



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

---

---

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN  
INGENIERÍA**

**APLICACIÓN DE LOS MODELOS ARIMA A LA  
ELABORACIÓN DE PRONÓSTICOS DE LA  
DESOCUPACIÓN FEMENINA EN MÉXICO  
2000-2012**

**T E S I S**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRO EN INGENIERÍA**

INGENIERÍA DE SISTEMAS – INVESTIGACIÓN DE OPERACIONES

P R E S E N T A :

**MARÍA DEL CARMEN MARCE VÁZQUEZ**

TUTOR:

**DR. MANUEL ORDORICA MELLADO**

**2012**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



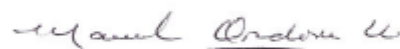
**JURADO ASIGNADO:**

Presidente: DR. RICARDO ACEVES GARCÍA  
Secretario: DR. MANUEL ORDORICA MELLADO  
Vocal: M. EN I. FRANCISCA IRENE SOLER ANGUIANO  
1<sup>er</sup> Suplente: DRA. COZUMEL ALLANEC MONROY LEÓN  
2<sup>o</sup> Suplente: M. EN I. ANN GODELIEVE WELLENS PURNAL

México, D. F.

**TUTOR DE TESIS:**

DR. MANUEL ORDORICA MELLADO



---

**FIRMA**



*A todas aquellas personas que confían  
en mí y que día con día me brindan  
su apoyo, comprensión y cariño.*

*¡Muchas gracias!*



# ÍNDICE

---

---

INTRODUCCIÓN .....	1
1. PROBLEMA DEL DESEMPLEO .....	5
1.1 Introducción .....	5
1.2 Conceptos relacionados con el empleo y el desempleo .....	6
1.2.1 Población económicamente activa .....	6
1.2.2 Ocupados y desocupados abiertos .....	7
1.2.3 Concepto de desempleo .....	8
1.3 Medición del empleo y la desocupación abierta .....	9
1.4 Tipos de desempleo .....	16
1.4.1 Desempleo friccional .....	16
1.4.2 Desempleo estructural .....	16
1.4.3 Desempleo cíclico .....	17
1.5 Consecuencias del desempleo .....	17
2. EMPLEO Y DESOCUPACIÓN ABIERTA EN MÉXICO .....	19
2.1 Introducción .....	19
2.2 Desempleo abierto en el contexto del mercado laboral .....	19
2.3 Factores que inciden en el nivel y variación de la desocupación abierta .....	23
2.4 Encuestas de ocupación en México .....	27



2.5	Características de la información para su análisis .....	30
2.5.1	Contraste entre las tasas de desocupación por niveles de escolaridad .....	30
2.5.2	Contraste entre las tasas de desocupación por regiones .....	32
2.5.3	Contraste entre las tasas de desocupación por sexo .....	33
2.5.4	Contraste entre las tasas de desocupación por edad .....	37
3.	ANÁLISIS DE SERIES DE TIEMPO .....	39
3.1.	Introducción .....	39
3.2.	Definición y objetivos del análisis de las series de tiempo .....	39
3.3.	Procesos estocásticos y series de tiempo .....	41
3.4.	Método de descomposición de series de tiempo .....	41
3.4.1	Componentes de las series de tiempo .....	42
3.4.2	Integración de los componentes de una serie de tiempo .....	46
3.4.3	Técnicas descriptivas simples .....	48
3.5.	Funciones de autocovarianza y autocorrelación .....	50
3.6.	Estacionariedad .....	59
3.7.	Diferenciación .....	64
3.8.	Metodología de Box-Jenkins .....	67
3.8.1	Proceso de modelización de Box-Jenkins .....	69
3.8.2	Modelos de promedios móviles (MA) .....	78
3.8.3	Dualidad entre los modelos de medias móviles (MA) y autorregresión (AR) .....	79
3.8.4	Modelos autorregresivos (AR) .....	80

3.8.5	Modelos mixtos autorregresivos y de medias móviles (ARMA) ..	81
3.8.6	Modelos autorregresivos integrados de medias móviles (ARIMA) .....	81
3.8.7	Comportamientos típicos de la FAC Y FACP .....	84
4.	PRONÓSTICOS DE LA DESOCUPACIÓN FEMENINA EN MÉXICO .....	85
4.1.	Introducción .....	85
4.2.	Modelo ARIMA para la tasa de desocupación femenina .....	95
4.2.1	Identificación .....	95
4.2.2	Estimación de los posibles modelos ARIMA .....	104
4.2.3	Diagnóstico al modelo elegido $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos .....	111
4.2.4	Análisis de intervención .....	113
4.2.5	Predicción .....	120
	CONCLUSIONES .....	125
	REFERENCIAS .....	131

# ÍNDICE DE FIGURAS

---

---

<b>Figura 1.1</b> Población por condición de actividad .....	6
<b>Figura 1.2</b> Población por condición de actividad .....	8
<b>Figura 1.3</b> Indicadores complementarios a la desocupación abierta .....	11
<b>Figura 2.1</b> El mercado de trabajo .....	21
<b>Figura 2.2</b> Porcentaje de los desocupados de acuerdo a su nivel de instrucción, en diciembre de 2010 .....	31
<b>Figura 2.3</b> Tasas de desocupación (mensual) para algunos estados de la República Mexicana, de marzo de 2005 a diciembre de 2010 .....	32
<b>Figura 2.4</b> Tasa de desocupación nacional (mensual) para hombres y para mujeres de abril de 2000 a diciembre de 2010 .....	33
<b>Figura 2.5</b> Tasa de ocupación parcial y desocupación TOPD1 (mensual) para hombres y para mujeres de enero de 2005 a diciembre de 2010 .....	34
<b>Figura 2.6</b> Tasa de condiciones críticas de ocupación TCCO (mensual) para hombres y para mujeres de enero de 2005 a diciembre de 2010 .....	35
<b>Figura 2.7</b> Tasa de presión general TPRG (mensual) para hombres y para mujeres de enero de 2005 a diciembre de 2010 .....	36
<b>Figura 2.8</b> Tasa de ocupación en el sector informal TOSI (mensual) para hombres y para mujeres de enero de 2005 a diciembre de 2010 .....	37
<b>Figura 2.9</b> Tasa de desocupación por grupos de edad, para los cuartos trimestres de los años 2005 al 2010 .....	38
<b>Figura 3.1</b> Ejemplo de serie de datos ( $Y_t$ ) con componente de tendencia .....	43
<b>Figura 3.2</b> Ejemplo de serie de datos ( $Y_t$ ) con componente de ciclicidad .....	43
<b>Figura 3.3</b> Ejemplo de serie de datos ( $Y_t$ ) con componente de estacionalidad .....	44

<b>Figura 3.4</b> Ejemplo de serie de datos ( $Y_t$ ) con componente de aleatoriedad .....	45
<b>Figura 3.5</b> Patrones de autocorrelación y de no autocorrelación .....	52
<b>Figura 3.6</b> Ejemplo de correlograma .....	59
<b>Figura 3.7</b> Ejemplo de serie ( $Y_t$ ) estacionaria .....	60
<b>Figura 3.8</b> Ejemplo de serie ( $Y_t$ ) con media no constante .....	60
<b>Figura 3.9</b> Ejemplo de serie ( $Y_t$ ) con varianza no constante .....	61
<b>Figura 3.10</b> Ejemplo de serie ( $Y_t$ ) con media y varianza no constantes .....	61
<b>Figura 3.11</b> Metodología Box-Jenkins .....	70
<b>Figura 4.1</b> Tasa de desocupación nacional (mensual) total, para hombres y para mujeres de abril de 2000 a diciembre de 2010 .....	86
<b>Figura 4.2</b> Tasa de ocupación parcial y desocupación TOPD1 (trimestral) total, para hombres y para mujeres del segundo trimestre de 2000 al cuarto trimestre de 2010 .....	87
<b>Figura 4.3</b> Tasa de condiciones críticas de ocupación TCCO (trimestral) total, para hombres y para mujeres del segundo trimestre de 2000 al cuarto trimestre del 2010 .....	88
<b>Figura 4.4</b> Porcentaje de los desocupados por sexo de acuerdo a su condición de experiencia laboral, en diciembre de 2010 .....	89
<b>Figura 4.5</b> Porcentaje de los desocupados por sexo de acuerdo a su nivel de instrucción, en diciembre de 2010 .....	90
<b>Figura 4.6</b> Porcentaje de la población por sexo subocupada, en diciembre de 2010 .....	92
<b>Figura 4.7</b> Porcentaje mensual de la población ocupada clasificada como subocupada, de enero de 2005 a diciembre de 2010 .....	93
<b>Figura 4.8</b> Porcentaje de los subocupados de acuerdo a su nivel de instrucción, en diciembre de 2010 .....	93
<b>Figura 4.9</b> Porcentaje de los hombres subocupados de acuerdo a su nivel de instrucción, en diciembre de 2010 .....	94

**Figura 4.10** Porcentaje de las mujeres subocupadas de acuerdo a su nivel de instrucción, en diciembre de 2010 ..... 94

**Figura 4.11** Serie de la tasa de desocupación mensual para mujeres (*MUJERES*) de abril de 2000 a diciembre de 2010 ..... 95

**Figura 4.12** Autocorrelación de la serie de la tasa de desocupación mensual para mujeres (*MUJERES*), generada con Eviews ..... 96

**Figura 4.13** Serie de la transformación logarítmica de la tasa de desocupación mensual para mujeres (*LMUJERES*) de abril de 2000 a diciembre de 2010, generado con Eviews ..... 97

**Figura 4.14** Resultados del estadístico de Dickey-Fuller aumentado (ADF) para las series *MUJERES*, *LMUJERES* y *D1LMUJERES*, generados con Eviews ... 98

**Figura 4.15** Serie de las primeras diferencias de la transformación logarítmica de la tasa de desocupación mensual para mujeres (*D1LMUJERES*), generada con Eviews ..... 99

**Figura 4.16** Autocorrelación de la serie de las primeras diferencias de la transformación logarítmica de la tasa de desocupación mensual para mujeres (*D1LMUJERES*), generada con Eviews ..... 100

**Figura 4.17** Serie de la transformación  $(1 - B)(1 - B^6)LMUJERES = D1D6LMUJERES$ , generada con Eviews ..... 101

**Figura 4.18** Autocorrelación de la serie de la transformación  $(1 - B)(1 - B^6)LMUJERES = D1D6LMUJERES$ , generada con Eviews ..... 102

**Figura 4.19** Ajuste del modelo  $ARIMA(1,1,1)(0,1,0)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews ..... 104

**Figura 4.20** Autocorrelación de los residuos del modelo  $ARIMA(1,1,1)(0,1,0)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews ..... 105

<b>Figura 4.21</b> Ajuste del modelo $ARIMA(1,1,1)(1,1,1)_6$ en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews .....	106
<b>Figura 4.22</b> Ajuste del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews .....	107
<b>Figura 4.23</b> Autocorrelación de los residuos del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews .....	108
<b>Figura 4.24</b> Histograma de los residuos del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews .....	111
<b>Figura 4.25</b> Correlograma de los residuos al cuadrado del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews .....	112
<b>Figura 4.26</b> Residuos del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews .....	114
<b>Figura 4.27</b> Ajuste del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos con análisis de intervención con variable artificial de impulso D0812, de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews .....	115
<b>Figura 4.28</b> Correlograma de los residuos del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews .....	117

**Figura 4.29** Histograma de los residuos del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews ..... 118

**Figura 4.30** Prueba ARCH al modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generada con Eviews ..... 118

**Figura 4.31** Residuos del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews ..... 119

**Figura 4.32** Valores reales, predicciones y límites de intervalo de confianza ( $\pm 2S.E.$ ) del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos, de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews ..... 120

**Figura 4.33** Valores reales, predicciones y límites de intervalo de confianza ( $\pm 2S.E.$ ) del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews ..... 121

# ÍNDICE DE TABLAS

---

---

<b>Tabla 3.1</b> Características de la FAC y la FACP de modelos generales no estacionales .....	84
<b>Tabla 4.1</b> Desocupados por sexo de acuerdo a grupo de edad, para el cuarto trimestre de 2010 .....	88
<b>Tabla 4.2</b> Desocupados por sexo de acuerdo a grupo de edad, para el cuarto trimestre de 2010 .....	91
<b>Tabla 4.3</b> Resultados de las series de diferentes transformaciones de la serie <i>LMUJERES</i> , generados con Eviews .....	101
<b>Tabla 4.4</b> Sobre ajuste al modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generados con Eviews .....	109
<b>Tabla 4.5</b> Parámetros de los modelos en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generados con Eviews.....	110
<b>Tabla 4.6</b> Resultados del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos con y sin intervención, de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generados con Eviews .....	116
<b>Tabla 4.7</b> Predicciones del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos, de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews .....	122
<b>Tabla 4.8</b> Predicciones del modelo $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews .....	123



# LISTA DE SIGLAS

---

<b>AC</b>	Autocorrelación
<b>ADF</b>	Dickey-Fuller aumentado
<b>AIC</b>	Akaike info criterion
<b>AR</b>	Modelo autorregresivo
<b>ARCH</b>	Heterocedasticidad condicional autorregresiva
<b>ARIMA</b>	Modelo autorregresivo integrado de medias móviles
<b>BIE</b>	Banco de Información Económica
<b>CIET</b>	Conferencia Internacional de Estadígrafos del Trabajo
<b>CT</b>	Cantidad de trabajo
<b>CTD</b>	Cantidad de trabajo demandada
<b>CTO</b>	Cantidad de trabajo ofrecida
<b>DA</b>	Desocupados abiertos
<b>DF</b>	Dickey Fuller
<b>DL</b>	Demanda laboral
<b>ECMO</b>	Encuesta Continua de Mano de Obra
<b>ECSO</b>	Encuesta Continua sobre Ocupación
<b>ENE</b>	Encuesta Nacional de Empleo
<b>ENET</b>	Encuesta Nacional de Empleo Trimestral
<b>ENEU</b>	Encuesta Nacional de Empleo Urbano
<b>ENH</b>	Encuesta Nacional en Hogares
<b>ENOE</b>	Encuesta Nacional de Ocupación y Empleo

- FAC** Función de autocorrelación
- FACP** Función de autocorrelación parcial
- I** Iniciadores (individuos que no trabajaron pero que declararon estar por iniciar una ocupación)
- INEGI** Instituto Nacional de Estadística y Geografía
- OBT** Ocupados que buscan un empleo adicional
- OBTA** Ocupados que buscan un empleo para abandonar el actual
- OBTT** Total de ocupados con búsqueda activa de trabajo
- OCDE** Organización para la Cooperación y el Desarrollo Económico
- OH15** Ocupados que trabajan de 1 a 15 horas a la semana
- OH35** Ocupados que trabajan menos de 35 horas a la semana
- OH35RM** Ocupados que trabajan menos de 35 horas a la semana por razones de mercado (es decir, por motivos involuntarios y no personales)
- OH35SM** Ocupados que laboran más de 35 horas en la semana de referencia con una percepción de ingresos, por su actividad, inferior al salario mínimo
- OH48SM** Ocupados que perciben por su actividad entre uno y dos salarios mínimos, pero con jornadas superiores a las 48 horas semanales
- OISM** Ocupados con ingresos inferiores al salario mínimo
- OIT** Organización Internacional del Trabajo
- OL** Oferta laboral
- PAC** Autocorrelación parcial
- PEA** Población económicamente activa
- PEID** Población económicamente inactiva disponible
- PIB** Producto Interno Bruto
- PM** Población mayor
- PNEA** Población no económicamente activa
- PO** Población ocupada

<b>POSI</b>	Población ocupada en el sector informal
<b>PT</b>	Población total
<b>SBC</b>	Schwarz criterion
<b>SR</b>	Salario real
<b>STPS</b>	Secretaría del Trabajo y Previsión Social
<b>TBP</b>	Tasa bruta de participación
<b>TCCO</b>	Tasa de condiciones críticas de ocupación
<b>TDA</b>	Tasa de desocupación abierta
<b>TDAA</b>	Tasa de desocupación o desempleo abierto alternativa
<b>TIID</b>	Tasa de ingresos inferiores al mínimo y desocupación
<b>TNP</b>	Tasa neta de participación
<b>TO</b>	Tasa de ocupación
<b>TOPD1</b>	Tasa de ocupación parcial y desocupación
<b>TOPD2</b>	Tasa de ocupación de menos de 35 horas y desocupación
<b>TOPRMD</b>	Tasa de ocupación parcial por razones de mercado y ocupación
<b>TOSI</b>	Tasa de ocupación en el sector informal
<b>TPEE</b>	Tasa de presión efectiva económica
<b>TPEP</b>	Tasa de presión efectiva preferencial
<b>TPRG</b>	Tasa de presión general

# INTRODUCCIÓN

---

---

El deterioro de los niveles de empleo representa un problema de gran preocupación para México, sobre todo en aquellas fases de contracción graves de los ciclos económicos, es decir, en las depresiones, donde la producción y el empleo disminuyen.

El desempleo es un fenómeno muy complejo donde interactúan variables sociales. Cuando una persona está desempleada no genera ingresos y por lo tanto no contribuye con impuestos. Como el desempleado no tiene capital, no demanda bienes ni servicios y esto provoca menos demanda, lo que ocasiona que los negocios reciban menos captación de ingresos o dinero y cierren gradualmente. Al cerrar los negocios provocan despidos y crece el desempleo. Es un problema que se va agrandando y agrandando. En México el tema es preocupante ya que existe un gran porcentaje de personas que se sabe trabaja en la economía informal, la cual es una forma de subsistencia, pero causa desequilibrios porque no aporta nada de impuestos.

El ritmo del crecimiento económico y el incremento de la demanda global son fundamentales para crear empleos productivos, sin embargo el problema del desempleo no depende únicamente del número de puestos de trabajo que se generen por la buena marcha de la economía; hay que considerar otras variables, como el crecimiento en las tasas de participación de la población económicamente activa, entre otras, que a su vez dependen de factores socioeconómicos.

El problema del desempleo es complejo, sin embargo esto no debe ser una razón para aminorar el esfuerzo en la búsqueda de opciones para combatirlo. Para tal fin es muy importante conocer los elementos y herramientas con las que se cuenta para una mejor comprensión del problema, así como para dar una estimación, con el fin de buscar opciones de manera anticipada para hacerle frente.

Dentro de lo complejo del problema del desempleo es importante señalar que hay diferencias que se pueden observar entre diversos grupos de la población, por regiones, nivel de escolaridad, edad y por supuesto por sexo.

El desempleo afecta tanto a hombres como a mujeres, pero el impacto que sufren las mujeres que pierden su trabajo es mayor, al tener más dificultad para encontrar otro que les permita equilibrar las responsabilidades de la familia y el trabajo.

La presencia de la mujer en el mundo laboral no es un hecho reciente, desde siempre las mujeres han participado activamente de diversas formas en la producción económica del país; para algunas los recursos obtenidos por sus actividades laborales representan un complemento, pero para muchas que viven en circunstancias diferentes es el único modo de supervivencia familiar.

## **Justificación**

Durante los últimos años el panorama laboral para las mujeres en México ha mejorado, pues han tenido una mayor participación en distintos ámbitos, incluyendo ocupaciones consideradas tradicionalmente masculinas. Las mujeres tienen más facilidad para adaptarse a las innovaciones, se caracterizan por ser disciplinadas y responsables, son más efectivas durante la jornada académica al administrar mejor el tiempo y hacer su trabajo con una alta calidad y clara orientación al detalle. Pese a lo anterior, las mujeres no han logrado alcanzar una proporción equitativa en salarios, prestaciones y posibilidades laborales. Es por todas estas razones, la inquietud por tratar el tema de la desocupación femenina en México.

## **Objetivo general**

El objetivo del presente trabajo es analizar la desocupación femenina en México, señalando las diferencias entre el comportamiento del desempleo de mujeres y hombres, y modelando la tasa de desocupación femenina a través de un modelo ARIMA, para pronosticar la desocupación femenina y contar con información, con cierto grado de probabilidad, de lo que puede pasar con el desempleo para las mujeres en México en un futuro.

## Objetivos específicos

- Buscar información sobre diferentes indicadores que miden la desocupación en México mediante la consulta en el BIE.
- Describir la evolución del desempleo en México durante los últimos años, a partir de la revisión de algunos de los indicadores más importantes.
- Describir el desempleo en México por nivel de instrucción, región y edad.
- Comparar la desocupación femenina y masculina, a través de contrastar algunos indicadores para señalar las diferencias existentes.
- Identificar posibles modelos ARIMA para la tasa de desocupación femenina mediante el proceso de modelación de Box-Jenkins.
- Estimar los parámetros de los modelos ARIMA identificados y comprobar la adecuación del modelo elegido a los datos históricos de la tasa de desocupación para mujeres.
- Pronosticar valores futuros de la tasa de desocupación femenina, a través del modelo elegido.
- Evaluar la capacidad predictiva de los modelos elegidos.

## Organización de la tesis

Para lograr los objetivos antes expuestos se ha distribuido la tesis de la siguiente manera:

Se inicia en el primer capítulo con la revisión de algunos conceptos relacionados con el empleo y los tipos de desempleo.

En el segundo capítulo se pretende describir cuál ha sido la evolución de la desocupación abierta en México durante los últimos años, a partir de la revisión de algunos de los indicadores más importantes. Dentro de estos indicadores, uno de los principales es la tasa de desocupación abierta, difundida por el INEGI.

En el tercer capítulo se estudian algunas técnicas de análisis cuantitativo. Este tipo de análisis consiste en la utilización de un conjunto de reglas matemáticas, estadísticas o lógicas con el objeto de construir y después examinar una representación matemática del mundo económico real, que es lo que se conoce como el modelo. Dentro de este capítulo se introduce el concepto de series de tiempo, sus componentes, varios métodos para corregir posibles problemas de las series, la aplicación de la econometría en ellas, así como la metodología Box-Jenkins que consiste en un proceso iterativo de cuatro etapas: identificación, estimación, comprobación del diagnóstico y predicción, obteniendo los parámetros de un modelo ARIMA.

En el cuarto capítulo se comparan diferentes indicadores sobre el desempleo correspondientes a mujeres y a hombres, y se señalan las diferencias entre las series a través del tiempo. También se identifica, estima y comprueba un modelo ARIMA a través del proceso de Box-Jenkins para la serie de la tasa de desocupación para mujeres. Finalmente se pronostican valores futuros de esta tasa a través del modelo elegido y se evalúa su capacidad predictiva, todo esto con la finalidad de contar con información futura de la desocupación femenina en México y adelantarse al futuro panorama laboral para las mujeres en México.

## **Alcances y limitaciones**

La aplicación de los modelos ARIMA a la modelación de este indicador es una propuesta metodológica que utiliza de forma eficiente la información completa de la serie que se está analizando y proporciona información futura que normalmente no se tiene con antelación. Pero las predicciones deben tener un horizonte de corto plazo, debido a que la relación estadística que vincula a la variable con los valores precedentes de la serie, generalmente contienen pocos términos y siempre los más recientes, por lo que ésta se diluye con el tiempo. Además dicha metodología supone que los datos históricos de la variable influyen en los datos actuales y futuros.

# 1

## PROBLEMA DEL DESEMPLEO

---

### 1.1 Introducción

Uno de los problemas de gran preocupación para muchos países, entre ellos México, es el deterioro de los niveles de empleo, sobre todo en aquellas fases de contracción graves de los ciclos económicos, es decir, en las depresiones, donde la producción y el empleo disminuyen.

Si bien, es cierto, que el ritmo del crecimiento económico y el incremento de la demanda global son fundamentales para crear empleos productivos, también es verdad que el problema del desempleo no depende únicamente del número de puestos de trabajo que se generen por la buena marcha de la economía; hay que considerar otras variables, como el crecimiento en las tasas de participación de la población económicamente activa, entre otras, que a su vez dependen de factores socioeconómicos.

La complejidad del problema del desempleo es muy grande, sin embargo esto no debe ser una razón para aminorar el esfuerzo en la búsqueda de opciones para combatirlo. Para tal fin es muy importante conocer los elementos y herramientas con las que se cuenta para una mejor comprensión del problema del desempleo. Este trabajo busca proporcionar una herramienta más para lograr este entendimiento.

Para ello se inicia en este capítulo con la revisión de algunos conceptos relacionados con el empleo y los tipos de desempleo.



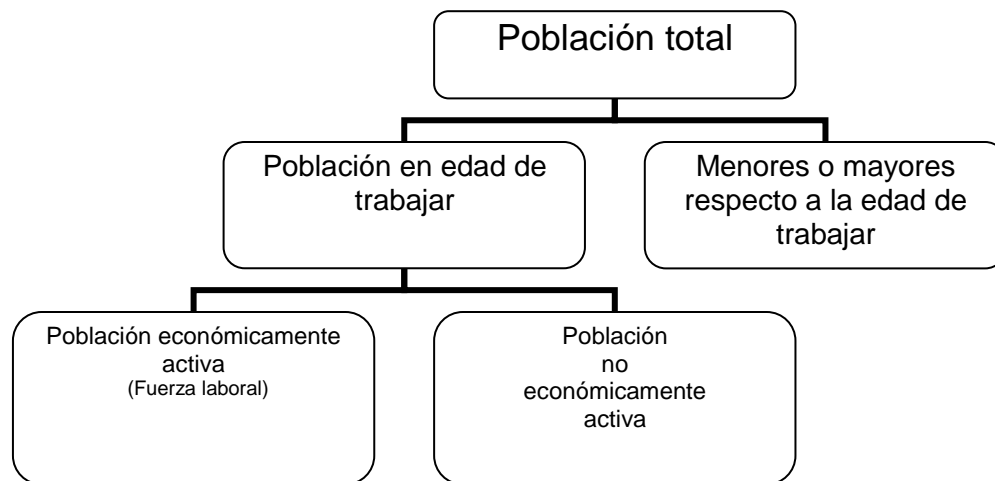
## 1.2 Conceptos relacionados con el empleo y el desempleo

### 1.2.1 Población económicamente activa

La fuerza de trabajo de un país no está constituida por toda la población, pues no todos están en capacidad de laborar, o bien, no todos desean trabajar. Así pues, se puede dividir la población total de un país o una región en dos grupos, la población en edad para trabajar y aquella que no tiene edad para laborar, pues es demasiado joven o vieja para hacerlo. Sin embargo, de las personas en edad para trabajar, no todos los miembros desean o pueden laborar, pues algunos deciden dedicarse a su hogar, a estudiar o a otras actividades, o bien, tienen algún tipo de enfermedad u otro tipo de impedimento. De acuerdo con lo anterior, la población en edad para trabajar puede ser separada en dos grandes grupos: la población económicamente activa y la no económicamente activa.

Por lo tanto, para formar parte de la población económicamente activa o fuerza de trabajo es necesario cumplir dos requisitos: tener edad para trabajar y desear trabajar.

Lo anterior puede resumirse en la figura 1.1.



**Figura 1.1** Población por condición de actividad. (INEGI, 2005)

Debido a su importancia, la información de la fuerza de trabajo de un país generalmente se amplía más asociándola con la estructura de la población, de la producción y de otras características del país. Así pues se puede tener información por:

- Zona: urbana o rural
- Sexo: hombres o mujeres
- Edad
- Rama de actividad económica: agricultura, industrias o servicios
- Sector institucional: sector privado o sector público
- Región: provincias, estados, *etc.*

### **1.2.2 Ocupados y desocupados abiertos**

Al grupo de aquellas personas que pertenecen a la fuerza laboral, se le puede dividir en individuos ocupados y desocupados abiertos. (INEGI, 2005) Aunque existen diversos criterios para determinar quién está ocupado y quién no, en este texto se considera como se hizo en la encuesta de Ocupación y Empleo de 2005 por el INEGI, los ocupados son las personas de 14 o más años de edad que durante la semana de referencia realizaron algún tipo de actividad económica, estando en cualquiera de las siguientes situaciones:

- Trabajando por lo menos una hora o un día para producir bienes y/o servicios de manera independiente o subordinada, con o sin remuneración.
- Ausentes temporalmente de su trabajo sin interrumpir su vínculo laboral con la unidad económica (vacaciones, permiso, enfermedad, *etc.*).

Por otro lado los desocupados abiertos son las personas de 14 o más años de edad que no estando ocupadas en la semana de referencia buscaron activamente incorporarse a alguna actividad económica en algún momento del último mes transcurrido. A su vez este grupo se puede subdividir en: la población cesante (los que han perdido su empleo y poseen experiencia laboral) y la población que busca trabajo por primera vez (no poseen ninguna experiencia laboral).

Por su parte la población no económicamente activa se puede subdividir en dos grupos, el primero integrado por amas de casa, estudiantes y jubilados o pensionados, que no buscan trabajo, pero que estarían dispuestos a aceptar uno; y el segundo por las amas

de casa, estudiantes de tiempo completo, jubilados y discapacitados, que no trabajan ni pueden o quieren buscar trabajo. Todo esto se presenta en la figura 1.2.

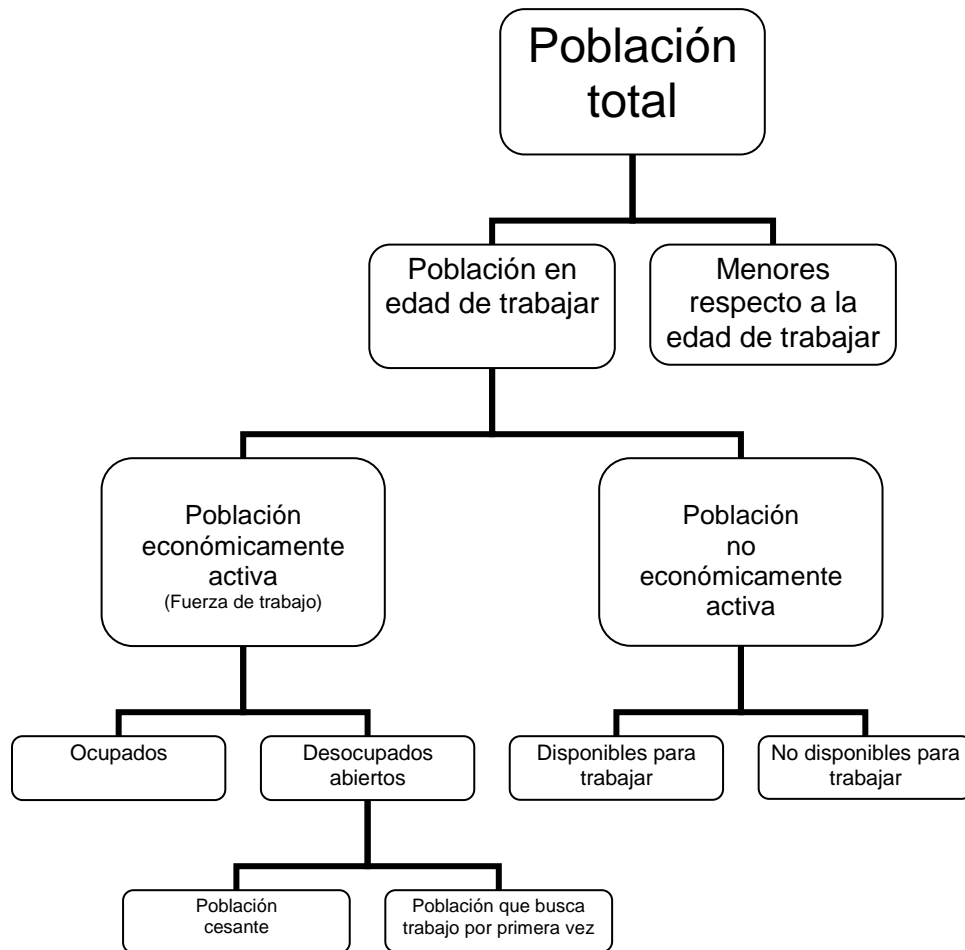


Figura 1.2 Población por condición de actividad. (INEGI, 2005)

### 1.2.3 Concepto de desempleo

El concepto de desempleo abierto quedó definido por la Organización Internacional del Trabajo (OIT) desde la XIII Conferencia Internacional de Estadígrafos del Trabajo (CIET): “El conjunto de individuos en edad de trabajar que en períodos de referencia especificados presentan tres características:

- a) **Estar sin trabajo**: significando esto que no contaban con un empleo ni tampoco realizaban alguna actividad por cuenta propia.

- b) **Disponibles para trabajar**: es decir con disponibilidad para tomar un empleo remunerado o iniciar una actividad por su cuenta.
- c) **En situación de búsqueda de trabajo**: lo que implica haber realizado alguna acción en el período de referencia con la intención de obtener un empleo o intentar comenzar alguna actividad por cuenta propia sin haberlo logrado. Ser miembro de la fuerza de trabajo, es decir, tener edad para trabajar y deseo de trabajar.” (INEGI, 2002)

### 1.3 Medición del empleo y la desocupación abierta

El empleo y la desocupación abierta se pueden medir a través de diferentes indicadores, los cuales se calculan con base en la información obtenida en las encuestas, las cuales se mencionan en el siguiente capítulo. Entre estos indicadores se pueden mencionar (INEGI, 2002):

**Tasa bruta de participación (TBP)**: Porcentaje que representa la población económicamente activa respecto a la población total.

$$TBP = \frac{PEA}{PT} (100)$$

*PEA* = Población económicamente activa

*PT* = Población total

**Tasa neta de participación (TNP)**: Porcentaje que representa la población económicamente activa respecto a la población mayor (conjunto de personas en edad de trabajar, es decir, de 12 y más años).

$$TNP = \frac{PEA}{PM} (100)$$

*PEA* = Población económicamente activa

*PM* = Población mayor

**Tasa de ocupación (TO):** Porcentaje de la población ocupada respecto a la población mayor.

$$TO = \frac{PO}{PM} (100)$$

*PO* = Población ocupada

*PM* = Población mayor

**Tasa de desocupación abierta (TDA):** Porcentaje de la población económicamente activa que se encuentra sin trabajar, pero que está buscando trabajo. Coloquialmente llamada tasa de desempleo.

$$TDA = \frac{DA}{PEA} (100)$$

*DA* = Desocupados abiertos

*PEA* = Población económicamente activa

En México como en otros países, el concepto de desempleo abierto exige el cumplimiento de ciertas condiciones, esto provoca que no todos los sectores modernos de la sociedad puedan describirse en estos términos. Para lograr esto último se crean las tasas complementarias (INEGI, 2009), las cuales se dividen en dos grupos dependiendo del objetivo que cubren, como se observa en el figura 1.3.

**Tasas de situación:** Ampliar la perspectiva situacional, sin que intervenga como restricción la condición de un comportamiento preciso en términos de búsqueda.

**Tasas de comportamiento:** Captar un radio más amplio de un comportamiento participativo en los mercados de trabajo, sin que la situación en términos de la semana de referencia sea una limitante.

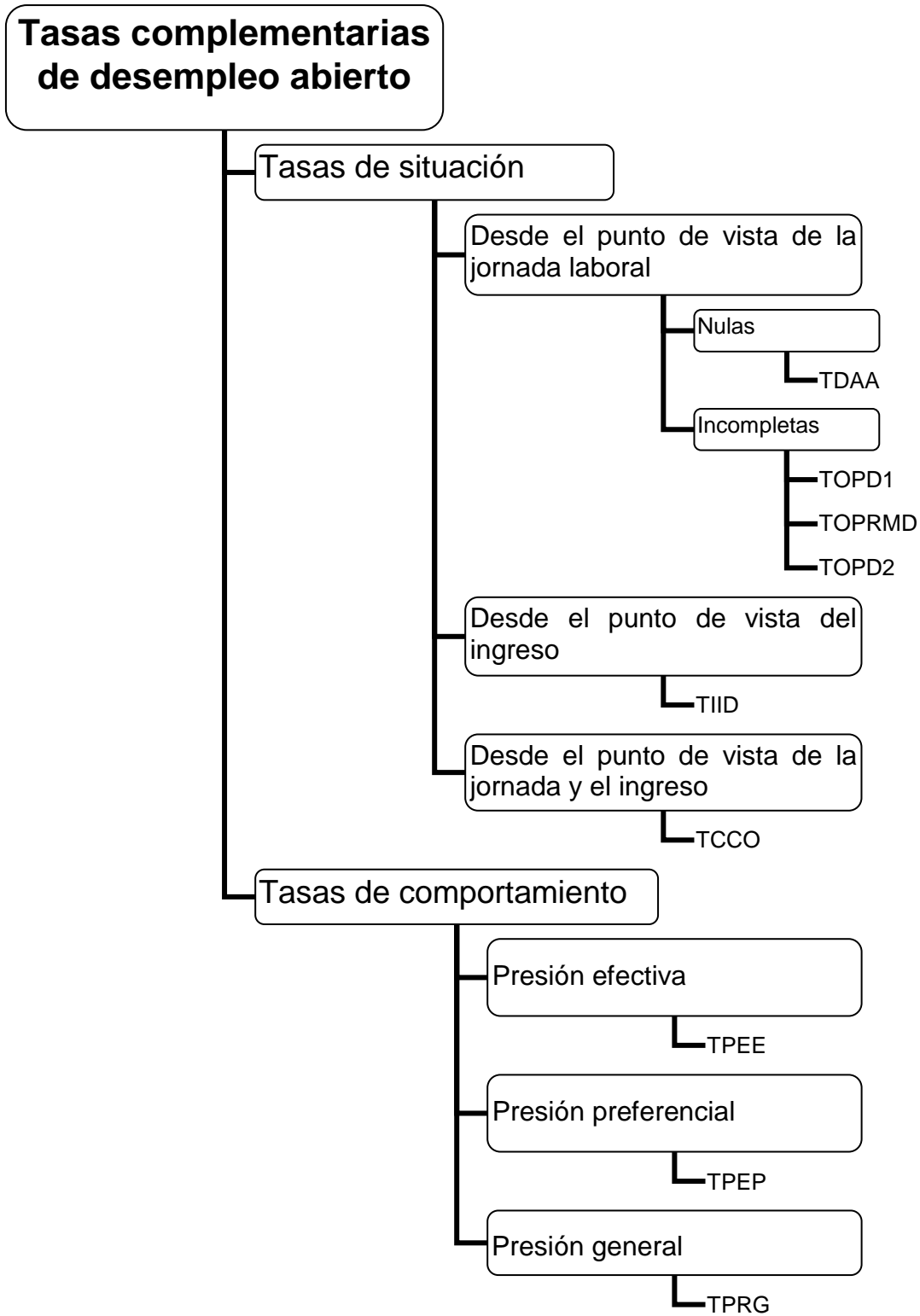


Figura 1.3 Indicadores complementarios a la desocupación abierta. (INEGI, 2009)

**Tasa de desocupación o desempleo abierto alternativa (TDAA):** Es la proporción de la población económicamente activa y de la población no económicamente activa que se encuentra desocupada o que está disponible para el trabajo aunque haya abandonado la búsqueda del mismo.

$$TDAA = \left( \frac{DA + PEID + I}{PEA + PEID} \right) 100$$

*DA* = Desocupados abiertos

*PEID* = Población económicamente inactiva disponible

*I* = Iniciadores (individuos que no trabajaron pero que declararon estar por iniciar una ocupación)

*PEA* = Población económicamente activa

**Tasa de ocupación parcial y desocupación (TOPD1):** Porcentaje de la población económicamente activa que se encuentra desocupada, más la ocupada que trabajó menos de 15 horas en la semana de referencia.

$$TOPD1 = \left( \frac{DA + OH15}{PEA} \right) 100$$

*DA* = Desocupados abiertos

*OH15* = Ocupados que trabajan de 1 a 15 horas a la semana

*PEA* = Población económicamente activa

**Tasa de ocupación parcial por razones de mercado y ocupación (TOPRMD):** Porcentaje de la población económicamente activa que se encuentra desocupada, más la ocupada que trabajó menos de 35 horas en la semana de referencia por razones de mercado.

$$TOPRMD = \left( \frac{DA + OH35RM}{PEA} \right) 100$$

*DA* = Desocupados abiertos

*OH35RM* = Ocupados que trabajan menos de 35 horas a la semana por razones de mercado (es decir, por motivos involuntarios y no personales)

*PEA* = Población económicamente activa

**Tasa de ocupación de menos de 35 horas y desocupación (TOPD2):** Porcentaje de la población económicamente activa que se encuentra desocupada, más el total de población ocupada con jornadas de trabajo incompletas en la semana de referencia, sea por razones voluntarias (ausentismo), personales involuntarias (salud) o por razones de mercado; incluye a los ocupados en empleos de medio tiempo.

$$TOPD2 = \left( \frac{DA + OH35}{PEA} \right) 100$$

*DA* = Desocupados abiertos

*OH35* = Ocupados que trabajan menos de 35 horas a la semana

*PEA* = Población económicamente activa

**Tasa de ingresos inferiores al mínimo y desocupación (TIID):** Porcentaje de la población económicamente activa que se encuentra desocupada, más el total de ocupados con percepciones inferiores al salario mínimo, englobando a la porción de la PEA que no se beneficia de ese mismo ingreso.

$$TIID = \left( \frac{DA + OISM}{PEA} \right) 100$$

*DA* = Desocupados abiertos

*OISM* = Ocupados con ingresos inferiores al salario mínimo

*PEA* = Población económicamente activa



**Tasa de condiciones críticas de ocupación (TCCO):** Porcentaje de la población ocupada que se encuentra trabajando menos de 35 horas a la semana por razones de mercado, más la que trabaja más de 35 horas semanales con ingresos mensuales inferiores al salario mínimo y la que labora más de 48 horas semanales ganando hasta dos salarios mínimos.

$$TCCO = \left( \frac{OH35RM + OH35SM + OH48SM}{PO} \right) 100$$

*OH35RM* = Ocupados que trabajan menos de 35 horas a la semana por razones de mercado (es decir, por motivos involuntarios y no personales)

*OH35SM* = Ocupados que laboran más de 35 horas en la semana de referencia con una percepción de ingresos, por su actividad, inferior al salario mínimo

*OH48SM* = Ocupados que perciben por su actividad entre uno y dos salarios mínimos, pero con jornadas superiores a las 48 horas semanales

*PO* = Población ocupada

**Tasa de presión efectiva económica (TPEE):** Porcentaje de la población económicamente activa que se encuentra desocupada, más el conjunto de ocupados en situación de búsqueda de un empleo adicional.

$$TPEE = \left( \frac{DA + OBT}{PEA} \right) 100$$

*DA* = Desocupados abiertos

*OBT* = Ocupados que buscan un empleo adicional

*PEA* = Población económicamente activa

**Tasa de presión efectiva preferencial (TPEP):** Porcentaje de la población económicamente activa que se encuentra desocupada, más el conjunto de ocupados que buscan un empleo para abandonar el actual.

$$TPEP = \left( \frac{DA + OBTA}{PEA} \right) 100$$

*DA* = Desocupados abiertos

*OBTA* = Ocupados que buscan un empleo para abandonar el actual

*PEA* = Población económicamente activa

**Tasa de presión general (TPRG):** Porcentaje de la población económicamente activa que se encuentra desocupada, más el total de ocupados que hayan realizado una búsqueda activa de trabajo en el período de referencia.

$$TPRG = \left( \frac{DA + OBTT}{PEA} \right) 100$$

*DA* = Desocupados abiertos

*OBTT* = Total de ocupados con búsqueda activa de trabajo

*PEA* = Población económicamente activa

**Tasa de ocupación en el sector informal (TOSI):** Porcentaje de la población ocupada, que trabaja para una unidad económica que opera a partir de los recursos del hogar, pero sin constituirse como empresa, de modo que la actividad no tiene una situación identificable e independiente de ese hogar. La manera operativa de establecer esto es que la actividad no lleva una contabilidad bajo las convenciones que permiten que sea auditada.

$$TOSI = \left( \frac{POSI}{PO} \right) 100$$

*POSI* = Población ocupada en el sector informal

*PO* = Población ocupada

En el siguiente capítulo se analizan algunos de estos indicadores para el caso particular de México.

## **1.4 Tipos de desempleo**

Existen tres tipos de desempleo dentro de la economía de una sociedad: desempleo friccional, desempleo estructural y el desempleo cíclico. Estos tipos de desempleo se asocian con las causas del desempleo como se explica enseguida. (Reyes, 1995)

### **1.4.1 Desempleo friccional**

El desempleo friccional se puede definir como el movimiento de los individuos, ya sea, por búsqueda de trabajo o por hallar mejores oportunidades laborales. Dentro de este tipo de desempleo se encuentran los casos donde el empleado tiene la libre elección para renunciar de su puesto, así como también aquellos donde los empresarios despiden al trabajador cuando lo consideren necesario, debido a las discrepancias entre las características del puesto del trabajo y el empleado.

Este tipo de desempleo no se puede erradicar por completo, pero como una política adecuada para disminuirlo se pueden mejorar los sistemas de información, para que sea más fácil a las personas hallar un nuevo empleo.

La cantidad de desempleo por fricción depende de la tasa a la cual la gente entra y sale de la fuerza de trabajo y de la tasa a la cual se crean y destruyan los empleos.

### **1.4.2 Desempleo estructural**

Técnicamente corresponde a un desajuste entre oferta y demanda de trabajadores, debido a la falta de preparación académica y de destrezas de las personas, pues al no haber una capacitación en el manejo de herramientas de trabajo, producidas por los cambios en las estructuras industriales, tecnológicas, ocupacionales y demográficas de la economía, se ve reducida la demanda de trabajo, y entonces es difícil a las personas con experiencia laboral que han perdido su trabajo, el encontrar empleo. Este tipo de desempleo se caracteriza porque las personas no logran satisfacer los requisitos de los empleadores, es fácil encontrar muchas ofertas de trabajo, pero muchas veces es difícil

llenar todos los requisitos (preparación académica, experiencia, manejo de idiomas y herramientas tecnológicas, edad, *etc.*).

Las políticas para hacer frente a este tipo de desempleo deben orientarse al fomento de programas de formación y reconversión profesional, y el establecimiento de nuevas actividades productivas con futuro.

### **1.4.3 Desempleo cíclico**

También conocido como desempleo por insuficiencia de la demanda agregada. Se refiere a la fluctuación en el desempleo causada por los ciclos económicos. Cuando el ritmo de crecimiento de la demanda agregada es mayor a lo normal, la demanda por trabajo es superior a lo corriente y el desempleo disminuye, pero ocurre lo contrario en las fases recesivas del ciclo económico, en las cuales se reduce la demanda y entonces el desempleo se incrementa.

Para enfrentar este tipo de desempleo puede considerarse como política económica los programas de reactivación económica, basados en políticas coyunturales o de estabilización, como por ejemplo, las políticas monetarias y fiscales.

## **1.5 Consecuencias del desempleo**

El trabajo les brinda a las personas la posibilidad de lograr ingresos, por lo tanto, pueden acceder a bienes y servicios, y de esta manera satisfacen sus necesidades. En la medida que logren satisfacer mejor sus necesidades, mejor será su nivel de vida. Es por eso que existe un fuerte vínculo entre el desempleo y la pobreza. Ahora bien, no sólo es importante el tener empleo, sino la calidad del empleo, es decir, el salario, la jornada laboral, las condiciones de empleo, la estabilidad, la seguridad social, entre muchos elementos.

Cabe destacar que el desempleo ocasiona a la sociedad tanto un costo económico como social. El costo económico corresponde a todo lo que se deja de producir y que será imposible de recuperar, esto no solo incluye los bienes que se pierden por no

producirlos sino también una cierta degradación del capital humano, que resulta de la pérdida de destrezas y habilidades. Por otro lado, el costo social abarca, la pobreza e "inquietud social y política" que implica el desempleo en grandes escalas. Las personas sometidas a una ociosidad forzosa padecen frustración, desmoralización y pérdida del amor propio. Aunque este costo social es muy difícil de medir, es objeto de una profunda y general preocupación por el impacto que tiene en la sociedad.

# 2

## EMPLEO Y DESOCUPACIÓN ABIERTA EN MÉXICO

---

### 2.1 Introducción

En este capítulo se pretende describir cuál ha sido la evolución de la desocupación abierta en México durante los últimos años, a partir de la revisión de algunos de los indicadores más importantes.

Para llevar a cabo la revisión antes señalada, previamente se ubica a la desocupación abierta en el contexto laboral, se identifican los factores que inciden en su nivel y variación, se dan a conocer algunas de las encuestas que se han utilizado en México para la medición del empleo y la desocupación, y se comentan algunas características de la información obtenida de estas encuestas para que sean consideradas en el análisis de los resultados.

### 2.2 Desempleo abierto en el contexto del mercado laboral

El desempleo es un fenómeno que aparece en los últimos 200 años de la historia de la civilización, en contraste con la pobreza que es un padecimiento milenario, además ha acompañado a esa transformación conocida como modernidad. Esto ha sido así porque al predominar una economía de autosuficiencia o una economía mercantil pero de autoempleo, pueden padecerse las peores limitaciones, pero no ocurre la

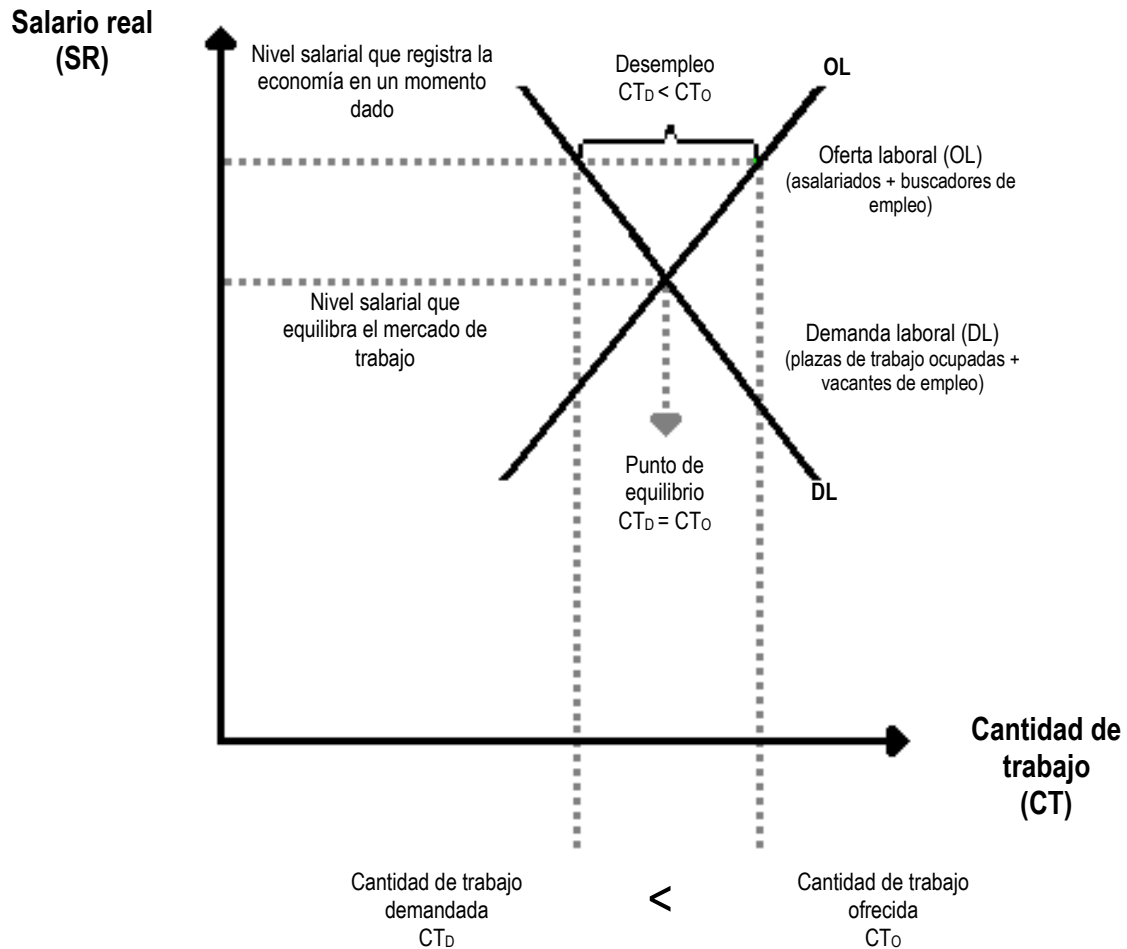
desocupación. Esto supone que el trabajo mismo se convierte en una mercancía que se ofrece a quien quiere utilizarla, del mismo modo que se ofrecen materias primas o insumos a un productor que requiera de ellos. Así, por un lado se tiene la oferta de un servicio, que es el trabajo, y la posible demanda que de él ejerce un empleador.

Estos elementos constituyeron un mercado nuevo que hasta entonces no se conocía: el mercado laboral. Cuando ese servicio que se ofrece no encuentra la demanda correspondiente, sucede entonces el fenómeno que se conoce como desempleo. No hay pues desocupación si no hay mercados laborales. No se es desempleado si no se hace un ofrecimiento activo de ese servicio que le pudiera interesar a un empleador, es decir, si no se busca esa demanda específica. Esta idea que pudiera parecer elemental, se olvida con frecuencia cuando se analizan los indicadores relativos al desempleo, o cuando se identifica indebidamente este fenómeno con otro tipo de carencias, haciéndolo incluso sinónimo de ellas. (INEGI, 2002)

En la figura 2.1 se ilustra cómo es que el fenómeno de la desocupación está necesariamente ligado al contexto de un mercado laboral. El eje de las abscisas (horizontal) representa la cantidad de empleo y el de las ordenadas (vertical) su precio o salario real; la curva DL corresponde a la demanda de los insumos laborales por parte de los empleadores y tiene pendiente negativa porque a más salario, menos demanda de dichos insumos; la curva OL representa a la oferta disponible de fuerza laboral por parte de la población y su pendiente es positiva porque, a más salario, más disponibilidad habrá para incorporarse a los mercados de trabajo. Un salario de equilibrio será aquél donde se cruzan ambas curvas, en cambio, un salario de desequilibrio provoca un desencuentro: mientras menos corresponda a la realidad del mercado mayor será la brecha existente entre DL y OL. Si la diferencia refleja más oferta por parte de la población que demanda por parte de los empleadores, gráficamente OL a la derecha de DL, esto representa desempleo, por el contrario si se observa que DL quedara a la derecha de OL, es decir, si la excediera, se tendrá escasez laboral. (INEGI, 2002)

Es importante señalar que mientras no haya un incremento sustantivo, sostenido y generalizado de la productividad laboral de toda la economía, siempre habrá un dilema

entre favorecer el salario a costa del nivel de empleo, o favorecer el nivel de empleo a costa del poder adquisitivo del salario.



**Figura 2.1** El mercado de trabajo. (INEGI, 2002)

Lo anteriormente expuesto es importante para entender lo que sucede con la desocupación abierta en México. La intención de integrarse a mercados de trabajo, es un comportamiento menos generalizado de lo que se cree por parte de la población económicamente activa en nuestro país. Esto es esencial para entender que no es el criterio de no trabajar ni una hora a la semana (situación), lo que hace comparativamente bajas las tasas de desocupación abierta en México. Además, el salario real que aproxima a un equilibrio en el mercado de trabajo (cantidad ofrecida de



trabajo igual a cantidad demandada), no es necesariamente el que garantiza cierto nivel de vida.

El que no toda la fuerza laboral en activo acuda a los mercados de trabajo en México significa que, a diferencia de las naciones desarrolladas, no existe en México una cultura de trabajo asalariado homogéneamente extendida a lo largo y ancho del país; esto aplica no sólo en el ámbito rural, sino también en el ámbito urbano.

Un mercado laboral, como cualquier otro mercado, cuando no se ajusta en términos de precios se ajusta en términos de cantidades físicas. Si el precio de un bien o servicio no responde a la baja en su demanda, o al exceso en su oferta, la cantidad colocada de ese bien o servicio disminuirá en el mercado.

Cuando los salarios presentan rigidez, porque los convenios salariales fueron acordados con mucha antelación, bajo circunstancias muy distintas a las presentes, o son indexados, es decir, son aumentados de manera mayor o igual a la inflación sin tomar en cuenta si ha habido mayor productividad, los mercados laborales terminan ajustando, por el lado de las cantidades, lo que no pudieron hacer por el lado de los precios, es decir se producen los despidos.

Los mercados laborales son flexibles cuando los salarios son los que se ajustan rápidamente ante los cambios en la oferta o en la demanda; por el contrario, son rígidos cuando no reaccionan a esos cambios manteniendo un poder adquisitivo, pero teniendo como consecuencia la pérdida de empleo. (INEGI, 2002)

Debido a lo anterior, se puede decir, que el desempleo abierto no es una medida del rezago social de un país, así como tampoco los cambios en su tasa significan que necesariamente esté mejor o peor que antes en términos de ingresos: la tasa de desocupación abierta lo único que señala es qué tan lejos o qué tan cerca está una economía de lograr el equilibrio en su mercado laboral.

Ver en el desempleo abierto un indicador de pobreza o marginación es un error de interpretación; la pobreza y la marginación históricamente, han existido mucho antes que se constituyeran los mercados de trabajo.

Ciertamente, la pobreza y la marginación, hablan de una carencia de oportunidades, pero el desempleo abierto habla de un comportamiento específico ante una carencia de oportunidades; si el individuo no tiene intención de vincularse a un mercado de trabajo, seguirá siendo un marginado, más no un desocupado abierto.

Los grupos sociales que guardan un mínimo de expectativas participan en los mercados de trabajo; los grupos que carecen de toda expectativa se apartan de los mercados de trabajo.

## **2.3 Factores que inciden en el nivel y variación de la desocupación abierta**

Además de la flexibilidad salarial y de una cultura laboral que no es homogéneamente participativa en mercados de trabajo, es importante considerar otros factores económicos, socio-demográficos e institucionales que afectan a la desocupación abierta.

El primero que se mencionará es la variación en el producto interno bruto (PIB), como un factor básicamente de comportamiento; sin duda, un crecimiento del PIB, se traduce en una tasa de desempleo a la baja.

“La relación existente entre la tasa de crecimiento y la tasa de desempleo se denomina “Ley de Okun” en honor al economista Arthur Okun. En la mayoría de países, existe una relación fiable entre la tasa de crecimiento del PIB y la variación de la tasa de desempleo. Okun, descubre esta relación en la década de los 60's. Se sostiene básicamente que, un elevado crecimiento crea un incremento en el empleo del país y por ende, una disminución en la tasa de desempleo. ¿Por qué? Porque las empresas deben contratar una mayor cantidad de trabajadores para producir más.

La tasa de desempleo es un objetivo básico de todo gobierno. Por ello, es necesario conocer la magnitud en la que una economía debe crecer si desea mantener o reducir la tasa de desempleo en un país. De esta manera, la tasa de desempleo para los macroeconomistas, representa la situación de la economía y da pautas para desear una

tasa de crecimiento del producto deseable para reducir la tasa de desempleo.” (Velasategui, 2006)

Para ahondar más específicamente en el caso de México se puede revisar un estudio económico realizado en El Colegio de México. (Loría y Ramos, 2007)

Ahí, donde termina el poder explicativo del crecimiento económico es donde entran todos los factores que tienen que ver con la flexibilidad o rigidez de los mercados de trabajo, tales como el marco institucional y el grado en que los salarios son sensibles al mercado laboral y la productividad; de las características de ambos depende que se logre o no una conexión eficaz entre crecimiento y ocupación.

La estructura social, por su parte, también influye en el nivel de las tasas de desempleo, en la medida en que hace de éste un comportamiento generalizable o no para toda la población económicamente activa.

Dado lo anterior, es importante no reducir al desempleo abierto a un fenómeno meramente de ciclo económico sin pasar por el examen de un mercado específico como el laboral; tampoco hay que verlo como un fenómeno meramente de coyuntura, porque en él también operan factores estructurales que le dan una ubicación en el contexto de cada sociedad.

Los factores que inciden en el nivel de la tasa de desocupación abierta y que no han sido mencionados se pueden clasificar en dos tipos, los institucionales y los socio-demográficos.

Dentro de los factores institucionales están los que tienen que ver con el marco regulatorio y normativo diseñado para la protección de la fuerza laboral. A estos factores se les puede clasificar en dos categorías, los que afectan a quienes ofrecen sus servicios laborales y los que afectan el proceder de quienes requieren de esos servicios.

Sobre los factores que afectan a los trabajadores destaca la figura del llamado seguro de desempleo. Este se resume en garantizar un ingreso corriente a quien ha perdido su empleo; lo que ha muchos hace pensar que mientras se cuenta con dicho respaldo, el comportamiento del individuo consiste en resistir periodos más largos de nula

ocupación, debido a que percibe el ingreso del seguro, mientras se dedica exclusivamente a buscar empleo bajo la cobertura de dicho seguro que, según cada legislación nacional, tiene cierta duración.

En México se han presentado diferentes iniciativas en relación al establecimiento de un seguro de desempleo sin haber fructificado hasta el momento; por mencionar alguna de ellas, se sabe que el presidente Felipe Calderón firmó el 9 de febrero de 2009 las iniciativas de reforma para diversas disposiciones de las leyes del IMSS, y del Infonavit, con las que se abre paso a una especie de seguro de desempleo por la vía de las afores. (Periódico Excélsior, 2009) En contraste, el gobierno del D. F. desde el 2006 puso en funcionamiento el programa de Seguro de Desempleo del Distrito Federal, sistema de protección para los trabajadores del D. F. que cumplan con ciertos requisitos<sup>1</sup>, que se traduce en beneficio económico mensual equivalente a 30 días de salario mínimo vigente en el D. F. entregado por un periodo hasta de 6 meses. (Gaceta oficial del D. F., 2009)

Cuando no se dispone de dicho seguro, ante la negativa de los mercados laborales, una mayor proporción de individuos se ven obligados a buscar otras opciones, como auto emplearse o tomar trabajos de carácter eventual, emigrar, etc.; ya que debe disminuirse al máximo la desocupación. Así, la ausencia del seguro de desempleo determina que haya sectores de la población que no puedan plantearse la disyuntiva de todo o nada, estar plenamente ocupados o no estarlo en absoluto, con lo que dicha ausencia abre un rango de situaciones posibles entre esas dos posiciones que es, precisamente, lo que consideran algunos de los indicadores complementarios al desempleo abierto.

En cuanto a los factores institucionales que inciden en el comportamiento de los empleadores cabe destacar lo mismo a las legislaciones laborales, que a los sistemas de seguridad y su cobertura. En algunas sociedades se ha observado como la necesidad de proteger a la clase trabajadora de todo tipo de abusos, ha obligado a la

---

<sup>1</sup> Haber laborado previamente a la pérdida del empleo, para una persona moral o física con domicilio fiscal en la Ciudad de México, al menos durante seis meses; haber perdido su empleo a partir del 1º de enero de 2006, por causas ajenas a su voluntad; no percibir ingresos económicos por concepto de jubilación, pensión, subsidio o relación laboral diversa, y ser demandante activo de empleo. (Gaceta oficial del D. F., 2009)

construcción, con el tiempo, de un marco legal y contractual en el que juegan un papel relevante los sindicatos, siendo su preocupación natural sus agremiados, no quienes no lo son, influyendo en muchas naciones en la introducción de cláusulas que contemplan desde prestaciones hasta indemnizaciones y despidos que, desde la perspectiva del empleador, han incrementado lo mismo los costos extra salariales del factor trabajo que el riesgo de contratarlo. En naciones con una cobertura mucho más extendida de la seguridad social que la de algunos países en desarrollo, se impide evadir esos costos, evasión que le da la fluidez al mercado de trabajo a costa de desprotegerlo. De ahí que en ciertos países ha habido disputas recurrentes entre gobierno y sindicatos para crear figuras de excepción dentro del marco normativo, que van desde la creación del estatuto de trabajador eventual, hasta disposiciones especiales que estimulen la contratación de trabajadores inexpertos por vía de un relajamiento en las obligaciones de las empresas.

En lo anterior se ilustran las disyuntivas que pueden caracterizar a las economías de mercado en un momento dado, dependiendo de la correlación de fuerzas a su interior, naciones que aún y cuando comparten un ambiente macroeconómico –mismo ritmo de crecimiento e inflación-, e incluso, una misma dinámica demográfica, pueden responder ante esta disyuntiva de manera diferente, unas inclinándose por la protección del trabajador, otras por la disminución del desempleo.

Sobre los factores socio-demográficos, uno que debe tomarse en cuenta al analizar el nivel de las tasas de desempleo abierto, es la migración internacional, en la medida en que ésta, al implicar un grupo de fuerza laboral que acude a mercados de trabajo en el extranjero, evita que dicho contingente impacte en el desempleo abierto del país de origen. Tal es el caso de la migración de mexicanos a Estados Unidos por motivos laborales, con una presencia mayoritaria de trabajadores rurales y hacia el sector agrícola, ésta afecta a las tasas de desempleo abierto e incluso explica el porqué las tasas de desempleo abierto en zonas rurales resultan menores a las urbanas en nuestro país. Así pues, el nivel del desempleo abierto en México no debe verse como una medida de déficit de oportunidades laborales, factor que origina la migración, sino como la expresión del momento que presenta el mercado laboral, cualquiera que sean las causas subyacentes que configuran el estado que en dicho momento manifiestan

tanto la oferta como la demanda de trabajo. En pocas palabras, cuando una nación es influida fuertemente por mercados laborales globalizados su tasa de desempleo debe mirarse como un efecto o consecuencia más que como un fenómeno causante de otros.

## **2.4 Encuestas de ocupación en México**

Es conocido que la contabilización del número de habitantes de un país se realiza a través de un censo o de un conteo de población y vivienda, en tanto que el seguimiento en el mediano y corto plazo de los cambios que presenta dicha población y de las situaciones específicas que le afectan, se realiza a partir de encuestas en hogares.

En México se tiene una plataforma censal renovada quinquenalmente, por medio del Conteo de Población y Vivienda, complementada por un sistema de encuestas con producción de datos bianual, trimestral y mensual que, además de posibilitar el análisis a mediano y corto plazo, incorporan la precisión o detalle propio de los instrumentos de captación más especializados. En materia de hogares, las dos grandes vertientes temáticas en México corresponden a las encuestas de ingreso-gasto y a la de empleo. (INEGI, 2005)

Desde 1972 se dio inicio a las encuestas en hogares, con una encuesta de propósitos múltiples denominada Encuesta Nacional en Hogares (ENH); poco después se separa el módulo de empleo, dando origen a la Encuesta Continua de Mano de Obra (ECMO 1973-1974), que posteriormente se convierte en la Encuesta Continua sobre Ocupación (ECSO 1974-1984).

Es en 1983-1984 cuando, después de un periodo de evaluación, surge la Encuesta Nacional de Empleo Urbano (ENEU) con la finalidad de contribuir al conocimiento y análisis del mercado de trabajo. Ésta se levantó en forma paralela con la ECSO durante un par de años.

Como resultado de la evaluación de estos dos proyectos estadísticos paralelos se diseñó un nuevo cuestionario que inició en 1985, el cual rompió con la estructura clásica de la ECSO, profundizó en algunos temas como las condiciones de trabajo, el sector de propiedad y tipo de establecimientos, la presión que ejercieron los ocupados

que buscaron un segundo trabajo, entre otros muchos aspectos. Este nuevo cuestionario mantuvo el mismo nombre de la ENEU, que tuvo una vigencia de 20 años, durante los cuales en varias ocasiones se ajustaron los instrumentos de captación. La cobertura geográfica de esta encuesta inició con las tres principales áreas metropolitanas del país (Ciudad de México, Guadalajara y Monterrey); con el paso de los años logró tener una cobertura de aproximadamente 93% de las áreas urbanas del país, llegando en su punto más alto al tener 48 ciudades en muestra, en el segundo trimestre de 2003.

A diferencia de muchos países, y en particular de los miembros de la Organización para la Cooperación y el Desarrollo Económico (OCDE), en México las encuestas de empleo en hogares comenzaron teniendo representatividad por ciudad, para después pasar a encuestas con cobertura nacional. Esto debido a la necesidad de contar con elementos que permitieran conocer la situación de ocupación en el área rural. Así en 1988 el INEGI, en convenio de colaboración con la Secretaría del Trabajo y Previsión Social (STPS), realizó por primera vez la Encuesta Nacional de Empleo (ENE), cubriendo todo el territorio nacional y desarrollándola con una periodicidad bianual hasta 1995. A partir de 1996 la ENE registra un importante avance al llevarse a cabo en forma anual (segundo trimestre), sin embargo, en virtud de que esta periodicidad no satisfacía los requerimientos de los usuarios, así como la importancia cada vez mayor que adquirió la información del empleo y desempleo, a partir del año 2000, el INEGI con el aval de la STPS, se dio a la tarea de levantar la encuesta en forma permanente, llevándose a cabo trimestralmente. (INEGI, 2005)

La Encuesta Nacional de Empleo Trimestral (ENET) tenía como propósito ofrecer una base de información estadística sobre las características ocupacionales de la población a nivel nacional, por entidad federativa, por tamaño de localidad y para las principales ciudades, así como sobre otras variables demográficas y económicas que permitieran profundizar en el análisis coyuntural de los aspectos laborales del país.

A partir de enero de 2005, la ENET fue sustituida por la Encuesta Nacional de Ocupación y Empleo (ENOE), la cual es resultado de una evaluación integral de varios años, durante los cuales el INEGI realizó una serie de actividades de carácter

conceptual, metodológico y de mejora de procesos, con la finalidad de instaurar una nueva encuesta que permitiera captar y conocer de mejor manera las características del mercado laboral mexicano.

El objetivo general de la ENOE es garantizar que se cuente con una base de información estadística sobre las características ocupacionales de la población a nivel nacional, así como una infraestructura socio-demográfica que permita profundizar en el análisis de los aspectos laborales. Mientras que los objetivos específicos son (INEGI, 2005):

- Dar una respuesta desde el plano de la información a temas emergentes en materia de ocupación, desocupación y empleo.
- Poner al día los instrumentos con respecto a los marcos conceptuales que surgieron en los últimos años.
- Ampliar la oferta de indicadores de carácter estratégico para el conocimiento cabal de la realidad nacional y la toma de decisiones con vías a la formulación de políticas laborales.
- Innovar tecnológicamente los procesos y los operativos de control de calidad de la encuesta de ocupación y empleo.

Además, el INEGI, tomando como referencia los 10 apartados que conforman los indicadores estratégicos trimestrales, se dio a la tarea de reconstruir la información captada por la ENE, entre 1995 y 2004, bajo los criterios de clasificación y de marco conceptual de la ENOE, para ofrecer series estadísticas más largas que las disponibles, incluyendo los años anteriores a 2005, a partir de los datos de la ENE. El usuario habrá de tomar en cuenta, sin embargo, que entre 1995 y la primera mitad del año 2000 la ENE sólo se levantaba los segundos trimestres y con representatividad estatal en los años pares, es decir, 1996, 1998 y 2000, en tanto que, en los años nones, la representatividad se limitó al ámbito nacional. De la segunda mitad del año 2000 en adelante, se tiene información trimestre a trimestre con representatividad tanto nacional como para cada una de las 32 entidades federativas, sin interrupción.



La conversión de 1995 hasta 2004 (período abarcado por la ENE) se pudo lograr en más del noventa por ciento de los ítems comprendidos por los Indicadores Estratégicos, en tanto que el remanente se refiere a desagregaciones específicas y novedosas que introduce ENOE para ciertos universos o subgrupos que no son replicables en ENE.

Los resultados de este nuevo modelo han prevalecido en la entrega y presentación de la información y los puntos de contacto con su predecesora la ENE, sólo han sido explorados, hasta ahora, para generar series estadísticas referidas a algunas tasas que forman parte de lo que se ha denominado "series unificadas", pero sin llevar más lejos el proceso de reconstrucción que permitiera al público usuario el disponer también de información armonizada en lo que a datos absolutos se refiere. (INEGI, 2005)

## **2.5 Características de la información para su análisis**

### **2.5.1 Contraste entre las tasas de desocupación por niveles de escolaridad**

Como se sabe, México es un país de enormes necesidades y la población con algún tipo de carencia se encuentra extendida a lo largo y ancho del país. Sin embargo, la presión, de aquellos que buscan activamente empleo, que se ejerce en los mercados de trabajo se centra en ciertos grupos de la población.

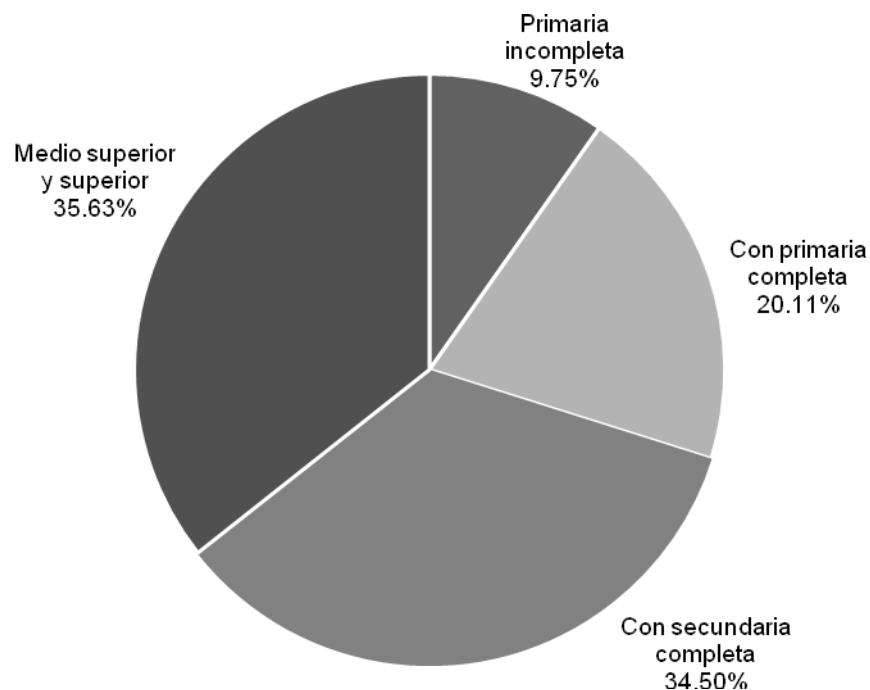
En efecto, la presión activa en los mercados de trabajo proviene de los sectores de población más dinámicos, pertenecientes a una cultura del trabajo asalariado y con cierto nivel de expectativas respecto a sus probabilidades de inserción en él, independientemente de que se cumplan o no dichas expectativas.

Una mirada al nivel de instrucción que muestran los desocupados abiertos indica algo al respecto. Es fácil observar aquí que los escolarizados predominan entre los buscadores activos de empleo, hay una relación ascendente entre escolaridad y presión en el mercado de trabajo.

Así, en diciembre de 2010, 35.63% de los desempleados contaban con educación media superior o superior, contra 20.11% que a lo sumo había cursado primaria

completa o 9.75% que no llegó a completar la primaria, como se observa en la figura 2.2.

Lo anterior refleja que los sectores en desempleo abierto, pese a padecer una situación adversa, guardan expectativas, por contar con el capital humano de la educación, y son más selectivos, es decir, no aceptan cualquier ocupación bajo cualquier condición sin haber empleado un periodo de búsqueda previamente.



**Figura 2.2** Porcentaje de los desocupados de acuerdo a su nivel de instrucción, en diciembre de 2010. (INEGI, 2012)

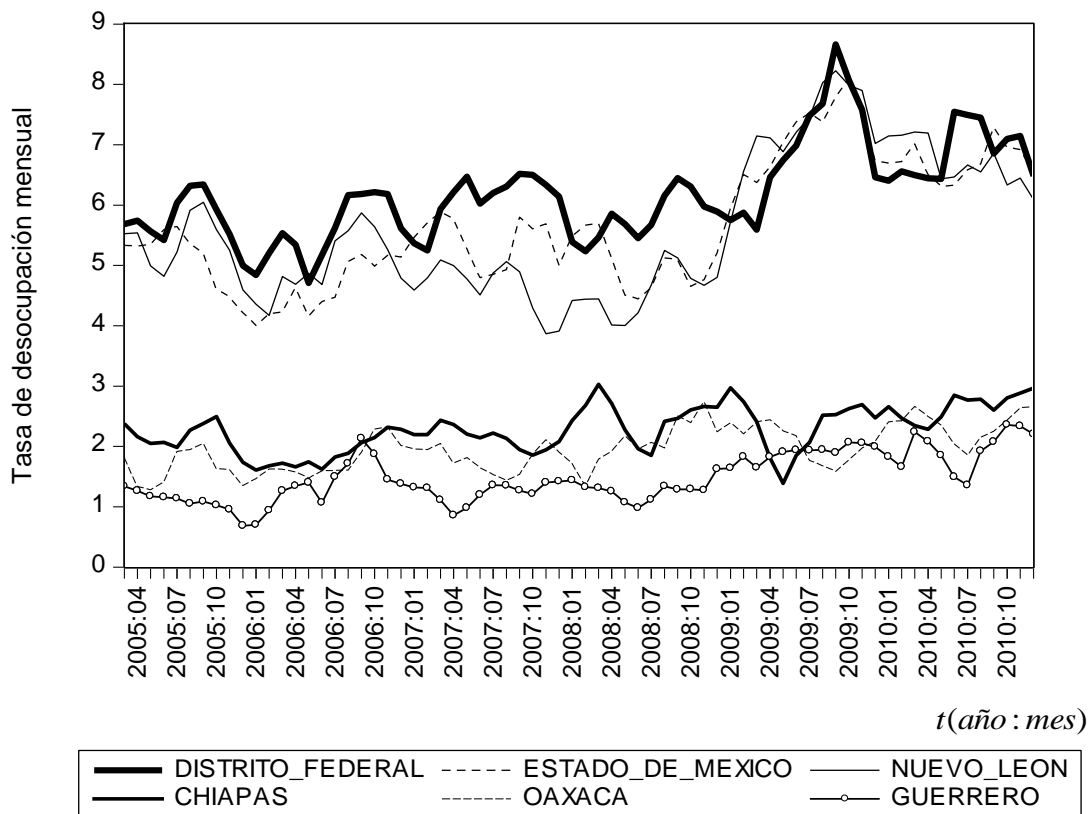
Dada una situación de carencia de trabajo, a mayor selectividad mayor tiempo de búsqueda y, con ello, mayor probabilidad de ser captado como desempleado abierto. El que un individuo esté dispuesto a asumir el costo de oportunidad de dejar de percibir, aunque sea ingresos marginales, quiere decir que para dicho individuo vale la pena esa inversión de tiempo completo en búsqueda y espera.

En contraste, un individuo no calificado no renuncia nunca a captar ingresos marginales, porque no tiene ni la expectativa ni la garantía de que obtendrá algo mejor a esos ingresos una vez que se presenta la oportunidad de obtenerlos, por carecer de

capital humano, ser menos selectivo y más maleable para asumir cualquier rol ocupacional.

### 2.5.2 Contraste entre las tasas de desocupación por regiones

La correlación entre desempleo abierto y expectativas no sólo da cierta forma social al desempleo, sino también una composición definida en términos regionales. En efecto, la cultura asalariada y las expectativas que genera explican que el desempleo abierto sea más alto en áreas urbanas que en rurales.

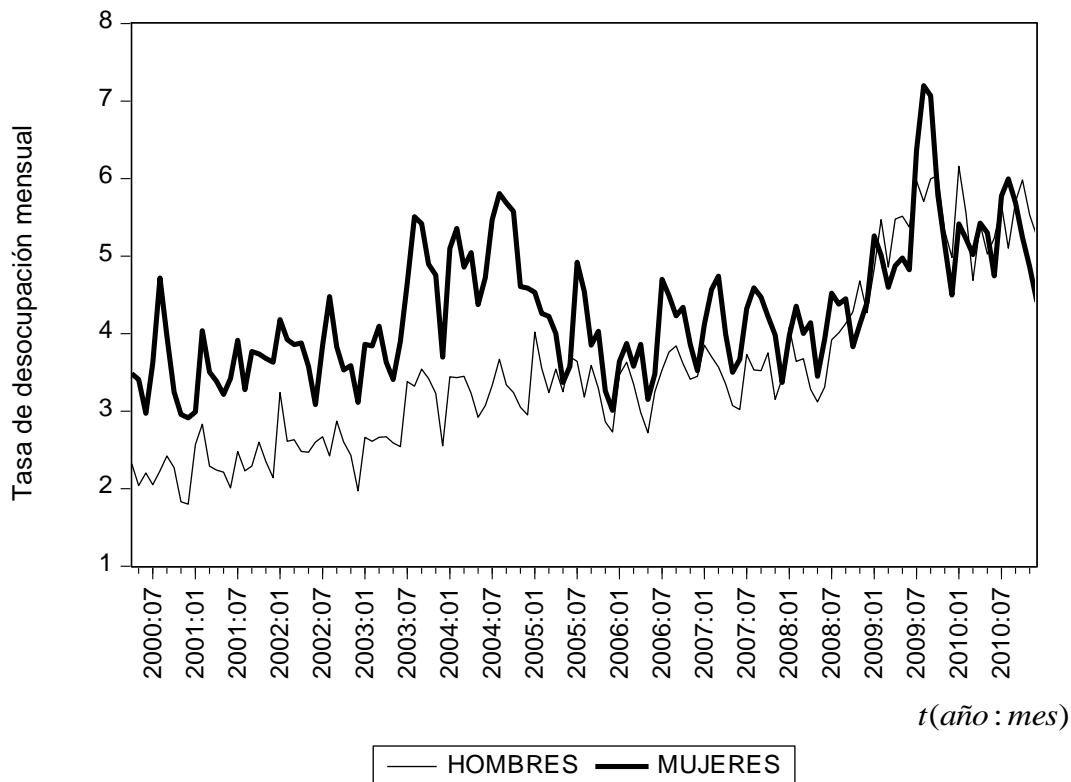


**Figura 2.3** Tasas de desocupación (mensual) para algunos estados de la República Mexicana, de marzo de 2005 a diciembre de 2010. (INEGI, 2012)

Así pues, la tasa de desocupación abierta, y en general cualquier otro indicador construido alrededor de la idea de presión en el mercado de trabajo, son más sensibles en el ámbito urbano y en aquellas entidades federativas (como Distrito Federal, Estado de México y Nuevo León) más estrechamente vinculadas al proceso de modernización

y los desequilibrios que le acompañan. En contraste, indicadores que sólo miran la situación y no el comportamiento de la población de estudio, resultan particularmente pertinentes para aquellas zonas del país donde el proceso de urbanización ha avanzado menos (como en los estados de Chiapas, Oaxaca y Guerrero) y la inserción por la vía asalariada a la actividad económica no está en el centro de las expectativas de los grupos de población. Esto se refleja en la figura 2.3. En ella se observa como las poblaciones de los estados de Nuevo León, Estado de México o el Distrito Federal, presentan mayores tasas de desocupación abierta que las observadas para el caso de Chiapas, Guerrero u Oaxaca.

### 2.5.3 Contraste entre las tasas de desocupación por sexo

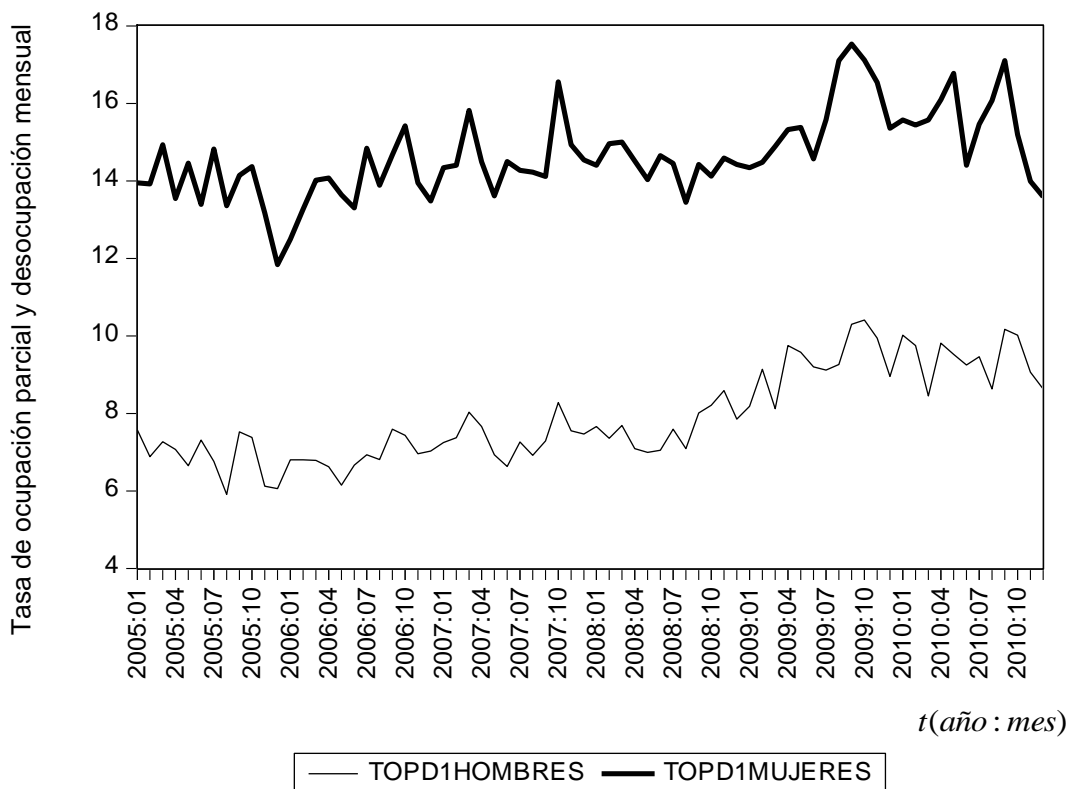


**Figura 2.4** Tasa de desocupación nacional (mensual) para hombres y para mujeres de abril de 2000 a diciembre de 2010. (INEGI, 2012)

Otro rasgo interesante de revisar del desempleo abierto en México es el que se refleja para los hombres y las mujeres, pues a lo largo del tiempo el nivel de desempleo entre las mujeres ha sido mayor con respecto al de los hombres. En la figura 2.4 puede

observarse tal comportamiento, aunque cabe resaltar que esta diferencia ha disminuido con el tiempo, incluso en los últimos meses del año 2010 la tasa de desocupación fue menor para las mujeres.

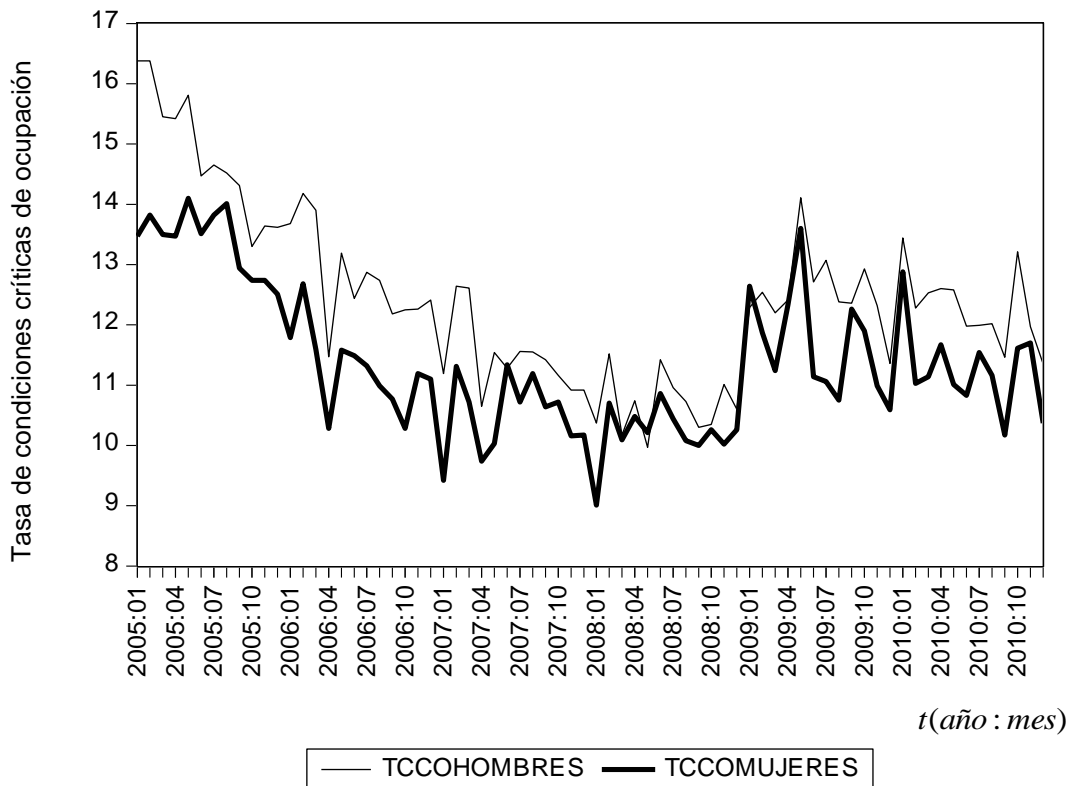
Aun cuando la diferencia entre las tasas de desocupación es cada vez menor, sigue existiendo una mayor diferencia en las tasas complementarias de desocupación, como se observa en la figura 2.5 de la TOPD1, que abarca tanto la ocupación parcial como la desocupación. La diferencia entre la TOPD1 obtenida por las mujeres y por los hombres es más acentuada que la correspondiente a la tasa de desocupación, lo que indica que muchas mujeres consiguen trabajos de ocupación parcial, debido seguramente a las actividades que tienen que desempeñar además en el hogar. Esta tasa fue de 8.63% para los hombres en diciembre de 2010, mientras que para la mujeres fue del 13.59%.



**Figura 2.5** Tasa de ocupación parcial y desocupación TOPD1 (mensual) para hombres y para mujeres de enero de 2005 a diciembre de 2010. (INEGI, 2012)

Otra tasa complementaria es la tasa de condiciones críticas de ocupación TCCO, indicador de condiciones inadecuadas de empleo desde el punto de vista del tiempo de

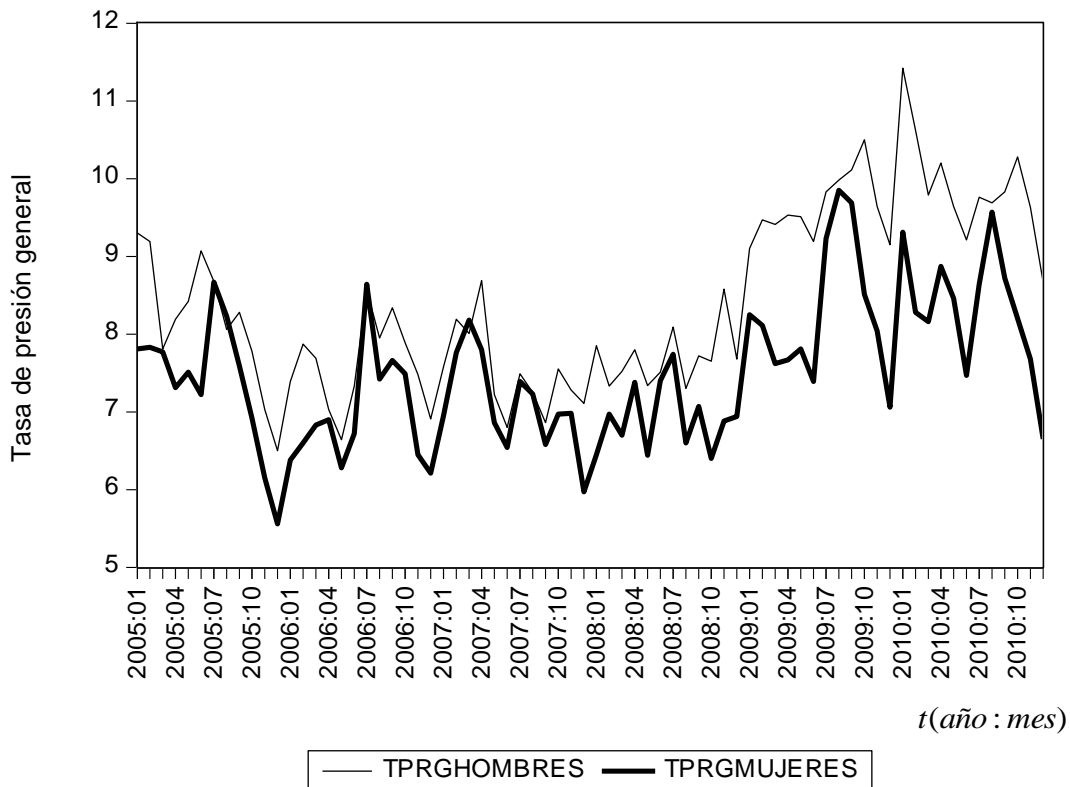
trabajo, los ingresos o una combinación insatisfactoria de ambos, que resulta particularmente sensible en las áreas rurales del país. Este indicador incluye a las personas que se encuentran trabajando menos de 35 horas a la semana por razones ajenas a sus decisiones, más las que trabajan más de 35 horas semanales con ingresos mensuales inferiores al salario mínimo y las que laboran más de 48 horas semanales ganando hasta dos salarios mínimos. En la figura 2.6 se puede observar que no hay diferencias muy marcadas entre la TCCO de hombres y de mujeres, siendo de 11.35% para los hombres y de 10.37% para las mujeres en diciembre de 2010, porcentajes sobre la población ocupada de hombres y mujeres respectivamente.



**Figura 2.6** Tasa de condiciones críticas de ocupación TCCO (mensual) para hombres y para mujeres de enero de 2005 a diciembre de 2010. (INEGI, 2012)

La tasa de presión general, otra tasa complementaria, incluye además de los desocupados, a los ocupados que buscan empleo, por lo que da una medida global de la competencia por plazas de trabajo conformada no sólo por los que quieren trabajar, sino por los que teniendo un empleo quieren cambiarlo o también los que buscan otro más para tener un segundo trabajo. Para los hombres esta tasa disminuyó de 9.15% de

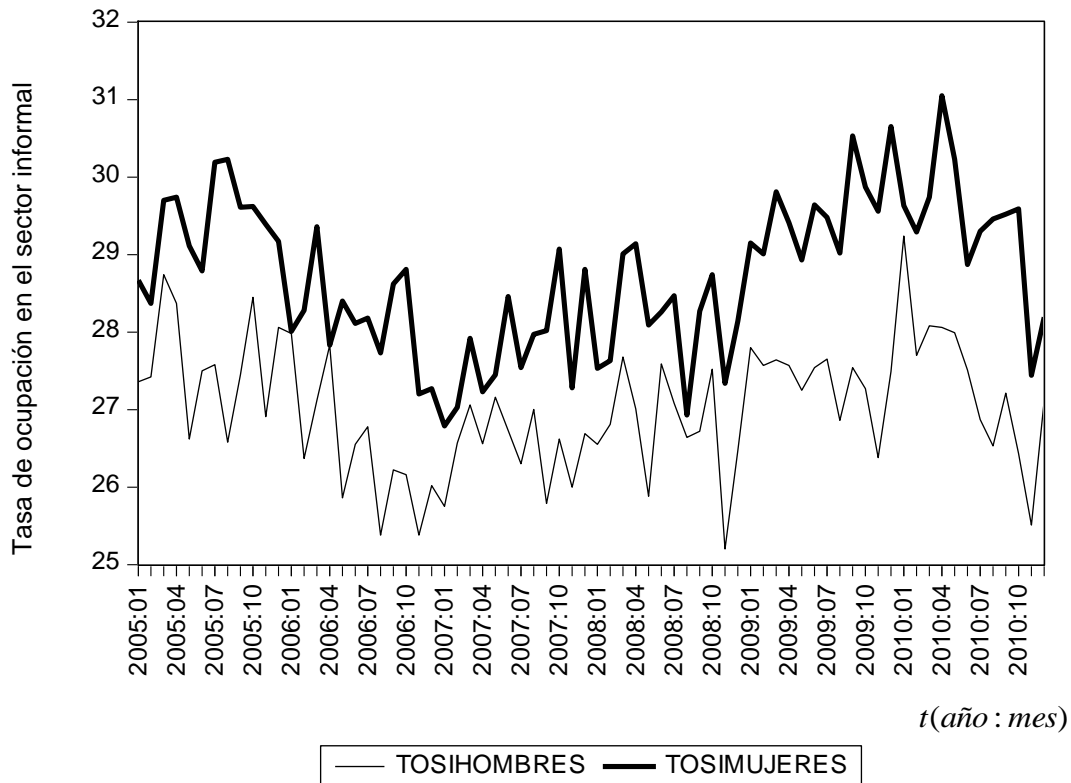
la PEA en el último mes de 2009 a 8.67% en el mismo mes de 2010, mientras que para las mujeres disminuyó de 7.06% a 6.65%, para los mismos meses, de acuerdo a la figura 2.7.



**Figura 2.7** Tasa de presión general TPRG (mensual) para hombres y para mujeres de enero de 2005 a diciembre de 2010. (INEGI, 2012)

En la figura 2.8 se presenta la tasa de ocupación en el sector informal, la cual se refiere a todas las personas que trabajan para unidades económicas no agropecuarias operadas sin registros contables y que funcionan a partir de los recursos del hogar o de la persona que encabeza la actividad sin que se constituya como empresa, de modo que la actividad en cuestión no tiene una situación identificable e independiente de ese hogar o de la persona que la dirige y que por lo mismo tiende a concretarse en una muy pequeña escala de operación. Dicha tasa representó 27.11% de la población ocupada de hombres en el mes de diciembre de 2010, mientras que para las mujeres fue de 28.19% de la población ocupada de mujeres para el mismo mes, es decir, más de la cuarta parte de la población ocupada de los hombres y de las mujeres seguramente se auto empleó o consiguió un empleo en el sector informal; en específico para la mujeres,

podría ser un indicativo de que buscan trabajos que les permitan desempeñar las actividades propias en el hogar.



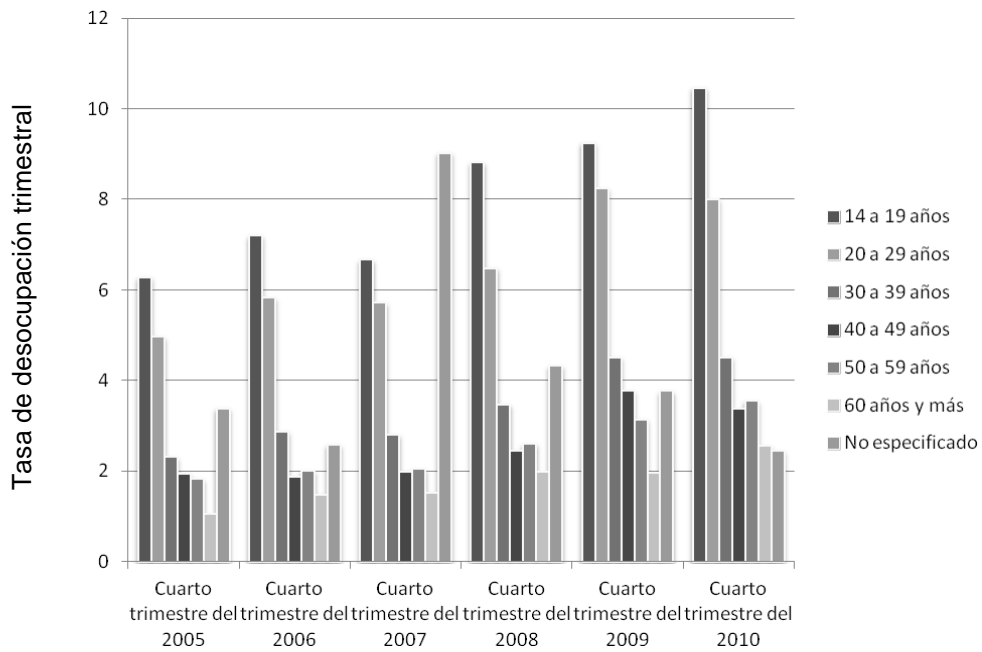
**Figura 2.8** Tasa de ocupación en el sector informal TOSI (mensual) para hombres y para mujeres de enero de 2005 a diciembre de 2010. (INEGI, 2012)

### 2.5.4 Contraste entre las tasas de desocupación por edad

De manera común las tasas más elevadas las encontramos entre los jóvenes que en el resto de la población. Así puede observarse en la figura 2.9, que las tasas de desocupación abierta para los grupos de 14 a 19 años y 20 a 29 años representan tres o más veces las tasas correspondientes a los grupos de 40 a 49 años, 50 a 59 años o de 60 años y más. El alto desempleo juvenil puede deberse a que estos grupos de la población se encuentran en un “estado de transición”, es decir, moviéndose de la escuela al mundo laboral y de la residencia en el seno familiar a su emancipación, es decir, a una independencia tanto en el plano económico y residencial, como en el de la toma de sus decisiones, razón que lo caracteriza como un grupo con frecuentes entradas y salidas del mercado de trabajo; se apunta también que su entrada en el



mercado laboral depende, entre otros factores, de la composición del empleo por sector, de los requerimientos mínimos de especialización para la generación y ocupación de nuevos puestos de trabajo, así como de la capacidad que tienen los empleadores para introducir y/o reemplazar trabajadores jóvenes con elementos sustitutos que intervienen en los procesos productivos.



**Figura 2.9** Tasa de desocupación por grupos de edad, para los cuartos trimestres de los años 2005 al 2010. (INEGI, 2012)

De acuerdo a todo lo observado anteriormente, puede deducirse que el desempleo abierto en México es un comportamiento correlacionado con los sectores modernos del país cuando éstos enfrentan un déficit de oportunidades.

Vivimos en un país donde la mayoría desea que las condiciones de vida mejoren, pero la situación no es clara, especialmente en lo que respecta a una recuperación sostenida de los salarios.

Por eso es importante estar conscientes de la necesidad de contar con información oportuna sobre diversos fenómenos, para así darnos cuenta de las diferentes situaciones que viven las personas en nuestro país, insistir en el mejoramiento y la diversificación de indicadores sobre aspectos claves como el comportamiento del mercado laboral.

# 3

## ANÁLISIS DE SERIES DE TIEMPO

---

### 3.1 Introducción

El análisis de series de tiempo es una metodología que tiene una fundamentación matemática ligada con el análisis estadístico y con los procesos estadísticos.

En la actualidad, el análisis de las series de tiempo juega un papel muy importante en el estudio económico, pues a través de éste se puede obtener información de variables a través del tiempo reduciendo así la incertidumbre de su comportamiento en el futuro y con esto se facilita la toma de decisiones.

### 3.2 Definición y objetivos del análisis de las series de tiempo

Una serie de tiempo es una sucesión de observaciones de cierto fenómeno o experimento (variable) en distintos momentos del tiempo y a intervalos iguales, que también se conoce como serie cronológica o histórica. Así pues, los datos de las series de tiempo recogen la evolución de la variable a lo largo del tiempo. (Hernández y Herrador, 2000)

Un ejemplo de serie de tiempo es la tasa de desocupación abierta de México.

Matemáticamente, una serie de tiempo se define por los valores  $Y_1, Y_2, \dots, Y_T$ , de una variable  $Y$  (por ejemplo, la tasa de desocupación abierta) en los momentos  $t_1, t_2, \dots, t_T$ . Así  $Y$  es una función de  $t$  y se simboliza por  $Y = F(t)$ .

El conjunto de observaciones históricas de la variable  $Y$  son obtenidas a intervalos regulares de tiempo y constituyen la serie temporal de datos  $Y_t$  con  $t = 1, 2, \dots, T$ .

El análisis de las series de tiempo trata de construir un modelo para explicar la estructura y prever la evolución de una variable que se observa a lo largo del tiempo. Una propiedad deseable en las series de tiempo es que sean estocásticas, es decir, que su comportamiento futuro sea determinado, en parte, por su pasado. Este supuesto se mantendrá a lo largo del presente trabajo.

A largo plazo se apreciará, por lo general, una evolución en las series de tiempo que puede ser creciente o decreciente, sin embargo hay series que presentan variaciones que no muestran un claro crecimiento o decrecimiento a largo plazo. Un estudio gráfico de los datos de una serie de tiempo proporciona bastante información de su evolución a corto y largo plazo. De hecho, el primer paso en el análisis de las series de tiempo es graficar los datos para obtener medidas descriptivas que permitan conocer sus principales propiedades.

El estudio de las series de tiempo tiene dos objetivos:

- a) Conocer el comportamiento pasado de la variable y efectuar análisis comparativos de unas series con otras, de lugares o tiempo diferentes.
- b) Predecir el valor de la variable en el futuro, es decir, a partir del conocimiento de los valores reales del pasado determinar su comportamiento futuro. Cabe destacar que dicho comportamiento futuro se considera como una clase de extensión del pasado, pero que difícilmente será exactamente igual.

### 3.3 Procesos estocásticos y series de tiempo

Un proceso estocástico es una sucesión de variables aleatorias, en correspondencia con una sucesión de valores de la variable de tiempo, y se denota como:  $\{Z(t); t \in T\}$ , en donde  $T$  es el conjunto índice de números reales que representan a la variable de tiempo y  $Z(t)$  es la variable aleatoria que corresponde al elemento  $t$  de  $T$ . Se dice que el proceso estocástico es continuo cuando  $T$  es un intervalo de números reales (cerrado o abierto), y se dice que es discreto cuando  $T$  es un conjunto finito o infinito numerable. (Guerrero, 2003)

De acuerdo con lo anterior, una serie de tiempo es la sucesión de observaciones de un proceso estocástico, cuyo conjunto índice está en relación al tiempo. Otra forma de interpretarlo es que una serie de tiempo está generada por una serie de choques aleatorios independientes  $\{\varepsilon_t\}$ . Con base en el supuesto de que dichos choques aleatorios son realizaciones independientes de una variable aleatoria con media constante (generalmente cero) y varianza  $\sigma^2$  constante, a esta sucesión de variables aleatorias se le conoce como proceso de *ruido blanco*, y un ejemplo de ello es la distribución normal con media constante  $\mu$  y varianza  $\sigma^2$ . (Guerrero, 2003)

### 3.4 Método de descomposición de series de tiempo

En el método de descomposición la serie se separa de forma horizontal, es decir, en varias componentes que llevan cada una información sobre la tendencia, el factor cíclico, la estacionalidad o la componente irregular. Para llevar a cabo la separación, se parte del supuesto de que sus componentes tienen alguna relación, ya sea aditiva, multiplicativa o mixta, las cuales se explican más adelante. (Serret, 1998)

Las series económicas son series que generalmente contienen componentes de manera aleatoria y determinista. Si se define por  $D_t$  el componente determinista y  $A_t$  el componente aleatorio, una serie temporal  $Y_t$  podría expresarse de la siguiente forma:

$$Y_t = D_t + A_t$$

En la expresión anterior, los dos componentes están relacionados de forma aditiva, pero también pueden concebirse otras formas distintas en la integración de ambos componentes.

El método de descomposición tiene las siguientes características:

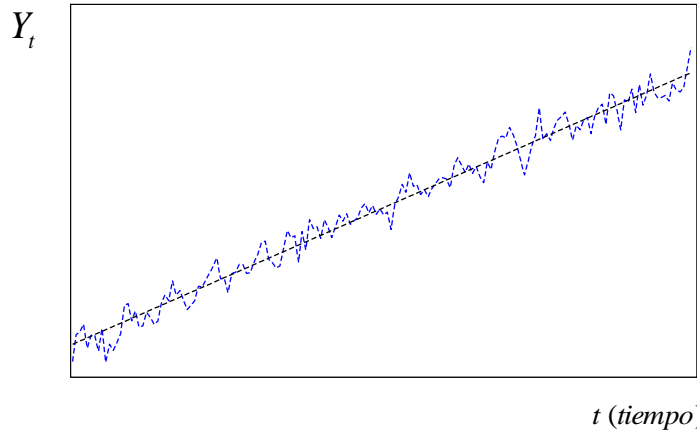
- Requiere disponer de la información estadística del pasado.
- Su misión es utilizar toda la información disponible para realizar análisis y/o predicciones, según las necesidades del estudio.
- Es de igual importancia tanto la cantidad como la calidad de la información de los datos.
- Realiza estudios de análisis y predicción de una variable, utilizando como información la contenida en los valores de la serie que mide la evolución de la variable objeto del estudio.

### 3.4.1. Componentes de las series de tiempo

Una serie de tiempo puede separarse en todos o algunos de los siguientes componentes: (Serret, 1998)

- a) **Tendencia secular** ( $T_t$ ). Se refiere a la dirección general a la que una serie de tiempo parece dirigirse en un intervalo grande de tiempo. Puede considerarse como la medida del cambio promedio de la serie por unidad de tiempo que indica la tendencia creciente, estacionaria o decreciente de los datos. En el futuro se hará referencia a ella únicamente como tendencia.

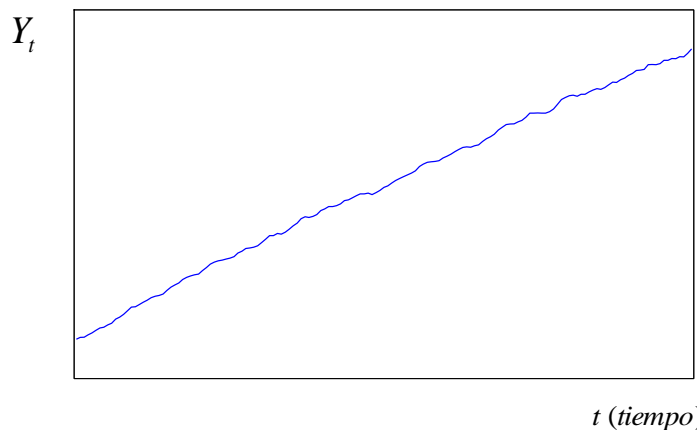
Las fuerzas básicas que producen o afectan la tendencia de una serie son: cambios en la población, inflación, cambios en los precios, cambios en el gusto del consumidor, cambios tecnológicos e incrementos en la productividad.



**Figura 3.1** Ejemplo de serie de datos ( $Y_t$ ) con componente de tendencia.

b) **Factor cíclico ( $C_t$ ).** Se refiere a las oscilaciones de larga duración alrededor de la curva de tendencia, las cuales pueden ser o no periódicas; es decir, pueden seguir o no caminos análogos en intervalos de tiempos iguales. Estas oscilaciones se dan en forma de onda o ciclos de más de un año de duración, producidas por cambios en las condiciones económicas. La duración de un ciclo se mide desde un pico hasta el siguiente pico y de un valle hasta el siguiente. En general los movimientos cíclicos se caracterizan por tener lapsos de expansión y contracción.

Como en general resulta difícil separar la tendencia y el factor cíclico se adopta la solución de englobar ambos componentes en uno solo, denominado ciclo-tendencia o tendencia generalizada.

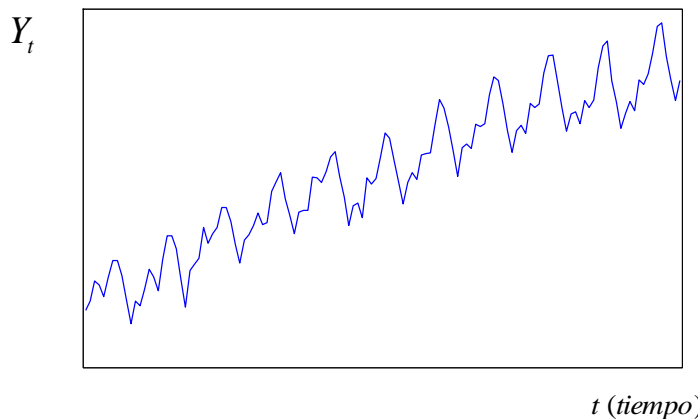


**Figura 3.2** Ejemplo de serie de datos ( $Y_t$ ) con componente de ciclicidad.

c) **Estacionalidad ( $E_t$ )**. Se refiere a las fluctuaciones periódicas que se observan en series de tiempo cuya frecuencia es menor a un año (trimestral, mensual, diaria, etc.), aproximadamente en las mismas fechas y casi con la misma intensidad. En general estas fluctuaciones se presentan de manera recurrente a través del tiempo. El componente estacional es de una importancia muy notable, debido a que el tiempo juega un papel que tiene mucho más sentido que en los otros componentes. El tiempo actúa en este caso de dos formas distintas:

1. Directamente, al afectar el clima propio de la estación a una serie de fenómenos económicos.
2. Indirectamente, al servir el calendario anual de marco a una serie de hechos y costumbres institucionales (vacaciones, fiestas, etc.).

La diferencia entre estas dos formas de incidencia del tiempo en la estacionalidad radica en que en la primera la intervención humana es mínima, en tanto que en la segunda las costumbres institucionales han sido establecidas por intervención humana, aunque en muchas ocasiones se ha tenido en cuenta para ello el clima de la zona.

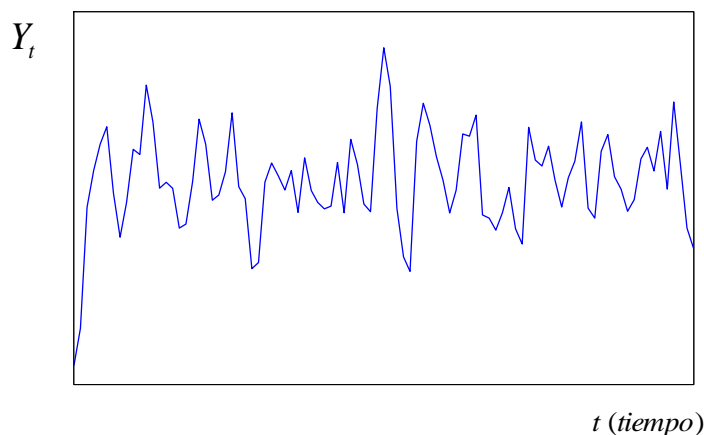


**Figura 3.3** Ejemplo de serie de datos ( $Y_t$ ) con componente de estacionalidad.

Por lo tanto, la estacionalidad de un fenómeno no debe ser difícil de predecir cuando ésta sea bastante fuerte a lo largo del tiempo.

Independientemente de que su predicción sea relativamente sencilla, es muy importante el conocimiento de este componente por la gran trascendencia que tiene desde el punto de vista económico.

- d) **Movimientos irregulares ( $I_t$ )**. Se refiere a movimientos esporádicos de corto plazo de las series de tiempo debido a sucesos que se producen de manera ocasional o imprevisible, tales como: elecciones, huelgas, clima poco usual, guerras, cambio de leyes, inundaciones, etc. que no están recogidas en la tendencia, en el factor cíclico o en la estacionalidad. Si bien pueden ser generados estos movimientos irregulares por factores de tipo económico, generalmente sus efectos producen variaciones que sólo duran un corto intervalo de tiempo, aunque debe reconocerse que en ocasiones sus efectos sobre el comportamiento de una serie pueden ser tan intensos que fácilmente podrían dar lugar a un nuevo ciclo o a otros movimientos. Así pues, el movimiento irregular tiene un carácter residual.



**Figura 3.4** Ejemplo de serie de datos ( $Y_t$ ) con componente de aleatoriedad.

El componente residual o término de error se puede descomponer en dos partes:

- Aleatoria, la cual recoge los pequeños efectos accidentales.
- Errática, que es consecuencia de hechos no siempre previsible, pero que pueden ser identificados a posteriori.



Si no se hace una mención expresa se considera que el componente residual está sólo formado por la parte aleatoria.

Con objeto de que todo el comportamiento sistemático de la serie esté recogido en los otros componentes, se suele pensar que la parte aleatoria, a la que se define por  $\varepsilon_t$ , no está relacionada con los valores anteriores o posteriores de la misma. Además se suele suponer también que su media teórica es nula, que su varianza teórica se mantiene constante para toda  $t$  y que su distribución es normal. A una variable que cumpla las anteriores hipótesis, como ya se había mencionado, comúnmente se le denomina variable de ruido blanco. Analíticamente se puede expresar de la siguiente forma: (Guerrero 2003)

Propiedades estadísticas de una variable de ruido blanco

$$E(\varepsilon_t) = 0 \quad \forall t$$

$$E(\varepsilon_t \varepsilon_{t'}) = 0 \quad \text{para } t \neq t'$$

$$E(\varepsilon_t^2) = \sigma^2 \quad \forall t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

Por sus características, una variable de ruido blanco es completamente impredecible.

El análisis de aquellos puntos en los cuales el término de error tiene un peso muy fuerte, permite corregir la serie original con el objeto de llegar a un conocimiento más preciso de los componentes restantes.

A continuación se examina la forma en que los componentes anteriores se pueden integrar para dar lugar a una serie de tiempo.

### 3.4.2. Integración de los componentes de una serie de tiempo

El valor que toma la serie de tiempo  $Y$  en cada momento del tiempo es una función de los componentes anteriormente mencionados. Entonces  $Y_t$  es el valor que toma la

variable  $Y$  en el momento  $t$  y está dado por la siguiente expresión:  
 $Y_t = f(T_t, C_t, E_t, I_t)$ , la cual puede seguir diferentes esquemas. (Serret, 1998)

### Esquema multiplicativo

Es el esquema que representa al valor de la serie de tiempo en cualquier instante como el producto de los valores correspondientes a los cuatro componentes, este se expresa como  $Y_t = T_t \times C_t \times E_t \times I_t$ . Es el que mejor describe fenómenos económicos, debido a que supone que los componentes están interrelacionados, es decir, conforme la tendencia aumenta (o disminuye), los efectos cíclicos se van amplificando (o reduciendo) conforme evoluciona la serie. En este modelo solo la tendencia se mide en las mismas unidades de los datos, mientras que los demás componentes se tratan como índices o porcentajes de aquella.

### Esquema aditivo

En este esquema el valor de la serie de tiempo para cualquier instante es igual a la suma de los valores correspondientes a los cuatro componentes, es decir,  $Y_t = T_t + C_t + E_t + I_t$ . En este modelo todos sus componentes se tratan como diferencias y se expresan en las unidades originales de los valores de los datos. Este esquema se utiliza cuando se supone que los componentes son independientes, por lo tanto la variación permanece constante dentro de todos los periodos base sin tener en cuenta la tendencia y de forma análoga las otras componentes, o sea, las oscilaciones alrededor de la tendencia siempre son parecidas.

### Esquema mixto

El esquema mixto le da el valor a la serie de tiempo multiplicando la tendencia, factor cíclico y estacionalidad, y sumando el movimiento irregular, esto puede expresarse como  $Y_t = T_t \times C_t \times E_t + I_t$ . Al igual que en el esquema multiplicativo, las componentes sistemáticas actúan como factores influyendo el nivel alcanzado por una sobre el alcance de las otras.

Es importante mencionar que en una serie concreta, no es necesario que todos los componentes estén presentes, de hecho, una serie con periodicidad anual carece de estacionalidad.

### 3.4.3. Técnicas descriptivas simples

El primer paso del análisis de las series de tiempo es graficar los datos, ya que esto permite detectar patrones tales como tendencias, ciclos, estacionalidad y movimientos irregulares; así como heterocedasticidad (varianza no constante), discontinuidades y discrepancias.

Por medio de las gráficas se puede descubrir si es recomendable realizar alguna transformación a los valores de la variable observada, con la finalidad de contar con una variable puramente estocástica, no estacional, con media y varianza constantes. Por lo tanto, existen tres principales razones para aplicar las transformaciones: estabilizar la varianza, hacer aditivo el efecto estacional y hacer que los datos tengan distribución normal. (Guerrero, 2003)

- **Estabilización de la varianza**

Para entender correctamente lo que se explica a continuación, es necesario conocer el término estacionariedad, ya que se utilizará con frecuencia.

Estacionariedad. Se dice que una serie de tiempo es estacionaria si su media y su varianza permanecen constantes y además las variaciones estrictamente periódicas han sido removidas.

La falta de estacionariedad puede ser descubierta al estudiar la gráfica de la serie de tiempo, en la cual se observará gran dependencia entre la variabilidad y el nivel de la serie.

Existen varias transformaciones para obtener una varianza constante y con ello una serie estacionaria. A continuación se enuncian algunas de ellas:

### Transformaciones logarítmicas y de raíces cuadradas

La transformación logarítmica consiste en trabajar con los logaritmos de la variable observada en lugar de los valores originales, mientras que en la transformación de raíces cuadradas se utilizan las raíces cuadradas de los valores de la variable en estudio.

Estas transformaciones resultan útiles cuando se presenta lo siguiente:

- a. La varianza es proporcional al crecimiento de la serie.
- b. El nivel medio de la serie se incrementa o decrece a una tasa constante, generalmente este supuesto es válido a corto y mediano plazo, ya que a largo plazo se pierde la confiabilidad del patrón de datos (si es que existe), puede cuestionarse el crecimiento con tasa constante.

### Transformaciones de Box Cox

Es también conocida como una transformación “normalizante”. Ésta está dada por:

$$Y^\lambda = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0, Y \geq 0 \\ \ln Y & \text{si } \lambda = 0, Y = 0 \end{cases}$$

En la práctica, cuando es necesario aplicar una transformación Box-Cox, en general, resulta adecuado un valor de  $\lambda = 0$  ó  $\lambda = \frac{1}{2}$ .

- **Hacer aditivo el efecto estacional**

Se sugiere aplicar este tipo de transformación cuando existe tendencia en la serie de tiempo y el efecto estacional aumenta con la media, dicho efecto puede ser de dos tipos: aditivo o multiplicativo. Cuando su tamaño es independiente de los demás factores se dice que el efecto es aditivo y multiplicativo cuando su tamaño es directamente proporcional a la media. Generalmente se utilizan los siguientes modelos, donde  $Y_t$  = observación al tiempo  $t$ ,  $m_t$  = media de las primeras  $t$  observaciones,  $E_t$  = efecto estacional y  $\varepsilon_t$  = error aleatorio:

a) Para el modelo completamente aditivo,  $Y_t = m_t + E_t + \varepsilon_t$ , ya no es necesario hacerle transformación alguna.

b)  $Y_t = m_t E_t + \varepsilon_t$

En este modelo existe un efecto estacional multiplicativo y error aditivo.

c)  $Y_t = m_t E_t \varepsilon_t$

Este modelo es totalmente multiplicativo y es recomendable aplicarle la transformación logarítmica. El tamaño de esos efectos determinan si se requiere o no una transformación.

- **Hacer que los datos se distribuyan de forma normal**

Al realizar estimaciones o llevar a cabo procedimientos de prueba es muy común suponer que los datos se distribuyen de forma normal. Utilizando la gráfica de probabilidad normal se puede constatar si la serie de observaciones independientes es normal o no, así como utilizando los coeficientes de asimetría y curtosis.

Se sabe que para verificar, gráficamente, que una variable aleatoria se distribuye normal, se debe tener el 95% de sus observaciones dentro del intervalo  $(-2\sigma, 2\sigma)$ . Otra alternativa es llevando a cabo un histograma para revisar cómo es la distribución y de esta forma descubrir si existe o no asimetría. En el caso de que la variable no se distribuya normal, se puede utilizar una transformación “normalizante” como la de Box-Cox.

### 3.5 Funciones de autocovarianza y autocorrelación

En cualquier análisis de series de tiempo, resulta necesario el conocimiento de la estructura particular de dependencia existente entre las variables del proceso, para poder concretar el modelo explicativo de la evolución histórica de la serie. Para ello se toma como instrumento estadístico de análisis de esta dependencia temporal entre variables, la función de autocovarianza, o bien, su expresión derivada, la función de

autocorrelación. Esta última generaliza la importante noción clásica de correlación al caso específico de las series de tiempo. (Gujarati, 2003)

A partir de esta sección, se utilizan los términos de proceso y serie indistintamente.

El término de autocorrelación se define como “la correlación entre los miembros de series de observaciones ordenadas en el tiempo”. (Kendall, 1971)

Es importante señalar que el modelo de regresión lineal supone que no existe dicha autocorrelación en las perturbaciones o errores aleatorios  $\varepsilon_t$ , es decir (Gujarati, 2003):

$$E(\varepsilon_t \varepsilon_{t'}) = 0 \quad t \neq t'$$

El modelo clásico supone que la perturbación de una observación cualquiera no está influenciada por la perturbación relacionada con cualquier otra observación.

Sin embargo, si existe dependencia se tiene autocorrelación, es decir:

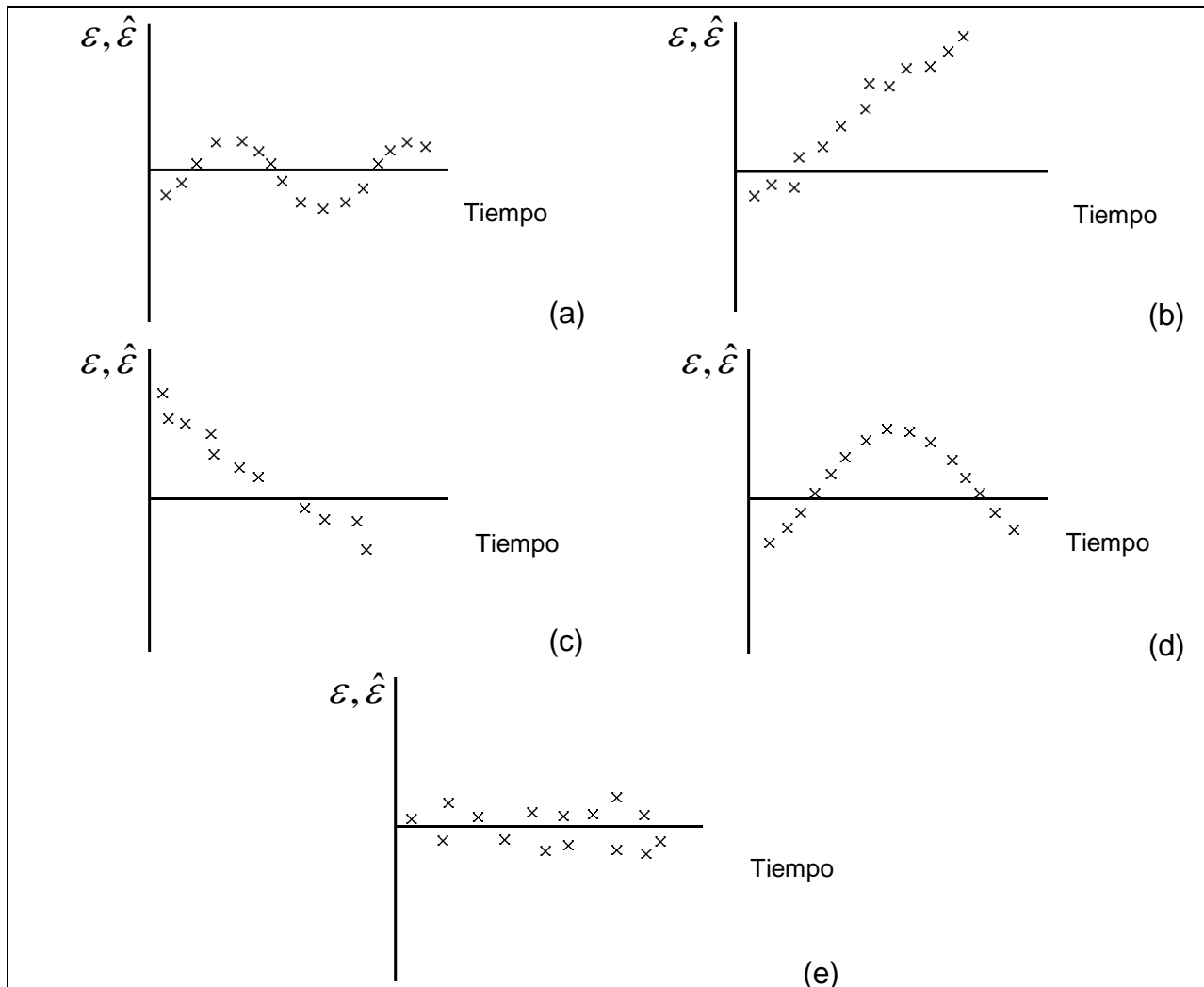
$$E(\varepsilon_t \varepsilon_{t'}) \neq 0 \quad t \neq t'$$

Existen algunos patrones razonables de autocorrelación y de no autocorrelación, donde  $\varepsilon$  es el error aleatorio y  $\hat{\varepsilon}$  es su estimación.

En la figura 3.5 se muestran los patrones de autocorrelación y no autocorrelación. En el inciso (a) se muestra un patrón cíclico, en el inciso (b) y (c) una tendencia lineal hacia arriba o hacia abajo en las perturbaciones, el inciso (d) muestra una tendencia cuadrática en las perturbaciones y en el inciso (e) no existe patrón sistemático apoyando el supuesto de no autocorrelación del modelo clásico de regresión lineal.

Existen diferentes razones por las cuales ocurre la autocorrelación, entre ellas están:

- Inercia. Es bien sabido que la mayoría de las series de tiempo económicas presentan ciclos, es decir, empezando en el fondo de la recesión, cuando se inicia la recuperación económica, las series comienzan a moverse hacia arriba, en este movimiento ascendente existe un impulso en la serie que continúa hasta que sucede algo que la hace descender nuevamente. Por consiguiente, es muy probable que las observaciones sucesivas sean interdependientes.



**Figura 3.5** Patrones de autocorrelación y de no autocorrelación. (Gujarati, 2003)

- Sesgo de especificación por usar variables excluidas. Este problema se presenta cuando al analizar un modelo donde se excluyeron variables necesarias, se observa un patrón de autocorrelación, el cual podría ser eliminado al incluir las variables debidas.
- Sesgo de especificación por usar una forma funcional incorrecta. Se presenta cuando no se considera el modelo correcto, y al analizarlo refleja una correlación.
- Fenómeno de la telaraña. Se da cuando la variable explicada  $Y$  reacciona ante cambios en las variables independientes  $X$ 's con un rezago de un periodo de tiempo, debido a que la implementación de las decisiones toman tiempo.

- Rezagos. Se producen cuando en un modelo, la variable dependiente está en función de sí misma pero con ciertos retrasos en el tiempo, entonces el término de perturbación (error) mostrará un patrón sistemático, debido a la influencia de la variable retrasada sobre sí misma.
- Manipulación de datos. Se presenta cuando se genera información para periodos mayores a través de los promedios, pues se suavizan los datos, eliminando así las fluctuaciones. Otra fuente de manipulación es la interpolación o extrapolación de la información.

Por otro lado, los efectos de la autocorrelación son:

- Se obtendrán estimadores insuficientes, por lo tanto los intervalos de confianza serán más anchos de lo necesario y la prueba de significancia menos fuerte.
- La varianza residual  $\hat{\sigma}^2$  tiende a subestimar al verdadero valor de  $\sigma^2$ ; incluso si no está subestimada esta varianza, la varianza de los estimadores pueden si subestimar a los verdaderos y como consecuencia de lo anterior, las pruebas usuales de significación t de Student y F de Fisher ya no son válidas, pues si se aplican tienden a conclusiones erróneas.
- Aunque los estimadores sean insesgados para una muestra en particular tienden a buscar una visión distorsionada de los verdaderos valores poblacionales. En otras palabras, los estimadores se vuelven sensibles a las fluctuaciones muestrales.

Para entender lo anteriormente expuesto, considérese una muestra aleatoria de tamaño  $T$ , de una población bivariante formada por las variables  $Y$  y  $X$ , con ella se pueden calcular los siguientes estadísticos:

- Las medias muestrales de  $Y$  ( $\bar{Y}$ ) y  $X$  ( $\bar{X}$ ) que se calculan como sigue:

$$\bar{Y} = \frac{\sum_{t=1}^T Y_t}{T} \qquad \bar{X} = \frac{\sum_{t=1}^T X_t}{T}$$



Las cuales son estimadores insesgados de las medias poblacionales, entonces si se definen las medias poblacionales de  $Y$  y de  $X$  como  $\mu_Y$  y  $\mu_X$ , respectivamente, se puede verificar que:

$$E(\bar{Y}) = \mu_Y \qquad E(\bar{X}) = \mu_X$$

- Las varianzas muestrales de  $Y$  ( $S_Y^2$ ) y de  $X$  ( $S_X^2$ ), que son momentos de segundo orden con respecto a la media, se calculan como sigue:

$$S_Y^2 = \frac{\sum_{t=1}^T (Y_t - \bar{Y})^2}{T} \qquad S_X^2 = \frac{\sum_{t=1}^T (X_t - \bar{X})^2}{T}$$

A las raíces cuadradas de las expresiones anteriores se les denominan desviaciones típicas de  $Y$  y  $X$ , respectivamente. En varias ocasiones, en lugar de la varianza muestral se utiliza la cuasivarianza muestral, la cual tiene la misma expresión que la varianza muestral, pero en lugar de dividirse entre  $T$  se divide entre  $T - 1$ , ya que se pierde un grado de libertad al tomar desviaciones con respecto a la media. Como se sabe, en las poblaciones infinitas, la cuasivarianza muestra un estimador insesgado de la varianza poblacional, debido precisamente a que se divide entre el número de grados de libertad.

Denotando a las varianzas poblacionales de  $Y$  y  $X$  por  $\sigma_Y^2$  y  $\sigma_X^2$  respectivamente, se puede verificar que:

$$E[S_Y^2] = \sigma_Y^2 \qquad E[S_X^2] = \sigma_X^2$$

Las varianzas y cuasivarianzas, y aún más las desviaciones y cuasidesviaciones típicas, sirven para medir la dispersión de las variables objeto de estudio.

- La covarianza muestral entre dos variables, en este caso entre  $X$  y  $Y$ , está dada por la siguiente expresión:

$$S_{X,Y} = \frac{\sum_{t=1}^T (Y_t - \bar{Y})(X_t - \bar{X})}{T}$$

La covarianza representa lo fuerte que pueden estar relacionadas dos variables. Así, cuanto mayor sea la covarianza entre  $X$  y  $Y$ , más fuerte será la relación entre ambas variables. Sin embargo, este estadístico tiene el inconveniente de que el valor que toma depende de las unidades de medida de las variables, es decir, es una medida de carácter absoluto. Por lo tanto, resulta difícil establecer lo fuerte que pueda ser una relación a partir del análisis de la covarianza.

Con el fin de evitar el inconveniente antes mencionado, se utiliza el coeficiente de correlación lineal muestral entre dos variables, para medir su grado de asociación. Este coeficiente es una medida de carácter relativo, cuyo valor está acotado entre -1 y 1. Donde el valor 1 indica que el grado de asociación entre dos variables es total y positivo, mientras que el valor -1 indica un grado de asociación total de carácter negativo. Por otra parte, el valor 0 refleja que no existe asociación entre las dos variables. A partir de las varianzas y de la correspondiente covarianza, se puede calcular de forma inmediata el coeficiente de correlación entre las variables  $X$  y  $Y$ , el cual se expresa como sigue:

$$r_{Y,X} = \frac{\sum_{t=1}^T (Y_t - \bar{Y})(X_t - \bar{X})}{\sqrt{\sum_{t=1}^T (Y_t - \bar{Y})^2 \sum_{t=1}^T (X_t - \bar{X})^2}}$$

Al estadístico muestral anterior, le corresponde el siguiente coeficiente de correlación lineal poblacional, que está dado por:

$$\begin{aligned} \rho_{YX} &= \frac{\sigma_{YX}}{\sigma_Y \sigma_X} \\ &= \frac{E[(Y - \mu_Y)(X - \mu_X)]}{\sqrt{E[(Y - \mu_Y)^2] E[(X - \mu_X)^2]}} \end{aligned}$$

El estadístico  $r_{Y,X}$  es un estimador consistente de  $\rho_{YX}$ , por lo que al aumentar el tamaño de la muestra  $r_{Y,X}$  converge al parámetro  $\rho_{YX}$ .

Para contar con este último concepto dentro del contexto de series de tiempo se define la variable  $X$  como  $X_t = Y_{t-1}$ , es decir, en la expresión anterior se ha definido la variable  $X$  como la variable  $Y$  desplazada un periodo.

Si la muestra es lo suficientemente grande, la media y la varianzas muestrales de  $Y$  serán prácticamente iguales a la media y a la varianzas muestrales desplazadas un periodo.

La covarianza de la variable  $Y_t$  y  $X_t$ , definida como  $Y_{t-1}$ , se le denota como  $g_1$  y está dada por:

$$S_{Y_t, X_{t-1}} = g_1 = \frac{\sum_{t=2}^T (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{T-1}$$

A la covarianza anterior se le denomina autocovarianza muestral de primer orden. La denominación de autocovarianza se debe a que es una covarianza entre una variable y la misma variable desplazada, y el calificativo de primer orden se debe al hecho de que se ha desplazado un periodo. Se utilizará la letra  $g$  para designar una autocovarianza, el subíndice de  $g_1$  indica que la autocovarianza es de primer orden.

La autocovarianza poblacional se debe definir en el contexto de los procesos estocásticos. Sin profundizar en la fundamentación estadística de los procesos estocásticos se señalará que en un proceso estocástico estacionario en donde la media y la varianzas se mantienen constantes a lo largo del tiempo, la autocovarianza de primer orden está definida por la siguiente expresión:

$$\gamma_1 = E[(Y_t - \mu_Y)(Y_{t-1} - \mu_Y)]$$

Análogamente, se define el coeficiente de correlación muestral de primer orden de la variable  $Y$  de la forma:

$$r_1 = \frac{\sum_{t=2}^T (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

$$= \hat{\rho}_1$$

Entonces, el coeficiente de autocorrelación teórico de primer orden está definido como sigue:

$$\rho_1 = \frac{\gamma_1}{\gamma_0}$$

$$= \frac{E[(Y_t - \mu_Y)(Y_{t-1} - \mu_Y)]}{E(Y_t - \mu_Y)^2}$$

De forma análoga a la autocovarianza muestral de primer orden se pueden definir las autocovarianzas muestrales de segundo orden, tercer orden, etc. Así, la autocovarianza muestral de orden  $h$  está dada por la siguiente expresión:

$$g_h = \frac{\sum_{t=h+1}^T [(Y_t - \bar{Y})(Y_{t-h} - \bar{Y})]}{T - h}$$

$$= \hat{\gamma}_h$$

Como se puede observar, cuanto mayor es el orden de la autocovarianza es menor el número de sumandos del numerador de la expresión anterior. