

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE CIENCIAS MATEMÁTICAS

E.A.P. DE ESTADÍSTICA

**Análisis Factorial Dinámico para la proyección de las
Exportaciones en el Perú**

TESIS

Para optar el Título Profesional de Licenciado en Estadística

AUTOR

Christian Fernando Flores Camacho

ASESOR

Emma Cambillo Moyano

Lima – Perú

2013

**ANÁLISIS FACTORIAL DINÁMICO PARA LA PROYECCIÓN DE LAS
EXPORTACIONES EN EL PERÚ**

CHRISTIAN FERNANDO FLORES CAMACHO

Tesis presentada a consideración del Cuerpo Docente de la Facultad de Ciencias Matemáticas, de la Universidad Nacional Mayor de San Marcos, como parte de los requisitos para obtener el Título Profesional de Licenciado en Estadística.

Aprobada por:

Mg. María Estela Ponce Aruneri
Presidente

Lic. Rosa Fátima Medina Merino
Miembro Jurado

Mg. Emma Norma Cambillo Moyano
Miembro Asesor

Lima – Perú
Octubre 2013

FICHA CATALOGRÁFICA

FLORES CAMACHO, CHRISTIAN FERNANDO

Análisis Factorial Dinámico para la Proyección de las Exportaciones en el Perú, (Lima) 2013.

Xi, 82p., 29,7 cm. (UNMSM, Licenciado en Estadística, 2013)

Tesis, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas. Estadística.

- i. UNMSM / F. de C.M.
- ii. Título (serie).

Bachiller Christian Fernando Flores Camacho

Análisis Factorial Dinámico para la Proyección de las Exportaciones en el Perú.

12 de octubre del 2013

La presente tesis es dedicada a Dios, a mi esposa y a mis padres, por su constante apoyo y dedicación; y, porque estuvieron siempre presentes en todo momento, brindándome su confianza en cada peldaño a subir en mi desarrollo profesional.

“Hay una fuerza motriz más poderosa que el vapor, la electricidad y la energía atómica: la voluntad.”; “La imaginación es más importante que el conocimiento.”

Albert Einstein

Agradezco mucho a mis profesores que a lo largo de mi carrera universitaria me enseñaron sus grandes saberes. En especial a mi asesora, profesora y amiga Emma Cambillo Moyano por su valioso y tenaz apoyo en el desarrollo de esta Tesis.

R E S U M E N

ANÁLISIS FACTORIAL DINÁMICO PARA LA PROYECCIÓN DE LAS EXPORTACIONES EN EL PERÚ

CHRISTIAN FERNANDO FLORES CAMACHO

Octubre – 2013

Orientador: Mg. Emma Norma Cambillo Moyano
Título Obtenido: Licenciado en Estadística

Este trabajo de tesis se realizó con el objetivo de proyectar las exportaciones mensuales del Perú mediante el Análisis Factorial Dinámico, modelo que consiste en la reducción de la dimensionalidad de las series a ser utilizadas como variables explicativas en la proyección de una serie macroeconómica desde un número grande a un número más pequeño de indicadores o factores. Este modelo fue desarrollado por Sargent y Sims (1977) e implementado por Aguirre y Céspedes (2004). En el Perú, las variables bajo estudio se concentraron en el primer factor al que se le denominó factor macroeconómico, posteriormente, se aplicó un modelo SARIMA para suavizar los errores de estimación y un Análisis de Intervención con el objetivo de disminuir el efecto producido por la caída de las exportaciones peruanas en el 2009 debido a la crisis mundial en ese periodo. Finalmente se obtuvo un modelo interesante el cual otorgaba mayor exactitud en la proyección de las exportaciones en el Perú explicando su variabilidad en un 97.5%.

PALABRAS CLAVES. ANÁLISIS FACTORIAL
TABLA MULTIWAY
ANÁLISIS DE INTERVENCIÓN
ANÁLISIS DE REGRESIÓN
MODELOS ARIMA

S U M M A R Y

DYNAMIC FACTOR ANALYSIS FOR THE PROJECTION OF THE EXPORTS IN PERÚ

CHRISTIAN FERNANDO FLORES CAMACHO

October – 2013

Guidance: Mg. Emma Norma Cambillo Moyano
Retrieved title: Graduate in Statistical

This thesis work was carried out with the aim of projecting monthly exports of Peru using Dynamic Factor Analysis model which consists in reducing the dimensionality of the series to be used as explanatory variables in the projection of a macroeconomic series from a number great in a smaller number of indicators or factors. This model was developed by Sargent and Sims (1977) and implemented by Aguirre and Céspedes (2004). In Peru, the variables under study concentrated on the first factor was called macroeconomic factor then SARIMA model was applied to smooth the estimation errors and Intervention Analysis in order to reduce the effect of the Peruvian exports decline in 2009 due to the global crisis in that period. Finally, an interesting pattern was obtained which gave more accuracy in projecting Peru exports in explaining the variability in a 97.5%.

KEY WORDS.

FACTOR ANALYSIS
TABLE MULTIWAY
ANALYSIS OF INTERVENTION
REGRESSION ANALYSIS
ARIMA MODELS

INDICE GENERAL

INTRODUCCIÓN	- 12 -
CAPÍTULO I	- 17 -
CONCEPTOS BÁSICOS	- 17 -
1.1 TABLAS “ <i>MULTIWAY</i> ”	- 17 -
1.1.1 Definición de Tablas “ <i>Multway</i> ”	- 18 -
A) Arreglo (“ <i>Array</i> ”).....	- 20 -
B) Conjunto (“ <i>Set</i> ”).....	- 20 -
1.1.2 Ejemplos de Tablas “ <i>Multway</i> ”	- 23 -
1.1.3 Estructuración de la Matriz de Datos.....	- 24 -
1.1.4 Representación Geométrica de la Matriz de Datos.....	- 27 -
1.1.5 Clasificación de los Métodos para el Análisis de Tablas “ <i>Multway</i> ”	- 30 -
A) Métodos de Análisis Tucker2 y Tucker3	- 30 -
B) Método de Análisis Parafac	- 31 -
C) Método de Análisis Statis	- 31 -
1.2 MODELOS DE SERIES DE TIEMPO	- 32 -
1.2.1 Procesos Autorregresivos: AR(p)	- 34 -
1.2.2 Procesos de Medias Móviles: MA(q)	- 34 -
1.2.3 Procesos Autorregresivos de Medias Móviles: ARMA(p,q).....	- 34 -
1.2.4 Modelos ARIMA(p,d,q)	- 35 -
1.3 ANÁLISIS DE INTERVENCIÓN	- 37 -
A) Ejemplo de Impulso.....	- 38 -
B) Ejemplo de Escalón	- 38 -

CAPÍTULO II	- 41 -
ANÁLISIS FACTORIAL DINÁMICO	- 41 -
2.1 INTRODUCCIÓN.....	- 41 -
A) Enfoque directo.....	- 42 -
B) Enfoque dual	- 45 -
2.2 MATRIZ DE VARIANZAS Y COVARIANZAS GLOBAL.....	- 49 -
2.2.1 Descomposición de la variabilidad	- 49 -
A) Para la parte Factorial	- 53 -
B) Para la parte Regresiva.....	- 53 -
CAPÍTULO III	- 54 -
ANÁLISIS FACTORIAL DINÁMICO PARA LA PROYECCIÓN DE LAS EXPORTACIONES EN EL PERÚ	- 54 -
3.1 ANALISIS EXPLORATORIO DE LAS EXPORTACIONES.....	- 56 -
3.2 PROYECCIÓN DE LAS EXPORTACIONES	- 57 -
3.2.1 Analisis Factorial Dinámico	- 58 -
3.2.2 SARIMA.....	- 59 -
CONCLUSIONES Y RECOMENDACIONES	- 64 -
BIBLIOGRAFÍA	- 66 -
ANEXO 1	- 68 -
ANEXO 2	- 70 -
ANEXO 3	- 76 -

LISTA DE FIGURAS

Figura N° 1: Ejemplo de matriz de 2 vías (unidades x variables)	- 18 -
Figura N° 2: Ejemplo de arreglo medido en 3 ocasiones	- 20 -
Figura N° 3: Ejemplo de conjunto medido en 3 ocasiones	- 21 -
Figura N° 4: “Two-way data”	- 22 -
Figura N° 5: “Three-way data”	- 22 -
Figura N° 6: Matriz de datos observada en el tiempo t considerando I filas.....	- 25 -
Figura N° 7: Matriz de datos observada en el tiempo t considerando J filas	- 26 -
Figura N° 8: Matriz de datos observada en cada variable j considerando T filas ..	- 27 -
Figura N° 9: Exportaciones del Perú (millones de US dólares)	- 56 -
Figura N° 10: Matriz de dispersión de la exportación y las variables macroeconómicas	- 57 -
Figura N° 11: Autocorrelaciones de los residuos del modelo de Regresión	- 59 -
Figura N° 12: Autocorrelaciones Parciales de los residuos del modelo de Regresión	- 60 -
Figura N° 13: Autocorrelaciones de los residuos del modelo SARIMA (3, 0, 1) (1, 0, 1)	- 60 -
Figura N° 14: Autocorrelaciones Parciales de los residuos del Modelo SARIMA (3, 0, 1) (1, 0, 1)	- 61 -
Figura N° 15: Proyección de las Exportaciones del Modelo SARIMA (3, 0, 1) (1, 0, 1)	- 62 -

LISTA DE CUADROS

Cuadro N° 1: Tabla “ <i>multiway</i> ” unidades, variables y tiempo.....	- 55 -
Cuadro N° 2: Matriz de Correlaciones de la exportación y las variables macroeconómicas.....	- 57 -
Cuadro N° 3: Porcentaje de Variación Explicada - Variables Macroeconómicas .-	58 -
Cuadro N° 4: Estimaciones del Modelo SARIMA (3, 0, 1) (1, 0, 1).....	- 61 -
Cuadro N° 5: Cuadro comparativo de la Exportación y su Estimación	- 63 -

INTRODUCCIÓN

En el Perú, en los últimos años, las exportaciones han crecido año tras año. Pero ¿qué es lo que el Perú exporta y quiénes son los que compran? El presente trabajo explora y analiza la evolución de las exportaciones Peruanas durante los últimos años. Ahora bien, para empezar, es necesario comprender las dos categorías o tipos principales que se utilizan para clasificar las exportaciones: estas son las exportaciones tradicionales y las exportaciones no tradicionales. Tal como indican los nombres, las exportaciones tradicionales comprenden aquellos productos que dominaron las exportaciones peruanas en el pasado (y aún ahora lo siguen haciendo). Las exportaciones no tradicionales, por el contrario, incluyen los productos que son exportados desde tiempos más recientes.

Cabe resaltar que dentro de las exportaciones no tradicionales las prendas de vestir (exportaciones textiles) forman el rubro que más contribuye a este tipo de exportación; le siguen los productos elaborados a base de cobre y las legumbres, sobre todo los espárragos, paltas y alcachofas. Se sabe que más del 50% de las exportaciones del Perú en textiles tienen como destino los Estados Unidos (EE.UU), además es este país quien compra, también, un tercio de las Exportaciones Peruanas agropecuarias (categoría que comprende a las frutas y legumbres).

La predominancia de los EE.UU se debe no solamente a su gran mercado interno, sino también al ATPDEA (Andean Trade Promotion and Drug Eradication Act) que permite la exportación de estos productos sin aranceles. El ATPDEA venció en diciembre del 2006, pero en la actualidad ha sido reemplazado por el Tratado de Libre Comercio entre Perú y EE.UU.

Como se sabe el comercio exterior es un intercambio comercial de bienes y servicios realizado entre países, siendo ello muy importante para el crecimiento y desarrollo sostenido a largo plazo; más aún, en la actualidad tiene un peso muy importante en la actividad económica de cada país, y todo esto es evidenciado por los países desarrollados, los cuáles han logrado su crecimiento y desarrollo económico gracias al incremento de sus exportaciones. Es decir, el desarrollo del comercio exterior o también llamado comercio internacional hace que los países prosperen, debido a que se realiza un aprovechamiento de lo que un país produce mejor intercambiándolo con los productos que otros países mejor producen y de este modo se genera un trueque comercial que los beneficia sustancialmente.

En el Perú la venta de las mercancías al mundo ha ido creciendo paulatinamente, no sólo en materias primas (exportaciones tradicionales como oro, cobre, zinc, algodón, etc.) sino también en nuevos productos (exportaciones no tradicionales como textiles, químicas, agropecuarias, metal mecánicas, etc.). Por lo antes expuesto, en el presente trabajo, se aplicará el Análisis Factorial Dinámico (AFD) para realizar la proyección de las exportaciones mensuales con buena precisión, dada su gran importancia en la economía de todo país considerando que es una componente importante del PBI¹.

Cabe mencionar que se han realizado diversos trabajos sobre proyección de exportaciones en el Perú, como el trabajo realizado por el Banco Central de Reserva del Perú (BCRP)², el cuál utiliza la información mensual desde Enero 2006 a Marzo 2007 para proyectar la Balanza Comercial (Proyección hacia Abril 2007) en miles de millones de dólares de las exportaciones e importaciones mediante el superávit comercial, este análisis se basa en la recopilación estadística procesada por la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT) hasta inicios del mes que se sujeta a regularizaciones de las DUA's de exportaciones.

³ Esta proyección adicionalmente toma en cuenta los precios internacionales de los principales "commodities" observados en el mes de abril del presente año, los cuales pueden diferir de los precios de venta efectivamente realizados.

¹ Producto Bruto Interno.

² "Proyección de precios de exportación utilizando tipos de cambio: Caso peruano" por Jesús Ferreyra y José L. Vásquez.

³ DUA es la Declaración Única de Aduana, cuya fecha de regularización de la mercancía puede darse luego de haber sido embarcada.

El Análisis Factorial Dinámico ha sido empleado como un innovador instrumento para realizar proyecciones en prestigiosos bancos centrales, como es el caso del Banco Central de Chile, el Banco de Canadá y el Banco Central Europeo, con la particularidad que los dos primeros han adoptado exitosamente esquemas de metas mínimas de inflación. Para el caso del Banco Central de Chile⁴, se analizaron 306 meses, desde marzo de 1986 a marzo de 2004, además realizaron cuatro tipos de modelos para realizar las proyecciones de la inflación (IPC). Se concluye de este trabajo que las proyecciones sobre la base de factores tienen una mayor calidad predictiva que el método Auto Regresivo (AR) para todos los horizontes considerados en el análisis (conforme el horizonte de proyección aumente, las estimaciones de los modelos con factores mejoran en relación con los modelos autorregresivos).

Para el Banco de Canadá⁵, se utilizaron datos trimestrales cuyo periodo fue desde el 1º trimestre de 1969 al 1º trimestre del 2000 con 334 variables macroeconómicas y financieras, complementariamente, se emplearon 110 variables de la economía de EE.UU. Estimaron cuatro modelos con y sin los datos de EE.UU, concluyéndose que cada uno de los modelos produce proyecciones igualmente precisos, desde el punto de vista estadístico; también se concluye la posibilidad de alguna evidencia de que los datos de EE.UU pueden ayudar a predecir cambios en la dirección de la inflación Canadiense. Contar con modelos que proyecten adecuadamente las exportaciones en el corto, mediano y largo plazo es necesario para realizar una adecuada administración de los ingresos monetarios por este régimen (Exportación Definitiva); de modo que se pueda distribuir correctamente dichos ingresos a los distintos organismos e instituciones gubernamentales, con la finalidad que prioricen los requerimientos de la nación. Es importante saber que si no se realiza una adecuada proyección de los ingresos por comercio exterior, ello ocasionaría un desequilibrio de proporciones considerables dependiendo de la coyuntura actual.

Como se había mencionado anteriormente el objetivo fundamental de este trabajo es desarrollar un modelo que pueda proyectar adecuadamente las exportaciones (variable macroeconómica) con la mayor exactitud posible.

⁴ Estudiado por Aguirre y Céspedes (2004).

⁵ Estudiado por Grosselin y Tkacz (2001).

Se sabe que las principales metodologías de proyección económica consideran, en la mayoría de los casos, solo un subconjunto del total de variables disponibles, tales como los modelos de vectores autoregresivos (VAR) que utilizan en general un número limitado de variables. Una alternativa que supera dicha limitación, son los modelos de proyección que involucran el uso de un número elevado de series económicas⁶. Esta metodología se enmarca dentro del Análisis Factorial Dinámico⁷ y consiste en la reducción de un número de variables a un número más pequeño de componentes o factores.

El método estadístico permite utilizar un análisis de componentes principales en el cálculo de los factores y estos son utilizados en las proyecciones de la variable dependiente. El Análisis Factorial es una metodología que mantiene una alta coherencia con los datos, es decir, ante la imposibilidad de apoyarse en una teoría económica para relacionar la variable dependiente con una gama amplia de variables disponibles, se pone énfasis en la correcta aplicación de la técnica para llegar al objetivo principal: **la proyección.**

En este trabajo se utiliza una metodología alternativa para la proyección de variables macroeconómicas, considerando un tercer ámbito o espacio de estudio (el tiempo), lo cual le otorga un análisis dinámico que considera la influencia del tiempo para el comportamiento de las unidades medidas en distintas variables; concibiendo, así, un análisis más exhaustivo debido a que se tiene un modelo que contiene la mayor variabilidad de las unidades medidas, ahora también, en el tiempo. La metodología utilizada se basa en la aplicación del Análisis Factorial Dinámico para obtener proyecciones sobre las exportaciones definitivas del Perú a partir de un conjunto de variables macroeconómicas medidas en diferentes ocasiones o tiempos que influyen en dicha proyección. La fuente de información para construir la matriz de datos provino de los Compendios Estadísticos correspondientes a los años 2000, 2002, 2004 y 2007 publicados por el INEI, además de la información publicada en la página web del BCRP, MINTRA y SUNAT.

⁶ Propuestos por Stock y Watson (1998).

⁷ Desarrollada por Sargent y Sims (1977)

Se consideró para el análisis los meses comprendidos entre Enero 1995 y Diciembre 2011 clasificados en variables del sector externo como *importaciones definitivas* y *exportaciones definitivas*, variables de ingresos como la *recaudación aduanera* e *ingresos tributarios*, variables del mercado laboral como el *empleo* y la *remuneración mínima* y variables del mercado financiero como el *PBI Global* y la *demanda interna*. La aplicación de la técnica estadística se ha realizado mediante el programa estadístico SPSS 18.0 y Stata 11.0, este último cuenta con una opción especializada en el cálculo del Análisis Factorial Dinámico.

Es importante especificar que todo este estudio se desarrollará en tres grandes y diferenciados capítulos: En el capítulo I se presentan las definiciones previas o conceptos básicos para comprender mejor en qué consiste el Análisis Factorial Dinámico, partiendo de conceptos geométricos como son las tablas “*multiway*” y la matriz de datos hasta la definición de técnicas de Series de Tiempo como son el Análisis de Intervención y los modelos ARIMA, técnicas que se utilizarán para disminuir los errores de estimación. En el capítulo II se procederá a detallar la técnica del Análisis Factorial Dinámico desde diversos enfoques para dilucidar como al considerar una tercera variabilidad, la del tiempo, contribuye en obtener estimaciones más naturales debido a que muchas veces al realizar un análisis no se toma en cuenta la influencia que pudiese tener el tiempo, lo cual representa una dimensión interesante en el Análisis Multivariante.

