



Centro de Investigación y
Desarrollo (CIDE)

CONTROL DE CALIDAD EN LOS PROCESOS ESTADÍSTICOS

Una aproximación basada
en los modelos ARIMA con
Análisis de Intervención

Lima, Agosto 2002

DIRECCIÓN Y SUPERVISIÓN

Econ. Mirlena Villacorta Olazabal
Directora Técnica del CIDE

Documento Elaborado por:

Lic. Est. Fernando César Camones Gonzales

Preparado	: Centro de Investigación y Desarrollo del Instituto Nacional de Estadística e Informática (INEI)
Impreso	: Talleres de la Oficina Técnica de Administración del INEI
Diagramación	: Centro de Edición de la Oficina Técnica de Difusión del INEI
Tiraje	: 200 Ejemplares
Domicilio	: Av. General Garzón 658, Jesús María. Lima - Perú
Orden de Impresión	: N° -OTA-INEI
Depósito Legal N°	: 150113-2002-4013

Presentación

El INEI pone a disposición la investigación metodológica: "**CONTROL DE CALIDAD EN LOS PROCESOS ESTADÍSTICOS Una aproximación basada en los modelos ARIMA con Análisis de Intervención**", que por su nivel de especialización está dirigida principalmente a los miembros de la comunidad académica, profesionales de las oficinas de estadística y los investigadores interesados en mantener la actualidad de sus procedimientos estadísticos.

Esta investigación metodológica tiene por finalidad, generar instrumentos y procedimientos que permitirán validar, mejorar y actualizar los procesos estadísticos. Se caracteriza por ser innovadora en su campo de aplicación, por contener un rigor científico en su desarrollo integral, por la validez de sus procesos, por la vigencia y actualidad de sus metodologías aplicadas.

En tal sentido, la investigación que se presenta, busca contribuir con la generación e incorporación de mecanismos de control de calidad basada en métodos y modelos estadísticos que respondan a las expectativas de evaluar la calidad de la información proveniente de las estadísticas básicas; así, como a la información que utiliza nuestra institución.

Este estudio, al igual que otros de carácter metodológico ha sido elaborado por profesionales del Centro de Investigación y Desarrollo (CIDE), en el marco del desarrollo y promoción de investigaciones estadísticas y socioeconómicas que permitan elevar la calidad de la información del INEI y el SEN.

El INEI espera como resultado de esta investigación, incorporarse en el circuito de la producción del conocimiento y elevar los estándares de calidad de sus procesos, sentando con ello las bases de la investigación metodológica en la institución.

Lima, Agosto 2002



Gilberto Moncada Vigo
Jefe del INEI

INDICE

Presentación	3
Prólogo	7
Introducción	9
I. ASPECTOS PRELIMINARES	11
1.1 Fundamentación del problema de investigación	11
1.1.1 El planteamiento del problema	11
1.1.2 La formulación del problema	12
1.1.3 La sistematización del problema	12
1.2 Objetivos de investigación	13
1.2.1 Objetivo general	13
1.2.2 Objetivos específicos	13
II. MARCO CONCEPTUAL	15
2.1 Las variables macroeconómicas	15
2.2 Búsqueda de un marco conceptual adecuado: los ARIMA con análisis de intervención	16
III. FORMULACION Y OPERACIONALIZACION DE LA HIPOTESIS	19
3.1 Formulación de la hipótesis	19
3.2 Operacionalización de la hipótesis	19
IV. METODOLOGIA DE MODELIZACION Y CONTROL DE CALIDAD EN VARIABLES MACROECONOMICAS	21
4.1 Identificación	21
4.1.1 Análisis exploratorio	21
4.1.2 Identificación de la estructura del modelo	24
4.2 Estimación de los parámetros del modelo	25
4.3 Diagnóstico del modelo	27
4.4 Análisis de intervención	28
4.5 Predicción	29
4.6 Control estadístico de calidad basado en los modelos	30

V. MEDICION, ANALISIS Y RESULTADOS	31
5.1 Descripción de las variables macroeconómicas	31
5.2 Desarrollo metodológico	32
5.2.1 Procedimiento de Modelización Automática (MODAUT)	32
5.2.2 Procedimiento de modelización con el programa Eviews 3.1	37
5.2.3 Pautas para identificar un perfil de las series macroeconómicas que no se adecuan a los ARIMA	58
VI. DISCUSION, CONCLUSIONES Y RECOMENDACIONES	63
6.1 Conclusiones	63
6.2 Recomendaciones	64
VII. REFERENCIAS BIBLIOGRAFICAS	65
ANEXOS	67
1. Modelización automática y control de calidad: Índices del PBI	68
2. Modelización automática y control de calidad: Volumen de producción Minera	75

Prólogo

La importancia de la utilización de procedimientos cuantitativos metodológicamente fiables y contrastados en todas las fases del trabajo estadístico es bien conocida y frecuentemente subrayada. En particular, la utilización de métodos y modelos estadísticos en el desarrollo del control de la calidad de la información de los censos y encuestas, es una de las materias de mayor actualidad en las reuniones internacionales de expertos estadísticos. Sin embargo, en estos momentos son todavía escasos los países cuyas oficinas de estadística utilizan estas técnicas de forma habitual.

Por este motivo, los trabajos emprendidos en el INEI, cuyas líneas directoras se muestran en el presente documento, resultan particularmente innovadores y en línea con los que se vienen desarrollando en las instituciones estadísticas más prestigiosas. En este sentido, es de resaltar el esfuerzo realizado por Fernando Camones, dentro del Centro de Investigación y Desarrollo (CIDE), para introducir esta metodología entre las herramientas de trabajo del citado organismo.

El documento presenta en primer lugar la formulación del problema al que se enfrenta el responsable del control de la calidad en las encuestas o censos que recogen información de variables cuantitativas y que se realizan de forma

continua, es decir, periódicamente. A continuación muestra los fundamentos teóricos rigurosos que sirven de base para la utilización de los modelos ARIMA univariantes con análisis de intervención (ARIMA-AI) en los trabajos del control de calidad. En particular, presenta algunas herramientas definidas a partir de las predicciones un periodo por delante obtenidas de los modelos. En la última parte del documento muestra algunos ejemplos prácticos de la utilización de los métodos en observaciones de series temporales de la economía peruana.

La utilización de los modelos de tipo econométrico en una oficina central de estadística, en particular en una oficina centrada en la coordinación y el análisis como puede ser el INEI, es un elemento esencial para un amplio conjunto de finalidades. Entre ellas, y en línea con los trabajos desarrollados por el autor de este artículo, pueden citarse las siguientes:

- a. Utilización de los modelos para el análisis de coyuntura. La idea central de su uso proviene del hecho de que los modelos ARIMA con análisis de intervención describen el comportamiento dinámico de las variables estudiadas. Así, mediante la parte estocástica de los modelos, pueden estudiarse las tendencias, las variaciones estacionales, las oscilaciones cíclicas, el grado de volatilidad o impredecibilidad, etc. Por

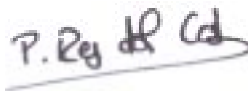
otra parte, mediante el análisis de intervención, pueden establecerse hipótesis sobre determinados acontecimientos económicos, que podrán contrastarse por medio de los modelos, pudiéndose comprobar si los datos validan o rechazan las hipótesis establecidas. Al mismo tiempo, puede llegarse a una medición empírica de los efectos ocasionados por determinados acontecimientos económicos y físicos, o por cambios metodológicos en la construcción de las series.

- b. Utilización de los modelos para mejorar las estimaciones del ajuste estacional. Los modelos obtenidos por estos procedimientos pueden utilizarse para mejorar los modelos de "default" que proporcionan los programas de ajuste (por ejemplo, el X12-ARIMA). En particular, las estimaciones puede mejorarse notablemente cuando se lleva a cabo un minucioso análisis de intervención y un cuidadoso tratamiento de los valores atípicos.
- c. Utilización de los modelos para construir herramientas de control estadístico de calidad. El principal uso de los modelos consiste en construir estas herramientas, las cuales pueden utilizarse para complementar los métodos de control de calidad basados en la experiencia y en el conocimiento de los diferentes sectores económicos. Entre los objetivos de estas herramientas de control estadístico debe señalarse los siguientes: detectar con mayor precisión los posibles

errores, integrar las fases del control de calidad, seleccionar los datos a depurar de acuerdo a su importancia en el cálculo de los agregados y establecer criterios objetivos y fundamentados estadística mente de control de calidad.

La utilización de estas modernas tecnologías, cuyas ventajas parecen actualmente fuera de toda duda, en las instituciones que se dedican a la elaboración de estadísticas, tanto en el sector privado como en el público, se encuentra a veces con el obstáculo de la carencia de personal especializado con la suficiente formación. En este caso, los grandes conocimientos técnicos en materia estadística del autor de este trabajo le han permitido introducirse con facilidad en los fundamentos necesarios para su utilización.

Querría terminar este Prólogo felicitando a Fernando Camones por su enorme éxito en la culminación de un trabajo que con toda seguridad le ha supuesto una gran dedicación y esfuerzo y animándole a seguir en el camino emprendido de introducir modernas técnicas estadísticas en los trabajos de las encuestas.



Pilar Rey del Castillo
Directora del Departamento
de Banco de Datos
Centro de Investigaciones Sociológicas
Madrid

Introducción

A lo largo de todos estos años los encargados del análisis de coyuntura han tenido problemas en evaluar la calidad de la información reportada mensualmente por los encargados de la producción de las estadísticas básicas y estos a su vez, también tenían la necesidad de desarrollar mecanismos de control de calidad, que evalué la información reportada como resultado de las encuestas mensuales realizadas a las empresas, este hecho ha motivado a desarrollar hipótesis alternativas a las actuales sobre la evaluación tradicional de la calidad del dato basada por un lado en el conocimiento exhaustivo del sector, y por otro lado en la comparación de solo dos datos a través de una tasa de variación sin tener en cuenta la historia completa de la serie, ante esta situación se plantea la alternativa que permita evaluar la calidad de un dato nuevo mediante herramientas de control estadístico de la calidad que pueden utilizarse para complementar los métodos tradicionales, los procedimientos que se proponen permiten el tratamiento univariado, mediante modelos ARIMA con Análisis de intervención, de un conjunto numeroso de series.

Dentro de las técnicas que se desarrollan en este estudio, en búsqueda de un marco teórico adecuado son los métodos de análisis univariado de series temporales ARIMA (Autoregressive Integrated Moving Average), basado en que una serie temporal obedece a un proceso estocástico¹ propuesto por Box - Jenkins (1970 y 1976), el Análisis de Intervención que permite la modelización ARIMA de una serie junto con la consideración de aquellos fenómenos que, sin formar parte de la dinámica básica de dicha serie, afectan a los valores que toma a lo largo del tiempo propuesto por Box - Tiao (1975) y el Análisis de control de Calidad que permite contrastar la calidad de la información coyuntural, a través de la variación relativa entre el dato realmente observado y la predicción con un período por delante denominado SORPRESA, Propuesto por Revilla, Rey y Espasa (1991).

En esta investigación se usa el método automático desarrollado por Revilla, Rey y Espasa (1991), eso encaja en la estrategia de Box-Jenkins y es una estrategia del modelado iterativo que identifica, estima y realiza un diagnóstico. Para este método

1/ Proceso estocástico, es una familia de variables aleatorias que corresponden a momentos sucesivos del tiempo. En economía y en general en las ciencias sociales, la información de una serie se obtiene mediante observación pasiva, es decir, sin control sobre los factores que influyen sobre la variable objeto de investigación. Por ello aunque se disponga de una serie muy larga, debe considerarse toda ella como una sola realización de un proceso estocástico.

automático usar los modelos ARIMA no es suficiente, para capturar los efectos de calendario y otras variaciones determinísticas (por ejemplo, la huelga, los fenómenos naturales: el fenómeno del niño, etc.), es el análisis de intervención que complementa para dar solución a este tipo de problemas. Por consiguiente todo el conjunto de modelos que utiliza el MODAUT en suma son los ARIMA con análisis de intervención.

En principio el trabajo intenta dar una síntesis teórica de los paper que circulan sobre los modelos ARIMA - AI y, a la luz de la bibliografía básica de la econometría. Y luego, mostrar la metodología de modelización y el uso de estos modelos para el desarrollo de herramientas de control de calidad. La idea es facilitar la tarea a los que con frecuencia trabajan con series de tiempo y tienen la necesidad de hacer un análisis exhaustivo y corrección de los datos estadísticos. Con este propósito, pienso que uno de los instrumentos más importantes para enfrentar los problemas que estamos viviendo en un mundo lleno de incertidumbre y riesgo, es la disponibilidad de una adecuada información estadística la que debe ser precisa, confiable, sostenida y sin retardos temporales; de fácil acceso e interpretación.

Uno de los aportes fundamentales se recoge en la sección V que ayuda al desarrollo de una metodología del control estadístico de la calidad a un dato en concreto aplicada a la serie Índice PBI en esta investigación, es gracias al tratamiento de la modelización automática que permite modelizar un gran número de series a la vez y por otro lado el tratamiento de la serie con el procedimiento siguiendo las decisiones y actuaciones de un experto usando el programa Eviews 3.1 ya que este paquete estadístico proporciona una variante para analizar el correlograma de los residuos al cuadrado, que se utiliza, principalmente, para analizar la presencia de varianzas heteroscedásticas.

Finalmente, el INEI puede jugar un rol líder en la promoción de buenas prácticas estadísticas en el desarrollo de la investigación metodológica y aplicaciones adecuadas a las necesidades de nuestra Institución, para luego dar a conocer a los sectores y las departamentales los conocimientos adquiridos durante el desarrollo de las investigaciones; Así, el INEI no dejaría de lado su responsabilidad como ente rector de las estadísticas Nacionales.

I. ASPECTOS PRELIMINARES

1.1 FUNDAMENTACION DEL PROBLEMA DE INVESTIGACION

1.1.1 El planteamiento del problema

En el INEI y los Sectores se observa cierta preocupación cada vez mayor, por contar con métodos estadísticos de control de calidad, que podrían usarse para mejorar la confiabilidad de la información, y al mismo tiempo, mejorar también la rapidez y oportunidad con la que se producen los datos; sin embargo, en la actualidad se puede decir que en los sectores se observa un elevado conocimiento sectorial por parte de los profesionales y expertos que trabajan en la elaboración de las estadísticas básicas mensuales, ello redundando en unos métodos de control de calidad que se basan fundamentalmente en las relaciones estructurales y productivas de los sectores, por lo cual se observa una reducida utilización de métodos estadísticos en las distintas fases de la elaboración de las estadísticas; así por ejemplo:

En el caso de las estadísticas del Sector Pesca, el método de control de calidad de la información se realiza, basándose fundamentalmente en el conocimiento del sector y en fuentes alternativas (por ejemplo, en cuanto a la materia prima, se contrasta con la información recibida diariamente vía fax desde los mismos establecimientos industriales, al mismo tiempo se estudia la coherencia de los rendimientos para cada establecimiento).

Si bien el control de calidad de la información se realiza de forma detallada y rigurosa, basándose en el elevado conocimiento sectorial con el que cuentan las personas que trabajan en la elaboración de estas estadísticas, se observa que no se utilizan en mejor medida métodos estadísticos de control de calidad, - En las estadísticas del Sector Energético y Minero, en la actualidad, el control de calidad se basa fundamentalmente, en la experiencia y en el amplio conocimiento sectorial de los expertos que trabajan en las encuestas, el control de calidad se realiza de manera minuciosa. La consistencia de la información se basa fundamentalmente en la comparación de la información con periodos anteriores, no siendo tan intensivos, en cambio, en la aplicación de métodos y modelos estadísticos que permitan complementar lo usual - Los datos que el INEI recibe regularmente de los sectores, y que sirven de base para la construcción de series agregadas, dado que cada mes y cada trimestre se reciben diverso número de datos nuevos, es preciso realizar un esfuerzo para detectar posibles errores en los mismos. Las causas de estos errores son múltiples y muy variadas; pueden ir desde un error en la colocación de una coma, hasta la confusión, a pesar del esfuerzo realizado en estos últimos años, la detección de errores que el INEI realiza en el presente es fundamentalmente tradicional y se basan en unos principios simples de comparar los dos últimos periodos o al

conocimiento exhaustivo del sector (ignorando pues la historia completa de la serie de que se trate), por lo cual se observa una reducida utilización de métodos estadísticos de control de calidad. Tal situación obedece a factores como debilidad técnica que existe en determinadas operaciones estadísticas y ausencia de visión de idea de proceso, como algo implícito en la producción estadística, aún existente en determinadas oficinas productoras de estadísticas.

Con el desconocimiento de métodos estadísticos apropiados por parte de los Sectores y del INEI para tratar los métodos de control de calidad que complementen los métodos actualmente utilizados, basados fundamentalmente en el conocimiento del sector, se desvirtúa el rol del INEI de lograr alcanzar el prestigio y gozar de plena credibilidad en el sistema estadístico nacional; si se mantienen estas condiciones el INEI perderá protagonismo en la conducción de las estadísticas y su liderazgo, siendo esta su mayor fortaleza, sé vera debilitada.

Por ello se considera conveniente, la utilización de métodos y modelos estadísticos en la fase de control de calidad, que puede complementar los métodos utilizados hasta ahora, basados fundamentalmente en el conocimiento del sector, con el objeto de mejorar la calidad y la rapidez de la información.

Dentro de los métodos estadísticos de control de calidad, se propone utilizar las técnicas estadísticas conocidas como métodos de depuración selectiva², basadas en modelos de series temporales

(modelos ARIMA con análisis de intervención).

En el caso de las estadísticas mensuales, el marco conceptual más adecuado para su estudio, son los procesos estocásticos discretos, que pueden ser interpretados como una sucesión de variables aleatorias.

Los modelos ARIMA son un caso particular de procesos estocásticos, discretos y lineales, la introducción de series temporales puede tener gran utilidad en la elaboración y análisis de encuestas e indicadores mensuales; en concreto, se propone la utilización de modelos de series temporales para el control de calidad.

1.1.2 La formulación del problema

- ¿Cómo afecta la ausencia de métodos y modelos de control de calidad en los procesos estadísticos y de qué manera podrían usarse para mejorar la acuracidad de la información, y, al mismo tiempo también la rapidez y oportunidad con el que se producen los datos?

1.1.3 La sistematización del problema

- ¿Cuáles son las características del método de control de calidad usadas actualmente en el INEI y los Sectores?
- La reducida utilización de métodos y modelos estadísticos apropiados ¿de qué manera afecta a la credibilidad y confiabilidad de la información?
- ¿Cuál es la relación entre el método de control de calidad usado actualmente y el desarrollo en cuanto a la acuracidad de la información,

2/ Selective editing methods based on time series modeling – Prepared by Pedro Revilla and Pilar Rey

rapidez y oportunidad con el que se producen los datos en el INEI y los Sectores?

1.2 OBJETIVOS DE LA INVESTIGACION

1.2.1 Objetivo general

Plantear e implementar la metodología de control de calidad en los datos macroeconómicos de la producción estadística básica mensual y en los indicadores coyunturales

1.2.2 Objetivos específicos

- Adecuar la metodología de los modelos ARIMA con Análisis de Intervención en la producción de las estadísticas básicas mensuales y los indicadores coyunturales

- Implementar los procedimientos de modelización en base a la estimación, validación y predicción de un conjunto dado de series macroeconómicas a la vez.
- Ejecutar una metodología de control estadístico de calidad basado en los modelos y que estén básicamente orientados a construir herramientas de control estadístico de calidad a nivel de los macro datos, con la finalidad de evaluar si es o no válido el dato real que llega como información.
- Proponer pautas para identificar un perfil de las series macroeconómicas donde no es adecuado utilizar la metodología de control de calidad basados en los modelos ARIMA - AI.

II. MARCO CONCEPTUAL

2.1 LAS VARIABLES MACROECONÓMICAS

Se puede considerar al menos dos factores que han limitado el estudio exhaustivo en los modelos macroeconómicos para la predicción y análisis de control de calidad en los procesos estadísticos en el caso Peruano. El más obvio probablemente sea el de la precariedad de la información en una escasa antigüedad y diferente año base³ de las series como en una generalizada falta de confiabilidad en los datos en sí. El otro factor reside en características particulares de la serie macroeconómica: la concurrente inestabilidad monetaria que se extendió hasta fines de la década de 1980 no creaba un ambiente propicio para la elaboración de proyecciones macroeconómicas de un valor significativo. Con la década de 1990, sin embargo, llegaron tanto una relativa estabilidad económica que brindaron un ambiente propicio para el desarrollo del estudio de modelos de predicción.

En este estudio se encara dos grupos de variables uno corresponde al Índice Mensual de la Actividad Productiva Nacional que muestra la evolución de la economía global y sectorial en el corto plazo, proporcionando a los usuarios un indicador sintético de la producción nacional, que consta de nueve variables

(PBI Global, Agropecuario, Pesca, Minería e Hidrocarburos, Manufactura, Electricidad y Agua, Construcción, Comercio y Otros Servicios) y el otro grupo de variables es el volumen de producción del Sector Minería, resultado de la actividad productiva de las empresas proporcionada por el INEI y las Oficinas Sectoriales.

Las series de datos agregados, tanto en los indicadores económicos elaborados por el INEI, como el de las estadísticas básicas mensuales elaboradas por las oficinas Sectoriales de estadística de los Ministerios han sido seleccionadas no solo por ser relevantes sino también para llevar a cabo el control de calidad de los datos; a diferencia de otras variables, tienen una frecuencia y antigüedad adecuada para los fines de este estudio. Sin embargo, como la calidad y precisión de las predicciones realizadas a partir de los modelos ARIMA-AI van estar íntimamente ligadas con características particulares de las series, es importante llevar a cabo un examen cuidadoso de las mismas; ya que las serie bajo estudio están afectadas por acontecimientos inusuales como huelgas que afectan directamente a la producción, clima más frío que lo normal y las sequías que afectan la estructura temporal de la información del Sector agricultura y por ende del Sector agroindustrial, el fenómeno del niño y las vedas tienen un efecto multiplicador sobre varios sectores

³/ Año Base, es el periodo en el que se establece el universo productivo de una economía y se define una estructura de precios que sirve para valorar la futura producción sectorial (para más información ver www.inei.gob.pe)

y en especial del Sector Pesca; a todos estos acontecimientos se suman los errores de medición de las series, referente al tratamiento de datos y los errores de muestreo que son afectadas al recoger la información a través de la encuesta mensual, donde este hecho, determina la irregularidad de las series (que esta constituida por las fluctuaciones residuales y erráticas de la serie), teniendo un comportamiento más errático las series sectoriales como Pesca y Minería, comparadas con las series del Índice del PBI, esto posiblemente ocurre por el nivel de agregación de la serie.

