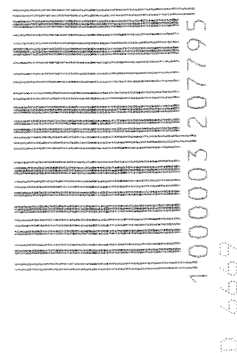


7803

EL PARO REGISTRADO NO AGRICOLA 1964 – 1976:
UN EJERCICIO DE ANALISIS ESTADISTICO UNIVARIANTE
DE SERIES ECONOMICAS

Antoni Espasa



Agradezco a Mercedes Montojo y Marisa Rojo su eficiente e infatigable colaboración como ayudantes de investigación en este trabajo.

S U M A R I O

	<u>Pág.</u>
I. Introducción	1
II. Tendencia en el paro registrado no agrícola (PRNA) y en el paro registrado total (PRT).	5
III. Heteroscedasticidad	9
IV. La estacionalidad en el PRNA	11
V. Un modelo ARIMA para PRNA	13
V.1. Identificación	13
V.2. Estimación y predicción	17
VI. Conclusión	22
APENDICE: Correlogramas y espectros de PRA, PR1, PRS, PRC, PRAD, PR1D, PRSD y PRCD.	26

-----00000000-----

I.- INTRODUCCION

El objeto de este documento es ilustrar en base a una serie concreta, pero registrado no agrícola, una metodología para el estudio univariante de las series económicas. Este enfoque no implica en absoluto convencimiento alguno por nuestra parte de que las series económicas han de estudiarse aisladamente. Todo lo contrario, compartimos la afirmación de que la explicación de los fenómenos económicos ha de ser por métodos multivariantes. Sin embargo, antes de introducirnos en ellos conviene conocer de forma aislada la estructura de cada una de las series. Sin este conocimiento previo, la modelación de estructuras económicas multivariantes es prácticamente imposible.

Dado este enfoque univariante, la metodología de este documento es puramente estadística y no considera el uso de los conocimientos a priori de la Teoría Económica. Esto no implica tampoco un menosprecio de las técnicas econométricas de las que estamos plenamente convencidos que son instrumento adecuado para la modelación de un sistema económico. Pero igualmente creemos que, dada la insuficiente información a priori de que se dispone para construir modelos econométricos, principalmente en cuanto a las estructuras dinámicas de los mismos, tanto en sus partes sistemáticas como en las referentes a las perturbaciones estocásticas que les afectan, un conocimiento previo de las series que entran en los modelos econométricos es un paso imprescindible en la construcción de éstos. Más aún, en la formulación de modelos macroeconométricos, principalmente para fines de predicción a corto plazo, es de suma importancia disponer de un conocimiento estadístico a fondo de las series económicas a un nivel de desagregación menor.

El estudio univariante de las series económicas, además de su importancia ya mencionada para ulteriores in

investigaciones cuantitativas de mayor complejidad, es también un método útil para el análisis de la coyuntura. En efecto, la consideración tradicional de las series económicas como compuestas de cuatro factores, tendencial, estacional, cíclico y errático, refleja cuatro centros de atención importantes para el analista de la coyuntura. La última parte de este documento se dedica a la formulación y estimación de modelos que incluyen éstos cuatro aspectos. Modelos que, como veremos, se pueden utilizar para predecir en un horizonte de unos seis meses los valores futuros de series mensuales. Previamente, las secciones II a IV se dedican a comentar los aspectos tendencial y estacional de las series.

Los instrumentos más utilizados en el análisis estadístico de este trabajo son el correlograma y el espectro. Un instrumento previo es el simple gráfico de la serie, tanto en su versión original como en primeras diferencias. En realidad, dichos gráficos y los del correlograma y espectro contienen la misma información y son, digamos, formas alternativas de presentarla que pueden ser más o menos útiles según los casos y objetivos. Así el examen previo del gráfico de la serie es importante para detectar si existen observaciones atípicas ("outliers") que quizás pueda convenir corregir de antemano, ya que dichas observaciones pueden afectar al correlograma y espectro. Advertamos que este tipo de correcciones en un análisis multivariante puede ser peligroso, pues dichas observaciones, consideradas inicialmente atípicas, pueden ser ocasionadas por el comportamiento de otras series. Incluso en el análisis univariante conviene ser muy cauto en cuanto a la corrección de observaciones atípicas y, en todo caso, debe realizarse cuando existan motivos a priori para suponer que es un factor muy especial la causa de la observación atípica.

Un primer examen del gráfico de la serie nos puede dar ya las primeras ideas sobre dos aspectos tendenciales,

estacionales y cíclicos de la serie. No obstante, es en base al correlograma y el espectro donde se pueden estudiar con más detalle todos los problemas, como veremos en las secciones posteriores. Para una exposición sencilla sobre series temporales véase Chatfield (1975) y para una exposición más rigurosa Box y Jenkins (1976) y Jenkins y Watts (1969) para el tema espectral. Aquí nos limitaremos a dar unas simples ideas sobre el correlograma y el espectro. El primero es el gráfico de los coeficientes de correlación de una serie con sus respectivos retardos y, como veremos luego, un examen del mismo, nos puede servir para contrastar si la serie es aleatoria, si tiene mucha o poca correlación serial, si es o no estacionaria, si contiene o no fluctuación estacional, etc.

El punto de partida en el análisis espectral consiste en que todo proceso estacionario (generador de una serie temporal) se puede representar como la suma de infinitos elementos, cada uno con una frecuencia distinta, con lo que es posible examinar la serie desde el punto de vista de su comportamiento en cada frecuencia o banda de frecuencias, en vez del examen clásico de su comportamiento a través del tiempo. El objetivo de tal descomposición es ver qué elementos, referentes a bandas de frecuencias específicas, tienen más importancia en el comportamiento de la serie. El análisis espectral se puede aplicar a una serie individual o a varias series conjuntamente. En el caso uniserial, el concepto básico es el del espectro. Este mide la contribución de un elemento, con una frecuencia determinada, en la varianza de la serie, de tal forma que, sumando los valores del espectro en todas las frecuencias obtenemos la varianza de la serie. Así, examinando el gráfico del espectro, podemos ver qué elementos pertenecientes a determinadas bandas de frecuencia, afectan más a la varianza de la serie y por tanto afectan más a ésta. No obstante, la utilización del análisis espectral en este documento es mínima y el lector que no esté familiarizado con el tema puede saltarse las partes del

texto que utilizan el análisis espectral sin peligro alguno de perder la continuidad del documento.

Respecto de las series de paro registrado, se han señalado desde antiguo grandes inconvenientes, ya que se limita a recoger a los que se registran como parados y no a proporcionar una estimación más adecuada del verdadero paro existente en el sistema. En este sentido, la información sobre el paro que proporciona la encuesta de población activa es, sin duda, una estimación mejor. No obstante, desde el punto de vista del análisis cuantitativo, la periodicidad, trimestral, de la encuesta de población activa es totalmente insuficiente. Para el análisis de la coyuntura y para la decisión de medidas económicas es indispensable conocer la evolución del paro mes a mes. Por otra parte, la metodología de la encuesta de población activa continúa cambiando y no disponemos de una serie histórica en la que sea posible basar un estudio sobre el comportamiento del paro.

Así, por ejemplo, al final del primer trimestre de 1977 todavía no conocemos el desglose del agregado total del paro para el segundo trimestre de 1976. También debido a este cambio de metodología el avance sobre el paro no agrícola para el tercer trimestre de 1976 representa una reducción en un 20 por ciento respecto al computado para el primer trimestre de ese año. Estas razones nos han llevado a escoger como serie de paro para este estudio, la de paro registrado.

II - TENDENCIA EN EL PARO REGISTRADO NO AGRICOLA (PRNA) Y EN EL PARO REGISTRADO TOTAL (PRT)

Ambas series son no estacionarias en la media. Es decir la media de dichas series evoluciona con el tiempo y en este caso de una forma creciente, por lo que decimos que ambas series tienen una tendencia positiva bastante marcada. El carácter no estacionario de estas series se aprecia simplemente viendo el gráfico de las mismas (véase por ejemplo el gráfico 1). Obviamente la no estacionariedad se aprecia también en el correlograma y en el espectro. Así en el gráfico 2 vemos que el correlograma del PRNA decae hacia cero muy lentamente. Igualmente en el examen del espectro de PRNA tenemos que en la serie original (*) los factores con periodicidad superior a cinco años contribuyen en un 73 por ciento a la varianza total, mientras que tomando desviaciones sobre una tendencia lineal la contribución de esos factores es del 15,19 por ciento. Igualmente ocurre con el PRT donde esos porcentajes son 96,7 y 11,87 respectivamente.

La importancia de la tendencia es distinta en ambas series. Así, las desviaciones sobre una tendencia polinomial de segundo orden tienen una varianza que es 57 por ciento (1) de la varianza de la serie original en PRT y sólo del 28

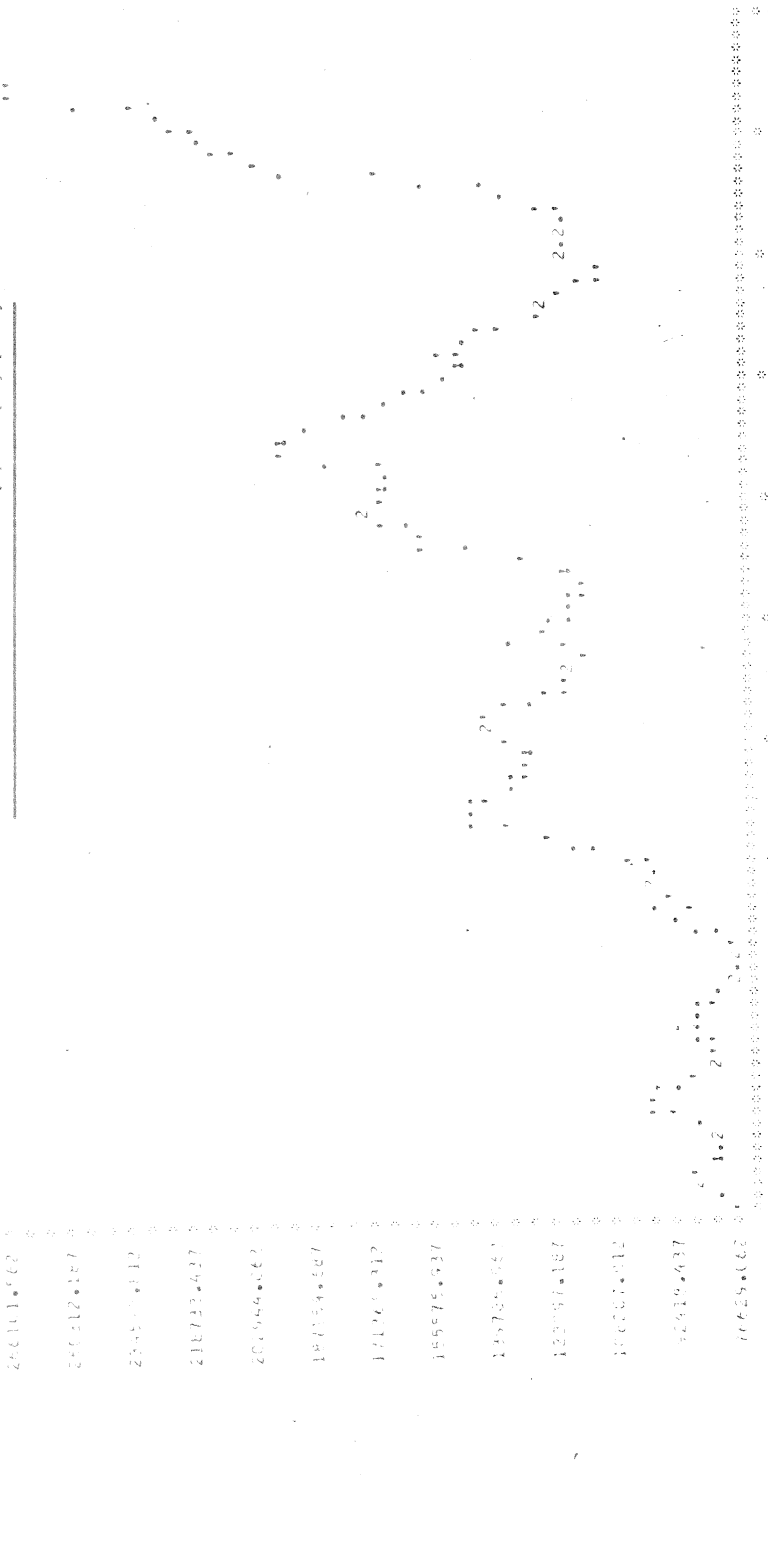
(*) En la estimación del espectro de la serie original se ha utilizado un programa que previamente corrige la tendencia de la serie mediante la aplicación de un filtro autorregresivo (preblanqueo) y posteriormente recolora los resultados obtenidos para la serie corregida. El filtro autorregresivo lo obtiene mediante una regresión por etapas de la serie original con sus retardos. No obstante el programa no contiene ningún test que verifique si la serie corregida es estacionaria por lo que las estimaciones espectrales pueden ser inconsistentes. Consecuentemente los resultados espectrales de este documento pueden estar sujetos a errores.

(1) Período de estimación 1963(XII)-1975(XII).



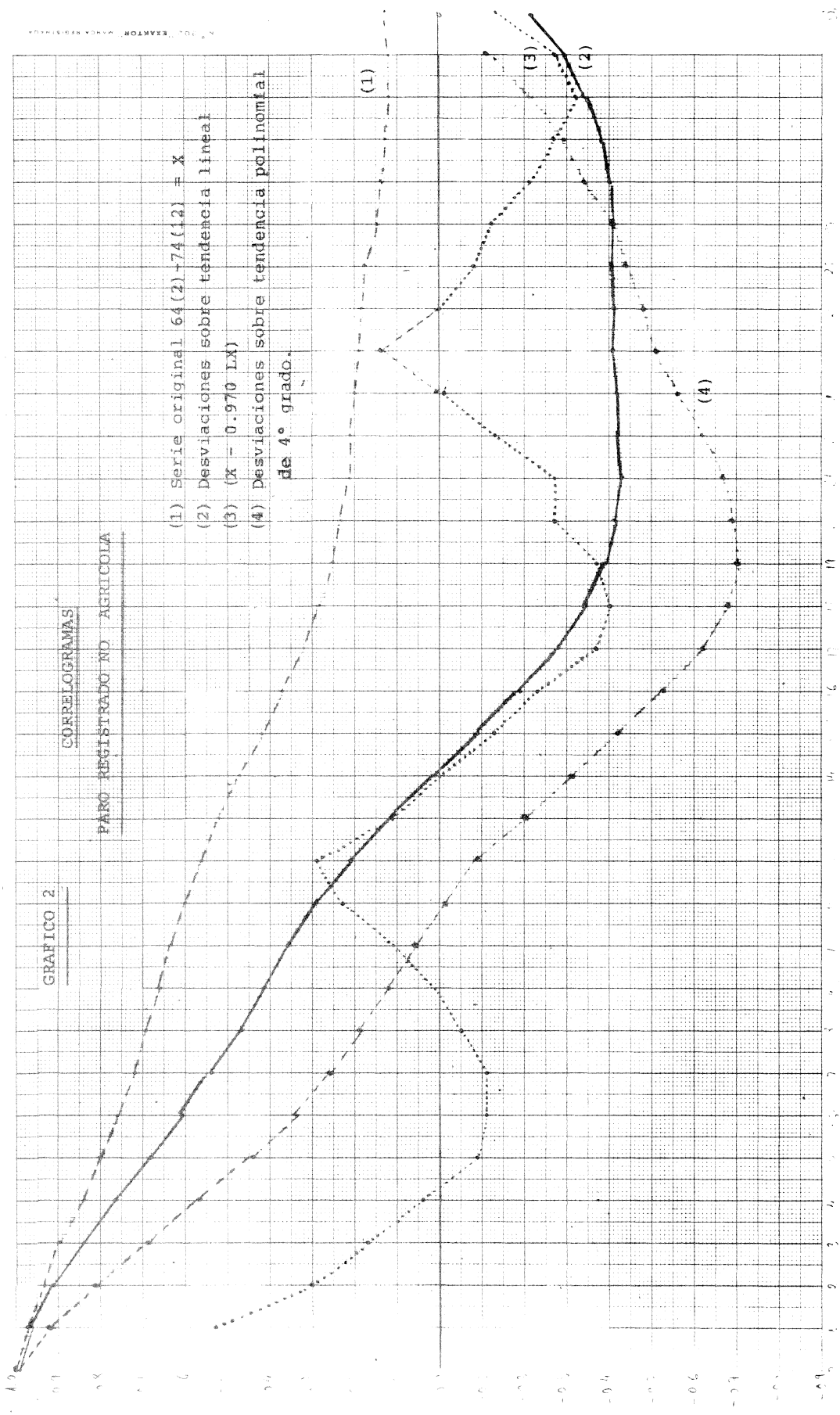
GRAFICO 1

PARO REGISTRADO NO AGRICOLA 1964(I)-1976(IX) (PRNA)



1964(I)

1976(IX)



por ciento ⁽²⁾ en el PRNA. Asimismo si tratamos la tendencia por medio de un filtro autorregresivo la varianza de la serie residual es el 4,38 ⁽¹⁾ por ciento de la original en PRT (filtro $(1 - .9L - .1L^{35})$ (**)) y sólo del 1,89 ⁽²⁾ en PRNA (filtro $(1-.96L)$). Asimismo, corregidas de tendencia (método polinomial o autorregresivo) las series PRT y PRNA la variabilidad por factores a corto plazo es mayor en PRT que en PRNA. Es decir, el paro agrícola actúa en el agregado aumentando la varianza de éste último. Resultado por otra parte lógico.

No obstante, la estimación de la tendencia de estas series es algo complejo. Primero la consideración del año 1975 aumenta el efecto tendencial. Así con un filtro autorregresivo (AR) de primer orden el coeficiente estimado ⁽¹⁾ (altamente significativo para ambas series) es de 1.019 en PRNA y significativamente distinto ⁽³⁾ de valores inferiores a la unidad. Sin embargo, si no se incluyen las observaciones correspondientes a 1975 el valor estimado es ligeramente inferior a la unidad con intervalos de confianza entre .945 y 1.016 ⁽⁴⁾. Asimismo si se ajusta una tendencia polinomial de segundo grado los coeficientes de T y T^2 son positivos si se incluye 1975 ⁽⁵⁾. No obstante esto último depende si el comienzo de la muestra está por debajo de la tendencia (1963) o al nivel de ésta (1964).

La identificación del grado del polinomio de tiempo tampoco es sencilla. Así en la serie PRT la inclusión de 1975 conduce a un polinomio de cuarto orden pero si se ex

(2) Período de estimación 1960(II)-1975(XII).

(**) L es el operador de retardos tal que aplicado j veces a una variable la retarda j períodos, $L^j(y_t) = y_{t-j}$.

(3) El nivel de significación empleado en este trabajo es del 5 por ciento.

(4) Período de estimación 1964(I)-1974(XII).

(5) Período de estimación 1962(XII)-1975(XII).

cluye sólo es significativa una tendencia de segundo grado. En la serie PRNA la tendencia polinomial de cuarto orden es significativa, incluso excluyendo 1975. Para esta serie, el test de Sargan (1964) nos lleva a preferir la tendencia exponencial a la polinomial.

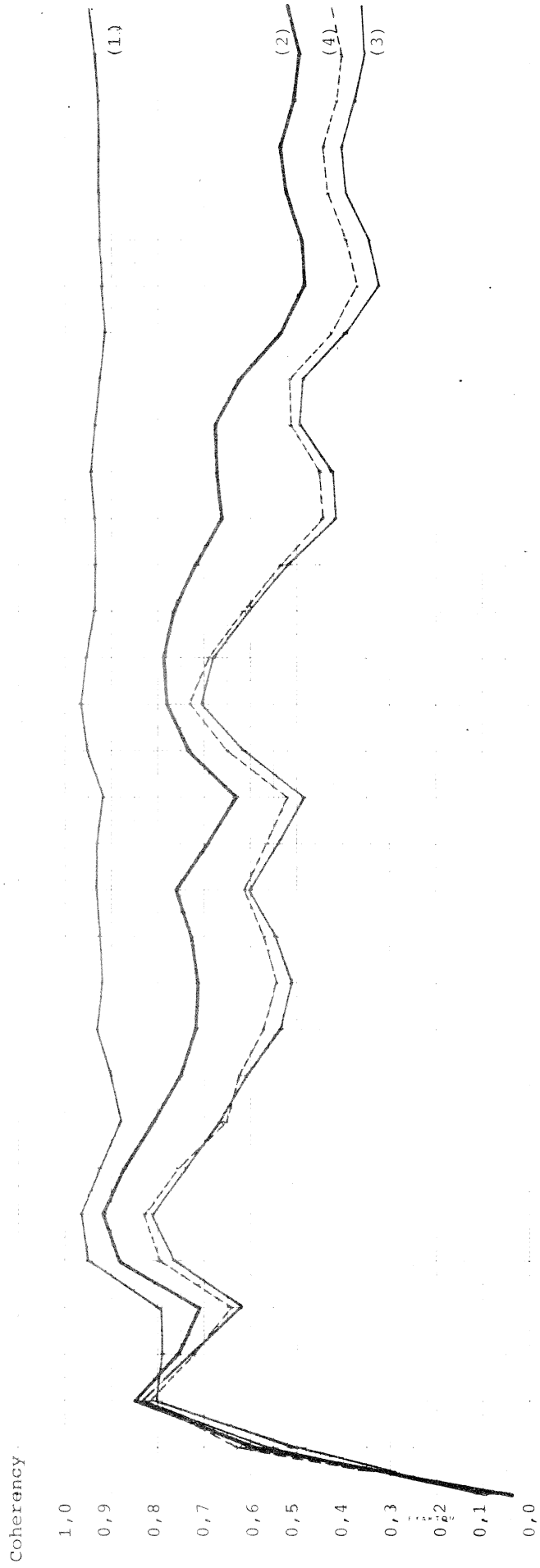
En series como PRNA en que además de un elemento tendencial contienen un elemento cíclico es de esperar que al ajustar un polinomio de tiempo en la serie se acabe con un polinomio de grado no muy bajo. Por otra parte, al ajustar una tendencia queremos que ésta remueva de la serie la variación a largo plazo y que no modifique mucho su variación a medio y corto plazo. Sin embargo, ambos objetivos son incompatibles. Así, cuanto mayor sea el grado del polinomio mayor será la variación a largo plazo que se quita a la serie pero también es mayor la modificación que se introduce en el comportamiento a medio y corto plazo. En consecuencia, se necesita un compromiso intermedio entre ambos objetivos. El gráfico 3 nos da el "coherency" de la serie original PRNA (1964(II)-1974(XII)) con sus desviaciones respecto a un polinomio de tiempo de 1º, 2º, 3º y 4º grado respectivamente. El "coherency" en el campo espectral es equivalente al ^{cuadrado del} coeficiente de correlación en el campo temporal. La ventaja del "coherency" es que se calcula para distintas bandas de frecuencia. En el gráfico 3 vemos que la correlación, de la serie original con todas las corregidas de tendencia, en las frecuencias alrededor de cero (largo plazo) es muy baja, como se deseaba. En las frecuencias entre 0,078 y 0,5 (corto plazo) la correlación disminuye con el grado del polinomio, como habíamos anticipado. Los espectros de estas series se encuentran representados en el gráfico 4, donde se aprecia la disminución del valor del espectro, en la banda alrededor de cero, con el grado del polinomio.

La importancia de los componentes de corto y largo en la varianza de la serie se puede ver en el cuadro I.

GRAFICO 3

"COHERENCY" ENTRE LA SERIE PRNA ORIGINAL Y:

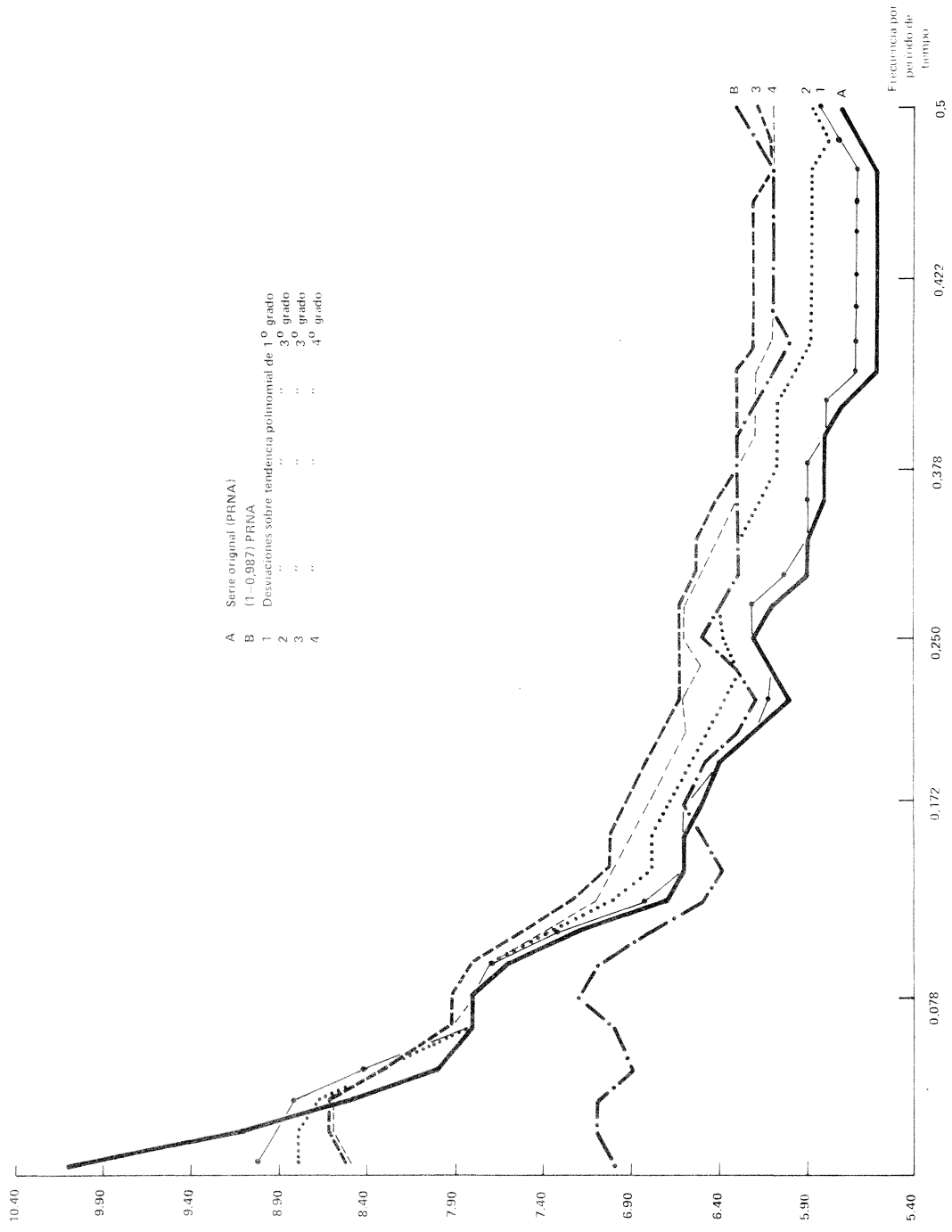
- (1) Sus desviaciones sobre un polinomio de 1° grado
- (2) Sus desviaciones sobre un polinomio de 2° grado
- (3) Sus desviaciones sobre un polinomio de 3° grado
- (4) Sus desviaciones sobre un polinomio de 4° grado



frecuencia por período de tiempo .5

GRAFICO 4

LOGARITMO DEL ESPECTRO DE PRNA



CUADRO I

PORCENTAJE DE VARIACION DE LA SERIE EXPLICADA POR
COMPONENTES DE PERIODICIDAD

Serie	Largo plazo (*)	Periodicidad	
		Superior a 64 meses	Superior a 21 meses
PRNA	77 %	92,5 %	97,4 %
Desviaciones sobre tendencia de			
1° grado	19,2 %	55,5 %	90,0 %
2° grado	15,6 %	48,2 %	85,27%
3° grado	10,3 %	36,5 %	76,6 %
4° grado	10,7 %	37,6 %	77,65%

(*) Son los componentes con frecuencia comprendida en las bandas $-0,008 (2\pi)$ y $0,008(2\pi)$ radianes.

A la vista de todos estos resultados la solución de compromiso puede ser el elegir un polinomio de segundo grado para estimar la tendencia. No obstante, conviene advertir que cualquier tendencia estimada no es más que una aproximación a una función del tiempo que nos es desconocida. Como hemos visto antes, esta aproximación depende mucho del período muestral que se utilice y en esto se debe procurar que el período muestral sea suficientemente largo pero suficientemente homogéneo. Como consecuencia de todo esto tenemos que el polinomio ajustado puede usarse para interpolar en el período de ajuste pero su utilización para predecir la tendencia fuera del período muestral puede conducir a graves errores por lo que no es aconsejable su utilización para tal fin.(*)

Con la finalidad de trabajar con una serie de comportamiento homogéneo en el período muestral considerado, en la mayor parte de este estudio utilizamos como muestra el período 1964(I)-1974(XII).

(*) La forma óptima de proceder es ajustar un modelo del tipo ARIMA (que veremos en secciones posteriores) y a partir de él estimar la tendencia de la serie. Sin embargo, en este documento no entraremos en el tema de la descomposición de los modelos ARIMA en tendencia, estacionalidad y error.

III - HETEROSCEDASTICIDAD

Como se aprecia en el gráfico I la serie PRNA no sólo tiene una tendencia en su nivel sino también en su varianza, es decir, la serie es heteroscedástica.

La heteroscedasticidad tiene un efecto considerable al estimar la tendencia en el nivel de la serie ajustando un polinomio de tiempo por mínimos cuadrados ordinarios (MCO). En efecto, cuando lo que está sometido a un efecto tendencial no es solamente la media de la serie sino también las desviaciones sobre ésta, el mejor ajuste de un polinomio temporal es el que minimiza la suma de los cuadrados de los residuos y este ajuste no tiene por qué darnos, y de hecho no nos da, una estimación adecuada de la tendencia en el nivel de la serie. Esta estimación se obtendría estableciendo hipótesis que, al menos de forma aproximada, sirviesen para corregir la heteroscedasticidad mediante una transformación de las variables del modelo.

En el caso del paro cabría esperar que al deflactar dicha variable por la población activa se corrigiese la heteroscedasticidad de la serie original. Sin embargo, no es así. Con esta corrección la nueva serie, porcentaje del paro registrado no agrícola, CPNA, tiene un efecto tendencial menor pero un carácter heteroscedástico similar a PRNA. Tenemos pues que la varianza del paro es principalmente, función de los niveles de actividad económica y no resulta sencillo ver qué corrección de la serie sería aceptable y corregiría el efecto heteroscedástico. En esta situación hemos recurrido a experimentar si simples transformaciones matemáticas nos servían para corregir la heteroscedasticidad. Box y Jenkins (1976) pag. 328, al tratar un problema similar sugieren utilizar las transformaciones propuestas por Box y Cox (1965). De entre éstas, la transformación uniparamétrica que dichos autores proponen a una serie z es:

$$z^{(\lambda)} = \begin{cases} z^\lambda & (\lambda \neq 0) \\ \log z & (\lambda = 0) \end{cases} \quad (1)$$

Se ha experimentado la transformación (1) para $\lambda = 0,5$; $0,25$ y 0 y el efecto heteroscedástico continúa presente en todas ellas. Esto parece indicar que una transformación que reduzca adecuadamente la tendencia de la varianza debe ser función de más de un parámetro. En tal caso el encontrar la transformación adecuada es un proceso bastante largo y, dado que, como veremos al tratar los modelos ARIMA, todo tipo de transformación no lineal presenta graves inconvenientes a la hora de predecir la serie original, en el presente estudio no se han intentado transformaciones distintas a (1).

Para considerar qué transformación uniparamétrica es más adecuada necesitamos un modelo sobre la formación de PRNA. En los modelos polinomiales utilizados en la sección anterior para el caso de polinomios de segundo grado tenemos que el valor del Kernel de la función de verosimilitud (Box y Cox (1965) fórmula (8)) para los distintos λ son:

	Función de verosimilitud
$\lambda = 1,0$	- 1.623
$\lambda = 0,5$	- 1.524
$\lambda = 0,25$	- 1.430
$\lambda = 0$	- 1.610

Dichos valores indican que, bajo los supuestos del modelo polinomial de 2º grado, $\lambda = 0,25$, es preferible a la serie original y también a las otras dos transformaciones experimentadas. No obstante, recordemos que con dicho valor de λ tampoco se corrige adecuadamente la heterocedasticidad.

IV - LA ESTACIONALIDAD EN EL PRNA

La estacionalidad en la serie PRNA no es muy marcada. Así en el gráfico 4 se observa que en las frecuencias estacionales no se aprecian apenas picos en el espectro. En el gráfico 5 se ve claramente que la variación estacional se debe principalmente a factores con periodicidad de 12 meses. Para poder comparar el espectro de PRNA con el de una serie con un componente estacional muy marcado en el gráfico 6 se presenta el espectro del índice de producción industrial.

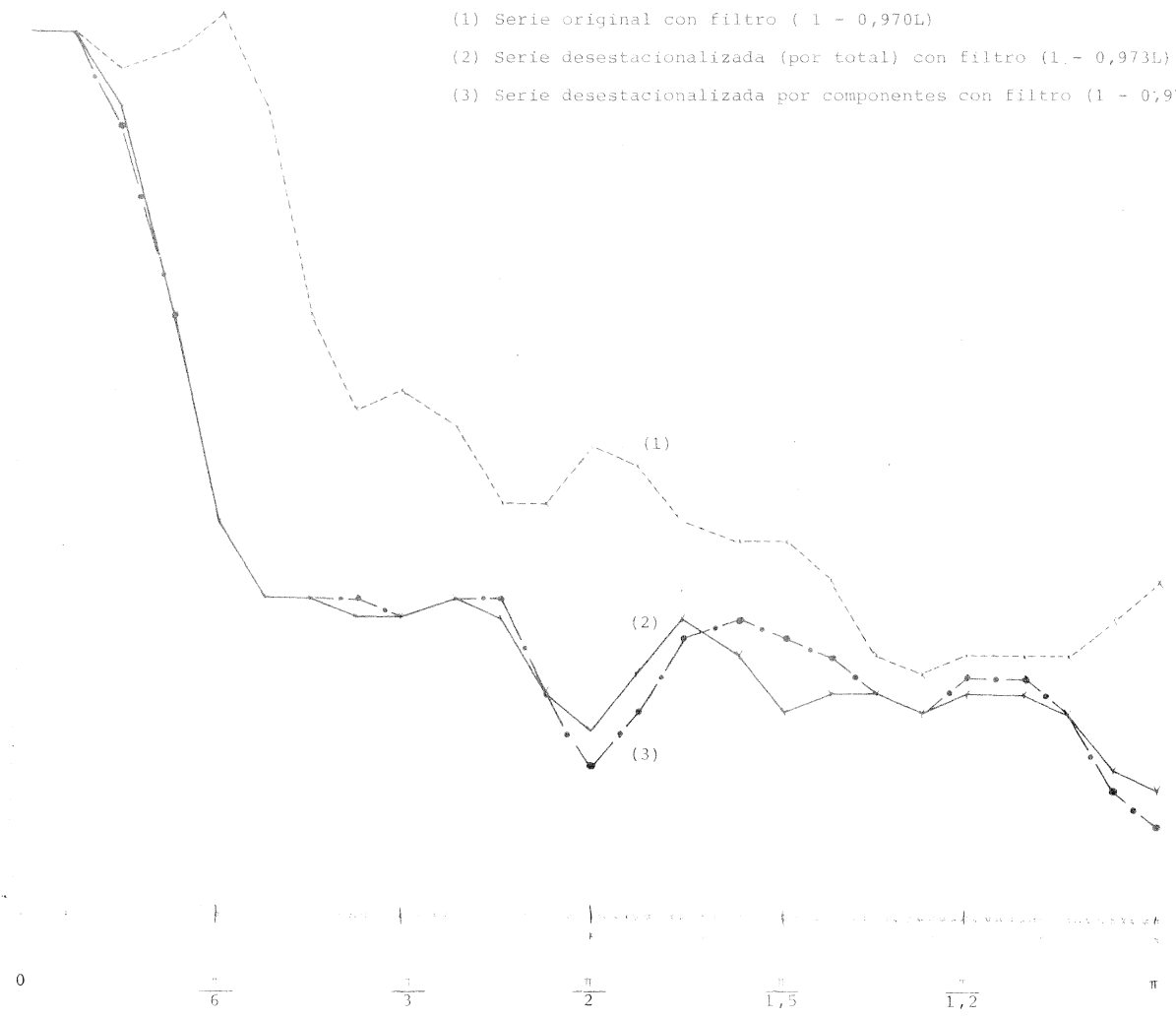
El método de desestacionalización empleado ha sido el llamado X-11 del "Bureau of the Census" (véase Shiskin (1967)) y se ha aplicado tanto a PRNA, obteniendo la serie ajustada que denominaremos PRNAD1, como a los componentes de PRNA (paro registrado en la industria, PRI, paro registrado en los servicios, PRS, y paro registrado en la construcción, PRC ⁽¹⁾) obteniendo una serie ajustada para PRNA como la suma de las series ajustadas para los componentes y que denominaremos PRNAD2.

El proceso de desestacionalización ha introducido alguna perturbación en las series ajustadas al corregir por exceso la variación estacional, como lo muestran los hundimientos que se aprecian en los espectros de PRNAD1 y PRNAD2 en el gráfico 5. Asimismo en el gráfico 7 ⁽²⁾ los "coherencias" de PRNA con PRNAD1 y PRNAD2 respectivamente en la banda de frecuencias alrededor de $\pi/6$ (periodicidad de doce meses) es muy bajo indicando que esta variación estacional - que como hemos

- (1) Estas series desestacionalizadas las denominaremos PRID, PRSD y PRCD respectivamente.
- (2) Para el "coherency" entre PRNA y PRNAD1 se dan dos estimaciones obtenidas a partir de estimaciones de los espectros y cross-espectros utilizando una ventana de Parzen con 24 y 32 retardos respectivamente.

ESPECTRO DE PRNA

- (1) Serie original con filtro (1 - 0,970L)
- (2) Serie desestacionalizada (por total) con filtro (1 - 0,973L)
- (3) Serie desestacionalizada por componentes con filtro (1 - 0,974L)

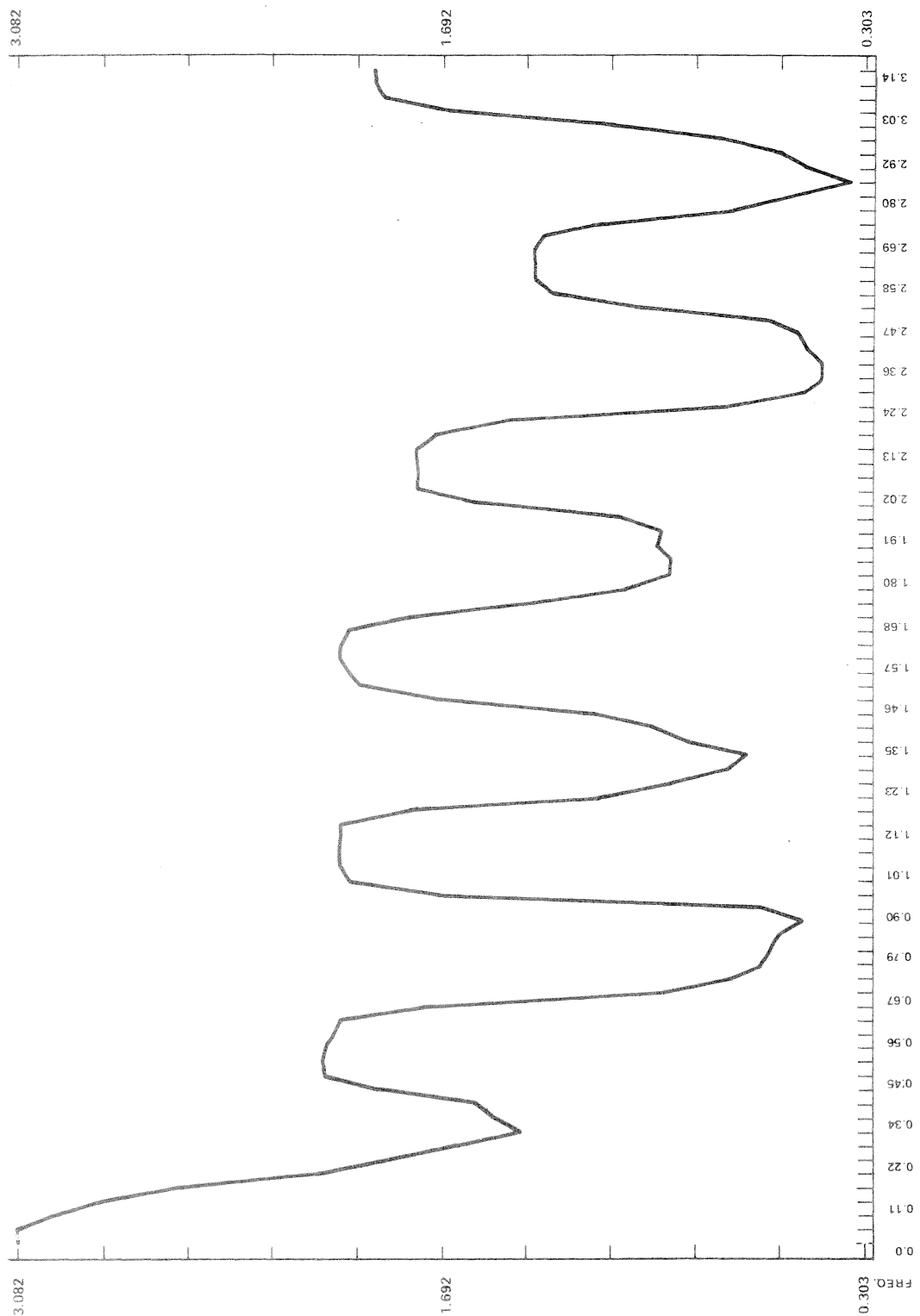


dicho antes es la principal - ha sido corregida. Igualmente el valor del "coherency" desciende también en las otras frecuencias estacionales. No obstante el "coherency" no es suficientemente alto para el resto de las frecuencias excepto en las proximidades de cero, lo que evidencia las perturbaciones introducidas por el procedimiento de ajuste. Esto es cierto tanto para PRNAD1 como para PRNAD2 por lo que es difícil elegir entre ambas. Para realizar esta elección habrá que tener presente que para las frecuencias entre 0 y $\pi/3$ radianes sus espectros son muy similares y en consecuencia también lo son sus "coherencias" con PRNA. Sin embargo, para las frecuencias $\pi/2$ y π (es decir las correspondientes a factores con periodicidad de 4 y 2 meses) PRNAD2 produce una corrección por exceso superior a la realizada por PRNAD1, mientras que en frecuencias de $\pi/1,5$ y $\pi/1,2$ (es decir las correspondientes a factores con periodicidad de 3, y 2,4 meses) ocurre lo contrario. Dependerá por tanto de qué banda se juzgue más importante para elegir uno u otro método de desestacionalización.

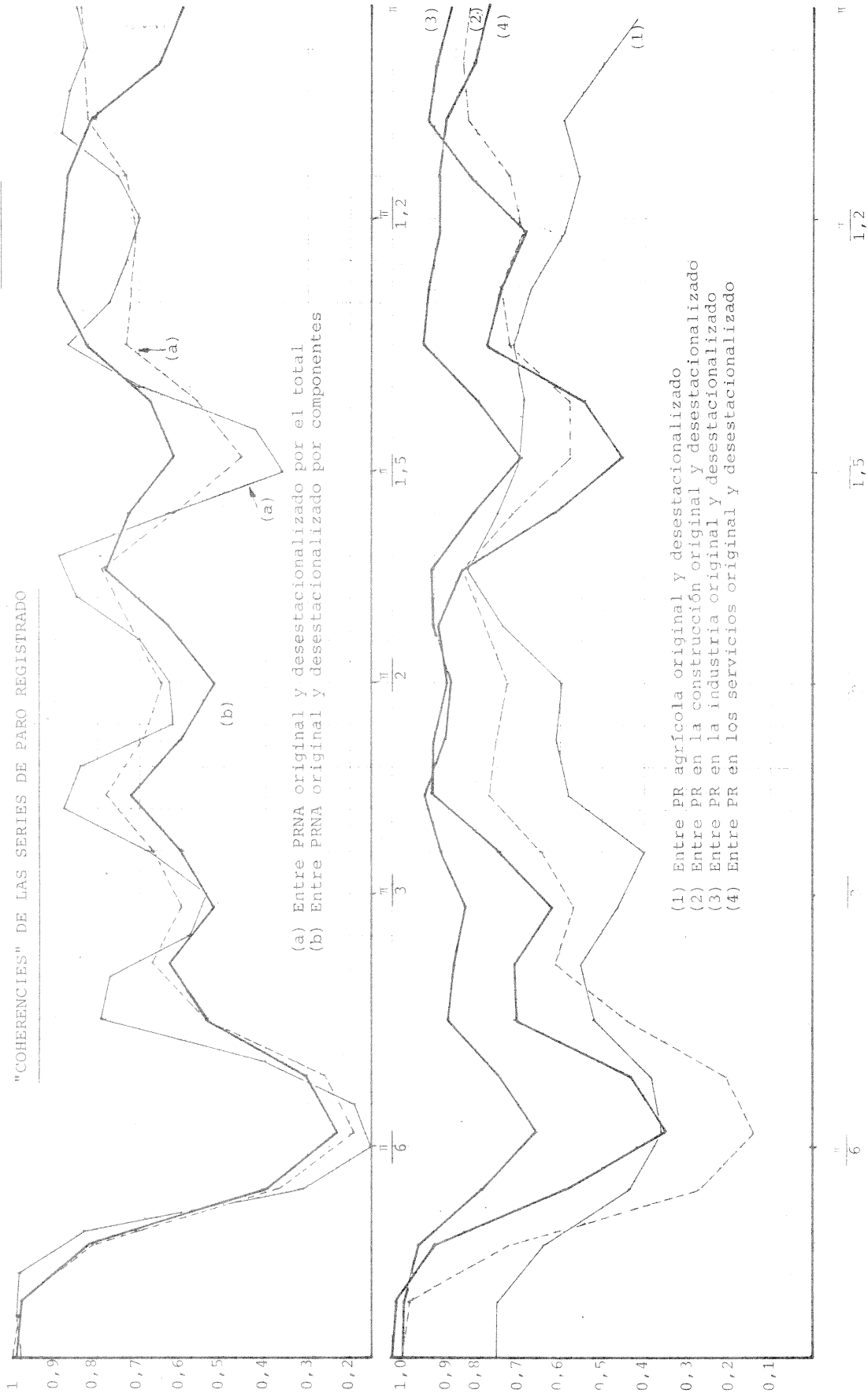
La segunda parte del gráfico 7 nos da los "coherencias" de PRI, PRC, PRS y del paro registrado agrícola (PRA) con sus correspondientes series desestacionalizadas. En él se aprecia que el proceso de desestacionalización es bastante correcto para PRS pero no lo es para las tres restantes. En el apéndice 1 se incluyen los correlogramas y espectros referentes a PRI, PRC, PRS y PRA y sus versiones ajustadas de variación estacional.

GRAFICO 6.

ESPECTRO DEL INDICE DE PRODUCCION INDUSTRIAL (IPI)
(Escala logarítmica)



"COHERENCIAS" DE LAS SERIES DE PARO REGISTRADO



V - UN MODELO ARIMA PARA PRNA

El objetivo de este apartado es identificar y estimar un modelo ARIMA del tipo:

$$\begin{aligned} & (1 - \phi_1 L - \dots - \phi_p L^p) (1 - \phi'_1 L - \dots - \phi'_{p_1} L^{p_1}) (1 - L^s)^{d_1} (1 - L)^{d_2} \text{PRNA}^* = \\ & = (1 - \theta_1 L - \dots - \theta_q L^q) (1 - \theta'_1 L - \dots - \theta'_{q_1} L^{q_1}) a_t \end{aligned} \quad (2)$$

en el que a_t es una perturbación aleatoria y PRNA^* es una transformación logarítmica de PRNA.

El problema de la transformación ha sido discutido previamente. Dado que el programa disponible no permite otra transformación que la logarítmica hemos utilizado ésta ya que como vimos en el estudio de la tendencia parecía mejor que la serie original. El período muestral utilizado para la identificación y estimación del modelo ha sido 1964(I)-1974(XII).

V.1.- IDENTIFICACION

El primer paso en la identificación es obtener una diferenciación de PRNA^* que sea estacionaria. Box y Jenkins (1976) proponen para ello examinar los correlogramas. Anderson (1976) sugiere también examinar las varianzas de PRNA^* y de sus distintas diferenciaciones para ver en cuál de ellas la varianza es menor y tomar dicha transformación como la requerida para la estacionariedad de PRNA^* . El fundamento de tal sugerencia es el siguiente. Sea z_t una serie determinada principalmente por una tendencia $f(t)$ y una perturbación aleatoria. Así

$$z_t = f(t) + a_t \quad (3)$$

si $f(t)$ es un polinomio de orden n , tenemos que $(1-L)^{n+1}f(t)=0$. En consecuencia la varianza de $(1-L)^j z_t$, $j=0, \dots, n-1$, vendrá dominada por la tendencia e irá disminuyendo a medida que aumenta el grado de diferenciación (j). Sin embargo para $j \geq n$ la varianza de z_t viene determinada por la de $(1-L)^j a_t$ y esta varianza aumenta con el grado de diferenciación, según la fórmula (véase Anderson (1971), sección 3.4).

$$\sigma^2 \binom{2j}{j}$$

donde σ^2 es la varianza de a_t . Tenemos pues que para el modelo (3) la aplicación de sucesivas diferenciaciones reduce la varianza de z_t mientras las series transformadas no son estacionarias, pero aumenta la varianza si se aplican diferenciaciones más allá del grado necesario para convertir en estacionaria la serie.

En (3) hemos supuesto un polinomio temporal de orden n que se aplica a toda la muestra. En muchos casos es más real suponer que un polinomio de grado n es válido para los subperíodos de la muestra pero no es necesariamente el mismo. En tal caso $(1-L)^n z_t$ tendrá una media que variará con el tiempo y obviamente para la estacionalidad necesitaremos $j=n+1$. Por ello, en la práctica, conviene examinar también si la media de la serie es o no significativamente distinta de cero. Con frecuencia nos encontramos con series en las que la menor varianza se logra para $j=n$ pero con una media significativa, mientras que para $j=n+1$ la media ya no es significativa. En tales casos, conviene utilizar $j=n+1$, a no ser que una quasidiferencia de la serie elimine la media.

En el cuadro 2 se presentan las desviaciones típicas para distintas transformaciones de PRNA* y PRNA según el esquema

$$(1-L)^{12} d_1 (1-L)^d Y = \frac{\text{PRNA}^*}{\text{PRNA}} \quad (4)$$

DESVIACIONES TÍPICAS DE DISTINTAS TRANSFORMACIONES DE PRNA*
Y PRNA SEGUN (4) PARA 1966(3)-74(12)

CUADRO 2

d	d ₁	0	1	2
0	PRNA	28.248,76	28.149,0	39.550,0
	PRNA*	0,2236	0,2123	0,2989
1	PRNA	5.379,52	5.883,0	9.047,0
	PRNA*	0,0397	0,0451	0,0728
2	PRNA	4.879,14	5.655,57	10.106,6
	PRNA*	0,0382	0,0487	0,0886

En el cuadro 2 se observa que tanto para PRNA como para PRNA* la menor varianza se obtiene con $d=2$, $d_1=0$ y el valor t para las medias ($t=\text{media}/\text{desviación típica de la media}$) está entre 0,23 y 0,29. Por otra parte las varianzas para $d=1$ $d_1=0$ son las que le siguen en valor, con coeficientes t inferiores a 2 ($t \approx 1,5$). Por ello será conveniente contrastar la hipótesis de que la estacionariedad se logra en $d=1$. Bajo el supuesto de que PRNA viene determinado por un esquema como (3) y que $(1-L)^2 f(t)=0$, Anderson (1971), pág. 76, propone un estadístico para realizar dicho contraste de hipótesis. Aplicado a nuestro caso nos lleva a rechazar $d=1$ en favor de $d=2$. No obstante para otros períodos muestrales, como 1969(I)-1976(X)

la varianza mínima se obtiene con $d=2$ y $d_1=1$. Con frecuencia el criterio de varianza mínima no nos conduce a un unívoco par de valores para d y d_1 sino a dos o tres de ellos. En nuestro caso los pares $(2,0)$, $(2,1)$ y quizás $(1,1)$ convendrá tenerlos presentes.

El método de Box-Jenkins (1976) para identificar d y d_1 se basa en el estudio de los correlogramas de la serie bajo consideración y el de sus respectivas diferenciaciones para combinaciones de valores de d , d_1 y s . En nuestro caso cogemos $s=12$ pues vimos al estudiar el espectro que el único pico estacional importante correspondía a factores de periodicidad de doce meses. En general, éste suele ser el valor requerido para s , no obstante pueden darse situaciones en las que el esquema expuesto en (4) tenga que ser más complejo, por ejemplo,

$$(1-L)^{s_2 d_2} (1-L)^{s_1 d_1} (1-L)^d y$$

El método del correlograma se basa (véase Box Jenkins (1976), sección 6.2.) en que para modelos ARIMA, es decir, modelos del tipo 1 con $d_1=d=0$, si son estacionarios y sus raíces no son cercanas a la unidad, el correlograma tiene de a cero moderadamente pronto mientras que si la serie tiene raíces en las proximidades de uno (es decir se aproxima a la no estacionariedad) el correlograma decrece muy lentamente.

En los gráficos 8 a 12 se recogen distintos correlogramas de PRNA para distintos valores de d y d_1 . En el correlograma de $(1-L)PRNA$ las autocorrelaciones tienen un esquema que indica que la serie es no estacionaria en el componente estacional. Tampoco decaen hacia cero suficientemente rápidos los correlogramas de $(1-L^{12})PRNA^*$ y quizás el de $(1-L)(1-L)^{12}PRNA$. El correlograma de $(1-L)^2 PRNA^*$ indica que el componente estacional es no estacionario, mientras que el correlograma de $(1-L)^2 (1-L^{12})PRNA^*$ parece responder al de una serie estacionaria. A la vista de estos correlogramas los valores elegidos para (d, d_1) son $(2,1)$ y quizás $(1,1)$.

GRAFICO 8.

CORRELOGRAMA DE (1 - L) PRNA*

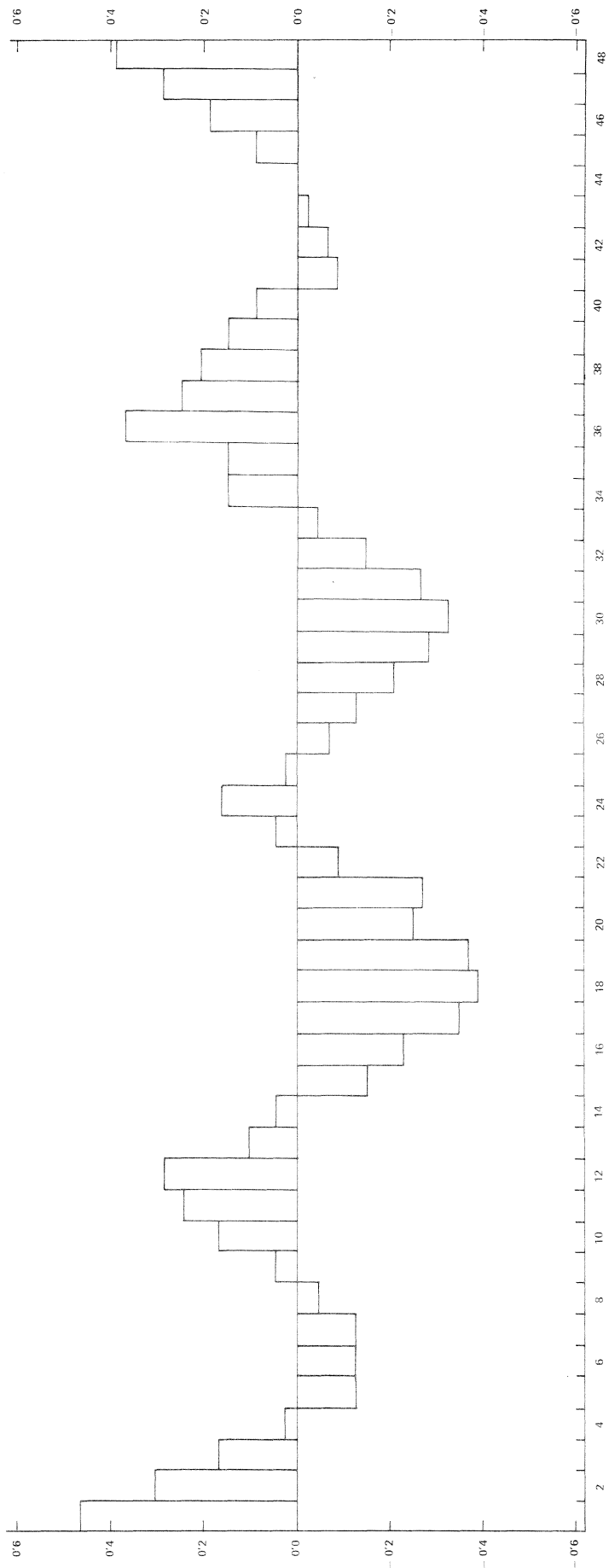


GRAFICO 9.

CORRELOGRAMA DE (1 - L¹²) PRNA*

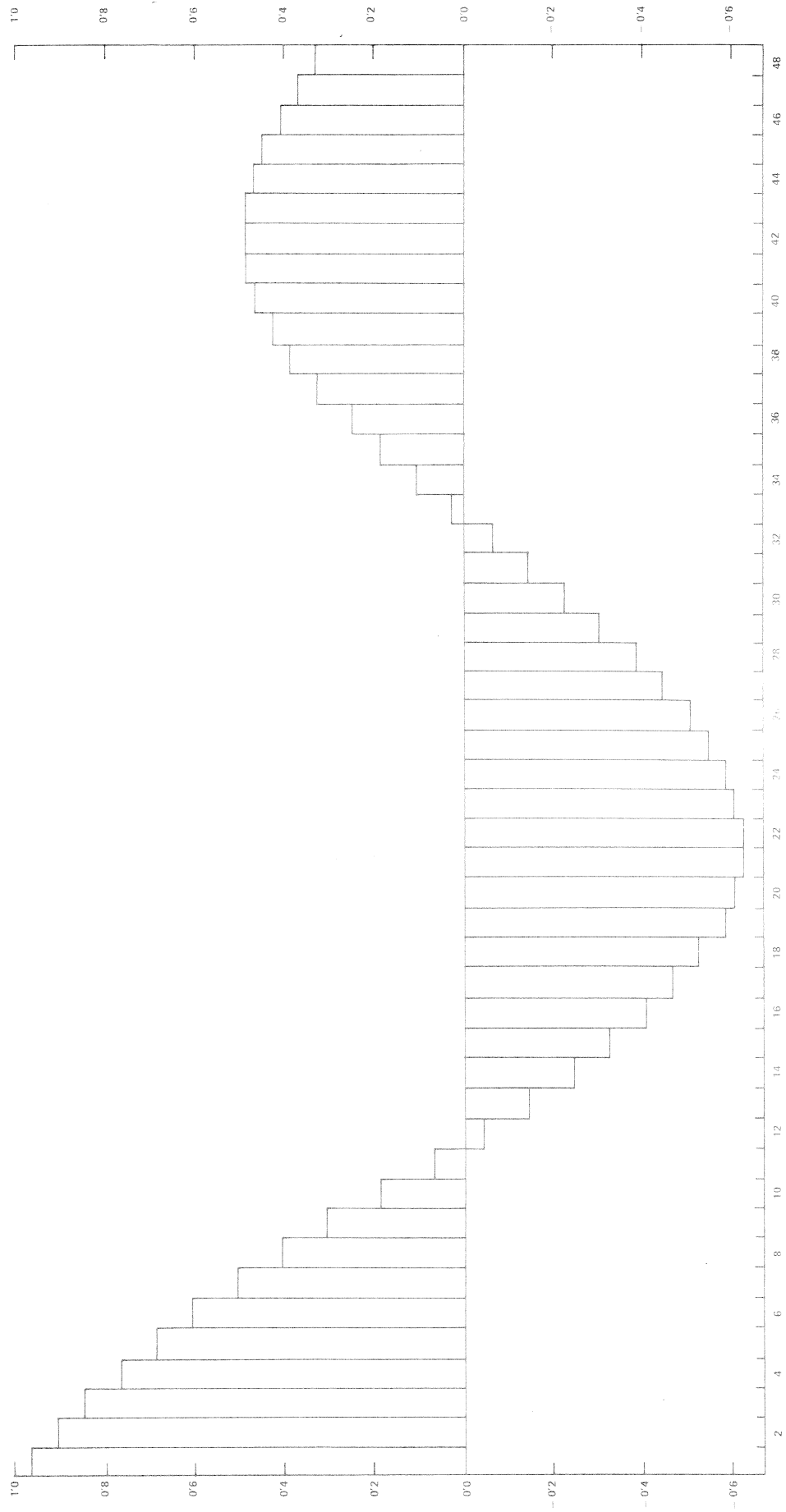


GRAFICO 10.

CORRELOGRAMA DE (1 - L)¹² PINA*

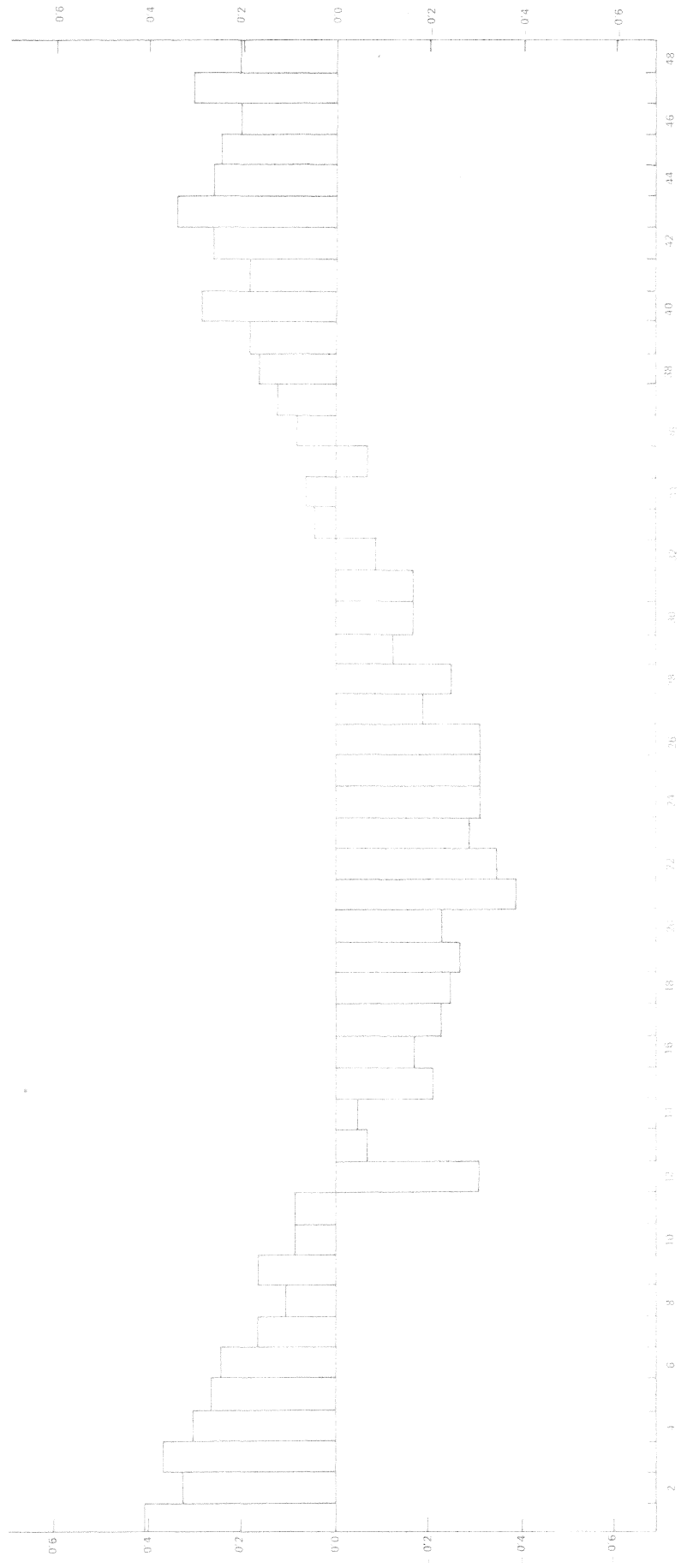


GRAFICO 10.A.
CORRELOGRAMA PARCIAL DE
(1 - L) (1 - L¹²) PRNA*

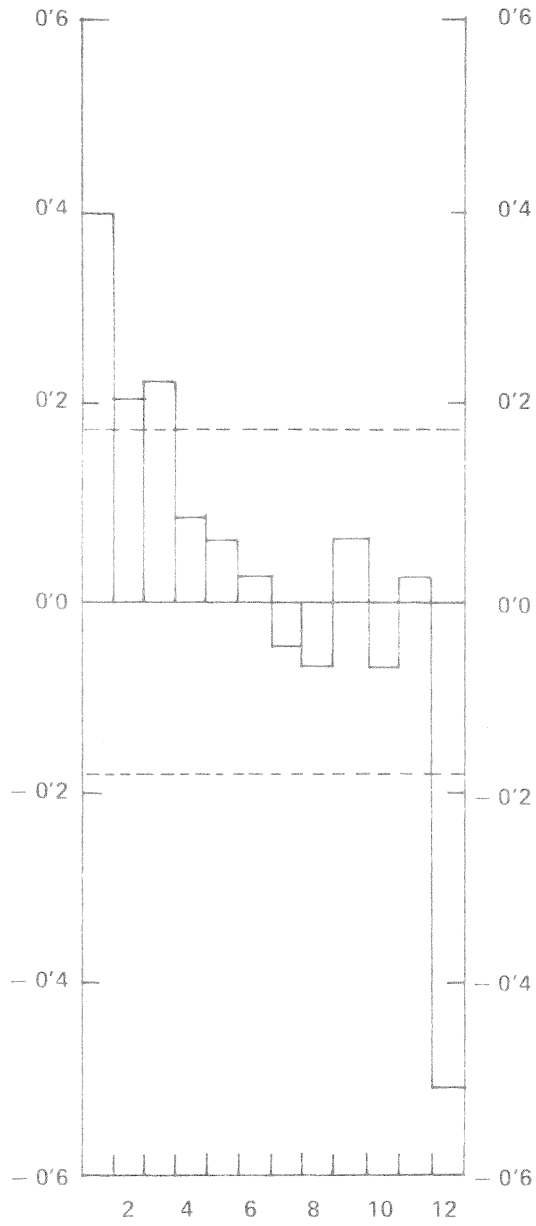
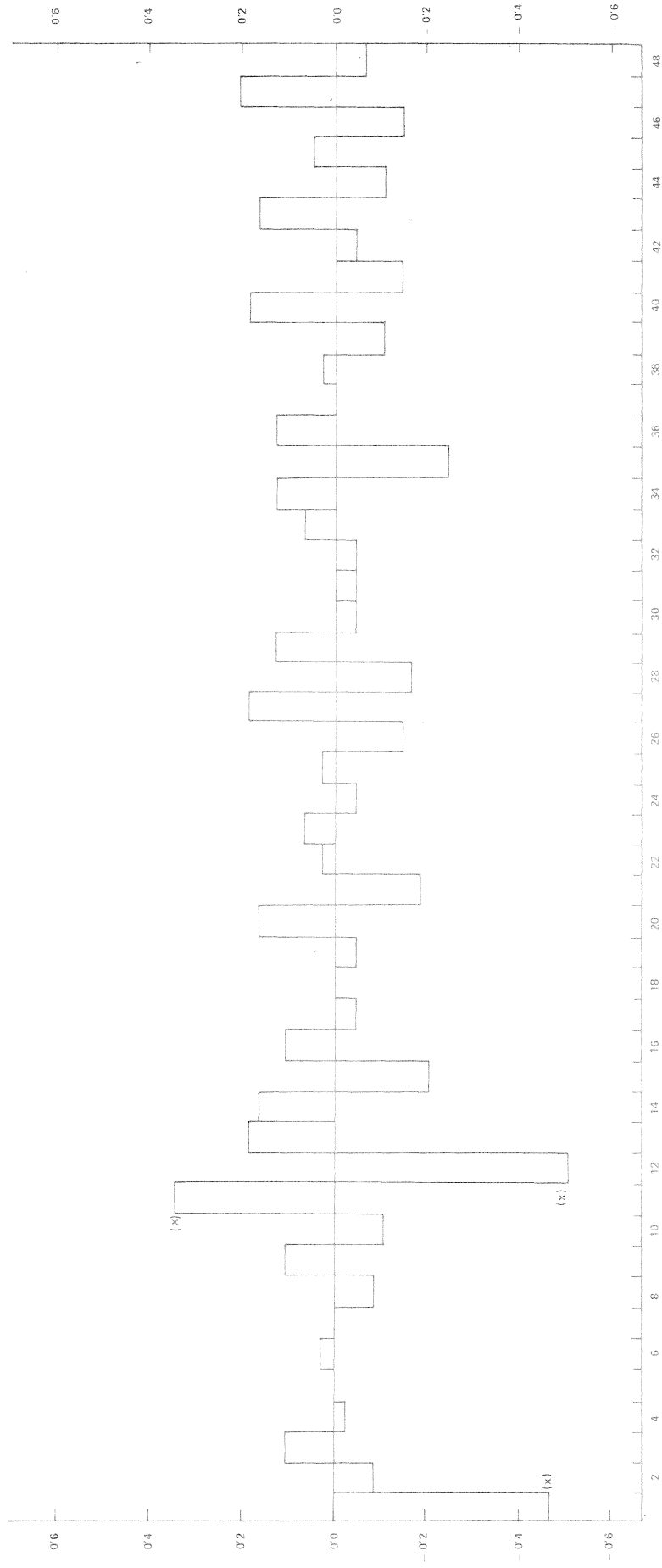


GRAFICO 11.

CORRELOGRAMA DE $(1 - L)^2 (1 - L^{12}) PRNA^*$



(x). Significativamente distinto de cero.

GRAFICO 11.A.

CORRELOGRAMA PARCIAL DE
 $(1 - L)^2 (1 - L^{12})$ PRNA*

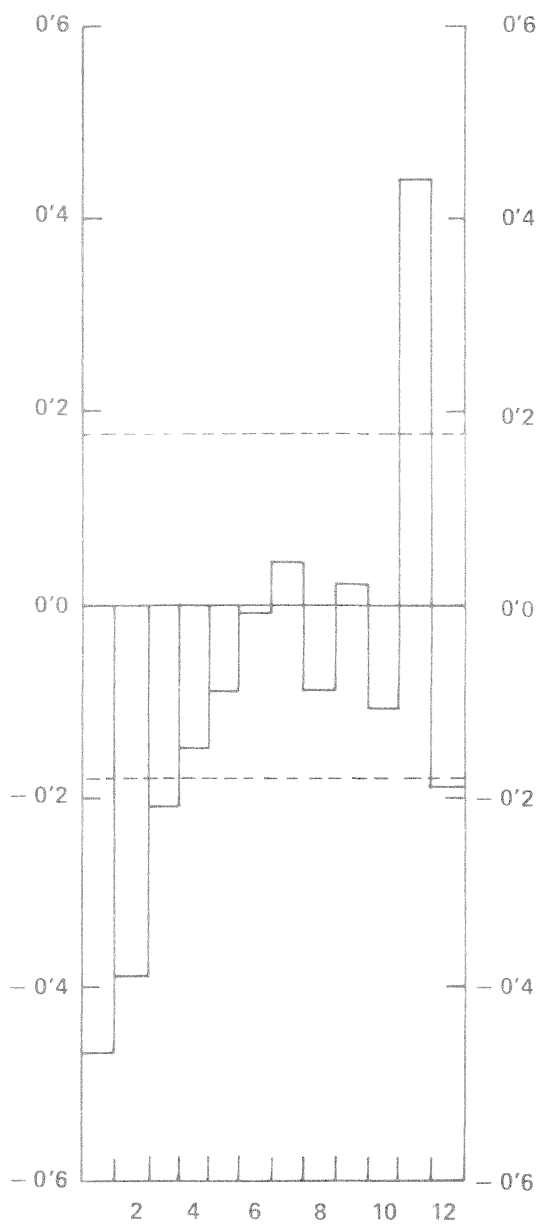
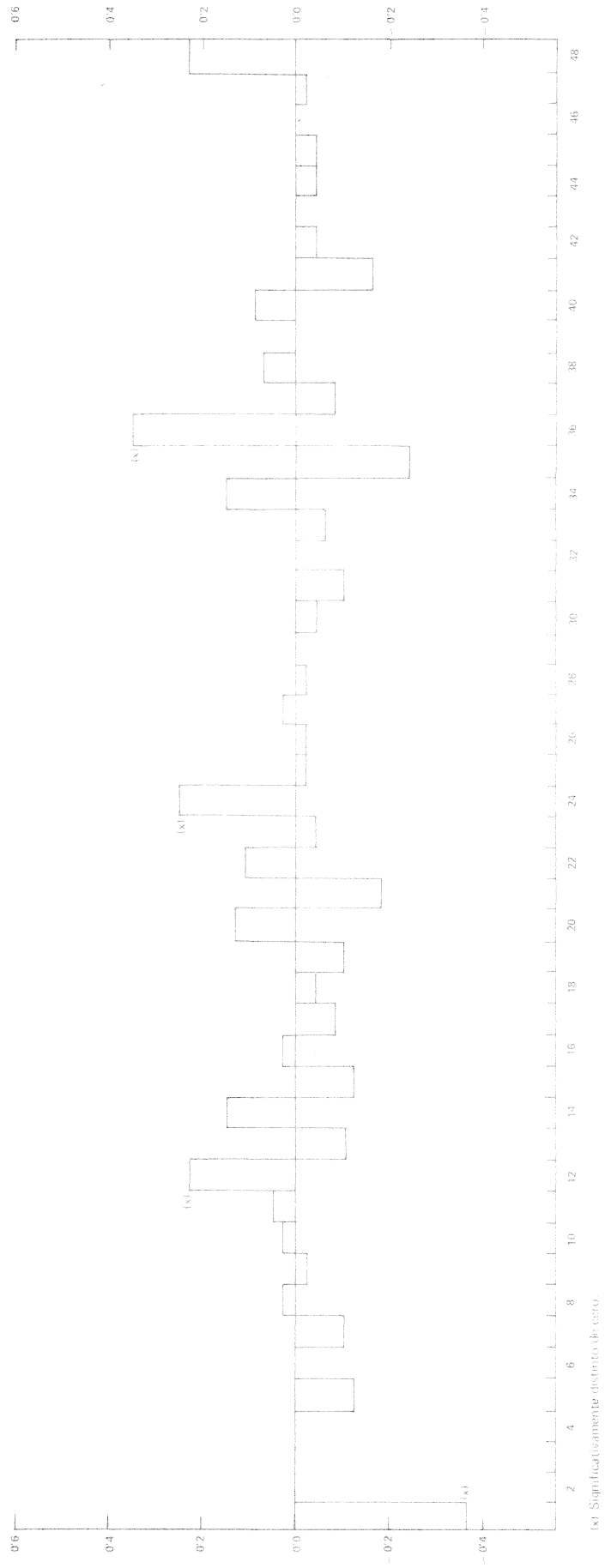


GRAFICO 12

CORRELOGRAMA DE $(1 - L)^2 PRINA$



(*) Significativamente diferente de zero.

V.2.- ESTIMACION Y PREDICCIÓN

Una vez indentificados s , d y d_1 tenemos que pasar a identificar p , p_1 , q y q_1 en base al correlograma y correlograma parcial correspondientes. Si partimos de $(1-L)(1-L^{12})PRNA^*$ como la serie estacionaria, su correlograma (gráfico 10) nos indica la presencia de un proceso AR regular que de acuerdo con el correlograma parcial (gráfico 10-A) se puede identificar como de orden tres. La estructura estacional a la vista del gráfico 10 la podemos identificar inicialmente como un proceso MA de primer orden. Con ello el modelo inicial sería

$$(1-\phi_1L-\phi_2L^2-\phi_3L^3)(1-L)(1-L^{12})PRNA^* = (1-\theta_{12}L^{12})a_t \quad (5)$$

Si partimos de $(1-L)^2(1-L^{12})PRNA^*$, a la vista de los gráficos 11 y 11-A el modelo inicial sería:

$$(1-L)^2(1-L^{12})PRNA^* = (1-\theta L)(1-\theta_{12}L^{12})a_t \quad (6)$$

El proceso de estimación se comenzó con el modelo (5). Los valores estimados para los coeficientes con sus valores t entre paréntesis fueron: $\phi_1=0,36(4,0)$, $\phi_2=0,10(1,0)$, $\phi_3=0,22(2,4)$ y $\theta_1 = 0,87(28,5)$. Las raíces correspondientes al polinomio AR fueron $0,82$ y $-0,23 \pm 0,47i$. Las raíces complejas generan un ciclo de $3,10$ meses de período, cifra ésta que no tiene gran sentido a la vista del comportamiento histórico de la serie. De acuerdo con las raíces obtenidas, el polinomio AR(3) se puede descomponer de la siguiente forma: $(1-.36L-.10L^2-.22L^3) = (1-.82L)(1+.46L+.27L^2)$, donde se observa que el componente AR(2) se puede aproximar por un proceso MA(1) del tipo $(1-.5L)$ así pues se pasó a estimar el modelo

$$(1-\phi_1L)(1-L)(1-L^{12})PRNA^* = (1-\theta_1L)(1-\theta_{12}L^{12})a_t \quad (7)$$

obteniéndose para ϕ_1 el valor de $0,92$ que no era significativamente distinto de uno. Así pues se pasó a estimar el modelo (6). Los resultados obtenidos fueron: $\theta_1=0,66(9,5)$ y