Finalmente, en el capítulo III, se desarrolla la proyección de la Exportaciones del Perú en valor FOB⁸ utilizando el Análisis Factorial Dinámico mediante un Análisis Exploratorio inicial de las variables macroeconómicas que formarán un factor macroeconómico el cual explicará el comportamiento de las Exportaciones en el Perú aplicando el Análisis de Intervención y el ARIMA para minimizar los errores de estimación y así obtener un modelo factorial más eficiente.

⁸ FOB: Cuyas iniciales significan: “Free on Board” (Libre a bordo). Es un Incoterm (International Commercial Terms o términos comerciales internacionales, en español) que se refiere al valor de venta de los productos en su lugar de origen, es decir que la mercancía es responsabilidad del vendedor hasta sobrepasar la borda del barco para la exportación.

CAPÍTULO I

CONCEPTOS BÁSICOS

En este capítulo se presentan las definiciones previas para comprender mejor el desarrollo del Análisis Factorial Dinámico, partiendo de conceptos como son las tablas “*multiway*” y la matriz de datos hasta la definición del Análisis de Intervención y los modelos ARIMA, técnicas estudiadas en Series de Tiempo que se utilizarán para disminuir los errores de estimación debido a que, como se mencionó anteriormente, las exportación en valor FOB es una variable macroeconómica muy dispersa y sensible a la ocurrencia de algún evento.

1.1 TABLAS “*MULTIWAY*”

En el análisis multivariante, generalmente, la matriz de datos se define como un conjunto de datos elementales a dos índices ($i = 1, \dots, I; j = 1, \dots, J$) concernientes a los valores tomados por J variables X_j en I unidades estadísticas. En este caso se trata de “*two-way data*” (matriz de 2 vías o direcciones), ya que cada elemento del conjunto de datos es caracterizado por un par de índices: i para el modo *unidades* y j para el modo *variables*. Estos datos se obtienen de estudios e investigaciones en diferentes áreas del conocimiento como las ciencias ambientales, sociales, del comportamiento, biomédicas, entre otras. Un ejemplo simple, al respecto, puede ser un conjunto de 10 gatos a quienes se les midió la edad en meses, el peso en gramos y el tamaño en centímetros; entonces se puede construir una matriz de datos a 2 vías en la cual i varía desde 1 hasta 10, es decir I toma el valor de 10 (número total de gatos) mientras que j tomaría los valores 1, 2, 3; es decir J sería 3 (número de variables medidas: edad, peso y tamaño).

Figura N° 1: Ejemplo de matriz de 2 vías (unidades x variables)

		Variables		
		Edad	Peso	Tamaño
Unidades	Gato 1 (i = 1)	17	978,5	40,5
	Gato 2 (i = 2)	15	987,9	41,2
	Gato 3 (i = 3)	19	1001,5	43,9
	Gato 4 (i = 4)	21	1217,1	42,7
	Gato 5 (i = 5)	17	1018,8	41,7
	Gato 6 (i = 6)	15	997,9	39,5
	Gato 7 (i = 7)	18	1425,5	39,9
	Gato 8 (i = 8)	20	1259,8	42,9
	Gato 9 (i = 9)	22	1359,3	42,3
	Gato 10 (i = 10)	18	1007,7	41,9

Se debe tener en cuenta, para la mejor comprensión del análisis “*multiway*”, que el modo es la clase particular de entidades como por ejemplo: individuos, atributos, lugares o tiempos, el índice es el criterio de clasificación más detallado del dato, como por ejemplo: n_{ijk} es la frecuencia con 3 índices y finalmente la dimensión es el espacio de la matriz sobre el cual se ubica uno o más modos o partes de un modo.

1.1.1 Definición de Tablas “*Multiway*”

Cuando el dato elemental se refiere a tres o más índices (vías o criterios de clasificación), entonces obtenemos la “*multiway data*” (matriz a muchas vías).

Particularmente, si consideramos 3 vías, los datos “*multiway*” serán de la forma siguiente:

$$X = \{x_{ijt}\} \quad ; i = 1, \dots, I; j = 1, \dots, J; t = 1, \dots, T$$

donde:

i : denota la vía del tipo *unidades*.

j : denota la vía del tipo *variables*.

t : denota la vía del tipo *ocasiones*.

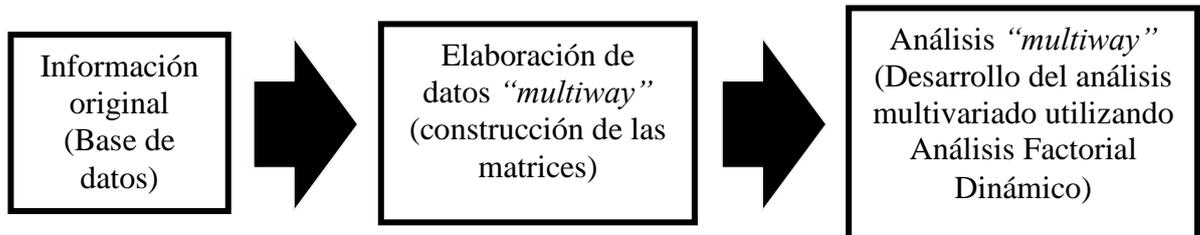
Los datos “*multiway*” se estructuran en: **unidades** (observaciones estadísticas), **variables** (medidas tomadas a cada una de las unidades) y finalmente **ocasiones** (momento en el cuál se analizan los valores de cada variable para cada unidad).

Para la vía o modo correspondiente a las ocasiones la perspectiva de su medición puede considerarse en diferentes formas, como por ejemplo: la medición del tiempo considerada en diferentes momentos de observación, la medición del tiempo considerada en posiciones diferentes, la medición del tiempo en diferentes condiciones experimentales, etc.

Los datos “*multiway*” tienen 3 o más índices (i, j, k,...), uno o más modos (unidad, variable, ocasión,...) y dos o más espacios (dimensiones). Además pueden ser obtenidos de diferentes maneras:

- a) Obtenerse directamente.
- b) Construirse especialmente mediante manipulación de los datos originales (lo que es más común).

El procedimiento para realizar un análisis “*multiway*” se da de la siguiente manera:



En cuanto a la clasificación de los datos “*multiway*” se debe tener en cuenta lo siguiente:

A. Arreglo (“Array”)

Son todas las combinaciones de los diferentes índices que se encuentran presentes.

Ejemplo: “Array” de 3 índices de la siguiente forma:

$$X = \{x_{ijt}\} \quad i = 1, \dots, I; j = 1, \dots, J; t = 1, \dots, T$$

donde consideramos I = 5 países, a los cuáles se les midió J variables económicas en 3 ocasiones o tiempos diferentes. La representación sería de la siguiente manera:

Figura N° 2: Ejemplo de arreglo medido en 3 ocasiones

	<u>Ocasión 1</u> J variables Económicas	<u>Ocasión 2</u> J variables Económicas	<u>Ocasión 3</u> J variables Económicas
<u>Unidad 1</u> Perú			
<u>Unidad 2</u> Ecuador			
<u>Unidad 3</u> Colombia			
<u>Unidad 4</u> Brasil			
<u>Unidad 5</u> Bolivia			

B. Conjunto (“Set”)

No todas las combinaciones de los valores de los diferentes índices se encuentran presentes. Ejemplo: “Set” de 3 índices.

$$X = \{X_1, \dots, X_t, \dots, X_T\}$$

Donde X_t es una matriz unidad*variables en la ocasión t , pero las variables son diferentes para cada una de las ocasiones (tenemos T grupos de variables diferentes).

Figura N° 3: Ejemplo de conjunto medido en 3 ocasiones

	<u>Ocasión 1</u> K ₁ variables Económicas	<u>Ocasión 2</u> K ₂ variables Sociales	<u>Ocasión 3</u> K ₃ variables Demográficas
<u>Unidad 1</u> Perú			
<u>Unidad 2</u> Ecuador			
<u>Unidad 3</u> Colombia			
<u>Unidad 4</u> Brasil			
<u>Unidad 5</u> Bolivia			

La situación más estudiada, tanto para “array” como para “set” es la de datos a tres vías del tipo: “unidades x variables x ocasiones”. Se puede afirmar que los datos “*multiway*” a tres vías se encuentran totalmente cruzados, es decir cada i-ésima unidad se encuentra medida con respecto a las mismas j-ésimas variables en diferentes t-ésimos tiempos u ocasiones. Así conseguimos el arreglo a tres vías ó tres modos.

A continuación se presenta un ejemplo sobre como un conjunto de datos medidos desde el enfoque con dos índices (I unidades x J variables) que puede ser observado en diferentes tiempos (“*multiway data*” de I unidades x J variables x T tiempos). Es decir se cuenta en un inicio, con un arreglo “*two-way data*” al cuál se le decide medir también un tercer modo (el tiempo) y así se obtiene el arreglo “*three-way data*”.

Figura N° 4: “Two-way data”

2001				
	Álgebra	Trigonometría	Aritmética	Geometría
Karen	117	137	125	115
Christian	115	139	125	115
Carlos	107	120	127	105
Yahaira	112	112	135	111
Gilberto	105	125	115	101
Débora	87	135	99	114
Pedro	99	117	121	107

Figura N° 5: “Three-way data”

2003				
	Álgebra	Trigonometría	Aritmética	Geometría
Karen	175	187	179	199
Christian	177	187	181	197

2002				
	Álgebra	Trigonometría	Aritmética	Geometría
Karen	125	157	159	122
Christian	125	155	155	125

2001				
	Álgebra	Trigonometría	Aritmética	Geometría
Karen	117	137	125	115
Christian	115	139	125	115
Carlos	107	120	127	105
Yahaira	112	112	135	111
Gilberto	105	125	115	101
Débora	87	135	99	114
Pedro	99	117	121	107

Nótese que en la figura N° 4 la matriz de datos sólo considera dos modos o vías correspondiendo las unidades para los alumnos y las variables para los puntajes obtenidos en cada curso; mientras que para la figura N° 5 la tabla “*multiway*” considera los tres modos o vías tomando en cuenta a los alumnos como unidades, los puntajes de los cursos como variables y finalmente el año de estudio como el tiempo u ocasión. Con este ejemplo se podría decir que el enfoque de tres modos o vías es el más adecuado si se desea tener en cuenta como evolucionaron las calificaciones de los estudiantes durante el 2001 al 2003, y de este modo ingresa la variabilidad ocasionada por el tiempo.

En conclusión, el análisis de datos tres vías se realiza cuando la información se presenta en arreglos donde cada dato se indexa con tres índices: uno que identifica el individuo i ($i = 1, \dots, I$), otro la variable j ($j = 1, \dots, J$) y un tercero que corresponde a la condición k ($k = 1, \dots, K$); representándose en un arreglo tridimensional: individuos, variables y condiciones.

Este tipo de datos ocurren frecuentemente, se destacan algunas aplicaciones en áreas como cromatografía, análisis ambiental y análisis de percepción⁹. Los métodos utilizados para el análisis de datos tres vías son de carácter exploratorio, porque identifican patrones de la estructura interna presente entre los elementos de las tres vías del arreglo, sin aplicar pruebas a estos patrones porque probar hipótesis supone un supuesto distribucional que en muchas situaciones se desconoce¹⁰.

Estos métodos describen la interacción entre las tres vías, prestando atención especial a las diferencias entre individuos, lo cual permite generar conclusiones más completas que las obtenidas al realizar análisis de dos vías aislados (individuos y variables).

1.1.2 Ejemplos de Tablas “*Multiway*”

- Se desea analizar la talla, peso, cantidad de glucosa, cantidad de triglicéridos, cantidad de colesterol mensual de 200 reclutas al ingresar a las fuerzas armadas desde su reclutamiento enero 1999 hasta su licenciamiento en diciembre del 2000. Para este ejemplo las observaciones son cada uno de los 200 reclutas ($i=1, 2, 3, \dots, 200$); las variables son talla, peso, cantidad de glucosa, cantidad de triglicéridos y cantidad de colesterol en su organismo ($j=1, 2, 3, 4, 5$) y por último los tiempos u ocasiones serían las mediciones de las variables mencionadas para cada uno de estos individuos en los meses comprendidos entre enero 1999 y diciembre 2000 ($t=1, 2, 3, \dots, 24$). Es decir para esta investigación se consideraría una matriz de datos de 200 filas y $5 \times 24 = 120$ columnas.

⁹ Bro (1998).

¹⁰ Kroonenberg 2008.

- Otro ejemplo práctico sería analizar la temperatura, humedad, afluencia de lluvias por metro cuadrado, cantidad de CO_2 enviado a la atmósfera y demanda de aerosoles en 100 ciudades a nivel mundial desde enero 1990 hasta diciembre del 2000. Para este ejemplo las observaciones o individuos serían las 200 ciudades a nivel mundial ($i=1, 2, 3, \dots, 200$); las variables medidas a estas ciudades serían la temperatura, humedad, afluencia de lluvias por metro cuadrado, cantidad de CO_2 enviado a la atmósfera y demanda de aerosoles ($j=1, 2, 3, 4, 5$) y las ocasiones serían los 132 meses considerados ($t=1, 2, 3, \dots, 132$). Es decir para este estudio interesante la matriz de datos estaría formada por 200 filas y $5 \cdot 132 = 660$ columnas.

1.1.3 Estructuración de la Matriz de Datos

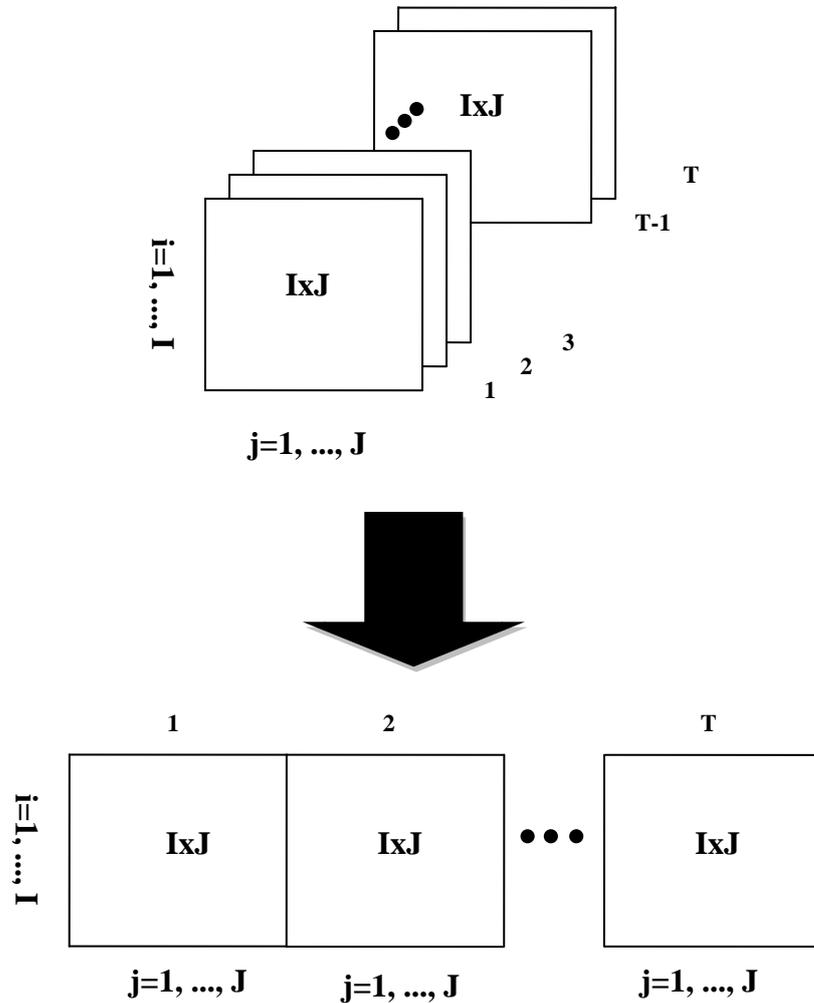
Existen varios enfoques o estructuraciones de la matriz de datos para los arreglos “*multiway*”, por ejemplo:

1. La matriz de datos, observada en el tiempo t , considerando I filas:

$$X_{(I \times J \times T)} \Rightarrow X_{(I, JT)}$$

Se tienen T matrices $I \times J$, es decir a I individuos se les midió J variables en T tiempos diferentes. Para este caso formamos una nueva matriz que tiene I filas y $J \cdot T$ columnas. Lo mencionado geoméricamente se representa del siguiente modo:

Figura N° 6: Matriz de datos observada en el tiempo t considerando I filas

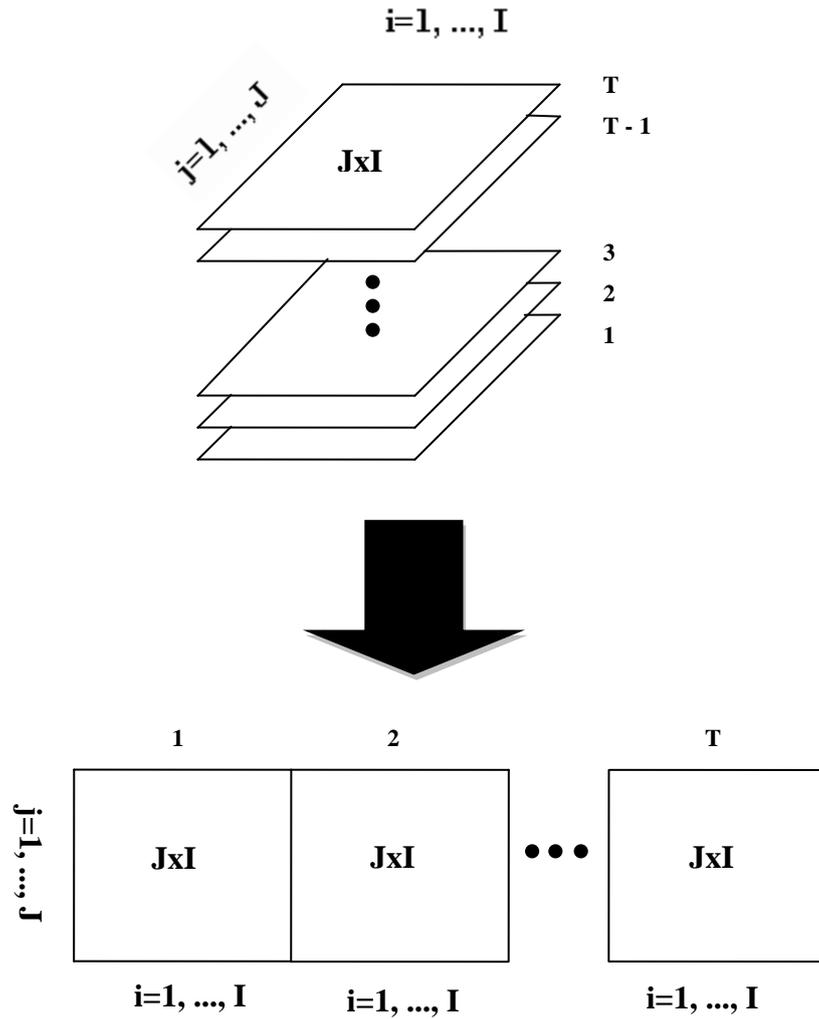


2. La matriz de datos, observada en el tiempo t , considerando J filas:

$$X_{(IxJ \times T)} \Rightarrow X_{(J, IT)}$$

En este caso, se tienen T matrices $J \times I$, es decir a J variables se les identifica cada i -ésimo individuo ($i = 1, 2, 3, \dots, I$) en T tiempos diferentes. Para este caso formamos una nueva matriz que tiene J filas e $I \times T$ columnas. Lo mencionado geoméricamente se representa del siguiente modo:

Figura N° 7: Matriz de datos observada en el tiempo t considerando J filas

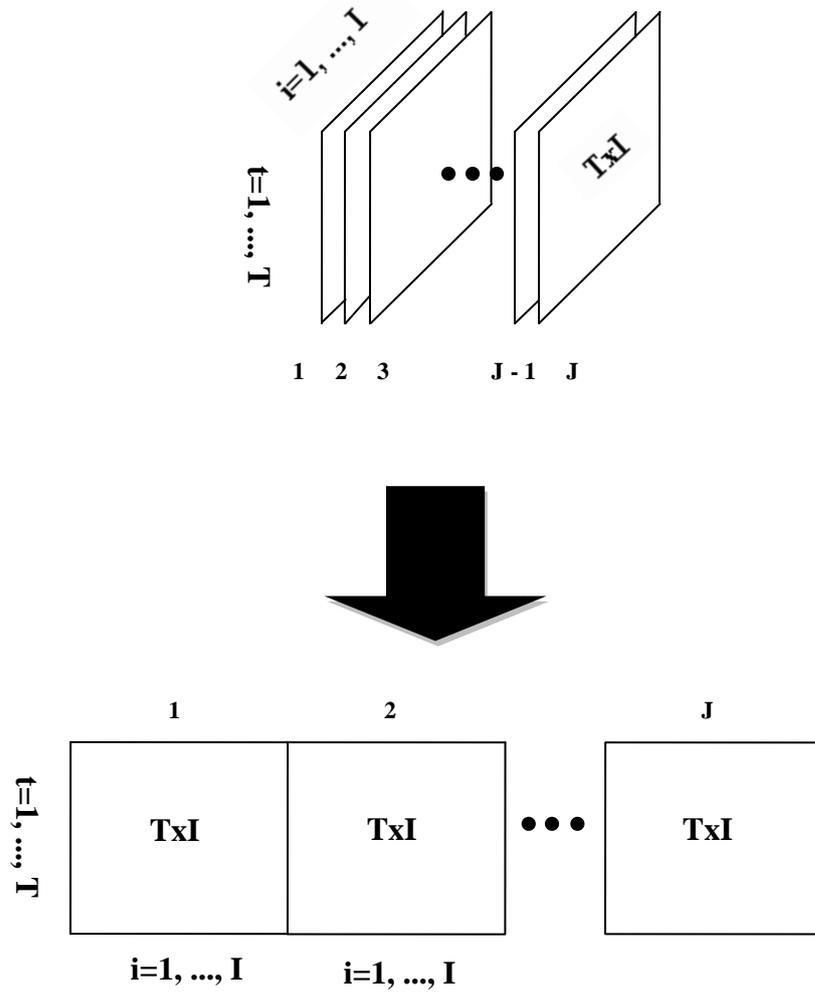


3. La matriz de datos, observada en cada variable j , considerando T filas:

$$X_{(I \times J \times T)} \Rightarrow X_{(T, IJ)}$$

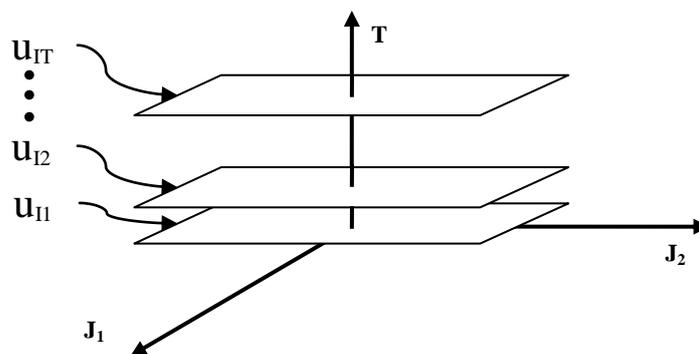
Ahora se tienen J matrices $T \times I$, es decir a T tiempos se les identifica cada i -ésimo individuo ($i = 1, 2, 3, \dots, I$) en J variables diferentes. Para este caso formamos una nueva matriz que tiene T filas e $I \times J$ columnas. Lo mencionado geoméricamente se representa del siguiente modo:

Figura N° 8: Matriz de datos observada en cada variable j considerando T filas



1.1.4 Representación Geométrica de la Matriz de Datos

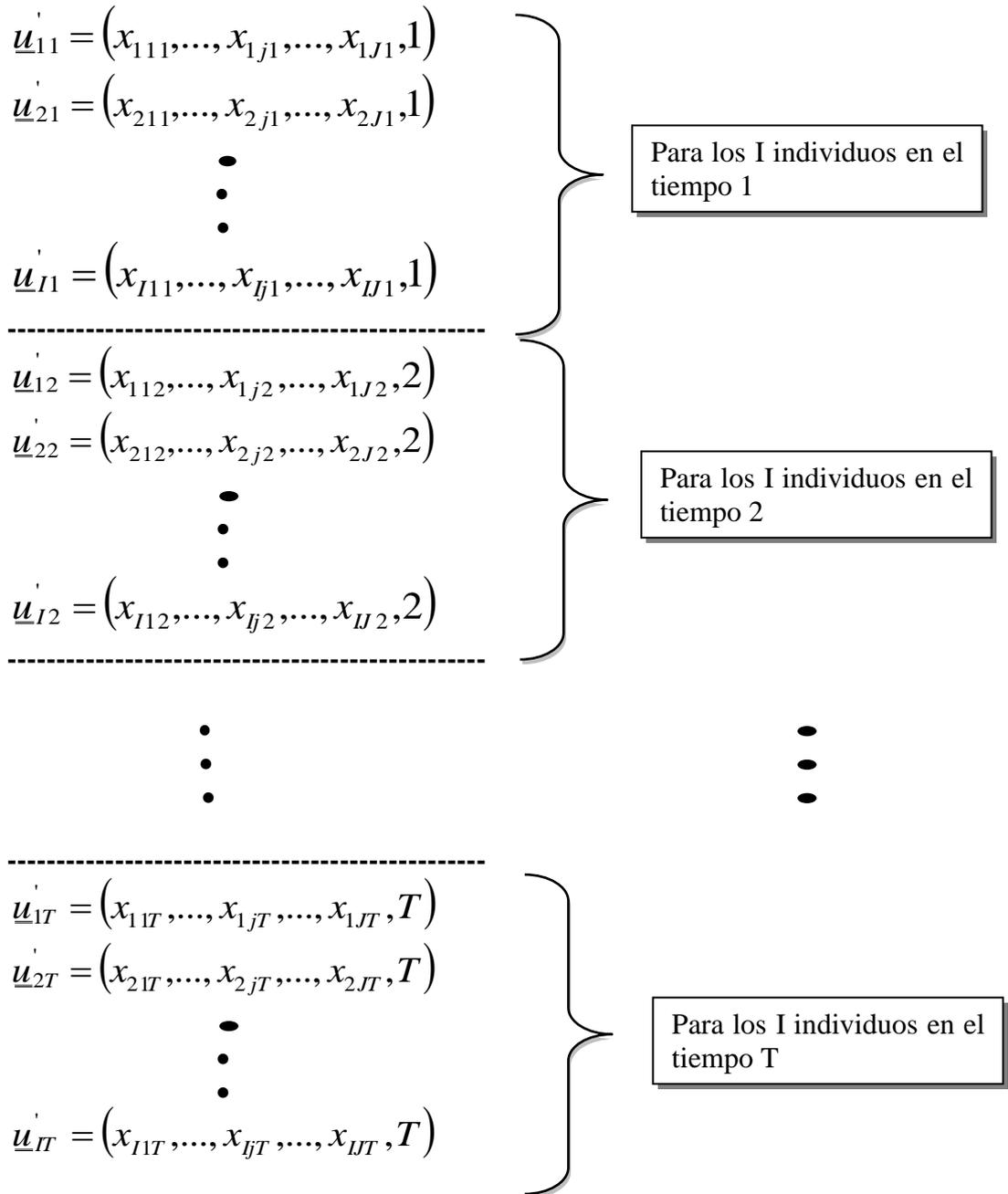
Si se considera $J = 2$ la representación geométrica del arreglo $X_{I,J,T}$ se representa de la siguiente manera:



Sea el espacio R^{J+1} que tiene como ejes coordenados las J variables y el tiempo. En R^{J+1} se considera el conjunto de vectores \underline{u}_{it} ordenados en J+1 elementos¹¹, contruidos del siguiente modo:

$$\underline{u}'_{it} = (x_{i1t}, \dots, x_{ijt}, \dots, x_{iJt}, t) \quad ; \quad i = 1, \dots, I; t = 1, \dots, T$$

Es decir:



¹¹ Coppi

En consecuencia la matriz de datos $X_{I,J,T}$ pasaría de la forma:

$$X_{(I,J,T)} = \begin{pmatrix} x_{111} & x_{121} & \dots & x_{1J1} & x_{112} & x_{122} & \dots & x_{1J2} & \dots & x_{11T} & x_{12T} & \dots & x_{1JT} \\ x_{211} & x_{221} & \dots & x_{2J1} & x_{212} & x_{222} & \dots & x_{2J2} & \dots & x_{21T} & x_{22T} & \dots & x_{2JT} \\ x_{311} & x_{321} & \dots & x_{3J1} & x_{312} & x_{322} & \dots & x_{3J2} & \dots & x_{31T} & x_{32T} & \dots & x_{3JT} \\ \dots & \dots \\ x_{I11} & x_{I21} & \dots & x_{IJ1} & x_{I12} & x_{I22} & \dots & x_{IJ2} & \dots & x_{I1T} & x_{I2T} & \dots & x_{IJT} \end{pmatrix}$$

a la forma siguiente, en función de \underline{u}_{it} :

$$X_{(I,J,T)} = \begin{pmatrix} \underline{u}_{11} & \underline{u}_{12} & \dots & \underline{u}_{1T} \\ \underline{u}_{21} & \underline{u}_{22} & \dots & \underline{u}_{2T} \\ \underline{u}_{31} & \underline{u}_{32} & \dots & \underline{u}_{3T} \\ \dots & \dots & \dots & \dots \\ \underline{u}_{I1} & \underline{u}_{I2} & \dots & \underline{u}_{IT} \end{pmatrix}_{kT}$$

Otro tipo de arreglo “*multiway*” son las matrices de proximidad, donde cada elemento es una medida de proximidad entre pares de individuos evaluados en T ocasiones independientes.

$$P = \{P_1, \dots, P_t, \dots, P_T\}$$

P_t es la matriz de proximidad en la ocasión t ($t = 1, \dots, T$). También P representa datos a 3 vías. P puede obtenerse directa o indirectamente.

1.1.5 Clasificación de los Métodos para el Análisis de Tablas “Multiway”

Las investigaciones sobre este tipo de datos muestran diferentes métodos de análisis provenientes de escuelas como la anglosajona y la francesa. Dentro de los métodos implementados por la escuela anglosajona se destaca la familia de modelos Tucker¹² y el modelo PARAFAC¹³. De la escuela francesa por su parte, sobresalen los métodos CANDECOM¹⁴, STATIS¹⁵ y AFM¹⁶. Entre algunos estudios que realizan comparaciones entre métodos, se destacan: la comparación entre STATIS y AFM realizado por Pagés (1996), la relación jerárquica entre algunos métodos tres vías presentado por Kiers (1991) y una comparación empírica entre algunos métodos tres vías planteado por Mures et al. (2006); estos dos últimos estudios se caracterizan por considerar métodos de las dos escuelas de análisis de datos.

En algunos métodos “multiway” para arreglos a tres vías, la información relacionada al orden del modo *ocasiones* no es considerado en el análisis (por ejemplo STATIS, TUCKER2 y TUCKER3). A continuación se detallarán algunos métodos:

A. Métodos de análisis TUCKER2 y TUCKER3

Se clasifican dentro de los modelos multilineales que pueden ser utilizados, en particular, para el estudio de arreglos cuantitativos de 3 índices. Para Tucker2 o modelo trilineal las ocasiones son consideradas en su expresión empírica es decir el conjunto efectivamente observado de las k ocasiones.

- Modelo cuadrilineal (Tucker3)

$$x_{ijk} = m_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} h_{pqr}$$

a_{ip} : puntaje de la unidad i en la dimensión latente p.

b_{jq} : puntaje de la unidad j en la dimensión latente q.

c_{kr} : puntaje de la ocasión k en la dimensión latente r.

¹² Tucker 1, 2 y 3 (Tucker 1966)

¹³ PARAllel FACtor analysis, Harshman (1970)

¹⁴ CANonical DECOMposition, Carroll & Chang (1970)

¹⁵ Structuration des Tableaux A Trois Indices de la Statistique, Lavit et al. (1994)

¹⁶ Análisis Factorial Múltiple, Escofier & Pagés (1982)

h_{pqr} : puntaje de la interacción entre las tres componentes de los tres modos.

- Modelo trilineal (Tucker2)

$$x_{ijk} = m_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q a_{ip} b_{jq} h_{pqk}$$

B. Método de análisis PARAFAC

El PARAFAC [Bro, 1997] es un método de descomposición que, conceptualmente, puede ser comparado a un PCA bilineal. El modelo fue propuesto independientemente por Harshman (1997) y por Carroll y Chang [Carroll, 1970] los cuales lo llamaron CANDECOMP (descomposición canónica). La descomposición de los datos se realiza en componentes trilineales, pero en vez de un vector de *scores* y uno de *loadings* como en el caso del Análisis de Componentes Principales (PCA) bilineal, cada componente consiste de un vector de *scores* y dos vectores de *loadings*.

$$x_{ijk} = m_{ijk} = \sum_{p=1}^P a_{ip} b_{jp} c_{kp}$$

En la práctica no se distingue entre *scores* y *loadings*, ya que son tratados numéricamente por igual. Por tanto, un modelo PARAFAC de una estructura tridimensional viene definido por tres matrices de *loadings*, A(I*P), B(J*P) y C(K*P), cuyos elementos son a_{ip} , b_{jp} y c_{kp} respectivamente. P representa el número de factores; I, J y K pueden ser considerados como el número de muestras, longitudes de onda y tiempos respectivamente.

C. Método de análisis STATIS

Structuration des Tableaux a Trois Indices de la Statistique, este método utiliza estrategias para el análisis en tres fases del tipo I.S.I:

- Interestructura.
- Compromiso.
- Intraestructura.

Statis fundamentalmente puede analizar datos “set” de 3 índices, pudiendo también ser utilizado para “array” de 3 índices. Statis considera tres casos, en referencia a la tipología:

- Mismas unidades, diferentes variables en diversas ocasiones (“set”) tratado con Statis.
- Mismas variables, diferentes unidades en diversas ocasiones (“set”) tratado con Statis Dual.
- Mismas variables, mismas unidades en diferentes ocasiones (“array”) tratado con la aproximación de los resultados combinando los otros 2 casos.

En el Análisis Factorial Dinámico (AFD), contrario a STATIS, Tucker2 y Tucker3, el orden de las ocasiones si es considerado explícitamente en el modelo del análisis. En el siguiente capítulo se presentará el Análisis Factorial Dinámico, su metodología, aplicaciones, ventajas y desventajas para su posterior utilización en la proyección de las exportaciones.

1.2 MODELOS DE SERIES DE TIEMPO

Un proceso estocástico es una familia de variables aleatorias que, en general, están relacionadas entre sí y siguen una ley de distribución conjunta. Se denota por:

$$Y_t(w), \quad t = \dots, t-2, t-1, t, t+1, t+2, \dots \quad \text{ó} \quad \dots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}, \dots \quad \text{ó} \quad \{Y_t\}_{-\infty}^{+\infty}$$

Si se fija el tiempo, $t = t_0$, se tiene una variable aleatoria determinada de la secuencia, $Y_{t_0}(w)$, con su distribución de probabilidad univariante.

Si se fija la variable aleatoria $w = w_0$, se tiene una realización del proceso, es decir, una muestra de tamaño 1 del proceso formada por una muestra de tamaño 1 de cada una de las variables aleatorias que forman el proceso:

$$\dots, Y_{t-2}(w_0), Y_{t-1}(w_0), Y_t(w_0), Y_{t+1}(w_0), Y_{t+2}(w_0), \dots$$

En el marco estadístico de los procesos estocásticos, una serie temporal, Y_1, Y_2, \dots, Y_T , se puede interpretar como una realización muestral de un proceso estocástico que se observa únicamente para un número finito de periodos, $t = 1, 2, \dots, T$. Por lo tanto una serie temporal es una sucesión de observaciones en la que cada una de ellas corresponde a una variable aleatoria distinta, y la ordenación de la sucesión de observaciones es esencial para el análisis de la misma, siendo no alterable, porque se cambiarían las características de la serie que se quiere estudiar.

El concepto de estacionariedad se puede caracterizar bien en términos de la función de distribución o de los momentos del proceso. En el primer caso, se hablará de estacionariedad en sentido estricto y, en el segundo, de estacionariedad de segundo orden o en covarianza.

Se dice que un proceso estocástico es estacionario en sentido estricto cuando al realizar un mismo desplazamiento en el tiempo de todas las variables de cualquier distribución conjunta finita, resulta que esta distribución no varía. Si el proceso es estacionario en sentido estricto se deberá verificar que

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

Se dice que un proceso es estacionario de primer orden, o en media, si se verifica que:

$$E(Y_t) = \mu, \quad \forall t$$

Por tanto, en un proceso estacionario en media, la esperanza matemática, o media teórica, permanece constante a lo largo del tiempo.

Se dice que un proceso es estacionario de segundo orden (o en sentido amplio) cuando se verifican las siguientes condiciones:

- i) $E(Y_t - \mu)^2 = \sigma^2 < \infty, \quad \forall t$
- ii) $E(Y_{t+k} - \mu)(Y_t - \mu) = \gamma_k, \quad \forall t$

1.2.1 Procesos Autorregresivos: AR(p)

El proceso autorregresivo de orden p, expresa Y_t en función de su pasado hasta el retardo $t - p$ y una innovación contemporánea:

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

En términos del operador de retardos,

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) Y_t = a_t \longrightarrow \phi_p(L) Y_t = a_t$$

donde $\phi_p(L)$ recibe el nombre de polinomio autorregresivo y $(\phi_1, \phi_2, \dots, \phi_p)$ es el vector de parámetros autorregresivos.

1.2.2 Procesos de Medias Móviles: MA(q)

El modelo medias móviles finito de orden q, expresa el valor de Y_t en función de la innovación contemporánea y de su pasado hasta el retardo q:

$$Y_t = a_t - \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}, \quad t = 1, 2, \dots$$

En términos del operador de retardos queda como sigue:

$$Y_t = (1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3 - \dots - \theta_q L^q) a_t \longrightarrow Y_t = \theta_q(L) a_t$$

donde $\theta_q(L)$ se denomina polinomio de medias móviles y $(\theta_1, \theta_2, \theta_3, \dots, \theta_q)$ es el vector de parámetros de medias móviles.

