77

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

ANALISIS EXPLORATORIO DE LA TASA DE DESEMPLEO ABIERTO VIA TABLAS DE DOBLE ENTRADA

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

CESAR MORENO GRANILLO

DIRECTOR DE TESIS: M. EN C. JOSE ANTONIO FLORES DIAZ

2002

VNIVERSIDAD NACIONAL
AVFNOMA DE
MEXICO

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

ANÁLISIS EXPLORATORIO DE LA TASA DE DESEMPLEO ABIERTO VÍA TABLAS DE DOBLE ENTRADA

realizado por **CÉSAR MORENO GRANILLO**

con número de cuenta **8637687-1**, quien cubrió los créditos de la carrera de **Actuaría**

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	**M. en C. José Antonio Flores Díaz**
Propietario	**M. en A.P. Ma. del Pilar Alonso Reyes**
Propietario	**Act. María Guadalupe Tzintzun Cervantes**
Suplente	**Act. Jaime Vázquez Alamilla**
Suplente	**Mat. Adrián Girard Islas**

Consejo Departamental de Matemáticas

M. en C. José Antonio Flores Díaz

FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMATICAS

A Adriana, fuente permanente de amor, apoyo y comprensión.

A mi extensa y maravillosa familia.

A mis suegros y todos mis cuñados con cariño.

A mis amigos de siempre, con quienes he compartido momentos importantes.

A mis amigos de Colgate, con la esperanza de que nuestra amistad se fortalezca con el tiempo.

Mi más sincero agradecimiento al Dr. Jaime Curts, quien me propuso el tema de este trabajo y que con paciencia lo dirigió en su etapa inicial.

A mi director de tesis, M. en C. José Antonio Flores por sus valiosos comentarios y el tiempo que dedicó a la revisión y corrección de este trabajo.

A ti Señor, por haberme permitido concluir esta etapa de mi vida profesional.

Análisis exploratorio de la tasa de desempleo abierto vía tablas de doble entrada

ÍNDICE

INTRODUCCIÓN

Una de las inquietudes que motivó escribir el presente documento fue el poder mostrar un enfoque de análisis estadístico de datos muy versátil, de fácil aplicación y relativamente nuevo: el Análisis Exploratorio de Datos (AED). Fue apenas en el año de 1977 cuando John W. Tukey publicó su libro titulado "Exploratory Data Analysis", en el cual aparecen de manera formal por vez primera muchas de las técnicas que se usan en este nuevo enfoque. De algunas otras, apenas bosquejadas entonces, se terminaría de formalizar su desarrollo en el transcurso de los siguientes ocho años.

A pesar de que ninguna de las técnicas del análisis exploratorio involucran procesos de inferencia estadística, éstas se pueden aplicar en una amplia variedad de situaciones. Poseen además, ciertas propiedades que las convierten en valiosas herramientas de análisis. Tal es el caso de las técnicas descriptivas, las cuales permiten representar el comportamiento de los datos de manera

mucho más fidedigna de lo que consiguen hacer sus contrapartes tradicionales cuando hay presencia de observaciones atípicas.

Existe otro tipo de herramientas que amplían la capacidad del investigador para comprender el comportamiento de sus datos, lo que a su vez le permite afinar sus hipótesis o incluso poder crear algunas nuevas, aún antes de aplicar cualquier técnica convencional de análisis de datos. El pulido de medianas (PM), es una de tales herramientas. El PM sirve para modelar el comportamiento de conjuntos de datos que puedan ser organizados en forma de tablas de doble entrada.

Sin embargo, pese a su valiosa utilidad es de sorprender que el uso del AED esté muy poco difundido en México. Esto resulta un tanto injusto ya que posee características que en ocasiones le llegan a colocar en ventaja frente a las técnicas clásicas de análisis de datos, siempre y cuando no estén involucrados procesos de inferencia.

Entre ellas se pueden mencionar:

1. En el AED, los residuos arrojados por cada modelo se someten a un minucioso análisis en la búsqueda de patrones que pongan en evidencia alguna debilidad en la calidad del ajuste.
2. La mayoría de sus aplicaciones no requieren que los datos deban cumplir con supuestos que con frecuencia no se satisfacen en la práctica.
3. Las herramientas exploratorias son altamente insensibles a la presencia de observaciones con comportamiento atípico.[1]

[1] Esta característica hace referencia a la propiedad de resistencia. Julian Besag (1981), la define como la cualidad que tienen ciertos estimadores y técnicas de análisis para proporcionar resultados que son, en el peor de los casos, ligeramente afectados por observaciones que no coinciden con el patrón general de los datos.

Por otro lado, al buscar información que pudiera usarse para mostrar algunas de las aplicaciones del AED, se detectó una evidente escasez de estudios donde se analizara el desempleo desde un punto de vista estadístico más que teórico.

Siendo el desempleo uno de los más sensibles problemas que México debe afrontar en la actualidad, se esperaba encontrar una gran variedad de estudios donde se le analizara desde distintas perspectivas. Sin embargo esto no fue así. Encontrar estudios con el enfoque deseado no fue fácil, lo cual reforzó la idea de elegirlo como objeto de análisis.

Es conveniente advertir que nunca se pretendió hacer de este documento un tratado sobre el desempleo, por ello aquí no se presenta ninguna teoría sobre este fenómeno. De igual forma, siempre estuvo fuera del alcance y de los objetivos de este trabajo el desarrollar algún modelo económico que pretendiera explicarlo, o evaluar las políticas gubernamentales que han sido implementadas para su control.

Existen muchas definiciones de desempleo, pero para este documento se decidió usar la que fue propuesta por la Organización Internacional del Trabajo, cuyo indicador, la tasa de desempleo abierto (TDA), es el más ampliamente usado y difundido por el gobierno de México.

En este trabajo se analiza el comportamiento de la TDA a través del tiempo con cuatro enfoques diferentes. Dos de ellos corresponden a trabajos previos en donde se usaron técnicas clásicas de análisis y cuyos resultados se comparan con los que se obtienen de la aplicación del PM.

Los objetivos que se definieron para el presente trabajo son los siguientes.

Objetivo General:

Presentar algunas de las técnicas exploratorias, en particular de aquellas usadas en el análisis de tablas de doble entrada, mostrando su aplicación en el estudio del comportamiento de la tasa de desempleo abierto en México.

Objetivos Específicos:

a) Presentar el algoritmo del PM, que es una técnica para el análisis de tablas de doble entrada con tres variables.

b) Mostrar su uso para ajustar modelos de tipo aditivo y no aditivo a este tipo de arreglos de datos.

c) Comentar sobre sus ventajas y desventajas en comparación con el análisis basado en medias.

d) Ejemplificar su aplicación y verificar la consistencia de los hallazgos que se obtienen al analizar datos usando el PM con los de las técnicas clásicas.

El presente documento está organizado en cuatro capítulos. En el primero se introducen los conceptos de empleo y desempleo, así como los distintos indicadores que se han desarrollado para medirlos. Asimismo, se presenta una síntesis de los resultados principales de dos estudios sobre la TDA de reciente desarrollo.

En el segundo capítulo se presenta la parte teórica del trabajo. Ahí se exponen la definición de una tabla de doble entrada; los algoritmos del pulido de medianas y de otras herramientas exploratorias; así como los distintos modelos que sirven para describir la estructura de este tipo de arreglo de datos.

En el capítulo tercero se ejemplifica el uso del pulido de medianas y de otras técnicas exploratorias en el análisis de la TDA, siguiendo para ello cuatro aproximaciones diferentes. Finalmente, en el cuarto capítulo se presentan las conclusiones y recomendaciones de este trabajo.

En este punto conviene señalar que son pocos los paquetes estadísticos que tienen incorporada una rutina de pulido de medianas. Con certeza sólo puede mencionarse al Minitab. Sin embargo, debido a que su alto costo lo hace de uso restringido, se decidió desarrollar un programa en lenguaje C++ para ejecutar el algoritmo, cuyo programa fuente se presenta en el anexo C.

CAPÍTULO 1

La tasa de desempleo abierto (TDA)

1.1 Los conceptos de empleo y desempleo

Un fenómeno que sin duda va a caracterizar al pasado siglo XX es el de la globalización. Nunca antes se había presentado evento alguno que de una u otra forma afectara en la vida de la gente de todas las regiones del planeta. Nunca antes también, la relación entre los países había sido tan estrecha. En la actualidad es cada vez menos factible para un país mantenerse al margen de lo que ocurre fuera de las propias fronteras, especialmente en materia económica.

En la época actual, las economías de países geográficamente distantes dependen entre sí en niveles nunca antes vistos. La gran crisis mexicana que inició a finales de 1994 y que arrastró consigo a economías de países de todos los continentes en el llamado "efecto tequila", o la crisis que sacudió a los países del sureste asiático a finales de la década de los noventa son un par de ejemplos que inmediatamente acuden a la mente. Otra expresión de este fenómeno

globalizador fue la formación de bloques de países como el de la Unión Europea o el establecimiento de tratados comerciales regionales como el tratado de libre comercio para América del norte.

Es en este contexto donde se puede apreciar la gran relevancia que tiene el disponer de información que sea comparable no sólo entre países, sino incluso, entre regiones distantes. Pero para que esto ocurra se requiere que dicha información sea generada usando criterios uniformes. Es por ello que para analizar cualquier fenómeno ahora resulta indispensable que en todos lados se usen los mismos conceptos e indicadores.

El desempleo como medida clave para evaluar el desempeño de una economía no podía ser diferente. Así lo entendieron un grupo de países que se unieron para conformar la Organización Internacional del Trabajo (OIT), cuya función es estudiar al empleo y desempleo en todos sus aspectos. En la VIII conferencia de estadígrafos del trabajo de la OIT fueron desarrolladas las definiciones de empleo y desempleo actualmente en vigor. Para su XIII conferencia celebrada en octubre de 1982, ya se discutía la conveniencia de modificarlas dada su desigual efectividad para medir los niveles de desempleo en los países miembros de la organización.

Al final se acordó mantener la definición de *persona con empleo* para todo individuo en edad de trabajar que durante el periodo de referencia, el cual puede ser un día o una semana, estuviera en cualquiera de las siguientes categorías:

Con empleo asalariado. Aquellos sujetos que recibían un sueldo o salario en metálico o en especie y se dice que podían estar *a) trabajando*, si lo hicieron al menos por una hora durante todo el periodo de referencia; o *b) sin trabajar*, que son quienes manteniendo un vínculo con su actual empleo, dejaron de hacerlo de forma temporal.

Con empleo independiente. Quienes durante el periodo de referencia, podían estar *trabajando*: personas que hubieran realizado algún trabajo para obtener beneficios o ganancia familiar, en metálico o en especie. *Con una empresa pero sin trabajar* (fuera ésta industrial, comercial, de explotación agrícola o de prestación de servicios), que estaban temporalmente ausentes del trabajo por cualquier razón especifica.

En el documento se aclara que por "razones prácticas" se considera que una persona trabaja, si lo hizo al menos *por una hora* durante todo el periodo de referencia.[1] En ciertos casos ésta resulta ser una definición bastante laxa, la cual tiene implicaciones muy importantes. Más adelante se regresará a este punto.

Asimismo se define como *persona desempleada* a todo individuo en edad de trabajar que durante el periodo de referencia cumpliera con los siguientes criterios:

1. Estuviera *sin empleo* (asalariado o independiente);
2. *disponible para trabajar*, es decir, con disposición inmediata para incorporarse a un empleo; y
3. *en busca de empleo*. Entendiendo por ello que hubiera tomado medidas concretas para buscar un empleo en un periodo de tiempo específico reciente.

Todavía en la XIV conferencia internacional llevada a cabo en 1987, se ratificó el criterio de "al menos una hora de trabajo", con el propósito de "no afectar los sistemas internacionales de estadísticas sobre empleo y desempleo y el de las Cuentas Nacionales de la ONU"[2] entre otros.

[1] OIT. XIII Conferencia Internacional de Estadígrafos del Trabajo. "Resolución sobre estadísticas de la población económicamente activa, del empleo, del desempleo y del subempleo". [s.p.]. http://www.ilo.org/public/spanish/bureau/stat/res/ecacpop.thm

[2] OIT. XIV Conferencia Internacional de Estadígrafos del Trabajo. "Directrices sobre la incidencia de los programas de promoción del empleo, sobre la medición del empleo y del desempleo". [s.p.] http://www.ilo.org/public/spanish

1.2 La medición del desempleo en México

Siguiendo las recomendaciones de la OIT, el gobierno de México define como *población desocupada abierta* a "todas las personas de 12 años y más que no estando ocupadas, buscaron activamente incorporarse a alguna actividad económica en el último mes previo a la semana del levantamiento o hasta en los últimos dos meses, siempre y cuando estén disponibles a incorporarse de inmediato a un empleo".[3]

A diferencia de los países del llamado "primer mundo", en las naciones en desarrollo esta definición tiene intrínsecos ciertos problemas de consideración. Esto es porque presupone la existencia de mercados laborales muy desarrollados, donde todos, o por lo menos la mayoría de los puestos de trabajo son asalariados, bien remunerados y formales.

Debe considerarse también que en estos países no existe un seguro de desempleo y que el grueso de su población carece de ahorros. En estas condiciones una persona que pierde su empleo no puede permanecer económicamente inactiva por largos periodos y muy pronto se ve obligado a subemplearse o incorporarse a la economía informal.

No hay cifras oficiales pero según algunas estimaciones, el sector informal en México contribuye aproximadamente con el 50% del total de la población económicamente activa y con más del 60% del PIB del país.[4] Pero existe un elemento adicional. Todos los indicadores de desempleo son proyecciones elaboradas por el INEGI a partir de una encuesta que se levanta en las principales zonas urbanas. Es decir, en sentido estricto la información no tiene el carácter de nacional.

---

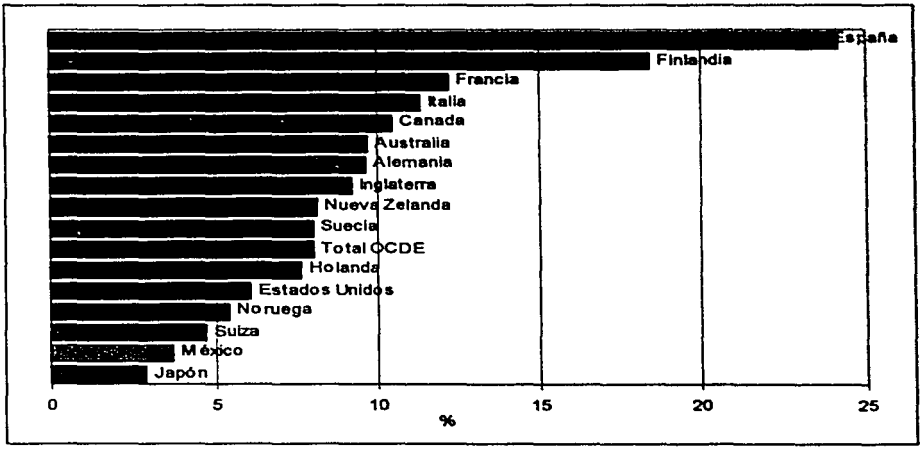
[3] INEGI. "Banco de información económica". [s.p.]. http://www.inegi.gob.mex

[4] Ibarra, Darío G. "Economía informal, saldo de la crisis". Macroeconomía. Año IV num 39, Octubre 1996.

Es evidente por lo tanto, que la TDA proporciona una medición del desempleo que no refleja la magnitud real del problema. Lo anterior se ha prestado a prácticas demagógicas –de forma deliberada o no- por parte del gobierno.

Por ejemplo, en el programa de empleo, capacitación y defensa de los derechos laborales, el cual se desprende del Plan Nacional de Desarrollo 1995-2000, se puede ver una gráfica (figura 1.1), donde se aprecia que en el año de 1994 México tuvo la tasa de desempleo abierto más baja de un selecto conjunto de países miembros de la Organización para la Cooperación y el Desarrollo Económico, OCDE.

Figura 1.1 TDA en países miembros de la OCDE en 1994



Esto se debe, se asienta en el documento, a que "la mayoría de los trabajadores desempleados no pueden enfrentar periodos largos de búsqueda de empleo...

por lo que pueden verse orillados a ingresar al sector informal o a trabajar jornadas menores a la máxima legal, aceptando remuneraciones relativamente bajas".[5]

A todas luces la comparación no es apropiada. Además no se aclara que a diferencia del resto de los países donde el indicador tiene una cobertura nacional, en el caso de México los resultados sólo eran representativos del comportamiento del fenómeno en 35 ciudades.

Esta situación resalta la necesidad de disponer de otros indicadores que ayuden a subsanar las insuficiencias de la definición oficial de desempleo.

1.3 Indicadores alternativos de desempleo en México

La OIT consciente de lo inoperante que resultó la definición de desempleo en ciertos países, en las conferencias XIV, XV y XVI de estadígrafos del trabajo emitió diversas recomendaciones para la evaluación del problema y desde entonces ha insistido en la necesidad de que los países calculen indicadores adicionales que ayuden a medir con mayor precisión la situación del empleo y desempleo.

Entre las últimas recomendaciones se encuentran las concernientes a la medición del subempleo por insuficiencia de horas laboradas; al empleo inadecuado; al problema del sector informal y al de los trabajadores estacionales.

Algunos de los indicadores adicionales que se calculan en México son los siguientes:

[5] México. Poder Ejecutivo Federal. Programa de empleo, capacitación y defensa de los derechos laborales 1995-2000. 1995. Pág. 34.

TPRG *Tasa de presión general.* Se refiere al porcentaje de la PEA que se encuentra desocupada y al de los ocupados que buscan trabajo con el propósito de cambiarse o tener un empleo adicional.

TDAA *Tasa de desempleo abierto alternativo.* Considera no sólo a los desocupados abiertos, sino también a la parte de la población económicamente inactiva que suspendió la búsqueda de empleo para realizar actividades del hogar o estudio, pero se encuentra disponible para aceptar un puesto de trabajo.

TOPD1 Tasa de desocupación abierta más ocupados que trabajan menos de 15 horas a la semana.

TOPD2 Tasa de desocupación abierta más ocupados que trabajan menos de 35 horas a la semana.

TCCO *Tasa de condiciones críticas de ocupación.* Porcentaje de la PEA que considera a los que trabajan menos de 35 horas a la semana por razones de mercado, que trabaja más de 35 horas a la semana y percibe ingresos mensuales inferiores al salario mínimo, o que trabajando más de 48 horas a la semana percibe ingresos inferiores a dos salarios mínimos.

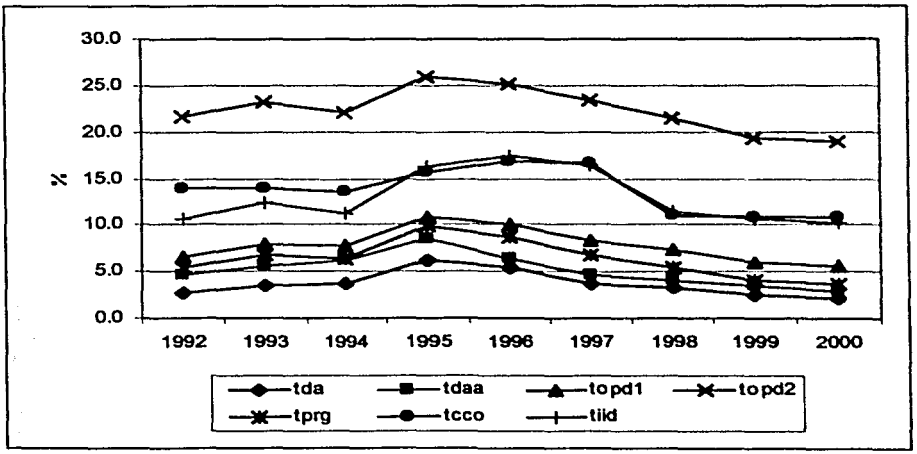
TIID *Tasa de ingresos insuficientes y desempleo.* Mide el porcentaje que representa la población desocupada y la ocupada que tiene ingresos inferiores al salario mínimo, respecto a la población económicamente activa.

En comparación con la TDA la mayor parte de estos indicadores complementarios representan una mejora en términos conceptuales, y arrojan

cifras que el sentido común indica están más acordes con la situación real del desempleo en el país. Sin embargo, en cada uno de ellos se pone énfasis en aspectos del problema muy específicos. Cuando fue necesario, en este trabajo se optó por usar la TIID ya que se considera que el nivel de percepciones, o su ausencia, es la característica más relevante del desempleo y subempleo.

En la figura 1.2 se muestra la tendencia del fenómeno del desempleo y del empleo en condiciones críticas. Los diferenciales en las tasas son enormes. De ahí la importancia de la correcta selección del indicador cuando se pretenda analizar la situación del problema en el país.

Figura 1.2 Tendencia de algunos indicadores del desempleo en México



1.4 La encuesta nacional de empleo urbano (ENEU)

El Instituto Nacional de Estadística Geografía e Informática (INEGI), es el organismo responsable de la medición continua de los niveles de empleo en México. Esto lo hace mediante el levantamiento mensual de una encuesta en hogares urbanos: la ENEU. En la actualidad, la encuesta se lleva a cabo en las 45 ciudades más importantes del país con lo que se cubre, según el Instituto, aproximadamente el 90% de la población que habita en localidades de 100,000 habitantes y más, lo cual equivale aproximadamente al 45% del total de la población.

El antecedente histórico de la ENEU se encuentra en la encuesta continua sobre ocupación (ECSO), la cual se empezó a levantar en 1973. Con la ECSO se dio inicio a la generación permanente de estadísticas sobre el empleo y el desempleo. Sin embargo, ni la cobertura conceptual de este estudio ni la geográfica ha sido siempre la misma.

En sus albores, la ECSO sólo disponía de información de las tres principales áreas metropolitanas del país: la Ciudad de México, Guadalajara y Monterrey. Para el año de 1983 ya proporcionaba información mensual de 16 áreas metropolitanas: Ciudad Juárez, Nuevo Laredo, Tijuana, Matamoros, Chihuahua, León, Mérida, Orizaba, Puebla, San Luis Potosí, Tampico, Torreón, Veracruz, además de las pioneras México, Guadalajara y Monterrey.

En 1985 el nombre de ECSO es sustituido por el actual de ENEU. En 1992 se incorporaron 18 ciudades más a la muestra; tres en 1993; dos en 1994; cuatro en 1996; una en 1998 y la última en 1999.

Finalmente en cuanto a la forma de selección de los hogares participantes en la encuesta, se tiene que ésta se hace bajo un esquema de muestreo polietápico y estratificado. Es decir, el procedimiento de selección de la muestra es por etapas, en la primera se eligen con probabilidad proporcional a su tamaño las áreas

geoestadísticas básicas (AGEB)[6], previamente estratificadas. Dentro de cada AGEB participante se seleccionan las unidades de segunda etapa que corresponden a una manzana o agrupamiento de manzanas, las cuales contienen a las viviendas que son las unidades de la tercera etapa.

Hasta este momento se ha abordado el tema de la definición del desempleo, de los distintos indicadores que se usan para medirlo, y del mecanismo empleado para obtener dicha información. En lo que resta del presente capítulo se bosquejan dos trabajos de tesis sobre la TDA, cuyos resultados serán comparados con los que se obtengan al aplicar la metodología de análisis de datos que aquí se propone.

1.5 Análisis de la TDA con una perspectiva estadística

Antes que nada se considera importante explicar la razón de insistir en analizar la TDA a pesar de lo que se ha comentado previamente. Las razones para ello son dos. En primer lugar está la disponibilidad de la información. La TDA es el indicador del que se cuenta con más historia; y la segunda, que está íntimamente relacionada con la primera, es que ese es el indicador que publica periódicamente el gobierno de México para evaluar la situación del problema en el país. Dicho lo anterior se procede a exponer los trabajos mencionados.

1.5.1 La convergencia de la TDA de México, Guadalajara y Monterrey

Este estudio corresponde al trabajo de investigación para tesis de licenciatura de la economista Jimena Roel.[7] Su objetivo era probar la hipótesis de que en el

[6] Las AGEB constituyen la unidad básica del marco geoestadístico del INEGI. Su perímetro está representado por calles, brechas o por rasgos físicos naturales y/o culturales inclusive, que son normalmente reconocibles y perdurables en el terreno.

[7] Roel Pavón, Jimena. "Consideraciones sobre la convergencia del desempleo en México". Inédita. México. Tesis presentada para aspirar al grado de licenciado en economía. Instituto Tecnológico Autónomo de México. 1995. 49 págs.

largo plazo las tasas de desempleo abierto de las ciudades de México, Monterrey y Guadalajara presentarán un mismo patrón de comportamiento.

En otras palabras, Roel pretendía probar que a partir de algún momento en el futuro, si la TDA de cualquiera de las tres ciudades presenta una tendencia creciente, las otras dos deberán crecer también; y/o que si la tendencia es decreciente o se mantiene estable, las otras deberán mostrar el mismo comportamiento. De esta forma es como Roel define el concepto de "convergencia" del desempleo.

Definiciones

Para entender este trabajo es necesario introducir algunos conceptos. El más importante es el de cointegración de variables económicas. Ésta es una idea que surgió hace menos de dos décadas, pero que ya se ha vuelto muy popular en los círculos de los econometristas. La cointegración exige la existencia de una *relación de equilibrio*[8] entre dos o más series económicas que sea *estacionaria*,[9] a pesar de que individualmente dichas variables no deban serlo. Si ello ocurre se dice entonces que las series están cointegradas.