En estos hechos recae la importancia del estudio de las series temporales⁴ que viene dada por la abundancia de este tipo de observaciones en el campo económico, ya que constituyen la mayor parte del material estadístico con el que trabajan los sectorialistas y los encargados de coyuntura económica. Estos profesionales necesitan a menudo instrumentos que le faciliten modelar y hacer uso de estos modelos para predecir y realizar un control de calidad sobre determinadas variables, esto les ayudará a justificar y sustentar las decisiones que tomen. Con el análisis de las series temporales se pretende básicamente, conocer una variable a través del tiempo para, a partir de este conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales, puedan construir modelos, luego pronosticar y hacer uso de estos modelos para el control de calidad al dato real que llega como información.

2.2 BÚSQUEDA DE UN MARCO CONCEPTUAL ADECUADO: LOS MODELOS ARIMA CON ANÁLISIS DE INTERVENCIÓN

Los modelos ARIMA desde los 70's han ganado un papel fundamental en la literatura sobre predicción. Originalmente introducida a través de los trabajos pioneros de Box y Jenkins, ésta generalización de los modelos autorregresivos y de medias móviles (ARMA) probó ser bastante efectiva a pesar de su sencilla computación y se instauró de ese modo como una alternativa relevante a los modelos estructurales, por entonces muy extendidos en la práctica de predicción. La forma tradicional de generar y aplicar modelos ARIMA es utilizando la metodología elaborada por Box y Jenkins⁵. Si bien existen distintas variantes de como esta se aplica, la forma más general es la que divide el proceso en cuatro pasos:⁶ identificación de la estructura del modelo, estimación de los parámetros del modelo, diagnóstico del modelo y la posterior predicción; luego, estas predicciones nos servirán para el control de calidad.

El Análisis de Intervención

La aplicación de los modelos ARIMA al nivel hasta ahora comentado (denominado "univariante") presenta algunas limitaciones como herramienta de predicción, por lo tanto también para el control de calidad. La más elemental es su incapacidad para detectar esos "efectos asociados a sucesos atípicos" que dan lugar a errores

4/ Series temporales, conjunto de observaciones generadas secuencialmente en el tiempo, una serie temporal en general tendrá un carácter aleatorio (sino tuviera carácter aleatorio, no tendría sentido en registrar los valores correspondientes a distintos periodos, ya que se podrían obtener simplemente por aplicación de una formula)

5/ Véase, Casas Tragodara, Carlos (Cáp. 15-18). *Econometría Moderna*

6/ Véase, por ejemplo Gujarati (1997, Cáp.22 Pág. 721-728). *Econometría Básica*. 3ra edición.

excepcionalmente elevados y pueden estar provocados por cuestiones tales como:

- Errores en la cuantificación de algún dato de la serie o cambios en el criterio de cálculo.
- Acontecimientos extraordinarios que afectan puntualmente al fenómeno en estudio (huelga, cambio de gobierno, devaluación, etc.)
- Variación en el comportamiento estacional (oscilaciones del período de semana santa, cambios climatológicos para un mismo período en diversos años, etc.)
- acciones especiales o "intervenciones" propiamente dicha (promoción especial, aumento de tarifas, reforma fiscal, lanzamiento de un nuevo producto, etc.)
- El variable impulso (AO: additive outlier), afecta solo en un punto aislado (o en diversos puntos, pero separados entre sí)
- Variable de escalón (LS: level shift), que afecta a todos los datos durante un periodo de tiempo.
- Cambio temporal (TC: temporary change), llamado también transitorio (en el programa TRAMO⁷), situación intermedia en que el efecto de la intervención anómala no es permanente pero persiste durante algún tiempo.

Predicción

En algunas ocasiones, el "valor atípico u observaciones anómalas" puede eliminarse (corrigiendo el error en la serie); en otras, es posible interpretar el posible error de predicción como una estimación del efecto que ha provocado una cierta "intervención". Sin embargo, hay ocasiones en el que es conveniente formalizar el estudio de esos impactos especiales: ése es el objetivo del llamado análisis de intervención que consiste en introducir variables en el modelo que permitan reflejar la presencia de cambios en la serie que no responden a ningún patrón sistemático de comportamiento. La estrategia básica consiste en, una vez identificado el modelo ARIMA, estimarlo incluyendo una o más variables ficticias que representen los cambios mencionados. Los tipos básicos de variables de intervención son las siguientes:

En la gran mayoría de los documentos de econometría se utilizan diversas técnicas de predicción cuantitativa. Pero los modelos Arima univariante en la práctica se ha visto que este tipo de modelos presenta mejores estimaciones que los modelos estructurales⁸. Por este motivo en los últimos años se han escrito abundante literatura al respecto. En la mayoría de los paper la predicción es realizada a partir del último dato del periodo muestral, donde el origen de predicción es el periodo t y se obtienen predicciones uno o varios periodos por delante, es decir $t+1$, $t+2$,... aunque también analizan cómo se habría comportado el modelo si hubiera tenido que hacer una predicción dentro del periodo histórico ya conocido, que ha servido de base a su estimación y contraste.

Pero como el objetivo del estudio es realizar control de calidad a una serie temporal de un periodo en concreto, la

7/ Gómez y Maravall (1996 1998a). Programa de modo automático.
8/ Casas Tragodara, Carlos (Cáp. 15 al 18). Econometría Moderna.

predicción que se desea obtener es con un periodo por delante, es decir, la predicción con origen en $t-1$ y por lo tanto es lo lógico cuando se quiere valorar el dato del mes t con toda la información de la serie que dispones, en eso radica la diferencia del uso de las predicciones para el control de calidad.

Control estadístico de calidad basado en los modelos

La información referente al tema del control de calidad de los datos agregados es bastante limitada, su información se restringe a unos cuantos paper elaborados por Pedro Revilla y Pilar Rey (1999)⁹.

El control estadístico de la calidad consiste en la aplicación de los modelos ARIMA-AI para construir herramientas de control estadístico de calidad, que van ha ser denominados SORPRESAS que vienen ha ser la variación relativa entre el dato realmente observado y la predicción para ese dato, se construyen basándose en la función de predicción, la idea básica es muy simple si los datos observados difieren considerablemente de los pronósticos ARIMA-AI, los datos pueden ser erróneos.

En general, en las encuestas continuas, la información de mayor utilidad para llevar a cabo el control de calidad, consiste en datos pasados de la misma población, obtenidos en encuestas de periodos anteriores.

Por ejemplo, en encuestas de periodicidad mensual, las tasas de variación, pueden mejorarse mediante el uso de las herramientas construidas sobre la base de los modelos ARIMA-AI.

Estas herramientas, presentan dos ventajas fundamentales respecto al uso de las tasas de variación:

- Las tasas de variación utilizan únicamente una observación del pasado de la series (el dato del mes anterior; en el caso de la tasa intermensual, o el dato del año anterior, en el caso de la tasa interanual). Por el contrario, las herramientas de depuración basadas en las funciones de predicción de los modelos ARIMA con análisis de intervención, utilizan (y de una forma óptima) todo pasado de la serie. De hecho, la predicción es una función lineal de todas las observaciones, presentes y pasadas.
- Las herramientas de depuración basadas en modelos ARIMA con análisis de intervención permiten llevar a cabo un control de calidad de tipo probabilístico.

A partir de los modelos pueden construirse un conjunto de herramientas de control de calidad, denominadas, genéricamente, sorpresas y los valores atípicos pueden ser definidos como aquellos indicadores cuya sorpresa queda fuera del intervalo.

9/ Consultores Españoles del INE y del CIS (Centro de investigaciones sociológicas) de España, respectivamente.

III. FORMULACION Y OPERACIONALIZACION DE LA HIPÓTESIS

3.1 FORMULACIÓN DE LA HIPÓTESIS

Los procedimientos utilizados por los especialistas del INEI y los Sectores, pueden ser complementados con técnicas estadísticas adecuadas, como los modelos ARIMA - AI que permiten recoger la información de la serie en el presente y en el pasado, basadas en criterios probabilísticos permitiendo obtener estimaciones e intervalos de confianza que posibilitan evaluar la calidad de los datos de un gran número de series a la vez.

3.2 OPERACIONALIZACIÓN DE LA HIPÓTESIS

- Los procedimientos utilizados actualmente por los especialistas del INE y los Sectores se lograrán mejorar a través de la aplicación y uso de métodos estadísticos en los procesos.
- Los modelos ARIMA con Análisis de Intervención tienen la capacidad de crear herramientas de control estadístico de calidad para mejorar la confiabilidad de la información y la rapidez con la que se producen los datos, y están, basadas en los procesos denominados sorpresa, que consiste en detectar valores atípicos en un intervalo de confianza.
- La evaluación de la calidad de los datos de un gran número de series a la vez, se logrará haciendo uso de los programas automáticos (MODAUT) que están basados en la metodología de Box - Jenkins, en concreto de los modelos ARIMA-AI. A su vez estas evaluaciones de la calidad del dato serán complementadas a través del programa Eviews 3.1 siguiendo las decisiones y actuaciones de un experto.

IV. METODOLOGÍA DE MODELIZACION Y CONTROL DE CALIDAD

El objetivo del presente estudio es desarrollar una metodología basada en los modelos Arima con Análisis de Intervención que sirva como herramientas de control de calidad a las series macroeconómicas; el propósito, es estudiar el dato de un mes en concreto a la que se quiere evaluar si es o no válido. Para tal caso se utilizaran las SORPRESAS (Revilla y Rey 1991).

A continuación se detalla la metodología a través de las siguientes etapas:

4.1 IDENTIFICACION

Esta primera etapa de la metodología de Box-Jenkins (1976) puede ser dividida en dos partes.

En una parte se busca corregir los problemas propios de la serie, mientras que en una segunda se busca identificar un modelo $ARIMA(p,d,q)(P,D,Q)_{12}$ que logre explicar adecuadamente la serie en estudio.

4.1.1 Análisis exploratorio

La mayoría de las series con las que se trabaja en economía presentan algún tipo de tendencia, y es fácil encontrar comportamientos estacionales cuando se trabaja con datos de frecuencia inferior a

la anual. Las series índices del PBI mensual y los volúmenes de producción de los sectores con las que se trabaja en esta investigación no son una excepción. Su representación gráfica permite afirmar que poseen tendencia y en muchos casos estacionalidad. Sin embargo a simple vista no es posible determinar el carácter estocástico y/o determinista de las mismas, por ello es relevante estudiar la estacionariedad de las series. La aplicación de modelos ARIMA- AI debe realizarse con series estacionarias. La estacionariedad exige que cumpla un doble requisito en la práctica:

- Una media aproximadamente constante en el tiempo, y
- Una varianza o dispersión también constante.

En términos gráficos, una media constante supone la no existencia de tendencia y una varianza constante corresponde a un gráfico en que las oscilaciones alrededor de la media sean similares, lo que técnicamente se conoce como homocedasticidad¹⁰.

Si la serie de tiempo no cumple con el requisito de estacionariedad será necesario realizar posiblemente las siguientes transformaciones:

10/ El estadístico de contraste para la prueba de Levene plantea como hipótesis nula la homogeneidad de varianzas:

H_0 = serie y_t con varianza homocedastica

H_1 = serie y_t con varianza heteroscedastica

Si el p-valúe asociado al estadístico de contraste es menor que 0.05, se rechaza la hipótesis nula de homogeneidad de varianzas, siendo aconsejable en este caso tomar logaritmos neperianos a la serie original.

• **Eliminación de la tendencia lineal.**

Una forma elemental de eliminar la tendencia, es decir, de conseguir una serie estacionaria en la media, es efectuar diferencias sucesivas de la misma, así:

$$\Delta y_t = y_t - y_{t-1}, d=1 \text{ (orden regular)}$$

y si la tendencia aún persiste, calcularemos una segunda diferencia:

$$\Delta^2 y_t = \Delta(\Delta y_t) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}),$$

d=2 (orden regular)

Y así sucesivamente. Generalmente es suficiente con una o dos diferencias sucesivas, para que la serie se establezca en medias.

• **Transformación previa de datos.**

Por su parte, una manera sencilla de reducir la heteroscedasticidad¹¹, es decir transformar una serie en otra con varianza relativamente constante, es tomar logaritmos, ya que ello supone trabajar con valores relativamente más homogéneos.

Las transformaciones logarítmicas de datos son muy frecuentes en los modelos ARIMA - AI, sin embargo, merece la pena hacer una breve reflexión sobre su significado.

Sea y_t la serie original e $y_t^* = \ln y_t$ la transformada. Al tomar ahora diferencias,

obtenemos:

$$y_t^* - y_{t-1}^* = \ln y_t - \ln y_{t-1} = \ln\left(\frac{y_t}{y_{t-1}}\right)$$

Y esta expresión es, aproximadamente, la tasa de variación de y_t , al menos para tasas de cambio relativamente pequeñas.

Hay una familia de posibles transformaciones, de la que los logaritmos constituyen un caso particular. Se trata de la denominada transformación de Box - Cox¹². La transformación sugerida es para, $\lambda = 0$ es decir, el logaritmo neperiano.

Para su mejor estudio se aplicaran contrastes de comprobación de la estacionariedad,¹³ entonces, dado que la metodología de Box y Jenkins, precisan que las series sean estacionarias, es importante comprobar esta situación antes de trabajar con ellas. El método formal para analizar si una serie es estacionaria es el test de raíces unitarias¹⁴.

El test de raíces unitarias más difundido es el test de Dickey Fuller (DF) que presenta una variante denominada test de Dickey-Fuller aumentado (ADF). Adicionalmente, puede considerarse también el test de Phillips-Perron (PP).

El test de Dickey-Fuller aumentado (ADF), que permite identificar la presencia de raíces unitarias, considera:

$$y_t = \mu + \rho y_{t-1} + a_t^{15}$$

11/ Es importante recalcar que la aplicación de logaritmos no corrige el problema de heteroscedasticidad, sino que es un método sencillo y efectivo de estabilizar la serie. Véase, Casas Tragodara, Carlos. (Cáp. 15 al 18). Econometría moderna.2001.
 12/ Véase con mayor detalle Antonio Aznar (1993, Pág. 94). Métodos de Predicción en Economía II y Ezequiel Uriel (2000, Pág. 98-100). Introducción a las series temporales.
 13/ Véase, por ejemplo, Ramón Mahía (1999). Revisión de los Procedimientos de Análisis de la Estacionariedad de las Series Temporales.
 14/ El orden de integración d, denotado por I(d) es el número de raíces unitaria que contiene la serie o el número de operaciones de diferenciación que hay que efectuar para convertir la serie en estacionaria.
 15/ proceso autorregresivo de primer orden AR(1)

Donde μ y ρ son parámetros a estimar y ϵ_t es un término del error que se supone cumple la propiedad de ruido blanco¹⁶. Dado que ρ es un coeficiente de autocorrelación que, toma valores $-1 < \rho < 1$, si $\rho = 1$, la serie y_t es no estacionaria, la varianza y_t de crece en el tiempo. Si el valor de $\rho > 1$, entonces se dice que la serie es explosiva. De esta forma, la hipótesis de estacionariedad de una serie puede ser evaluada analizando si el valor absoluto de ρ es estrictamente menor que uno. Tanto el test de Dickey-Fuller como test de Phillips Perron plantean como hipótesis nula la existencia de una raíz unitaria:

$$H_0 : \rho = 1^{17} \quad (\text{problema de raíces unitarias o serie no estacionaria})$$

$$H_1 : \rho < 1$$

La interpretación al test estadístico ADF sería:

Se rechaza la hipótesis nula si el estadístico de Dickey-Fuller Aumentado (ADF) es mayor que los valores críticos de MacKinnon¹⁸. Por lo tanto afirmaríamos

$$\nabla_{12} y_t = (1 - B^{12}) y_t = y_t - y_{t-12} \quad , \quad D=1 \quad (\text{orden estacional})$$

$$\nabla_{12}^2 y_t = (1 - B^{12})^2 y_t = (y_t - y_{t-12}) - (y_{t-12} - y_{t-24}), \quad D=2 \quad (\text{orden estacional})$$

generalmente será suficiente con una única diferencia estacional $D=1$

Una vez estudiados los rasgos más importantes de la serie de tiempo, se puede proceder a calcular la desviación estándar a las transformaciones realizadas a la serie original, por ello, se elegirá entre

que la serie es estacionaria en la componente regular; caso contrario hay un problema de raíces unitarias que amerita una primera diferencia para resolver el problema de no estacionariedad.

- **Eliminación de la tendencia estacional.**

El tratamiento de la Estacionalidad en los modelos ARIMA-AI, se da habitualmente cuando los datos son mensuales, como en el caso de esta investigación, es decir es un contexto de predicción a corto plazo. Por ello, el tratamiento de la estacionalidad debe ocupar un papel central en el proceso de modelización.

A tales efectos, los modelos ARIMA-AI tratan de captar el comportamiento estacional en forma similar al componente regular. En este sentido, el componente estacional puede exigir una o varias diferencias de orden estacional (D), es decir, por ejemplo, para nuestro caso, para datos mensuales sería de la siguiente manera:

ellas la que posea menor desviación estándar y se procederá a la identificación del modelo ARIMA a través de esta selección.

Hasta ahora hemos revisado algunos conceptos básicos para analizar una serie. Sin embargo, aún no sabemos cómo elegir el modelo más conveniente para cada

^{16/} white noise, media nula, varianza constante y ausencia de autocorrelación.

^{17/} Podría calificarse como test de no estacionariedad, pues es, precisamente, lo que se chequea en su hipótesis nula.

^{18/} MacKinnon (1991) estima los valores de respuesta del test utilizando los resultados de la simulación, permitiendo que se calculen los valores críticos de Dickey y Fuller para cualquier tamaño muestral y cualquier número de variables explicativas. El programa e-views 3.1 proporciona de manera automática esta salida.

caso, pasamos a la búsqueda del modelo más conveniente:

4.1.2 Identificación de la Estructura del Modelo

El primer paso en el análisis de una serie es la identificación del modelo. Si la serie no presenta estacionalidad, determinar los órdenes del modelo ARIMA(p,d,q), y si la serie presenta estacionalidad de periodo $s=12$, los órdenes del modelo será de tipo multiplicativo ARIMA(p,d,q)*(P,D,Q)₁₂

Una vez determinado los parámetro d y D en el análisis exploratorio, el siguiente paso será identificar los parámetros p y q y, si la serie presentara el caso de estacionalidad, Los parámetros P y Q. El instrumento técnico básico para identificar

un modelo ARIMA es la **denominada función de autocorrelación simple y parcial** (que mide el grado de correlación entre cada valor de la variable y los desfasados 1,2,3..., h periodos), la función de autocorrelación parcial no tiene la misma potencia identificadora que la función de autocorrelación simple y se sugiere usarla solo en caso de duda acerca de la identificación del modelo. Las autocorrelaciones parciales tienen un comportamiento contrario a las autocorrelaciones simples: se cortan para los procesos autorregresivos y decaen exponencialmente o en forma sinusoidal para los modelos de media móvil. Esto prueba la dualidad entre los modelos autorregresivos y de media móvil, en el siguiente cuadro 1 se resume los criterios que hacemos referencia:

Cuadro 1. Criterios para determinar el valor de los parámetros para un modelo ARMA(p,q)¹⁹

Modelo	Patrón típico ACF	Patrón típico PACF
AR(p)	Cae de forma exponencialmente o de forma sinusoidal convergente, o ambos	Picos significativos hasta el rezago p
MA(q)	Picos significativos hasta el rezago	Cae exponencialmente
ARMA(p,q)	Cae exponencialmente	Cae exponencialmente

En la práctica cuando se estudia las autocorrelaciones simples y parciales obtenidas a través de un conjunto de datos, se observan las primeras autocorrelaciones que son las más confiables y la identificación se realiza a partir de estas. Sin embargo, es importante la preocupación del análisis de la componente estacional, debido a que hasta ahora solo hemos resuelto el problema de cómo identificar un ARMA o ARIMA. Por tanto, si al calcular las

autocorrelaciones notaremos que, luego de un corte o decaimiento inicial correspondiente al comportamiento de un AR o MA, las autocorrelaciones reaparecen nuevamente tomando valores altos alrededor del periodo estacional (12 para series mensuales). Esto significa pues, que hay dos enlaces: uno entre periodos sucesivos como en cualquier modelo ARMA y otro entre meses distintos. Por lo tanto, necesitamos usar un modelo de tipo multiplicativo y se representa como

19/ Fuente Damodar N. Gujarati: Econometría Básica (1999. Pág. 725)

la combinación de dos modelos ARIMA(p,d,q)*(P,D,Q)₁₂, donde el primero corresponde a la parte regular y el segundo a la parte estacional²⁰.

No resulta fácil seleccionar un modelo ARIMA incluso para personas con experiencia. En muchas ocasiones se tendrán dudas razonables sobre el modelo más adecuado. Pero no debemos preocuparnos en exceso por ello, ya que una posibilidad de seleccionar varios modelos alternativos y comprobar posteriormente cuál de ellos resulta finalmente más idóneo.

En otros términos, no debe considerarse la etapa de identificación del modelo como una decisión irreversible, sino como el inicio de un proceso de selección que tendremos ocasión de revisar en etapas sucesivas.

En esta etapa es importante aclarar el significado del estadístico Q que aparece en la última línea del gráfico de las funciones de autocorrelación simple y autocorrelación parcial.

