
1977	enero	0.86		julio	1.06
	febrero	0.77		agosto	0.92
	marzo	0.84		setiembre	0.89
	abril	0.86		octubre	0.78
	mayo	0.99		noviembre	0.78
	junio	1.02		diciembre	0.86
	julio	0.99	1980	enero	0.71
	agosto	1.05		febrero	0.72
	setiembre	0.91		marzo	0.80
	octubre	0.78		abril	0.82
	noviembre	0.78		mayo	0.82
	diciembre	0.77		junio	1.27
1978	enero	0.81		julio	1.27
	febrero	0.66		agosto	1.01
	marzo	0.75		setiembre	0.83
	abril	0.78		octubre	0.80
	mayo	0.89		noviembre	0.71
	junio	0.97		diciembre	0.75
	julio	1.00	1981	enero	0.75
	agosto	1.03		febrero	0.66
	setiembre	0.80		marzo	0.84
	octubre	0.79		abril	0.82
	noviembre	0.81		mayo	0.84
	diciembre	0.75		junio	0.94
1979	enero	0.85		julio	1.02
	febrero	0.64		agosto	0.86
	marzo	0.77		setiembre	0.83
	abril	0.79		octubre	0.82
	mayo	0.93		noviembre	0.69
	junio	1.01		diciembre	0.78

(continuación apéndice 6)

<i>Año</i>	<i>Mes</i>	<i>Tasas</i>	<i>Año</i>	<i>Mes</i>	<i>Tasas</i>
1982	enero	0.71		julio	1.07
	febrero	0.65		agosto	1.04
	marzo	0.72		setiembre	0.87
	abril	0.72		octubre	0.81
	mayo	0.82		noviembre	0.76
	junio	0.94		diciembre	0.82
	julio	1.03	1985	enero	0.76
	agosto	0.94		febrero	0.68
	setiembre	0.77		marzo	0.78
	octubre	0.78		abril	0.72
	noviembre	0.72		mayo	0.85
	diciembre	0.85		junio	0.93
1983	enero	0.78		julio	1.14
	febrero	0.60		agosto	0.91
	marzo	0.75		setiembre	0.82
	abril	0.75		octubre	0.78
	mayo	0.86		noviembre	0.77
	junio	1.09		diciembre	0.58
	julio	1.07	1986	enero	0.77
	agosto	1.05		febrero	0.69
	setiembre	0.83		marzo	0.78
	octubre	0.77		abril	0.76
	noviembre	0.80		mayo	0.88
	diciembre	0.82		junio	0.92
1984	enero	0.80		julio	0.83
	febrero	0.69		agosto	0.89
	marzo	0.82		setiembre	0.87
	abril	0.87		octubre	0.79
	mayo	0.90		noviembre	0.75
	junio	1.01		diciembre	0.84

Apendice 7.- MODELO LOGISTICO MODIFICADO (MLM)

El MLM aplicado a poblaciones humanas (Meade,1988), es una modificación de la curva logística considerada por Leach

$$P_t = \frac{P_\infty}{1+(P_\infty/P_0-1) \exp(-ct^*)} + d \quad 7.1.$$

donde t^* es el año t menos el año base.

P_t es una variable aleatoria que representa el tamaño de la población luego de t años,

P_0 parámetro que indica la población en el año base,

P_∞ parámetro que indica el nivel de saturación de la población cuando $t \rightarrow \infty$,

c es la tasa de crecimiento.

Los cambios introducidos ponen énfasis en el ajuste de la tendencia logística para la obtención de buenas proyecciones de la misma. Estudios realizados determinan que las proyecciones obtenidas del MLM muestran ser superiores a las del modelo logístico usado por Leach.

Apendice 8.- MODELOS ARIMA DE BOX Y JENKINS

Las proyecciones de series de tiempo consisten en hacer uso de un modelo de series de tiempo ajustado a datos históricos para extrapolar en el futuro. Este método toma la serie observada Z_1, Z_2, \dots, Z_T como una realización de un proceso estocástico.