1.2.3 Procesos Autorregresivos de Medias Móviles: ARMA(p,q)

Los procesos autorregresivos de medias móviles determinan Y_t en función de su pasado hasta el retardo p, de la innovación contemporánea y el pasado de la innovación hasta el retardo q:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad , \quad t = 1, 2, \dots$$

Este modelo se puede escribir en términos del operador de retardos como sigue:

$$(1 - \phi_1 L - \dots - \phi_p L^p) Y_t = (1 - \theta_1 L - \dots - \theta_q L^q) a_t$$

$$\phi_p(L) Y_t = \theta_q(L) a_t$$

donde $\phi_p(L)$ es el polinomio autorregresivo y $\theta_q(L)$ es el polinomio medias móviles.

1.2.4 Modelos ARIMA(p,d,q)

Sea el siguiente modelo ARMA(p,q):

$$\phi_p(L) Y_t = \theta_q(L) a_t \tag{1.1}$$

donde el polinomio AR se puede factorizar en función de sus p raíces L_1, L_2, \dots, L_p ,

$$\phi_p(L) = (1 - L_1^{-1}L)(1 - L_2^{-1}L) \dots (1 - L_p^{-1}L)$$

Supongamos que (p - 1) raíces son estacionarias (con módulo fuera del círculo unidad) y una de ellas es unitaria, $L_i = 1$. Entonces, el polinomio AR, se puede reescribir como sigue:

$$\phi_p(L) = (1 - L_1^{-1}L)(1 - L_2^{-1}L) \dots (1 - L_p^{-1}L) = \varphi_{p-1}(L)(1 - (1)^{-1}L)$$

$$\phi_p(L) = \varphi_{p-1}(L)(1 - L)$$

donde el polinomio $\varphi_{p-1}(L)$ resulta del producto de los (p - 1) polinomios de orden 1 asociados a las raíces L_i con módulo fuera del círculo unitario.

Sustituyendo en el modelo ARMA(p,q) (1.1) se tiene que:

$$\varphi_{p-1}(L)(1 - L) Y_t = \theta_q(L) a_t \longrightarrow \varphi_{p-1}(L) \Delta Y_t = \theta_q(L) a_t \tag{1.2}$$

donde el polinomio $\phi_{p-1}(L)$ es estacionario porque todas sus raíces tienen módulo fuera del círculo unitario y el polinomio $\Delta = (1-L)$ es el que recoge la raíz unitaria. El modelo (1.2) representa el comportamiento de un proceso Y_t que no es estacionario porque tiene una raíz unitaria. A un proceso Y_t con estas características se le denomina proceso integrado de orden 1.

En general, el polinomio AR del modelo (1.1) puede contener más de una raíz unitaria, por ejemplo, d , entonces se puede descomponer como:

$$\phi_p(L) = \phi_{p-d}(L)(1-L)^d$$

Y sustituyendo, de nuevo, en el modelo ARMA(p,q) (1.1), se tiene:

$$\phi_{p-d}(L)\Delta^d Y_t = \theta_q(L)a_t$$

donde el polinomio $\phi_{p-d}(L)$ es estacionario porque sus $(p-d)$ raíces tienen módulo fuera del círculo unitario, y el polinomio $\Delta^d = (1-L)^d$, de orden d , contiene las d raíces unitarias no estacionarias. A un proceso Y_t con estas características se le denomina proceso integrado de orden d y se denota por $Y_t \sim I(d)$.

Un proceso Y_t es integrado de orden d , $Y_t \sim I(d)$, si Y_t no es estacionario, pero su diferencia de orden d , $\Delta^d Y_t$, sigue un proceso ARMA($p-d$, q) estacionario e invertible.

El orden de integración del proceso es el número de diferencias que hay que tomar al proceso para conseguir la estacionariedad en media, o lo que es lo mismo, el número de raíces unitarias del proceso. En la práctica, los procesos que surgen más habitualmente en el análisis de las series temporales económicas son los $I(0)$ e $I(1)$, encontrándose los $I(2)$ con mucha menos frecuencia.

En general, si una serie Y_t es integrada de orden d , se puede representar por el siguiente modelo:

$$\phi_p(L)\Delta^d Y_t = \delta + \theta_q(L)a_t \tag{1.3}$$

donde el polinomio autorregresivo estacionario $\phi_p(L)$ y el invertible de medias móviles $\theta_q(L)$ no tienen raíces comunes.

El modelo (1.3) se denomina modelo Autorregresivo Integrado de Medias Móviles de orden (p, d, q) o ARIMA(p, d, q), donde p es el orden del polinomio autorregresivo estacionario, d es el orden de integración de la serie, es decir, el número de diferencias que hay que tomar a la serie para que sea estacionaria, y q es el orden del polinomio de medias móviles invertible.

1.3 ANÁLISIS DE INTERVENCIÓN

La Intervención es la acción y efecto de dirigir o influenciar los asuntos, decisiones o comportamiento de alguna persona, empresa, institución o fenómeno determinado. Por ejemplo un caso de intervención estatal está relacionado a que los bancos centrales de los países también participan en los mercados de divisas cuando consideran necesario estabilizar el precio de la moneda nacional. Asimismo una intervención puede ser interpretada como la ocurrencia de un evento externo al comportamiento usual o histórico de una variable en estudio¹⁷.

Una limitación de la modelización ARIMA univariante está en que, provocados por distintos motivos, aparecen con frecuencia puntos atípicos al comportamiento natural de la variable en estudio que dan lugar a errores excepcionalmente elevados. Se trata de circunstancias excepcionales que perturban la dinámica general de la serie que deseamos modelar. Para incorporar tales fenómenos al modelo contamos con el Análisis de Intervención con dos posibilidades principales: variables de impulso y variables de escalón.

Es decir, las series con frecuencia se ven afectadas por sucesos puntuales a veces conocidos. Ejm: una huelga, un año bisiesto, un cambio legal o un cambio en feriado. Al incluir estos efectos en la serie se pueden obtener mejoras significativas en la precisión de la estimación de los parámetros y de las predicciones.

¹⁷ [1] Victor M. Guerrero, 1991 y O. D. Anderson, 1980

A. Ejemplo de Impulso

En la predicción de la serie de producción de un producto en una empresa se ha presentado una huelga en un momento t . Se puede incluir ese efecto a través de una variable ficticia que llamaremos variable impulso. Es decir sea la serie de producción Z_t de modo tal que:

$$Z_t = \mu + \omega I_t + h_t$$

donde:

μ : media poblacional.

h_t : modelo ARIMA

w : magnitud de I_t

I_t : es la variable impulso tal que:

$$I_t = \begin{cases} 1 & \text{si } t = \text{huelga} \\ 0 & \text{si } t \neq \text{huelga} \end{cases}$$

B. Ejemplo de Escalón

En la predicción de la serie correspondiente al desempleo de un país determinado se produce un cambio legal el cual modifica la definición de desempleo a partir de un determinado momento. Es decir sea la serie del desempleo Z_t de modo tal que:

$$Z_t = \mu + \omega S_t + h_t$$

donde:

μ : media poblacional.

h_t : modelo ARIMA

w : magnitud de S_t

S_t : es la variable escalón tal que:

$$S_t = \begin{cases} 1 & \text{si } t < \text{cambio legal} \\ 0 & \text{si } t \geq \text{cambio legal} \end{cases}$$

De acuerdo a Guerrero (1991) [1], el análisis de intervención podría considerarse como una extensión del análisis de procesos ARIMA, ya que un modelo para una serie que contenga los efectos de una intervención, puede expresarse como:

$$(X_t) = \varepsilon_{I,t} + N_t$$

donde:

X_t = Serie de tiempo de interés.

$\varepsilon_{I,t}$ = Función que representa los efectos de la intervención.

N_t = Modelo ARIMA, estacionario e invertible, que representa a la parte estocástica de la serie (no se consideran efectos estacionales sobre la serie, para evitar complicaciones en la notación) y se escribe como:

$$\phi(B)\nabla^d N_t = \theta_0 + \theta(B)a_t$$

Con N_t medida como desviaciones con respecto a su media si $d=0$, mientras que es $\varepsilon_{I,t}$ una función que permite representar los efectos que se presentaron no sólo en el momento en que ocurrió la intervención, sino después de esta.

La función de $\varepsilon_{I,t}$ permite medir los cambios ocurridos en la serie de tiempo cuando se presenta la intervención como posterior a ésta. Es necesario, que $\varepsilon_{I,t}$ sea un modelo dinámico de la intervención.

Para lograr la especificación de $\varepsilon_{I,t}$, se hace uso de la siguiente función de pulso:

$$P_{I,t} = \begin{cases} 1 & \text{si } t = I \\ 0 & \text{si } t \neq I \end{cases}$$

I se refiere al momento en que ocurrió la intervención. Con esta función, que define a una serie de tiempo toma el valor de 1 en el momento $t = I$ y 0 en cualquier otro momento, es entonces razonable suponer que puede satisfacer una ecuación del tipo:

$$(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_p B^p) \varepsilon_{I,t} = (\omega_0 - \omega_1 B - \omega_2 B^2 - \dots - \omega_q B^q) P_{I,t} \quad (1.4)$$

Cuya solución depende de la condición:

$$\varepsilon_{I,t} = 0 \quad \text{para } t < I$$

Ello indica que antes de la intervención no hay efectos atribuibles a ella. El caso más sencillo de una función de intervención es cuando $p = q = 0$, en este caso se tendría $\varepsilon_{I,t} = \omega_0 P_{I,t}$ que gráficamente luciría como una función de pulso, en donde la magnitud del “brinco” sería de tamaño ω_0 .

Ahora si se quiere encontrar la solución (o forma explícita) para otros valores de p y q , es necesario resolver la ecuación (1.4). Por ejemplo, si $p = 1$ y $q = 1$, se tiene que:

$$\varepsilon_{I,t} = \delta_1 \varepsilon_{I,t-1} + \omega_0 P_{I,t} + \omega_1 P_{I,t-1} \quad (1.5)$$

Lo que conduce explícitamente a:

$$\begin{aligned} \varepsilon_{I,I} &= \omega_0 \\ \varepsilon_{I,I+1} &= \delta_1 \omega_0 - \omega_1 \\ \varepsilon_{I,I+2} &= \delta_1^2 \omega_0 - \delta_1 \omega_1 \\ &\bullet \bullet \bullet \\ \varepsilon_{I,I+m} &= \delta_1^m \omega_0 - \delta_1^{m-1} \omega_1 \end{aligned} \quad (1.6)$$

Es decir:

$$P_{I,t} = \begin{cases} 0 & \text{si } t < I \\ \omega_0 & \text{si } t = I \\ \delta_1^{t-I} \omega_0 - \delta_1^{t-I-1} \omega_1 & \text{si } t > I \end{cases}$$

A partir del modelo (1.4) se postula un modelo dinámico más general, en el cual se consideran raíces unitarias para el polinomio de retraso que opera sobre $\varepsilon_{I,t}$, de esta forma se obtiene:

$$\phi(B) \nabla^d \varepsilon_{I,t} = \omega(B) P_{I,t} \quad (1.7)$$

con

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_p B^p$$

$$\omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \dots - \omega_q B^q$$

y la ecuación (1.7) representa el Modelo Dinámico General de Intervención.

CAPÍTULO II

ANÁLISIS FACTORIAL DINÁMICO

2.1 INTRODUCCIÓN

El Análisis Factorial Dinámico (AFD) propuesto por Coppi R. y Zannella F. en 1978 permite analizar arreglos de 3 vías del tipo “unidades * variables * tiempos” cuando se tienen variables cuantitativas de registro cronológico.

$$X = \{x_{ijt}\} \quad ; i = 1, \dots, I \quad ; j = 1, \dots, J \quad ; t = 1, \dots, T$$

En este tipo de arreglo los “tiempos” juegan el rol de ocasiones, en consecuencia se trata de un arreglo con 3 vías del tipo “unidades, variables y ocasiones”, en el cual las ocasiones se encuentran ordenadas. Este método “*multiway*” analiza no sólo las características estructurales de un conjunto de variables, sino también su evolución. El AFD constituye toda una metodología del análisis de dispersión, medido a través de la matriz de varianzas y covarianzas (S) y las matrices de proximidades entre unidades (P) y entre tiempos (Q), cada una de dichas matrices se descompone en la suma de tres matrices que miden aspectos diferentes de la variabilidad total, utilizándose para ello simultáneamente análisis de componentes principales y análisis de regresión. El arreglo $X = \{X_{ijt}\}$ con: $i = 1, \dots, I$; $j = 1, \dots, J$ y $t = 1, \dots, T$ puede ser analizado en modo bidimensional mediante dos enfoques: el enfoque directo y el enfoque dual.

A. Enfoque directo

Considera las matrices estandarizadas de la matriz de datos $X(IT,J)$ (estructurada verticalmente en una matriz de $I \times T$ filas y J columnas) o $X(TI,J)$ (estructurada también verticalmente sólo que ahora se tienen $T \times I$ filas y J columnas) denotadas por $Z(IT,J)$ ó $Z(TI,J)$ respectivamente. La dispersión total se mide mediante la matriz de varianzas y covarianzas:

$$S = \{s_{jj'}\} \quad ; j, j' = 1, \dots, J$$

El AFD es un método para el análisis de dispersión medido a través de S , para lo cual S se descompone en la suma de tres matrices, cada una de las cuales mide un aspecto diferente de la variabilidad total. Se puede demostrar que S puede expresarse del siguiente modo:

$$S = S_i^* + S_t^* + S_{it}$$

donde:

1) S_i^* , variabilidad sintética, es la matriz de varianzas y covarianzas de las unidades medias estandarizadas \bar{Z}_{ij} , centradas respecto a las medias estandarizadas relativas a cada variable $\bar{Z}_{.j}$, en ambos casos no se considera la influencia del tiempo. Expresa la variabilidad entre unidades medias, es decir no aparece la influencia del tiempo, que es **sintetizado**, para cada unidad, por la media de las diferentes observaciones relativas a dicha unidad. Geométricamente sería:

$$Para : t = 1; j = 1 \rightarrow \bar{Z}_{.11} = \frac{1}{T} (Z_{(1)11} + Z_{(2)11} + \dots + Z_{(I)11})$$

$$Para : t = 1; j = 2 \rightarrow \bar{Z}_{.21} = \frac{1}{T} (S_{(1)21} + S_{(2)21} + \dots + S_{(I)21})$$

• •
• •
• •

$$Para : t = 1; j = J \rightarrow \bar{Z}_{.J1} = \frac{1}{T} (S_{(1)J1} + S_{(2)J1} + \dots + S_{(I)J1})$$

$$Para : t = 2; j = 1 \rightarrow \bar{Z}_{.12} = \frac{1}{T} (Z_{(1)12} + Z_{(2)12} + \dots + Z_{(I)12})$$

$$Para : t = 2; j = 2 \rightarrow \bar{Z}_{.22} = \frac{1}{T} (S_{(1)22} + S_{(2)22} + \dots + S_{(I)22})$$

• •
• •
• •

$$Para : t = 2; j = J \rightarrow \bar{Z}_{.J2} = \frac{1}{T} (S_{(1)J2} + S_{(2)J2} + \dots + S_{(I)J2})$$

•
•
•

$$Para : t = T; j = 1 \rightarrow \bar{Z}_{.1T} = \frac{1}{T} (Z_{(1)1T} + Z_{(2)1T} + \dots + Z_{(I)1T})$$

$$Para : t = T; j = 2 \rightarrow \bar{Z}_{.2T} = \frac{1}{T} (S_{(1)2T} + S_{(2)2T} + \dots + S_{(I)2T})$$

• •
• •
• •

$$Para : t = T; j = J \rightarrow \bar{Z}_{.JT} = \frac{1}{T} (S_{(1)JT} + S_{(2)JT} + \dots + S_{(I)JT})$$

3) S_{it} variabilidad dinámica diferencial, es la matriz de varianzas y covarianzas de los valores estandarizados $Z_{ijt} - \bar{Z}_{ij.} - \bar{Z}_{.jt} + \bar{Z}_{.j.}$ donde Z_{ijt} son las observaciones estandarizadas propiamente dichas. Mide la variabilidad debida a la interacción entre unidades y tiempos.

El AFD analiza estadísticamente las tres componentes de variabilidad de la matriz de datos $X_{(IxJxT)}$ desde su estructuración a una matriz de 2 vías, que puede ser:

$$X_{(IxJxT)} \Rightarrow X_{(I,JT)} \quad (\text{Estructuración 1})$$

$$X_{(IxJxT)} \Rightarrow X_{(J,IT)} \quad (\text{Estructuración 2})$$

$$X_{(IxJxT)} \Rightarrow X_{(T,IJ)} \quad (\text{Estructuración 3})$$

B. Enfoque dual

El enfoque dual considera las matrices estandarizadas de $X(JT,I)$ o $X(TJ,I)$, denotadas por $Z(JT,I)$ ó $Z(TJ,I)$, por lo que se invierten los roles de unidades y variables en relación al enfoque directo. En el caso dual lo que se descompone es la matriz P de proximidad global entre las unidades.

$$P = P_j^* + P_t^* + P_{jt}$$

1) P_j^* es la matriz de proximidad de las $\bar{Z}_{ij.}$, centradas respecto a $\bar{Z}_{i..}$

Los grupos en este caso son las variables; la matriz expresa la dispersión de las unidades medias, en los diferentes grupos-variable. Esta fuente de dispersión expresa la estructura sintética del sistema.

2) P_t^* es la matriz de proximidad de las $\bar{Z}_{i.t}$ centradas respecto a $\bar{Z}_{i..}$

Expresa la dispersión de las unidades respecto al tiempo, es decir la dinámica media de las unidades.

3) P_{jt} es la matriz de proximidad de los valores $Z_{ijt} - \bar{Z}_{ij.} - \bar{Z}_{i.t} + \bar{Z}_{i..}$

Expresa la dinámica diferencial de las variables.

Tal como se mostró anteriormente, la matriz de datos X puede ser observada de dos maneras:

$$\text{i. } X_t = \{x_t\} = X_{I,J \times T} \quad ; t = 1, \dots, T$$

donde X_t es la matriz *unidadesXvariables* observada en el tiempo t .

$$\text{ii. } X_i = \{x_i\} = X_{J,I \times T} \quad ; i = 1, \dots, I$$

donde X_i es la matriz *tiemposXvariables* observada en la unidad i .

De tal manera que permite evaluar dos tipos de **variabilidad**:

- i. La variabilidad debida a las unidades estadísticas en cada tiempo t (en x_t), denominada variabilidad estructural.
- ii. La variabilidad debida al tiempo, para cada unidad estadística (en x_i), denominada variabilidad dinámica.

Estas variabilidades: estructural y dinámica, no están separadas, pues existe una interacción entre las dos, esencialmente debida a la diferente dinámica de cada unidad (interacción **unidades-tiempos**).

El AFD analiza los dos tipos de variabilidades y la interacción entre ellas utilizando dos modelos: el modelo factorial y el modelo regresivo.

Factorial: Para la parte estructural (en x_t)

Regresivo respecto al tiempo: Para la parte dinámica (en x_i)

En el AFD la matriz de varianzas y covarianzas global es calculada a partir de la matriz de datos, mediante una reestructuración del arreglo X.

La estructuración 1 de la matriz de datos $X_{(I \times J \times T)}$ es la matriz a dos vías siguiente:

$$X_{(I, J \times T)}$$

Siendo esta matriz de dos dimensiones de la forma siguiente:

$$X_{(I, J \times T)} = \begin{pmatrix} x_{111} & x_{121} & \dots & x_{1J1} & x_{112} & x_{122} & \dots & x_{1J2} & \dots & x_{11T} & x_{12T} & \dots & x_{1JT} \\ x_{211} & x_{221} & \dots & x_{2J1} & x_{212} & x_{222} & \dots & x_{2J2} & \dots & x_{21T} & x_{22T} & \dots & x_{2JT} \\ x_{311} & x_{321} & \dots & x_{3J1} & x_{312} & x_{322} & \dots & x_{3J2} & \dots & x_{31T} & x_{32T} & \dots & x_{3JT} \\ \dots & \dots \\ x_{I11} & x_{I21} & \dots & x_{IJ1} & x_{I12} & x_{I22} & \dots & x_{IJ2} & \dots & x_{I1T} & x_{I2T} & \dots & x_{IJT} \end{pmatrix}$$

En esta matriz las $J \times T$ columnas son las variables observadas en los diferentes tiempos u ocasiones.

En cuanto a la metodología de proyección considere la siguiente variable macroeconómica de interés, Y_t , a ser proyectada utilizando X_t , un grupo de variables explicativas de dimensión N , disponibles en el periodo t . Generalmente los métodos de proyección utilizan sólo un subconjunto del amplio grupo de variables X_t , potencialmente útiles para la proyección debido a los problemas estadísticos que implica un alto número de variables explicativas; en consecuencia, el desempeño de estos métodos tradicionales de proyección está determinado por el restringido grupo de variables elegidas. Una alternativa a estas metodologías de proyección consiste en agrupar la información contenida en el conjunto completo de variables (X_t) eliminando el contenido característico de cada serie. La idea es que la información relevante y por lo tanto la capacidad predictiva de un número mayor de variables macroeconómicas pueda ser concentrada o resumida en un número menor de factores no observables. Siendo estos factores las variables subyacentes que explicarían la variación conjunta en el tiempo de las variables observables. Asumiendo que (X_t, Y_{t+1}) puede ser representado por una estructura factorial dinámica con \bar{r} factores dinámicos comunes (f_t) se obtiene que:

$$X_{it} = \lambda_i(L)f_t + e_{it} \quad (i)$$

$$X_{t+h} = \beta_h(L)f_t + \gamma_h(L)Y_t + \zeta_{t+h} \quad (ii)$$

Donde h es el horizonte de proyección para la variable Y_t , e_{it} y ζ_{t+h} son shocks característicos; $\lambda_i(L)$, $\beta_h(L)$ y $\gamma_h(L)$ son operadores de rezagos, los últimos dos dependientes del horizonte de proyección elegido. La ecuación (i) indica que el comportamiento de las variables contenidas en el vector X_t puede ser explicado por un grupo común de factores, f_t , sus rezagos, y un componente característico. La segunda ecuación indica que estos factores junto con rezagos de la variable a ser proyectada son utilizados en la estimación. Si se asume adicionalmente que el número de rezagos es menor a q , un número finito, es posible expresar el sistema anterior a través de la siguiente representación estática:

$$X_t = \theta F_t + e_t \quad (iii)$$

$$Y_{t+h} = \beta_h' F_t + \gamma_h(L)Y_t + \varepsilon_{t+h} \quad (iv)$$

Donde $F_t = (f_t', \dots, f_{tq}')$ es una matriz de factores de orden $(r \times 1)$, con $r \leq (q+1)\bar{r}$. Stock y Watson (1998) muestran que asumiendo ciertas restricciones sobre ciertos momentos y condiciones de estacionariedad en las variables utilizadas, el espacio comprendido por los factores dinámicos, f_t , puede ser estimado consistentemente utilizando análisis de componentes principales sobre la matriz de covarianza de X_t . Adicionalmente, el uso de componentes principales permite que estos estimadores puedan ser calculados fácilmente, incluso cuando N , el número de variables explicativas, es grande, y permite ser generalizado para lidiar con irregularidades en los datos. Adicionalmente, Stock y Watson (1998) muestran que las proyecciones de Y_t obtenidas a partir del uso de los factores son asintóticamente eficientes si el número de factores incluidos en el modelo a estimar es igual o mayor al número de factores dinámicos comunes \bar{r} , y si la dimensión transversal de X_t es mucho mayor a la dimensión temporal.

2.2 MATRIZ DE VARIANZAS Y COVARIANZAS GLOBAL

Sea S la matriz de varianzas y covarianzas de $X_{(I,JxT)}$. Como se había explicado en el punto 2.1 el AFD busca explicar **la variabilidad total medida por S** , en particular la parte lineal de dicha variabilidad, es decir la traza de la matriz de varianzas y covarianzas $S_{J,J}$ que se denota como (trS) .

Sea:

$$S = S_{J,J} = \{s_{jj'}\} \quad j, j' = 1, \dots, J$$

Entonces:

$$S_{(JxJ)} = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1J} \\ s_{21} & s_{22} & \dots & s_{2J} \\ s_{31} & s_{32} & \dots & s_{3J} \\ \dots & \dots & \dots & \dots \\ s_{J1} & s_{J2} & \dots & s_{JJ} \end{pmatrix}$$

2.2.1 Descomposición de la variabilidad

Según Coppi, El AFD es un método para el análisis de la dispersión de la matriz $X_{(I,JxT)}$ mediante la descomposición de la variabilidad total S en la suma de tres matrices, cada una de las cuales mide un aspecto diferente de la variabilidad total. Se puede comprobar que S puede expresarse del siguiente modo:

$$S_{JxJ} = (S_i^*)_{JxJ} + (S_t^*)_{JxJ} + (S_{it})_{JxJ} \quad (2.1)$$

donde:

- 1) S_i^* , variabilidad sintética, es la matriz de varianzas y covarianzas de las \bar{X}_{ij} centradas respecto a las $\bar{X}_{.j}$.

Expresa la variabilidad entre unidades medias. Es decir que no aparece la influencia del tiempo, que es **sintetizado** -para cada unidad- por la media de las diferentes observaciones relativas a dicha unidad.

2) S_t^* , variabilidad dinámica media, es la matriz de varianzas y covarianzas de las $\bar{X}_{.jt}$ centradas respecto a las $\bar{X}_{.j}$. Mide por lo tanto la variabilidad entre los tiempos medios. Aquí no aparece la influencia de cada una de las unidades que son sustituidas por una “unidad media.

3) S_{it} , variabilidad dinámica diferencial, es la matriz de varianzas y covarianzas de los valores $x_{ijt} - \bar{x}_{ij} - \bar{x}_{.jt} + \bar{x}_{.j}$ y mide la variabilidad debida a la interacción entre unidades y tiempos.

El elemento genérico puede ser considerado como la suma de cuatro componentes distintos:

$$x_{ijt} = \bar{x}_{.j} + (\bar{x}_{ij} - \bar{x}_{.j}) + (\bar{x}_{.jt} - \bar{x}_{.j}) + (x_{ijt} - \bar{x}_{ij} - \bar{x}_{.jt} + \bar{x}_{.j}) \quad (2.2)$$

donde:

$\bar{x}_{.j}$ = Promedio general de la j-ésima variable.

$(\bar{x}_{ij} - \bar{x}_{.j})$ = Efecto debido a la estructura estática de las unidades.

$(\bar{x}_{.jt} - \bar{x}_{.j})$ = Efecto debido a la dinámica media.

$(x_{ijt} - \bar{x}_{ij} - \bar{x}_{.jt} + \bar{x}_{.j})$ = Efecto debido a la dinámica diferencial, que es la interacción entre las unidades y los tiempos.

La variabilidad total, puede expresarse como:

$$S = \sum_i \sum_j \sum_t (x_{ijt} - \bar{x}_{ij} - \bar{x}_{.jt} + \bar{x}_{.j})(x_{ijt} - \bar{x}_{ij} - \bar{x}_{.jt} + \bar{x}_{.j})'$$

Para la descomposición de la variabilidad total, Coppi, propuso tres modelos:

a) **Modelo 1:** Se basa en una descomposición particular de la variabilidad lineal de la matriz S:

$$trS = tr\bar{S}_t + trS_t^*$$

b) **Modelo 2:** Se basa en una descomposición particular de la variabilidad lineal de la matriz S:

$$trS = tr\bar{S}_i + trS_i^*$$

c) **Modelo 3:** Es una combinación de los modelos 1 y 2 partiendo del modelo 1

La relación (2.2) representa un modelo de dos factores para el análisis de varianza: el denominado Modelo 1 del Análisis Factorial Dinámico, considera que los diferentes componentes de (2.2) y los elementos relativos de la variabilidad total (2.1) en términos de un Análisis de Componentes Principales y un modelo de Regresión Lineal.

La descomposición fundamental en AFD en el enfoque directo es la siguiente:

$$S = \bar{S}_t + S_t^* \quad , \text{ donde:} \quad \bar{S}_t = S_i^* + S_{it}$$

Esta descomposición es una generalización al caso multivariado “*multiway*” de la tradicional descomposición de la variabilidad “*within*” (S_t^*) y “*between*” (\bar{S}_t) .

\bar{S}_t es la media de la variabilidad dentro de cada tiempo y en este sentido es denominada variabilidad estructural, ya que se compone de la estructura sintética (S_i^*) de las unidades y de la dinámica diferencial (S_{it}) de las mismas.

El elemento genérico de \bar{S}_t es:

$${}_t\bar{S}_{jj'} = \frac{1}{T} \sum_t \left\{ \frac{1}{I} \sum [(X_{ijt} - \bar{X}_{.jt})(X_{ij't} - \bar{X}_{.j't})] \right\}$$

y el elemento genérico de S_t^* es:

$${}_tS_{jj'}^* = \frac{1}{T} \sum_t [(\bar{X}_{.jt} - \bar{X}_{.j.})(\bar{X}_{.j't} - \bar{X}_{.j'.})]$$

Las tres componentes de la variabilidad son estudiadas a través de:

- Un Análisis de Componentes Principales (ACP) aplicado a la matriz \bar{S}_t que permite describir la estructura factorial media de las unidades (estructura sintética y dinámica diferencial).
- Un sistema regresivo que permite describir la dinámica temporal de los centros medios de las variables. Para ello, se utiliza un modelo de regresión lineal del tipo:

$$\bar{X}_{.jt} = \sum_{l=0}^k b_{lj} t^l + e_{.jt} \quad ; j = 1, \dots, J; t = 1, \dots, T$$

donde:

b_{lj} son los parámetros a estimar.

$e_{.jt}$ son los residuos.

Además satisfacen las condiciones:

$$\text{cov}(e_{jt}, e_{j't'}) = \begin{cases} w_j \in R^+ \text{ si } j = j'; t = t' \\ 0; \text{c.c} \end{cases}$$

Utilizando la representación geométrica de la matriz $X_{I,J \times T}$ observada en el tiempo t (X_t), se tiene una nube de vectores $\{\underline{u}_{it}\}$; $i=1, \dots, I$, de modo que al ir variando t las nubes se ubican sobre hiperplanos paralelos al sub-espacio coordinado R^J .

Mientras que utilizando la representación geométrica de la matriz $X_{J,I \times T}$ observada en la unidad i (X_i), se tiene una nube de vectores $\{\underline{u}_{it}\}$; $t=1, \dots, T$; las cuales describen la **trayectoria de la unidad i** en el tiempo.

Sea R^{J+1} , una métrica euclídea definida por una matriz M definida a su vez simétrica y positiva (Coppi). M sería de la siguiente forma:

$$M_{(J+1), (J+1)} = \begin{pmatrix} \underline{m}_{11} & \underline{m}_{12} & \dots & \underline{m}_{1(J+1)} \\ \underline{m}_{21} & \underline{m}_{22} & \dots & \underline{m}_{2(J+1)} \\ \underline{m}_{31} & \underline{m}_{32} & \dots & \underline{m}_{3(J+1)} \\ \dots & \dots & \dots & \dots \\ \underline{m}_{(J+1)1} & \underline{m}_{(J+1)2} & \dots & \underline{m}_{(J+1)(J+1)} \end{pmatrix}_{(J+1) \times (J+1)}$$

Entonces la distancia euclídiana entre los vectores \underline{u}_{it} y $\underline{u}_{i't'}$ está dada por:

$$d_M^2(\underline{u}_{it}, \underline{u}_{i't'}) = \|\underline{u}_{it} - \underline{u}_{i't'}\|_M^2 ; \quad \begin{cases} i, i' = 1, \dots, I \\ t, t' = 1, \dots, T \end{cases}$$

Es decir:

$$d_M^2(\underline{u}_{it}, \underline{u}_{i't'}) = \left((x_{it} - x_{i't'})^2 + \dots + (x_{ijt} - x_{i'jt'})^2 + \dots + (x_{iJt} - x_{i'Jt'})^2 + (t - t')^2 \right)$$

La variabilidad total, TrS , puede ser interpretada en términos de tales distancias:

A. Para la parte Factorial

En términos de distancias proyectadas sobre una o más variedades lineales de R^{J+1} (subespacios vectoriales afines, en particular se determinan subespacios obtenidos con el método de componentes principales).

B. Para la parte Regresiva

En términos de distancias de los vectores a oportunas funciones de regresión respecto al tiempo. (Se usarán, particularmente, funciones lineales).

CAPÍTULO III

ANÁLISIS FACTORIAL DINÁMICO PARA LA PROYECCIÓN DE LAS EXPORTACIONES EN EL PERÚ

Las variables del sector externo: importaciones para el consumo, variables de ingresos: la recaudación aduanera e ingresos tributarios, variables del mercado laboral: empleo y remuneración mínima y variables del mercado financiero: PBI Global y demanda interna; tienen una incidencia directa sobre el comportamiento de las exportaciones, como se verá en el desarrollo de este capítulo. Esta relación ha sido confirmada por diversos estudios macroeconómicos como el caso del Banco Central de Chile para la proyección de su inflación. Las variables macroeconómicas utilizadas, para proyectar las exportaciones mensuales, según valor FOB en millones de dólares, son las siguientes:

1. Empleo - Lima Metro. - Total - Empr. de 100 y más Trab. (Índice Base Mayo 2004 = 100) (**EMPLEO**).
2. Total Importaciones (Millones de US Dolares) (**IMPORT**).
3. PBI Global (Millones de nuevos soles a precios de 1994) (**PBI**).
4. Total Recaudación Aduanera (Millones de nuevos soles) (**REC_ADU**).
5. Remuneración Mínima Vital Nominal Mensual (US\$) (**REM_MIN**).
6. Demanda interna (índice 1994=100) (**DEM_INT**).
7. Ingresos tributarios (mill. S/.) (**ING_TRIB**).

La información sobre las variables macroeconómicas fue extraída de los Compendios Estadísticos y de la página web del Instituto Nacional de Estadística e Informática (INEI), del Banco Central de Reserva del Perú (BCRP) y de la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT).

Las variables macroeconómicas han sido consideradas desde Enero de 1995 a Diciembre del 2011, para el análisis forman un arreglo de (17 años)*(12 meses) = 204 observaciones, a los cuáles se les ha registrado las 7 variables macroeconómicas, obteniéndose un arreglo de (12 meses)*(7 variables)*(17 años). Es decir se tiene una tabla de dimensiones I, J y T de la siguiente forma:

- *I* varía del 1 al 12 (cada año tiene 12 meses de análisis).
- *J* varía del 1 al 7 (a cada mes se le miden 7 variables macroeconómicas).
- *T* varía del 1 al 17 (se está analizando desde el año 1995 al 2011).

La tabla “*multiway*” utilizada es la siguiente:

Cuadro N° 1: Tabla “*multiway*” unidades, variables y tiempo

TIEMPO	UNIDADES	VARIABLES						
		EMPLEO	IMPORT	PBI	REC_ADU	REM_MIN	DEM_INT	ING_TRIB
1995	Enero	X ₁₁₁	X ₁₂₁	X ₁₃₁	X ₁₄₁	X ₁₅₁	X ₁₆₁	X ₁₇₁
	Febrero	X ₂₁₁	X ₂₂₁	X ₂₃₁	X ₂₄₁	X ₂₅₁	X ₂₆₁	X ₂₇₁
	Marzo	X ₃₁₁	X ₃₂₁	X ₃₃₁	X ₃₄₁	X ₃₅₁	X ₃₆₁	X ₃₇₁

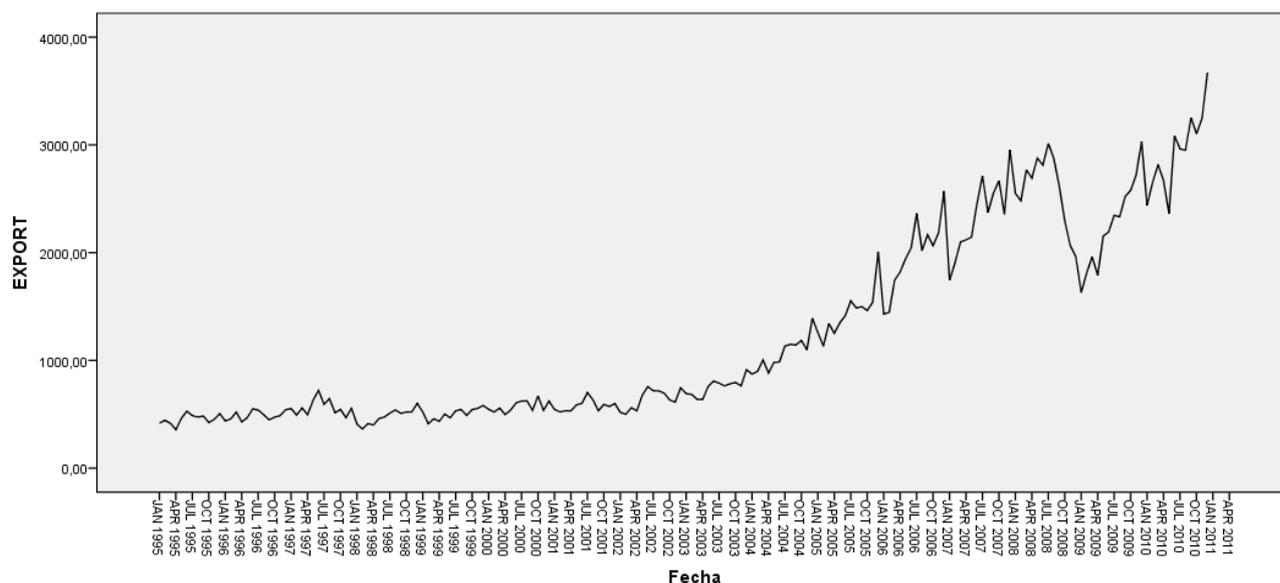
	Diciembre	X ₍₁₂₎₁₁	X ₍₁₂₎₂₁	X ₍₁₂₎₃₁	X ₍₁₂₎₄₁	X ₍₁₂₎₅₁	X ₍₁₂₎₆₁	X ₍₁₂₎₇₁
.	
.	
.	
2011	Enero	X ₁₁₍₁₇₎	X ₁₂₍₁₇₎	X ₁₃₍₁₇₎	X ₁₄₍₁₇₎	X ₁₅₍₁₇₎	X ₁₆₍₁₇₎	X ₁₇₍₁₇₎
	Febrero	X ₂₁₍₁₇₎	X ₂₂₍₁₇₎	X ₂₃₍₁₇₎	X ₂₄₍₁₇₎	X ₂₅₍₁₇₎	X ₂₆₍₁₇₎	X ₂₇₍₁₇₎
	Marzo	X ₃₁₍₁₇₎	X ₃₂₍₁₇₎	X ₃₃₍₁₇₎	X ₃₄₍₁₇₎	X ₃₅₍₁₇₎	X ₃₆₍₁₇₎	X ₃₇₍₁₇₎

	Diciembre	X ₍₁₂₎₁₍₁₇₎	X ₍₁₂₎₂₍₁₇₎	X ₍₁₂₎₃₍₁₇₎	X ₍₁₂₎₄₍₁₇₎	X ₍₁₂₎₅₍₁₇₎	X ₍₁₂₎₆₍₁₇₎	X ₍₁₂₎₇₍₁₇₎

3.1 ANALISIS EXPLORATORIO DE LAS EXPORTACIONES

Al realizar el análisis de la curva de exportación se observa que presenta una tendencia creciente de comportamiento exponencial. Sin embargo debido a la caída de las más importantes Bolsas de valores del mundo en Octubre 2008 las exportaciones presentan una caída significativa, que altera el comportamiento creciente, entre Octubre 2008 y Junio 2009, para luego superar la crisis mundial y retomar su tendencia creciente y exponencial.

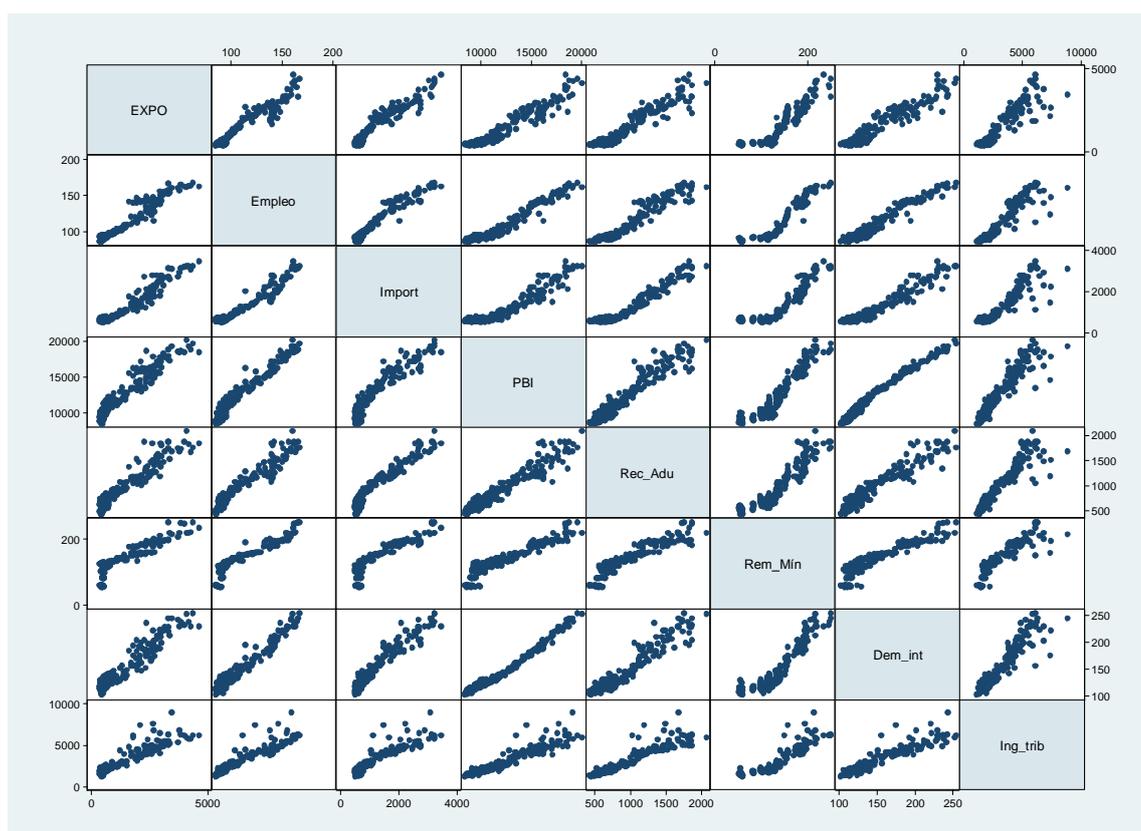
**Figura N° 9: Exportaciones del Perú (millones de US dólares)
Enero 1995 – Diciembre 2011**



Existe relación directa entre la Exportación mensual en valor FOB y las variables macroeconómicas (Figura N° 10). El incremento del Empleo total en Lima Metropolitana, Total Importaciones, PBI Global, Total Recaudación Aduanera, Remuneración Mínima Vital Nominal Mensual, Demanda interna e Ingresos tributarios inciden directamente en el incremento de la Exportación, ello a consecuencia que existe una alta correlación directa entre las mismas, sustentado en que todas las correlaciones son mayores a 0.8. (Cuadro N° 2).

Dado que las variables macroeconómicas están relacionadas con las Exportaciones se utilizó un Modelo Factorial Dinámico con mayor ajuste a los datos, permitiendo obtener estimaciones más precisas para las exportaciones mensuales.

Figura N° 10: Matriz de dispersión de la exportación y las variables macroeconómicas



Cuadro N° 2: Matriz de Correlaciones de la exportación y las variables macroeconómicas

	EXPORT	Empleo	Import	PBI	Rec_Adu	Rem_Min	Dem_int	Ing_trib
EXPORT	1	,956**	,947**	,953**	,940**	,876**	,943**	,909**
Empleo	,956**	1	,959**	,969**	,941**	,890**	,972**	,917**
Import	,947**	,959**	1	,930**	,957**	,839**	,954**	,896**
PBI	,953**	,969**	,930**	1	,937**	,913**	,989**	,925**
Rec_Adu	,940**	,941**	,957**	,937**	1	,899**	,938**	,912**
Rem_Min	,876**	,890**	,839**	,913**	,899**	1	,884**	,866**
Dem_int	,943**	,972**	,954**	,989**	,938**	,884**	1	,921**
Ing_trib	,909**	,917**	,896**	,925**	,912**	,866**	,921**	1

** . La correlación es significativa al nivel 0,01 (bilateral).

3.2 PROYECCIÓN DE LAS EXPORTACIONES

En los países en vías de desarrollo la Balanza Comercial, que se define como la diferencia entre las Exportaciones e Importaciones, es positiva lo que implica que un país se encuentra en superávit comercial.

Por ello la proyección de las Exportaciones es esencial para monitorear el desarrollo del país en base a un factor que concentre variables macroeconómicas, cuya incidencia directa con las exportaciones, permitan realizar una proyección eficaz obteniendo entonces un menor margen de error.

3.2.1 Análisis Factorial Dinámico

Dada la existencia de una alta correlación entre las variables macroeconómicas y las exportaciones, el AFD permitió obtener un factor global, que concentra el 94.21% de la variabilidad de los indicadores macroeconómicos (Cuadro N° 3).

Cuadro N° 3: Porcentaje de Variación Explicada - Variables Macroeconómicas

FACTOR	AUTOVALOR	DIFERENCIA	PROPORCION	PROPORCION ACUMULADA
Factor1	6,5947	6,53239	0,9421	0,942
Factor2	0,0623	0,02158	0,0089	0,951
Factor3	0,0407	0,03935	0,0058	0,956
Factor4	0,0014	0,00579	0,0002	0,957
Factor5	-0,0044	0,00333	-0,0006	0,956
Factor6	-0,0078	0,01168	-0,0011	0,955
Factor7	-0,0194		-0,0028	0,952

El factor global obtenido, denominado en adelante como Factor Macroeconómico, permitió obtener las puntuaciones factoriales bajo el modelo siguiente:

$$A = 0.99 * empleo + 0.97 * import + 0.98 * pbi + 0.97 * rec_adu$$

$$B = 0.93 * rem_min + 0.98 * dem_int + 0.94 * ing_trib$$

$$Factor = A + B$$

Una vez determinado el factor macroeconómico, se utilizó a este como variable exógena para poder explicar las Exportaciones mediante el Análisis de Regresión:

$$Exportaciones = 1369.95 + (Factor \text{ Macroeconómico}) * 999.56 \quad (3.1)$$

Obteniéndose un coeficiente de determinación de 0.929, es decir, el factor macroeconómico explica la variabilidad de las exportaciones en un 92.9%; lo cual puede ser interpretado como un promedio ponderado de las variables macroeconómicas. Las variables macroeconómicas Empleo total en Lima Metropolitana, Total Importaciones, PBI Global, Total Recaudación Aduanera, Remuneración Mínima Vital Nominal Mensual, Demanda interna e Ingresos tributarios, concentradas en el “factor macroeconómico”, realmente guardan relación directa con las Exportaciones mensuales en valor FOB. Sin embargo, el análisis de residuos detectó la presencia de autocorrelación, que es corregido mediante un modelo de series de tiempo SARIMA con Intervención a raíz de la caída de la Bolsa de Valores de Lima – Perú, en el 2009, tal como se comentó anteriormente.

3.2.2 SARIMA

Para seleccionar el modelo SARIMA se utilizaron las autocorrelaciones totales y parciales de los residuos del Modelo de Regresión (3.1) y se determinó que se debía utilizar el Modelo SARIMA (3, 0, 1) (1, 0, 1).

Figura N° 11: Autocorrelaciones de los residuos del modelo de Regresión (3.1)

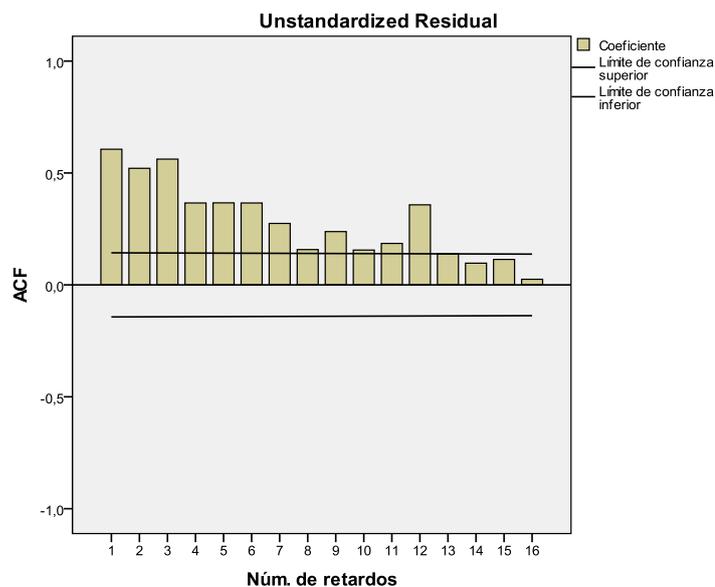
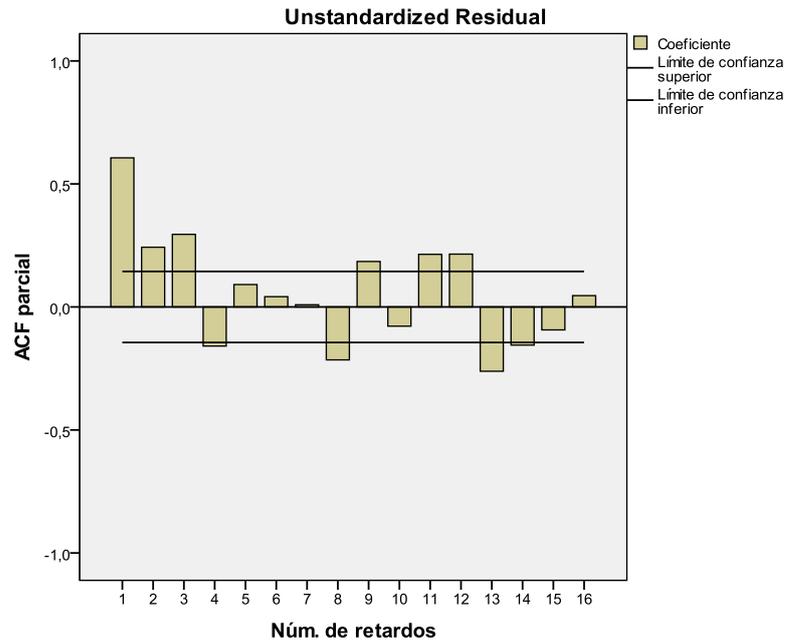


Figura N° 12: Autocorrelaciones Parciales de los residuos del modelo de Regresión (3.1)



Bajo el Modelo SARIMA(3, 0, 1) (1, 0, 1) las autocorrelaciones de los residuos se encuentran dentro de los límites establecidos; por ello se puede afirmar que, utilizando Intervención, se encuentran incorrelacionados. (Figura N° 13 y N° 14).

Figura N° 13: Autocorrelaciones de los residuos del modelo SARIMA (3, 0, 1) (1, 0, 1)

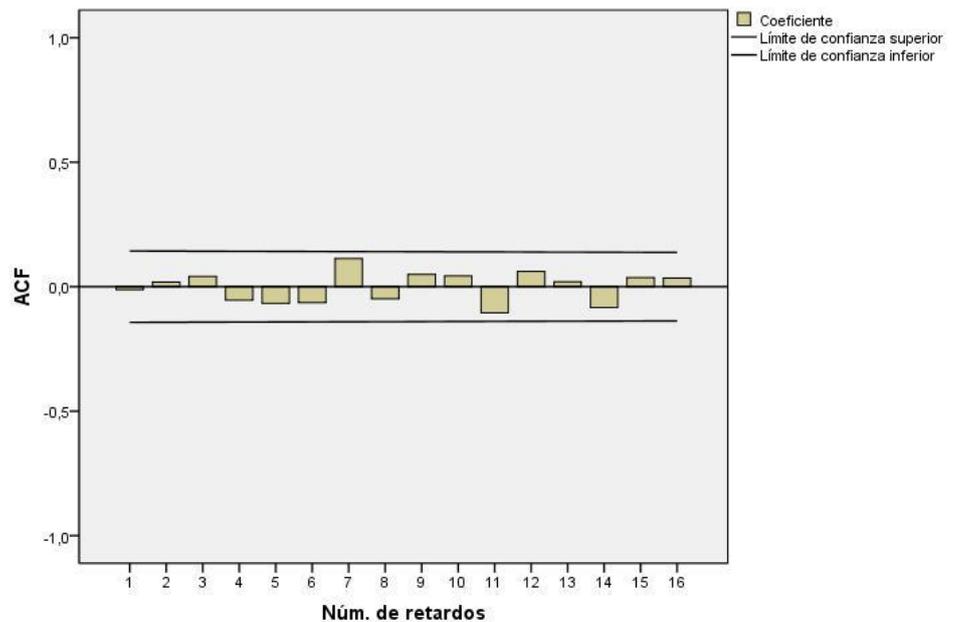
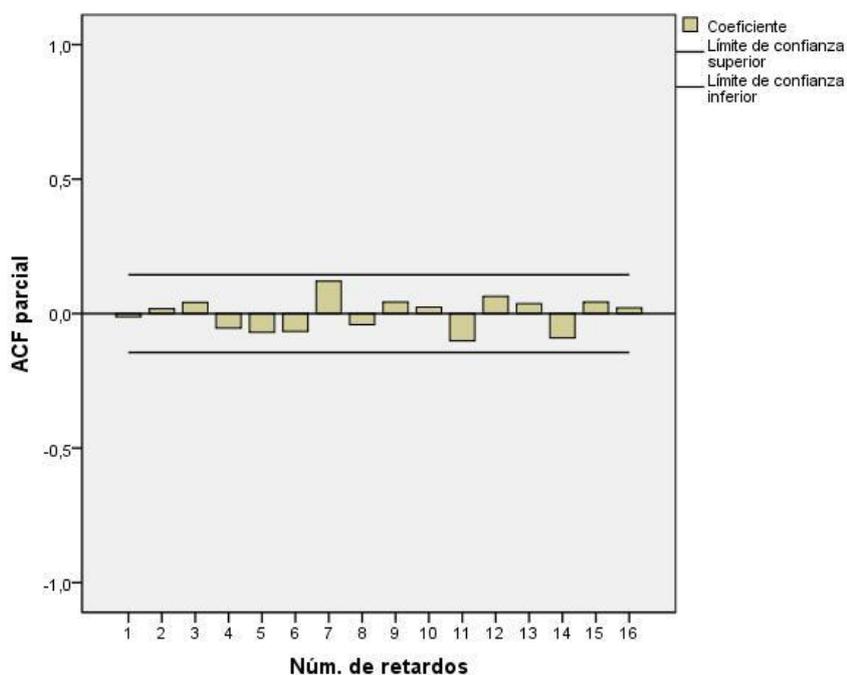


Figura N° 14: Autocorrelaciones Parciales de los residuos del Modelo SARIMA (3, 0, 1) (1, 0, 1)



El modelo de serie temporal utilizado es el SARIMA (3, 0, 1) (1, 0, 1) seleccionado luego de realizar un análisis de las autocorrelaciones parciales de los residuos, considerando como series independientes: el “Factor Macroeconómico” y el “Impulso Oct 2008”. Esta última variable, dicotómica, fue utilizada por los acontecimientos suscitados en el 2009 los cuales ya fueron expuestos anteriormente.

Cuadro N° 4: Estimaciones del Modelo SARIMA (3, 0, 1) (1, 0, 1)

		Estimaciones	Error típico	t	Sig. aprox.
Retardos no estacionales	AR1	,180	,165	1,086	,279
	AR2	,335	,120	2,793	,006
	AR3	,453	,078	5,828	,000
	MA1	-,313	,182	-1,721	,087
Retardos estacionales	Seasonal AR1	,886	,090	9,845	,000
	Seasonal MA1	,567	,144	3,922	,000
Coeficientes de regresión	REGR factor score 1 for analysis 1	327,523	95,142	3,442	,001
	Impulso Oct 2008	-396,865	73,363	-5,410	,000
Constante		1790,803	887,215	2,018	,045

Se ha utilizado el algoritmo de Melard para la estimación.