El contraste fue propuesto por Box y Pierce (1970), y parte de una suma de coeficientes de autocorrelación. Para n datos y m coeficientes, el estadístico Q se define como:

$$Q = n \sum_{k=1}^m r_k^2$$

y puede demostrarse que se distribuye asintóticamente como una χ^2 con m grados de libertad, bajo la hipótesis nula:

$$H_0: (\rho_1 = \rho_2 = \dots = \rho_m = 0)$$

Dicho de otra manera, solo si el valor de Q calculado supera el teórico de las tablas, podrá rechazarse la hipótesis nula de todos los coeficientes de autocorrelación y tendrá sentido buscar el orden del proceso ARMA correspondiente. En caso contrario, la serie analizada parece corresponder a un ruido blanco, es decir a un ARMA(0,0).

En programa Eviews 3.1, incorpora una variante del estadístico Q, conocida como estadístico Q de Ljung-Box (1978):

$$Q = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-m}$$

Que se distribuye aproximadamente como una χ^2 con (n-m-p) grados de libertad (número de coeficientes o retardos considerados).

4.2 ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO²¹

Una vez determinados los órdenes p, d, q, P, D y Q de un proceso ARIMA(p,d,q)(P,D,Q)_s con periodo conocido s=12, se trata de, a partir de una realización del proceso u observación de la serie, {y₁...y_n} estimar los parámetros del modelo con componente estocástica (Lny_t) que sigue una representación autorregresiva, integrada y de medias móviles (ARIMA) de tipo multiplicativo (Box y Jenkins, 1976):

$$Lny = \frac{\theta_q(B)\Theta_Q(B^{12})}{\phi_p(B)\Phi_P(B^{12})(1-B)^{d=0,1,2}(1-B^{12})^{D=0,1}} * a_t$$

20/ Véase Mg. Isela Agüero Pág. 51. Análisis de Series Temporales

21/ El MODAUT, estima los parámetros con máxima verosimilitud (filtro de Kalman), y el Eviews 3.1 estima por MCO

es decir, se trata de estimar los parámetros, $i=0, \dots, p+P+q+Q$,

donde:

$$\beta_i = \phi_i, i = 1, \dots, p$$

$$\beta_i = \Phi_{i-p}, i = p+1, \dots, P+p$$

$$\beta_i = \theta_{i-(p+P)}, i = p+P, \dots, p+P+q$$

$$\beta_i = \Theta_{i-(p+P+q)}, i =$$

$$p+P+q+1, \dots, p+P+q+Q$$

$$\beta_i = c, i = 0$$

Si $\hat{\beta}_i$ es la estimación de los parámetros β_i , la primera etapa en la validación del modelo consistirá en comprobar si los coeficientes β_i son significativamente distintos de cero. Para ello, sobre cada parámetro, se planteará la hipótesis nula:

$$H_0: \beta_i = 0$$

Dicha hipótesis puede ser interpretada como que la variable asociada al parámetro β_i no mejora el ajuste con respecto al obtenido con las restantes variables incluidas en el modelo. Si el p-value asociado al valor del estadístico de contraste t es menor que α , se rechazará la hipótesis nula al nivel de significación α , y se concluirá que el parámetro es significativo y será considerado en el modelo.

El hecho de que el modelo identificado sea adecuado no debe hacer descartar la posibilidad de que otro modelo algo más

complejo pueda ajustarse mejor a la serie observada. Para comprobar si algún otro modelo se ajusta mejor conviene utilizar la técnica de **sobre ajuste**²².

Una cuestión importante en el análisis de los coeficientes es el examen de estacionariedad e invertibilidad, en las salidas que nos proporciona el Eviews 3.1 suele aparecer junto al coeficiente estimado los coeficientes de estacionariedad e invertibilidad; con respecto que, si el modelo estimado cumple con la estacionariedad, procede de factorizar la parte autorregresiva y se concluye que si el coeficiente de estacionariedad esta próximo a 1, es un indicio de que el modelo no es estacionario, siendo aconsejable, en este caso, tomar una diferencia adicional. Por otra parte, para comprobar si el modelo estimado es invertible se factoriza la parte media móvil y se concluye, si algún valor del coeficiente de invertibilidad está próximo a 1, será indicativo de que el modelo estimado es no invertible. En muchos casos, este problema se ha podido producir como consecuencia de una sobrediferenciación. Para completar el análisis de los coeficientes, es conveniente examinar la matriz de correlación entre los coeficientes estimados con el objeto de detectar si existe el problema de multicolinealidad. Cuando la correlación entre dos coeficientes estimados esté próximo a 1, sería un indicio suficiente para considerar que existe un problema grave de multicolinealidad. Si esto ocurre los coeficientes estimados serian muy inestables con lo que podría cambiar mucho de una realización a otra. Para

22/ El sobre ajuste consiste en estimar cada uno de los modelos obtenidos al aumentar en una unidad cada uno de los órdenes p, q, Q y P en el modelo original. En cualquier caso, se recomienda que los órdenes no alcancen valores superiores a dos. Si un modelo admite algún orden superior a dos, puede ser un indicio de que el modelo originalmente seleccionado no es el más adecuado.

evitar este problema puede ser conveniente eliminar algún parámetro del modelo con el objeto de conseguir unos estimadores más estables aún a costa de que el grado de ajuste sea pequeño, y en consecuencia, lo que se debe entender en esta primera etapa²³, que los coeficientes deben ser estadísticamente significativos, que el modelo estimado cumpla con la estacionariedad e invertibilidad y que los coeficientes estimados no tengan el problema de multicolinealidad.

4.3 DIAGNÓSTICO DEL MODELO

La validación del modelo consistirá en el estudio de los residuos, los cuales ofrece un vasto número de análisis que consisten principalmente en lo siguiente; como comprobar si se satisface la hipótesis del modelo ARIMA relativa a que los residuos a_t del modelo estimado se aproximen al comportamiento de un ruido blanco. Por consiguiente, la comprobación de la hipótesis de proceso de ruido blanco se realizará sobre una estimación de los mismos. Si \hat{y}_t es la estimación de y_t mediante el modelo ajustado, una estimación de la realización del proceso de ruido blanco a_t será el residuo:

$$e_t = \hat{a}_t = y_t - \hat{y}_t$$

La validación consistirá que el residuo \hat{a}_t tiene media cero, que es estable en varianza²⁴, que sus observaciones están incorrelacionadas y que procede de una distribución normal.

Respecto de la nulidad de la media y a la estabilidad de la varianza de la serie \hat{a}_t la primera es consecuencia del método de estimación de los coeficientes del modelo, mientras que la segunda es consecuencia de la estabilidad en varianza de la serie y_t .

Los residuos deben ser incorrelacionados, es decir, no deben existir ningún tipo de correlación en la serie de residuos y para este fin podemos utilizar las funciones de autocorrelación y las funciones de autocorrelación parcial. Sin embargo, un valor fuera de los límites de confianza en los retardos iniciales debería ser considerado como un indicio claro de que el modelo no es correcto. Un estadístico adecuado para contrastar la hipótesis de independencia, aleatoriedad o ruido blanco de los residuos es el estadístico de Box-Ljung que, para cada coeficiente de la función de autocorrelación simple, permite contrastar la hipótesis nula de que todos los coeficientes anteriores, hasta el correspondiente, son nulos. La interpretación es, si el nivel de significancia crítico o el p-value es menor que α , se rechazará la hipótesis nula al nivel de significación α , caso contrario aceptaríamos que los residuos tienen un comportamiento independiente, aleatorio o de ruido blanco.

El estadístico de Jarque - Bera se utiliza para determinar si la serie sigue una distribución normal. Un valor bajo de la probabilidad conduce a rechazar la hipótesis nula de una distribución normal.

23/ Estos análisis son considerados como la primera etapa de validación del modelo, por lo tanto, deben ser considerados en el momento de elegir el modelo que mejor represente el comportamiento de la serie estudiada

24/ Es decir que cumple con el requisito esencial de ruido blanco, pues, en caso contrario, el modelo debe ser rechazado o reestimado, ya que es un indicativo que el residuo contienen información relevante que debe ser extraída para su posterior modelización y para la predicción.

Si el valor obtenido del estadístico J-B es inferior al valor de referencia de tablas (aproximadamente, 6) y la probabilidad es distinta de cero, podemos suponer que la serie presenta características normales.

Otros estadísticos que nos ayudan a determinar las características de los errores son los test de diagnóstico de modelos disponibles en el Eviews 3.1, como identificación de los residuos al cuadrado, normalidad de los residuos, correlación serial de los residuos y heterocedasticidad autorregresiva condicional en los residuos.

4.4 ANÁLISIS DE INTERVENCIÓN

El objetivo del llamado análisis de intervención consiste en introducir variables en el modelo que permitan reflejar la presencia de cambios en la serie que no responden a ningún patrón sistemático de comportamiento. La estrategia básica consiste en, una vez identificado el modelo ARIMA, estimarlo incluyendo una o más variables ficticias que representen cambios inusuales. Los tipos básicos de variables de intervención son:

- La variable impulso, se trata de un atípico aditivo, el efecto de la observación anómala es inmediato y dura solo un periodo.

$$Ln y_t = \frac{\theta_q(B)\Theta_Q(B^{12})}{\phi_p(B)\Phi_P(B^{12})(1-B)^{d=0,1,2}(1-B^{12})^{D=0,1}} * a_t + \omega_0 DF_t$$

donde: D: impulso

F_t : es el mes que toma el valor 1

- variable escalón, conocido también como level shift, es el caso opuesto, en el que dicho efecto es permanente,

se representa introduciendo una raíz unitaria en el denominador del filtro racional, lo que equivale a una integración de impulso registrado en

$$DF_t$$

$$Ln y_t = \frac{\theta_q(B)\Theta_Q(B^{12})}{\phi_p(B)\Phi_P(B^{12})(1-B)^{d=0,1,2}(1-B^{12})^{D=0,1}} * a_t + \omega_0 \frac{DF_t}{1-B}$$

- Finalmente la situación intermedia en que el efecto de la de la observación anómala no es permanente pero

persiste durante algún tiempo, se recoge mediante un filtro de tipo siguiente:

$$Ln y_t = \frac{\theta_q(B)\Theta_Q(B^{12})}{\phi_p(B)\Phi_P(B^{12})(1-B)^{d=0,1,2}(1-B^{12})^{D=0,1}} * a_t + \omega_0 \frac{DF_t}{1-\delta B} \quad , \quad 0 < \delta < 1$$

Este tipo de atípicos recibe el nombre de "transitorio o "temporary change", poseyendo una respuesta al impulso monótona y convergente controlada por el parámetro δ , estando su signo

determinado por el de DF_t . Naturalmente, si $\delta = 0$ o si $\delta = 1$ se obtienen, respectivamente, los atípicos aditivo y de cambio de nivel antes comentado. En esta investigación se asume $\delta = 0.7$ ²⁵

25/ El MODAUT como el TRAMO utilizan este valor para la interpretación del coeficiente de la variable ficticia TC o transitorio.

A efectos de incorporar a los modelos aspectos de tipo cuantitativo, se ha hecho habitual definir las denominadas variables ficticias o Dummy que toman el valor de 1 si ocurre el fenómeno y 0 en caso contrario.

La selección de las variables de intervención y su modelización vendrán guiadas por el fenómeno a modelizar y, por supuesto, por las pruebas de validación que tratamos en la fase diagnóstico del modelo

4.5 PREDICCIÓN

Una vez seleccionado el modelo definitivo (al menos por el momento) que cumple satisfactoriamente los criterios de evaluación establecidos, puede pasarse a la etapa de predicción. En la mayoría de los documentos escritos sobre predicción, el problema consiste en obtener estimaciones de las predicciones futuras, aunque también puede resultar útil analizar cómo se habría comportado el modelo si hubiera tenido que hacer una predicción dentro del período histórico ya conocido, que ha servido de base a su estimación y contraste.

Como el objetivo de la investigación es hacer control de calidad a un dato en concreto, la predicción que se calcula a través del MODAUT y del Eviews 3.1 es de manera automática, y consiste hacer la predicción con un periodo por delante donde el origen de la predicción es el periodo $t-1$ es decir, esto hace que utilices toda la información disponible hasta el periodo $t-1$ y, por lo tanto es lógico porque el objetivo del estudio es valorar el mes t , con toda la información de la serie que dispones. Es importante en esta etapa de predicción evaluar la capacidad

predictiva del modelo, esta evaluación dependerá que los errores de predicción un período por delante en media, sean menores comparadas con la desviación teórica²⁶, El error relativo es en realidad una aproximación al error de predicción un período por delante en términos relativos.

$$\text{error Relativo} = \frac{y_t - \hat{y}_t}{y_t}$$

El error relativo se calcula así por que es más correcto para que se pueda comparar la media de estos errores con la desviación teórica o desviación típica residual del modelo.

Para el periodo histórico puede realizarse dos tipos de predicción²⁷:

- La estática, o paso a paso. Y
- La dinámica, o en cadena.

Las predicciones estática y dinámica coinciden solo para el primer periodo, pero, a partir del segundo periodo, la predicción dinámica utiliza el valor estimado y no el valor real del periodo precedente.

La predicción estática nos informa, pues, de los errores que hubiéramos cometido de utilizar el modelo para predecir solo un periodo por delante. La predicción dinámica deja al modelo que vaya realimentando sus propias predicciones.

Naturalmente, a efectos de una autentica predicción hacia futuro, solo será posible el segundo tipo de predicción. Cuando se disponga de nueva información real sobre los periodos de predicción, puede comprobarse el grado de error de nuestras estimaciones (un criterio a posteriori básico para enjuiciar la validez del modelo). Aunque puede reestimarse el modelo con cada nueva estimación disponible, la

26/ En el MODAUT, la acuracidad es la desviación teórica, que es lo mismo que desviación típica residual.

27/ Son proporcionadas por el Eviews 3.1, sin embargo el MODAUT también tienen salidas automáticas como el predis (predicciones un periodo por delante o varios periodos, estas predicciones son dentro del periodo histórico ya conocido y sirven para evaluar como se comportan los modelos obtenidos) y el predimod (proporciona predicciones para varios periodos por delante desde un único origen) y las sorpresas (que proporciona predicciones con un periodo por delante).

práctica habitual es continuar con los mismos parámetros estimados durante varios periodos, aunque alimentando el modelo con los valores conocidos de los errores realmente cometidos.

4.6 CONTROL ESTADÍSTICO DE CALIDAD BASADO EN LOS MODELOS

Las herramientas de control de calidad, para un dato en concreto se basan en los

$$(1 - B)^{d=0,1,2} (1 - B^{12})^{D=0,1} \ln y_t = \frac{\theta_q(B)\Theta_Q(B^{12})}{\phi_p(B)\Phi_P(B^{12})} a_t + \frac{\omega_s(B)}{\delta_r(B)} DF_t$$

donde:

- $\ln y_t$ es el logaritmo neperiano del indicador objeto de estudio.
- B es el operador de retardos, $B^s(y_t) = y_{t-s}$, $s=12$
- $\theta_q(B), \phi_p(B), \Theta_Q(B^{12}), \Phi_P(B^{12}), \omega_s(B), \delta_r(B)$ son polinomios en el operador de retardos.
- a_t son variables de ruido blanco i.i.d. $N(0, \sigma_a)$
- DF_t son las variables de intervención

La **Sorpresa simple**, S_t es la variación relativa entre el dato realmente observado y la predicción para ese dato \hat{y}_t :

$$S_t = \frac{y_t - \hat{y}_t}{\hat{y}_t}$$

donde: \hat{y}_t es la predicción un periodo por delante.

Dado que el error de predicción de un periodo por delante $e_t = \hat{y}_t$ es un proceso de ruido blanco con distribución $N(0, \sigma_a)$

modelos ARIMA con análisis de intervención que permiten llevar a cabo un control de calidad de tipo probabilística.

A partir de los modelos pueden construirse un conjunto de herramientas de control de calidad, denominadas, genéricamente, sorpresas.

A continuación, se describe la sorpresa simple dado el modelo ARIMA con Análisis de Intervención:

$$y \ln y_t - \ln \hat{y}_t \cong \frac{y_t - \hat{y}_t}{\hat{y}_t} \text{ se tiene que } S_t$$

es aproximadamente $N(0, \sigma_a)$. donde σ_a es la desviación típica residual o la raíz del error cuadrático medio. Por tanto, se puede construir un intervalo de confianza (por ejemplo al 95%):

$$P\{-\lambda_\alpha \sigma_a \leq S_t \leq \lambda_\alpha \sigma_a\} = 1 - \alpha$$

$$P\{-1.96 \sigma_a \leq S_t \leq 1.96 \sigma_a\} = 0.95$$

Y los valores atípicos pueden ser definidos como aquellos indicadores cuya sorpresa queda fuera del intervalo.

Las sorpresas pueden aplicarse tanto en la macrodepuración (si se hace sobre datos agregados) como para la microdepuración (si se hace sobre los datos elementales proporcionados por las empresas).

Sin embargo, la herramienta de control de calidad que se utilizará en esta investigación será la denominada Sorpresa Simple.

V. MEDICIONES, ANÁLISIS Y RESULTADOS

5.1 DESCRIPCIÓN DE LAS VARIABLES MACROECONÓMICAS

La serie que se estudiará en este acápite es la del Índice del Producto Bruto Interno, esta serie Peruana proviene de una encuesta que se realiza mensualmente a las empresas, y que en rigor esta a cargo de los sectores involucrados en la obtención de la misma, este índice mensual de la actividad productiva nacional muestra la evolución de la economía global y sectorial en el corto plazo, proporcionando a los usuarios un indicador sintético de la producción nacional; sin embargo, es importante resaltar el proceso de obtención y estimación del indicador del PBI mensual, la que básicamente tiene las siguientes características²⁸:

La clasificación de los sectores productivos participantes en la muestra, se basa en la clasificación industrial internacional uniforme (CIIU-Revisión 3), a fin de homogenizar el Sistema de Contabilidad Nacional.

Los encargados de realizar las encuestas mensuales a los principales agentes productivos del sector son las Oficinas

Sectoriales de Estadística e Informática (SNEI) las cuales recogen el volumen de producción mensual en cada rama de la actividad económica y son los responsables de la calidad de la información básica que llega al INEI (ente rector de las estadísticas) canalizada fundamentalmente a través de estas.

Hay Sectores como Construcción, Comercio y Otros debido a su escaso desarrollo de su infraestructura informativa, son estimados mediante métodos indirectos²⁹ y los Sectores como Agropecuaria, Pesca, Minería, Manufactura, Electricidad y Agua son estimados por métodos directos³⁰.

La valoración de la producción sectorial, se obtiene multiplicando el precio promedio del año base de cada producto por su respectivo volumen de producción. Luego, se calcula el Índice del sector mediante la comparación del valor de producción en el periodo investigado, con el valor del periodo base; por último, se calcula el Índice global de la Producción Nacional (PBIG), que se obtiene mediante la suma ponderada de los Índices Sectoriales, utilizando como factor de ponderación, la estructura porcentual del año base³¹.

28/ Para mayor detalle ver www.INEI.gob.pe informe de producción N°02 – Febrero 2002

29/ Son aquellos que carecen de registros estadísticos apropiados.

30/ Son los que tienen información básica periódica y oportuna.

31/ El año base de la serie PBIG mensual es 1994=100

5.2 DESARROLLO METODOLOGICO

Las series a desarrollar en este estudio a efectos de aplicar la metodología para el control de calidad basada en los modelos ARIMA - AI, objetivo fundamental del estudio, es el índice del PBI con año base 1994=100 que consta de nueve variables (PBI Global, Agropecuario, Pesca, Minería e Hidrocarburos, Manufactura, Electricidad y Agua, Construcción, Comercio y Otros Servicios) de 133 observaciones que va desde Enero de 1991 y terminan en Enero del 2002, el volumen de producción del Sector Minería³²: cobre, zinc, plata y petróleo que muestran 205 observaciones que van desde enero de 1985 a enero del 2002 y las otras variables son el oro, estaño, molibdeno, y gas natural con 157 observaciones que va desde 1989:01 al 2002.01. La frecuencia mensual de los datos introduce problemas de estacionalidad y un comportamiento errático, debido a que proceden de encuestas tomadas con periodicidad mensuales, la fuente³³ original de los datos es el INEI y el Ministerio de Minería.

5.2.1 Procedimiento de Modelización Automática: MODAUT

A continuación se ofrecen los resultados de la aplicación del programa MODAUT en modo automático al grupo³⁴ de las series Índices del PBI. Los detalles acerca del funcionamiento de este programa se encuentran en el informe final de consultoría de Pilar Rey y Pedro Revilla (Diciembre 2001-CIDE-INEI)

Para el análisis correcto de este grupo de variables, Índices de PBI, el análisis será

en tres etapas de acuerdo a la metodología planteada en los programas disponibles, primero será el MODAUT, luego el PREDIS y por último el Control de calidad: las SORPRESAS que se muestran a continuación:

I MODAUT.

Según la arquitectura del procedimiento automático, en esta etapa se fusionan los pasos de Identificación, Estimación, validación hasta encontrar el modelo adecuado ARIMA con Análisis de Intervención. Por lo tanto, la modelización automática (MODAUT) está basada en la metodología de modelización de Box-Jenkins & Box-Tiao.

Para correr el programa MODAUT en modo automático se consideró el tratamiento de atípicos antes y después, que no incluya las variables de calendario, que los modelos si estén afectados por la transformación en logaritmos y se seleccionó el periodo de ajuste del modelo desde enero de 1991 a julio del 2001; bajo estas especificaciones el programa nos da los resultados en pocos segundos; este a su vez, esta acompañado de un resumen de modelización correspondiente a las sucesivas interacciones o etapas de modelización de Box -Jenkins & Box-Tiao. En cada uno de estos ficheros que resumen los modelos, llamados WORK, aparecen especificaciones para cada serie, como las características más relevantes del modelo estimado en esa interacción. Además, se encuentran los test de validación no superados.

A continuación presentamos la modelización de la serie Índice del PBI que consta de nueve variables, a efectos de análisis tomaremos solo la variable Índice del PBI Global:

32/ Las salidas en modo automático de estas series se encuentran en el anexo II, los modelos, los Predis y las sorpresas.