El orden de integración de una serie es un concepto relacionado de manera muy cercana al de la cointegración, y se define como el número de veces (d) que debe diferenciarse una serie no estacionaria para alcanzar la estacionariedad en nivel y se denota por I(d). Por ejemplo, una serie no estacionaria es I(1) si al ser diferenciada una vez la serie de estas diferencias resulta ser estacionaria. Una serie estacionaria es por lo tanto integrada de orden cero I(0).

[8] Se define como una combinación lineal de las variables bajo estudio, que en el largo plazo converge a un valor al que tiende a regresar incluso después de la ocurrencia de perturbaciones en el sistema.
[9] Aquí se hace referencia a la estacionariedad de primer orden, la cual consiste en que la media de la serie nunca depende del tiempo.

De manera más formal. Sean dos series X_t y Y_t integradas de orden d > 0. Es decir son I(d). Si X_t y Y_t están cointegradas, entonces:

1. Existe una combinación lineal $Z_t = X_t + bY_t$, estacionaria, para alguna $b \neq 0$. Se dice que Z_t describe la relación de equilibrio entre ambas series.

2. Es posible describir esta relación por medio de un modelo que incluye un término residual estacionario, es decir I(0), el cual reflejará las desviaciones del equilibrio. Esto último es de gran relevancia ya que es precisamente lo que garantiza que las desviaciones sean temporales.

Objetivo del estudio y descripción de los datos

De lo anterior se desprende que el trabajo de Roel se circunscribió en demostrar que las tasas de desempleo abierto de las ciudades de Guadalajara, México y Monterrey están cointegradas. Los datos que utilizó para su análisis son las tasas trimestrales de estas tres ciudades, las cuales comprenden un periodo de 21 años, de1973 a 1993.

Metodología y conclusiones

Usando la prueba aumentada de Dicky-Fuller[10] comprobó que las tres series son no estacionarias I(1). Por medio de mínimos cuadrados ordinarios (MCO), estimó los residuales de todas las combinaciones de las tasas de las tres ciudades, las cuales tienen la forma:

- $TDA\_CD\_1_t = C + a_1(TDA\_CD\_2_t) + a_2(TDA\_CD\_3_t) + a_3(\text{tendencia}) + R_t$

[10] Para mayor información sobre la prueba y la metodología en general consultar: Dickey, D. A. y Fuller, W. A. "Likelihood ratio statistics for autoregressive time series with a unit root". Econometrica, 49. 1981. Págs. 1057-72. Engle, R. F. y Granger, C. W. J. "Cointegration and error correction: representation, estimation and testing". Econometrica, 55, 1987. Págs. 251-76.

Finalmente verificó la estacionariedad de la serie de los residuales con un rezago, usando de nuevo la prueba de Dicky-Fuller.

La conclusión del estudio es que sí existe una relación de equilibrio entre las tres series a la cual convergerán en el largo plazo. Sin embargo, se admite que no es posible determinar el momento en el que esto ocurrirá.

1.5.2 Tres modelos de series de tiempo para la TDA de doce ciudades

Este segundo estudio corresponde al trabajo de tesis de licenciatura de la actuaria Martha Morales.[11] En él su autora desarrolla algunos modelos de series de tiempo para estudiar el comportamiento de esta tasa.

Definición de conceptos

Morales afirma que en toda serie de tiempo se presentan dos tipos de fenómenos que afectan su comportamiento: los de tipo sistemático y los aleatorios; y que la combinación de ambos determinan el valor de la serie en cualquier punto en el tiempo.

Los efectos sistemáticos son aquellos que se presentan con cierta regularidad y que es factible predecir su ocurrencia futura. Éstos son la tendencia y las variaciones cíclica y estacional. De tipo aleatorio sólo se tiene a la variación irregular. Sus definiciones son las siguientes.

La tendencia (T) es el comportamiento que la serie presenta a lo largo de los años, la cual puede ser creciente, decreciente o constante. A los patrones de comportamiento repetitivos que se presentan en periodos de tiempo mayores a un año se les define como la variación cíclica (C) de la serie. La variación

[11] Morales Alvarez, Martha E. "El desempleo en el estado neoliberal: un análisis estadístico". Inédita. México. Tesis presentada para obtener el título de Actuario. Universidad Nacional Autónoma de México. 1998. 104 págs.

estacional (E) considera los patrones de comportamiento en periodos menores a un año. En términos generales éstos se presentan con regularidad en una misma época del año.

Finalmente, la variación irregular (I) de una serie se refiere a alteraciones en su comportamiento que se presentan sin ninguna periodicidad y son el resultado de la ocurrencia de eventos no predecibles.

Metodología y descripción de los datos

En su documento Morales expone una metodología para ajustar tres modelos diferentes de series de tiempo a los datos de cada una de las doce ciudades de manera individual. Dichos modelos son: Y=TxCxExI; Y=TxE con tendencia lineal y Y=TxE con tendencia parabólica.

La información usada en este análisis corresponde a la tasa trimestral de desempleo abierto de las ciudades de México, Chihuahua, Guadalajara, León, Mérida, Monterrey, Orizaba, Puebla, San Luis Potosí, Tampico, Torreón y Veracruz durante el período de 1983 a 1995.

Modelo Y=TxCxExI

El proceso de separación de las cuatro componentes inicia con el suavizamiento de la serie usando promedios móviles centrales. Para ello, primero se calcula la media aritmética de grupos de cuatro observaciones consecutivas; después se obtiene el promedio de cada par de valores obtenidos en el paso anterior. Con esto se supone eliminada del valor de la serie la aportación de las variaciones estacional e irregular.

La aportación de la componente de tendencia corresponde al valor esperado del conjunto de promedios móviles obtenidos en los pasos previos.

En otras palabras, T corresponde a la recta de regresión de los puntos TxC, ó $T = E(TxC) = B_0 + B_1t$ con t=tiempo (años). Una vez conocidos los valores se pueden obtener los de C mediante su despeje simple en la expresión T=TxC.

De igual forma se tiene que ExI = TxCxExI / TxC. El siguiente paso consiste en obtener el valor promedio de cada uno de los trimestres, el cual corresponde al de la aportación de la componente E.

Adicionalmente indica cómo calcular un índice estacional, el cual se obtiene al encontrar la E promedio de los cuatro trimestres multiplicados por un factor de ajuste, cuyo valor dependerá de la periodicidad de las cifras originales (doce si es mensual; seis si es bimestral, cuatro si es trimestral, etc.).

Modelo Y=TxE con tendencia lineal

En este modelo se ignoró la variación cíclica debido a que se juzgó que el periodo comprendido del estudio corresponde a un mismo ciclo, por lo que no aporta mayores elementos la estimación de C. Por ser consecuencia de eventos impredecibles tampoco se consideró la variación irregular en el modelo.

El procedimiento de ajuste es similar al expuesto previamente. La diferencia radica en la forma de estimar el efecto de la tendencia, que en principio será anual y después se hará un ejercicio para convertirlo a una base trimestral. Se empieza por calcular la suma de las cuatro tasas trimestrales que conforman cada año, luego se estima la recta de regresión para estos valores anuales, para después ajustarlos a trimestres. Los valores así generados se dividen entre los originales obteniéndose la E para cada uno de los periodos de la serie. El índice estacional se calcula de igual forma que en el modelo anterior.

Modelo Y=TxE con tendencia parabólica

El procedimiento es exactamente el mismo al usado para el modelo lineal, salvo que usando mínimos cuadrados se ajusta una curva de la forma $Y = a + bX + cX^2$, en lugar de una recta a la suma de las tasas trimestrales.

Conclusión

Se encontró que para el periodo estudiado, usando como criterio de evaluación al error cuadrático medio, el modelo que mejor se ajustó a los datos observados en la mayoría de las ciudades es el Y=TxE con tendencia parabólica. Sin embargo, para los ocho trimestres de los años 1996 y 1997, éste es también el modelo que proporciona las peores estimaciones debido precisamente a la naturaleza de la curva usada para el ajuste. Ver tabla siguiente.

Tabla 1.1 Error cuadrático medio de los tres modelos ajustados

Ciudad	TxCxExI	T. Lineal	T. Parabólica
México	80.6	84.1	33.7
Chihuahua	204.6	219.3	42.1
Guadalajara	136.6	158.7	25.9
León	40.4	40.5	18.6
Mérida	72.4	76.1	16.8
Monterrey	164.8	186.9	31.4
Orizaba	46.1	47.9	22.2
Puebla	63.7	65.1	32.5
San Luis Potosí	49.5	56.4	22.2
Tampico	78.07	84.7	72.0
Torreón	131.4	140.3	27.1
Veracruz	66.8	67.0	30.2

CAPÍTULO 2

Análisis resistente de tablas de doble entrada

2.1 Definición de tablas de doble entrada

Una tabla de doble entrada es un arreglo matricial de observaciones de una variable respuesta ordenada de acuerdo a dos criterios de clasificación o *efectos* columna y renglón, los cuales pueden tomar I y J valores respectivamente. A la intersección del renglón i con la columna j se le conoce como la entrada o celda (i,j). Ver siguiente figura.

Figura 2.1 Estructura de una tabla de doble entrada

		Efecto columna		
		1	. .	J
Efecto renglón	1	y_{11}	. .	y_{1J}
	:	:	y_{ij}	:
	I	y_{I1}	. .	y_{IJ}

Las tablas aquí estudiadas poseen dos características importantes: a) sólo cuentan con una observación por celda y b) mientras la escala de la variable observada es de tipo intervalar, la de los efectos puede ser incluso de tipo nominal.

Analizar información con esta estructura tiene como propósito entender la relación que guarda la variable observada con los efectos. Específicamente se desea entender su comportamiento ante un cambio de nivel de cualquiera de los dos criterios de clasificación.

Para conseguir lo anterior es necesario identificar una función **g** que asocie cada entrada (i,j) de la matriz con los valores y_{ij} de la variable observada. Como en cualquier caso donde se pretende describir un fenómeno real por medio de un modelo matemático es necesario incluir un término de error o residual e_{ij}, el cual representa las desviaciones de los valores estimados de los observados.

$$y_{ij} = g(i,j) + e_{ij}.$$

Al proceso de identificación de la función **g** aquí se le denomina como el ajuste de un modelo a la tabla de datos. El modelo más simple que se le puede ajustar a una tabla de doble entrada es conocido como aditivo simple, el cual es el objeto del siguiente apartado.

2.2 El modelo aditivo simple (MAS)

Éste se caracteriza porque las contribuciones individuales de ambos factores al valor de la observación y_{ij} se expresa mediante una suma. Además de los dos anteriores este modelo considera un tercer término, el cual aparece en todas las entradas de la tabla y cuyo valor permanece constante, por lo que se le denomina término común. En este caso la función g es de la forma:

$$g(i,j) = \mu + \alpha_i + \beta_j$$

Para una tabla de doble entrada con I renglones y J columnas el modelo aditivo simple queda descrito por la siguiente ecuación

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \text{con } i = 1,\ldots,I \text{ y } j = 1,\ldots,J$$

donde

- y_{ij} - Valor de la variable observada para el nivel i del factor renglón y el nivel j del factor columna
- μ - Término común en toda la tabla
- α_i - Efecto del factor renglón en su nivel i
- β_j - Efecto del factor columna en su nivel j
- ε_{ij} - Residuos

En este modelo tanto las α_i, las β_j, así como los ε_{ij} están centrados en cero.

En muchas ocasiones el MAS no alcanza a describir adecuadamente el comportamiento de los datos debido a que la estructura de la tabla no es aditiva, o por lo menos no lo es completamente. En estos casos se tendrá que buscar algún modelo más complejo. Este problema se aborda un poco más adelante, en el siguiente apartado se presenta una técnica exploratoria que sirve para ajustar un modelo aditivo simple a una tabla de doble entrada.

2.3 Ajuste de un MAS usando un enfoque exploratorio

De manera más específica a lo comentado en el punto 2.1, ajustar un modelo aditivo a una tabla de doble entrada significa obtener los estimadores para cada uno de los términos que aparecen en la ecuación anterior. Usando el enfoque "clásico", la estimación de los parámetros se basaría en el cálculo de las medias aritméticas por renglón y por columna.

En el enfoque exploratorio para obtener los estimadores de los parámetros se hace uso de otra de las medidas de tendencia central pero que tiene la gran

ventaja de ser resistente.[1] El algoritmo que se describe a continuación es la base sobre la cual descansa la propuesta de análisis que se expone en este documento.

2.3.1 El pulido de medianas (PM)

El pulido de medianas (median polish), propuesto por John W. Tukey a mediados de los años setenta, es un algoritmo que de forma iterativa genera los estimadores para los parámetros del modelo aditivo simple. En cada iteración del PM se obtienen y sustraen del cuerpo de la tabla las medianas de los renglones y las columnas, actualizando con ello los estimadores de todos los términos del modelo.

Aunque es indistinto si el pulido se inicia por filas o por columnas, los resultados que se obtienen no serán necesariamente iguales. En este caso se decidió describir el proceso empezando por filas.

Cada iteración del algoritmo consta básicamente de tres etapas. En la primera se obtienen las medianas por renglón, las cuales se restan al cuerpo de la tabla, de ahí la idea del "pulido". En la segunda etapa se calculan las medianas por columna y se restan del cuerpo de la tabla. Los valores que quedan en el cuerpo de la tabla constituyen los residuales. En la última se actualizan el valor de los estimadores del término común y de los efectos. Estas tres etapas se repiten hasta que las medianas de los efectos renglón y columna se hacen cero.

<u>El algoritmo del PM</u>

Los pasos que constituyen una iteración completa del algoritmo son los siguientes.

[1] Recordar que se habla de métodos o estadísticos, como la mediana en este caso, resistentes si éstos generan resultados que cambian sólo ligeramente cuando unos pocos datos son reemplazados por números muy diferentes a los originales.

0. Antes de empezar a ejecutar el algoritmo tanto la columna de efectos-renglón como la fila de efectos-columna, así como el término común son iguales a cero. A los datos observados, ya ordenados en forma tabular, se les considera como los residuos con los que trabajará el algoritmo la primera vez que se efectúe el paso uno.
1. Se obtienen las medianas renglón de los residuos arrojados en la iteración anterior y se guardan en una nueva columna que se denominará "vector columna de medianas-renglón".
2. Se calcula la mediana de la fila de "efectos-columna", arrojados por la iteración anterior. Dicho número será denominado "aportación de los efectos-columna".
3. Se actualiza el valor de la matriz de residuos al sustraérsele el vector columna de medianas-renglón del paso (1).
4. Se suman el vector columna de medianas-renglón y la columna de efectos-renglón de la iteración anterior, lo cual arroja el nuevo valor para la columna de efectos-renglón.
5. Se calculan las medianas de las columnas de la matriz del paso (3) y se guardan en una nueva fila que se denominará "vector renglón de medianas-columna".
6. Se obtiene la mediana de la columna de efectos-renglón obtenida en el paso (4). Dicho número será denominado "aportación de los efectos-renglón".
7. Se actualiza el valor de la matriz de residuos al sustraerle el vector renglón de medianas-columna del paso (5).
8. Se suman el vector renglón de medianas-columna y el renglón de efectos-columna arrojado por la iteración anterior, cuyo resultado arroja el nuevo valor para el renglón de efectos-columna.
9. Se actualiza el valor del efecto común al sumar las aportaciones de los efectos columna y renglón obtenidos en los pasos (2) y (6) al correspondiente que se obtuvo en la iteración anterior.

10. Se considera como la columna de efectos-renglón de esta iteración, a la que resulta de sustraer la aportación de los efectos-renglón del vector columna obtenido en el paso (4).
11. Se considera como el renglón de efectos-columna de esta iteración, al que resulta de sustraer la aportación de los efectos-columna del vector fila generado en el paso (8).
12. Se calculan las medianas de los efectos renglón y columna. Si éstas son igual a cero se detiene el algoritmo. Si no es así se regresa al paso 1.

A continuación se presenta el desarrollo formal elaborado por Hoaglin y Emerson[2].

Las condiciones frontera del algoritmo son:

$$m^{(0)} = a_i^{(0)} = b_j^{(0)} = 0 \quad \text{y} \quad e_{ij}^{(0)} = y_{ij}$$

Después de k-1 iteraciones el PM arroja una y_{ij} dada por la expresión:

$$y_{ij} = m^{(k-1)} + a_i^{(k-1)} + b_j^{(k-1)} + e_{ij}^{(k-1)}$$

La iteración k puede ser dividida en tres partes. En la primera se calculan:

$$\alpha_i^{(k)} = \underset{i=1,\ldots,I}{mediana}\left\{e_{ij}^{(k-1)} \mid j = 1,\ldots,J\right\}$$
$$\mu_b^{(k)} = mediana\left\{b_j^{(k-1)} \mid j = 1,\ldots,J\right\}$$
$$d_{ij}^{(k)} = e_{ij}^{(k-1)} - \alpha_i^{(k)}$$
$$\alpha_i^{(k)} + a_i^{(k-1)}$$

[2] Emerson, John D. y Hoaglin, David C. "Analysis of Two-Way Tables by Medians", en Understanding Robust and Exploratory Data Analysis, comp. por Hoaglin, D.C., Mosteller, F. y Tukey, J.W. New York. John Wiley & Sons. 1983. Pág. 171

En la segunda parte:

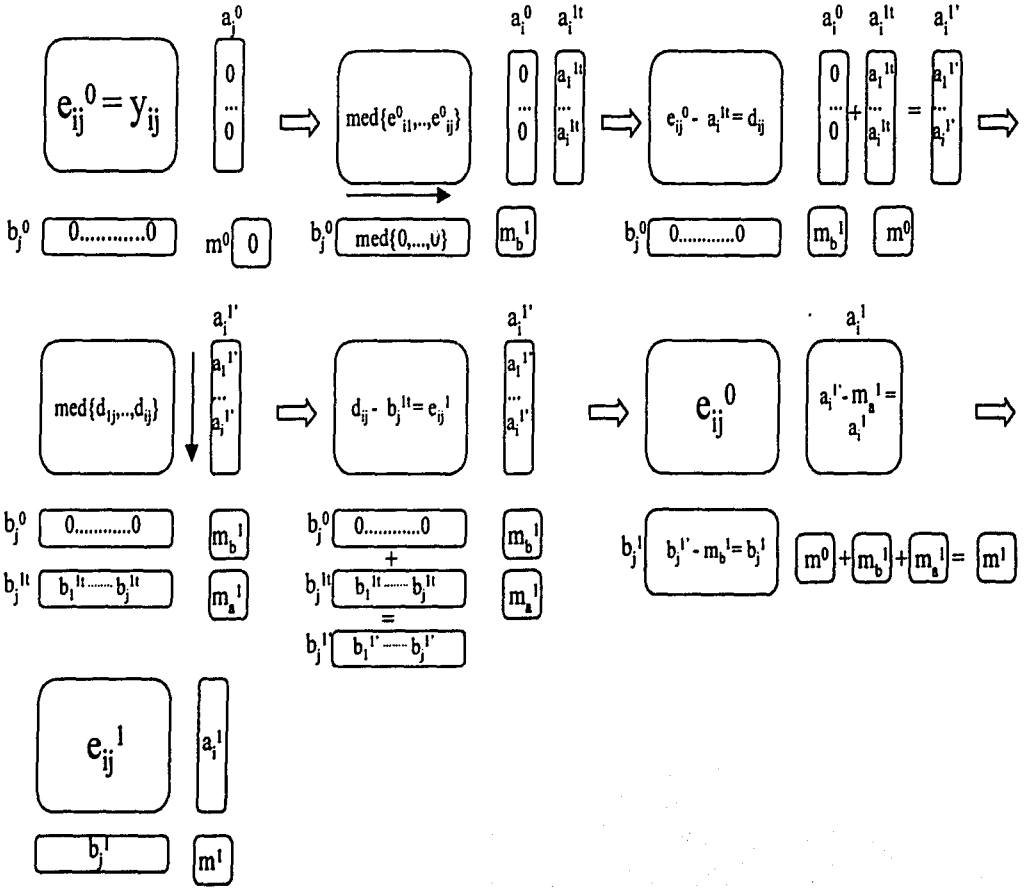
$$\beta_j^{(k)} = \underset{j=1,\ldots,J}{mediana}\left\{d_{ij}^{(k-1)} | i = 1,\ldots,I\right\}$$
$$\mu_a^{(k)} = mediana\left\{a_i^{(k-1)} | i = 1,\ldots,I\right\}$$
$$e_{ij}^{(k)} = d_{ij}^{(k)} - \beta_j^{(k)}$$
$$\beta_j^{(k)} + b_j^{(k-1)}$$

Finalmente:

$$m^{(k)} = m^{(k-1)} + \mu_a^{(k)} + \mu_b^{(k)}$$
$$a_i^{(k)} = a_i^{(k)} + \alpha_i^{(k)} - \mu_a^{(k)}$$
$$b_j^{(k)} = b_j^{(k-1)} + \beta_j^{(k)} - \mu_b^{(k)}$$

En la siguiente página se muestra el diagrama de flujo de una iteración completa del algoritmo.

Primera iteración del algoritmo del pulido de medianas (empezando por filas)



2.3.2 Características del modelo arrojado por el PM

En este apartado se analizarán algunos aspectos sobresalientes de los ajustes que se obtienen al usar el algoritmo. Se discute sobre si el modelo es o no óptimo bajo algún criterio y sobre los límites de su resistencia.

No es único

Como se mencionó en la sección anterior, el algoritmo puede iniciarse de manera indistinta por renglones o por columnas. Sin embargo, los resultados que se obtienen al empezar el pulido por renglones no son necesariamente iguales a los que se generan si se empieza por calcular las medianas por columna. Pese a ello, esta diferencia no resulta ser muy grande y para este enfoque de análisis, esto no representa una complicación mayor. Ambos resultados son útiles, pero es importante recordar esto siempre que se desee replicar el proceso.

No es óptimo

Otra de sus "limitaciones" es que, a diferencia del enfoque clásico, el PM no proporciona un ajuste óptimo en el sentido de minimizar alguna función de los residuos.

En efecto, siempre que se tiene un conjunto de n observaciones de una misma variable y_i, generalmente uno se pregunta por el valor **v** que indica la ubicación del centro del conjunto de datos. Existen diferentes aproximaciones para encontrar la **v**. La más difundida es la que propone a la media aritmética como dicho valor.

Pero la media aritmética tiene una propiedad extremadamente relevante y ésta es la de minimizar la expresión siguiente, la cual es la base del criterio de mínimos cuadrados:

$$\sum_i^n (y_i - v)^2$$

Otra alternativa es seleccionar la **v**, de tal suerte que se minimice la siguiente suma

$$\sum_{\omega}^{n} |y_i - v|$$

que de forma equivalente al caso anterior, ésta constituye la base del criterio de mínimos residuales absolutos (MRA). Para este enfoque el problema se resuelve eligiendo como el valor **v** a la mediana del conjunto de datos. Ahora bien, para ajustar un modelo de tipo aditivo a una tabla de dimensión (IxJ) usando el criterio de MRA, se requiere encontrar los valores de μ, α_i y β_j con i=1,...,I y j=1,...,J tal que se minimice la expresión

$$\sum_i \sum_j |y_{ij} - \mu - \alpha_i - \beta_j|$$

A pesar de lo que en primera instancia pudiera pensarse, los estimadores del modelo que arroja el PM no satisfacen el criterio de MRA. Para verificarlo es necesario recordar el algoritmo. Si se empieza por pulido por filas, durante la primera parte del algoritmo cuando se obtienen las medianas renglón y se restan del cuerpo de la tabla, la suma de esas diferencias es mínima de acuerdo al MRA. Es decir, se obtienen a_i tales que

$$\min_{a_i} \sum_j |y_{ij} - m - a_i - b_j| \qquad \forall \quad i = 1,...,I$$

Observar que en todas estas I sumas las b_j permanecen fijas. Al final de esta primera parte del algoritmo se obtienen los primeros residuales. En la segunda se cambia el orden del pulido y ahora se buscan los valores que cumplan con

$$\min_{b_j} \sum_i |y_{ij} - m - a_i - b_j| \qquad \forall \quad j = 1,...,J$$

Pero ahora se mantienen fijas las a_i. Al término de esta parte se obtienen los residuos finales. En la última etapa se procede a actualizar los valores del término común y de los efectos renglón y columna. Con ello se ha completado la primera iteración del algoritmo.

En la siguiente iteración se vuelve a cambiar el orden del pulido de la tabla y se busca minimizar las sumas de valores absolutos por renglón. Aquí ya no se cumple necesariamente que las a_i del paso anterior minimicen las sumas de los nuevos residuos, por lo que es necesario obtener los nuevos valores de las a_i. En la siguiente parte ocurrirá lo mismo con las b_j; después de actualizar los estimadores del modelo se cambiará el orden del pulido nuevamente y así sucesivamente.

En teoría el proceso se repite hasta que todas las a_i y b_j son iguales a cero. Esto es posible toda vez que en cada iteración la suma de valores absolutos de los residuos se va haciendo cada vez más chica. Esta suma puede entenderse como una sucesión no creciente de números positivos, la cual se sabe converge en el límite por estar acotada por el cero.

Emerson y Hoaglin afirman que en la práctica, los resultados arrojados por el PM con frecuencia son muy cercanos a los óptimos en el sentido de MRA.[3]

Es resistente

Por el contrario, la principal virtud del PM, es sin duda su insensibilidad a la presencia de observaciones extremas. Para abordar este tema, primero se estudiarán los límites de la resistencia de la mediana, lo cual servirá como referencia y guía para determinar los límites para esta propiedad.

Sea $x_1, x_2, \ldots, x_n$ una muestra de n observaciones y sea $k > 0$, $k \in E$. Usando la notación de estadísticas de orden, se define a la mediana como

$$\text{mediana } \{x_1, \ldots, x_k, x_{k+1}, \ldots, x_n\} = \begin{cases} x_{k+1} & \text{para } n = 2k+1 \\ \frac{1}{2}(x_k + x_{k+1}) & \text{para } n = 2k \end{cases}$$

[3] Emerson, J.D. y Hoaglin, D.C. Op. cit. Pág. 184

Supóngase que en la muestra existen t observaciones que pueden considerarse como extremas. Debe notarse que por definición, para obtener la mediana de los datos, estos tienen que ser ordenados por lo que las observaciones anómalas siempre quedarán en los extremos. En estas condiciones $t \leq k-1$ si n=2k o $t \leq k$ si n=2k+1 para que el valor de la mediana no se altere por la presencia de estas observaciones atípicas.

Para relacionar el número anterior con el tamaño de muestra se puede construir una función Ir(n) que represente la proporción máxima permisible de observaciones anómalas en la muestra, la cual puede definirse como

$$Ir(n) = \begin{cases} \dfrac{k-1}{n} & para \quad n = 2k \\ \dfrac{k}{n} & para \quad n = 2k+1 \end{cases}$$

El caso optimista del límite de la propiedad de resistencia de la mediana, supondría que la única observación "buena" es la $x_{(k+1)}$ para el caso de n impar y las $x_{(k)}$ y $x_{(k+1)}$ para una n par. Sin embargo este dato carece de utilidad práctica. En primer lugar porque siempre es más conveniente asumir la posición más conservadora; y en segundo porque de ocurrir ello, se podría discutir sobre si las "verdaderas" observaciones atípicas no serían en realidad la $x_{(k)}$ y/o la $x_{(k+1)}$.

Siguiendo el mismo orden de ideas, para conocer los límites de la resistencia a observaciones extremas del pulido de medianas, es necesario notar que para el caso de tablas de doble entrada de dimensión (IxJ), la localización más desafortunada de éstas ocurre cuando todas caen en un mismo renglón, si I<J, o en una sola columna en el caso de J<I.

Como IJ es el número de celdas de la matriz de datos, se tiene que la función *lr* para el caso matricial se define como:

$$lr(IJ) = \begin{cases} \dfrac{k-1}{IJ} & para \quad min(I,J) = 2k \\ \dfrac{k}{IJ} & para \quad min(I,J) = 2k+1 \end{cases}$$

Como en el caso de la mediana, no se considera de utilidad práctica presentar los límites para el caso optimista.

2.4 Diagnóstico de aditividad de los datos

Hasta este momento se ha expuesto un método para ajustar un modelo aditivo a una tabla de doble entrada usando el pulido de medianas y se han discutido algunas de sus principales características. Pero, ¿qué pasa si los datos no tienen estructura aditiva simple? y ¿cómo saber cuando ocurre esto?. Las respuestas a ambas preguntas se encuentran en el resto del capítulo. A continuación se revisarán algunas técnicas para detectar la no aditividad de los datos. Más adelante se presentan otros algoritmos que sirven para ajustar modelos más complejos a la tabla de datos.

Inspección visual de los residuales

Una forma fácil de verificar si los datos poseen una estructura no aditiva consiste en la revisión "a ojo" de los residuos tratando de detectar algún patrón en su comportamiento. El que esto ocurra sugiere que existe cierto tipo de interacción entre los factores.

2.4.1 La gráfica de diagnóstico

A menudo ocurre que no es fácil percibir la presencia de tales patrones a simple vista, especialmente si la tabla es muy grande. En general siempre será necesario revisar el comportamiento de los residuos con una base más objetiva. Precisamente para ello Tukey propuso la denominada gráfica de diagnóstico (diagnostic plot).

Esta gráfica está definida por los puntos (cv_{ij}, e_{ij}), donde $cv_{ij} = \frac{a_i b_j}{m}$. A los valores cv_{ij} se les denomina "valores de comparación" (comparison values).

La gráfica de diagnóstico de un modelo aditivo que describe adecuadamente la estructura de la tabla no debe mostrar patrón alguno. Entre más se parezca a una recta, se dispone de mayor evidencia de que existe interacción entre los efectos, algo que un modelo de este tipo no puede explicar.

En la figura 2.2 de la siguiente página, se muestran las gráficas de diagnóstico de (a) una tabla con datos ficticios que posee estructura aditiva y (b) una con estructura claramente no-aditiva. En este segundo caso la información tabulada consiste en mediciones del volumen específico del caucho cuando es sometido a seis niveles diferentes de temperatura y presión.[4]

2.4.2 La recta resistente

Además de servir para detectar la existencia de interacción entre los factores, la gráfica de diagnóstico también es de utilidad, como se detallará más adelante, cuando se tiene que ajustar un modelo no aditivo a los datos. Sin embargo, para ello es necesario conocer el valor de su pendiente.

[4] Tomado de Emerson, J.D. y Wong, G.Y. "Resistant Nonadditive Fits for Two-Way Tables", en Exploring Data Tables, Trends and Shapes, Nueva York, John Wiley & Sons, 1985. Pág.72.

Figura 2.2 (a). Gráfica de diagnóstico para una tabla con estructura aditiva

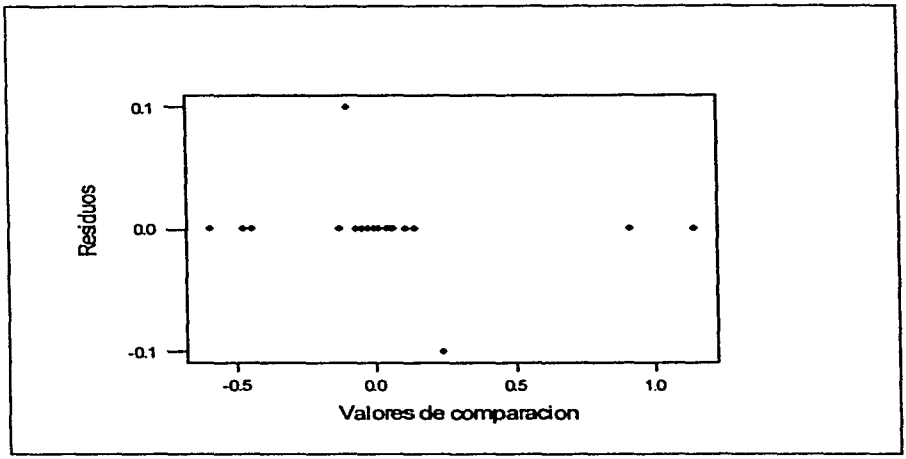
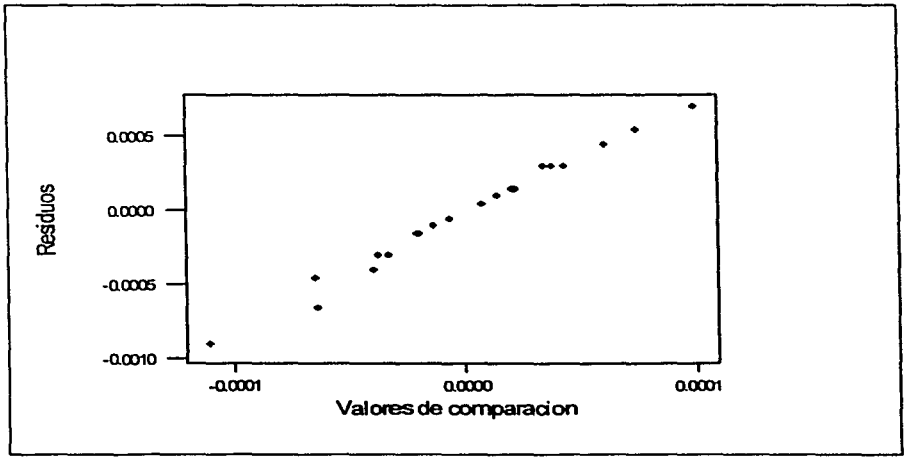


Figura 2.2 (b). Gráfica de diagnóstico para una tabla con estructura no aditiva



Para obtener el valor de la pendiente, según el enfoque clásico, se tendría que determinar la recta que mejor se ajusta a la gráfica usando el método de regresión. En contraste, en lugar de construir dicha recta usando mínimos cuadrados, Emerson y Hoaglin sugieren usar una aproximación a la cual se le conoce como el algoritmo de la recta resistente.

Como las demás técnicas exploratorias, este algoritmo parte de un ajuste inicial, el cual va mejorando en cada iteración. Naturalmente, en el proceso se usa la mediana en lugar de la media.

El punto de partida del algoritmo consiste en dividir al conjunto de puntos a graficar en tres secciones (izquierda, centro y derecha), y en encontrar los pares ordenados más "representantivos" de cada grupo. Con ellos se calcula una recta inicial. Posteriormente se obtienen los residuos del modelo y se calcula su pendiente. Si ésta es cero se detiene el algoritmo. En caso contrario, se usa su valor para "corregir" la ecuación de la recta. Para esta nueva recta se calculan sus residuos y se verifica su pendiente y así sucesivamente. A continuación se presentan los pasos del algoritmo.

Algoritmo para construir la recta resistente

1. Ordenar los pares ordenados de la gráfica de acuerdo a las abscisas y dividir al conjunto en tres secciones: izquierda, centro y derecha; procurando asignar igual número de elementos a cada grupo. Aquellos pares ordenados que tengan la misma abscisa deberán ser ubicados en una misma sección.

2. En cada sección se calcula la mediana de las abscisas, obteniéndose los valores x_I, x_C y x_D. Para cada grupo se obtienen también las medianas de las ordenadas: y_I, y_C y y_D. Estas tres parejas (x_I, y_I), (x_C, y_C), (x_D, y_D) son los puntos resumen de cada sección.

3. Usando los puntos resumen de las secciones izquierda y derecha se calcula el valor de la pendiente:

$$b_1 = \frac{y_D - y_I}{x_D - x_I}$$

4. Con esta pendiente y los tres puntos resumen se calcula el valor de la ordenada al origen de la ecuación de la recta (a_1) el cual será el promedio de los tres números siguientes:

$$a_I = y_I - b_1 x_I$$

$$a_C = y_C - b_1 x_C$$

$$a_D = y_D - b_1 x_D$$

5. Calcular los residuos de la recta estimada para todos los puntos de los tres grupos, usando la ecuación:

$$r_i = y_i - \hat{y}_i = y_i - (a_1 + b_1 x_i)$$

6. Obtener los puntos resumen para los residuos de las secciones izquierda y derecha (x_I, r_I) y (x_D, r_D) donde r_I y r_D representan las medianas de los residuos. Calcular la pendiente de la recta de los residuos:

$$b'_1 = \frac{r_D - r_I}{x_D - x_I}$$

7. Terminar el algoritmo si $b'_1 \approx 0$. En otro caso, actualizar el valor de la pendiente haciendo $b_2 = b_1 + b'_1$.

   Ejecutar los pasos 5 y 6 con lo que se obtiene una b'_2 que es necesario verificar sea igual a cero, paso 7.

En este proceso se puede presentar el caso de haber sobrecorregido la recta, lo cual se detecta cuando b'_{j-1} y b'_j tienen signos opuestos. En esta situación se puede mejorar la estimación de la pendiente al hacer:

$$b_{j+1} = b_j - b'_{j-1}\left(\frac{b_j - b_{j-1}}{b'_j - b'_{j-1}}\right)$$

Para este nuevo valor se calcula nuevamente la b'_{j+1}, terminando el algoritmo si ésta es igual a cero.

Con esto se termina la parte del diagnóstico de aditividad. A continuación se revisarán los modelos más apropiados para tablas que poseen una estructura más compleja.

2.5 Modelos para datos con estructura no aditiva

Una vez que ya se ha confirmado que la tabla de datos posee estructura no aditiva, existen dos posibles caminos para ajustarle un modelo. El primero consiste en transformarlos de tal suerte que éstos posean estructura aditiva en la nueva escala. En el segundo se pretende incorporar al modelo por lo menos un término adicional que sirva para explicar la interacción entre los efectos.

Una característica singular de la gráfica de diagnóstico es que en ambos casos resulta ser de mucha utilidad. En el siguiente apartado se hará una breve exposición del tipo de transformaciones que se usan con mayor frecuencia para remover la no aditividad en los datos y la forma como se usa la gráfica con este objetivo. En un apartado posterior se explicará como se usa para ajustarle a la tabla un modelo que considera el término por interacción.

2.5.1 Familia de transformaciones potencia

Transformar un conjunto de datos consiste en aplicar una función que cambie la escala en que están medidos. Algunas transformaciones cambian también la forma de su distribución, mientras otras son de tipo lineal, esto es que no implican que ésta cambie.

Dependiendo del tipo de datos con los que se esté trabajando, la transformación puede perseguir alguno de los siguientes objetivos:

1. Facilitar la interpretación/análisis de los datos;
2. Promover la simetría de la distribución de los datos;
3. Estabilizar su varianza;
4. Promover una relación lineal entre dos variables;
5. Simplificar la estructura de los datos de tablas de dos y más entradas

Aquí únicamente se revisarán las transformaciones del tipo cinco de la lista anterior, entendiendo por "simplificar la estructura" de una tabla el promover la aditividad de los datos.

De este grupo sobresale la llamada "familia de transformaciones potencia", la cual es una colección de funciones que pueden ser descritas como $y^p = az + b$ donde a, b y p son constantes arbitrarias. Cuando p = 0 se reemplaza y^0 por log(y) con objeto de contar con una familia de curvas que cambien ligeramente conforme p se aproxima a cero.

La familia de transformaciones potencia queda formalmente definida como

$$T(x) = \begin{cases} ax^p + b & p \neq 0 \\ c\log(x) + d & p = 0 \end{cases}$$

Donde a, b, c, d y p son números reales.

Si a > 0 cuando p > 0 y a < 0 para p < 0, las funciones que pertenecen a esta familia poseen algunas propiedades muy interesantes:

1. Son funciones estrictamente crecientes que preservan el orden de los datos originales. Esto es que si $x_1 < x_2$ entonces $T(x_1) < T(x_2)$.
2. Las medianas son transformadas en medianas, y en general cuartiles son transformados en cuartiles.
3. Son funciones continuas. Ello implica que datos "cercanos" en la escala original, bajo la transformación continuarán estando "cercanos".
4. Tienen derivadas de todos los órdenes.
5. Son funciones o cóncavas o convexas en todo su dominio.

Salvo por la restricción impuesta al coeficiente "a", la selección de las constantes b, c y d se hace a conveniencia. En cambio el valor p dependerá de la estructura de los datos a transformar.[5]

Se dice que la elección de las constantes a, b, c y d es a conveniencia debido a que para toda p fija, cualquier selección de las constantes a y b, ó (c y d cuando p=0), representa una transformación lineal de cualquier otra selección de constantes.

Esto es, que si $t(x)$ y $T(x)$ son dos transformaciones distintas dadas por $t(x) = ax^p + b$ y $T(x) = Ax^p + B$ y a es una constante distinta de cero, entonces es posible transformar $t(x)$ en $T(x)$.

Demostración:

$$T(x) = \left(\frac{A}{a}\right)t(x) + \left(B - \frac{A}{a}b\right) =$$

[5] Emerson, J. D. y Stoto, M. A. "Transforming Data" en Understanding Robust and Exploratory Data Analysis, New York, John Wiley & Sons, 1983. Pág. 99

$$\frac{A}{a}\left(ax^p + b\right) + B - \frac{A}{a}b =$$

$$Ax^p + B$$

De lo anterior se deduce que aplicar cualquier transformación lineal a un conjunto de datos implica cambiar su origen y su escala pero de manera uniforme, lo cual no afecta la forma de su distribución.

De esta familia Tukey sugiere considerar sólo un grupo de elementos que tienen la ventaja de que los datos en las nuevas unidades son relativamente más fáciles de interpretar en comparación con la escala generada por otras transformaciones lineales. Este grupo, muy utilizado en la práctica, es conocido como la "Escalera de Transformaciones de Tukey" y está integrado por funciones del tipo:

$$Tl(x) = \begin{cases} x^p & p = \ldots -2, -1, -\frac{1}{2}, -\frac{1}{3}, -\frac{1}{4}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, \ldots \\ \log(x) & p = 0 \end{cases}$$

2.5.2 La gráfica de diagnóstico y la transformación de los datos

Para usarla con este fin, es necesario obtener la recta resistente y calcular su pendiente s de acuerdo a como fue explicado en un apartado anterior. Una transformación que sirve para simplificar la estructura de la tabla de datos es aquella que usa como exponente a p, con p = 1-s.[6]

Un problema con la transformación de datos es que en muchas ocasiones, aún cuando se ha tenido éxito en remover la no-aditividad, la interpretación de los resultados en la nueva escala se dificulta o carece de significado práctico.

[6] Una justificación a esto se encuentra en Emerson, John D. "Mathematical Aspects of Transformation" en Understanding Robust and Exploratory Data Analysis. Págs. 267-273

Otra dificultad se presenta al momento de comparar los modelos para elegir al más adecuado para representar al conjunto de datos. Esto se hace comparando sus residuos, pero para ello se requiere que todos se encuentren medidos en la misma escala. En los casos donde se usó una transformación potencia es necesario expresar los residuos en la escala original usando la función inversa.

Debido a lo anterior en muchas situaciones puede ser preferible optar por ajustar un modelo más complejo en lugar de transformar los datos. A continuación se revisarán algunos de estos modelos.

2.5.3 El modelo extendido

En 1949 Tukey propuso un modelo para el caso en que la interacción de los efectos está presente en toda la tabla y ésta puede ser explicada agregando un término adicional, en el cual aparecen multiplicadas las aportaciones de ambos efectos para cada celda.[7] Tukey lo llamó de "un grado de libertad por no aditividad", aunque aquí se le denomina simplemente como modelo extendido:

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma\alpha_i\beta_j + \rho_{ij}$$

Es conveniente aclarar que en este trabajo únicamente se considera el caso más simple de interacción entre los efectos, el cual ocurre cuando ésta se presenta en todo el cuerpo de la tabla. Méndez señala que es factible que la interacción se presente sólo en ciertas celdas, o incluso en una sola, situación que suele confundirse, advierte, con la presencia de observaciones atípicas.[8]

Para ajustar un modelo de tipo extendido a una tabla de datos es necesario obtener en primer lugar los estimadores para μ, α y β del modelo aditivo simple.

[7] Tukey, John W. "One Degree of Freedom for Nonadditivity", Biometrics, 5, 1949. Págs. 232-42.
[8] Méndez, Ignacio. "Descomposición de la interacción en tablas de doble entrada" en Memoria del X Foro Nacional y II Congreso Iberoamericano de Estadística. México INEGI, 1996. Pág. 161

Con ellos se construye la gráfica de diagnóstico y su recta resistente, cuya pendiente **s** sirve para obtener $\hat{\rho} = k$, mediante:

$$k = \frac{\mathbf{s}}{\hat{\mu}}$$

Esto es así ya que los residuos del modelo aditivo simple se definen como:

$$e_{ij} = y_{ij} - (m + a_i + b_j) \qquad \ldots (1)$$

y la ecuación de la recta que le fue ajustada a la gráfica de diagnóstico es

$$e_{ij} = cte + \mathbf{s}cv_{ij} + err_{ij} \qquad \ldots (2)$$

donde cv_{ij} son los valores de comparación. Restando (2) de (1) se obtiene

$$0 = y_{ij} - (m + a_i + b_j) - cte - \mathbf{s}cv_{ij} - err_{ij} \qquad \ldots (3)$$

Por otro lado, se tiene que si

s = pendiente de la gráfica de los puntos $\left(\frac{a_i b_j}{m}, e_{ij}\right)$

k = pendiente de la gráfica de los puntos $(a_i b_j, e_{ij})$

Entonces

$$\mathbf{s} = mk \qquad \ldots (4)$$

Sustituyendo (4) en (3)

$$0 = y_{ij} - (m + a_i + b_j) - cte - mk\frac{a_i b_j}{m} - err_{ij} \quad \Rightarrow$$

$$cte + err_{ij} = y_{ij} - (m + a_i + b_j + ka_i b_j) \quad \Rightarrow$$

$$r_{ij} = y_{ij} - (m + a_i + b_j + ka_i b_j)$$

Donde r_{ij} son los estimadores para los residuales ρ_{ij} del modelo extendido.

La relación entre los residuos ε_{ij} del aditivo simple y los ρ_{ij} está dada por $\varepsilon_{ij} = \gamma\alpha\beta_j + \rho_{ij}$. En contraste con los ε_{ij}, los residuos ρ_{ij} no necesariamente estarán centrados en cero, por lo cual es recomendable centrarlos usando el pulido de medianas.

Si se conviene en distinguir con un "0" a los estimadores del modelo extendido, antes de que sus residuos estén centrados en cero y con un "1" al término común y los efectos arrojados al pulir la tabla de residuos e_{ij}, los estimadores para el modelo extendido se obtienen haciendo $a_i = a_i^0 + a_i^1$, $b_j = b_j^0 + b_j^1$ $m = m^0 + m^1$. Con estos valores se actualiza el estimador k y con ello finalmente se generan los nuevos residuos r_{ij}.

Este es un proceso iterativo, como el resto de las herramientas del AED, sin embargo, para efectos prácticos en la mayoría de los casos bastará con efectuar el procedimiento completo una sola vez. El algoritmo para ajustar un modelo extendido a una tabla de doble entrada sería el siguiente:

1. Usar el PM en la tabla original de datos para obtener los valores iniciales del término común, los efectos y residuos: m^0, a_i^0, b_j^0 y e_{ij}^0.
2. Construir la gráfica de diagnóstico y obtener $k^0 = \frac{s^0}{m^0}$.
3. Construir la tabla de residuales $r_{ij}^0 = e_{ij} - k^0 a_i^0 b_j^0$.
4. Usar el PM en la tabla de residuales r_{ij}^0 y obtener m^1, a_i^1 y b_j^1.
5. Actualizar el valor de los estimadores, mediante $m = m^0 + m^1$, $a_i = a_i^0 + a_i^1$ y $b_j = b_j^0 + b_j^1$.
6. Construir una nueva gráfica de diagnóstico usando los valores de los estimadores del paso anterior y obtener k. Actualizar el término ka_ib_j.
7. Finalmente, calcular los residuos $r_{ij} = y_{ij} - (m + a_i + b_j + ka_ib_j)$.

2.5.4 Los modelos multiplicativos

Otro tipo de estructura no aditiva de los datos para la cual el PM resulta ser de utilidad en ciertas condiciones, es aquella que puede ser descrita por la ecuación:

$$y_{ij} = \nu\phi_i\lambda_j + \varepsilon_{ij}$$

Para que el PM pueda ser usado en tablas con esta estructura se requiere que todas las entradas de la tabla sean estrictamente positivas. Si ello ocurre, usando logaritmos se obtiene $\ln(y_{ij}) = \ln(\nu\phi_i\lambda_j) = \ln(\nu) * \ln(\phi_i) * \ln(\lambda_j)$, con lo cual se provee de estructura aditiva a los datos. Tomando antilogartimos se regresan éstos a las unidades originales y se obtienen los estimadores para el modelo anterior, el cual es conocido como multiplicativo simple.

Otro tipo de estructura no aditiva de una tabla de doble entrada queda descrita por la siguiente ecuación:

$$y_{ij} = \mu + \nu\phi_i\lambda_j + \varepsilon_{ij}$$

Este modelo corresponde a un conjunto de datos con estructura multiplicativa simple que han sido trasladados μ unidades del origen. Sería fácil obtener los estimadores para cada uno de los términos si todas las entradas de la matriz fueran estrictamente positivas y se conociera a priori el valor de μ,. Bastaría con ejecutar el algoritmo descrito previamente.

La estimación para este modelo denominado modelo multiplicativo está determinada por

$$\hat{y}_{ij} = q + hc_i d_j$$

Es claro que se presenta un aparente serio problema cuando no es posible usar la transformación logaritmo. Es decir, cuando por lo menos una de las entradas

de la matriz es cero o de signo negativo. Para estos casos es de gran utilidad saber que el conjunto de modelos de tipo multiplicativo y los de tipo extendido son en realidad el mismo, solo que descritos de diferente forma.

En efecto, conociendo los valores q, h, c_i y d_j es posible encontrar el modelo extendido equivalente. Inversamente, una vez conocidos la m, a_i, b_j y k es sencillo encontrar el correspondiente modelo multiplicativo.[9]

Con esto se dispone de un método indirecto para ajustarle un modelo a una matriz de datos con estructura multiplicativa cuando alguna de sus entradas es cero o tiene signo negativo.

2.5.5 El modelo mixto

Cuando se trabaja con los residuos de un modelo aditivo simple con la intención de mejorar la calidad del ajuste, es factible que se tenga que enfrentar la dificultad de que éstos carezcan por completo de estructura aditiva, pero que posean estructura multiplicativa perfecta. Este hecho acarrea ciertas complicaciones.

En primer lugar se tiene que el término común y los efectos son cero, esto naturalmente impide la construcción de la gráfica de diagnóstico y por consecuencia obtener el modelo extendido. Por otro lado sería igualmente imposible usar la función logaritmo para transformar los datos, toda vez que más de la mitad de las entradas no-cero tendrían signo negativo, al estar los residuos centrados alrededor de este valor.

[9] La demostración es muy simple y puede consultarse en Emerson, John D. y Wong, George Y. Op. cit. Págs. 84-85.

Basándose en la propiedad de resistencia del PM, Emerson y Wong sugieren que en estos casos se identifiquen primeramente los renglones y columnas donde el número de entradas negativas sea mayor al de las positivas con el propósito de cambiarles su signo. De igual forma sugieren identificar las entradas con ceros y sustituirlas por una cantidad positiva lo suficientemente pequeña como para poder aplicar logaritmos sin que ello cause problemas.

Después de la operación de cambio de signo todas las entradas negativas restantes pueden ser sustituidas por la misma cantidad positiva, o bien pueden ser consideradas como datos faltantes. En ambos casos es importante tener en mente los límites para la resistencia del PM, especialmente en los casos de tablas de dimensiones pequeñas, de hasta cinco renglones y/o columnas, ya que dos o más entradas negativas en un mismo renglón y/o columna pueden afectar de manera considerable el modelo ajustado.

Hecho lo anterior, la tabla ya está en condiciones de ser transformada. Una vez obtenidos los estimadores para el modelo multiplicativo simple, se procede a "regresarles" su signo a todos los renglones y columnas en cuestión. Finalmente, al combinar los resultados del modelo aditivo, el cual fue ajustado a los datos originales, con los del modelo multiplicativo ajustado a los residuos del primero, se obtiene un modelo del tipo:

$$y_{ij} = \mu + \alpha_i + \beta_j + v\phi_i\lambda_j + \rho_{ij}$$

que aquí se le denominará como modelo mixto (additive-plus-multiplicative fit).

Antes de presentar el algoritmo completo para el modelo mixto, es conveniente comentar que éste puede entenderse como una generalización de todos los que se han venido mostrando. Es decir, todos son susceptibles de ser estudiados como casos particulares de este último.

Condición	Modelo resultante
$\nu = 0$	Aditivo simple
$\mu = \alpha_i = \beta_j = 0$	Multiplicativo simple
$\alpha_i = \beta_j = 0$	Multiplicativo
$\alpha_i = \phi_i \quad \beta_j = \lambda_j$	Extendido

Otros dos casos especiales que no han sido comentados son

Condición	Modelo resultante
$\alpha_i = \phi_i$ y $\beta_j \neq \lambda_j$	Lineal por columnas
$\alpha_i \neq \phi_i$ y $\beta_j = \lambda_j$	Lineal por renglones

El modelo lineal por columnas es útil cuando al graficar (e_{ij}, a_i) para cada j, se obtiene una línea recta. De manera equivalente, se usa el lineal por renglones cuando al graficar los puntos (e_{ij}, b_j), se obtiene una recta.

A continuación se presenta el algoritmo para ajustar un modelo mixto a una tabla de datos.

1. *Ajuste de un modelo aditivo simple.* Usar para tal efecto el PM y revisando visualmente la matriz de residuos $\{e_{ij}\}$, así como por medio de la gráfica de diagnóstico. Evaluar la conveniencia de transformar los datos para ajustarles un modelo multiplicativo en lugar de uno de tipo extendido.
2. *Generación de la matriz ajustada de residuos.* Identificar las columnas y renglones en los que el número de entradas negativas sea superior al de las positivas y cambiarles de signo. Después sustituir todas las entradas cero por una cantidad positiva, de orden varias veces menor al valor más chico de la

tabla. Decidir si las entradas que continúan siendo negativas se considerarán como observaciones faltantes o se les aplicará el mismo tratamiento que a los ceros. A esta matriz de residuos $\{u_{ij}\}$, con entradas estrictamente positivas es a la que se le llama "matriz ajustada de residuos".

3. Aplicar la transformación logaritmo a la matriz $\{u_{ij}\}$. Usar el PM para obtener un ajuste aditivo simple. $Log(u_{ij}) = m' + a_i' + b_j'$
4. Tomar antilogaritmos para regresar a la escala original de medición de los datos, obteniéndose $u_{ij} = k'c'_i d_j'$
5. Cambiar el signo de los renglones y columnas a los que se les cambió de signo en el paso 2. En este momento se tiene $e_{ij} = kc_i d_j$. Con ello se completa el modelo $y_{ij} = m + a_i + b_j + kc_i d_j + r_{ij}$
6. Centrar en cero los residuales r_{ij} del modelo, usando nuevamente el PM para tal fin.
7. De juzgarse necesario iterar el algoritmo y actualizar el valor de los estimadores.

2.6 Comparación de modelos

En el proceso de análisis exploratorio de tablas de datos es común que se tenga que decidir entre dos o más modelos porque parecen tener un desempeño similar. Con ello surge la pregunta, ¿cuál de todos ellos es el más adecuado para representar a los datos?. La respuesta a esta pregunta se encuentra naturalmente en los residuos. El modelo que describe mejor la estructura de la tabla es aquel que además de arrojar los residuos más chicos, éstos muestran el mejor comportamiento, entendiendo por ello que no sigan ningún patrón específico.

Con este fin, Tukey desarrolló dos diagramas que en la actualidad son, tal vez, las herramientas exploratorias de uso más difundido: el Diagrama de Tallo y Hoja

(Stem & Leaf Display), y el Diagrama de Caja (Box Plot) y su extensión el de Cajas Paralelas (Paralel Box Plot).

2.6.1 El Diagrama de tallo y hoja (DTH)

De manera similar que el histograma, este diagrama proporciona de manera gráfica valiosa información sobre un conjunto de datos. Permite identificar que tan simétrica es la distribución de los datos; si éstos están centrados y alrededor de qué valor; asimismo da una idea sobre su dispersión.

Sin embargo, a diferencia del histograma, el DTH conserva los dígitos más significativos de los datos lo que permite observar su distribución *dentro* de cada intervalo y así poder detectar patrones en los mismos. Con ello el analista está en mejores condiciones para evaluar la calidad de la representación y en caso de requerirlo, éste pueda aumentar o disminuir el número de intervalos considerados en el diagrama. Otra ventaja que tiene el DTH es la de facilitar la detección de observaciones atípicas o extremas (outliers)

Para su construcción es necesario en primer lugar, identificar los valores máximo y mínimo del conjunto de datos. Después se requiere descomponer en dos partes cada una de las observaciones. Una parte representará el "tallo" y la otra formará la "hoja". La selección de la partición tiene que ver con el número de tallos con que contará el diagrama, lo que a su vez, requiere tomar en consideración el número total de observaciones y el intervalo que cubren.

Cuando se trabaja con números enteros, en general las hojas se forman con las unidades, dejando las decenas y/o centenas para formar los "tallos". Cuando se trabaja con racionales, o con números enteros relativamente grandes que hayan sido divididos entre cien, mil, etc., para una mayor facilidad en su manejo, es menester determinar el número de posiciones decimales que se considerarán

significativas y, por lo tanto, que vale la pena conservar. La menor de ellas será la que represente la "hoja" de cada tallo.

Una regla que ha demostrado funcionar bien para determinar el número más conveniente de líneas (tallos) en el diagrama, $20 \leq n \leq 300$ es la misma que se usa para determinar el máximo número de intervalos en un histograma. Este cantidad está determinada por

$$L = [10 * \log_{10}(n)] \qquad para \quad 20 \leq n \leq 300$$

Donde: n - número de datos y [ ] - función parte entera

En ciertos casos puede ser deseable mostrar un menor número de intervalos a los recomendados por esta regla. Velleman (1976), propuso considerar de manera alternativa $L = 2\sqrt{n}$ la cual recomienda menor número de líneas que la anterior para $n < 100$.

Una vez establecido el valor de L, se requiere determinar el intervalo de valores a considerar en cada línea. La manera más simple de hacerlo es dividir el rango del conjunto (observación más grande – observación más chica), entre el valor de L. El cociente será redondeado (hacia arriba) a la potencia de 10, con exponente entero más próximo.

Una vez construido el diagrama es posible entonces ordenar las hojas, lo cual representa otra de las aplicaciones del DTH, la de poder ordenar con facilidad un conjunto de datos.

Variantes del diagrama

En algunos casos después de construir el DTH se puede encontrar que algunas de las líneas del diagrama quedan muy saturadas (con muchas hojas), lo cual dificulta la inspección de los datos. El fenómeno anterior es un signo que revela la existencia de intervalos no representados por el diagrama básico.

Una alternativa de solución es partir en dos la longitud de cada línea. Es decir, obtener dos tallos por cada uno de los tallos originales. Es costumbre usar el símbolo "*" para identificar las líneas cuyas hojas van del 0 al 4; y el símbolo "·" para las líneas con hojas del 5 al 9.

En ciertos casos duplicar el número de tallos es aún insuficiente. Es posible entonces quintuplicar el número original de tallos, usando "*" para identificar las líneas de 0 y 1; "t" para líneas de 2 y 3 (del inglés two & three); "f" para las de 4 y 5 (de four & five); "s" para las de 6 y 7 (de six & seven); y "·" para los 8 y 9.

En estas dos últimas variantes la longitud de los intervalos correspondería a 5 y 2 veces la potencia de 10 más cercana, respectivamente.

Por último, cuando se trabaja con números positivos y negativos (un ejemplo claro son los residuos, los cuales tienden a estar centrados en cero), no es extraño que deban incluirse los tallos: "+0" y "-0". En estos casos se recomienda distribuir de manera equitativa en ambos tallos los valores exactamente iguales a cero con objeto de no afectar la simetría del diagrama.

2.6.2 El diagrama de caja (DC)

Esta es una representación gráfica de algunas medidas resistentes de localización y de dispersión de un conjunto de datos. Este diagrama desarrollado también por Tukey, muestra la localización donde se acumula la mitad de los datos; de la mediana; de la observación más chica y de la más grande; así como la de las observaciones extremas.

Así, de un solo vistazo se puede identificar la forma de la distribución de los datos, su simetría, la longitud y "pesadez" de las colas. En algunas variantes del

diagrama, es factible determinar incluso si la distribución se aleja de la normalidad.

El lugar donde se concentra la mitad de las observaciones está representado por la "caja" en el diagrama, sus extremos están localizados justamente en los cuartiles primero y tercero. La mediana, o segundo cuartil, queda ubicada por lo tanto dentro de la caja y comúnmente se le señala por medio de una raya.

Existen diferencias entre algunos autores sobre la forma como deben marcarse los datos que quedan fuera de la caja. Emerson y Strenio sugieren marcar con "x", a todos los valores que se ubican a más de 1.5 veces de la dispersión cuartil de los extremos de la caja, a los que definen como observaciones extremas. Esta dispersión cuartil hace referencia a la distancia que hay entre el primero y tercer cuartil.

El resto de las observaciones, las que caen dentro de los rangos [primer cuartil - 1.5*(dispersión cuartil)] y [tercer cuartil + 1.5*(dispersión cuartil)], quedan unidas a los extremos de la caja por medio de una línea continua.

Por su parte, Hartwig y Dearing[10] sugieren marcar con una "x" al valor más chico y al más grande que se encuentren a menos de 1.0 veces la dispersión cuartil de la caja y unirlos a sus extremos por medio de una línea punteada. Al resto sugieren marcarlos de manera individual de dos formas diferentes: aquellos que se ubiquen entre 1.0 y 1.5 veces la dispersión cuartil de los extremos de la caja, con puntos claros; y aquellos que estén a más de 1.5 veces la dispersión cuartil por medio de puntos oscuros.

[10] Hartwig, Frederick y Dearing, Brian E. Exploratory Data Analysis. 12ª ed. California, E.U. SAGE Publications, 1990, pág. 23

La razón de esta distinción es porque así se permite detectar si una distribución se aleja de la normalidad. En efecto, en la distribución normal aproximadamente sólo un 5% de las observaciones quedan a una distancia mayor de 1.0 rangos intercuartiles a la izquierda y derecha del primer y tercer cuartil respectivamente. Por lo que un indicador de que una distribución se aleja de la normalidad, es que más del 5% de sus observaciones aparezcan marcadas de manera individual en el diagrama.

2.6.3 El porcentaje (P) de reducción de la variación absoluta total

Esta es una medida numérica que permite comparar la calidad del ajuste de dos modelos diferentes. La idea fundamental detrás de este indicador, es la de comparar la dispersión de los residuos del modelo con respecto a su mediana, con la dispersión de los datos originales con respecto a su propia mediana.

Sea y_{ij} un conjunto de observaciones a las que les ha sido ajustado un modelo como los discutidos en este trabajo. Sean t_{ij} sus valores estimados y sean r_{ij} sus residuos. Entonces se tiene que

$$y_{ij} = t_{ij} + r_{ij}$$

De manera análoga al criterio de la R^2 o "fracción de la suma de cuadrados de los residuos explicada por el ajuste", la cual en el análisis de varianza sirve como medida de bondad de ajuste, donde

$$R^2 = \left[1 - \frac{\sum_i \sum_j r_{ij}^2}{\sum_i \sum_j (y_{ij} - \bar{y})^2}\right] * 100$$

En el enfoque exploratorio se define a P como el "porcentaje de reducción de la variación absoluta total alcanzada por el modelo". Donde

$$P = \left[1 - \frac{\sum_i \sum_j |r_{ij}|}{\sum_i \sum_j y_{ij} - med\{y_{ij}\}}\right] * 100$$

Con el tema de la comparación de modelos concluye la presentación de la metodología para el análisis exploratorio de tablas de doble entrada. En el siguiente capítulo se muestran y discuten cuatro ejemplos de aplicación de todas estas herramientas, los cuales están orientados a analizar la tasa de desempleo abierto desde distintos enfoques.

CAPÍTULO 3

Aplicaciones del pulido de medianas: análisis exploratorio de la TDA

El presente capítulo consta de cuatro secciones, en cada una de las cuales se estudia el comportamiento de la tasa de desempleo abierto[1] con diferentes enfoques, empleando para ello las técnicas expuestas anteriormente.

En la primera aplicación se desea averiguar si el comportamiento del desempleo abierto es geográficamente homogéneo. En la segunda se aborda el problema de la cointegración de la TDA de las Ciudades de México, Guadalajara y Monterrey. En la tercera se construye un modelo para la TDA de doce áreas urbanas, comparándose su calidad con la de tres ajustes obtenidos con un enfoque de series de tiempo. Finalmente, en un contexto de pronóstico de series de tiempo se presenta la que es, tal vez, la propuesta de aplicación más relevante del pulido de medianas.

[1] En un caso se consideró conveniente estudiar también a la tasa de ingresos insuficientes y desempleo (TIID). Ver página 20 para detalles de su definición.

Como ya ha sido mencionado, la TDA es el objeto de estudio en estos cuatro ejemplos de aplicación. Sin embargo, la cobertura geográfica y temporal de la información es diferente para cada uno de ellos, por lo que es necesario especificar en cada caso las ciudades y el periodo de tiempo contemplados.

Con relación al proceso de análisis de los datos se juzgó que era suficiente explicarlo en forma detallada sólo para el primer ejemplo, ya que siempre es el mismo. De forma similar, aún cuando las gráficas y diagramas son herramientas indispensables para el análisis, sólo se presentan las más importantes en cada caso. Lo anterior con la intención de que el documento no creciera de manera innecesaria.

3.1 Análisis regional del desempleo

México es un país tan heterogéneo que en muchos aspectos parece como si en él coexistieran dos o más naciones. Esto es particularmente notorio en términos de desarrollo donde se aprecian fuertes diferencias en el ámbito regional. En este contexto se plantea la interrogante de si el desempleo refleja o no dichas diferencias. Este ejemplo está orientado precisamente a tratar de responder tal pregunta.

Para esta aplicación, además de trabajar con la TDA trimestral se decidió analizar también la TIID, por razones que se explican más adelante. Las zonas urbanas consideradas fueron las 16 de las que se dispone de información para el periodo comprendido entre los años de 1989 al 2000. Estas son: Chihuahua, Ciudad Juárez, Guadalajara, León, Matamoros, Mérida, México, Monterrey, Nuevo Laredo, Orizaba, Puebla, San Luis Potosí, Tampico, Tijuana, Torreón y Veracruz, así como un total de zonas urbanas. En este caso los datos fueron tomados del banco de datos del INEGI, el cual puede consultarse en el sitio de internet del Instituto[2].

[2] INEGI. "Banco de Información Estadística". [s.p.] http://www.inegi.gob.mx

Antes de empezar a detallar el proceso es importante hacer un par de convenciones. La primera se refiere a la estructura de la tabla de datos. De ahora en adelante los renglones de la matriz corresponderán al periodo de tiempo observado mientras que las columnas representarán a las ciudades.

Como se vio en el capítulo anterior, a pesar de la gran versatilidad y facilidad de aplicación del pulido de medianas, tiene el inconveniente de no arrojar un modelo único, toda vez que sus resultados pueden variar ligeramente dependiendo de si el pulido se inició por filas o por columnas.

La segunda convención se refiere a si el pulido de la tabla se empezará por filas o por columnas. Como fue señalado previamente este orden es indistinto; sin embargo, dado que los resultados que arroja el PM no son necesariamente los mismos, se conviene que siempre se empezará por filas.

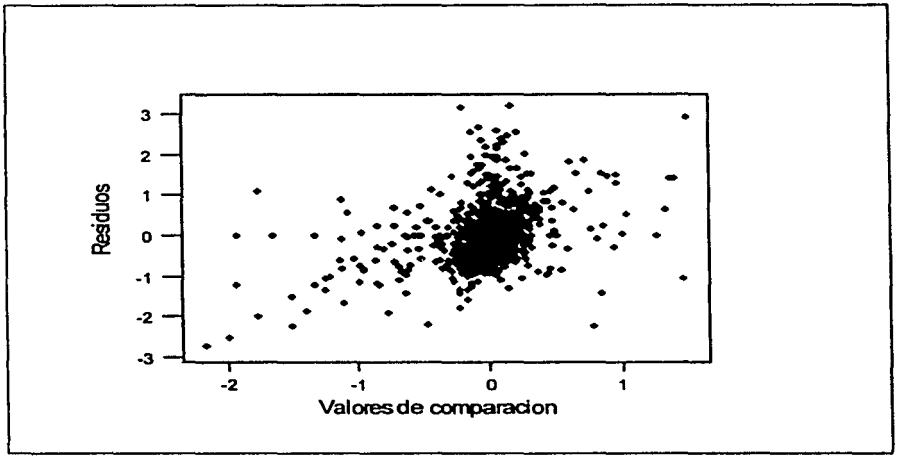
Establecido lo anterior se procede a detallar el análisis al que fueron sometidos los datos. El primer paso consiste en ordenarlos en forma de tabla. Para este caso la matriz consta de 48 filas por 17 columnas. Esta información se muestra en la tabla 1.1 del anexo A.

Una vez obtenidos los estimadores que arrojó el algoritmo, el cual requirió de ocho iteraciones, se procede a evaluar su calidad de ajuste. Para lo cual es menester verificar que los residuos no muestren ningún patrón específico. El primer paso consiste en tratar de identificar alguno mediante la simple inspección visual. Sin embargo, en todos los casos de estudio del presente trabajo, hacer esta revisión resulta de nula utilidad, dado el gran número de observaciones. Por esta razón se omitirá este primer paso de ahora en adelante.

El siguiente paso, y de hecho el más efectivo, consiste en la construcción de la gráfica de diagnóstico, la cual muestra la regresión de los residuos en los valores

de comparación y cuya pendiente sugerirá la transformación potencia más adecuada para remover la no aditividad de los datos.

Figura 3.1 Gráfica de diagnóstico de un modelo aditivo simple para la TDA.



Aunque se aprecia una gran cantidad de puntos alrededor del origen, la pendiente de la recta resistente[3] sugiere que los datos poseen cierta estructura no aditiva, la cual podría removerse al aplicarles una transformación potencia $p = 1 - 0.78 \approx 0.2$. La nueva tabla que se obtuvo al transformar los datos requirió de siete iteraciones del PM y la pendiente de su gráfica de diagnóstico fue cero.

Un tercer modelo a probar es el extendido, el cual intenta explicar la interacción de los efectos mediante la inclusión de un término adicional. Dicho término es de la forma $\frac{s}{m} a_i b_j$ donde s es la pendiente de la gráfica

[3] La descripción detallada de esta herramienta se presenta en la página 43.

de diagnóstico. La tabla 1.2 (a), (b) y (c) que aparece en el anexo A muestra los estimadores para el término común y los efectos ciudad de los tres modelos.

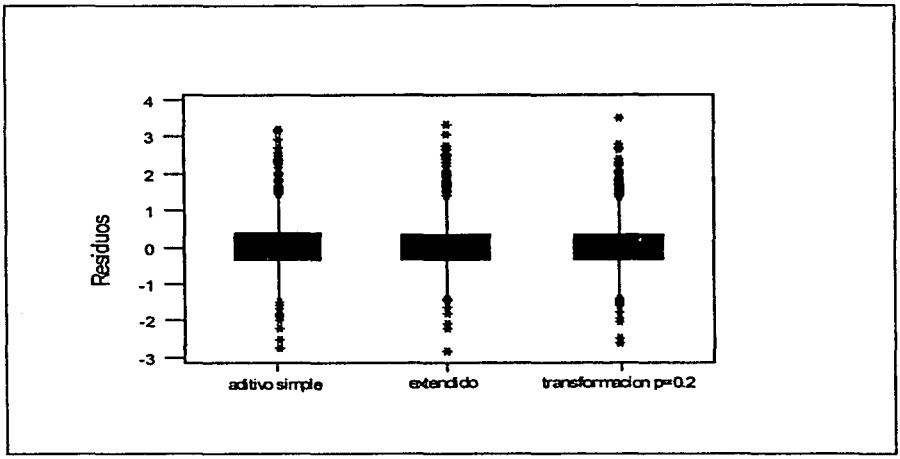
Para la selección del modelo siempre se tienen en consideración dos criterios los cuales se basan en el comportamiento exhibido por los residuos. Uno de ellos consiste en inspeccionar la forma de su distribución, para lo cual los diagramas de cajas paralelas son de gran utilidad. En este caso, aún cuando las diferencias entre los tres grupos no son muy grandes, sí se percibe que la caja correspondiente al modelo de datos transformados es un poco más chica y que las colas de esta distribución son ligeramente más simétricas que las de sus contrapartes. Ver figura 3.2 mostrada en la siguiente página.

El otro criterio es el del porcentaje (P) de variación absoluta que es explicada por los modelos, el cual se exhibe en la tabla 3.1 De nuevo las diferencias no son grandes, pero sigue siendo el modelo donde se transformaron los datos el que arroja la P más grande, lo que lo confirmaría como la mejor elección para describir la estructura de la tabla.

Tabla 3.1 Porcentaje de variación absoluta de los datos que alcanzan explicar los modelos

P	Modelo ajustado a la TDA		
	Aditivo simple	Extendido	Transformación P = 0.2
% de variación absoluta explicada por el modelo	56.3%	59.0 %	60.1%

Figura 3.2 Diagrama de cajas para los residuos de los modelos para la TDA



Pese a lo anterior, usar el modelo que se obtuvo tomándole raíz quinta a las tasas obliga a trabajar en una escala diferente a la original, y aunado al hecho de que no existe una gran diferencia en la calidad de ambos ajustes, se considera más conveniente optar por el modelo extendido.

Una vez seleccionado el modelo se procedió a revisar sus efectos ciudad, encontrándose que naturalmente existen diferencias en los niveles de desempleo abierto entre las ciudades, pero éstas se deben a circunstancias específicas de cada ciudad, más que a condiciones asociadas a los distintos niveles de desarrollo regional.

Por ejemplo, las ciudades que durante el periodo mostraron los niveles de desempleo abierto más altos son Tampico, México, Monterrey y Matamoros. Por

el contrario, las ciudades con los índices más bajos de desempleo abierto se encontraban Mérida, Ciudad Juárez, León y Tijuana. En cuanto a la localización geográfica de las ciudades que los conforman ambos grupos son muy heterogéneos. Ver tabla siguiente.

Tabla 3.2 Modelo ajustado a la TDA y los efectos ciudad

$TDA_{ij} = 2.6 + trimestre_i + ciudad_j + (0.3) \times (trimestre_i) \times (ciudad_j) + r_{ij}$					
$Ciudad_j$	**Efecto**	**$Ciudad_j$**	**efecto**	**$Ciudad_j$**	**efecto**
General	0.40	Matamoros	0.61	San Luis P.	-0.75
Cd Juárez	-1.40	Mérida	-1.20	Tampico	1.05
México	1.00	Monterrey	1.08	Tijuana	-1.57
Chihuahua	0.25	N. Laredo	-0.89	Torreón	0.25
Guadalajara	0.00	Orizaba	-0.60	Veracruz	0.16
León	-1.40	Puebla	-0.35		

En términos de la TDA por lo tanto, no parece ser factible detectar patrones regionales de desempleo en el país.

En el primer capítulo se mencionó que la tasa de ingresos insuficientes y desempleo (TIID), parece describir de manera más adecuada la situación de vivir en el desempleo: el individuo no tiene una fuente de ingresos que le permitan satisfacer sus necesidades primarias. Las personas con empleo que son consideradas en esta tasa, aunque sí tienen un ingreso, éste es insuficiente para subsistir. Por esta razón se pensó en replicar el análisis anterior usando este indicador y los hallazgos son muy interesantes.

En este caso sólo se presentan los resultados más relevantes, toda vez que el procedimiento para la construcción del modelo para este nuevo conjunto de datos es el mismo. Los datos se presentan en la tabla 1.3 del anexo A.

Para la TIID ocurre lo mismo que para la TDA en cuanto a la calidad de los modelos. El mejor es el que se obtiene mediante la transformación de los datos y el segundo es el extendido. Sin embargo se prefirió usar éste último dado que todos sus términos están expresados en las unidades originales. Ver la figura y tabla siguientes.

Figura 3.3 Diagrama de cajas de los residuos de tres modelos para la TIID

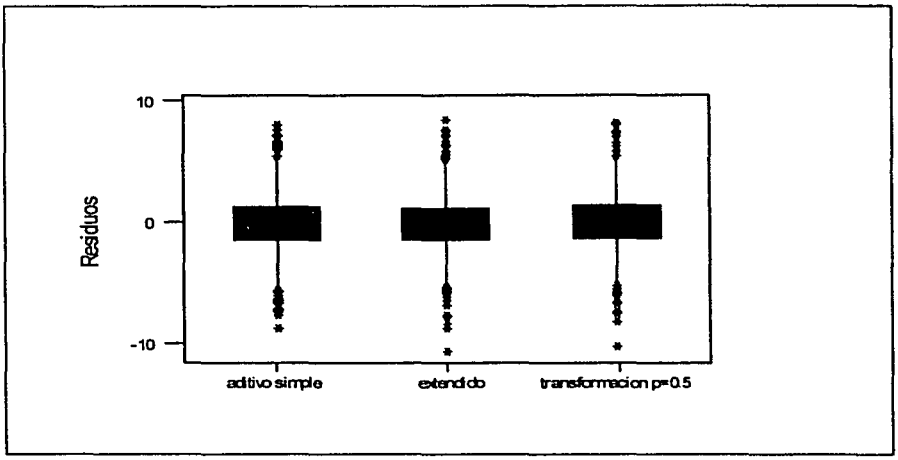


Tabla 3.3 Porcentaje de la variación absoluta de los datos que alcanzan explicar los modelos

P	Modelo ajustado a la TIID		
	Aditivo simple	Extendido	Transformación P = 0.5
% de la variación absoluta explicada por el modelo	61.0%	61.8 %	62.6%

El término común en este tipo de modelos generalmente carece de interés por sí mismo, ya que sólo representa el elemento con base en el cual se pueden distinguir las diferencias entre los efectos. No obstante, dado que en este caso se dispone de dos indicadores de desempleo a los que se les ajustó un mismo modelo es interesante notar la gran diferencia que hay en dicho término para ambas formas de abordar el fenómeno del desempleo.

A grandes rasgos se puede decir que el nivel de la TIID osciló alrededor del 11.4%, mientras que el de la TDA estuvo alrededor del 2.6%, lo cual representa una diferencia enorme (8.8 puntos porcentuales), pero más aún cuando se interpretan estos números en términos de seres humanos viviendo en condiciones realmente críticas.

Al ordenar los valores de los efectos ciudad, los cuales se exhiben en la tabla 3.4, se aprecia claramente que la TIID sí muestra un comportamiento diferenciado de acuerdo a la zona geográfica. En lo que respecta a este indicador la ciudad de Guadalajara sirve para dividir al grupo. Las ciudades con efectos de signo positivo son las que tuvieron la tasa con los niveles más altos. De manera equivalente, los efectos de signo negativo identifican a las zonas urbanas donde el fenómeno se presentó con menor intensidad.

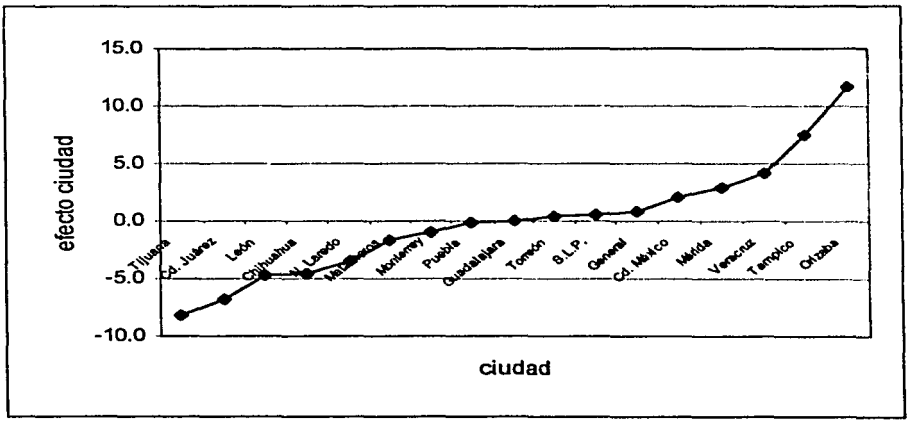
Tabla 3.4 Modelo ajustado a la TIID y sus efectos ciudad

$TIID_{ij} = 11.39 + trimestre_i + ciudad_j + (0.04) \times (trimestre_i) \times (ciudad_j) + r_{ij}$					
$Ciudad_j$	**Efecto**	**$Ciudad_j$**	**Efecto**	**$Ciudad_j$**	**Efecto**
General	0.80	Matamoros	-1.70	San Luis P.	0.54
Cd Juárez	-6.80	Mérida	2.85	Tampico	7.51
México	2.09	Monterrey	-0.95	Tijuana	-8.22
Chihuahua	-4.53	N. Laredo	-3.43	Torreón	0.40
Guadalajara	0.00	Orizaba	11.67	Veracruz	4.20
León	-4.73	Puebla	-0.20		

La figura 3.4 permite comprobar que existe un claro patrón en el comportamiento de este indicador de acuerdo a la región geográfica. Las ciudades del bajío y el norte del país, especialmente las ubicadas en la franja fronteriza, son las que tuvieron los porcentajes más bajos de su PEA en desempleo abierto o percibiendo ingresos inferiores al mínimo. Mientras que México, Mérida y las ciudades de la costa del golfo muestran niveles arriba del valor mediano (Guadalajara).

Es importante subrayar que el consolidado del total de las zonas urbanas cae también en el segundo grupo, y que hay dos ciudades que rompen con este esquema. Estas son Puebla que se comporta como las ciudades del norte y Torreón que queda clasificada dentro del grupo del sur-golfo.

Figura 3.4 Gráfica de los efectos ciudad del modelo extendido para la TIID

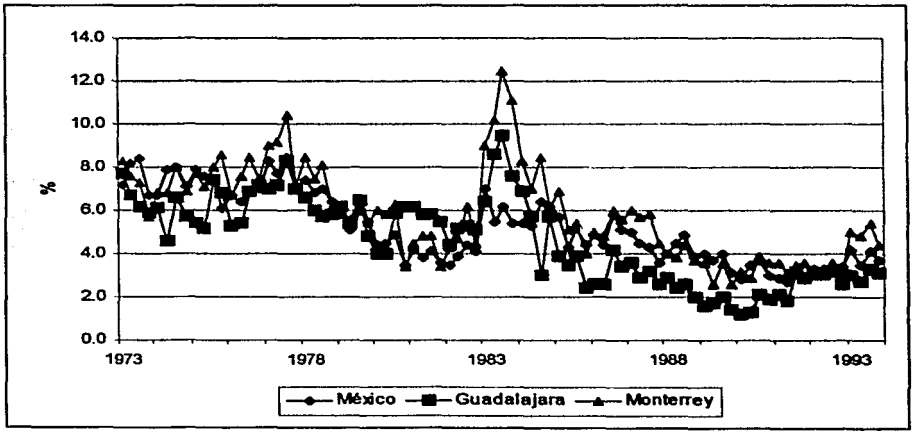


3.2 La convergencia de la TDA

Para este análisis se utilizó información trimestral de las ciudades de México, Guadalajara y Monterrey para el periodo comprendido entre 1973 a 1993. En este caso, la información fue tomada del trabajo de Roel[4].

Una primera aproximación de análisis consiste en construir la gráfica de los datos originales, figura 3.5. En términos generales el nivel de desempleo en las tres ciudades muestra un comportamiento muy parecido, sólo que con ritmos e intensidades diferentes. En comparación con las otras dos ciudades, Guadalajara parece responder con un trimestre de rezago a los fenómenos que hacen recaer o repuntar el nivel de desempleo. En Monterrey por otro lado, las fluctuaciones de la tasa son mucho más marcadas a las observadas en la Ciudad de México y Guadalajara, y de hecho su dispersión es mayor.

Figura 3.5 Tasa de desempleo trimestral para el periodo 1973 - 1993



[4] Roel Pavón, Jimena. Op. cit. Págs. 49 y 50.

Después de inspeccionar la gráfica anterior se entiende por qué Roel sospechó de la posible convergencia de las tasas. Sin tomar en cuenta ningún tipo de consideración económica, sería difícil descartar la posibilidad de que en algún momento en el futuro éstas lleguen a converger incluso a un mismo valor.

Sin embargo, eso no es lo que se desea probar. La hipótesis a verificar es que es posible encontrar una combinación lineal de las series que sea estacionaria. Del tipo de

$$\eta_1 x = \eta_0 + \eta_2 y + \eta_3 z + \varepsilon$$

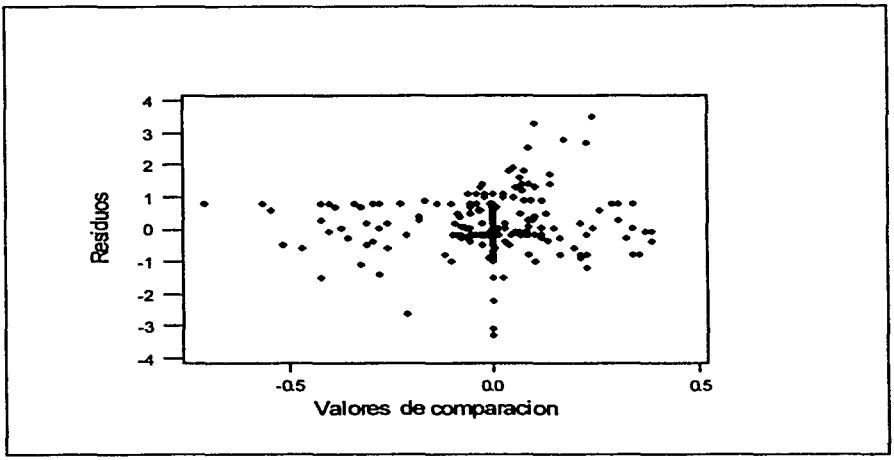
Una vez demostrado que las series son integradas de primer orden, I(1), para que estén cointegradas "no es necesario que los residuos (ε) sean puramente aleatorios, basta que sean un proceso estacionario en general".[5]

Los datos estudiados se presentan en la tabla 1.4 del anexo A. Después de ajustarle un modelo aditivo simple a los datos, se procedió a construir su gráfica de diagnóstico, fig. 3.6, la cual sugiere la existencia de una ligera interacción entre los factores, toda vez que su pendiente es distinta de cero (s=-0.19).

De lo anterior se desprende que la transformación que usa como exponente $p = 1 - s = 1.19$ podría ayudar a remover la no aditividad de los datos. La segunda alternativa es optar por el modelo extendido, el cual se obtiene agregando el término adicional cuyo coeficiente es $k = \frac{s}{m} = \frac{-0.19}{5.05} \approx -0.04$.

[5] Troncoso Viniegra, Alfredo. "Co-integración en series de tiempo". Inédita. México. Tesis presentada para aspirar al grado de Actuario. Instituto Tecnológico Autónomo de México. 1997. Pág. 47.

Figura 3.6 Gráfica de diagnóstico del modelo aditivo simple



Para seleccionar el modelo es muy útil construir el diagrama de cajas paralelas toda vez que permite comparar el comportamiento de los residuos. En él, se aprecia con facilidad que la transformación potencia efectuó un mejor trabajo para remover la no-aditividad de los datos del que hizo el término de interacción de los factores. Ver figura 3.7 de la siguiente página.

En la tabla 3.5 se muestra el indicador P del porcentaje de variación total explicada por cada uno de los tres ajustes. Aquí se confirma que es mediante la transformación de los datos como se obtienen los mejores resultados. Aún así, el modelo extendido tiene la enorme ventaja de trabajar en las unidades originales. En muchas situaciones esto puede llegar a cargar la balanza a favor de esta segunda alternativa.

Figura 3.7 Diagrama de cajas paralelas para los tres modelos ajustados

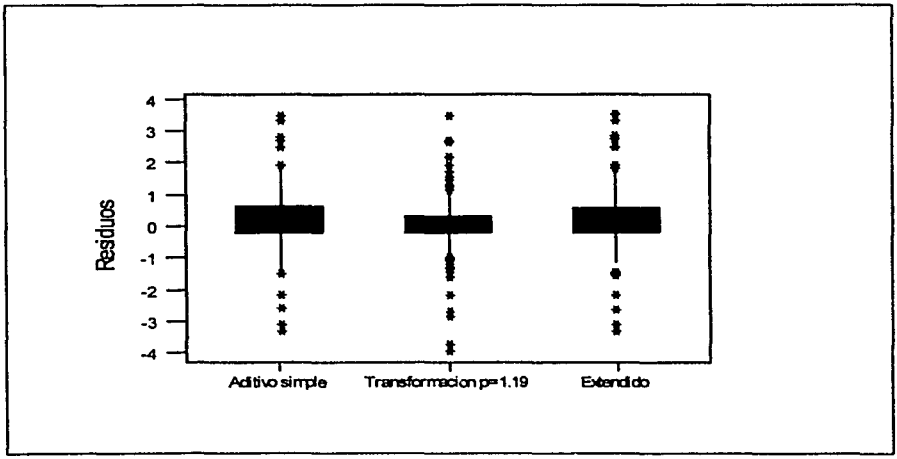


Tabla 3.5 Porcentaje de variación absoluta total explicada por el modelo

	Modelo Ajustado		
	Aditivo Simple	Con datos transformados	Extendido
P=	66.5%	70.3%	66.6%

Un criterio también importante está dado por el principio de parsimonia, el cual señala que en caso de disponer dos modelos que expliquen igual de bien un fenómeno, siempre resultará preferible trabajar con el que involucre el menor número de parámetros. Con esto en mente se decidió seleccionar finalmente al que usa la transformación potencia, aún cuando ésta sea poco común.

Bajo este modelo, el comportamiento de los datos para la tabla completa está dado por: $y_{ij}^{1,2} = 6.9 + \alpha_i + \beta_j + \varepsilon_{ij}$. Donde α_i = efecto tiempo y β_j = efecto ciudad.

La expresión que describe la estructura de la tabla para cada ciudad es:

$$v_i = 6.9 + \alpha_i + 0 = 6.9 + \alpha_i$$

$$z_i = 6.9 + \alpha_i - 1.2 = 5.7 + \alpha_i$$

$$w_i = 6.9 + \alpha_i + 0.3 = 7.2 + \alpha_i$$

Donde:

v_i = estimador para la TDA-Mex elevada a la potencia $p = 1.2$

z_i = estimador para la TDA-Gdj elevada a la potencia $p = 1.2$

w_i = estimador para la TDA-Mty elevada a la potencia $p = 1.2$

De estas últimas tres ecuaciones no es difícil verificar que se cumplen las siguientes igualdades:

$$v_i = \frac{1}{2} z_i + \frac{1}{2} w_i + 0.45$$

$$z_i = \frac{1}{2} v_i + \frac{1}{2} w_i - 1.35$$

$$w_i = \frac{1}{2} v_i + \frac{1}{2} z_i + 0.9$$

Que es en parte lo que se quería probar. Lo que faltaría por verificar es que si $TDAmx_i$, $TDAgd_i$ y $TDAmt_i$ representan las tasas de desempleo *observadas* en cada una de las tres ciudades elevadas a la potencia $p = 1.2$, los residuos definidos como

$$\varepsilon v_i = TDAmx_i - v_i$$

$$\varepsilon z_i = TDAgd_i - z_i$$

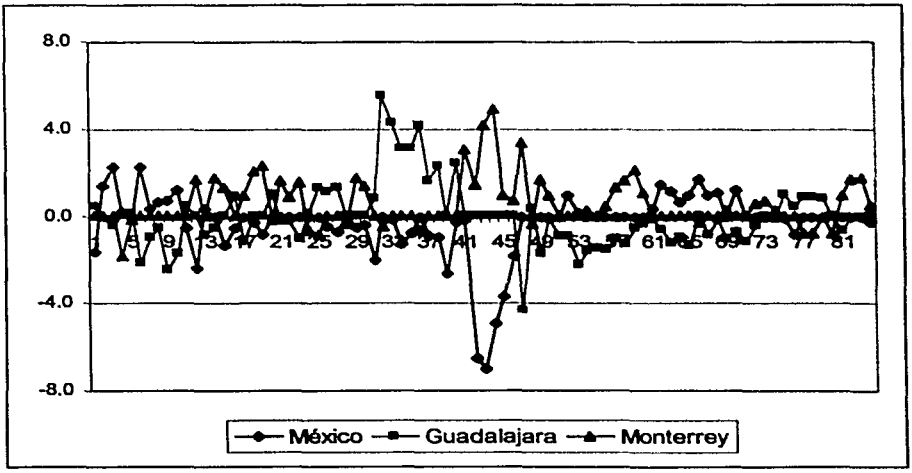
$$\varepsilon w_i = TDAmt_i - w_i$$

muestran un comportamiento estacionario.

En la figura 3.8 es posible apreciar que los residuos para las ciudades de México y Monterrey oscilan alrededor del cero, a excepción del periodo comprendido entre los puntos 41 y 50, que corresponden al tercer trimestre de 1982 y el primer trimestre de 1985, donde las tasas de México fueron sobrestimadas y para Monterrey fue subestimada. Ambas series no presentan tendencia, lo cual está de acuerdo con lo encontrado por Roel. Para estas dos ciudades por lo tanto, sí fue posible construir un modelo con las características deseadas.

El caso de Guadalajara presenta problemas porque existen dos periodos donde los residuos muestran tendencia, es decir donde la serie es no estacionaria. El primero es entre los puntos 32 al 46 donde ésta es decreciente. El segundo es del 47 al 75 cuando ésta es creciente. Para apoyar el diagnóstico de estacionariedad de estas series se presentan en el apéndice B de este documento las gráficas de la FAC (función de autocorrelación) y PAC (función de autocorrelación parcial).

Figura 3.8 Residuos de las combinaciones lineales que estiman las raíces de las tasas para México, Guadalajara y Monterrey.



3.3 Ajuste de un modelo para la TDA de doce grandes ciudades

En este ejemplo se tratará de identificar y ajustar un modelo para la TDA trimestral de las ciudades de México, Chihuahua, Guadalajara, León, Mérida, Monterrey, Orizaba, Puebla, San Luis Potosí, Tampico, Torreón y Veracruz para el periodo comprendido entre el primer trimestre de 1983 y el segundo de 1996. La matriz de datos se presenta en la tabla 1.5 del anexo A.[6]

Como en los casos anteriores, la estrategia para seleccionar el modelo consiste en empezar por ajustarle a los datos uno de tipo aditivo simple y revisar el comportamiento de sus residuos. En caso de ser necesario, buscar algún modelo alternativo que describa de mejor manera la estructura de la tabla.

Obtener los estimadores del primer modelo requirió de seis iteraciones del algoritmo. La pendiente de la gráfica de diagnóstico sugiere el uso de una transformación potencia (p=1-0.6=0.4) para remover la no-aditividad de los datos.

Sin embargo, el avance obtenido al transformar los datos no es muy significativo. El tamaño de los residuos de ambos ajustes es aproximadamente el mismo. La suma de sus valores absolutos es de 379 para el primero y de 371 para el segundo

La segunda alternativa consiste en incluir el término que intenta explicar la interacción de los efectos. El estimador para su coeficiente es $k = \frac{0.61}{3.25} \approx 0.19$.

Sin embargo, este modelo tampoco mejora de forma sustancial la calidad del ajuste. La suma de valores absolutos de sus residuos es de 373 aproximadamente.

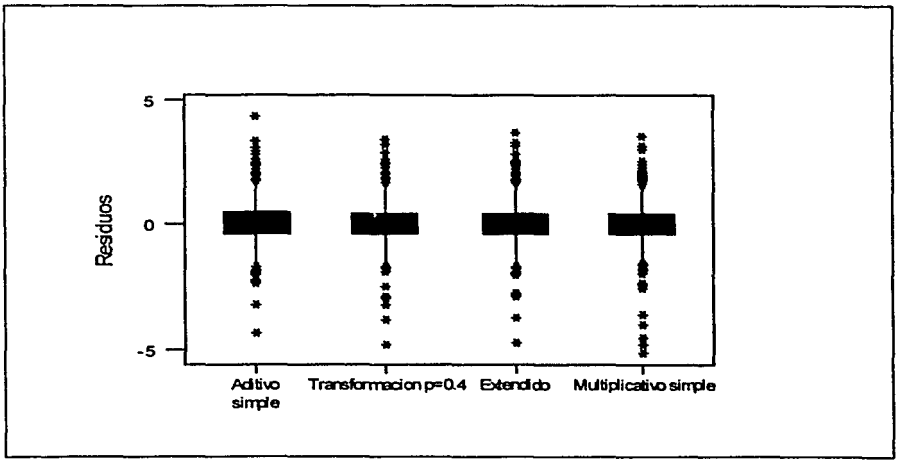
[6] Datos tomados de Morales Alvarez, Martha E. Op. cit. Págs. 3 y 4 Anexo I.

Dados los magros avances conseguidos hasta ahora, es conveniente intentar ajustar otro modelo aún a pesar de no disponer de información que lo sugiera. Transformar los datos usando la función logaritmo es una valiosa estrategia que se usa muy frecuentemente cuando se trabaja con series de tiempo económicas, entre otras razones por su reconocida efectividad para estabilizar la varianza.

Por esta razón, se decidió probar con un modelo multiplicativo simple. Pero contrario a lo esperado, se observa que arroja los residuos más grandes. Ver tabla 3.6 en la siguiente página.

Al revisar el diagrama de cajas paralelas para los cuatro ajustes, figura 3.9, se encontró que los residuos obtenidos al usar una transformación potencia son los que muestran el mejor comportamiento. Poseen las colas más simétricas y es este modelo el que genera el menor número de observaciones extremas. Por estas razones se decidió adoptarlo para explicar el comportamiento de los datos.

Figura 3.9 Diagrama de cajas paralelas de los residuos de los modelos ajustados



Al comparar la calidad del ajuste de este modelo con la de los construidos por Morales, se puede verificar en la tabla 3.6, que este ajuste se compara ventajosamente con dos de los tres propuestos por ella y es prácticamente igual de bueno que el tercero. Es importante subrayar el hecho de que cualquiera de los modelos generados usando el pulido de medianas, posee una R^2 (porcentaje de la varianza explicada por el modelo) que es prácticamente el doble a la alcanzada por dos de sus contrapartes de series de tiempo.

Tabla 3.6 Variabilidad de los datos que alcanzan a explicar los distintos modelos

Modelo	Ajustados con el pulido de medianas				Ajustados con series de tiempo		
	Aditivo simple	Con transformación	Extendido	Multiplicativo simple	TxCxExI	TxExI t. lineal	TxExI t parabólica
$\Sigma_i\Sigma_j\|r_{ij}\| =$	379.3	370.7	373.0	382.1	-	-	-
$P =$	60.3%	61.2%	61.0%	60.0%	-	-	-
$R^2 =$ (*)	80.8%	81.8%	81.5%	78.8%	41.9%	-	-
$R^2 =$ (**)	79.0%	79.9%	79.7%	77.0%	-	42.2%	82.4%

(*)- Para el periodo comprendido entre Ene-Mar '83 a Oct-Dic '95
(**)- Para el periodo comprendido entre Jul-Sep '83 a Oct-Dic '95

Adicionalmente existe un detalle que es importante resaltar, el cual está relacionado a la gran sensibilidad de los modelos de Morales a los datos extremos. Para mostrarlo se procedió a identificar ciudad por ciudad, los periodos de mayor desempleo abierto y sustituir sus valores observados por los esperados según el modelo aditivo simple ajustado a la tabla.

Con estas nuevas series se recalculó, a manera de ejemplo, el modelo TxCxExI para las doce ciudades con la intención de averiguar en qué medida cambian los resultados una vez que se han eliminado las observaciones que podrían considerarse como atípicas.

Los resultados de este ejercicio son muy interesantes. Como puede observarse en la tabla 3.7 y a pesar de que el máximo de observaciones sustituidas para cada ciudad fue tres, los resultados del modelo, medidos en términos de R^2, mejoraron sustancialmente en la mayoría de los casos. Los cambios más

dramáticos ocurren para las ciudades de Puebla y Chihuahua, donde el porcentaje de varianza que el modelo alcanza a explicar aumenta en 156 y 131% respectivamente. Otras ciudades donde se observa un fuerte incremento en su R^2 son León, con +82% y Veracruz con +89%.

Tabla 3.7 Error cuadrático medio y R^2 para los dos conjuntos de datos

Modelo T x C x E x I					
Ciudad	**Observaciones remplazadas**	**ECM (a)**	**ECM(b)**	**R^2 (a)**	**R^2 (b)**
México	95III / 95IV / 96I	81	67	25.6	38.6
Chihuahua	84III / 95II	205	170	11.3	26.2
Guadalajara	83II / 83III / 95III	137	115	24.4	36.3
León	95II / 95III	40	35	14.8	27.0
Mérida	83III	72	69	22.5	25.7
Monterrey	83II / 83III / 95III	165	129	27.8	43.6
Orizaba	84III / 84IV	46	42	41.8	46.5
Puebla	85I / 95II / 95III	64	52	10.1	26.0
San Luis Potosí	83II	50	49	39.6	39.6
Tampico		78	-	19.3	-
Torreón	83III	131	120	12.5	20.0
Veracruz	84IV	67	59	11.1	21.1

(a)- Series con los datos originales (b)- Series sustituyendo datos

3.4 Predicción de la TDA usando el pulido de medianas

Siempre que se estudia el comportamiento de una variable económica surge la inquietud por conocer su comportamiento futuro en el corto plazo. En este sentido el caso del desempleo abierto no podía ser diferente. Para abordar este problema los econometristas suelen desarrollar modelos de series de tiempo como el usado por Alvarez, donde se propone descomponer las series en sus componentes de tendencia y de variaciones estacional, cíclica y aleatoria.

Otra aproximación y sin duda, la más socorrida debido a su probada efectividad para elaborar pronósticos, es la propuesta por Box & Jenkins que consiste en usar modelos autorregresivos. Pero para poder identificar el modelo más adecuado para una serie usando este enfoque se requiere de cierto nivel de conocimiento técnico y de experiencia en su manejo.

Con el único propósito de mostrar el poder de las herramientas resistentes aquí presentadas, se propone usar el PM para elaborar pronósticos de corto plazo para la TDA. En este ejemplo se usarán 128 observaciones de la TDA mensual, a partir de enero de 1990, de lo que INEGI define como el total de zonas urbanas, y con ello se pronosticarán los últimos cuatro meses del año 2000.

Para este ejercicio se optó por seguir la estrategia sugerida por Besag[7] debido a su sencillez, en lugar de la propuesta por Stoto y Wong[8]. En la aproximación seguida aquí los datos se organizan en forma matricial donde las filas representan a los años y las columnas a los meses.

Conviene señalar que en este caso fue necesario usar la aplicación del pulido de medianas del Minitab en lugar de usar el programa desarrollado en C++, dado que los resultados arrojados por este último se ven seriamente afectados por la ausencia de información para el tercer cuatrimestre del 2000.

Para ajustarle un modelo aditivo simple a los datos fue necesario iterar en cuatro ocasiones el algoritmo. La recta resistente de la gráfica de diagnóstico de los residuos tiene una pendiente igual a 0.57, lo cual sugiere transformar los datos con una $p = 1 - 0.57 = 0.43$ para remover la no-aditividad. Como se ha venido haciendo, el segundo camino consiste en probar con un modelo extendido, el cual mejora ligeramente lo alcanzado por el aditivo simple.

Una tercera opción es el modelo multiplicativo, a pesar de que la gráfica de diagnóstico no sugería su utilidad. De manera sorprendente se encontró que en términos de porcentaje de variación absoluta este último es el que mejor se ajusta a los datos, aunque sólo por un margen muy pequeño.

---

[7] Besag, Julian. "On Resistant Techniques and Statistical Analysis". Biometrika, 68. 2. 1981. Págs.463-469.

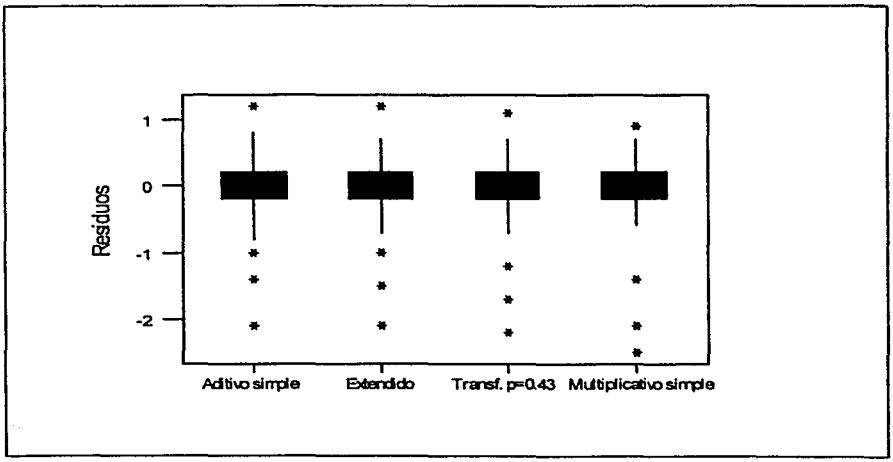
[8] Stoto, Emerson y Wong, George Y. Op.cit. Págs. 75-79. En realidad, en este caso se extendería la idea esbozada por ellos y en lugar de pensar en estimar valores faltantes en el cuerpo de la tabla, se usaría el conocimiento de la estructura de la tabla para elaborar pronósticos de valores de celdas inexistentes.

Tabla 3.8 Porcentaje de la variación absoluta de los datos que alcanzan explicar los modelos

	Modelo ajustado a la TDA mensual 90-00			
	Aditivo simple	Extendido	Transformación P = 0.43	Multiplicativo
% de la variación absoluta explicada por el modelo	70.5%	71.4 %	71.0%	71.7 %

El diagrama de cajas paralelas muestra asimismo, que sus residuos se comportan por lo menos igual de bien que los de los otros.

Figura 3.10 Diagrama de cajas paralelas de residuos de los modelos ajustados



Por otro lado, con la ayuda del programa X-12-ARIMA[9] se encontró que el modelo ARIMA(2 1 2)(0 1 1) sirve para obtener los pronósticos para los meses de septiembre a diciembre del 2000.

En la tabla 3.9 se muestra los resultados de comparar los pronósticos arrojados por cada modelo versus los observados para ese periodo. Se observa que los cinco modelos tienden a subestimar el valor de la tasa en el cuatrimestre, con excepción del mes de octubre cuando la tasa es sobrestimada.

En la siguiente tabla se presenta un comparativo de la suma de residuos absolutos de los cinco modelos.

Tabla 3.9 Valores observados y pronosticados de la TDA para Sep-Dic 2000.

Tasa de desempleo abierto (total áreas urbanas)						
Periodo	Datos Observados	Aditivo Simple	Extendido	Transfor-Mación	Multiplicativo simple	ARIMA
Sep '00	2.51	2.28	2.28	2.27	2.25	2.34
Oct '00	1.97	2.30	2.29	2.27	2.21	2.54
Nov '00	2.00	1.78	1.86	1.89	1.92	1.89
Dic '00	1.90	1.60	1.71	1.72	1.77	1.74
Suma de residuos absolutos		1.07	0.88	0.83	0.71	1.01

Aquí se puede comprobar que tres de los modelos que fueron ajustados usando el PM arrojan pronósticos que se aproximan más a los datos observados de lo que se consigue al usar el modelo ARIMA.

[9] X-12-ARIMA Monthly Seasonal Adjustment Method, Oficina de Censos, Departamento de Comercio, E.U.

CAPÍTULO 4

Conclusiones

Las conclusiones del presente trabajo son de dos tipos. Por un lado están las que se desprenden de la revisión del comportamiento de la tasa de desempleo abierto per se. Cabe aclarar que se deja para los especialistas el análisis detallado de las implicaciones económicas y sociales de los resultados aquí obtenidos dado que eso queda fuera del alcance y de los objetivos de este trabajo. Por otro lado están las relacionadas con la metodología de análisis de datos propuesta. Son éstas últimas en las que se pondrá mayor énfasis.

El análisis del desempleo abierto de dieciséis grandes zonas urbanas mostró que de enero de 1989 a diciembre del 2000, el comportamiento de las tasas presentó notables diferencias por ciudad; aunque éstas son más bien atribuibles a las circunstancias particulares de cada urbe.

Por el contrario, la tasa de ingresos insuficientes y desempleo confirmó la existencia de fuertes diferencias regionales. De las ciudades analizadas, son las

ubicadas en el norte y en la franja fronteriza las que muestran los porcentajes más bajos de PEA en desempleo o con ingresos inferiores al mínimo. En el otro extremo quedan comprendidas la ciudad de México, Mérida y las tres situadas en la costa del golfo: Tampico, Orizaba y Veracruz.

Es importante subrayar que también se encontraron fuertes diferencias en los niveles observados de la TIID en las ciudades. Entre Orizaba la ciudad con mayores problemas y Tijuana, que presentó la tasa más baja de las ciudades estudiadas, media una diferencia de casi veinte puntos porcentuales.

La fuerte presencia de la industria maquiladora en el norte del país y el tipo de producción que se realiza allá, debe ser un factor relevante para estas diferencias. Por el contrario, en las otras ciudades los trabajadores enfrentan condiciones de empleo mucho más críticas, y ante la falta de alternativas tienen que conformarse con realizar actividades productivas muy mal remuneradas.

En el segundo ejemplo de aplicación se encontró que el estudio del comportamiento de la TDA de México, Guadalajara y Monterrey usando la metodología de análisis aquí propuesta, arroja resultados que apoyan la conclusión de J. Roel sobre la cointegración de las tasas de esas ciudades.

Lo anterior no implica sin embargo, que durante el periodo comprendido de 1973 a 1993 las tres urbes hubieran experimentado un desarrollo equilibrado. Por el contrario, como lo advierte la misma Roel en su trabajo "otros indicadores de bienestar como educación y salud, infraestructura y grado de industrialización, sugieren que este desarrollo ha sido heterogéneo". [1]

En este sentido el reto para el Gobierno de México es el implementar políticas efectivas que promuevan un desarrollo regional equilibrado. La aplicación de planes de promoción del empleo localizados en ciudades específicas ha

[1] Roel Pavón, Jimena. Op. cit. Pág. 44

demostrado generar únicamente flujos migratorios que a la larga tienen el efecto contrario al esperado.

Desde otro punto de vista, este estudio también mostró que a nivel exploratorio, el análisis resistente de tablas de doble entrada puede ser de gran utilidad al abordar el problema de la cointegración de variables económicas. Su aplicación más importante consistiría en proponer las ecuaciones que describen la relación de equilibrio, es decir, las combinaciones lineales de las variables en estudio.

En el tercer ejemplo presentado en el capítulo anterior, se vio que la R^2 de cualquiera de los modelos arrojados por el PM fueron casi tan altas como la del mejor de los ajustes que obtuvo Martha Morales al usar un enfoque de series de tiempo.

Con objeto de mostrar la alta sensibilidad a la presencia de observaciones extremas que tienen el tipo de modelos usado por Morales, siguiendo la metodología que ella propone, se ajustó uno de ellos a un conjunto de datos en donde los valores extremos habían sido sustituidos por otros que se acercaban más al comportamiento general del conjunto. La calidad de ajuste del modelo mejoró de manera sensible en la mayoría de las ciudades, pero de manera sustancial en cuatro de ellas.

Una característica que tienen la mayoría de las herramientas clásicas es que dependen del cálculo de la media y la desviación estándar de los datos, y en muchos casos también requieren que éstos cumplan con algunos supuestos distribucionales. El problema de trabajar con estas dos medidas es que son extremadamente sensibles a la presencia de observaciones atípicas, las cuales aparecen con mucha frecuencia en los lotes de datos en estudio.

Lo anterior da como resultado que con mucha frecuencia se corre el riesgo de que los datos atípicos provoquen la violación de algún supuesto, lo cual pondría en

duda la validez de los resultados. O bien, que sin llegar a ese extremo, su presencia pueda modificar de manera importante el comportamiento aparente del conjunto completo y por lo tanto, que las conclusiones obtenidas de su análisis puedan resultar si bien no necesariamente incorrectas, sí por lo menos inexactas.

Es precisamente en este contexto donde se aprecia con mayor claridad la importancia de usar medidas o técnicas resistentes. Y es justo ahí donde reside la gran utilidad del análisis exploratorio de datos, el cual aprovecha la mucho menor sensibilidad de la mediana a la presencia de observaciones extremas.

Finalmente, en el último ejemplo de aplicación se usó el PM para generar pronósticos para la TDA de los últimos cuatro meses del año 2000, y se mostró que al compararlos con los datos reales éstos compiten ventajosamente con los de un modelo de tipo ARIMA. Un resultado similar a éste lo obtuvo Besag, cuando comparó los pronósticos para ciertos datos de ventas arrojados por un modelo Box & Jenkins versus los que obtuvo él usando el PM.[2]

De ninguna manera se pretende sugerir la aplicación del PM para el pronóstico de series de tiempo. Por muchas razones los modelos ARIMA son la mejor herramienta de que se dispone para este fin. El ejemplo tuvo como objetivo exhibir la gran efectividad del PM para identificar y describir la estructura que subyace en una tabla de doble entrada, aún incluso cuando ésta tenga observaciones faltantes.

En efecto, la estrategia seguida para predecir algunos valores de la TDA consistió simplemente en organizar la información en forma de tabla, dejando los datos a pronosticar en la última fila. Para el PM dichas observaciones fueron consideradas como "hoyos" de la tabla. Los pronósticos obtenidos pueden entenderse como los valores esperados de las celdas sin información, los cuales pueden ser usados para "completar" la tabla si fuera necesario.

[2] Besag, Julian. Op.cit. Págs. 467-468

En comparación con la idea de los pronósticos, lo anterior representa una aplicación menos pretenciosa, aunque no menos útil. En muchos casos y por diversas razones no se puede contar con la información de la variable observada para todas las entradas de la tabla. Ya sea porque en realidad no se dispone de ella, o bien porque se trata de observaciones atípicas que es preferible manejarlas como datos faltantes en lugar de dejarlas y correr el riesgo de distorsionar de alguna manera los resultados del análisis.

Conclusiones y recomendaciones finales

La revisión del concepto de la tasa de desempleo abierto y de los niveles que ésta ha alcanzado a lo largo de las últimas tres décadas en las principales zonas urbanas del país, pone en evidencia su incapacidad para describir de manera realista la situación del problema en México.

La sola difusión de la TDA y el poco cuidado que se tiene en su interpretación, existe una marcada tendencia a olvidar la cobertura geográfica de la ENEU, demerita el esfuerzo que realiza el INEGI para informar sobre la evolución de una variable clave para evaluar el desempeño de la economía de cualquier región o país.

La TDA por sí misma no sirve para apoyar procesos de planeación y toma de decisiones serios. Incluso se ha llegado a utilizar para abordar el tema del desempleo de forma demagógica. Debido a lo anterior, es importante motivar la divulgación y el uso de los indicadores complementarios los cuales, en conjunto, proporcionan un panorama que se aproxima mucho más a la situación real del problema.

Otro de los objetivos del presente trabajo era mostrar la utilidad de analizar tablas de doble entrada usando el PM. A lo largo de este documento se han resaltado

las ventajas y desventajas que representa el trabajar con esta aproximación en lugar del enfoque clásico de análisis que se basa en la media.

Entre sus ventajas se encuentran:

1. El enfoque de análisis aquí propuesto permite ajustar, prácticamente a cualquier tabla de doble entrada con una variable observada y dos de clasificación, un modelo del tipo $y_{ij} = \mu + \alpha_i + \beta_j + \nu\varphi\lambda_j + \rho_{ij}$. El único requisito es que la tabla conste de por lo menos 3 columnas y/o renglones.
2. La validez de sus resultados no depende de que la variable observada ni los residuos del modelo cumplan con ningún tipo de supuesto distribucional.
3. Dado que todas las herramientas utilizadas se basan en la mediana, los modelos ajustados son altamente resistentes. Esto es, insensibles a la presencia de observaciones atípicas y/o de observaciones faltantes en la tabla de datos.

De sus desventajas es importante mencionar:

1. No arroja un ajuste o modelo único.
2. Dicho modelo tampoco es óptimo necesariamente.
3. Aunque su aplicación es bastante simple, ajustar un modelo a un conjunto de datos requiere de la iteración del algoritmo, mientras que con el análisis basado en medias en una sola oportunidad se obtienen los estimadores para todos los términos del modelo. Las tablas aquí analizadas en muchas ocasiones requirieron de seis o más iteraciones.

La metodología propuesta es extremadamente flexible y puede ser utilizada en muchas circunstancias y con objetivos muy diversos. Asimismo, aquí se mostró que los modelos obtenidos usando el PM llegan a competir incluso de forma ventajosa frente a los tradicionales en términos de precisión o calidad del ajuste.

En este sentido cabe señalar la conveniencia que en una futura investigación se validen estos hallazgos usando técnicas de simulación. En otras palabras, se propone que se generen diversas tablas partiendo de modelos conocidos por el investigador, de hecho definidos por él, las cuales serían analizadas usando tanto la aproximación resistente como la clásica con objeto de determinar cual de las dos es capaz de describir con mayor precisión el comportamiento de los datos.

Por último, es importante no perder de vista que el PM, así como el resto de las técnicas expuestas en este documento, pertenecen a un conjunto de aplicaciones que en buena medida desarrollara John W. Tukey y que bautizó con el nombre de análisis *exploratorio* de datos.

En efecto, el uso exclusivo del enfoque de análisis aquí propuesto, o bien su aplicación como complemento de la aproximación clásica, debe depender en cada caso de la naturaleza de los datos y del objetivo del estudio. La decisión siempre quedará naturalmente en manos del investigador.

ANEXO A
Tablas de datos

TABLA 1.1 TASA TRIMESTRAL DE DESEMPLEO ABIERTO EN DIECISEIS CIUDADES PARA EL PERIODO DE 1989 A 2000

Periodo		General	Cd. Juárez	México	Chihuahua	Guadalajara	León	Matamoros	Mérida	Monterrey
1989	I	3.2	0.7	4.0	1.6	1.6	0.9	2.7	0.6	3.6
	II	3.0	0.5	3.7	1.4	1.7	0.9	2.5	0.4	2.6
	III	3.3	0.8	4.0	2.1	2.0	0.8	2.9	0.8	3.6
	IV	2.5	1.7	3.1	1.8	1.4	0.8	2.7	0.6	2.6
1990	I	2.5	2.0	2.8	2.7	1.2	1.4	3.1	0.6	3.2
	II	2.8	1.2	3.5	2.3	1.3	1.4	2.8	0.4	2.9
	III	3.1	1.5	3.7	2.2	2.1	1.3	3.3	1.4	3.9
	IV	2.6	2.3	3.0	2.6	1.9	1.0	3.8	0.9	3.6
1991	I	2.7	2.1	2.9	2.1	2.1	1.1	5.3	0.6	3.6
	II	2.3	0.8	2.7	2.1	1.6	0.9	3.6	0.4	2.1
	III	2.9	1.0	3.2	2.5	3.2	1.4	4.6	1.3	3.5
	IV	2.6	0.8	2.8	1.9	2.9	0.7	4.3	1.0	3.6
1992	I	2.9	0.7	3.3	1.8	3.1	1.0	5.6	0.9	3.0
	II	2.8	1.0	3.3	1.7	3.1	0.7	4.6	1.3	3.0
	III	2.9	1.1	3.4	1.9	3.2	1.4	5.1	1.2	3.6
	IV	2.7	1.0	3.4	2.0	2.6	0.8	4.2	1.4	3.1
1993	I	3.5	2.4	4.2	3.7	3.0	1.0	6.0	1.3	5.0
	II	3.2	2.6	3.5	3.3	2.7	1.5	4.7	1.1	4.8
	III	3.7	2.9	4.1	3.7	3.3	1.7	4.4	1.4	5.4
	IV	3.3	1.8	3.7	5.1	3.1	1.4	6.1	1.1	4.4
1994	I	3.7	1.7	4.3	5.8	3.6	1.3	7.1	1.3	4.7
	II	3.6	2.0	4.2	5.4	3.1	1.8	6.3	1.4	4.9
	III	3.9	2.9	3.8	4.9	3.8	2.3	6.4	2.0	5.2
	IV	3.6	1.9	3.9	5.8	3.1	1.6	5.0	2.3	5.5
1995	I	5.1	2.5	5.6	7.5	6.2	3.2	6.5	3.6	6.0
	II	6.3	2.5	6.9	7.7	6.9	5.6	4.3	3.2	8.4
	III	7.4	3.6	8.6	7.0	7.6	4.8	5.4	5.0	10.2
	IV	6.1	2.5	7.4	5.9	6.0	2.5	4.7	4.3	7.3
1996	I	6.2	3.7	7.6	5.4	5.7	2.5	5.4	3.3	6.7
	II	5.6	2.3	7.3	5.7	5.1	1.9	4.1	3.5	5.7
	III	5.5	2.5	7.0	5.2	5.0	1.8	3.9	2.8	6.8
	IV	4.7	2.0	5.6	4.1	4.4	1.5	3.8	3.3	5.5
1997	I	4.3	2.4	4.8	4.1	3.8	1.0	3.7	2.8	5.0
	II	3.9	1.8	5.0	4.5	3.4	1.2	2.5	2.1	3.9
	III	3.7	1.4	4.5	4.2	3.1	1.3	2.8	2.2	3.8
	IV	3.1	1.4	3.6	2.9	2.7	1.1	2.4	1.9	3.0
1998	I	3.5	1.2	4.4	3.0	3.2	1.3	2.9	1.4	3.4
	II	3.2	0.7	4.1	3.4	3.1	1.2	3.0	1.7	2.8
	III	3.2	1.0	3.8	2.6	2.7	1.2	4.0	1.6	3.3
	IV	2.8	0.7	3.7	2.7	2.2	0.9	2.2	1.3	2.9
1999	I	2.9	0.5	3.6	2.4	2.4	1.1	2.6	1.6	3.0
	II	2.6	0.6	3.3	1.9	2.5	1.3	2.9	1.6	2.5
	III	2.3	0.8	2.8	2.6	1.9	0.9	1.9	1.4	1.7
	IV	2.2	0.9	2.8	2.3	1.5	0.8	1.7	1.2	1.7
2000	I	2.3	0.6	2.9	2.2	1.7	1.2	2.2	1.1	2.4
	II	2.2	0.8	2.7	2.0	1.9	1.1	2.2	1.3	2.1
	III	2.4	0.9	2.9	1.8	2.4	1.3	2.3	1.2	2.2
	IV	2.0	0.8	2.3	2.0	1.6	1.0	2.0	0.9	1.8

TABLA 1.1 TASA TRIMESTRAL DE DESEMPLEO ABIERTO EN DIECISEIS CIUDADES PARA EL PERIODO DE 1989 A 2000 (Cont.)

Periodo		Nuevo Laredo	Orizaba	Puebla	San Luis Potosí	Tampico	Tijuana	Torreón	Veracruz
1989	I	0.8	1.9	2.1	1.6	2.5	1.7	2.7	2.4
	II	1.1	1.9	1.5	2.0	3.3	0.9	2.4	2.2
	III	0.8	2.3	2.0	1.6	2.9	1.6	2.3	2.6
	IV	1.5	1.4	1.7	1.7	2.8	1.0	1.7	1.9
1990	I	1.6	2.0	1.8	1.3	2.6	1.4	2.4	2.2
	II	2.2	1.5	1.5	1.4	3.2	1.3	3.0	1.5
	III	2.1	1.8	1.7	2.0	3.3	2.8	3.3	2.4
	IV	1.5	2.0	2.2	1.3	3.4	1.1	2.1	1.8
1991	I	1.8	1.9	2.5	0.8	5.3	1.9	2.6	2.5
	II	1.2	0.7	1.5	1.4	4.4	1.4	2.0	2.9
	III	1.6	1.1	2.6	1.4	5.9	1.0	2.4	3.2
	IV	1.0	0.8	1.4	1.0	6.3	1.1	2.0	2.2
1992	I	0.9	1.1	2.0	1.2	5.9	1.8	2.6	2.3
	II	1.1	0.8	2.2	1.3	5.3	0.7	2.4	2.6
	III	1.9	1.1	2.8	1.2	4.9	0.6	2.5	2.6
	IV	1.3	1.7	2.4	1.0	4.9	0.6	2.0	2.5
1993	I	1.4	2.3	1.9	1.5	5.7	1.8	3.2	3.1
	II	1.2	1.7	1.9	1.4	5.0	1.8	3.0	3.0
	III	1.8	1.8	2.4	1.7	5.3	1.7	5.0	4.0
	IV	1.4	2.0	2.1	1.9	5.8	1.1	4.6	4.0
1994	I	2.9	1.9	2.6	1.6	5.9	1.1	5.5	4.8
	II	3.0	1.9	1.7	1.8	5.7	1.6	5.3	3.4
	III	2.7	2.0	2.9	3.0	5.9	1.1	6.3	4.4
	IV	1.8	1.7	3.0	1.9	4.6	1.1	5.7	3.5
1995	I	4.5	1.9	3.9	3.2	7.3	1.8	6.6	4.4
	II	5.9	3.4	6.7	4.3	7.6	1.8	6.7	5.4
	III	4.3	4.4	6.2	4.9	6.2	1.9	6.9	6.3
	IV	3.7	4.0	4.8	3.3	5.7	1.5	5.9	5.1
1996	I	3.6	4.5	4.0	4.6	6.2	1.3	5.9	5.2
	II	2.6	3.6	3.9	3.7	6.6	1.2	4.7	4.5
	III	3.3	3.4	3.6	3.8	5.9	1.8	4.8	3.9
	IV	2.7	3.5	3.4	3.0	4.0	1.4	4.0	3.7
1997	I	3.1	3.8	3.6	2.7	4.6	1.6	3.7	3.3
	II	2.4	2.1	2.6	2.6	3.2	0.8	3.0	3.1
	III	2.8	3.1	2.0	2.3	3.7	1.1	3.4	2.8
	IV	2.9	3.0	2.9	2.4	3.4	1.2	3.1	3.1
1998	I	2.1	4.1	2.8	2.8	3.4	1.0	3.2	3.6
	II	2.4	3.2	2.2	2.1	3.7	1.4	2.9	2.5
	III	2.1	2.9	2.1	2.3	4.2	1.0	3.2	2.9
	IV	1.5	2.9	1.9	1.9	2.6	0.8	2.3	2.9
1999	I	1.4	3.1	2.1	1.8	2.7	1.2	3.0	3.0
	II	1.5	2.5	1.7	1.8	2.6	0.5	2.2	2.6
	III	1.3	2.4	2.5	1.5	1.9	0.9	2.8	2.1
	IV	1.1	2.2	2.0	2.2	2.3	1.2	2.1	2.4
2000	I	1.4	2.4	2.8	1.4	3.0	1.0	2.0	2.2
	II	0.9	1.7	2.9	1.7	1.9	1.2	1.6	2.8
	III	0.7	2.7	2.2	1.9	2.1	1.1	2.1	2.8
	IV	1.3	2.0	1.7	1.8	1.3	1.0	1.8	2.9

TABLA 1.2 ESTIMADORES PARA EL TERMINO COMUN Y LOS EFECTOS CIUDAD PARA LA TDA DE DIECISEIS CIUDADES PARA TRES MODELOS DIFERENTES