Cada Z_t de la serie es considerada como una variable aleatoria de manera que la serie tendrá una función de probabilidad conjunta $p(Z_1 \dots Z_T)$. Por lo general no interesa la forma específica de la distribución conjunta, sino que se expresa la variable aleatoria Z_t en forma de un modelo lineal discreto conocido como modelo ARIMA (p, d, q) (autorregresivo / integrado / prom. móviles) que puede expresarse de la siguiente manera

$$\phi(B) (1-B)^d Z_t = \theta(B) u_t + \mu \quad 8.1.$$

donde

d : número de diferencias consecutivas para que el modelo sea estacionario,

B : es el operador de rezago (por ej. $BX_t = X_{t-1}$),

μ : media general,

u_t : disturbios o errores aleatorios con $E(u_t)=0$,

$Var(u_t)$ constante y $Cov(u_t, u_{t+j})=0$,

$\phi(B)=(1-\phi_1B-\dots-\phi_pB^p)$: es el polinomio autorregresivo de orden p , las raíces deben caer fuera del círculo unitario para que el proceso sea estacionario,

$\theta(B)=(1-\theta_1B-\dots-\theta_qB^q)$: es el polinomio promedios móviles de orden q , las raíces deben caer fuera del círculo unitario para que el proceso sea invertible.

Para obtener buenas estimaciones de los parámetros el proceso debe ser estacionario e invertible. En las aplicaciones generalmente se requiere para ajustar un modelo ARIMA trabajar con series "débilmente estacionarias" esto significa que no presentan cambios sistemáticos en la media o variancia. Es decir:

$$E(Z_1) = \dots = E(Z_T) = \mu \quad 8.2.$$

$$\text{Var}(Z_1) = \dots = \text{Var}(Z_T) = \tau_0 \quad 8.3.$$

$$\text{Cov}(Z_t, Z_{t+j}) = \text{Cov}(Z_{t+m}, Z_{t+m+j}) = \tau_j \quad 8.4.$$

en la práctica la mayoría de las series son no estacionarias pero se pueden transformar para que lo sean aplicando por ejemplo diferencias, raíz cuadrada o logaritmo a la serie original.

En el caso de los modelos SARIMA $(p, d, q)(P, D, Q)$, donde se agrega la parte estacional, se puede escribir el modelo general como una combinación de polinomios determinísticos estacionales con tendencia en el tiempo junto con procesos estocásticos promedios móviles y autorregresivos estacionales

$$\Gamma(B)_s \phi(B) (1-B)^d (1-B)^D Z_t = \Omega(B)_s \theta(B) u_t \quad 8.5.$$

donde

D : número de diferencias de orden estacional,

d : número de diferencias de orden no estacional,

B : operador retroactivo o de rezago,

s : período estacional,

$\Gamma(B)_s = (1 - \Gamma_1 B^s - \dots - \Gamma_P B^{Ps})$: polinomio autorregresivo estacional de orden P , cuyas raíces deben caer fuera del círculo unitario para que el proceso sea estacionario,

$\Omega(B)_s = (1 - \Omega_1 B^s - \dots - \Omega_Q B^{Qs})$: polinomio promedios móviles estacional de orden Q , cuyas raíces deben caer fuera del círculo unitario para que el proceso sea invertible,

$\phi(B)$, $\theta(B)$ y u_t están definidos como antes.

Un modelo estacionario puede ser caracterizado por sus dos primeros momentos esperanza y variancia-covariancia.

La función de autocovariancia :

$$\tau_j = \text{cov}(Z_t, Z_{t+j}) = E[(Z_t - \mu)(Z_{t+j} - \mu)] \quad 8.6.$$

depende de las unidades de medida de las variables pero se la puede standardizar haciendo

$$f_j = \tau_j / \tau_0 \quad 8.7.$$

que es llamada función de autocorrelación (fac). Esta fac tiene tres propiedades:

- 1- es una función par,
- 2- $|f_j| \leq 1$,
- 3- no es única (a varios modelos le puede corresponder la misma fac), pero a un modelo le corresponde una única fac.

Otra característica de un proceso es la función de autocorrelación parcial (facp), que mide la correlación que existe entre dos valores rezagados sin tener en cuenta los intermedios.

Para la construcción y elección de los modelos lineales de series de tiempo, Box y Jenkins presentan un procedimiento iterativo que básicamente consiste en cuatro etapas:

- 1) "Identificación" de un modelo tentativo,
- 2) "Estimación" de los parámetros del modelo,
- 3) "Chequeo" del modelo para ver si éste es adecuado o no,
- 4) "Previsiones" de valores futuros.