33/ La información de las series Índice del PBI, el volumen de producción de Pesca y Minería, son proporcionadas de manera gratuita en la página web del INEI www.inei.gob.pe.

34/ La modelización automática del volumen de producción del sector Minería, puede ser consultada en el anexo II.

Cuadro 1. Salida del fichero work serie PBI Global

SERIE	MODELO ARIMA	REG	MEDIA	BIC
1001=PBIG	$(0,1,1)(0,1,1)_{12}$	0	0	-7.166
TIPOS DE INTERVENCION				
AO 25 TC 76 LS 16 AO 29				
DESVIACION TIPICA RESIDUOS			0.0245358	
P-VALORES B-P RESIDUOS			10 0.980 N 22 0.980 N 34 0.973 N	
P-VAL. B-P RESIDUOS CUADRADO			12 0.987 N 24 0.980 N 36 0.936 Y	

En el cuadro 1, podemos observar el fichero Work que trata la serie Índices del PBI Global, el modelo hasta ahora estimado es el $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos, $REG=0$ significa que el modelo ARIMA no tiene ninguna variable de intervención, la $MEDIA=0$ significa que en el modelo la media constante es cero y el estadístico BIC es el Criterio de Información Bayesiana, su utilidad es puesta de manifiesto cuando se quiere elegir entre dos modelos. Por tanto, se elige la que tenga menor BIC entre ellas, en este caso - 7.166 es el menor de los BIC que eligió este modelo.

La desviación típica de los residuos es de 2.4%, lo que significa que el modelo no esta siendo explicada en un 2.4%, también nos indica el grado de volatilidad de la serie, en este caso como es relativamente pequeña la desviación típica residual se puede considerar homogénea o se podría decir que es poco errática.

Los p-valores de Box-Pierce-Ljung (BPLs) de los residuos comprueba si este BPL es estadísticamente significativo para el caso, de los retardos 10, 22 y 34 no son significativos, lo que me indica que el residuo tiene un comportamiento de ruido blanco, osea no hay información que extraer.

El p-valor Box-Pierce-Ljung residuos cuadrados, me indican si hay o no presencia de heteroscedasticidad en los residuos; en los retardos 12 y 24 no son significativos, (no hay heteroscedasticidad), pero entre el rezago 24 y 36 existen uno o varios retardos de la función de autocorrelación altos (si hay heteroscedasticidad). El modelo se da por válido aún sin superar todos los BPLs (Obs: todo lo que aparece en estos ficheros Work es una ayuda para seguir como ha realizado el programa la modelización automática, pero no explica todo los pasos exhaustivamente).

Cuadro 2. Modelo ARIMA-AI de la serie PBI Global³⁵

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1001=PBIG	12=mensual	0=Ln	0.0245358	$(0,1,1)(0,1,1)_{12}$	0=sin cte.	4
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-		0.508534729		-		0.643329199
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS: COMPONENTE DETERMINISTICA		
2=AUTOMATICA	25(ene-93)	1=impulsos (AO)		-0.075639762		
2=AUTOMATICA	76(abr-97)	3=cambios temporales (TC)		0.066835081		
2=AUTOMATICA	16(abr-92)	2=escalones (LS)		-0.067716102		
2=AUTOMATICA	29(may-93)	1=impulsos(AO)		-0.063439989		

35/ Los cuadros de los modelos de Agro, Pesca, Minería, Industria, Electricidad, Construcción, comercio y Otros servicios pueden ser interpretados con similar detalle (véase las salidas en el anexo I).

En el cuadro 2. se observa el modelo ARIMA-AI con desviación típica residual de 2.45% la que posteriormente servirá

para la predicción y el análisis de las sorpresas. La siguiente ecuación resume los datos contenidos en el cuadro 2:

$$(1-B)(1-B^{12})LnPBIG_t = (1-0.508B)(1-0.643B^{12})a_t - 0.075(1-B)(1-B^{12})AO_{ene93} + 0.066(1-B)(1-B^{12})TC_{abr97} - 0.067(1-B)(1-B^{12})LS_{abr92} - 0.063(1-B)(1-B^{12})AO_{may93}$$

En la ecuación del modelo Índice del PBI Global presenta la componente estocástica (ARIMA) y la componente determinista (Intervención) al analizar la variable de intervención AO de enero de 1993 con parámetro igual a menos 0.0756 de la variable PBI Global se interpreta como una reducción en el nivel de ese mes del 7.56%, por causa del fenómeno atípico ocasionado en el mes de enero de 1993, sólo ha sido afectado el mes de enero, sin afectar a los meses sucesivos.

Al analizar la variable de intervención TC para abril de 1997 con un parámetro igual a 0.0668, la variable se incrementa en esa fecha en un 6.68%, y ese incremento continua en los meses posteriores en forma decreciente (en el siguiente mes de mayo de 1997 el incremento es de $6.68\% * 0.7 = 4.68\%$, al siguiente mes de junio de 1997, es de $4.68\% * 0.7 = 3.37\%$, etc donde: $\delta = 0.7$)

Al analizar abril de 1992 un LS con valor del parámetro igual a -0.0677, se puede decir que el nivel de la serie PBIG disminuye en 6.77%, disminución que se mantiene en todos los periodos sucesivos.

Así para mayo de 1993, el AO con parámetro igual a -0.063 se interpreta como una reducción en el nivel de ese mes del 6.33%, siendo la bajada únicamente en ese mes de mayo, sin afectar a los meses sucesivos.

II PREDIS.

El programa PREDIS se ejecuta de manera similar, como resultado se obtiene, un fichero de salida que incluye, para cada serie a la que se ha ejecutado los valores reales, los errores relativos de predicción³⁶ y la media de los errores de predicción³⁷ (en valor absoluto). La observación de estos datos permite evaluar el comportamiento de los modelos para predicción, el objetivo del programa PREDIS es comprobar la capacidad del modelo para predecir observaciones en un período futuro.

En este caso para correr el programa PREDIS, se utiliza el modelo identificado y validado en la etapa anterior; por lo tanto, se eligió el periodo de validación o evaluación del modelo para el periodo agosto del 2001 a enero del 2002, 6 meses a predecir, este periodo de predicción esta dentro del periodo histórico.

36/ El error relativo es en realidad una aproximación a los errores de predicción un periodo por delante en términos relativos (para que se pueda comparar directamente con la desviación teórica o desviación típica residual)

37/ La media de los errores (en valor absoluto), es lo mismo que el promedio de los errores relativos de predicción en valor absoluto.

Cuadro 3.1. Predicciones para el Índice del PBI Global³⁸

PREDICCIONES UN PERIODO POR DELANTE:1001=PBIG			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	123	121.2	1.5
Sep-01	117.8	119	-1
Oct-01	123.5	122	1.2
Nov-01	122	123	-0.8
Dic-01	124.1	124	0.1
Ene-02	122	120.4	1.3
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			1
DESVIACION TEORICA			2.5

En el cuadro 3.1 y 3.2, Se observa que la serie Índices del PBI Global tiene errores de predicción un periodo por delante que son bastante pequeños, mejor de lo esperado en media igual a 1 frente a 2.5 de la desviación teórica, mientras el modelo para la serie Pesca no es tan bueno, pues se obtienen en medias de errores mayores de lo esperado 27 frente a 20.8 de la desviación teórica, lo mismo les ocurre a las series Minería y Agropecuaria pero en menor medida que sin embargo, aunque no supere la desviación teórica a la media de errores,

se podrían dar como válidos esos modelos obtenidos para la predicción. Lo contrario ocurre a las series Industria, Construcción, Comercio y Otros servicios donde en todas estas series la media de errores es menor que la desviación teórica, porque eso es lo que se debe esperar de un modelo que sea útil para la predicción de observaciones futuras; esto es así, porque la desviación teórica que observamos ahí es la desviación típica de los residuos del modelo y estos residuos coinciden con el error de predicción un periodo por delante.

Cuadro 3.2 Estadísticos de validación de los modelos del Índice del PBI

PREDICCIONES UN PERIODO POR DELANTE:1002=AGRO	
MEDIA DE ERRORES(EN VALOR ABSOLUTO)	4.6
DESVIACION TEORICA	2.6
PREDICCIONES UN PERIODO POR DELANTE:1003=PESCA	
MEDIA DE ERRORES(EN VALOR ABSOLUTO)	27
DESVIACION TEORICA	20.8
PREDICCIONES UN PERIODO POR DELANTE:1004=MINERIA	
MEDIA DE ERRORES(EN VALOR ABSOLUTO)	3.6
DESVIACION TEORICA	2.7
PREDICCIONES UN PERIODO POR DELANTE:1005=INDUSTRIA	
MEDIA DE ERRORES(EN VALOR ABSOLUTO)	1.1
DESVIACION TEORICA	1.5
PREDICCIONES UN PERIODO POR DELANTE:1007=CONSTRUCCION	
MEDIA DE ERRORES(EN VALOR ABSOLUTO)	4.8
DESVIACION TEORICA	6.3
PREDICCIONES UN PERIODO POR DELANTE:1006=ELECTRICIDAD	
MEDIA DE ERRORES(EN VALOR ABSOLUTO)	1.1
DESVIACION TEORICA	1.5
PREDICCIONES UN PERIODO POR DELANTE:1008=Comercio	
MEDIA DE ERRORES(EN VALOR ABSOLUTO)	1.1
DESVIACION TEORICA	3.8
PREDICCIONES UN PERIODO POR DELANTE:1009=Otros serv.	
MEDIA DE ERRORES(EN VALOR ABSOLUTO)	1.1
DESVIACION TEORICA	3.8

38/ Los cuadros de las predicciones de Agro, Pesca, Minería, Industria, Electricidad, Construcción, comercio y Otros servicios son presentados completos en el anexo I.

III CONTROL DE CALIDAD: SORPRESA.

Las herramientas de control estadístico de calidad que van a ser denominadas SORPRESA, se construyen basándose en la función de predicción de los modelos ARIMA-AI, la cual constituye la herramienta más importante en la depuración selectiva de datos.

En estas series del índice del PBI, el objetivo es estudiar el mes de enero del 2002, para

evaluar si este dato real es válido o no; sin embargo, esto solo servirá para evaluar los datos de enero del 2002 de las nueve variables nada más, suponiendo que tienes valores válidos todos de la serie hasta diciembre del 2001. Por lo tanto para correr el programa se utiliza el modelo identificado y validado en la primera etapa, donde el mes y el año para el cálculo de la sorpresa será enero del 2002 y la sorpresa simple que se le solicita al MODAUT será en series originales.

**Cuadro 4. Sorpresas en predicciones un período por delante:
Índice del PBI (enero del 2002)**

SORPRESAS EN PREDICCIONES UN PERIODO POR DELANTE:PERIODO=ENERO-2002					
SERIE	VALOR REAL	PREDICCION	SORPRESA	DESVIACION TEORICA	INTERVALO DEL 95%
1001=PBIG	122	120.6	1.18	2.41	[115.0; 126.4]
1002=AGRO	128.4	121.9	5.33	2.81	[115.4; 128.8]
1003=Pesca	73.1	77.1	-5.2	21.58	[50.5; 117.7]
1004=Minería	178.1	180.9	-1.58	2.98	[170.7; 191.8]
1005=Industria	113.4	111.3	1.92	4.19	[102.5; 120.8]
1006=Electricidad	153.1	151.8	0.9	1.46	[147.5; 156.2]
1007=Construcción	120.3	116.5	3.31	6.33	[102.9; 131.9]
1008=Comercio	113.5	113.3	0.21	3.71	[105.3; 121.8]
1009=Otros serv.	113.5	113.3	0.21	3.71	[105.3; 121.8]

En el cuadro 4, podemos observar las sorpresas de las nueve series del Índice del PBI, por comodidad y por ser el objetivo del estudio, realizar control estadístico de calidad a un dato en concreto es decir en este caso, enero del 2002 el análisis de las nueve series a la vez se medirá a través de la variación relativa entre el valor real y la predicción un período por delante llamada sorpresa; si la sorpresa calculada está dentro del intervalo de $1.96 \cdot$ desviaciones teóricas (intervalo del 95% de confianza), se concluirá que no es un dato atípico, concluyéndose que el dato para el mes de análisis es válido.

Basándonos en estos principios básicos sobre la calidad de un dato, la serie Índices del PBIG, admite que el mes enero del 2002 no es un dato atípico al 95 % de confianza debido a que las sorpresas son menores que $1.96 \cdot$ desviación teórica³⁹ o que estas dentro del intervalo al 95% de confianza, sin embargo se podrían decir que hay una sorpresa relativa en la serie Índice Agropecuario, debido a que la sorpresa es un tanto cercano al valor de $1.96 \cdot 2.81 = 1.96 \cdot$ desviación teórica = 5.50 encontrándose dentro del intervalo de confianza, por lo tanto también es válido el dato para este período, y en el resto de las series se observa que las

39/ La desviación típica residual es lo mismo que la desviación teórica.

sorpresas son menores que una desviación típica, por lo tanto es obvio que será menor que $1.96 \times$ Desviación teórica. Concluyéndose de manera categórica que la calidad del dato para el mes de enero del 2002 en las nueve series son considerados válidos, suponiendo que se tienen como valores válidos todos de las series hasta diciembre del 2001. Finalmente, si observamos el Sector Pesca la sorpresa (-5.2) es menor que una desviación teórica (-21.58), con mayor motivo será para $1.96 \times$ desviación teórica concluyendo que el dato enero del 2002 no es un valor atípico, sin embargo. Su desviación teórica es de consideración, es decir la desviación típica residual a captado que el comportamiento de esta serie Índice del PBI Pesca es bastante errática y gran parte de esta información en un 21.58% no esta siendo explicada por este modelo; en el resto de las series las desviaciones teóricas son bastantes reducidas comparadas con el Sector Pesquero.

5.2.2 Procedimiento de modelización con el programa Eviews 3.1

Metodología de control de calidad basadas en los modelos ARIMA-AI, aplicado al Índice del PBI Global (PBIG)

La serie que se estudia en esta sección es la del Índice Mensual del Producto Bruto Interno Global (PBIG), los datos son obtenidos como la agregación ponderada de Los índices sectoriales que están comprendidos en el campo coyuntural, utilizando como factores de ponderación la estructura porcentual del año base $1994=100$. El índice mensual de la actividad productiva Nacional muestra la

evolución de la economía global y sectorial en el corto plazo, proporcionando a los usuario un indicador sintético de la producción nacional. La muestra utilizada va desde 1991:01 al 2002:01, es decir 133 observaciones.

1. IDENTIFICACIÓN

• Análisis exploratorio

La serie PBIG muestra una evolución creciente en la serie, con el paso del tiempo, especialmente a partir del año 1994, después de superar una leve caída en el año 1993. Esto significa que la serie no presenta un valor medio constante en todo el periodo muestral, es decir, no oscila en torno al mismo valor. Por tanto, podemos suponer ya a priori que, probablemente no será estacionaria, luego presentara al menos una raíz unitaria (ver Figura 1).

En la figura 2, podemos apreciar, que la función de autocorrelación decrece exponencialmente y de forma lenta, mientras que la función de autocorrelación parcial presenta un valor significativo en el retardo uno, con un coeficiente de autocorrelación cercano a la unidad (0.934).

Este gráfico puede considerarse como indicativo de la no estacionariedad de la serie, es decir, presenta al menos una raíz unitaria (ver Figura 2).

• Estudio de la estacionariedad

En el gráfico 1 de la serie PBIG, se observa cierto comportamiento homogéneo en la varianza y es confirmada por el estadístico de Levene de homocedasticidad de la

Figura 1. Serie PBIG: 1991:01 2002:01
(Año base 1994=100)

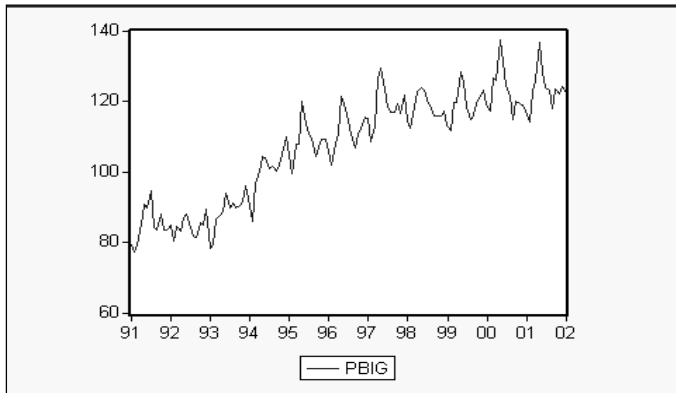


Figura 2. Autocorrelación de la Serie PBIG: 1991:01 2002:01
(Año base 1994=100)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
██████████	██████████	1 0.936	0.936	119.27	0.000
██████████		2 0.876	-0.003	224.56	0.000
██████████		3 0.828	0.061	319.21	0.000
██████████	█	4 0.808	0.214	410.13	0.000
██████████	█	5 0.823	0.309	505.20	0.000
██████████	█	6 0.814	-0.123	598.89	0.000
██████████		7 0.798	0.026	689.53	0.000
██████████		8 0.757	-0.114	771.88	0.000
██████████	█	9 0.732	0.118	849.50	0.000
██████████	█	10 0.738	0.164	929.11	0.000
██████████	█	11 0.754	0.096	1012.7	0.000
██████████	█	12 0.763	-0.029	1099.2	0.000
██████████	█	13 0.713	-0.374	1175.2	0.000
██████████		14 0.659	-0.022	1240.7	0.000
██████████		15 0.611	-0.048	1297.6	0.000
██████████		16 0.589	0.001	1350.8	0.000
██████████	█	17 0.602	0.088	1407.0	0.000
██████████	█	18 0.582	-0.198	1459.9	0.000
██████████		19 0.554	-0.021	1508.1	0.000
██████████		20 0.507	0.001	1549.0	0.000
██████████		21 0.474	0.048	1584.9	0.000
██████████		22 0.472	0.004	1621.0	0.000
██████████		23 0.481	0.086	1658.9	0.000
██████████		24 0.481	-0.081	1696.9	0.000

varianza al no rechazar la hipótesis nula de homocedasticidad (ver tabla 1), por tanto se puede decir que la serie es estacionaria en varianza, a efectos de

interpretación se decidió trabajar con la transformación logarítmica que literalmente denominamos LPBIG = $\log(\text{PBIG})$.

Tabla 1. Prueba de homocedasticidad en varianza a la serie PBIG (Año base 1994=100)

Test of Homogeneity of Variance ^a					
		Levene Statistic	df1	df2	Sig.
PBIG	Based on Mean	.713	10	121	.710
	Based on Median	.516	10	121	.876
	Based on Median and with adjusted df	.516	10	100.463	.875
	Based on trimmed mean	.672	10	121	.748
a. PBIG is constant when YEAR, not periodic = 2002. It has been omitted.					

Los resultados del estadístico de Dickey-Fuller Aumentado (ADF) mostradas en la Tabla 2 son la que nos permiten identificar la presencia de raíces unitarias. Como se puede apreciar, en los casos de los niveles del PBIG y el logaritmo del PBIG (LPBIG) no se puede rechazar la hipótesis nula de que la serie es no estacionaria. Al tomar las primeras diferencias al logaritmo del PBIG (DLPBIG) si se puede rechazar la H_0 , lo cual indica que estas son efectivamente estacionarias, confirmando que la serie en

primera diferencia del PBIG con transformación logarítmica es integrada de orden $I(1)$ o $d=1$, por lo que se decidió utilizar las primeras diferencias de los logaritmos y no de la serie original PBIG básicamente para evitar problemas de escala; también, este resultado se puede observar gráficamente que la serie DLPBIG = $(1-B)LPBIG$ es estacionaria en medias, es decir no hay presencia de tendencia (véase figura 3).

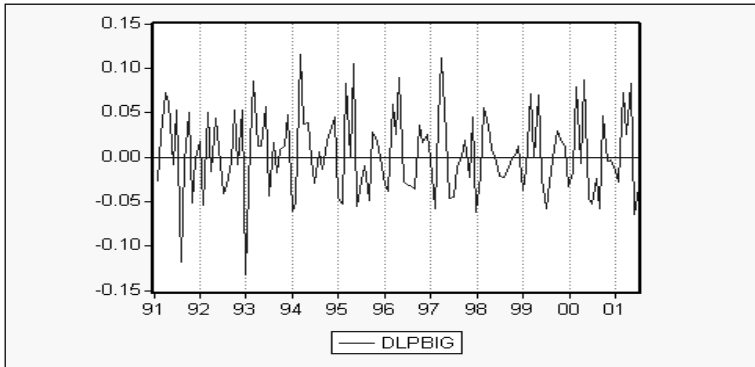
Tabla 2. resultados del estadístico de Dickey-Fuller Aumentado (ADF) en la serie PBIG, LPBIG Y D(LPBIG).

Augmented Dickey-Fuller Unit Root Test on PBIG				
ADF Test Statistic	-1.837906	1% Critical Value*	-3.4839	
		5% Critical Value	-2.8847	
		10% Critical Value	-2.5790	
*MacKinnon critical values for rejection of hypothesis of a unit root.				

Augmented Dickey-Fuller Unit Root Test on LPBIG				
ADF Test Statistic	-1.877031	1% Critical Value*	-3.4839	
		5% Critical Value	-2.8847	
		10% Critical Value	-2.5790	
*MacKinnon critical values for rejection of hypothesis of a unit root.				

Augmented Dickey-Fuller Unit Root Test on D(LPBIG)				
ADF Test Statistic	-9.141820	1% Critical Value*	-3.4843	
		5% Critical Value	-2.8849	
		10% Critical Value	-2.5791	
*MacKinnon critical values for rejection of hypothesis of a unit root.				

Figura 3. Serie DLPBIG = (1-B)LPBIG



En la figura 4 se incluyen los valores de los treinta y seis primeros coeficientes de autocorrelación y autocorrelación parcial, así como su representación gráfica, correspondiente a la serie DLPBIG. Se aprecia, en el gráfico de la función de autocorrelación, un coeficiente

sistemáticamente significativo para los retardos estacionales 12, 24 y 36, indicándonos la existencia de tendencia en la parte estacional, que habrá que corregir para proceder a la correcta identificación de los procesos ARIMA presentes en esta serie.