En el gráfico de pronósticos la estimación E (Figura A.7 del Anexo 2) corresponde al modelo SARIMA (3, 0, 1) x (1, 0, 1) y como se aprecia es el modelo que se ajusta mejor a la exportación observada. Finalmente las exportaciones pueden ser proyectadas mediante el siguiente modelo de regresión múltiple:

$$X'_{(t)} = 1790.8 + 327.5 * F - 396.9 * I + E_{(t)}$$

$$E_{(t)} = 0.18 * E_{t-1} + 0.34 * E_{t-2} + 0.45 * E_{t-3} - 0.31 * a_{t-1} + 0.88 * E_{t-12} + a_t + 0.56 * a_{t-12} + \varepsilon_t$$

$$X_{(t)} = X'_{(t)} + E_{(t)}$$

donde:

X: exportaciones mensuales

F: factor macroeconómico

I: variable impulso

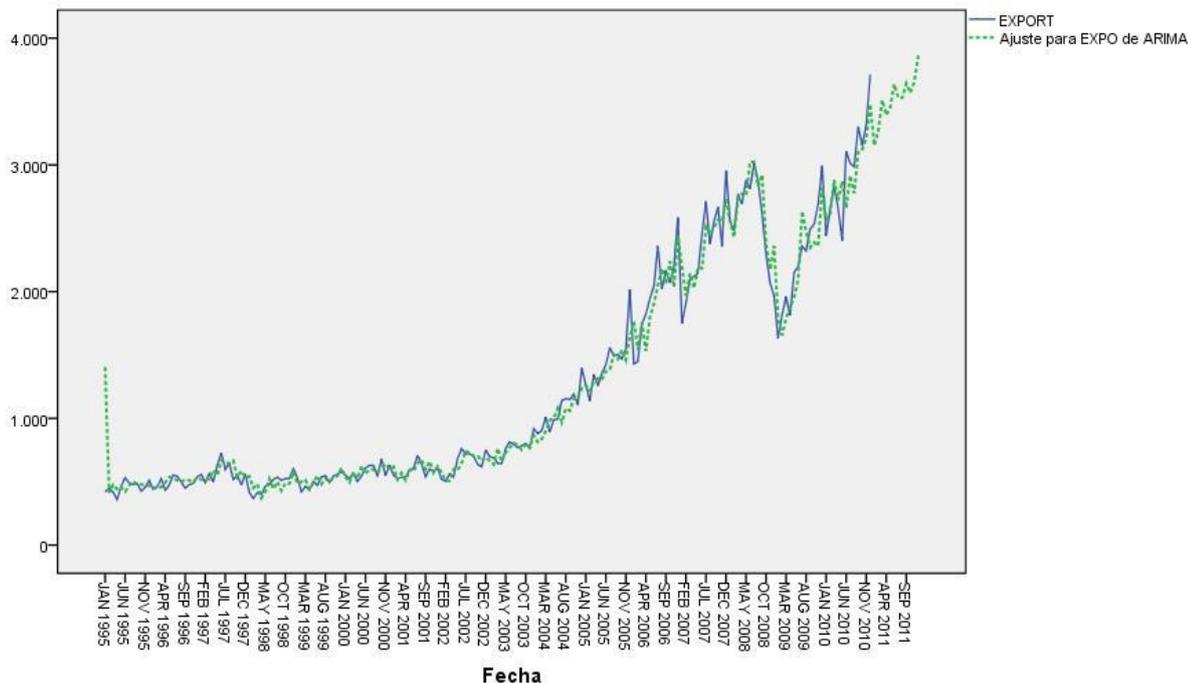
E: error de estimación del modelo inicial

a: ruido blanco

ε : error de estimación al modelar *E*

Las proyecciones utilizando el modelo SARIMA se presentan en el Figura N° 15 y se observa que las exportaciones estimadas se asemejan a las observadas.

Figura N° 15: Proyección de las Exportaciones del Modelo SARIMA (3, 0, 1) (1, 0, 1)



El análisis de residuos justifica el modelo SARIMA debido a que las autocorrelaciones totales y parciales deben estar dentro de las bandas de confianza para afirmar que los residuos están incorrelacionados, lo cual no ocurre utilizando sólo un Análisis de Regresión tal como se observa en las figuras N° 11 y N° 12, razón por la cual las proyecciones de las Exportaciones Definitivas mediante el modelo SARIMA (3, 0, 1)(1, 0, 1) si permitió obtener residuos incorrelacionados (Figura N° 13 y N° 14), asimismo, el haber utilizado una variable de intervención permitió obtener mejores proyecciones, obteniendo finalmente un coeficiente de determinación de 0.975 es decir las variables macroeconómicas explican el 97.5% de la variabilidad de las exportaciones en el Perú.

Cuadro N° 5: Cuadro comparativo de la Exportación y su Estimación

MES - AÑO	EXPORTACIÓN DEFINITIVA 1/	ANÁLISIS DE REGRESIÓN 2/	SARIMA (3,0,1)(1,0,1) 3/
Ene 2011	\$ 2.966	\$ 3.267	\$ 3.154
Feb 2011	\$ 3.367	\$ 3.066	\$ 3.266
Mar 2011	\$ 3.773	\$ 3.462	\$ 3.512
Abr 2011	\$ 3.442	\$ 3.801	\$ 3.393
May 2011	\$ 4.098	\$ 3.784	\$ 3.453
Jun 2011	\$ 4.212	\$ 3.591	\$ 3.638
Jul 2011	\$ 4.272	\$ 3.460	\$ 3.531
Ago 2011	\$ 4.642	\$ 3.655	\$ 3.529
Sep 2011	\$ 3.986	\$ 3.667	\$ 3.643
Oct 2011	\$ 3.828	\$ 3.660	\$ 3.569
Nov 2011	\$ 3.299	\$ 3.770	\$ 3.654
Dic 2011	\$ 4.385	\$ 3.819	\$ 3.863

NOTA:

1/ Exportación definitiva en millones de US\$.

2/ Proyección mediante el Análisis de Regresión Simple considerando sólo al factor macroeconómico como variable regresora.

3/ Proyección mediante el modelo SARIMA(3,0,1)(1,0,1) con residuos incorrelacionados.

CONCLUSIONES Y RECOMENDACIONES

En base a los resultados obtenidos se concluye que las estimaciones precisas descansan sobre cuatro pilares: variables macroeconómicas altamente correlacionadas con las exportaciones, factor macroeconómico altamente correlacionado con las exportaciones, análisis de intervención y modelo SARIMA.

El AFD, apoyándose en el modelo SARIMA y el Análisis de Intervención, permitió crear un modelo de proyección de las exportaciones en que los residuos se encuentren dentro de los límites establecidos; razón por la cual, se obtuvieron residuos incorrelacionados que coadyuvaron al cálculo de buenas proyecciones.

Como se había mencionado anteriormente el objetivo fundamental de este trabajo es el desarrollar un modelo que pueda proyectar adecuadamente las exportaciones con la mayor exactitud posible, lo cual se logró debido a que finalmente se obtuvo un modelo interesante el cual otorgaba mayor exactitud en la proyección de las exportaciones en el Perú explicándose su variabilidad en un 97.5%.

Este trabajo pretende dar una descripción del Análisis Factorial Dinámico desde una perspectiva macroeconómica; sin embargo, no hay una guía detallada que explique exactamente cómo se va a desarrollar dicho modelo ya que son muchas las situaciones que podrían suscitarse. Por ejemplo, para este trabajo, debido a que se utilizaron variables de comercio exterior y la crisis mundial del 2009 forzó necesariamente al apoyo del Análisis de Intervención que permite aislar aquellos datos atípicos y a la vez exógenos que influyen en el comportamiento de una determinada variable, en este caso, el Análisis de Intervención permitió controlar el comportamiento atípico de algunos meses del 2009 en que se tuvo una caída significativa de las exportaciones en el Perú.

Existen algunas consideraciones relevantes que podrían ser tratadas con más detenimiento, como por ejemplo enfocar la investigación a algún sector de exportación ya sea tradicional o no tradicional y que a su vez tenga una participación considerable sobre el total de exportaciones.

Un punto muy importante que debe considerarse es el conocer en un buen grado la naturaleza y posibles factores exógenos de la variable en análisis, si no se tenía conocimiento de la crisis mundial y de su incidencia sobre el comercio exterior peruano, no se habría utilizado también el Análisis de Intervención. Es decir, debe existir un buen conocimiento del fenómeno o variable en análisis, de ser así, las proyecciones calculadas tendrán un menor margen de error.

BIBLIOGRAFÍA

- [1]. ADAME SALVADOR, RAMÍREZ MARTHA E. (2005), Análisis de Intervención de series de tiempo y su aplicación a la producción de sedimentos, en la cuenca del río Texcoco – México.
- [2]. AGUIRRE ALVARO; CÉSPEDES LUIS FELIPE. (2004), Uso del Análisis Factorial Dinámico para Proyecciones Macroeconómicas.
- [3]. AMAYA L. JEANNETTE L.; PACHECO D. PEDRO N. (2002) Análisis Factorial Dinámico mediante el método de Tucker3.
- [4]. BELLAGAMBA ALICIA; FURTADO MAGDALENA; NALBARTE LAURA. (2000), Descripción y Análisis desde la perspectiva del Análisis Factorial Dinámico.- El Mercado de trabajo Uruguayo.
- [5]. BLANCO JORGE, CASTRILLEJO ANDRÉS, NALBARTE LAURA. (2002), Variabilidad en la clasificación de las trayectorias temporales diferenciales en el análisis factorial dinámico, según diferentes aproximaciones, jerárquicas y fuzzy.
- [6]. DURÁN VÍQUEZ RODOLFO; ANA CECILIA KIKUT VALVERDE (2006); Pronóstico de Inflación mediante el uso de Análisis factorial.
- [7]. FEDERICI ALESSANDRO. (199-); L' Analisi Fattoriale Dinamica in STATA; Dipartimento di Scienze Economiche.- Università di Roma La Sapienza.

- [8]. GONZÁLEZ CASIMIRO MARÍA DEL PILAR (20--), Análisis de series temporales: Modelos ARIMA.
- [9]. JIMENEZ URIEL, PEIRÓ EZEQUIEL (2005), Introducción al Análisis de Series Temporales.
- [10]. KHOLODILIN KONSTANTIN A. (20--), Dynamic Factor Analysis as a Methodology of Business Cycle Research. (consulta el 21 de noviembre del 2012). Disponible en: <http://www.tesisenred.net/bitstream/handle/10803/4043/kk1de1.pdf?sequence=1>
- [11]. MAURICIO JOSÉ ALBERTO. (2005), Introducción al Análisis de Series Temporales. (consulta el 24 de noviembre del 2012). Disponible en: <http://www.ucm.es/info/ecocuan/jam/JAM-IAST-Libro.pdf>
- [12]. PEÑA DANIEL (2005), Análisis de Series Temporales.
- [13]. PONCELA PILAR, SENRA EVA (2003), A Two Factor Model to Combine US Inflation Forecasts. (consulta el 22 de noviembre del 2012). Disponible en: <http://www.alde.es/encuentros/anteriores/viieea/autor/es/S/125.pdf>
- [14]. SENRA DÍAZ EVA. (s.a), Análisis de Intervención. (consulta el 21 de noviembre del 2012). Disponible en: http://www2.uah.es/esd/ST/Apuntes_AI.pdf
- [15]. COPPI, Curso Taller en Brasil: Métodos de “Analisi Fattoriale Dinamica” (AFD).

ANEXO 1

La Exportación presenta mucha variabilidad, además de tener una distribución asimétrica positiva o también llamada con cola a la derecha. Además, las variables macroeconómicas seleccionadas para proyectar las Exportaciones también tienen distribución asimétrica con presencia de outliers para las variables Importación, Recaudación aduanera, Demanda interna e Ingresos tributarios.

Figura A.1: Gráfico de Cajas para todas las variables

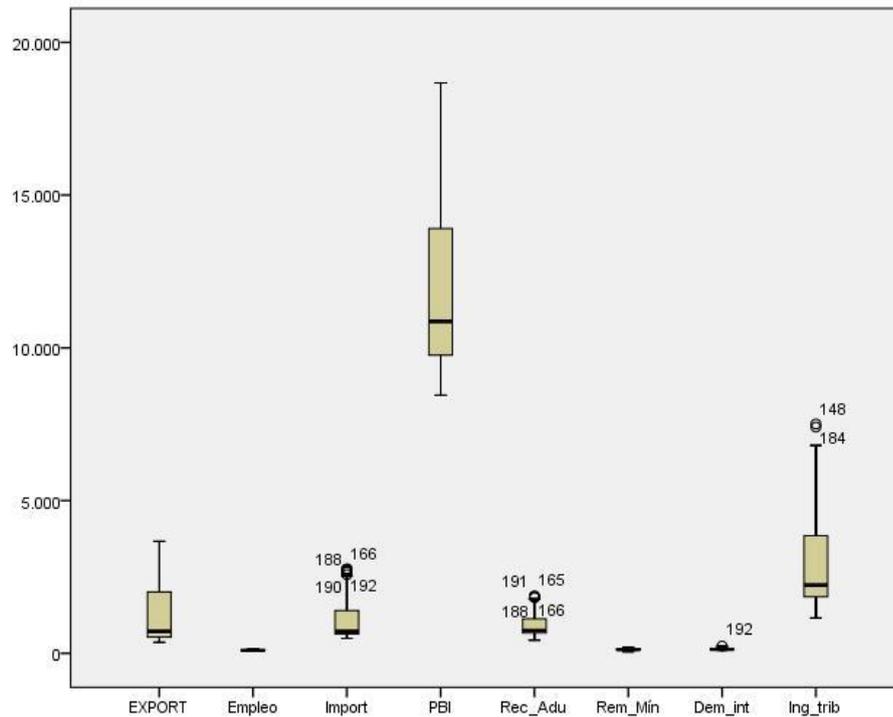
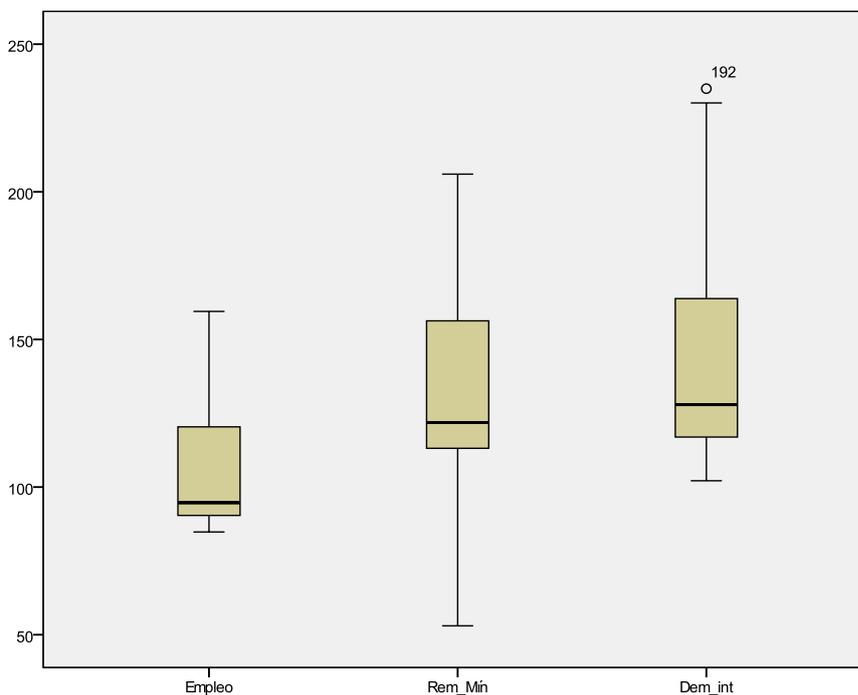


Figura A.2: Gráfico de Cajas de las variables con menor escala de medición



Se observa claramente que la variable Exportación presenta mucha variabilidad, al igual que las Importaciones, el PBI, la Recaudación aduanera y los Ingresos tributarios. Además se observa que todas las variables en conjunto presentan distribución asimétrica.

Cuadro A.1: Estadísticas de resumen de la exportación y las variables macroeconómicas

Variables	Media	Mediana	Desv. típ.	Mínimo	Máximo
EXPORT	1385,9686	768,9000	1066,63635	359,80	4641,50
Empleo	109,2760	95,6500	24,08417	84,80	167,30
Import	1180,7244	732,8200	776,17362	492,30	3457,14
PBI	12313,6327	11123,9180	3102,75011	8448,18	20055,21
Rec_Adu	956,1966	788,3500	387,49335	428,90	2074,60
Rem_Mín	135,5240	126,9350	44,82653	53,01	250,35
Dem_int	147,8723	129,6003	39,49447	102,15	254,06
Ing_trib	3022,3407	2286,0718	1551,47784	1156,38	8909,23

ANEXO 2

Cuadro A.2: Inversa de la Matriz de correlaciones

	Empleo	Import	PBI	Rec_Adu	Rem_Mín	Dem_int	Ing_trib
Empleo	36,628	-16,655	-19,762	2,799	-2,894	,619	-,905
Import	-16,655	36,260	31,059	-16,953	2,227	-34,422	-,984
PBI	-19,762	31,059	96,367	-12,471	-9,831	-81,658	-3,799
Rec_Adu	2,799	-16,953	-12,471	22,713	-4,820	10,339	-1,721
Rem_Mín	-2,894	2,227	-9,831	-4,820	9,765	6,700	-,631
Dem_int	,619	-34,422	-81,658	10,339	6,700	99,082	-,615
Ing_trib	-,905	-,984	-3,799	-1,721	-,631	-,615	9,022

Cuadro A.3: KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,884
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	3539,742
	gl	21
	Sig.	,000

La Medida de adecuación muestral de Kayser – Meyer – Olkin (KMO) compara las magnitudes de los coeficientes de correlación observados con las magnitudes de sus coeficientes de correlación parcial. Debido a que el KMO es de 0.884 se puede afirmar que es bueno aplicar el Análisis Factorial Dinámico.

La prueba de esfericidad de Bartlett se utiliza para verificar la hipótesis:

$$H_0 : \rho = I$$

$$H_1 : \rho \neq I$$

Rechazar esta hipótesis significa que es pertinente aplicar el Análisis Factorial, en caso contrario habría que optar por otro método multivariado. En el Cuadro N° 6 se observa que el $p - \text{valor} = 0.000 < 0.005$ por lo tanto rechazamos la hipótesis nula y se puede afirmar que la matriz de correlaciones es diferente de la matriz identidad, lo cual implica que existe correlación entre las variables macroeconómicas.

Cuadro A.4: Matrices anti-imagen

		Empleo	Import	PBI	Rec_Adu	Rem_Min	Dem_int	Ing_trib
Covarianza anti-imagen	Empleo	,027	-,013	-,006	,003	-,008	,000	-,003
	Import	-,013	,028	,009	-,021	,006	-,010	-,003
	PBI	-,006	,009	,010	-,006	-,010	-,009	-,004
	Rec_Adu	,003	-,021	-,006	,044	-,022	,005	-,008
	Rem_Min	-,008	,006	-,010	-,022	,102	,007	-,007
	Dem_int	,000	-,010	-,009	,005	,007	,010	-,001
	Ing_trib	-,003	-,003	-,004	-,008	-,007	-,001	,111
Correlación anti-imagen	Empleo	,939 ^a	-,457	-,333	,097	-,153	,010	-,050
	Import	-,457	,818 ^a	,525	-,591	,118	-,574	-,054
	PBI	-,333	,525	,811 ^a	-,267	-,320	-,836	-,129
	Rec_Adu	,097	-,591	-,267	,900 ^a	-,324	,218	-,120
	Rem_Min	-,153	,118	-,320	-,324	,944 ^a	,215	-,067
	Dem_int	,010	-,574	-,836	,218	,215	,830 ^a	-,021
	Ing_trib	-,050	-,054	-,129	-,120	-,067	-,021	,992 ^a

a. Medida de adecuación muestral

La Correlación Anti-imagen, es el negativo del coeficiente de correlación parcial, el objetivo es que la mayoría de las correlaciones anti-imagen sean pequeños para que la aplicación del análisis factorial dinámico sea válida. En el cuadro N° 7 la mayoría de las correlaciones anti-imagen son pequeñas.

Las Medidas de Adecuación de la Muestra (MSA), se calculan para cada variable y se cumple que coeficientes MSA bajos (ceranos a cero), indican que no es conveniente aplicar el análisis factorial. Del cuadro mostrado se observa que los coeficientes MSA son altos (ceranos a 1) lo cual especifica que es válido aplicar el Análisis Factorial Dinámico.

Cuadro A.5: Comunalidades

	Inicial	Extracción
Empleo	1,000	,974
Import	1,000	,948
PBI	1,000	,975
Rec_Adu	1,000	,956
Rem_Min	1,000	,896
Dem_int	1,000	,975
Ing_trib	1,000	,916

Método de extracción: Análisis de Componentes principales.

Evidentemente la comunalidad nos indica el porcentaje de variabilidad de cada variable explicada por el factor común. Del cuadro N° 8 se confirma que cada variable macroeconómica aporta al **factor macroeconómico** de manera significativa.