1) Identificación de un modelo de la clase general de modelos

Esto se hace con la ayuda de la función de autocorrelación muestral (facm) y la función de autocorrelación parcial muestral (facpm).

En la práctica, nunca se conocen los valores poblacionales de las autocorrelaciones y autocorrelaciones parciales de los procesos estocásticos. En consecuencia, cuando se quiere identificar un modelo se

deben utilizar la función de autocorrelación muestral (facm) y la función de autocorrelación parcial muestral (facpm). Una estimación de $\{r_k\}$ se calcula usando la fórmula:

$$r_k = c_k/c_0 \quad 8.8.$$

$$\text{donde } c_k = (1/n) \sum_{t=1}^{n-k} (Z_t - Z)(Z_{t+k} - Z) ; k \geq 0 \quad 8.9.$$

es la estimación de la auto-covariancia r_k , Z_t es la serie estacionaria y Z es la media de la serie.

Para identificar los modelos es necesario saber si estas autocorrelaciones son significativas de cero, es por esto que se debe conocer el error standard de las mismas. Considerando que la variancia de r_k es

$$\text{Var}(r_k) \approx (1/n)(1 + 2 \sum_{i=1}^q r_i^2) ; k > q \quad 8.10.$$

puede ser estimada y determinar qué autocorrelaciones son significativas de cero. Se grafica la función de autocorrelación estimadas con sus respectivos intervalos de confianza y se obtiene lo que se conoce como correlograma.

Es importante notar que la facpm no tiene una fórmula simple, la misma se despeja de las llamadas ecuaciones de Jules-Walker.

La forma de la facm junto con la de facpm dan una idea del modelo que puede postularse.

2) Estimación de los parámetros del modelo

Una vez identificado el modelo se deben estimar los parámetros propuestos. Cuando hay modelos autorregresivos puros, es posible realizar buenas estimaciones, sin embargo, cuando los modelos son mixtos o promedios móviles hay que recurrir a un método iterativo (condicional o incondicional). Los software utilizados emplean métodos iterativos aunque los modelos propuestos sean autorregresivos puros. Para encontrar los estimadores de los parámetros se supone $u_t \sim N(0, \sigma^2_u)$ i.i.d., se plantea la función de verosimilitud conjunta de los residuos. Al calcular las

derivadas de la misma para encontrar los estimadores de los parámetros que la hagan máxima, quedan planteadas funciones no lineales, por lo que se debe recurrir a métodos iterativos. Los estimadores hallados son, entonces, aproximadamente máximo verosímiles por lo que asintóticamente cumplen con las propiedades de dichos estimadores. Esto hace que sea posible encontrar su variancia y plantear test de hipótesis para probar la significación de cada uno.

3) Chequeo del modelo planteado

Identificado el modelo y estimados sus parámetros, se debe verificar si el modelo es adecuado. Esto quiere decir entre otras cosas que si hay parámetros no significativos, hay que eliminarlos y reestimar el modelo. Además se debe comprobar la aleatoriedad de los residuos, esto es, los supuestos sobre $E(u_t)=0$, $Var(u_t)$ constante y la no correlación entre los mismos. Esto se detecta en el correlograma de los residuos donde se testa la significación individual de las autocorrelaciones r_k . Este test se basa en la fórmula de Bartlett que para $t \rightarrow \infty$, r_k se distribuye normal con $E(r_k)=0$ y $Var(r_k)=1/T$. Por otra parte se realiza el test conjunto para las K primeras autocorrelaciones de los residuos donde:

H_0) los u_k son aleatorios

H_1) los u_k no son aleatorios

se utiliza la estadística Q de Ljung-Box definida como

$$Q(K) = n(n+2) \sum_{k=1}^K (1/n-k) r_k^2(u) \quad 8.11.$$

que bajo la hipótesis nula tiene una distribución aproximada χ^2 con $K-p-q-P-Q$ grados de libertad

Además para la elección y chequeo del modelo deben tenerse en cuenta los objetivos planteados. En base a éstos se determinan si se privilegia

obtener un buen ajuste, lograr buenas previsiones o encontrar un modelo parsimonioso.