$\theta_{12} = 0,88(29,8)$; el error estándar de los residuos fué 0,02878. En base al correlograma de los residuos no se rechaza la hipótesis de ruido blanco tanto en base a los test t para cada retardo como en base al test Box-Pierce sobre los retardos 12, 24 y 36, que toma los valores 7,4; 20,0; y 30,3 respectivamente. Para validar definitivamente el modelo (6) se estimó el siguiente modelo más amplio:

$$(1 - \phi_{12}L^{12} - \phi_{24}L^{24}) (1-L)^2 (1-L^{12}) PRNA^* = (1 - \theta L) a_t \quad (8)$$

En él se observó que el polinomio AR(2) genera un ciclo de 2,97 años de período lo que parece bastante consistente con la muestra. Sin embargo, en el correlograma de los residuos el retardo 36 era significativo. Esto nos llevó a ensayar con el modelo

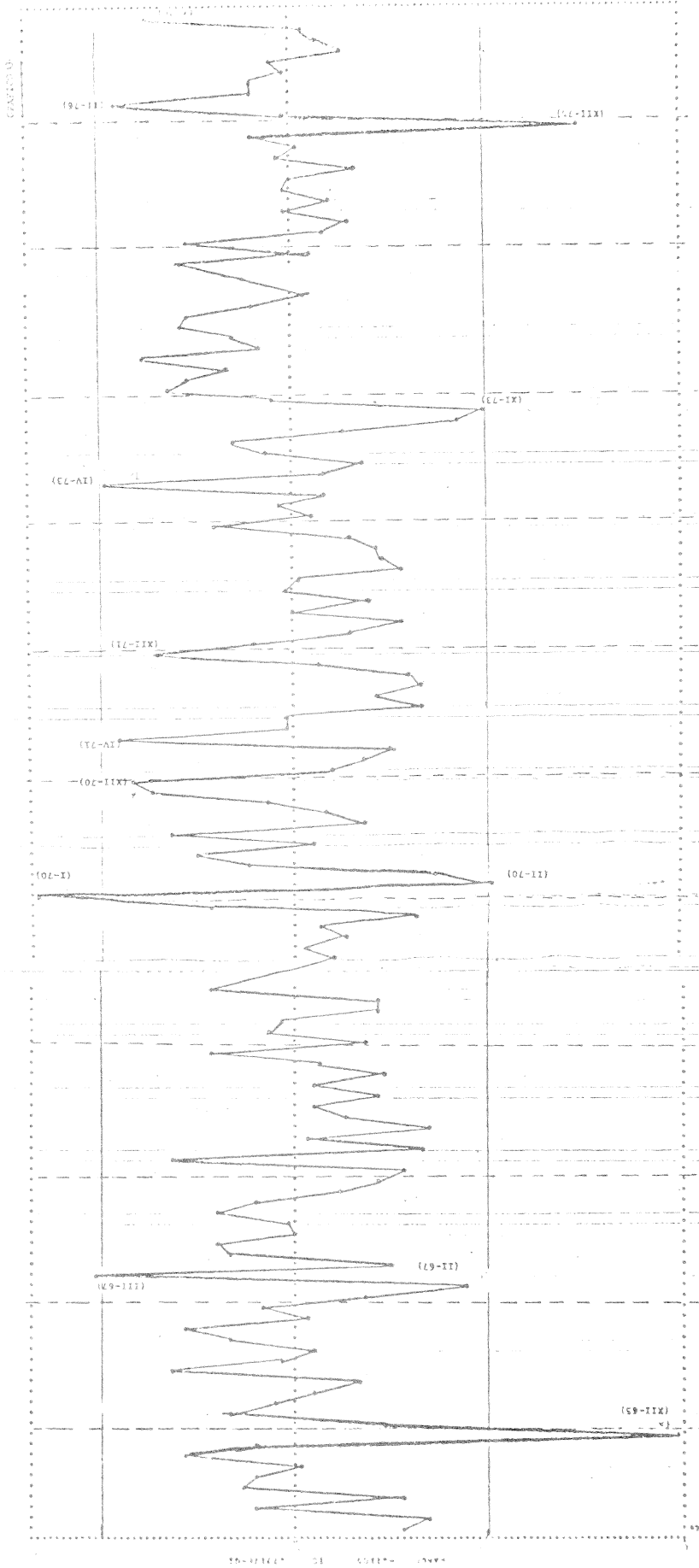
$$(1 - \phi_{12}L^{12} - \phi_{24}L^{24}) (1-L)^2 (1-L^{12}) PRNA^* = (1 - \theta L) (1 - \theta_{12}L^{12}) a_t \quad (9)$$

que se mostró como el modelo más adecuado y fué por tanto el elegido. Los modelos comentados se reestimaron extendiendo el período muestral hasta octubre de 1976 y fué también el modelo (9) el elegido. Los resultados de esta estimación fueron

$$(1 + \underbrace{0,17L^{12}}_{(2,03)} + \underbrace{0,25L^{24}}_{(3,06)}) (1-L)^2 (1-L^{12}) PRNA^* = (1 - \underbrace{0,70L}_{(11,4)}) (1 - \underbrace{0,86L^{12}}_{(27,4)}) a_t \quad (10)$$

$R^2 = 0,629$	Box-Pierce 12 = 6,9
	24 = 16,2
$\sigma = 0,02752$	36 = 24,7

las correlaciones entre los coeficientes todas fueron inferiores a 0,30. Las raíces de AR(2) son $0,085 \pm 0,49i$ lo que implica un ciclo de 3,85 años. Los residuos de este modelo se dan en el gráfico 13.



COUNTS

COUNTS PER MINUTE

Respecto a las propiedades predictivas del modelo tenemos que el error estándar en la predicción de un período hacia delante es de 2,75 % pero este error aumenta rápidamente a medida que aumenta el número de períodos por delante a predecir. En el cuadro 3 se dan estos errores estándar y se observa que para predicciones de tres intervalos por delante es del 6,3 % y para seis de 12,33%. Tenemos pues que el modelo puede ser adecuado para predecir hasta tres periodos por delante, menos adecuado para predicciones hasta cinco meses por delante y totalmente inadecuado para predicciones de mayor horizonte.

CUADRO 3

ERROR ESTANDAR EN LA PREDICCIÓN DE 1
PERIODOS POR DELANTE

Período	error estándar	Coefficientes ψ
1	2,75 %	1,30
2	4,52 %	1,60
3	6,31 %	1,91
4	8,21 %	2,21
5	10,21 %	2,51
6	12,33 %	2,81

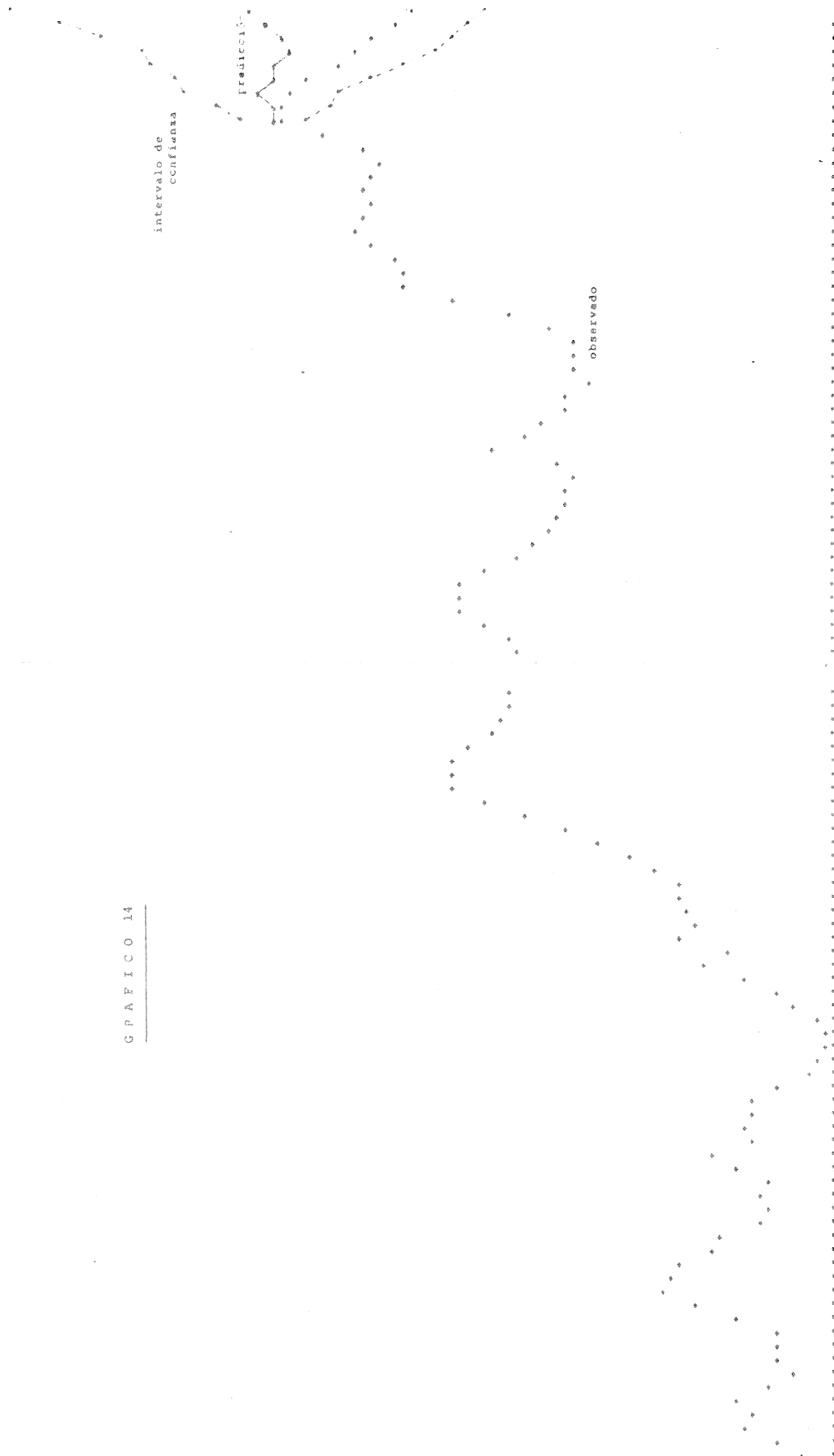
El modelo se muestra también inadecuado para predecir los puntos de cambio en los ciclos. Así, por ejemplo, si nos colocamos en diciembre de 1971, momento en que el paro está llegando al valor máximo del ciclo, y utilizamos el modelo para predecir hasta septiembre de 1972, observamos en el gráfico 14 que el modelo no capta la caída. Por el contrario, si nos situamos en diciembre de 1972 (en pleno descenso del paro) y utilizamos el modelo para predecir todo el año 1973, vemos en el gráfico 15 que obtenemos una predicción bastante ajusta

G R A F I C O 14

intervalo de
confianza

Predicción

observado



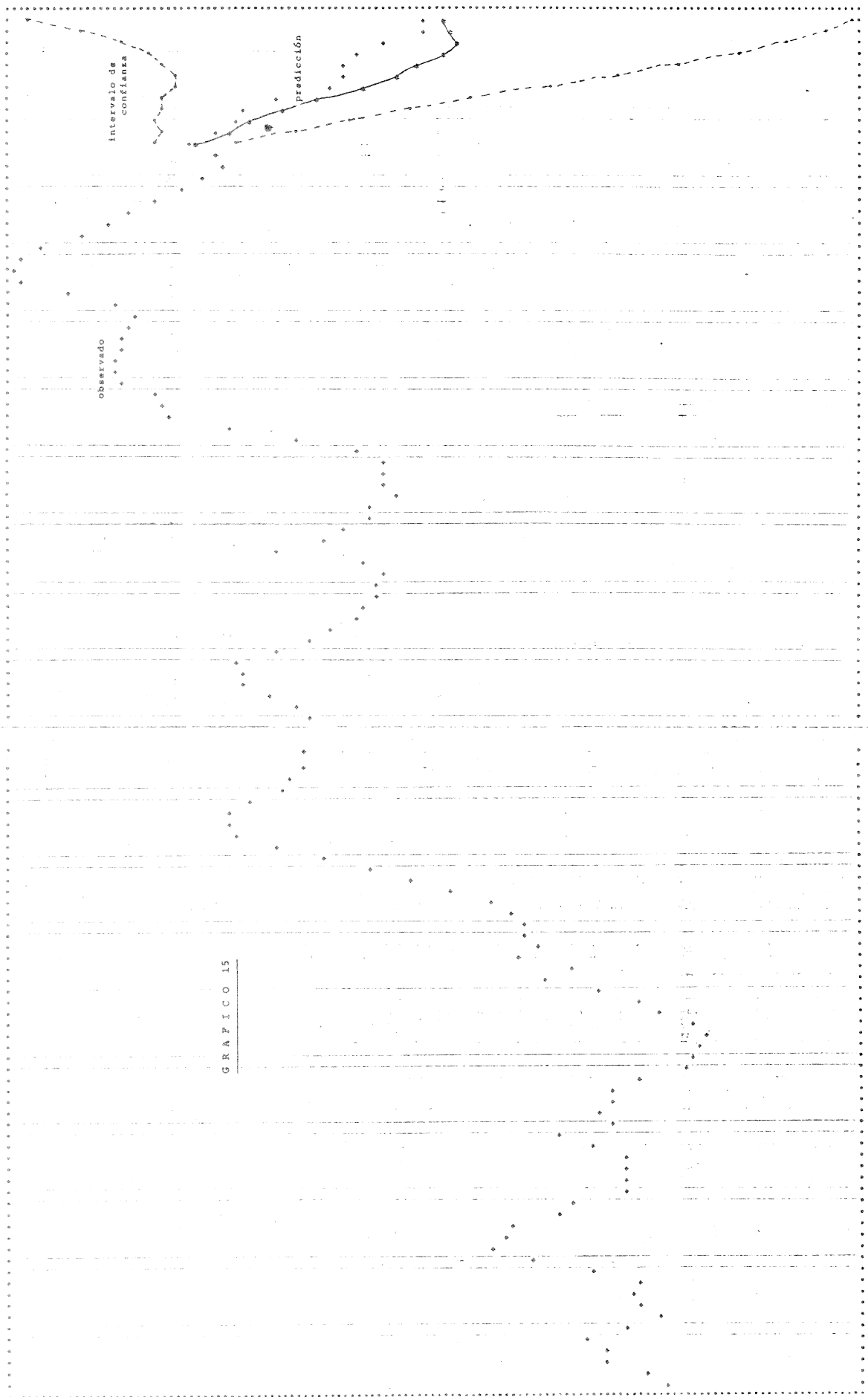


GRAFICO 15

da. Esto indica que el comportamiento cíclico del paro viene determinado por otras variables que habría de tener en cuenta. La conclusión de todo esto es que un modelo ARIMA es poco útil para predecir el paro y que debe pasarse a considerar modelos con funciones de transferencia que recojan las principales variables explicativas de éste.

Pasemos a comparar los resultados del modelo (10) con un modelo de regresión para el paro no agrícola estimado en Espasa (1976). En dicho trabajo se propone el siguiente modelo trimestral (con variables desestacionalizadas):

$$\text{PRNAT}^* = 30,14 + 0,037T - 2,52\text{PIB}^* + 0,95\text{PRNAT}^*_{-1} - 0,22\text{PRNAT}^*_{-4} + 0,48\text{WR}^*$$

$$R^2 = .9732 \quad \sigma = 0,0355 \quad \text{D.W.} = 1,48$$

donde PRNAT es el paro no agrícola trimestral

PIB es el producto interior bruto no agrícola

WR es el salario real para esos sectores

y el asterisco indica logaritmos

Vemos que el error en la predicción de PRNAT un trimestre por delante es de 3,6 %. Para comparar este error con el cometido con el modelo (10) necesitamos calcular el error en la predicción de la media trimestral con el modelo (10). Para ello hay que predecir tres periodos por delante con el modelo (10) y obtener su media. La varianza de esta media ($\overline{\text{PRNA}}$) será

$$\text{Var}(\overline{\text{PRNA}}) = 1/9 \{ \text{Var}(\text{PRNA}_1) + \text{Var}(\text{PRNA}_2) + \text{Var}(\text{PRNA}_3) + 2\text{Cov}(\text{PRNA}_1, \text{PRNA}_2) + 2\text{Cov}(\text{PRNA}_1, \text{PRNA}_3) + 2\text{Cov}(\text{PRNA}_2, \text{PRNA}_3) \} \quad (12)$$

donde PRNA_j es la predicción j periodos por delante de PRNA con el modelo (10).