Figura 4. Autocorrelación de la serie DLPBIG = (1-B)LPBIG

Autocorrelación	Partial Correlation	AC	PAC	Q-Stat	Prob
█	█	1 -0.085	-0.085	0.9318	0.334
█	█	2 -0.127	-0.136	3.0141	0.222
█	█	3 -0.220	-0.250	9.3445	0.025
█	█	4 -0.259	-0.360	18.208	0.001
█	█	5 0.190	0.023	23.041	0.000
█	█	6 0.151	0.038	26.116	0.000
█	█	7 0.064	-0.010	26.677	0.000
█	█	8 -0.181	-0.206	31.133	0.000
█	█	9 -0.204	-0.186	36.844	0.000
█	█	10 -0.132	-0.267	39.251	0.000
█	█	11 0.083	-0.214	40.221	0.000
█	█	12 0.587	0.434	68.945	0.000
█	█	13 -0.005	0.136	68.949	0.000
█	█	14 -0.068	0.108	69.617	0.000
█	█	15 -0.259	-0.005	99.368	0.000
█	█	16 -0.305	-0.198	113.04	0.000
█	█	17 0.336	0.137	129.72	0.000
█	█	18 0.043	-0.111	130.00	0.000
█	█	19 0.082	-0.014	131.00	0.000
█	█	20 -0.158	-0.067	134.81	0.000
█	█	21 -0.233	-0.029	143.12	0.000
█	█	22 -0.122	-0.184	145.44	0.000
█	█	23 0.188	-0.008	150.95	0.000
█	█	24 0.386	0.020	174.52	0.000

█	█	25 0.000	0.005	175.50	0.000
█	█	26 0.040	0.065	176.90	0.000
█	█	27 -0.257	0.108	186.68	0.000
█	█	28 -0.282	-0.042	187.97	0.000
█	█	29 0.278	-0.005	210.64	0.000
█	█	30 0.045	-0.094	210.88	0.000
█	█	31 0.054	0.075	212.37	0.000
█	█	32 -0.132	-0.037	215.37	0.000
█	█	33 -0.218	0.088	223.64	0.000
█	█	34 0.024	0.005	223.86	0.000
█	█	35 0.132	-0.009	226.96	0.000
█	█	36 0.382	0.117	290.41	0.000

Por lo tanto es necesario extraer el comportamiento estacional con una diferencia $D=1$ en la parte estacional, entonces la transformación será de la siguiente manera $(1-B)(1-B^{12})LPBIG = DDLPBIG$, el resultado de la transformación es confirmado por el gráfico del correlograma. En la Figura 5, se observa que el efecto

sistemático en los periodos de estacionalidad han sido extraídos; por lo tanto se puede decir que se ha corregido la existencia de tendencia en la parte estacional. A continuación, procedemos, de nuevo, a la obtención de las funciones de autocorrelación y autocorrelación parcial, pero ahora de la última serie generada DDLPBIG o DLOG(PBIG,1,12)

Figura 5. Serie DDLPBIG = $(1-B)(1-B^{12})LPBIG$

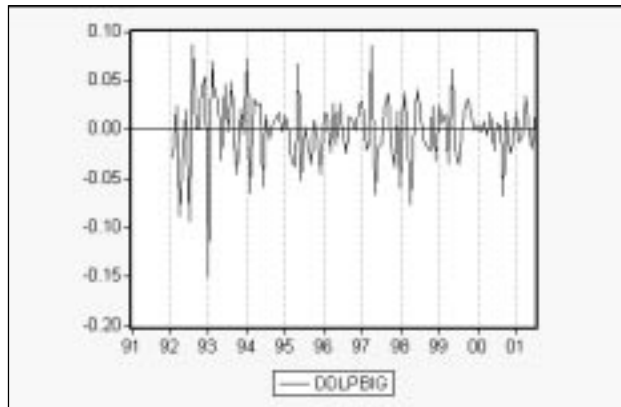


Figura 6. Autocorrelación y autocorrelación parcial de la Serie DDLPBIG = $(1-B)(1-B^{12})LPBIG$

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
█	█	1 -0.370	-0.370	16.047	0.000
█	█	2 -0.061	-0.229	16.483	0.000
█	█	3 0.080	-0.039	17.254	0.001
█	█	4 0.036	0.053	17.408	0.002
█	█	5 -0.112	-0.071	18.922	0.002
█	█	6 0.187	0.151	23.222	0.001
█	█	7 -0.206	-0.122	28.492	0.000
█	█	8 0.046	-0.050	28.753	0.000
█	█	9 0.151	0.128	31.621	0.000
█	█	10 -0.071	0.046	32.261	0.000
█	█	11 0.045	0.137	32.518	0.001
█	█	12 -0.149	-0.192	35.415	0.000
█	█	13 -0.036	-0.174	35.588	0.001
█	█	14 0.016	-0.155	35.621	0.001
█	█	15 0.038	-0.062	35.809	0.002
█	█	16 -0.160	-0.105	39.266	0.001
█	█	17 0.184	0.079	43.858	0.000
█	█	18 -0.086	0.023	44.885	0.000
█	█	19 0.005	-0.030	44.888	0.001
█	█	20 -0.010	-0.066	44.902	0.001
█	█	21 -0.023	-0.052	44.977	0.002
█	█	22 -0.108	-0.084	46.661	0.002
█	█	23 0.144	0.047	49.690	0.001
█	█	24 -0.213	-0.216	56.328	0.000

█	█	25 0.180	-0.080	57.800	0.000
█	█	26 0.187	0.014	59.512	0.000
█	█	27 -0.030	0.066	59.648	0.000
█	█	28 -0.120	-0.109	61.855	0.000
█	█	29 0.153	0.037	65.481	0.000
█	█	30 -0.126	-0.042	67.977	0.000
█	█	31 -0.005	-0.103	67.982	0.000
█	█	32 0.044	-0.094	68.293	0.000
█	█	33 0.013	0.065	68.322	0.000
█	█	34 -0.032	-0.035	68.482	0.000
█	█	35 0.127	0.066	71.252	0.000
█	█	36 -0.077	-0.127	72.217	0.000

esto lo podemos observar en la figura 6 que nos muestra que los retardos estacionales 12, 24 y 36 han sido corregidos. En la figura 5 se destaca un valor anormalmente alto para el periodo 1993:01 y otros como el de 1993:05, 1997:04 más leves que sería importante darle algún tipo de tratamiento más adelante.

En la tabla 3 se observa diferentes transformaciones que convierte a LPBIG en estacionaria, los criterios de desviación estándar mínima y de correlograma más simple nos llevan a elegir a la transformación $DDL\text{PBIG}=(1-B)(1-B^{12})\text{LPBIG}$ frente a las transformaciones $DD2\text{LPBIG}=(1-B)^2(1-B^{12})$, $D\text{LPBIG}=(1-B)\text{LPBIG}$ y $D2\text{LPBIG}=(1-B)^2\text{LPBIG}$ como la mejor transformación para LPBIG.

Tabla 3. Desviación estándar de las transformaciones a LPBIG

	DLPBIG	D2LPBIG	DD2LPBIG	DDLBPBIG
Mean	0.003497	0.006676	-0.000231	-5.99E-06
Median	-0.002005	0.004073	0.003919	0.001961
Maximum	0.115950	0.153171	0.104006	0.086184
Minimum	-0.132128	-0.117322	-0.105504	-0.149910
Std. Dev.	0.043752	0.059428	0.040766	0.036169
Skewness	0.293307	0.088633	-0.184305	-0.695161
Kurtosis	3.131806	2.514384	3.050293	5.114837

• **Identificación de posibles modelos ARIMA.**

Hemos comprobado que la serie LPBIG presenta tendencia creciente ($d=1$). Además, al presentar estacionalidad de periodo $s=12$, ha sido necesario aplicar una diferencia en la parte estacional para eliminarla ($D=1$). En consecuencia, el modelo que se ajustará será un $ARIMA(p,1,q)(P,1,Q)_{12}$ en logaritmos.

La identificación de los parámetros p y q , órdenes de los polinomios autorregresivos y de medias móviles de la parte regular del modelo, se realizará a partir de las funciones de autocorrelación simple y parcial de la serie diferenciada regular y estacionalmente (véase figura 6), en la figura 6 se observa que el primer coeficiente de la autocorrelación (AC) es

no nulo de -0.370 y que decrece al aumentar los retardos. Por otro lado, en la autocorrelación parcial (PAC) aunque los dos primeros retardos están fuera de los límites de la banda de confianza, el único significativamente grande es el primero con -0.37 . Una primera tentativa sería suponer que la serie LPBIG diferenciada regular y estacionalmente presenta, en la parte regular, la estructura de un modelo de medias móviles de orden q igual a 1. sin embargo, el hecho de que al coeficientes del segundo retardo de la PAC no puede ser considerado nulo hace pensar en la posibilidad de un modelo autorregresivo de medias móviles de ordenes p y q distintos de cero. Supondremos entonces que la estructura es la de este segundo modelo y, en particular, que $p=q=1$. si, al ajustarlo, el coeficiente asociado al parámetro autorregresivo no fuera significativamente distinto de cero se

recurriría al modelo de medias móviles de orden $q=1$.

La identificación de los parámetros P y Q , órdenes de autorregresión y de medias móviles de la parte estacional del modelo, se realizará a partir de las funciones de autocorrelación simple y parcial para la serie diferenciada regular y estacionalmente (véase la figura 6), sobre los 3 primeros retardos estacionales, de la serie DDLPBIG, tanto en la parte de la autocorrelación simple como en la autocorrelación parcial, respectivamente, todos los coeficientes están dentro o muy próximos, a los límites de la banda de confianza, por lo que no hay evidencia de ninguna estructura estacional. Por ello, inicialmente, consideraremos que la serie DDLPBIG diferenciada regular y estacionalmente no presenta ninguna

estructura en la parte estacional y, en consecuencia, ajustaremos un modelo en logaritmos. Una vez ajustada, analizaremos las AC y las PAC de los residuos en busca de alguna estructura en la parte estacional. Caso de no encontrarla, el análisis proseguirá sobre el modelo identificado.

2. Estimación de los posibles modelos ARIMA

Una vez identificados los posibles órdenes empezaremos a estimar el modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos, la tabla 4 el diálogo proporciona la estimación de los parámetros del modelo, en la que el parámetro de la componente autorregresivo no es significativo, entonces el modelo se reduce a un modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos.

Tabla 4. Ajuste del modelo $ARIMA(1,1,1)(0,1,0)_{12}$ en logaritmos

Dependent Variable: DLOG(PBIG,1,12)				
Method: Least Squares				
Date: 06/27/02 Time: 11:45				
Sample(adjusted): 1992:03 2001:07				
Included observations: 113 after adjusting endpoints				
Convergence achieved after 7 iterations				
Backcast: 1992:02				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.071833	0.178910	0.401504	0.6888
MA(1)	-0.539077	0.150049	-3.592680	0.0005
R-squared	0.184628	Mean dependent var	-0.000374	
Adjusted R-squared	0.177283	S.D. dependent var	0.037018	
S.E. of regression	0.033576	Akaike info criterion	-3.932446	
Sum squared resid	0.125139	Schwarz criterion	-3.884174	
Log likelihood	224.1832	Durbin-Watson stat	2.011409	
Inverted AR Roots	.07			
Inverted MA Roots	.54			

Se procede a estimar el modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos y que sea el examen de los residuos de este modelo el que nos oriente en la

identificación del factor estacional. (Véase la tabla 5), al analizar el diálogo se observa que el parámetro de la media móvil es significativo.

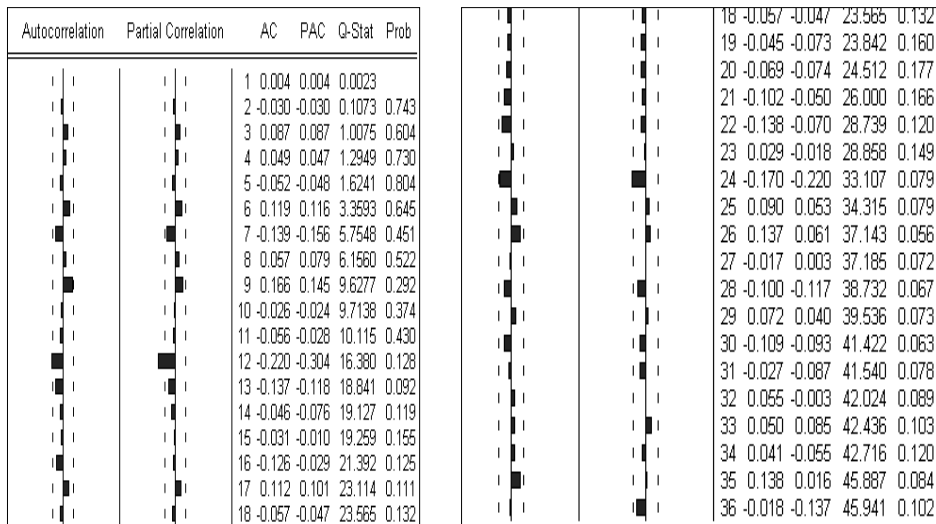
Tabla 5. Ajuste del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos

Dependent Variable: DLOG(PBIG,1,12)				
Method: Least Squares				
Date: 06/27/02 Time: 11:49				
Sample(adjusted): 1992:02 2001:07				
Included observations: 114 after adjusting endpoints				
Convergence achieved after 4 iterations				
Backcast: 1992:01				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.470561	0.082161	-5.727277	0.0000
R-squared	0.179515	Mean dependent var	-0.000617	
Adjusted R-squared	0.179515	S.D. dependent var	0.036945	
S.E. of regression	0.033465	Akaike info criterion	-3.947926	
Sum squared resid	0.126546	Schwarz criterion	-3.923924	
Log likelihood	226.0318	Durbin-Watson stat	1.984640	
Inverted MA Roots	.47			

Al analizar las autocorrelaciones AC y la PAC de la serie residual del modelo

$ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos (véase figura 7) sobre los primeros retardos

Figura 7. Correlograma residual del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos



Estacionales en las autocorrelaciones simple y parcial, proporcionadas por la figura 7, se observa que, tanto en la parte AC como en la PAC, únicamente el primer coeficiente está un tanto fuera de los

limites de la banda de confianza, siendo más significativo en el caso de la autocorrelación parcial (PAC). Sin embargo podemos observar que estos retardos estacionales no son significativos, tal como

se observa en el correlograma los retardos 12, 24 y 36. Por lo tanto los residuos a través del análisis de la función de autocorrelación simple y parcial tiene un comportamiento de ruido blanco (puramente aleatorios, comprobando que todos cumplen con la hipótesis nula de la no presencia de autocorrelación). Podríamos estar frente al posible modelo ARIMA seleccionado.

El hecho de que el modelo identificado sea adecuado no debe hacer descartar la

posibilidad de que otro modelo algo más complejo pueda ajustarse mejor a la serie observada. Entonces se procede realizar la técnica del sobreajuste:

De los cuatro sobreajustes al modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos (ver tabla 6), solo aparecen aquellos que sus parámetros al ser estimados son significativos, dejando de lado aquellos que no lo son.

Tabla 6. Sobreajuste al modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos

Periodo de ajuste 1991:01 al 2001.07			
Criterios de validez	MODELOS		
	MA(1)	MA(1)SMA(12)	SAR(12)MA(1)
Adjusted R-squared	0.179515	0.529461	0.19035
Sum squared resid	0.126546	0.071931	0.080904
Log likelihood	226.0318	258.2313	219.381
Akaïke info criterion	-3.947926	-4.495286	-4.262372
Schwarz criterion	-3.923924	-4.447283	-4.210902
Durbin-Watson stat	1.98464	2.003222	1.879893
Q-7(p-valor)	(5.754)0.451	(9.98)0.076	(7.32)0.198
Q-14(p-value)	(19.12)0.119	(23.69)0.022	(14.59)0.265
Q-21(p-value)	(26.00)0.166	(31.76)0.033	(20.21)0.381

Como podemos observar en la tabla 6 y 7 los tres modelos que se presentan como posibles seleccionados poseen coeficientes significativamente diferentes a cero, y cumplen con la estacionariedad e invertibilidad como se puede observar el modelo $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos posee una menor suma de residuos al cuadrado y un mayor R-cuadrado ajustado, por lo que este modelo presenta un mejor grado de ajuste. También podríamos analizar el estadístico de D-W que se observa en la Tabla 6, los tres modelos ARIMA me están indicando que sus residuos tienen un comportamiento de ruido blanco y por lo tanto no presentan autocorrelación de

ningún orden, el valor ideal es $D-W=2$, también esa cualidad esta siendo absorbida por los otros dos modelos, sin embargo este estadístico solo podemos utilizarlo de forma aproximada, es más conveniente realizar una identificación de los residuos a través del estadístico de Q de Ljung-Box.

En lo que respecta a los indicadores de parsimonia, el AIC y SBC favorece también al modelo $ARIMA(0,1,1)(0,1,1)_{12}$. Finalmente, sin embargo este modelo falla en el análisis del Estadístico desarrollado por Ljung-Box en la que presenta valores mayores para este modelo, por lo que podemos decir que hay dependencia entre

el residuo y su valor en el periodo anterior. De esta manera los valores de $Q(7)$, $Q(14)$ Y $Q(21)$ no cumplen con la hipótesis nula de la no presencia de autocorrelación, en este caso el modelo esta rechazando esta condición, por lo tanto queda descartado por el momento el modelo $ARIMA(0,1,1)(0,1,1)_{12}$ ⁴⁰ por lo que se debería mejorar su especificación, mientras que entre el modelo $ARIMA(0,1,1)(0,1,0)_{12}$ y $ARIMA(0,1,1)(1,1,0)_{12}$, por el momento

seguiremos el análisis con el modelo $ARIMA(0,1,1)(0,1,0)_{12}$ con un sólo parámetro en la parte regular de media móvil. Los otro dos modelos tanto el modelo $ARIMA(0,1,1)(0,1,1)_{12}$ como el modelo $ARIMA(0,1,1)(1,1,0)_{12}$ lo usaremos en la etapa de predicción para comparar sus resultados y elegiremos aquel modelo cuya capacidad predictiva sea mayor.

Tabla 7. Parámetros de los modelos en logaritmos

Dependent Variable: DLOG(PBIG,1,12)				
Method: Least Squares				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.469417	0.083404	-5.628261	0
SMA(12)	-0.885445	0.000105	-8428.096	0
Inverted MA Roots	0.99	.86 -.49i	.86+.49i	.49 -.86i
	.49+.86i	0.47	.00+.99i	-.00 -.99i
	-.49+.86i	-.49 -.86i	-.86+.49i	-.86 -.49i
-0.99				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.470561	0.082161	-5.727277	0
Inverted MA Roots	0.47			
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	-0.249971	0.082625	-3.02538	0.0032
MA(1)	-0.489957	0.083052	-5.899393	0
Inverted AR Roots	.86+.23i	.86 -.23i	.63+.63i	.63 -.63i
	.23 -.86i	.23+.86i	-.23+.86i	-.23 -.86i
	-.63 -.63i	-.63 -.63i	-.86 -.23i	-.86+.23i
Inverted MA Roots	0.49			

3. Diagnóstico al modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos elegido

- Incorrelación**

En la figura 7, el correlograma del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos analizada por el estadístico de Q de Ljung-Box, determina que hay ausencia de autocorrelación en los residuos, es decir

el comportamiento se asemeja al de un ruido blanco, Se observa también que todos los coeficientes caen dentro de la banda de confianza o en casos extremos están en el limite de la banda al 95% de confianza, además todos los p-valores asociados al estadístico de Ljung-Box para cada retardo (p-value) son lo suficientemente grandes como para no rechazar la hipótesis nula de que, en cada caso, todos los coeficientes anteriores, hasta el correspondiente, son nulos. En

40/ Véase Tragodara, Carlos. Econometría moderna. Sep-2001.Universidad Pacifico.

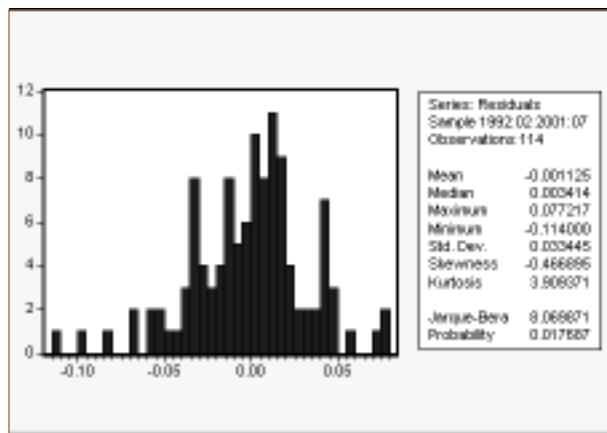
consecuencia podemos concluir que los residuos del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos están incorrelados.