(a) Modelo aditivo simple

Efecto ciudad	General	Cd. Juárez	México	Chihuahua	Guadalajara	León	Matamoros	Mérida	Monterrey
	0.40	-1.40	1.00	0.25	0.00	-1.40	0.61	-1.20	1.08
Efecto ciudad	**Nvo. Laredo**	**Orizaba**	**Puebla**	**San Luis P.**	**Tampico**	**Tijuana**	**Torreón**	**Veracruz**	
	-0.89	-0.60	-0.35	-0.75	1.05	-1.57	0.25	0.16	
Termino común	2.61								

(b) Modelo usando transformación potencia (p=0.2)

Efecto ciudad	General	Cd. Juárez	México	Chihuahua	Guadalajara	León	Matamoros	Mérida	Monterrey
	0.04	-0.17	0.08	0.02	0.00	-0.17	0.06	-0.14	0.08
Efecto ciudad	**Nvo. Laredo**	**Orizaba**	**Puebla**	**San Luis P.**	**Tampico**	**Tijuana**	**Torreón**	**Veracruz**	
	-0.10	-0.06	-0.03	-0.07	0.09	-0.19	0.02	0.02	
Termino común	1.20								

(c) Modelo extendido

Efecto ciudad	General	Cd. Juárez	México	Chihuahua	Guadalajara	León	Matamoros	Mérida	Monterrey
	0.40	-1.40	1.00	0.25	0.00	-1.40	0.60	-1.20	1.07
Efecto ciudad	**Nvo. Laredo**	**Orizaba**	**Puebla**	**San Luis P.**	**Tampico**	**Tijuana**	**Torreón**	**Veracruz**	
	-0.88	-0.60	-0.35	-0.75	1.05	-1.60	0.25	0.16	
Termino común	2.61								

TABLA 1.3 TASA TRIMESTRAL DE INGRESOS INSUFICIENTES Y DESEMPLEO EN DIECISEIS CIUDADES PARA EL PERIODO DE 1989 A 2000

Periodo		General	Cd. Juárez	México	Chihuahua	Guadalajara	León	Matamoros	Mérida	Monterrey
1989	I	21.2	7.8	24.2	7.8	16.1	16.2	10.7	17.4	16.3
	II	17.9	5.6	21.2	5.3	12.5	12.5	7.6	13.4	11.5
	III	16.5	6.1	18.7	5.9	14.5	10.7	6.0	13.1	13.1
	IV	17.4	8.2	19.9	6.0	15.4	11.6	6.6	12.6	12.1
1990	I	17.7	8.3	20.6	7.7	14.2	10.5	7.5	11.9	11.8
	II	14.4	6.0	17.6	7.0	11.9	8.0	6.0	10.0	10.1
	III	13.4	7.7	16.3	5.8	9.0	7.2	7.4	12.3	9.9
	IV	12.9	8.3	15.6	6.4	7.9	6.2	8.2	11.8	11.4
1991	I	13.1	7.9	15.2	6.4	10.2	6.5	11.7	11.1	11.3
	II	11.3	5.6	13.3	4.8	8.4	6.2	9.2	8.7	10.0
	III	11.5	4.9	13.1	5.4	10.5	6.5	10.6	11.1	9.9
	IV	10.9	4.9	11.8	4.4	11.0	4.2	12.3	10.5	10.2
1992	I	12.0	3.8	12.4	5.0	10.8	5.4	12.6	16.1	10.2
	II	11.2	4.6	11.8	4.1	11.5	4.4	12.2	13.1	8.3
	III	10.4	4.7	10.9	4.3	10.5	5.5	12.1	11.1	9.5
	IV	9.8	4.5	10.7	4.3	9.0	3.9	11.5	12.5	8.6
1993	I	13.0	7.1	14.9	6.1	10.9	5.5	13.9	15.0	11.4
	II	12.4	8.1	13.8	6.4	9.0	6.6	13.2	14.0	12.6
	III	12.3	8.1	13.6	6.0	9.4	5.6	12.7	11.3	11.9
	IV	11.7	6.6	12.6	9.3	9.4	5.0	14.4	13.0	11.7
1994	I	11.4	7.1	11.8	9.9	9.7	4.3	16.6	11.4	10.8
	II	11.6	7.3	12.6	9.6	9.0	4.8	16.7	11.1	11.6
	III	11.7	8.1	12.4	8.8	11.0	5.8	14.7	11.4	10.7
	IV	10.6	7.9	10.8	10.1	9.8	4.3	13.3	11.1	11.0
1995	I	13.2	7.7	13.3	13.0	14.7	5.5	14.0	17.6	13.8
	II	17.5	9.2	17.1	14.7	18.8	11.7	13.8	21.1	16.5
	III	18.1	9.4	18.9	13.5	19.3	10.9	15.0	21.7	19.5
	IV	16.0	7.4	16.9	12.5	17.4	9.1	16.2	19.2	15.6
1996	I	17.6	9.3	19.5	11.6	17.9	7.4	15.8	20.8	15.8
	II	18.8	7.3	20.6	15.3	18.5	7.7	18.4	23.6	15.2
	III	17.4	5.9	19.3	13.1	15.5	8.4	17.7	19.7	16.5
	IV	15.8	5.2	17.2	10.8	15.0	7.1	13.9	17.9	14.8
1997	I	19.0	6.8	19.8	14.1	18.2	10.7	13.2	25.5	17.5
	II	17.2	7.5	18.5	13.5	16.0	9.7	10.9	20.7	15.3
	III	16.2	6.1	16.3	10.0	13.6	9.3	10.6	17.9	12.1
	IV	14.1	4.4	15.0	9.3	12.5	8.0	8.1	20.2	10.8
1998	I	12.9	3.6	13.7	7.9	11.6	7.2	7.5	16.3	10.6
	II	11.6	2.3	12.0	7.8	11.7	6.4	9.3	15.2	9.5
	III	10.9	2.2	11.3	6.1	11.1	6.3	8.2	13.0	9.0
	IV	10.0	1.6	10.5	6.5	9.6	5.5	6.6	12.0	7.4
1999	I	11.4	1.7	12.7	6.4	10.9	6.7	8.5	16.0	7.4
	II	10.2	2.5	11.0	5.4	10.7	5.2	8.5	14.5	6.3
	III	9.5	2.4	10.8	5.1	9.4	4.3	6.7	14.1	4.4
	IV	11.3	3.2	12.7	5.5	9.3	6.8	5.9	16.0	5.5
2000	I	11.3	2.7	13.1	4.8	9.7	6.1	6.7	16.1	6.6
	II	10.3	3.9	12.1	4.8	10.0	7.0	6.9	15.0	6.5
	III	10.0	2.9	11.5	3.7	10.2	5.7	5.6	13.6	5.4
	IV	8.8	2.7	10.1	4.3	8.3	5.0	5.4	12.4	4.4

TABLA 1.3 TASA TRIMESTRAL DE INGRESOS INSUFICIENTES Y DESEMPLEO EN DIECISEIS CIUDADES PARA EL PERIODO DE 1989 A 2000 (Cont.)

Periodo		Nuevo Laredo	Orizaba	Puebla	San Luis Potosí	Tampico	Tijuana	Torreón	Veracruz
1989	I	18.3	35.0	24.2	19.0	16.0	5.6	19.5	18.0
	II	11.4	28.9	20.0	13.7	13.4	5.4	16.4	14.6
	III	9.4	29.1	17.6	16.8	14.1	5.4	15.4	12.9
	IV	16.5	26.8	17.2	16.7	16.0	5.4	13.9	15.3
1990	I	18.4	24.2	16.1	16.8	16.0	5.6	13.1	15.0
	II	11.3	20.5	13.9	13.1	15.3	5.0	12.1	10.3
	III	12.4	18.2	12.0	12.0	13.6	6.0	12.1	12.1
	IV	11.0	20.9	12.7	11.5	13.2	2.9	8.9	12.5
1991	I	8.9	23.3	14.6	11.6	19.6	4.4	10.7	13.6
	II	5.5	20.3	12.0	9.0	18.3	3.5	9.2	10.6
	III	8.4	18.3	10.8	9.1	21.9	2.5	9.9	11.1
	IV	9.6	18.1	9.5	11.1	24.9	2.1	11.9	10.1
1992	I	10.8	21.1	12.8	12.5	24.9	3.5	13.5	13.9
	II	9.2	18.7	14.1	8.7	21.4	2.7	12.1	11.1
	III	7.9	21.2	13.6	8.1	19.4	2.0	9.8	12.1
	IV	7.2	18.0	11.9	7.7	18.2	1.8	9.2	12.1
1993	I	9.0	25.1	13.8	8.8	20.4	3.5	12.9	15.6
	II	6.4	24.2	11.1	7.7	20.3	3.0	10.9	15.4
	III	7.6	21.8	10.0	8.6	20.9	3.7	14.3	15.8
	IV	7.8	20.6	10.8	9.0	22.2	2.6	16.1	15.8
1994	I	11.3	19.4	8.0	9.2	19.7	4.2	15.1	17.5
	II	11.2	17.8	6.3	9.8	21.6	4.3	14.4	14.7
	III	11.0	18.5	7.2	10.6	19.1	3.0	14.2	12.8
	IV	11.6	16.6	7.8	8.8	18.8	2.5	13.9	12.8
1995	I	12.5	19.0	10.6	12.0	26.5	3.7	17.2	18.0
	II	15.1	27.7	16.0	18.3	27.8	3.8	21.6	23.5
	III	12.5	28.2	14.6	17.6	26.4	3.5	21.2	22.2
	IV	11.2	27.8	12.7	15.0	24.3	3.5	18.8	21.9
1996	I	10.3	27.4	13.2	17.8	27.1	2.8	17.7	22.8
	II	11.8	30.6	16.7	18.8	28.4	2.9	19.7	22.8
	III	13.6	28.7	14.5	19.4	25.4	4.0	17.3	21.9
	IV	10.4	28.1	12.9	15.5	22.2	2.8	15.7	20.1
1997	I	11.4	35.9	16.0	25.3	27.5	3.4	18.7	23.6
	II	11.8	32.6	12.1	21.0	21.4	3.3	16.5	19.2
	III	9.6	29.5	9.8	18.7	20.1	3.3	13.7	17.7
	IV	9.1	30.3	11.5	14.6	20.0	2.7	11.9	20.9
1998	I	7.8	29.6	10.8	15.5	18.7	1.8	12.2	19.6
	II	7.1	26.4	9.4	12.9	17.4	2.5	11.1	18.9
	III	6.6	24.8	9.4	11.4	16.7	1.9	9.5	18.5
	IV	5.4	23.5	9.2	11.6	13.1	1.8	7.9	17.0
1999	I	6.0	24.5	9.4	11.1	13.6	2.4	8.7	18.9
	II	5.0	21.9	10.0	9.9	11.1	1.9	7.9	17.0
	III	5.6	21.1	8.4	9.3	10.2	2.2	7.8	13.2
	IV	6.2	24.3	11.8	12.0	13.0	2.7	7.6	17.7
2000	I	5.9	25.9	11.4	11.2	14.2	3.0	8.1	17.4
	II	3.5	23.3	11.2	11.2	12.9	2.8	6.3	15.5
	III	4.2	21.1	10.8	11.8	12.5	2.7	5.8	14.8
	IV	4.7	20.8	9.2	10.9	8.7	1.6	5.0	14.9

TABLA 1.4 TASA TRIMESTRAL DE DESEMPLEO ABIERTO EN LAS CIUDADES DE MEXICO, GUADALAJARA Y MONTERREY EN EL PERIODO DE 1973 - 1993.

Periodo		México	Guadalajara	Monterrey
1973	I	7.2	7.7	8.3
	II	8.2	6.7	7.6
	III	8.4	6.2	7.3
	IV	6.7	5.9	5.7
1974	I	6.7	6.1	6.8
	II	7.9	4.6	6.8
	III	8.0	6.6	8.0
	IV	7.1	5.7	6.9
1975	I	7.9	5.4	7.7
	II	7.6	5.2	7.1
	III	7.5	7.4	8.0
	IV	6.1	6.8	8.6
1976	I	6.7	5.3	6.7
	II	6.4	5.4	7.6
	III	6.8	6.9	8.5
	IV	7.0	7.2	7.5
1977	I	8.3	7.0	9.0
	II	7.7	7.2	9.2
	III	8.5	8.3	10.4
	IV	7.1	7.0	7.2
1978	I	7.4	6.6	8.5
	II	6.8	6.0	7.5
	III	7.0	5.7	8.1
	IV	6.4	5.8	6.2
1979	I	6.1	6.2	5.8
	II	5.1	5.4	5.6
	III	6.0	6.5	6.6
	IV	5.5	4.8	5.5
1980	I	4.4	4.0	6.0
	II	4.5	4.0	5.8
	III	4.9	5.9	6.3
	IV	3.5	6.2	3.4
1981	I	4.2	6.2	4.5
	II	3.8	5.8	4.8
	III	4.1	5.8	4.8
	IV	3.6	5.5	3.4
1982	I	3.5	4.4	4.3
	II	3.9	5.2	4.7
	III	4.4	5.3	6.2
	IV	4.1	5.1	4.5
1983	I	7.0	6.4	9.0
	II	5.5	8.6	10.2
1983	III	6.2	9.5	12.5
	IV	5.4	7.6	11.1
1984	I	5.4	6.9	8.3
	II	5.3	5.7	7.0
	III	6.4	3.0	8.5
	IV	6.2	5.7	6.2
1985	I	5.7	3.9	6.9
	II	4.3	3.5	5.1
	III	5.2	3.9	5.4
	IV	4.4	2.4	4.0
1986	I	4.9	2.6	5.0
	II	4.4	2.6	4.8
	III	5.8	4.2	6.0
	IV	5.1	3.4	5.6
1987	I	5.0	3.6	6.0
	II	4.5	2.9	5.7
	III	4.3	3.2	5.8
	IV	3.6	2.6	4.5
1988	I	4.0	2.9	4.0
	II	4.5	2.4	3.8
	III	4.9	2.6	4.4
	IV	3.9	2.0	3.7
1989	I	4.0	1.6	3.6
	II	3.7	1.7	2.6
	III	4.0	2.0	3.6
	IV	3.1	1.4	2.6
1990	I	2.8	1.2	3.2
	II	3.5	1.3	2.9
	III	3.7	2.1	3.9
	IV	3.0	1.9	3.6
1991	I	2.9	2.1	3.6
	II	2.7	1.8	3.1
	III	3.2	3.2	3.5
	IV	2.8	2.9	3.6
1992	I	3.3	3.1	3.0
	II	3.3	3.1	3.0
	III	3.4	3.2	3.6
	IV	3.4	2.6	3.1
1993	I	4.2	3.0	5.0
	II	3.5	2.7	4.8
	III	4.1	3.3	5.4
	IV	3.7	3.1	4.4

TABLA 1.5 TASA TRIMESTRAL DE DESEMPLEO ABIERTO EN DOCE CIUDADES PARA EL PERIODO DE 1983 I A 1996 II

Periodo		México	Chihuahua	Guadalajara	León	Mérida	Monterrey
1983	I	6.8	7.0	6.3	1.6	2.3	8.8
	II	5.5	6.6	8.1	1.3	4.2	9.8
	III	6.4	6.2	8.6	3.3	5.2	11.4
	IV	6.3	7.4	6.7	1.2	4.2	9.1
1984	I	5.4	7.2	6.9	1.6	3.9	8.3
	II	5.3	6.9	5.7	2.1	3.7	7.0
	III	6.4	9.6	6.0	3.0	5.0	8.5
	IV	6.2	6.1	5.7	2.1	4.0	6.2
1985	I	5.7	5.4	3.9	1.7	3.4	6.9
	II	4.3	3.3	3.5	1.2	2.6	5.1
	III	5.2	4.2	3.9	1.4	2.8	5.4
	IV	4.4	3.6	2.4	1.1	2.2	4.0
1986	I	4.9	3.3	2.6	1.2	2.4	5.0
	II	4.4	2.7	2.6	0.8	2.5	4.8
	III	5.8	3.6	4.2	1.3	1.9	6.0
	IV	5.1	3.4	3.4	1.2	1.7	5.6
1987	I	5.0	2.5	3.4	1.2	1.9	5.9
	II	4.6	2.2	2.9	1.6	.9	5.4
	III	4.4	2.9	3.1	1.5	1.5	5.5
	IV	3.6	2.6	2.5	.9	1.7	4.3
1988	I	4.0	2.2	2.9	1.2	1.4	4.0
	II	4.5	2.0	2.4	1.2	1.2	3.8
	III	4.9	3.0	2.6	.8	1.4	4.4
	IV	3.9	1.8	2.0	.6	1.6	3.7
1989	I	4.0	1.6	1.6	.9	.6	3.6
	II	3.7	1.4	1.7	.9	.4	2.6
	III	4.0	2.1	2.0	.8	.8	3.6
	IV	3.1	1.8	1.4	.8	.6	2.6
1990	I	2.8	2.7	1.2	1.4	.6	3.2
	II	3.5	2.3	1.3	1.4	.4	2.9
	III	3.7	2.2	2.1	1.3	1.4	3.9
	IV	3.0	2.6	1.9	1.0	.9	3.6
1991	I	2.9	2.1	2.1	1.1	.6	3.6
	II	2.7	2.1	1.8	.9	.4	2.1
	III	3.2	2.5	3.2	1.4	1.3	3.5
	IV	2.8	1.9	2.9	.7	1.0	3.6
1992	I	3.3	1.8	3.1	1.0	.9	3.0
	II	3.3	1.7	3.1	.7	1.3	3.0
	III	3.4	1.9	3.2	1.4	1.2	3.6
	IV	3.4	2.0	2.6	.8	1.4	3.1
1993	I	4.2	3.7	3.0	1.0	1.3	5.0
	II	3.5	3.3	2.7	1.5	1.1	4.8
	III	4.1	3.7	3.3	1.7	1.4	5.4
	IV	3.7	5.1	3.1	1.4	1.1	4.4
1994	I	4.3	5.8	3.6	1.3	1.3	4.7
	II	4.2	5.4	3.1	1.8	1.4	4.9
	III	3.8	4.9	3.8	2.3	2.0	5.2
	IV	3.9	5.8	3.1	1.6	2.3	5.5
1995	I	5.7	7.3	6.0	2.9	3.5	6.0
	II	7.4	7.8	6.7	5.6	3.2	8.1
	III	8.8	7.0	7.3	4.9	5.0	9.6
	IV	7.4	5.9	5.9	2.6	4.2	7.2
1996	I	7.6	5.4	5.7	2.5	3.3	6.7
	II	7.3	5.7	5.1	1.9	3.5	5.7

TABLA 1.5 TASA TRIMESTRAL DE DESEMPLEO ABIERTO EN DOCE CIUDADES PARA EL PERIODO DE 1983 I A 1996 II (Cont.)

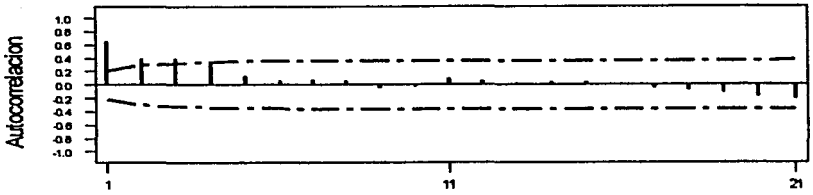
Periodo		Orizaba	Puebla	San Luis Potosí	Tampico	Torreón	Veracruz
1983	I	4.3	3.4	3.0	1.8	5.4	4.5
	II	4.5	4.1	5.6	4.6	6.3	3.5
	III	4.9	5.0	4.7	4.8	7.9	4.8
	IV	3.8	3.4	3.3	4.5	5.1	4.7
1984	I	4.5	3.2	4.4	3.4	5.6	4.5
	II	4.6	3.0	4.7	5.7	5.1	4.6
	III	5.6	3.8	4.3	7.4	6.7	5.5
	IV	5.4	3.4	3.5	5.8	5.8	6.8
1985	I	3.8	5.3	4.0	6.5	4.2	5.6
	II	1.8	3.6	2.8	4.2	3.1	3.6
	III	2.9	4.2	3.2	6.1	3.7	3.7
	IV	2.4	3.4	2.9	3.4	3.2	2.6
1986	I	2.4	2.7	3.0	4.6	3.2	3.1
	II	2.7	2.5	2.5	5.0	4.2	3.6
	III	3.1	3.1	4.3	5.9	5.1	4.5
	IV	3.4	2.6	3.0	6.2	4.1	3.4
1987	I	2.9	3.0	3.6	6.3	4.0	2.6
	II	2.5	2.6	3.4	5.2	2.2	4.4
	III	3.2	3.3	2.7	4.8	3.3	3.9
	IV	3.1	3.0	2.6	4.3	2.4	3.9
1988	I	2.4	3.0	1.9	4.2	2.7	3.7
	II	2.5	3.0	1.4	3.9	2.0	2.5
	III	2.7	2.4	2.8	3.1	2.0	3.0
	IV	2.7	1.9	1.3	2.4	2.4	2.2
1989	I	1.9	2.1	1.6	2.5	2.7	2.4
	II	1.9	1.5	2.0	3.3	2.4	2.2
	III	2.3	2.0	1.6	2.9	2.3	2.6
	IV	1.4	1.7	1.7	2.8	1.7	1.9
1990	I	2.0	1.8	1.3	2.6	2.4	2.2
	II	1.5	1.5	1.4	3.2	3.0	1.5
	III	1.8	1.7	2.0	3.3	3.3	2.4
	IV	2.0	2.2	1.3	3.4	2.1	1.8
1991	I	1.9	2.5	.8	5.3	2.6	2.5
	II	.7	1.5	1.4	4.4	2.0	2.9
	III	1.1	2.6	1.4	5.9	2.4	3.2
	IV	.8	1.4	1.0	6.3	2.0	2.2
1992	I	1.1	2.0	1.2	5.9	2.6	2.3
	II	.8	2.2	1.3	5.3	2.4	2.6
	III	1.1	2.8	1.2	4.9	2.5	2.6
	IV	1.7	2.4	1.0	4.9	2.0	2.5
1993	I	2.3	1.9	1.5	5.7	3.2	3.1
	II	1.7	1.9	1.4	5.0	3.0	3.0
	III	1.8	2.4	1.7	5.3	5.0	4.0
	IV	2.0	2.1	1.9	5.8	4.6	4.0
1994	I	1.9	2.6	1.6	5.9	5.5	4.8
	II	1.9	1.7	1.8	5.7	5.3	3.4
	III	2.0	2.9	3.0	5.9	6.3	4.4
	IV	1.7	3.0	1.9	4.6	5.7	3.5
1995	I	2.0	4.0	2.7	6.6	6.5	4.4
	II	3.4	6.7	4.2	7.3	6.9	5.1
	III	4.3	6.3	5.0	6.1	7.0	6.1
	IV	4.0	4.8	3.4	5.8	6.0	5.0
1996	I	4.5	4.0	4.6	6.2	5.9	5.2
	II	3.6	3.9	3.7	6.6	4.7	4.5

TABLA 1.6 TASA MENSUAL DE DESEMPLEO ABIERTO TOTAL ZONAS URBANAS PARA EL PERIODO DE ENERO 1990 A DICIEMBRE 2000

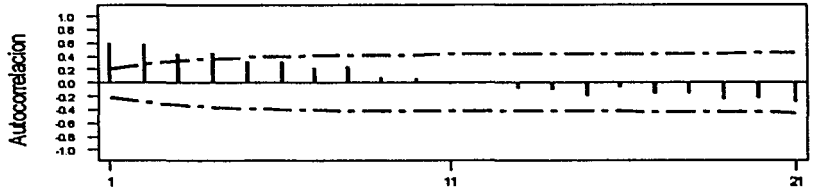
Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
1990	2.60	2.40	2.40	2.70	2.70	3.00	3.60	3.00	2.80	3.30	2.30	2.10
1991	2.80	2.50	2.90	2.60	2.30	2.10	2.70	3.20	3.10	3.20	2.70	2.20
1992	2.90	3.20	2.70	2.70	2.90	2.70	3.10	2.50	2.80	2.80	2.90	2.20
1993	3.20	3.50	3.60	3.00	3.30	3.20	3.60	3.90	4.00	3.70	3.20	2.90
1994	3.80	3.70	3.60	3.80	3.20	3.30	3.90	3.60	3.80	3.90	3.90	3.20
1995	4.50	5.30	5.70	6.30	6.60	6.60	7.30	7.60	7.30	6.70	5.80	5.50
1996	6.40	6.30	6.00	5.90	5.40	5.60	5.80	5.30	5.50	5.20	4.80	4.10
1997	4.50	4.20	4.20	4.30	3.90	3.40	4.10	3.50	3.40	3.20	3.30	2.80
1998	3.59	3.53	3.39	3.05	3.20	3.36	3.20	3.02	3.28	3.11	2.59	2.60
1999	2.85	3.20	2.71	2.70	2.44	2.58	2.26	2.49	2.24	2.50	2.09	2.00
2000	2.28	2.43	2.15	2.45	2.14	2.11	2.03	2.58	2.51	1.97	2.00	1.90

ANEXO B
Gráficas del análisis de series de tiempo

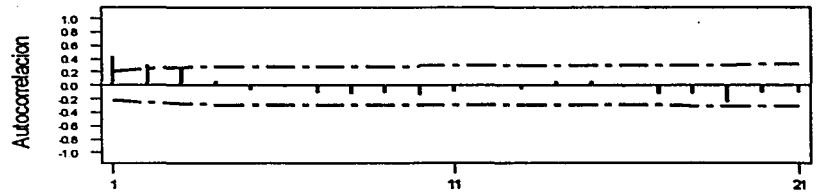
FAC para los residuos de la Cd. de México