Para calcular los valores de los componentes de la fórmula (12) necesitamos conocer los valores de los coeficientes ψ_j , que son los coeficientes que se obtienen a partir del modelo (10) expresado PRNA* de la forma

$$\text{PRNA}^* = \psi_1 a_t + \psi_2 a_{t-1} + \dots,$$

estos coeficientes se dan en la columna 3 del cuadro 3.

Las fórmulas para calcular los distintos componentes de (12) se dan en Box-Jenkins (1976), capítulo 5. Aplicándolas obtenemos:

$$\text{Var}(\text{PRNA}_2) = 2,695 \sigma_a^2$$

$$\text{Var}(\text{PRNA}_3) = 5,268 \sigma_a^2$$

$$\text{Cov}(\text{PRNA}_1, \text{PRNA}_2) = 1,3 \sigma_a^2$$

$$\text{Cov}(\text{PRNA}_1, \text{PRNA}_3) = 1,6 \sigma_a^2$$

$$\text{Cov}(\text{PRNA}_2, \text{PRNA}_3) = 3,38 \sigma_a^2$$

$$\begin{aligned} \text{Así } \text{Var}(\overline{\text{PRNA}}) &= 1/9(1+2,695+5,268+2,6+3,2+6,76) \sigma_a^2 = \\ &= 2,39 \times 0,000757 = 0,001809 \end{aligned}$$

Así el error estandar en la predicción de $\overline{\text{PRNA}}$ con (10) es del 4,25 %, que es ligeramente superior al 3,6 que se obtiene con el modelo de regresión mencionado.

CONCLUSION

Las principales conclusiones del análisis univariante realizado en este trabajo son:

- 1.- La estimación de la tendencia por medio de polinomios temporales es bastante imprecisa y, generalmente, inadecuada. El método de diferenciaciones sucesivas parece más correcto para series económicas.
- 2.- Con frecuencia las series económicas no tienen una varianza homogénea a lo largo del período muestral. En dichos casos conviene aplicar una transformación del tipo Cox-Box que convierta a la serie original en homogénea.
- 3.- La desestacionalización de las series económicas por el método X-11 con frecuencia elimina de la serie original más variación que la puramente estacional. Las técnicas espectrales son muy útiles para estudiar los efectos de la desestacionalización. En el caso del PRNA tanto la desestacionalización de la serie agregada como la obtenida a través de la desestacionalización de los componentes son inadecuadas. La razón de esto está en que el método X-11 supone un modelo Arima específico (véase Espasa (1977a)) para las series que trata y este modelo es distinto al que según los resultados de la sección V parece seguir el PRNA.
- 4.- Un marco suficientemente amplio para modelar las series económicas es el de los modelos Arima (véase Espasa (1977 b)). En este trabajo se ha estimado un modelo Arima para el PRNA que deja un error residual con una desviación estándar del 2,75 %. El modelo contiene un factor cíclico con período de 3,85 años. No obstante el modelo no ^{es} adecuado para captar convenientemente los cambios en los ciclos

del PRNA. Estos pues parecen debidos a factores exógenos. La consecuencia es que el análisis univariante es poco útil para el estudio del PRNA y se debe pasar a modelos del tipo de funciones de transferencia, en los que se relacionan variables entre sí. No obstante los resultados del modelo Arima estimado en este trabajo no son mucho peores a los obtenidos con un modelo de regresión desarrollado anteriormente por el autor. Antes de pasar a considerar un modelo que relaciona distintas variables entre sí convendrá decidir mediante principios de teoría económica si la variable a considerar es el paro o si más bien son el empleo y la mano de obra disponible, obteniéndose el paro de forma residual.

BIBLIOGRAFIA

- ANDERSON, O.D., 1976, Time Series Analysis and Forecasting, The Box-Jenkins Approach, Butterworths, London
- ANDERSON, T.W., 1971, The Statistical Analysis of Time Series. John Wiley.
- BOX, G.E.P. y D.R. COX, 1964, "An Analysis of Transformations" Journal of the Royal Statistical Society, serie B, págs. 211-252.
- BOX, G.E.P. y G.M. JENKINS, 1976, Time Series Analysis Forecasting and Control, Holden Day, 2° edición
- CHATFIELD, C., 1975, The Analysis of Time Series: The Theory and Practice, Chapman y Hall, London .
- CHATFIELD, C y D.L. PROTHERO , 1973, "Box-Jenkins Seasonal Forecasting: Problemx in a Case Study", Journal of the Royal Statistical Society, Serie A págs. 295-310.
- ESPASA, A., 1976, "La ecuación agregada de paro: una primera aproximación", Banco de España, trabajo no publicado.
- ESPASA, A., 1977a, "El problema de la desestacionalización de las series económicas: Métodos utilizados y su interpretación", Boletín de estudios económicos, V. XXXII, n°101, págs. 461-478. Agosto.
- ESPASA, A., 1977 b, "La predicción económica ". Banco de España, trabajo no publicado.
- JENKINS, G.H. y D.G. WATTS, 1969, Spectral Analysis and its Applications, Holden Day.
- NELSON, C.R., 1976, "The Interpretation of R^2 in Autoregressive-Moving Average Time Series Models", The American Statistician, Noviembre, págs. 175-180.

SARGAN, J.D., 1964, "Wages and Prices in the United Kingdom: a Study in Econometric Methodology" in Econometric Analysis for National Economic Planning, de P.E. Hart y otros, Colston paper, Butterworths, London.

SHISKIN, J.; A.H. YOUNG y J.C. MUSGRAVE, 1967, "The X-11 variant of the Census Method II Seasonal Adjustment Program", Bureau of the Census Technical Paper, n° 15 (revisado), Washington, D.C., U.S., Department of Commerce.

APENDICE 1

CORRELOGRAMAS Y ESPECTROS DE PRA, PRI, PRS, PRC

PRAD, PRID, PRSD y PRCD

En los gráficos de los correlogramas las bandas establecidas alrededor de cero son:

$$0 \pm 2 \times (1/\sqrt{T}) \quad (A1.1.)$$

donde T es el número de observaciones.

La desviación típica para cada coeficiente de autocorrelación sería $1/\sqrt{T}$ si las series representadas fueran aleatorias. En consecuencia, los valores fuera de dichas bandas se pueden tomar como significativamente distintos de cero.

En los gráficos de los espectros los intervalos de confianza se han calculado mediante la fórmula (Ver Jenkins y Watts (1969), sección 6.4):

$$\log_{10} F_x(\omega) + \log_{10} \frac{\nu}{\chi_{\nu}^2 (1-(\alpha/2))},$$

$$\log_{10} F_x(\omega) + \log_{10} \frac{\nu}{\chi_{\nu}^2 (\alpha/2)}$$

donde $F_x(\omega)$ es el valor estimado del espectro para la frecuencia ω , ν es el número de grados de libertad y χ_{ν}^2 es el valor tabulado de la distribución χ_{ν}^2 cuadrado. El valor utilizado de

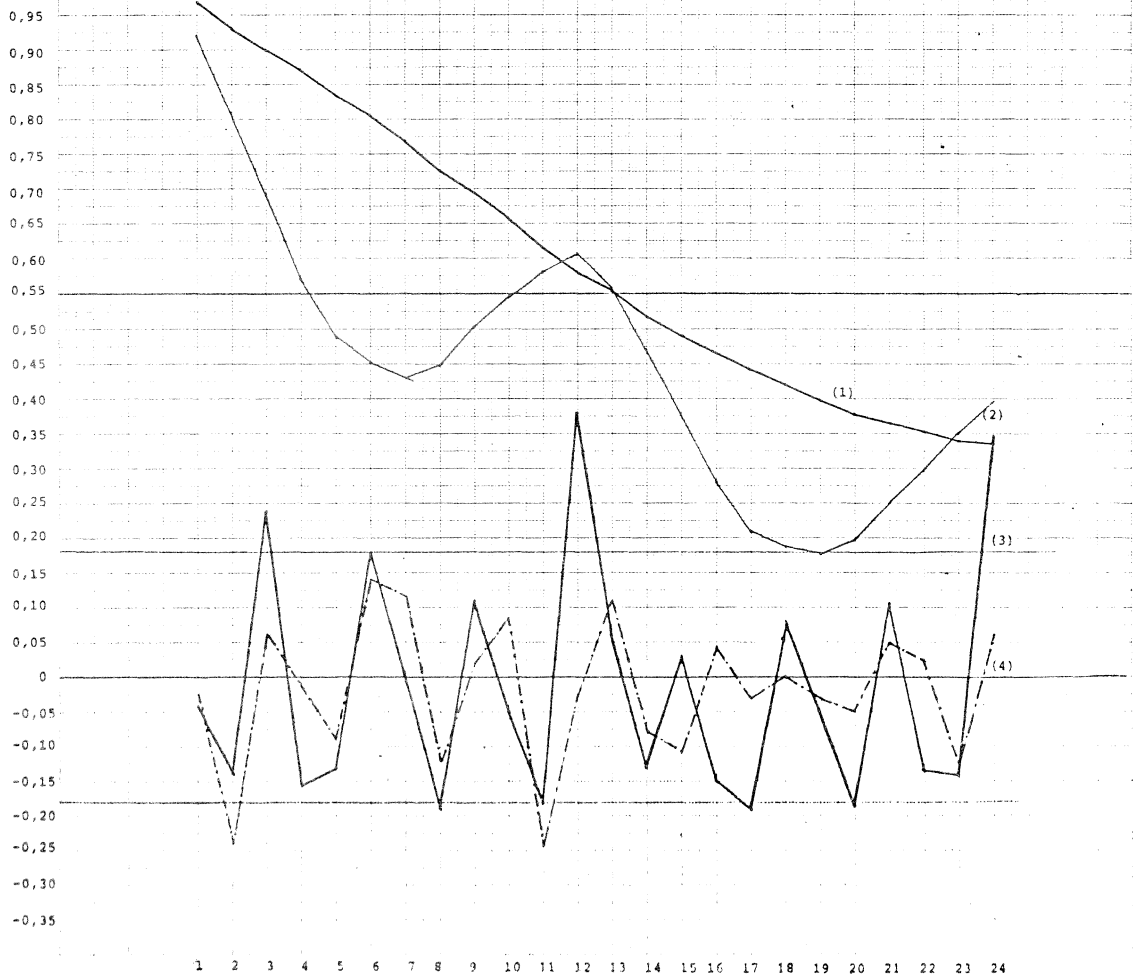
α es 0,01 y los grados de libertad (ν) en la ventana de Parzen son $3.71 T/M$ (Jenkins y Watts (1969) pág. 252) donde T es el número de observaciones y M el número de covarianzas empleadas en la estimación del espectro. En nuestro caso, M es 24 y T 131, excepto para PRA que es 123. En consecuencia,

$$\log_{10} \frac{\nu}{\chi_{\nu}^2 (.995)} = \begin{cases} -.2956 & \text{para PRI, PRC y PRS} \\ -.3073 & \text{para PRA} \end{cases}$$

$$\log_{10} \frac{\nu}{\chi_{\nu}^2 (.005)} = \begin{cases} .4352 & \text{para PRI, PRC y PRS} \\ .4438 & \text{para PRA} \end{cases}$$

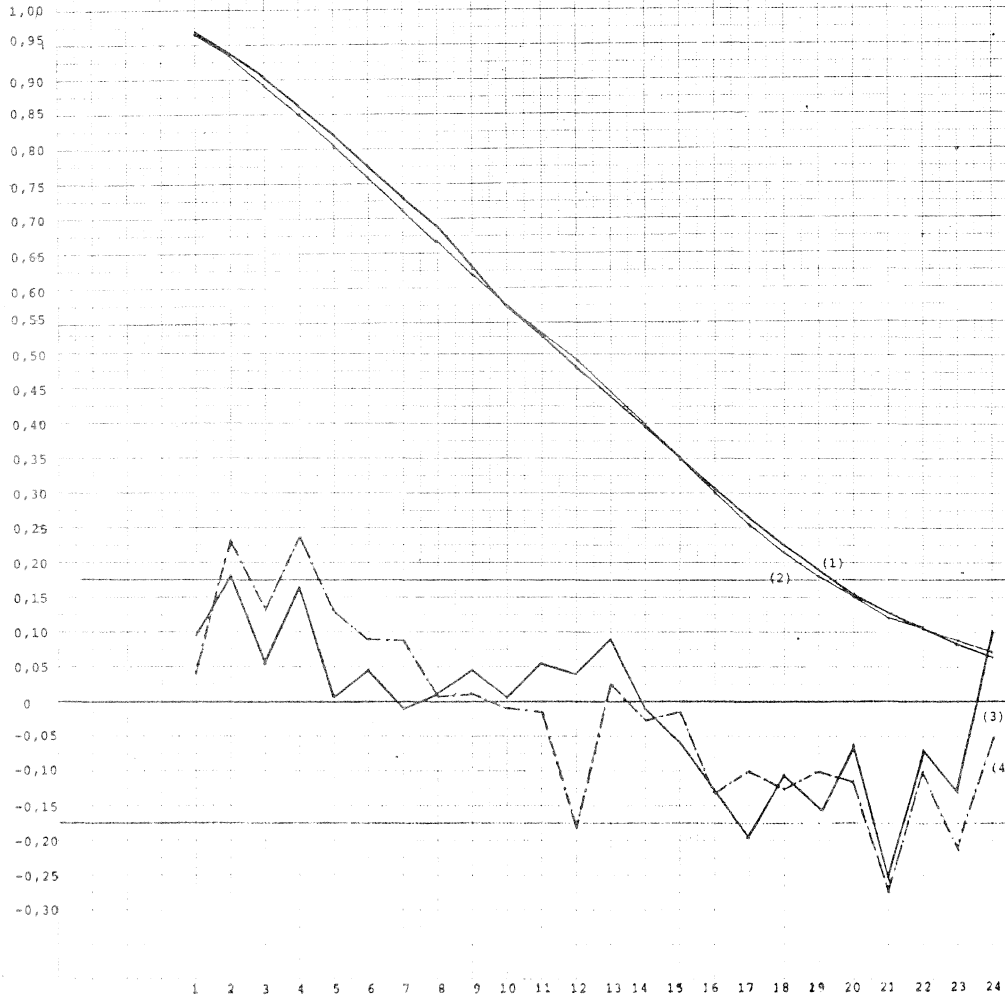
CORRELOGRAMA DEL PARO REGISTRADO AGRICOLA (PRA)

- (1) Serie original
- (2) Serie desestacionalizada (PRAD)
- (3) $(1 - 1.16L + .31L^2 - .11L^9)PRA$
- (4) $(1 - .97L)PRAD$



CORRELOGRAMA DEL PARO REGISTRADO EN LA INDUSTRIA (PRI)

- (1) Serie original
- (2) Serie desestacionalizada (PRID)
- (3) (1 - .968L)PRI
- (4) (1 - .970L)PRID