4) Previsiones

Sólo se pasa a este punto si en el anterior se encontró que el modelo es adecuado, caso contrario debe volverse al punto 1) para identificar otro modelo. Cada proyección tiene una distribución probabilística y por lo tanto se pueden construir intervalos de confianza. Las proyecciones que tienen el menor ECM son

$$Z_t(h) = E(Z_{t+h} / H_t) \quad 8.12.$$

donde H_t es la "historia hasta el período t " y h el horizonte de previsión. Sabiendo que

$$Z_{t+h} = Z_t(h) + e_{t+h} \quad 8.13$$

y suponiendo que $u_t \sim N(0, \sigma^2_u) \Rightarrow e_{t+h} \sim N(0, \text{Var}(e_{t+h})) \Rightarrow$

$$(Z_{t+h} / H_t) \sim N(Z_t(h), \text{Var}(e_{t+h})) \quad 8.14.$$

Con lo que los intervalos de confianza del 95% serán

$$Z_t(h) \pm 1,96 \text{Var}(e_{t+h}) \quad 8.15.$$

Cuando se trabaja con series logaritmadas las previsiones están dadas por (Nelson, 1973)

$$Z_t(h) = \exp \{ \tilde{Z}_t(h) + \frac{1}{2} \text{Var}(e_t(h)) \} \quad 8.16.$$

donde: $Z_t(h)$ es la previsión de la serie original,

$\tilde{Z}_t(h)$ es la previsión de la serie logaritmada,

$V(e_t(h))$ es la variancia del error de previsión.

Los intervalos de confianza se hacen tomando

$$\text{límite inferior} = \exp(Z_{t1}^*(h)) \quad 8.17.$$

$$\text{límite superior} = \exp(Z_{ts}^*(h)) \quad 8.18.$$

donde $Z_{t1}^*(h)$ y $Z_{ts}^*(h)$ son los límites inferiores y superiores de la serie logaritmada.

Debido a que el mejor modelo es elegido entre varios posibles modelos, una desventaja de esta metodología a tener en cuenta es que las previsiones varían de acuerdo al modelo seleccionado. Por este motivo la elección del mejor modelo debe hacerse cuidadosamente teniendo en cuenta todos los criterios mencionados.

Para la aplicación de esta metodología los especialistas consideran que se debe contar con un mínimo de 50 observaciones equiespaciadas en el tiempo.

Apéndice 9.- DESESTACIONALIZACION DE SERIES. METODO DE MEDIAS MOVILES

El análisis de las series mensuales se puede realizar estudiando en cada una de ellas el comportamiento de la tendencia y de la estacionalidad. El modelo aditivo que representa una serie estacional con tendencia es el siguiente:

$$Z_t = T_t + S_t + a_t \quad 9.1.$$

donde T_t : componente de tendencia.

S_t : componente de estacionalidad.

a_t : error aleatorio con $E(a_t)=0$ y $\text{Var}(a_t)=cte$.

Se aplica el procedimiento de medias móviles con el fin de desestacionalizar la serie, es decir, destruir la correlación entre meses para poder así reconocer e interpretar movimientos importantes no

estacionales que perturban la serie. Esto también permite reconocer el comportamiento de la estacionalidad.

En este método lo que se intenta primeramente es estimar la tendencia con un proceso de medias móviles centrado en 12 meses de la forma

$$T^*_t = [z_{t-6} + 2 \sum_{j=-5}^5 z_{t+j} + z_{t+6}] / 24 \quad t=7, \dots, N-6 \quad 9.2.$$

luego, considerando la serie sin tendencia ($Y_t = Z_t - T^*_t$) se estima la estacionalidad haciendo

$$Y_{.j} = (1/n_{.j}) \sum_{i=1}^{n_{.j}} Y_{tj} \quad j=1, \dots, 12 \quad , \quad 9.3.$$

luego

$$Y_{..} = (1/12) \sum_{j=1}^{12} Y_{.j} \quad , \quad 9.4.$$

y por último

$$S_j = Y_{.j} - Y_{..} \quad j=1, \dots, 12 \quad , \quad 9.5.$$

obteniéndose de este modo la estimación de la estacionalidad para cada mes.