- **Normalidad**

El estadístico de Jarque-Bera se utiliza para determinar si la serie de residuos sigue una distribución normal; en la figura 8 se muestra el resultado del histograma residual, en este caso el valor bajo de la probabilidad conduce a rechazar la hipótesis nula de una distribución normal

en definitiva, como el valor obtenido del estadístico del Jarque-Bera es superior al valor de referencia de tablas (aproximadamente, 6) y la probabilidad es menor que un $\alpha = 5\%$, podemos suponer que la serie residual no presenta características que se asemeja a una normal. Sin embargo; amparándonos en el teorema central del limite, podemos decir que el trabajar con series de tiempo lo que implícitamente supone el uso de una muestra grande, asegura que nuestros errores sean asintóticamente normales.)⁴¹

Figura 8. Histograma residual del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos



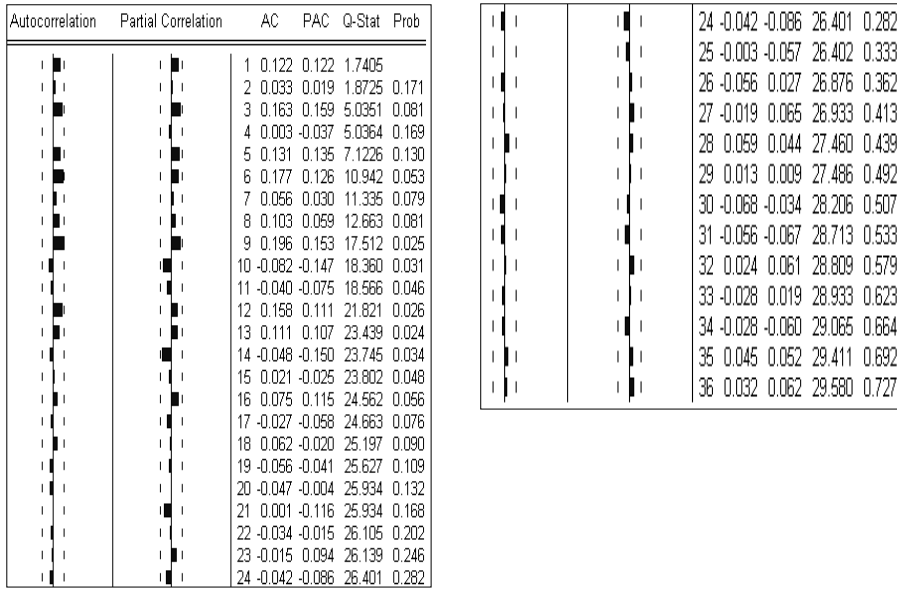
- **Heteroscedasticidad residual.**

Una variante al estadístico Q de Ljung-Box, es el test de los residuos al cuadrado, que se utiliza, principalmente, para analizar la presencia de una heteroscedasticidad condicional autorregresiva (ARCH) en los residuos, en la figura 9 presentamos el resultado del correlograma residual al cuadrado del modelo $ARIMA(0,1,1)(0,1,0)_{12}$

en logaritmos, al analizar el estadístico Q, podemos aceptar la hipótesis nula que establece que los residuos no siguen un comportamiento ARCH, también se observa que los residuos al cuadrado son incorrelacionados. Por lo tanto no hay rezagos iniciales significativos que recoger, por lo tanto no se tiene un comportamiento ARCH en los residuos de este modelo.

41/ Véase, Casas Tragodara, Carlos. Econometría moderna. Teorema central del limite. Sep-2001.

Figura 9. Correlograma de los residuos al cuadrado del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos



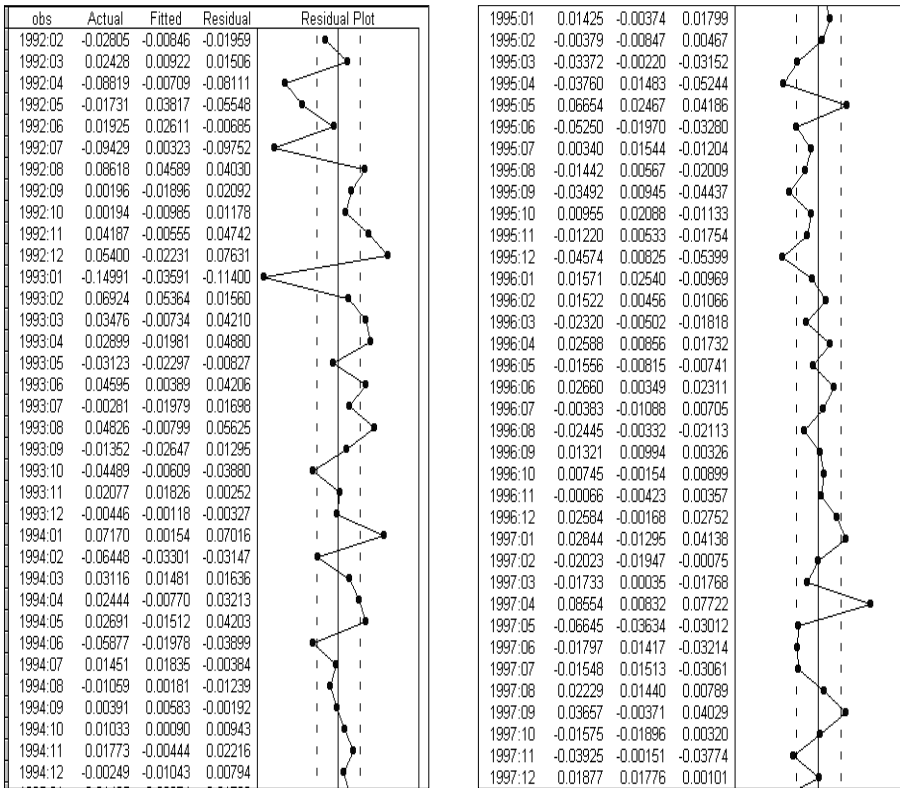
3.1 Análisis de Intervención

En la Figura 7 del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos, se comprobó que los residuos siguen un comportamiento de ruido blanco, es decir los residuos son incorrelacionados.

En la figura 10 los residuos del modelo en análisis presenta movimientos bruscos que no será posible ser captados por una dependencia sistemática de su pasado, con respecto a los residuos del modelo $ARIMA(0,1,1)(0,1,0)_{12}$, resulta de especial interés centrarse en los puntos raros o valores atípicos. En este caso, el principal error del modelo se sitúa en el dato enero del año 1993, en el que se

observa un importante descenso en el Índice del Producto Bruto Interno Global, el modelo proporciona un dato ajustado muy superior para enero de 1993 (columnas actual y fitted, respectivamente), dando lugar a un error cuantificado de -0.114 ¿Por qué se produce un error tan importante es ese momento? ¿Puede ser a un error de transcripción del dato? ¿Se produjo en esta fecha algún fenómeno climático como el fenómeno del niño? ¿O se debe al cambio del año base?. En consecuencia es importante estudiar su causa; para esta fecha enero de 1993 y sería importante saber exactamente que ocurrió, esto podría ayudarnos a mejorar o incluso a replantear el modelo de predicción.

Figura 10. Residuos del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos



El tratamiento de este problema lo haremos mediante el análisis de intervención que consiste en introducir variables artificiales afectadas por un filtro

ARMA igual a $\frac{\omega_s(B)}{\delta_r(B)}$ (la estimación se

realizará conjuntamente con el modelo $ARIMA(0,1,1)(0,1,0)_{12}$, en logaritmos y con parámetro $\theta_1 = -0.47$), con ello el modelo ahora toma la forma:

$$Lny_t = \frac{(1 + 0.47B)}{(1 - B)(1 - B^{12})} a_t + \omega_0(1 - B)(1 - B^{12})DF_t$$

donde :

ω_0 = es el parámetro en el filtro DF_t , en el filtro $r=0$ y $s=0$

D = es la variable impulso (AO)

F_t = enero de 1993

Entonces la variable artificial OA utilizada que denominaremos $D931$, toma el valor uno en enero de 1993 y el valor cero en todos los demás puntos.

Los resultados de este ajuste se pueden observar en la Tabla 8, donde la variable de intervención Impulso $D931$ es significativa con parámetros $D931$ igual a menos 0.095, que puede ser interpretada

como en una reducción del PBI Global en el nivel de ese mes del 9.5%, siendo la bajada únicamente en ese mes, sin afectar a los meses sucesivos Y $\theta_1 = -0.35$, hasta

el momento podemos afirmar que el modelo estimado es adecuado, en el sentido de que todos sus coeficientes son significativamente distintos de cero.

Tabla 8. Resultados del modelo ARIMA con Análisis de Intervención

Dependent Variable: DLOG(PBIG,1,12)				
Method: Least Squares				
Date: 07/02/02 Time: 09:32				
Sample(adjusted): 1992:02 2001:07				
Included observations: 114 after adjusting endpoints				
Convergence achieved after 8 iterations				
Backcast: 1992:01				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D931	-0.095720	0.031456	-3.042985	0.0029
MA(1)	-0.351228	0.091248	-3.849164	0.0002
R-squared	0.233749	Mean dependent var	-0.000617	
Adjusted R-squared	0.226908	S.D. dependent var	0.036945	
S.E. of regression	0.032484	Akaike info criterion	-3.998768	
Sum squared resid	0.118181	Schwarz criterion	-3.950764	
Log likelihood	229.9297	Durbin-Watson stat	2.036838	
Inverted MA Roots	.35			

En este caso si comparamos el modelo $ARIMA(0,1,1)(0,1,0)_{12}$ con y sin intervención mostradas en la Tabla 9 se puede comentar que, se logra un mejor ajuste en el modelo con intervención, debido a que la desviación típica residual es un tanto menor que el modelo sin intervención, los estadísticos de parsimonia el AIC Y SC también son menores y el R cuadrado ajustado es mayor. Ahora que sabemos que el ajuste ha mejorado entonces a este modelo ARIMA con intervención le haremos un diagnóstico al residuo.

En la figura 11 al analizar los diferentes rezagos todos aceptan la hipótesis de aleatoriedad es decir los residuos son incorrelados.

Si analizamos la figura 12 podemos observar que la probabilidad del estadístico de Jarque y Bera no rechaza la hipótesis nula y podemos afirmar que se asemeja a una normal y el estadístico ARCH prueba que la relación cuadrática entre el residuo y su valor retardado no tienen una relación de dependencia. Por lo tanto, no es conveniente construir un modelo ARCH (ver tabla 10)

Tabla 9. Resultados del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos con y sin intervención

Periodo de ajuste 1991.01 al 2001:07		
criterios de validez	MODELOS	
	MA(1)D931	MA(1)
Adjusted R-squared	0.226908	0.179515
S.E. of regression	0.032484	0.033465
Sum squared resid	0.118181	0.126546
Log likelihood	229.9297	226.0318
Akaike info criterion	-3.998768	-3.947926
Schwarz criterion	-3.950764	-3.923924
Durbin-Watson stat	2.036838	1.98464
Q-7(p-value)	7.65(0.264)	5.75(0.451)
Q-14(p-value)	15.38(.0284)	19.12(0.119)
Q-21(p-value)	20.82(0.408)	26.00(0.166)

Figura 11. Correlación del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ con intervención (AO=D931)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.022	-0.022	0.0559	
		2 0.021	0.021	0.1093	0.741
		3 0.144	0.145	2.5730	0.276
		4 0.070	0.078	3.1670	0.367
		5 0.018	0.017	3.2071	0.524
		6 0.106	0.086	4.5893	0.468
		7 -0.158	-0.179	7.6552	0.264
		8 0.004	-0.021	7.6571	0.364
		9 0.070	0.049	8.2759	0.407
		10 -0.039	-0.004	8.4699	0.488
		11 -0.038	-0.016	8.6579	0.565
		12 -0.177	-0.209	12.737	0.311
		13 -0.142	-0.140	15.382	0.221
		14 0.005	-0.015	15.386	0.284
		15 0.023	0.089	15.456	0.348
		16 -0.123	-0.017	17.497	0.290
		17 0.112	0.150	19.212	0.258
		18 -0.064	-0.052	19.778	0.286
		19 -0.035	-0.085	19.948	0.336
		20 -0.041	-0.127	20.188	0.383
		21 -0.067	-0.071	20.821	0.408
		22 -0.111	-0.050	22.591	0.366
		23 0.075	0.066	23.417	0.379
		24 -0.184	-0.172	28.413	0.201
		25 0.094	0.071	29.715	0.194
		26 0.114	0.103	31.675	0.168
		27 -0.053	0.009	32.105	0.190
		28 -0.085	-0.107	33.223	0.190
		29 0.070	0.029	33.990	0.201
		30 -0.138	-0.121	36.997	0.146
		31 -0.053	-0.150	37.449	0.164
		32 0.020	-0.062	37.516	0.195
		33 0.028	0.126	37.647	0.226
		34 -0.003	-0.005	37.649	0.265
		35 0.104	0.106	39.462	0.239
		36 -0.037	-0.084	39.700	0.268

Figura 12. Histograma residual del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ con intervención (AO=D931)

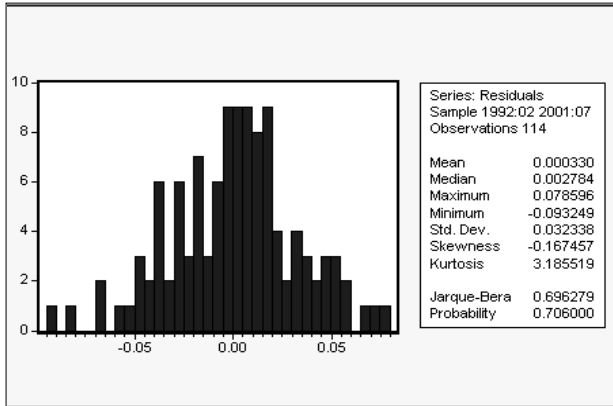


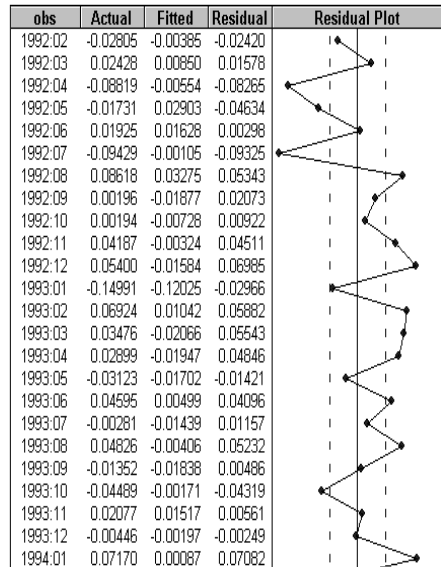
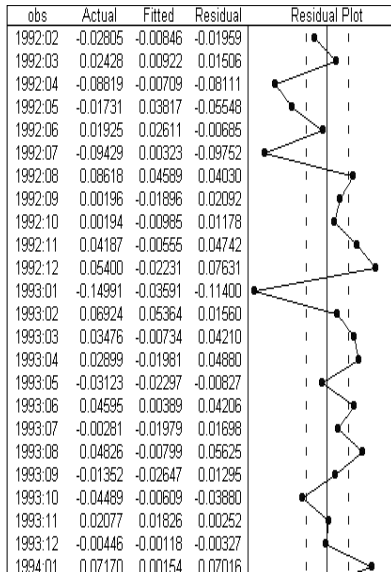
Tabla 10. Prueba ARCH al modelo $ARIMA(0,1,1)(0,1,0)_{12}$ con intervención (AO D931)

ARCH Test:				
F-statistic	0.489090	Probability	0.485796	
Obs*R-squared	0.495719	Probability	0.481387	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 07/03/02 Time: 11:56				
Sample(adjusted): 1992:03 2001:07				
Included observations: 113 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000971	0.000176	5.520265	0.0000
RESID^2(-1)	0.066297	0.094799	0.699350	0.4858
R-squared	0.004387	Mean dependent var	0.001041	
Adjusted R-squared	-0.004583	S.D. dependent var	0.001543	
S.E. of regression	0.001547	Akaike info criterion	-10.08785	
Sum squared resid	0.000266	Schwarz criterion	-10.03958	
Log likelihood	571.9636	F-statistic	0.489090	
Durbin-Watson stat	1.997989	Prob(F-statistic)	0.485796	

Los residuos de la estimación del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos con intervención D931 se dan en el figura 13, obsérvese que el error atípico de enero

de 1993 ha desaparecido o sea ha sido recogido por el modelo y no aparecen atípicos alrededor de esta fecha donde se realizó la intervención.

Figura 13. Residuos de estimación del modelo sin intervención y con intervención (AO=Enero-1993)



4. Predicción

Para construir modelos con fines de hacer control de calidad, el método consiste en desarrollar un modelo basado en las observaciones en el período histórico (1991:01 al 2001:07) y generar, a partir de él, predicciones para el período de validación (2008:01 2002:01). La comparación entre las predicciones y las observaciones en el periodo de validación permitirá evaluar la capacidad predictiva del modelo. Finalmente tenemos tres modelos elegidos para una misma serie (Índice del PBI Global) a la que evaluaremos su capacidad predictiva; los modelos son $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos y con intervención en enero de 1993, $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos y $ARIMA(0,1,1)(1,1,0)_{12}$ en logaritmos, se elegirá aquel modelo cuya capacidad predictiva sea mayor. A

continuación presentamos los resultados para evaluar la capacidad predictiva de los modelos elegidos, el análisis de las tablas 11, 12 y 13 se puede apreciar a primera vista que, en forma bastante general, los tres modelos generan predicciones relativamente buenas para el PBI Global. Utilizando el primer indicador que nos permite comparar fácilmente resultados entre diferentes modelos se observa que los errores relativos de predicción con un período adelante son bastante pequeños en media, pero sin embargo el modelo $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos, tiene mejor capacidad predictiva, debido a que la media de errores en valor absoluto (0.86) es menor de lo esperado frente a la desviación teórica (2.63), y esta a su vez es igual a la desviación típica residual, también al observar los gráficos 14, 15 y 16 la raíz del error cuadrático medio del modelo $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos es

un tanto menor (1.07), en comparación con los otros modelos que tienen raíz de errores cuadráticos medios de 4.44 y 2.94 respectivamente. Por lo tanto, el modelo que mejor comportamiento tiene para

predicción es el modelo $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos, que posteriormente se utilizará para el control estadístico de la calidad a través de las sorpresas (Revilla y Rey 1991)

Figura 14. Predicciones y límites del intervalo de confianza del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos con análisis de intervención (AO=Enero-93)

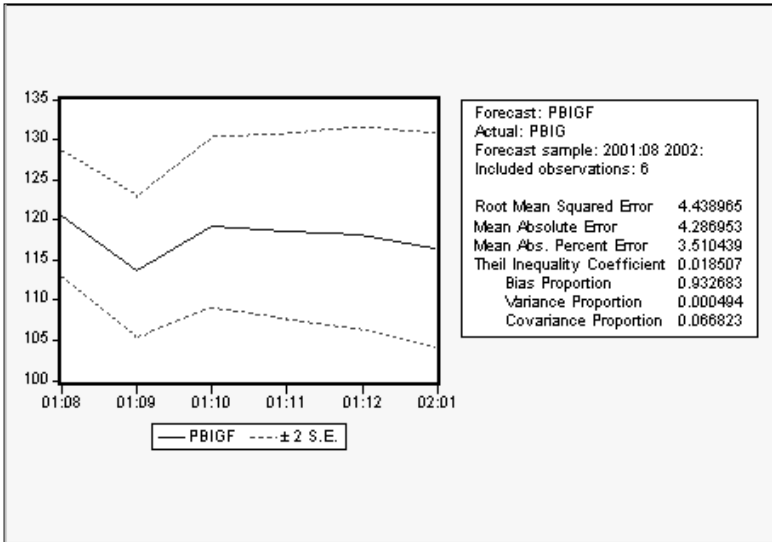


Tabla 11. Predicciones del modelo $ARIMA(0,1,1)(0,1,0)_{12}$ en logaritmos y con intervención (AO=enero-93)

Predicciones un periodo por delante: Índice PBIG			
periodo	valor real	predicción	error relativo
Ago-01	122.977439	120.475963	2.03409367
Sep-01	117.745021	113.726577	3.41283481
Oct-01	123.47929	119.152045	3.50442969
Nov-01	122.009623	118.532832	2.84960379
Dic-01	124.068379	118.205861	4.7252309
Ene-02	122.017332	116.482086	4.53644246
medias de errores(valor absoluto)			3.51043922
desviación teórica			3.2484

Figura 15. Predicciones y limites del intervalo de confianza del modelo $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos

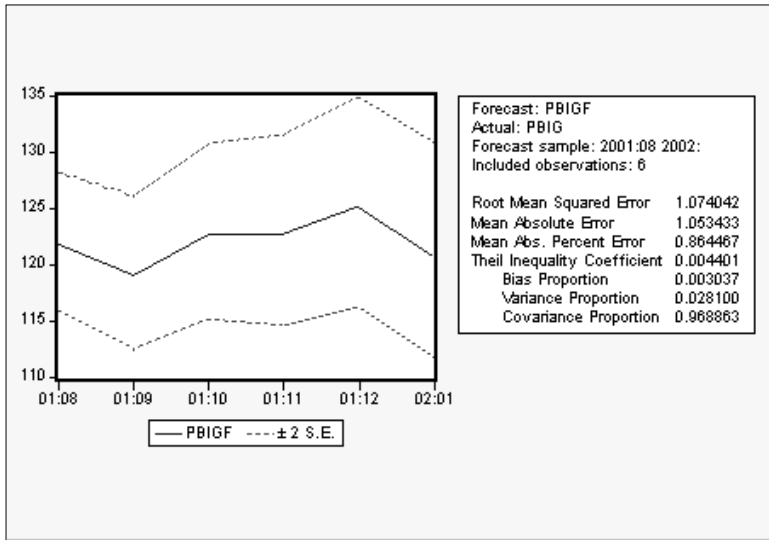


Tabla 12. Predicciones del modelo $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos

Predicciones un periodo por delante: Índice PBIG			
periodo	valor real	predicción	error relativo
Ago-01	122.977439	121.782525	0.97165361
Sep-01	117.745021	118.986922	-1.05473807
Oct-01	123.47929	122.607436	0.70607265
Nov-01	122.009623	122.706548	-0.5712051
Dic-01	124.068379	125.11229	-0.84139973
Ene-02	122.017332	120.746239	1.04173179
medias de errores(valor absoluto)			0.86446682
desviación teórica			2.5342

Figura 16. Predicciones y limites del intervalo de confianza del modelo $ARIMA(0,1,1)(1,1,0)_{12}$ en logaritmos

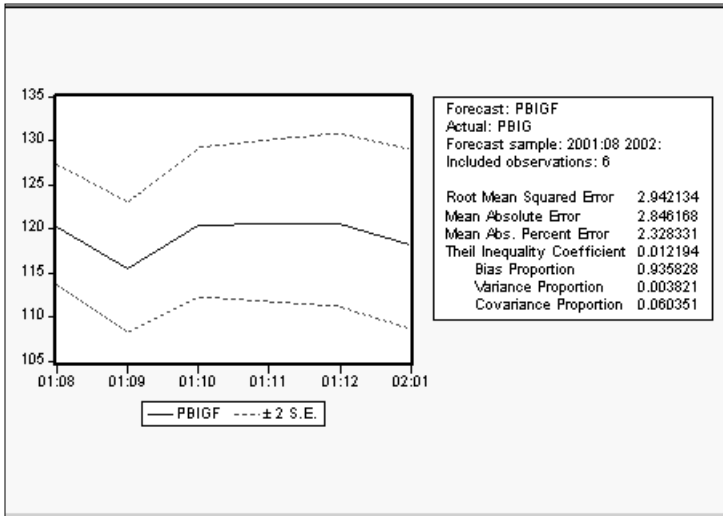


Tabla 13. Predicciones del modelo $ARIMA(0,1,1)(1,1,0)_{12}$ en logaritmos

Predicciones un periodo por delante:Indice PBIG			
periodo	valor real	predicción	error relativo
Ago-01	122.977439	120.205208	2.25425966
Sep-01	117.745021	115.389473	2.00054983
Oct-01	123.47929	120.369759	2.5182611
Nov-01	122.009623	120.467263	1.26412948
Dic-01	124.068379	120.540995	2.84309654
Ene-02	122.017332	118.247377	3.08968859
medias de errores(valor absoluto)			2.32833086
desviación teórica			2.8444

5. Control de calidad

Como el objetivo del estudio es evaluar la calidad del dato de un mes en concreto, se eligió al mes de enero del 2002, como periodo de validación. Para este análisis se realizó primeramente la predicción \hat{y}_t

con un período por delante, es decir, la predicción con origen en t-1. Esto hace que utilices toda la información disponible hasta el momento de t-1 y, por tanto, es lógico cuando se quiera valorar por ejemplo enero del 2002, el origen de la predicción será diciembre del 2001, utilizando toda la información desde 1991:01 hasta

2001:12, para la predicción utilizaremos el modelo que mejor capacidad predictiva tuvo para el periodo de validación, ese modelo es el $ARIMA(0,1,1)(0,1,1)_{12}$ en logaritmos.

Para lograr evaluar la calidad de un periodo en concreto utilizaremos la sorpresas simples, que es la variación relativa entre el dato realmente observado y la predicción para ese dato.

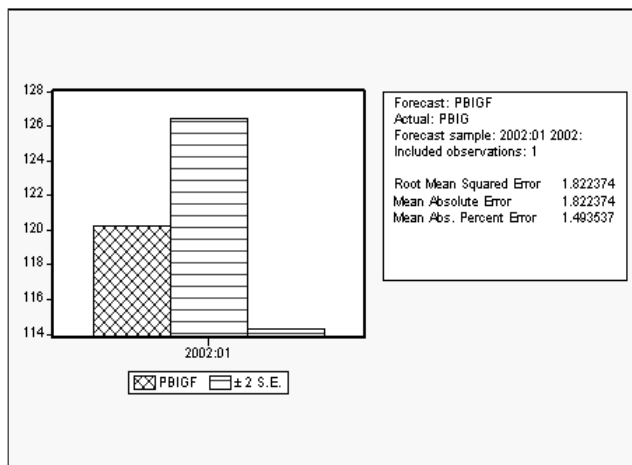
A continuación, presentamos los resultados en la Tabla 14, donde se puede

observar que el Índice del PBI Global para el periodo enero- 2002 es de buena calidad o es considerado como válido, debido a que la sorpresa esta dentro del intervalo de 1.96desviaciónesteóricas (intervalo del 95% confianza), donde se puede observar que el dato real tiene un comportamiento aleatorio y la predicción es el dato deterministico o fijo (es importante aclarar que los estadísticos desviación teórica, desviación típica residual y la raíz del error cuadrático medio son aproximadamente iguales)

Tabla 14. Evaluación de la calidad del período: enero-2002

SORPRESA EN PREDICCIONES UN PERIODO POR DELANTE:Índice PBI Global=ENERO 2002					
VALOR REAL	PREDICCION	SORPRESA	MEDIA DE ERRORES(VALOR ABSOLUTO)	DESVIACION TEORICA	INTERVALO DEL 95%
122.02	120.19	1.52	1.49	1.82	[116.62,123.77]

Figura 17. Predicción y los limites de intervalo de confianza



5.2.3 Pautas para identificar un perfil de las series macroeconómicas que no se adecuan a los ARIMA

Los modelos ARIMA son válidos para casi todas las series pero, por distintos motivos, hay algunas para las que no se pueden construir los modelos porque fallan las mínimas restricciones que exigen los ARIMA, sobretodo en el Sector Pesquero son especialmente erráticas, esto se debe a múltiples problemas, como pueden ser las técnicas de recolección de información o también puede surgir como resultado de la presencia de factores atípicos en la que se observa datos muy diferentes (la observación es muy pequeña o muy grande) con relación a las demás observaciones de la muestra, a esto se conoce como factores atípicos. Así, (las variables Enlatado y continental fresco, entre otros), presentan comportamientos que se alejan de la distribución normal de media cero y varianza constante; principal supuesto de los modelos estadísticos de Box-Jenkins (1976). En estas series temporales se observa que para diferentes intervalos de tiempo, presenta diferentes patrones de inestabilidad (volatilidad), concluyendo que no son estacionarias. Esto se debe a que la varianza de las observaciones cambia con el tiempo (presencia de heteroscedasticidad). Ante este fenómeno lo que se hace es aplicar logaritmos para estabilizar las series. Pero, es importante aclarar que la aplicación de logaritmos no va a corregir el problema de heteroscedasticidad sino que simplemente representa un método sencillo y efectivo de estabilizar a la serie. Por lo que se puede concluir que estamos

frente a procesos de Heteroscedasticidad Condicional Autorregresiva, en este tipo de series la evolución de la varianza condicional cobra una mayor importancia que el de media condicional, y que no es captada o modelada por los modelos clásicos del tipo ARIMA.

La identificación de un proceso ARCH⁴² es complementada por los modelo ARIMA en la etapa de validación de los residuos (donde los ARIMA estiman la media y los ARCH⁴³ la varianza condicional) hay una serie de test de diagnósticos de modelos disponibles en el programa Eviews 3.1 (identificación de los residuos- correlograma Q-Ljung-Box, correlograma de los residuos al cuadrado, correlación serial de los residuos y el test ARCH LM) que ayudan a identificar las características en los residuos que se asemejan a los modelos ARCH (un método que sirve para corregir la heteroscedasticidad).

Las características principales que definen los perfiles de series macroeconómicas que se complementan con los ARCH son las siguientes:

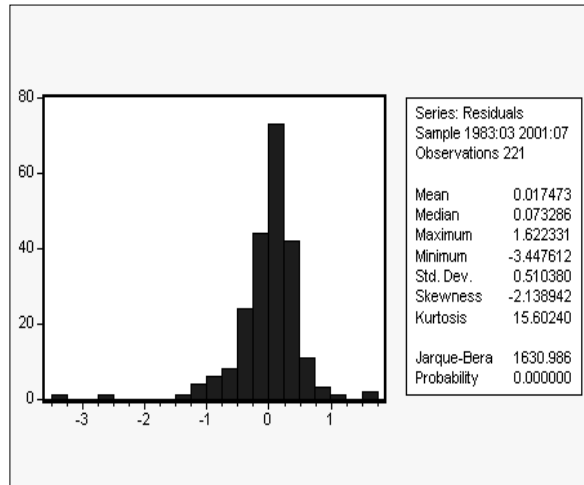
- La distribución de los residuos ε_t es leptocúrtica con colas anchas o pesadas de gran densidad, la verificación es realizada, calculando la curtosis y probando que su valor es superior a 3.

Ver figura 1. las características de la serie residual dista mucho de una distribución normal, y es leptocúrtica, esto es, el coeficiente de curtosis es mayor que 3.

42/ Véase, Pulido Antonio y López Ana María. Predicción y simulación aplicada a la economía y gestión de empresas.1999

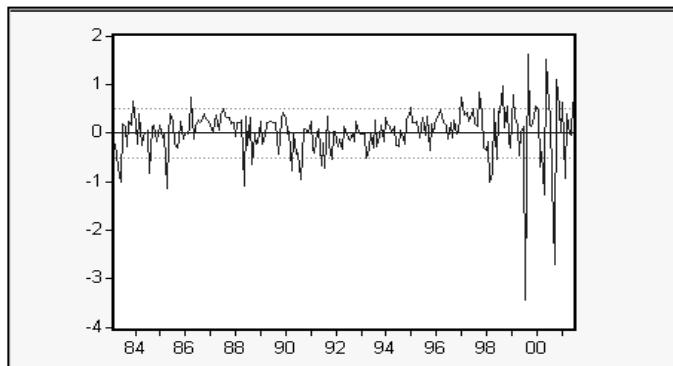
43/ Véase, Engle, Robert. ARCH.1995

Figura 1. distribución de la serie residual ε_t



- Existen agrupamientos en la volatilidad (varianza heteroscedastica). En la Figura 2, la serie de los residuos, presenta volatilidad cambiante en el tiempo (heteroscedasticidad), agrupamientos y persistencia.

Figura 2. serie de los residuos ε_t



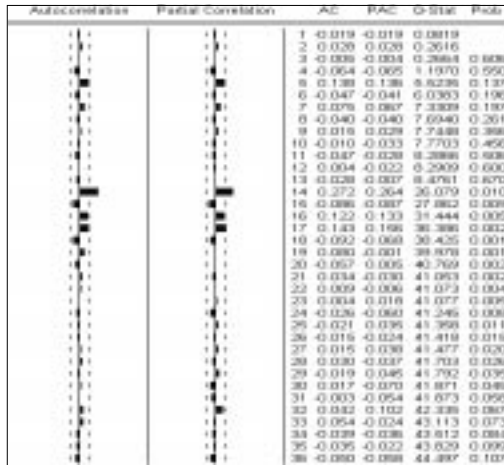
- El correlograma de los residuos ε_t es incorrelacionado. su función de autocorrelación es la de un proceso Gaussiano de Ruido Blanco, es decir que los residuos de hoy no son afectados por los residuos pasados por tratarse de la existencia de incorrelación.

Como se observa en la Figura 3. el correlograma de los residuos presenta que

Pero, un verdadero proceso Gaussiano Ruido Blanco debe ser también incorrelacionado no lineal, esto es, si tomamos al cuadrado los residuos las

autocorrelaciones no deben ser significativas. Sin embargo, ocurre lo contrario para este tipo de series (Sector Pesca) ver Figura 4.

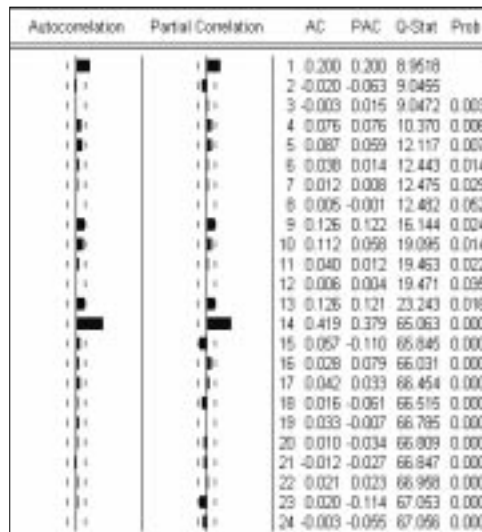
Figura 3. correlograma de los residuos ε_t



- Los residuos al cuadrado ε_t^2 , están correlacionados con los $\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots, \varepsilon_{t-q}^2$. En la Figura 4. notamos que la serie

residual al cuadrado considera un valor significativo en los primeros retardos lo cual verifica que hay correlación y la serie no seguirá un proceso Gaussiano de Ruido Blanco.

Figura 4. Autocorrelación de los residuos al cuadrado ε_t^2



- La existencia de heteroscedasticidad es confirmada por la prueba de verosimilitud en el siguiente cuadro 1, donde se verifica la existencia del problema de heteroscedasticidad (osea se rechaza la hipótesis de homocedasticidad) Las autocorrelaciones simple y parcial de los residuos al cuadrado indican una estructura ARCH.

Cuadro 1. Test para detectar la heteroscedasticidad

ARCH Test:				
F-statistic	9.077524	Probability	0.002894	
Obs*R-squared	8.794596	Probability	0.003021	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 07/06/02 Time: 19:06				
Sample(adjusted): 1983:04 2001:07				
Included observations: 220 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.208795	0.067467	3.094756	0.0022
RESID^2(-1)	0.199935	0.066360	3.012893	0.0029
R-squared	0.039975	Mean dependent var	0.260536	
Adjusted R-squared	0.035572	S.D. dependent var	0.985425	
S.E. of regression	0.967740	Akaike info criterion	2.781342	
Sum squared resid	204.1614	Schwarz criterion	2.812193	
Log likelihood	-303.9476	F-statistic	9.077524	
Durbin-Watson stat	1.974625	Prob(F-statistic)	0.002894	

Bajo estas características, lo que se hace es modelar la varianza condicional con el enfoque ARCH (Heterocedasticidad Condicional AutoRegresiva).

El tratamiento de series con estas características con alta heteroscedasticidad será tratado posteriormente en otro estudio que muy bien se podrían

complementar a la metodología de los modelos ARIMA con análisis de intervención, para el control de calidad planteadas en este estudio.

Sin embargo, la presencia de heteroscedasticidad nunca ha sido razón para descartar un buen modelo. Pero, esto no significa que ella deba ser ignorada.

VI. CONCLUSIONES Y RECOMENDACIONES

6.1 CONCLUSIONES

- La idea de detectar errores en la producción de datos por medio de los errores de predicción en series temporales, nos lleva al análisis exhaustivo en el desarrollo de una metodología para el control estadístico de calidad, al llegar una nueva observación se determina si pertenece al intervalo de confianza y, en caso contrario, se propone para su posterior revisión; en este estudio se han utilizado de manera óptima las sorpresas (Revilla y Rey, 1991), como herramienta de depuración.

En líneas generales esta investigación contiene dos secciones fundamentales uno que da a conocer de manera objetiva la metodología del control de calidad basada en los modelos ARIMA con análisis de intervención, que ha permitido que se logre resolver el problema de la falta de usos de métodos estadísticos y técnicas para evaluar la calidad de un dato en concreto y el otro es el uso de dos programas, uno por un lado de uso en modo automático que permitió el logro de modelizar un conjunto de series a la vez y el otro programa el EVIEWS 3.1 que permitió la modelización paso a paso hasta llegar al objetivo fundamental de medir la calidad de la información reportada o al dato de un mes en concreto, esta investigación ha sido desarrollada muy detalladamente para que le permita al lector llevar a

cabo el análisis de cualquier variable que tenga los requisitos mínimos exigidos por este método que es el de la estacionariedad en media y en varianza y que sus residuos al cuadrado en los primeros rezagos del correlograma no sean significativos, en concreto se ha logrado obtener una metodología que permite describir paso a paso la herramienta estadística para el control estadístico de calidad para los lectores de esta investigación.

- Con respecto al análisis del Índice del PBI Global bajo la modelización automática y el uso del EvIEWS 3.1 no es con la finalidad de comparar los resultados de uno u otro lado sino llevar a cabo una metodología iterativa de modelización y hacer uso de estos modelos para construir herramientas de control de calidad a un dato agregado en concreto, lográndose él propósito en esta investigación. Uno con un gran número de series a la vez y el otro siguiendo la estrategia del experto con una sola variable.
- La sustitución del procedimiento tradicional por las sorpresas implica una diferencia metodológica importante. Mientras que la detección tradicional considera 2 observaciones de la serie o sea la tasa de variación, las sorpresas utilizan de forma eficiente la información completa de la serie que se esta analizando.

- Finalmente, es necesario aclarar que el uso y dominio de la modelización de modo automático, proporciona al usuario una facilidad y confiabilidad para evaluar la calidad de cualquier otro dato que crea conveniente. En este caso se eligió al mes de enero del 2002, sin embargo podría hacerse para el mes que uno sospeche que probablemente sea outlier o atípico.

6.2 RECOMENDACIONES

- La idea viene del hecho que nosotros necesitamos tanta información como sea posible sobre el fenómeno que nosotros intentamos medir. Por otro lado, los subconjuntos diferentes de datos muestran a menudo una variabilidad muy diferente. Por ejemplo, al revisar la producción mensual del sector pesca, nosotros podemos encontrar el extremo de tener datos pequeños (incluso ceros) de la producción valorada para estos periodos de veda en el Perú. Claro, está que estos datos no deben ser considerados como outlier (es decir como datos sospechosos) si nosotros no tenemos la información sobre este problema estacional no sería necesario corregirlo, sino aceptarlo como un dato que si realmente ocurrió. Pero un problema adicional es que, este periodo estacional ocasionado por algún tipo de fenómenos como el que citamos, es muy diferente de una rama a otra, igual hay ramas en el que la producción disminuye pero fuertemente crece en

esa temporada de problemas climáticos, por esta razón, es importante adquirir la información sobre las características dinámicas diferentes de cada uno de las ramas y mejorar reglas de corrección y estrategias.

- Por otro lado el hecho mismo que las características propias del sector pesca es especialmente erráticas será necesario complementar a la metodología de los modelos ARIMA-AI, con métodos que corrijan la heteroscedasticidad que probablemente esta siendo causada por los datos atípicos que son ocasionadas más que todo por el fenómeno del niño y por la recogida de información. De acuerdo al perfil de la serie macroeconómica del Sector Pesca, que se estudió en la sección (V-5.2.3), el método para corregir la heteroscedasticidad es en efecto los ARCH (modelos autorregresivos con heteroscedasticidad condicional), método que debe ser puesta en marcha para modelizar la varianza condicional con la finalidad de obtener intervalos de predicción más precisos y estimadores más eficientes.
- Por último se recomienda la implantación de la metodología de los modelos ARIMA-AI, para el control de calidad, iniciándose por el INEI, para luego dar a conocer a los sectores y a las departamentales los conocimientos adquiridos durante la investigación.

VII. REFERENCIAS BIBLIOGRAFICAS

- "METODOLOGÍA".
Méndez Álvarez, Carlos.
Bogotá.edit. McGraw - Hill.1995
- "INFORME FINAL DE LAS ACTIVIDADES DE CONSULTORÍA PARA EL MEJORAMIENTO DE LAS ESTADÍSTICAS BÁSICAS".
Revilla, Pedro.
INEI-CIDE-2001
- "SELECTIVE EDITING METHODS BASED EN TIME SERIES MODELLING. ROMA. 1999 Y ANALYSIS AND QUALITY CONTROL FROM ARIMA MODELLING".
Revilla, Pedro y Rey, Pilar.
Europe. 2000
- "ÍNDICE DE DISPONIBILIDAD DE BIENES DE EQUIPO: MODELIZACIÓN ARIMA - ANÁLISIS DE INTERVENCIÓN COMPONENTES SUBYACENTES Y PATRÓN CÍCLICO".
Quilis M, Enrique.
INE.1998.
- "UN NUEVO MÉTODO PARA EL CONTROL DE CALIDAD DE LOS DATOS EN SERIES DE DATOS EN SERIES TEMPORALES".
Maravall, Agustín y Luna, Cristina.
Banco de España. 1998.
- "MODELOS ARIMA-AI PARA LAS SERIES AGREGADAS MONETARIAS (SALDOS MEDIOS MENSUALES) M3 Y M2 (1990)".
Espasa, Antoni.
- Servicios de estudio del Banco de España. (7901)
- "INFORME FINAL DE LAS ACTIVIDADES DE CONSULTORÍA PARA LA IMPLANTACIÓN DE LOS PROCESOS Y SOFTWARE EN LOS PROCEDIMIENTOS ESTADÍSTICOS DE CONTROL DE CALIDAD EN LA PRODUCCIÓN DE LA ESTADÍSTICA BÁSICA".
Revilla, Pedro y Rey, Pilar (dic.-2001).
INEI-CIDE.
- "ECONOMETRÍA BÁSICA" (Tercera edición).
Gujarati, Damodar N. (1997).
Mc Hill INTERAMERICANO S.A
- "MODELOS ECONÓMICOS DE PREDICCIÓN MACROECONÓMICOS EN LA ARGENTINA".
McCandless, Gabrieli, Florencia y Murphy, Tomas (1999).
Banco Central de la República Argentina.
- "PREDICCIÓN Y SIMULACIÓN APLICADA A LA ECONOMÍA Y GESTIÓN DE EMPRESAS".
Pulido, Antonio y López, Ana María (1999).
Edición pirámide.
- "ECONOMETRÍA MODERNA" (Cáp. 15 al 18). **Casas Tragodara, Carlos** (sep. 2001).
Universidad del Pacífico.