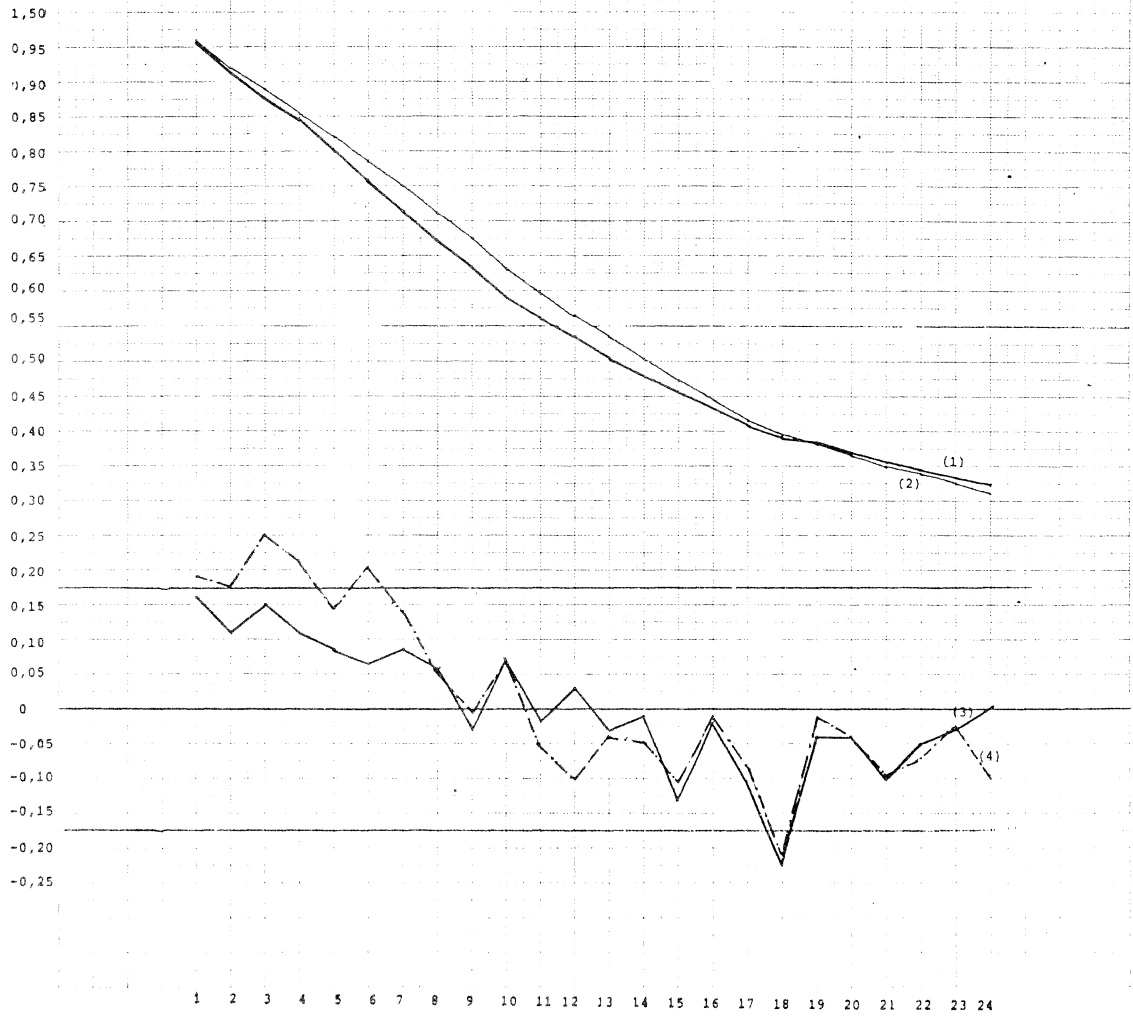
CORRELOGRAMAS DEL PARO REGISTRADO EN LA CONSTRUCCION (PRC)

- (1) Serie original
- (2) Serie desestacionalizada (PRCD)
- (3) $(1 - .964L)PRC$
- (4) $(1 - .971L)PRCD$

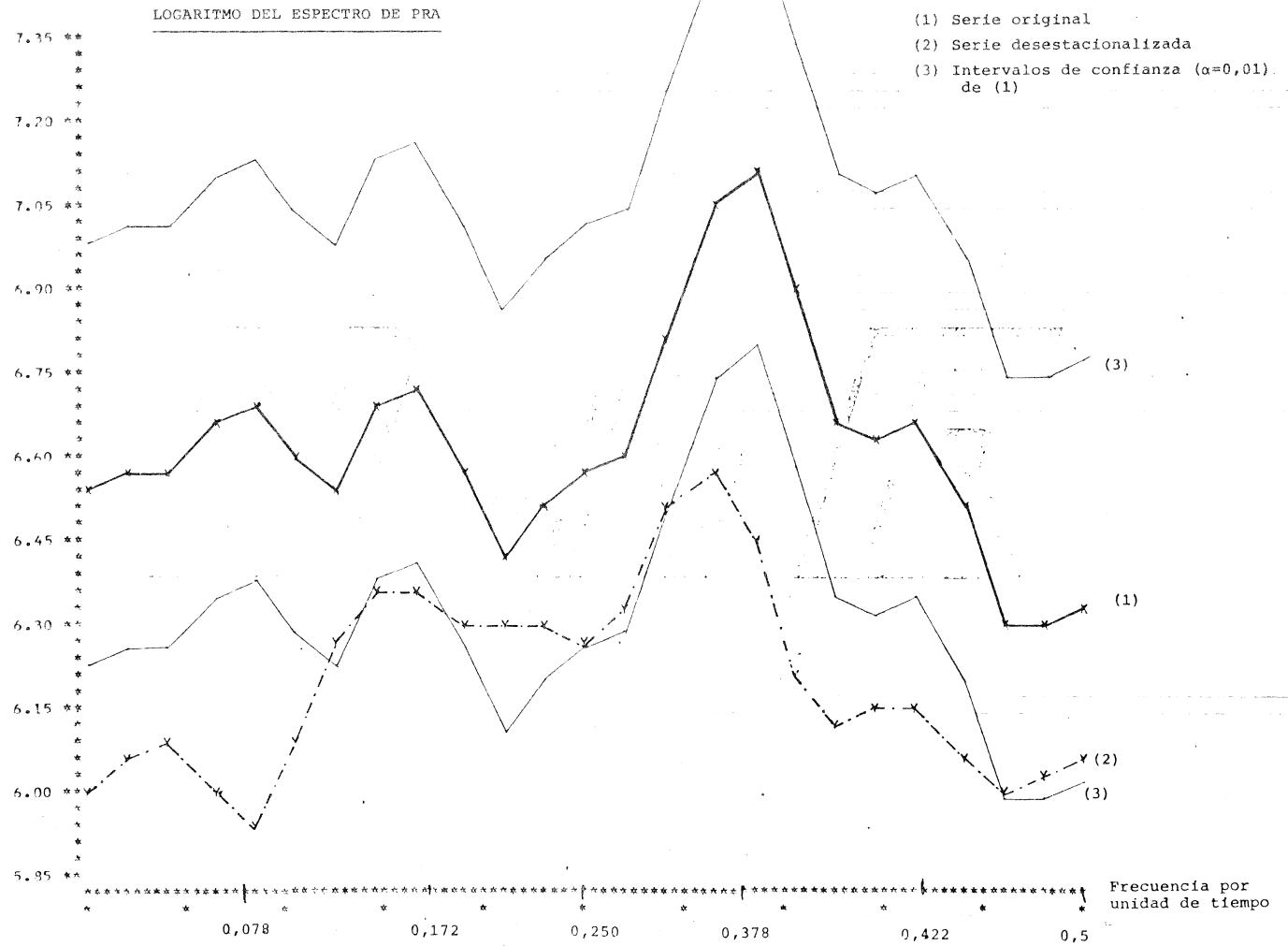


CORRFLOGRAMA DEL PARO REGISTRADO EN LOS SERVICIOS (PRS)

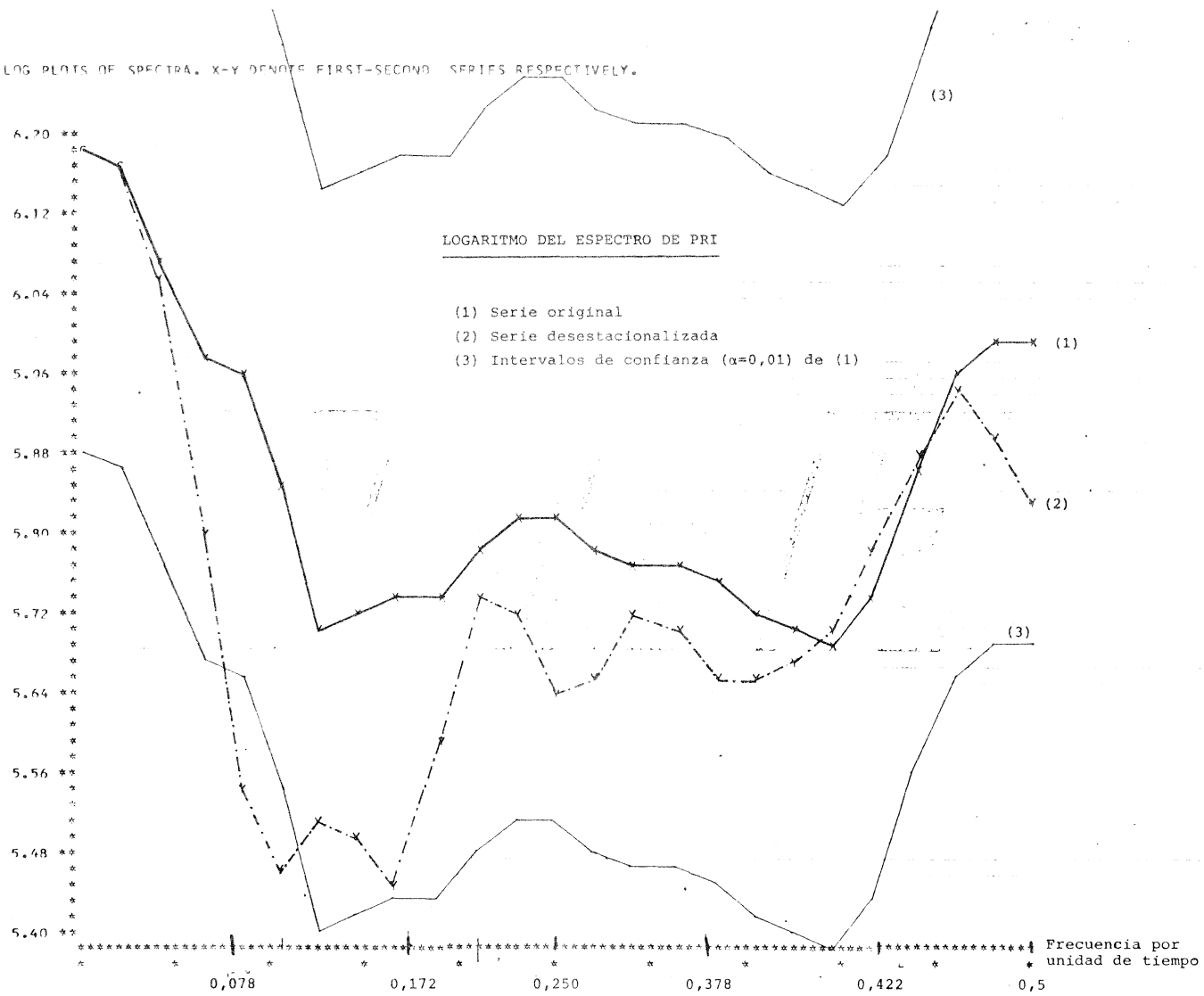
- (1) Serie original
- (2) Serie desestacionalizada (PRSD)
- (3) $(1 - .959L)PRS$
- (4) $(1 - .955L)PRSD$



SEMI LOG PLOTS OF SPECTRA. X-Y DENOTE FIRST-SECOND SERIES RESPECTIVELY.

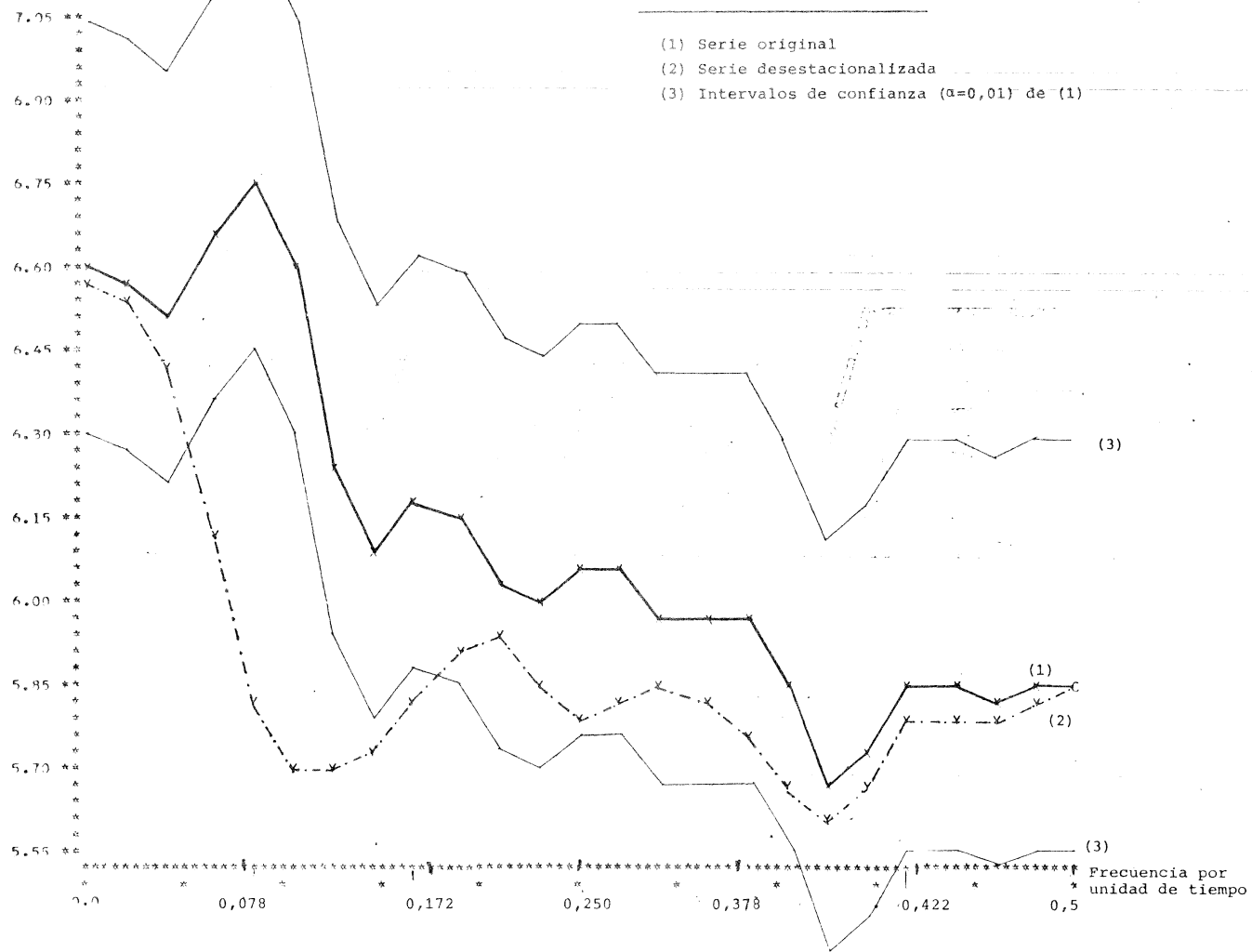


SEMI LOG PLOTS OF SPECTRA. X-Y DENOTE FIRST-SECOND SERIES RESPECTIVELY.



SEMI LOG PLOTS OF SPECTRA. X-Y DENOTE FIRST-SECOND SERIES RESPECTIVELY.

LOGARITMO DEL ESPECTRO DE PRC



SEMI LOG PLOTS OF SPECTRA X-Y DENOTE FIRST-SECOND SERIES RESPECTIVELY.

LOGARITMO DEL ESPECTRO DE PRS