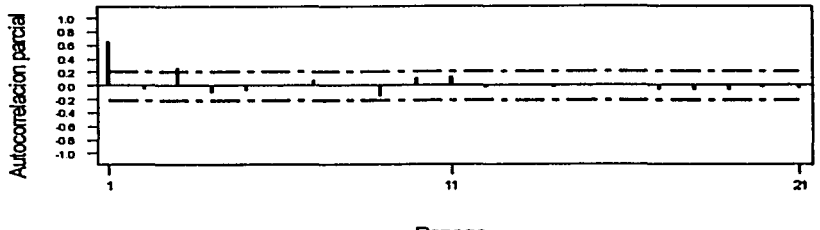
FAC para los residuos de Guadalajara



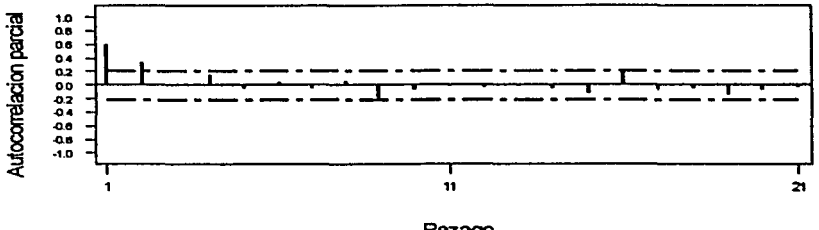
FAC para los residuos de Monterrey



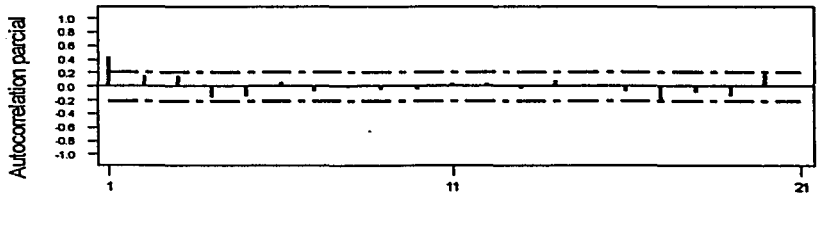
FACP para los residuos de la Cd. de México



FACP para los residuos de Guadalajara



FACP para los residuos de Monterrey



ANEXO C
Programa fuente del programa del pulido de medianas

```
#include<conio.h>
#include<iostream.h>
#include<fstream.h>
#include<stdlib.h>
#define TRUE 1
#define FALSE 0

class Matriz;                    //Declaración de la clase matriz

//INICIO DE LA CLASE Vector
class Vector{
        float *vec;              //Atributos: vec, nombre del vector
          int longt;             // longt, longitud del vector
  public:
    Vector(int tam);
//              ~Vector(){delete []vec;}
    friend float mediana_vec(Vector &b);
    friend Vector operator +(Vector &a,Vector &b);
    Vector operator =(Vector b);
    friend Vector operator -(Vector &a,float num);
    float dimevec(int i){return vec[i];}
    int archCV(Vector &b,float num);
    int archvec(char nom[7]);
    void exhibe_vec(char nom[3]);
              friend Vector & mediana_matren(Matriz &B,Vector &a);
    friend Vector & mediana_matcol(Matriz &B,Vector &b);
};

Vector::Vector(int tam){         //Constructor de Vector inicializando a cero
        int i;
        longt=tam;           // tam es un parámetro que inicializa a longt
  vec=new float [tam+1];  // vec se le está asignando memoria
  if(!vec){
        cout<<"\nFallo en la asignación de memoria.";
    exit(1);
  }
  for(i=0;i<tam;i++)
        vec[i]=0;       //inicialización a cero de todos los elementos
}
float mediana_vec(Vector &b){
  int j,pass,switched=TRUE;
  float hold;
  Vector a(b.longt);
  a=b;
  for(pass=0;pass<a.longt-1 && switched==TRUE;pass++){ //Inicio método de
        switched=FALSE;                                //la burbuja
    for(j=0;j<a.longt-pass-1;j++)
        if(a.vec[j]>a.vec[j+1]){
        switched=TRUE;
        hold=a.vec[j];
        a.vec[j]=a.vec[j+1];
        a.vec[j+1]=hold;
      }
  }     //fin método de la burbuja
```

```
  pass=a.longt%2;    //comprobación de si es par o no el vector
  j=a.longt/2;
  if(pass==0)
        return ((a.vec[j-1]+a.vec[j])/2); //regresa el valor de la mediana (par)
  else
        return a.vec[j];  //regresa el valor de la mediana si es impar
}

Vector operator +(Vector &a, Vector &b){
        int i;
  if(a.longt!=b.longt){
        cout<<"La dimension de los vectores no concuerda";
    exit(1);
  }
  Vector temp(a.longt);
  for(i=0;i<temp.longt;i++)
        temp.vec[i]=a.vec[i]+b.vec[i]; //Suma los elementos de los vectores 1 a 1
  return temp;  //regresa el vector
}

Vector Vector::operator =(Vector b){
        int i;
  if(longt!=b.longt){      //Compara la dimensión de los vectores
        cout<<"La dimension de los vectores no concuerda";
    exit(1);
  }
  for(i=0;i<longt;i++)
                vec[i]=b.vec[i];    //Iguala los elementos 1 a 1
  return *this; //regresa el apuntador al vector
}

Vector operator -(Vector &a, float num){
        int i;
  Vector temp(a.longt);
  for(i=0;i<a.longt;i++)
        temp.vec[i]=a.vec[i]-num;  //Se resta num a los elementos del vector
  return temp;
}

int Vector::archCV(Vector &b,float num){
        int i,j;
  float temp;
  ofstream flujo("CV.txt");   //Se abre el flujo para crear el archivo
  if(!flujo){
        cout<<"\nNo se puede crear el archivo: CV.txt";
    return 1;
  }
  for(i=0;i<longt;i++)
        for(j=0;j<b.longt;j++){
        temp=vec[i]*b.vec[j]/num;     //Se calcula CV
      flujo<<temp<<endl;            //Se manda al archivo
    }
  flujo.close();  //Se cierra el flujo
  return 0;
}
```

```
int Vector::archvec(char nom[7]){
        int i;
  ofstream flujo(nom);    //Se abre el flujo para el archivo
  if(!flujo){
        cout<<"\nNo se puede crear el archivo: "<<nom;
    return 1;
  }
  for(i=0;i<longt;i++)
        flujo<<vec[i]<<endl;    //Se manda la matriz al archivo
  flujo.close();            //Se cierra el flujo
        return 0;
}

void Vector::exhibe_vec(char nom[3]){
        int i;
  for(i=0;i<longt;i++)
        cout<<nom<<"["<<(i+1)<<"]="<<vec[i]<<endl; //Se manda a pantalla el vector
  cout<<endl;
}
//Fin de la clase Vector

//INICIO DE LA CLASE Matriz
class Matriz{
        private:
        int ren,col; //Atributos: ren.  Numero de renglones de la matriz
        float *mat;       //                          col.  Numero de columnas de la matriz
                //                  mat.  Nombre del arreglo que representa la matriz
  public:
        Matriz(void){}
    Matriz(int row,int colum);
//                  ~Matriz(){delete []mat;}
                    int leemat(void);
    int corregmat(void);
    int numren(void){return ren;}
    int numcol(void){return col;}
    friend istream & operator >>(istream &istream, Matriz &A);
    friend ostream & operator <<(ostream &stream, Matriz A);
    friend Vector & mediana_matren(Matriz &B,Vector &a);
    friend Vector & mediana_matcol(Matriz &B,Vector &b);
    Matriz restamat_vecren(Vector &a);
    Matriz restamat_veccol(Vector &b);
    int archmat(char nom[7]);
    Matriz matVA(Vector &a,Vector &b,float m);
    Matriz operator =(Matriz B);
};

Matriz::Matriz(int row,int colum){
        ren=row;                  //Parametro row que inicializa a ren
  col=colum;  //Paramentro colum que inicializa a col
  int i,j;
  if(ren>0 && col>0){
        mat=new float [ren*col+1]; //A mat se le asigna memoria
    for(i=0;i<ren;i++)
        for(j=0;j<col;j++)
        mat[i*col+j]=0;  //los elementos de la matriz se igualan a cero
  }
```

```
  else{
        cout<<"\nNumero de renglones o columnas no validos para matriz";
    exit(1);   //matriz de dimensiones mayor a cero
  }
}

int Matriz::leemat(void){
        int i,j;
  cout<<"\nMetodo que lee el archivo MATRIZ.TXT ";
  cout<<"\nEl archivo debe encontrarse en la MISMA ruta que el EJECUTABLE";
  cout<<"\n\nInserta la DIMENSION de la Matriz";
  cout<<"\n\nDime los RENGLONES de la Matriz: ";
  cin>>i;  //lee del teclado el renglon
  ren=(int)i;    // variable i que inicializa al atributo ren
  while(ren<=0){
        cout<<"\nNumero de renglones no VALIDOS.";
    cout<<"\nDigite el NUEVO numero de RENGLONES: ";
    cin>>i;
    ren=(int)i;
  }
  cout<<"\nDime las COLUMNAS de la Matriz: ";
  cin>>j;  //lee del teclado la columna
  col=(int)j;    //variable j que inicializa al atributo  col
  while(col<=0){
        cout<<"\nNumero de columnas no VALIDOS.";
    cout<<"\nDigite el NUEVO numero de COLUMNAS: ";
    cin>>j;
    col=(int)j;
  }
  mat=new float [ren*col+1];  // A mat se le asigna memoria
  ifstream flujo("matriz.txt");   // Se abre el flujo para leer el archivo
  if(!flujo){                   //Se comprueba si existe el archivo matriz.txt
        cout<<"\nNo se puede abrir el archivo: MATRIZ.TXT";
    return 1;
  }
  cout<<"\n\nLeyendo el archivo MATRIZ.txt";
  for(i=0;i<ren;i++)
        for(j=0;j<col;j++)
        flujo>>mat[i*col+j];        //Se lee del archivo la matriz
  cout<<"\nArchivo MATRIZ.TXT ya fue leido\n";
  return 0;
}

int Matriz::corregmat(void){
        int i,j;
        char resp='s';
  cout<<"\nDeseas modificar algun valor de la matriz  (s/n): ";
  cin>>resp;
        while(resp=='s'){
        cout<<"\nDigita el numero de RENGLON: ";
    cin>>i;
    while(i<=0 ||i>ren){
        cout<<"\Digita un numero de RENGLON valido: ";
      cin>>i;
    }
```

```
    cout<<"\nDigita el numero de COLUMNA: ";
    cin>>j;
    while(j<=0 ||j>col){
        cout<<"\Digita un numero de COLUMNA valido: ";
      cin>>j;
    }
    cout<<"\nEl valor actual es e["<<i<<"]["<<j<<"]="<<mat[(i-1)*col+(j-1)];
    cout<<"\nInserta el NUEVO valor: ";
    cin>>mat[(i-1)*col+(j-1)];   //modifica el valor del elemento
    cout<<"\nDeseas modificar otro valor (s/n): ";
    cin>>resp;
  }
  return 0;
}

istream &operator >>(istream &istream, Matriz &A){
        int i,j;
  cout<<"\nIngresa la dimension de la Matriz.";
        cout<<"\nIngresa el numero de RENGLONES de la Matriz: ";
  istream>>A.ren;   //Lee desde el teclado el numero de renglones y lo asigna
  while(A.ren<=0){  // al atributo ren de la matriz
        cout<<"\nError numero de RENGLONES no valido";
    cout<<"\nIngresa Nuevo numero de RENGLONES: ";
    istream>>A.ren;
  }
  cout<<"\nIngresa el numero de COLUMNAS de la Matriz: ";
  istream>>A.col;   //Lee desde el teclado el numero de renglones y lo asigna
  while(A.col<=0){       // al atributo col de la matriz
        cout<<"\nError numero de COLUMNAS no valido";
    cout<<"\nIngresa Nuevo numero de COLUMNAS: ";
    istream>>A.col;
  }
  A.mat=new float [A.ren*A.col+1]; // A  A.mat  se le asigna memoria
  cout<<"\nIngresa los VALORES de la Matriz por renglones."<<endl;
  for(i=0;i<A.ren;i++)
        for(j=0;j<A.col;j++){
        cout<<"\ne["<<(i+1)<<"]["<<(j+1)<<"]= ";
      istream>>A.mat[i*A.col+j];  //Se leen los valores desde el teclado
    }                       // y se asignan a la matriz
  return istream;
}

ostream & operator <<(ostream &stream, Matriz A){
        int i,j;
  stream<<"\n";
  for(i=0;i<A.ren;i++){
        stream<<"[ ";
    for(j=0;j<A.col;j++)
        stream<<A.mat[i*A.col+j]<<" ";  //despliega la matriz en la pantalla
          stream<<" ]\n";
  }
  cout<<endl;
  return stream;
}
```

```
Vector & mediana_matren(Matriz &B,Vector &a){
        int i,j,pass,switched;
  float hold;
  Matriz A(B.ren,B.col);
  A=B;
  for(i=0;i<A.ren;i++){  //Inicia metodo de la burbuja
        switched=TRUE;
        for(pass=0;pass<A.col-1 && switched==TRUE;pass++){
        switched=FALSE;
      for(j=0;j<A.col-pass-1;j++)
        if( A.mat[i*A.col+j] > A.mat[i*A.col+(j+1)] ){
        switched=TRUE;
          hold=A.mat[i*A.col+j];
          A.mat[i*A.col+j]=A.mat[i*A.col+(j+1)];
          A.mat[i*A.col+(j+1)]=hold;
        }
    }
  } //termina metodo de la burbuja
  pass=A.col%2;  //Se comprueba si es par o impar la matriz
  j=A.col/2;
  if(pass==0)
        for(i=0;i<A.ren;i++)   //Se obtiene un vector de medianas cuando es par
        a.vec[i]=(A.mat[i*A.col+(j-1)]+A.mat[i*A.col+j])/2;
  else
        for(i=0;i<A.ren;i++)   //Se obtiene un vector de medianas cuando es impar
        a.vec[i]=A.mat[i*A.col+j];
  return a; //Se regresa el vector
}

Vector & mediana_matcol(Matriz &B,Vector &b){
  int i,j,pass,switched;
  float hold;
  Matriz A(B.ren,B.col);
  A=B;
  for(j=0;j<A.col;j++){          //Inicia metodo de la burbuja
        switched=TRUE;
        for(pass=0;pass<A.ren-1 && switched==TRUE;pass++){
        switched=FALSE;
      for(i=0;i<A.ren-pass-1;i++)
        if( A.mat[i*A.col+j] > A.mat[(i+1)*A.col+j] ){
        switched=TRUE;
          hold=A.mat[i*A.col+j];
          A.mat[i*A.col+j]=A.mat[(i+1)*A.col+j];
          A.mat[(i+1)*A.col+j]=hold;
        }
    }
  } //termina metodo de la burbuja
  pass=A.ren%2;         //Se comprueba si es par o impar la matriz
  i=A.ren/2;
  if(pass==0)
        for(j=0;j<A.col;j++)   //Se obtiene un vector de medianas cuando es par
        b.vec[j]=( A.mat[(i-1)*A.col+j] + A.mat[i*A.col+j] )/2;
  else
        for(j=0;j<A.col;j++)   //Se obtiene un vector de medianas cuando es impar
        b.vec[j]=A.mat[i*A.col+j];
  return b;  //Se regresa el vector
}
```

```
Matriz Matriz::restamat_vecren(Vector &a){
        int i,j;
  for(i=0;i<ren;i++)
        for(j=0;j<col;j++)  //Se resta a la matriz los elementos de un vector
        mat[i*col+j]=mat[i*col+j]-a.dimevec(i);  //Se efectua por renglones
  return *this;  //Se regresa el apuntador a la matriz
}

Matriz Matriz::restamat_veccol(Vector &b){
        int i,j;
  for(j=0;j<col;j++)
        for(i=0;i<ren;i++)  //Se resta a la matriz los elementos de un vector
        mat[i*col+j]=mat[i*col+j]-b.dimevec(j); //Se efectua por columnas
  return *this;  //Se regresa el apuntador a la matriz
}

int Matriz::archmat(char nom[7]){
  int i,j;
  ofstream flujo(nom);  //Se abre el flujo para crear el archivo
  if(!flujo){
        cout<<"\nNo se puede crear el archivo: "<<nom;
    return 1;   //Se valida si se puede crear el archivo
  }
        for(i=0;i<ren;i++){
        for(j=0;j<col;j++)
        flujo<<mat[i*col+j]<<"\t"; //Se manda la matriz al archivo
    flujo<<endl;

  }
  flujo.close();   //Se cierra el flujo
  return 0;
}


Matriz Matriz::matVA(Vector &a,Vector &b,float m){
        int i,j;
  for(i=0;i<ren;i++)
        for(j=0;j<col;j++) //Se obtiene la matriz de valores ajustados
        mat[i*col+j]=mat[i*col+j]+a.dimevec(i)+b.dimevec(j)+m;
  return *this;  //Se regresa el apuntador a la matriz
}

Matriz Matriz::operator =(Matriz B){
        if(ren!=B.ren || col!=B.col){      //Se comprueba la igualdad de dimensiones
        cout<<"La dimension de las matrices no concuerda";
    exit(1);
  }
  int i,j;
  for(i=0;i<ren;i++)
        for(j=0;j<col;j++)
        mat[i*col+j]=B.mat[i*B.col+j]; //Se iguala uno a uno los valores de
  return *this;                                                    // los elementos de la matriz
}
//fin de la clase Matriz
```

```
//Funcion accesoria
int archval(float num){
        ofstream flujo("m.txt");   //Se abre un flujo para el archivo m.txt
  if(!flujo){               //Se valida la creación del archivo
        cout<<"\nNo se puede crear al archivo: m.txt";
    return 1;
  }
  flujo<<num;      //Se manda el numero al archivo
  flujo.close();   //Se cierra el flujo
  return 0;
}
//Fin funcion accesoria

//FUNCION PRINCIPAL main
int main(void){
        Matriz e;          //Creación del objeto de tipo Matriz  "e"
  int opcion,itera=0,i,j,estatus=0;
  clrscr();
  cout<<"\n Programa que realiza el pulido de medianas de una matriz.";
  cout<<"\n\nMENU\n( 1 ) Leer la matriz desde un archivo\n";
  cout<<"( 2 ) Insertar la matriz desde el teclado\n( 3 ) Salir.\n";
  cout<<"\nDigita una opcion: ";
  cin>>opcion;              //Se lee la opcion
  if(opcion==1){
        estatus=e.leemat();    //lee la matriz de un archivo
        if(estatus==1)return 1;
  }else  if(opcion==2)
                          cin>>e;          //lee la matriz desde el teclado
      else
        return 0;
        e.corregmat();           //Se pueden modificar los valores de la matriz
  while(itera<=0){
        cout<<"\nDigita el numero de iteraciones (mayor a cero): ";
    cin>>itera;           //Se lee el numero de iteraciones
  }
  cout<<"\nPara empezar por RENGLONES ( 1 )";
  cout<<"\nPara empezar por COLUMNAS  ( 2 )";
  cout<<"\n\nDigita la opcion: ";
  cin>>opcion;      // lee opcion de efectuar el algoritmo por reng. o column.
  while(opcion!=1 && opcion!=2){
        cout<<"\nOpcion no valida.\nDigita la opcion: ";
    cin>>opcion;
  }
  i=e.numren();    //numero de renglones de la matriz
  j=e.numcol();    //numero de columnas de la matriz
  Vector ai(i),a1p(i),apr(i),bj(j),b1p(j),bpr(j); //Se crean los objetos Vector
  float mb=0,ma=0,mpr=0,m=0;    //Se crean variables de tipo float
  int cont; //contador
  if(opcion==1)
        for(cont=0;cont<itera;cont++){
        mediana_matren(e,a1p);
            mb=mediana_vec(bj);
          e.restamat_vecren(a1p);
                        apr=ai+a1p;
            mpr=m+mb;
          mediana_matcol(e,b1p);
        ma=mediana_vec(apr);
```

```
            e.restamat_veccol(b1p);
          bpr=bj+b1p;
        m=ma+mpr;
            bj=(bpr-mb);
          ai=(apr-ma);
    }


  else if(opcion==2)
        for(cont=0;cont<itera;cont++){
                        mediana_matcol(e,b1p);
        ma=mediana_vec(ai);
        e.restamat_veccol(b1p);
        bpr=bj+b1p;
        mpr=m+ma;
        mediana_matren(e,a1p);
        mb=mediana_vec(bpr);
        e.restamat_vecren(a1p);
        apr=ai+a1p;
        m=mb+mpr;
        bj=(bpr-mb);
        ai=(apr-ma);
        }
  cout<<"\nSalida de archivo ai.txt";
  ai.archvec("ai.txt");
  cout<<"\nSalida de archivo bj.txt";
  bj.archvec("bj.txt");
  cout<<"\nSalida de archivo m.txt";
  archval(m);
  cout<<"\nSalida de archivo e.txt";
  e.archmat("e.txt");
  cout<<"\nSalida de archivo cv.txt";
  ai.archCV(bj,m);
  cout<<"\nSalida de archivo va.txt";
  e.matVA(ai,bj,m);
  e.archmat("va.txt");
  cout<<"\n\nFIN del programa.";
  return 0;
}
//Fin de la función principal MAIN
```

BIBLIOGRAFÍA

1. Besag, Julian. "On resistant techniques and satistical analysis". Biometrika Vol. 68, 2. 1981. Págs 463-469.

2. Byers, J.D. "Testing for common trends in regional unemployment". Applied Economics. 23. 1991. Págs. 1087-1092.

3. Flores Díaz, José Antonio. "Comparación de métodos para análisis de datos en tablas de doble entrada con una observación por celda". Inédita. México. Tesis presentada para aspirar al grado de Actuario. UNAM. 1977. 72 págs.

4. Hartwig, F. y Dearing, B. "Exploratory Data Analysis". Sage University Paper No. 16, California, E.U. 1990. 81 págs.

5. Hoaglin, David C. "Direct Approximations for Chi-Squared Percentage Points". Journal of the American Statistical Association. Vol. 72, 359, Septiembre 1977. Págs. 508-515.

6. Hoaglin, David C., Frederick Mosteller, y John W. Tukey eds. Understanding Robust and Exploratory Data Analysis. Nueva York, John Wiley & Sons, 1983.

7. Hoaglin, David C., Frederick Mosteller, y John W. Tukey eds. Exploring Data Tables, Trends and Shapes. Nueva York, John Wiley & Sons, 1985.

8. Hoaglin, David C., Frederick Mosteller y John W. Tukey, eds. Fundamentals of Exploratory Analysis of Variance. Nueva York, John Wiley & Sons. 1991.

9. Hoaglin, David C. y P. F. Velleman. Applications, Basics and Computing of Exploratory Data Analysis. Boston, Massachusetts, Duxbury Press, 1981.

10. Huber, Peter J. Robust Statistics. Nueva York, John Wiley & Sons. 1981.

11. Ibarra, Darío G. "Economía informal, saldo de la crisis". Macroeconomía. Año IV num 39, Octubre 1996.

12. INEGI. "Banco de Información Estadística". [s.p.] http:// www.inegi.gob.mx

13. Marsh, Catherine. Exploring Data. An Introduction to Data Analysis for Social Scientists. Polity Press.

14. Marston, Stephen T. "Two views of the geographic distribution of unemployment". Quaterly Journal of Economics. Vol 100. 1985. Págs. 57-79.

15. Méndez, Ignacio. "Descomposición de la Interacción en Tablas de Doble Entrada" en Memoria del X Foro Nacional de Estadística y II Congreso Iberoamericano de Estadística. Ed. por Ma. de Lourdes de la Fuente, Eduardo Gutiérrez y Jorge Olguín, México, INEGI, 1996. Págs. 161-166.

16. México. Poder Ejecutivo Federal. Programa de empleo, capacitación y defensa de los derechos laborales 1995-2000. 1995.

17. Morales Alvarez, Martha Edith. "El desempleo en el estado neoliberal: un análisis estadístico". Inédita. Tesis presentada para obtener el título de Actuario. UNAM. 1998. 104 págs.

18. Mosteller, Frederick y John W. Tukey. Data Analysis and Regression. E.U. Addison Wesley, 1977.

19. OIT. XIII Conferencia Internacional de Estadígrafos del Trabajo. "Resolución sobre estadísticas de la población económicamente activa, del empleo, del desempleo y del subempleo". [s.p.] 2000. http://www.ilo.org/public/spanish/ bureau /stat/res

20. OIT. XIV Conferencia Internacional de Estadígrafos del Trabajo. "Directrices sobre la incidencia de los programas de promoción del empleo, sobre la medición del empleo y del desempleo". [s.p.] http://www.ilo.org/public/spanish/ bureau /stat

21. Organización para la Cooperación y el Desarrollo Económico. Economic Outlook. No. 58. Diciembre 1995.

22.Roel Pavón, Jimena. "Consideraciones sobre la convergencia del desempleo en México". Inédita. México. Tesis presentada para aspirar al grado de Licienciado en Economía. ITAM.1995. 49 págs.

23.Salgado Ugarte, Isaías. El análisis exploratorio de datos biológicos, Fundamentos y aplicaciones. México, ENEP–Zaragoza UNAM, 1992. 234 págs.

24.Troncoso Viniegra, Alfredo. "Co-integración en series de tiempo". Inédita. México. Tesis presentada para aspirar al grado de Actuario. Instituto Tecnológico Autónomo de México. 1997. 126 págs.

25.Tukey, John W. Exploratory Data Analysis. Londres, Addison Wesley, 1977.