- "REVISIÓN DE LOS PROCEDIMIENTOS DE ANÁLISIS DE LA ESTACIONARIEDAD DE LAS SERIES TEMPORALES".
Mahía, Ramón (1999).
- "MÉTODOS DE PREDICCIÓN EN ECONOMÍA II".
Aznar, Antonio (1993)

Anexos

1. Modelización Automática y control de calidad:
Índices del PBI
2. Modelización Automática y control de calidad:
volumen de producción Minera

ANEXO 1

MODELIZACIÓN AUTOMÁTICA Y CONTROL DE CALIDAD: ÍNDICES DEL PBI

1.1 Parte 1: Salidas del programa MODAUT

**Tabla 1. Modelo ARIMA-AI Índice del PBI agropecuario
(1991:01 2002.01)**

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1002=AGRO	12=mensual	0=Ln	0.0263938	(0,1,1)(0,1,1) ₁₂	0=sin cte	10
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-		-0.267651039		-		0.241701865
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS:COMPONENTE DETERMINISTICA		
2=AUTOMATICA	89(may-98)	3=cambios temporales(TC)		-0.174196063		
2=AUTOMATICA	91(jul-98)	3=cambios temporales(TC)		-0.091525106		
2=AUTOMATICA	66(jun-96)	1=impulsos(AO)		0.065434967		
2=AUTOMATICA	68(ago-96)	1=impulsos(AO)		0.075112598		
2=AUTOMATICA	43(jul-94)	1=impulsos(AO)		-0.071698147		
2=AUTOMATICA	25(Ene-93)	1=impulsos(AO)		-0.071199076		
2=AUTOMATICA	73(ene-97)	3=cambios temporales(TC)		0.079139927		
2=AUTOMATICA	108(dic-99)	1=impulsos(AO)		-0.052484572		
2=AUTOMATICA	54(jun-95)	2=escalones(LS)		-0.07436194		
2=AUTOMATICA	21(set-92)	2=escalones(LS)		0.064827203		

**Tabla 2. Modelo ARIMA-AI del Índice PBI Pesca
(1991:01 2002.01)**

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1003=PESCA	12=mensual	0=Ln	0.207914	(0,0,3)(0,1,1) ₁₂	0=sin cte	4
PARAMETROS: COMPONENTE ESTOCASTICA						
MA(1) REGULAR		MA(2) REGULAR		MA(3) REGULAR		MA ESTACIONAL
-0.289355623		0.110207755		0.192341866		0.559206811
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS:COMPONENTE DETERMINISTICA		
2=AUTOMATICA	84(dic-97)	3=cambios temporales(TC)		-0.943975885		
2=AUTOMATICA	17(may-92)	2=escalones(LS)		0.307242896		
2=AUTOMATICA	111(mar-200)	1=impulsos(AO)		0.831169527		
2=AUTOMATICA	39(mar-94)	3=cambios temporales(TC)		0.556409309		

**Tabla 3. Modelo ARIMA - AI Índice PBI Minería
(1991:01 2002.01)**

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1004=Minería	12=mensual	0=Ln	0.0268355	(0,1,1)(0,1,1) ₁₂	0=sin cte	0
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-		0.368769206		-		0.999727533
VARIABLE OUTLIER	PERIODO DE INTERVENCION		TIPO DE INTERVENCION		PARAMETROS:COMPONENTE DETERMINISTICA	

**Tabla 4. Modelo ARIMA - AI Índice PBI Industrial
(1991:01 2002.01)**

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1005=Industria	12=mensual	0=Ln	0.0423283	(0,1,1)(1,0,0) ₁₂	0=sin cte	2
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-		0.522161129		0.483669777		-
VARIABLE OUTLIER	PERIODO DE INTERVENCION		TIPO DE INTERVENCION		PARAMETROS:COMPONENTE DETERMINISTICA	
2=AUTOMATICO	7(jul-91)		1=impulso(AO)		0.15778869	
2=AUTOMATICO	76(abr-97)		3=cambios temporales(TC)		0.113419963	

**Tabla 5. Modelo ARIMA - AI Índice del PBI Electricidad
(1991:01 2002.01)**

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1006=Electricidad	12=mensual	0=Ln	0.0151858	(0,1,1)(0,1,0) ₁₂	0=sin cte	11
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-		0.232647149		-		-
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS:COMPONENTE DETERMINISTICA		
2=AUTOMATICO	17(may-92)	2=escalones(LS)		-0.170302759		
2=AUTOMATICO	26(feb-93)	2=escalones(LS)		0.119131697		
2=AUTOMATICO	14(feb-92)	1=impulsos(AO)		-0.063088049		
2=AUTOMATICO	79(jul-97)	2=escalones(LS)		0.079827933		
2=AUTOMATICO	49(ene-95)	2=escalones(LS)		-0.081933979		
2=AUTOMATICO	20(ago-92)	3=cambios temporales(TC)		-0.047165235		
2=AUTOMATICO	110(feb-00)	1=impulsos(AO)		0.037829655		
2=AUTOMATICO	37(ene-94)	1=impulsos(AO)		-0.053721059		
2=AUTOMATICO	33(set-93)	1=impulsos(AO)		-0.03302078		
2=AUTOMATICO	62(feb-96)	1=impulsos(AO)		0.038903267		
2=AUTOMATICO	54(jun-95)	1=impulsos(AO)		-0.031699154		

**Tabla 6. Modelo ARIMA - AI Índice PBI Construcción
(1991:01 2002.01)**

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1007=Construcción	12=mensual	0=Ln	0.0634651	(0,1,1)(1,0,0) ₁₂	0=sin cte	0
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-		0.471104415		0.321247694		-
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS:COMPONENTE DETERMINISTICA		
-	-	-		-		

**Tabla 7. Modelo ARIMA - AI Índice PBI Comercio
(1991:01 2002.01)**

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1008=Comercio	12=mensual	0=Ln	0.0377568	(0,1,1)(1,0,0) ₁₂	0= sin cte.	3
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
		0.588120428		0.586770754		
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS:COMPONENTE DETERMINISTICA		
2=AUTOMATICO	25(ene-93)	3=cambios temporales(CT)		-0.15072364		
2=AUTOMATICO	73(ene-97)	1=impulsos(AO)		0.099938302		
2=AUTOMATICO	16(abr-92)	2=escalones(LS)		-0.095475249		

**Tabla 8. Modelo ARIMA - AI Índice PBI Otros servicios
(1991:01 2002.01)**

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1009=OTROS SER.	12=MENSUAL	0=Ln	0.0377568	(0,1,1)(0,1,1) ₁₂	0=sin cte.	3
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-		0.588120428		-		0.586770754
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS:COMPONENTE DETERMINISTICA		
2=AUTOMATICO	25(ene-93)	3=cambios temporales(CT)		-0.15072364		
2=AUTOMATICO	73(ene-97)	1=impulsos(AO)		0.099938302		
2=AUTOMATICO	16(abr-92)	2=escalones(LS)		-0.095475249		

1.2 Parte 2: Salidas del programa Predis

PREDICCIONES UN PERIODO POR DELANTE:1003=PESCA			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	42.2	46.9	-11.1
Sep-01	43.3	50.1	-15.8
Oct-01	79.2	76.2	3.9
Nov-01	61.3	118.9	-94.2
Dic-01	69	93	-34.8
Ene-02	73.1	74.7	-2.2
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			27
DESVIACION TEORICA			20.8

PREDICCIONES UN PERIODO POR DELANTE:1004=MINERIA			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	179.9	161	10.5
Sep-01	175.2	167.5	4.4
Oct-01	179.2	180.6	-0.8
Nov-01	180.7	175.1	3.1
Dic-01	184	182.3	0.9
Ene-02	178.1	180.9	-1.6
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			3.6
DESVIACION TEORICA			2.7

PREDICCIONES UN PERIODO POR DELANTE:1005=INDUSTRIA			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	111.9	114.2	-2
Sep-01	111.1	110.1	0.8
Oct-01	119.7	114.5	4.4
Nov-01	114.9	118.8	-3.4
Dic-01	110.4	113.4	-2.7
Ene-02	113.4	111.3	1.9
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			2.5
DESVIACION TEORICA			4.2

PREDICCIONES UN PERIODO POR DELANTE:1007=Contrucción			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	111.3	105.2	5.4
Sep-01	108.7	107.3	1.3
Oct-01	118.3	108.3	8.5
Nov-01	120.8	112.2	7.1
Dic-01	117.8	114.2	3
Ene-02	120.3	116.5	3.2
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			4.8
DESVIACION TEORICA			6.3

PREDICCIONES UN PERIODO POR DELANTE:1008=Comercio			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	117.4	114.2	2.7
Sep-01	110.4	111.3	-0.9
Oct-01	116.3	114.4	1.6
Nov-01	119.2	118.7	0.5
Dic-01	117.6	118.6	-0.9
Ene-02	113.5	113.7	-0.2
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			1.1
DESVIACION TEORICA			3.8

PREDICCIONES UN PERIODO POR DELANTE:1009=Otros serv.			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	117.4	114.2	2.7
Sep-01	110.4	111.3	-0.9
Oct-01	116.3	114.4	1.6
Nov-01	119.2	118.7	0.5
Dic-01	117.6	118.6	-0.9
Ene-02	113.5	113.7	-0.2
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			1.1
DESVIACION TEORICA			3.8

ANEXO 2

MODELIZACIÓN AUTOMÁTICA Y CONTROL DE CALIDAD: VOLUMEN DE PRODUCCIÓN MINERA

2.1 Parte 1: Salidas del programa MODAUT

Modelo ARIMA - AI Volumen de producción Minera
(1985:01 2002:01= Cu, Zn, Ag, Pb y Petróleo) y (1989:01 2002:01=Au, Sn, Mo y GAS Natural)

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1018=cobre	12=mensual	0=Ln	0.0574924	(0,1,1)(1,0,0) ₁₂	0=sin cte.	10
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL	MA ESTACIONAL	
		0.348022782		0.56091071		
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS: COMPONENTE DETERMINISTICA		
2=AUTOMATICA	47	1		-1.773090212		
2=AUTOMATICA	64	1		-1.079405684		
2=AUTOMATICA	64	1		0.49301414		
2=AUTOMATICA	48	1		-0.536979394		
2=AUTOMATICA	63	1		-0.417754692		
2=AUTOMATICA	46	1		-0.446895834		
2=AUTOMATICA	14	1		-0.222585312		
2=AUTOMATICA	35	1		-0.217988717		
2=AUTOMATICA	44	1		-0.17492166		
2=AUTOMATICA	194	1		-0.169450815		

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1020=plata	12=mensual	0=Ln	0.0441977	(2,1,0)(1,0,0) ₁₂	0=sin cte.	10
PARAMETROS: COMPONENTE ESTOCASTICA						
AR(1) REGULAR		AR(2) REGULAR		AR ESTACIONAL	MA ESTACIONAL	
-0.371077873		-0.273453009		0.351938917	-	
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS: COMPONENTE DETERMINISTICA		
2=AUTOMATICA	47	1		-1.116942488		
2=AUTOMATICA	49	2		0.436282098		
2=AUTOMATICA	45	1		0.430598654		
2=AUTOMATICA	44	2		-0.356343235		
2=AUTOMATICA	56	1		-0.189685518		
2=AUTOMATICA	15	3		-0.190735248		
2=AUTOMATICA	85	2		-0.175002899		
2=AUTOMATICA	94	3		0.138788193		
2=AUTOMATICA	50	1		-0.112696335		
2=AUTOMATICA	17	2		0.131492964		

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1021=plomo	12=mensual	0=Ln	0.0454462	(1,1,0)(1,0,0) ₁₂	0=sin cte.	9
PARAMETROS: COMPONENTE ESTOCÁSTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-0.460535311				0.345295636		
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS: COMPONENTE DETERMINISTICA		
2=AUTOMÁTICA	47	1		-1.935149611		
2=AUTOMÁTICA	44	1		-0.715572037		
2=AUTOMÁTICA	15	3		-0.436202254		
2=AUTOMÁTICA	48	1		-0.421698203		
2=AUTOMÁTICA	46	3		-0.447474574		
2=AUTOMÁTICA	43	1		-0.270339218		
2=AUTOMÁTICA	37	1		-0.172212776		
2=AUTOMÁTICA	55	3		-0.204610152		
2=AUTOMÁTICA	17	3		0.19612156		

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1023=oro	12=mensual	0=Ln	0.0455751	(0,1,0)(1,1,0) ₁₂	0=sin cte.	14
PARAMETROS: COMPONENTE ESTOCÁSTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
-		-		0.227730271		-
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS: COMPONENTE DETERMINISTICA		
2=AUTOMÁTICA	60	1		-0.169135747		
2=AUTOMÁTICA	25	2		-0.438834992		
2=AUTOMÁTICA	126	3		0.204032137		
2=AUTOMÁTICA	71	1		-0.117823429		
2=AUTOMÁTICA	119	1		-0.119837799		
2=AUTOMÁTICA	56	2		0.163613376		
2=AUTOMÁTICA	145	3		-0.201882632		
2=AUTOMÁTICA	109	2		-0.151864835		
2=AUTOMÁTICA	150	1		0.150061635		
2=AUTOMÁTICA	37	2		-0.298199711		
2=AUTOMÁTICA	49	3		-0.178234224		
2=AUTOMÁTICA	79	1		-0.083928594		
2=AUTOMÁTICA	122	1		-0.073324133		
2=AUTOMÁTICA	59	2		0.117833535		

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1022=petróleo	12=trimestral	0=Ln	0.0330916	(0,1,1)(1,0,0) ₁₂	0=sin cte.	6
PARAMETROS: COMPONENTE ESTOCASTICA						
AR REGULAR		MA REGULAR		AR ESTACIONAL		MA ESTACIONAL
		0.721386052		0.241468557		
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS: COMPONENTE DETERMINISTICA		
2=AUTOMATICA	60	3		-0.298704201		
2=AUTOMATICA	53	1		-0.157823429		
2=AUTOMATICA	52	3		-0.114884377		
2=AUTOMATICA	198	3		-0.147581509		
2=AUTOMATICA	34	2		-0.106138452		
2=AUTOMATICA	55	2		0.10334337		

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO ARIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1024=estaño	12=trimestral	0=Ln	0.0364561	(0,1,3)(0,1,1) ₁₂	0=sin cte.	18
PARAMETROS: COMPONENTE ESTOCASTICA						
MA(1) REGULAR		MA(2) REGULAR		MA(3) REGULAR		MA ESTACIONAL
-0.166172453		-0.559147125		0.259155581		-0.665905988
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION		PARAMETROS: COMPONENTE DETERMINISTICA		
2=AUTOMATICA	77	3		-0.320444581		
2=AUTOMATICA	119	3		-0.186937623		
2=AUTOMATICA	87	1		0.335787958		
2=AUTOMATICA	84	1		0.155476904		
2=AUTOMATICA	78	1		-0.146240552		
2=AUTOMATICA	91	1		0.167109607		
2=AUTOMATICA	82	1		0.139978111		
2=AUTOMATICA	148	1		0.117219266		
2=AUTOMATICA	121	2		0.156053821		
2=AUTOMATICA	126	3		-0.104052484		
2=AUTOMATICA	115	1		0.092513353		
2=AUTOMATICA	132	3		-0.148350664		
2=AUTOMATICA	25	2		0.234331654		
2=AUTOMATICA	37	2		0.187838503		
2=AUTOMATICA	111	3		0.082037508		
2=AUTOMATICA	151	1		0.207332968		
2=AUTOMATICA	64	3		-0.060229368		
2=AUTOMATICA	68	3		0.058743041		

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA	MODELO A RIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1025=molibdeno	12=mensual	0=Ln	0.1080335	(3,1,0)(0,1,0) ₁₂	0=sin cte.	15
PARAMETROS: COMPONENTE ESTOCASTICA						
AR(1) REGULAR		AR(2) REGULAR		AR(3) REGULAR		MA ESTACIONAL
-0.387153195		-0.261737858		-0.589797173		-
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION	PARAMETROS:COMPONENTE DETERMINISTICA			
2=AUTOMATICA	86	1	-1.002908658			
2=AUTOMATICA	37	2	0.829709106			
2=AUTOMATICA	120	1	-0.459974338			
2=AUTOMATICA	94	3	0.736843961			
2=AUTOMATICA	92	1	0.59110881			
2=AUTOMATICA	107	3	0.524749506			
2=AUTOMATICA	91	3	-0.439033679			
2=AUTOMATICA	73	2	-0.356630886			
2=AUTOMATICA	125	3	0.428964926			
2=AUTOMATICA	117	2	-0.425960998			
2=AUTOMATICA	146	2	-0.762041474			
2=AUTOMATICA	102	2	-0.279525285			
2=AUTOMATICA	75	1	-0.204608986			
2=AUTOMATICA	134	3	-0.396223504			
2=AUTOMATICA	143	3	-0.399658441			

SERIE	PERIODO	TRANSFORMACION	DESVIACION TIPICA RESIDUAL	MODELO A RIMA	INCLUYE CONSTANTE	# DE INTERVENCIONES
1026=gas natural	12=mensual	0=Ln	0.0579484	(2,1,0)(1,0,0) ₁₂	0=sin cte.	17
PARAMETROS: COMPONENTE ESTOCASTICA						
AR(1) REGULAR		AR(2) REGULAR		AR ESTACIONAL		MA ESTACIONAL
0.506991782		0.165902792		0.852092172		-
VARIABLE OUTLIER	PERIODO DE INTERVENCION	TIPO DE INTERVENCION	PARAMETROS:COMPONENTE DETERMINISTICA			
2=AUTOMATICA	133	1	-0.370631971			
2=AUTOMATICA	131	1	0.428914704			
2=AUTOMATICA	144	2	-0.605518016			
2=AUTOMATICA	61	2	-0.476840957			
2=AUTOMATICA	147	1	0.399087636			
2=AUTOMATICA	109	1	-0.25193172			
2=AUTOMATICA	123	1	-0.255244502			
2=AUTOMATICA	127	1	-0.251797777			
2=AUTOMATICA	12	3	-0.355872767			
2=AUTOMATICA	112	1	-0.169064531			
2=AUTOMATICA	84	2	0.23849548			
2=AUTOMATICA	136	1	0.137945481			
2=AUTOMATICA	91	3	-0.143960026			
2=AUTOMATICA	122	2	-0.21901983			
2=AUTOMATICA	5	1	-0.134944296			
2=AUTOMATICA	119	1	-0.145103166			
2=AUTOMATICA	117	1	0.074298649			

2.2 PARTE 2: Salidas del programa PREDIS

PREDICCIONES UN PERIODO POR DELANTE:1018=cobre			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	70.2	55.4	21.1
Sep-01	78.7	61.5	21.8
Oct-01	83	68.6	17.4
Nov-01	63.5	74.3	-16.9
Dic-01	93.3	69.9	25.1
Ene-02	72	78.2	-8.5
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			18.5
DESVIACION TEORICA			5.7

PREDICCIONES UN PERIODO POR DELANTE:1019=zinc			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	86.9	84.3	3
Sep-01	87	86.3	0.7
Oct-01	90	87.8	2.4
Nov-01	125.3	86.5	30.9
Dic-01	104.5	106.5	-1.9
Ene-02	105.9	105	0.8
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			6.6
DESVIACION TEORICA			3.5

PREDICCIONES UN PERIODO POR DELANTE:1020=plata			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	241.8	234.5	3
Sep-01	239.8	235.6	1.7
Oct-01	246.2	242.8	1.4
Nov-01	221.4	240.8	-8.7
Dic-01	229.2	229.5	-0.1
Ene-02	220.8	232.3	-5.2
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			3.4
DESVIACION TEORICA			4.4

PREDICCIONES UN PERIODO POR DELANTE:1021=plomo			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	25.4	24.3	4.1
Sep-01	25	24.9	0.5
Oct-01	26.4	25.5	3.1
Nov-01	25.3	25.8	-2
Dic-01	25	26.1	-4.5
Ene-02	24.7	24.8	-0.4
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			2.4
DESVIACION TEORICA			4.5

PREDICCIONES UN PERIODO POR DELANTE:1022=petróleo			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	3033.8	2943.6	3
Sep-01	2902.4	2873.7	1
Oct-01	3035	2961.7	2.4
Nov-01	2851.2	2960	-3.8
Dic-01	3140	2941.9	6.3
Ene-02	3111.8	3060.5	1.6
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			3
DESVIACION TEORICA			3.3

PREDICCIONES UN PERIODO POR DELANTE: Au			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	13.6	10.6	21.9
Sep-01	11.5	13.4	-16.3
Oct-01	11.1	13.5	-21.9
Nov-01	12.7	10.4	18
Dic-01	10.2	11.1	-8.9
Ene-02	12.3	9.7	21.5
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			18.1
DESVIACION TEORICA			4.6

PREDICCIONES UN PERIODO POR DELANTE: Sn			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	3.1	2.5	19.8
Sep-01	3.2	4.6	-43.4
Oct-01	3.6	2.8	22.7
Nov-01	3.2	2.3	30.4
Dic-01	3.2	5.3	-66.8
Ene-02	3.1	3.9	-24.5
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			34.6
DESVIACION TEORICA			3.6

PREDICCIONES UN PERIODO POR DELANTE: mb			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	0.8	0.7	15.3
Sep-01	0.9	0.8	10
Oct-01	1	0.9	11.4
Nov-01	0.8	0.8	2.6
Dic-01	1.6	0.9	42.1
Ene-02	0.3	1.5	-363.4
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			74.1
DESVIACION TEORICA			10.8

PREDICCIONES UN PERIODO POR DELANTE:Gas			
PERIODO	VALOR REAL	PREDICCION	ERROR RELATIVO
Ago-01	2114.4	2481.6	-17.4
Sep-01	1946.8	2100.5	-7.9
Oct-01	1622.1	1988.8	-22.6
Nov-01	937.1	1280.9	-36.7
Dic-01	695.9	417.7	40
Ene-02	899.3	646.9	28.1
MEDIA DE ERRORES(EN VALOR ABSOLUTO)			25.4
DESVIACION TEORICA			5.8

2.3 Parte 3: Salidas del programa SORPRESA

SORPRESAS EN PREDICCIONES UN PERIODO POR DELANTE:ENERO-2002					
SERIE	VAOR REAL	PREDICCION	SORPRESA	DESVIACION TEORICA	INTERVALO DEL 95%
1018=cobre	72	77.8	-7.34	7.3	[67.4; 89.7]
1019=zinc	105.9	105.1	0.72	4.35	[96.5; 114.5]
1020=plata	220.8	232.3	-4.97	4.41	[213.1; 253.3]
1021=plomo	24.7	24.8	-0.39	4.52	[22.7; 27.1]
1022=petróleo	3111.8	3084.1	0.9	3.31	[2890.3; 3290.8]

SORPRESAS EN PREDICCIONES UN PERIODO POR DELANTE:ENERO-2002					
SERIE	VAOR REAL	PREDICCION	SORPRESA	DESVIACION TEORICA	INTERVALO DEL 95%
1023: Au	12.3	12.4	-0.92	5.66	11.1 ; 13.9
1024:Sn	3.1	3.2	-2.48	3.81	3.0 ; 3.4
1025:Mb	0.3	1.5	-77.99	11.98	1.2 ; 1.9
1026:Gas	899.3	498.7	80.34	6.6	438.2 ;567.5