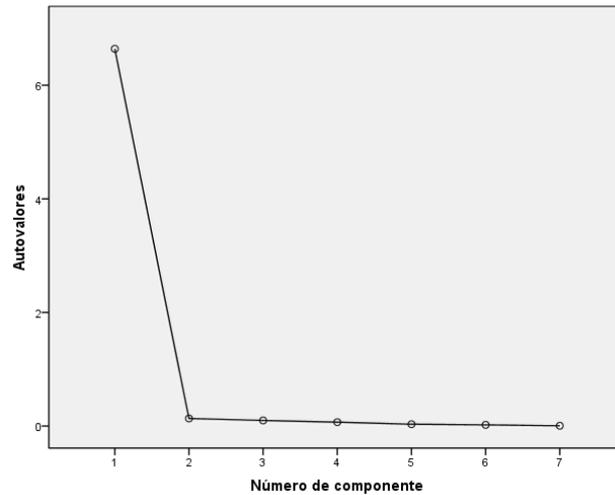
Cuadro A.6: Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,641	94,869	94,869	6,641	94,869	94,869
2	,134	1,908	96,777			
3	,099	1,410	98,187			
4	,069	,986	99,173			
5	,032	,459	99,632			
6	,021	,296	99,928			
7	,005	,072	100,000			

Método de extracción: Análisis de Componentes principales.

El **factor macroeconómico** explica en buena proporción (94.8%) el total de la varianza para las 7 variables macroeconómicas. Lo cual es otro indicador de que un solo factor (el factor macroeconómico) es suficiente para explicar a las 7 variables macroeconómicas.

Figura A.3: Gráfico de sedimentación



El gráfico de sedimentación es contundente para especificar que se debe extraer un solo factor (denominado **factor macroeconómico**) debido a que el punto de inflexión se observa claramente en un solo factor.

Cuadro A.7: Matriz de componentes

	Componente
	1
Empleo	,987
Import	,974
PBI	,988
Rec_Adu	,978
Rem_Mín	,946
Dem_int	,988
Ing_trib	,957

Método de extracción:
Análisis de componentes
principales.

La matriz de componentes especifica la solución factorial propiamente dicha, es decir, contiene las correlaciones entre las variables originales (o saturaciones) y cada uno de los factores. Del cuadro N° 10 es evidente que todas las variables macroeconómicas están altamente correlacionadas con un sólo factor.

Cuadro A.8: Correlaciones reproducidas

		Empleo	Import	PBI	Rec_Adu	Rem_Min	Dem_int	Ing_trib
Correlación reproducida	Empleo	,974 ^a	,961	,975	,965	,934	,975	,944
	Import	,961	,948 ^a	,962	,952	,922	,962	,932
	PBI	,975	,962	,975 ^a	,966	,935	,975	,945
	Rec_Adu	,965	,952	,966	,956 ^a	,925	,966	,936
	Rem_Min	,934	,922	,935	,925	,896 ^a	,935	,906
	Dem_int	,975	,962	,975	,966	,935	,975 ^a	,945
	Ing_trib	,944	,932	,945	,936	,906	,945	,916 ^a
Residual ^b	Empleo		,006	,002	-,010	-,015	,004	-,014
	Import	,006		-,017	,011	-,040	,004	-,017
	PBI	,002	-,017		-,013	-,002	,014	-,009
	Rec_Adu	-,010	,011	-,013		-,004	-,013	-,014
	Rem_Min	-,015	-,040	-,002	-,004		-,024	-,017
	Dem_int	,004	,004	,014	-,013	-,024		-,011
	Ing_trib	-,014	-,017	-,009	-,014	-,017	-,011	

Método de extracción: Análisis de Componentes principales.

a. Comunalidades reproducidas

b. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 0 (,0%) residuales no redundantes con valores absolutos mayores que 0,05.

Las comunalidades reproducidas son altas para todas las variables macroeconómicas consideradas, ello debido a que sus coeficientes tienen valores muy altos, todos por encima de 0.9.

Cuadro A.9: Matriz de coeficientes para el cálculo de las puntuaciones en las componentes

	Componente
	1
Empleo	,149
Import	,147
PBI	,149
Rec_Adu	,147
Rem_Min	,143
Dem_int	,149
Ing_trib	,144

Método de extracción:
Análisis de componentes principales.
Puntuaciones de componentes.

Se observa que la matriz de coeficientes para el cálculo de las puntuaciones en la componente (factor macroeconómico) tiene valores muy similares.

Por todo lo especificado en el subcapítulo 3.2 se puede afirmar que las variables macroeconómicas Empleo, Importaciones, PBI Global, Recaudación Aduanera, Remuneración Mínima, Demanda interna e Ingresos tributarios se concentran en un solo factor denominado “**factor macroeconómico**” el cuál será la variable latente que explicará a la variable de interés para los pronósticos (Exportaciones mensuales en valor FOB).

Cuadro A.10: Cargas Factoriales

VARIABLE	FACTOR1	FACTOR2	FACTOR3	FACTOR4	UNIDAD
empleo	0,9862	-0,0256	-0,0259	-0,0309	0,0251
import	0,9733	-0,173	0,036	-0,0004	0,0215
pbi	0,9896	0,0807	-0,0882	0,0038	0,0064
rec_adu	0,9748	-0,0227	0,1123	0,0118	0,0365
rem_mn	0,9349	0,1515	0,0755	-0,0044	0,0972
dem_int	0,9895	-0,0315	-0,1121	0,0134	0,0071
ing_trib	0,9445	0,0268	0,0092	0,0069	0,1072

ANEXO 3

Una vez determinado el factor macroeconómico, se analiza que tanto (dicho factor) explica a la variable de interés “Exportaciones mensuales en valor FOB”, para ello se muestran los siguientes resultados:

Cuadro A.11: Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	,964 ^a	,929	,929	236,56089	,770

a. Variables predictoras: (Constante), REGR factor score 1 for analysis 1

En este cuadro resumen se observa que el coeficiente es elevado (0.929) es decir el “factor macroeconómico” explica la variabilidad de las Exportaciones mensuales en valor FOB en un 92.9%.

Cuadro A.12: Correlaciones aplicando Análisis de Regresión

		EXPORT	REGR factor score 1 for analysis 1
Correlación de Pearson	EXPORT	1,000	,964
	REGR factor score 1 for analysis 1	,964	1,000
Sig. (unilateral)	EXPORT	.	,000
	REGR factor score 1 for analysis 1	,000	.
N	EXPORT	192	192
	REGR factor score 1 for analysis 1	192	192

El factor macroeconómico sí explica el comportamiento de las exportaciones al realizar un modelo de regresión lineal simple se comprueba que el coeficiente de correlación de Pearson es elevado (0.964) lo cual especifica que las variables macroeconómicas Empleo total en Lima Metropolitana, Total Importaciones, PBI Global, Total Recaudación Aduanera, Remuneración Mínima Vital Nominal Mensual, Demanda interna e Ingresos tributarios concentradas en el “factor macroeconómico” realmente guardan relación directa con las Exportaciones mensuales en valor FOB.

Cuadro A.13: ANOVA

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1,398E8	1	1,398E8	2497,482	,000 ^a
	Residual	1,063E7	190	55961,053		
	Total	1,504E8	191			

a. Variables predictoras: (Constante), REGR factor score 1 for analysis 1

El p valor del Análisis de Varianza (ANOVA) es $0.000 < 0.05$. Es decir el modelo:

$$Exportaciones = (Factor\ Macroeconómico) + E$$

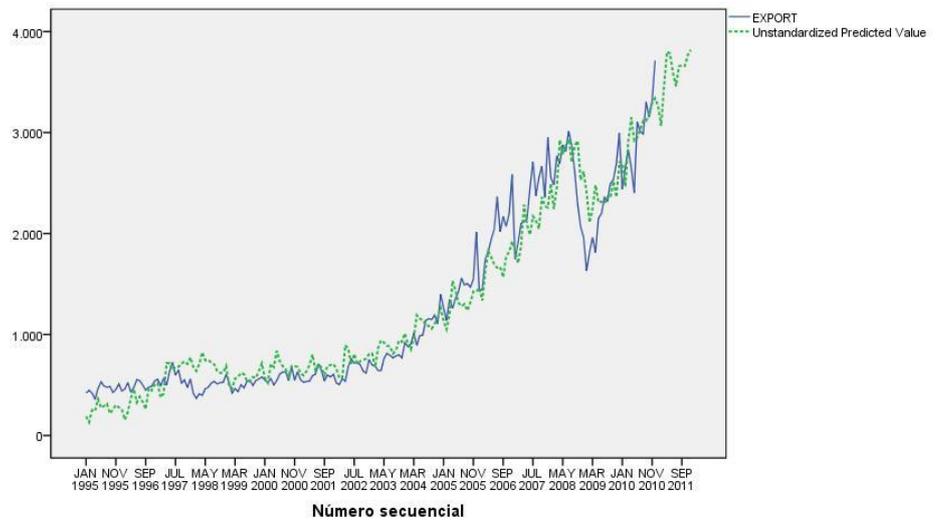
Donde E es el error de estimación, es significativo considerando un nivel de significación de 0.05.

Cuadro A.14: Tabla de Coeficientes

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.	Intervalo de confianza de 95,0% para B	
	B	Error típ.	Beta			Límite inferior	Límite superior
1 (Constante)	1369,952	17,295		79,210	,000	1335,836	1404,067
REGR factor score 1 for analysis 1	999,563	20,001	,964	49,975	,000	960,109	1039,016

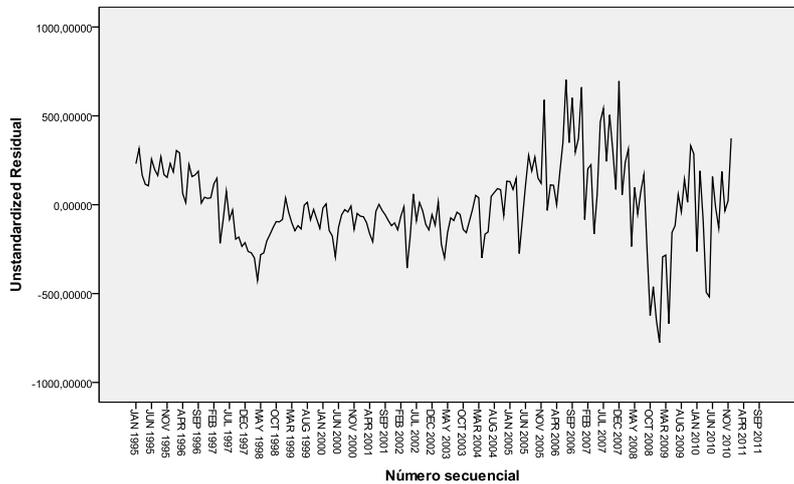
La tabla de coeficientes muestra que el “factor macroeconómico” y la constante son significativos para el modelo de regresión.

Figura A.4: Gráfico de secuencia Exportación Observada vs. Exportación Estimada



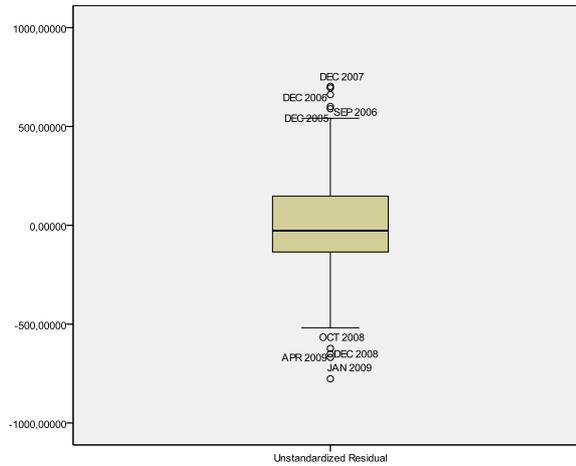
Las curvas para la variable Exportación observada (—————) y la variable Exportación estimada (- - - - -) no están muy superpuestas, es decir aplicar sólo un análisis de regresión simple no determina una buena precisión en las estimaciones, sin embargo ambas curvas tienen la misma tendencia.

Figura A.5: Residuos



La variabilidad de los residuos es considerable en el periodo Octubre 2008 – Junio 2009 en donde se encuentran las mayores diferencias entre la Exportación Observada y la Exportación estimada mediante el Análisis de Regresión.

Figura A.6: Gráfico de Cajas para los residuos



La distribución de los errores es casi simétrica pero con la presencia de algunos datos atípicos. Influenciado principalmente por un acontecimiento exógeno, como se mencionó en el inicio del capítulo III (la caída de las más importantes Bolsas de valores del mundo).

Cuadro A.15: Autocorrelaciones de los errores por Regresión

Autocorrelaciones

Serie:Unstandardized Residual

Retardo	Autocorrelación	Típ. Error ^a	Estadístico de Box-Ljung		
			Valor	gl	Sig. ^b
1	,606	,072	71,625	1	,000
2	,521	,071	124,816	2	,000
3	,562	,071	187,044	3	,000
4	,366	,071	213,584	4	,000
5	,367	,071	240,344	5	,000
6	,366	,071	267,158	6	,000
7	,275	,070	282,340	7	,000
8	,158	,070	287,376	8	,000
9	,238	,070	298,912	9	,000
10	,156	,070	303,865	10	,000
11	,185	,070	310,928	11	,000
12	,358	,070	337,407	12	,000
13	,138	,069	341,397	13	,000
14	,097	,069	343,365	14	,000
15	,114	,069	346,089	15	,000
16	,025	,069	346,222	16	,000

a. El proceso subyacente asumido es la independencia (ruido blanco).

b. Basado en la aproximación chi cuadrado asintótica.

Cuadro A.16: Autocorrelaciones Parciales de los errores por Regresión

Autocorrelaciones parciales

Serie:Unstandardized Residual

Retardo	Autocorrelación parcial	Típ. Error
1	,606	,072
2	,243	,072
3	,295	,072
4	-,159	,072
5	,091	,072
6	,042	,072
7	,009	,072
8	-,215	,072
9	,185	,072
10	-,078	,072
11	,214	,072
12	,215	,072
13	-,261	,072
14	-,155	,072
15	-,093	,072
16	,046	,072

Cuadro A.17: Autocorrelaciones del modelo SARIMA (3, 0, 1) (1, 0, 1)

Serie>Error para EXPO de ARIMA, MOD_49, CON

Retardo	Autocorrelación	Típ. Error ^a	Estadístico de Box-Ljung		
			Valor	gl	Sig. ^b
1	-,012	,072	,028	1	,868
2	,018	,071	,092	2	,955
3	,041	,071	,426	3	,935
4	-,054	,071	1,011	4	,908
5	-,067	,071	1,895	5	,863
6	-,064	,071	2,721	6	,843
7	,113	,070	5,301	7	,623
8	-,048	,070	5,777	8	,672
9	,049	,070	6,274	9	,712
10	,044	,070	6,664	10	,757
11	-,105	,070	8,935	11	,628
12	,061	,070	9,716	12	,641
13	,020	,069	9,798	13	,710
14	-,083	,069	11,250	14	,666
15	,036	,069	11,528	15	,714
16	,035	,069	11,785	16	,759

a. El proceso subyacente asumido es la independencia (ruido blanco).

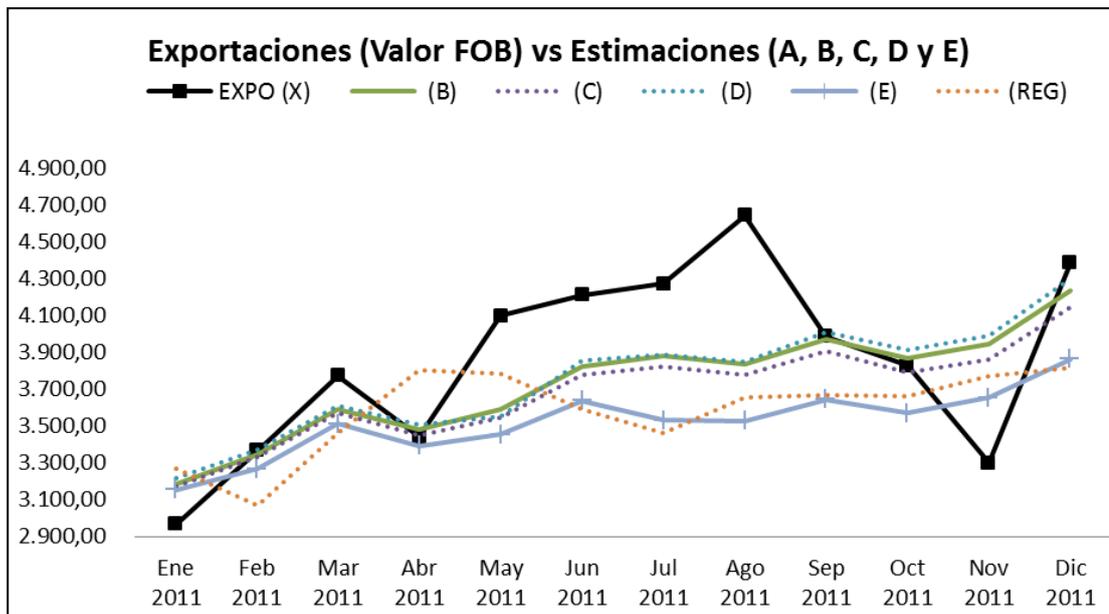
b. Basado en la aproximación chi cuadrado asintótica.

Cuadro A.18: Autocorrelaciones Parciales del modelo SARIMA (3, 0, 1) (1, 0, 1)

Serie:Error para EXPO de ARIMA,
MOD_49, CON

Retardo	Autocorrelación parcial	Típ. Error
1	-,012	,072
2	,018	,072
3	,042	,072
4	-,054	,072
5	-,070	,072
6	-,066	,072
7	,120	,072
8	-,041	,072
9	,043	,072
10	,024	,072
11	-,101	,072
12	,064	,072
13	,037	,072
14	-,091	,072
15	,043	,072
16	,021	,072

Figura A.7: Gráfico de Pronósticos



Donde:

X: Montos mensuales reales de la Exportación definitiva.

B: Proyecciones del modelo SARIMA (3,1,0)(0,1,1) considerando constante.

C: Proyecciones del modelo SARIMA (3,1,0)(0,1,1) sin constante.

D: Proyecciones del modelo SARIMA (3,1,1)(1,1,1) considerando constante.

E: Proyecciones del modelo SARIMA (3, 0, 1)(1, 0, 1) considerando constante.

REG: Proyecciones de la aplicación del Análisis de Regresión Simple considerando como variable regresora sólo el factor macroeconómico.

Como se detalló en el capítulo III el modelo SARIMA (3, 0, 1)(1, 0, 1) es el seleccionado para la proyección de las exportaciones mensuales en el Perú mediante el Análisis Factorial Dinámico, luego de realizar un análisis de las autocorrelaciones parciales y totales de los residuos, considerando como series independientes: el “Factor Macroeconómico” y el “Impulso Oct 2008”.

Cuadro A.19: Resumen del modelo SARIMA (3, 0, 1) (1, 0, 1)

Resumen del modelo ^b					
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	,987 ^a	,975	,975	141,47900	1,771

a. Variables predictoras: (Constante), Ajuste para EXPO de ARIMA, MOD_1, CON

b. Variable dependiente: EXPORT

Se utilizó como variable dependiente a la Exportación y como variable independiente la exportación ajustada mediante el modelo SARIMA.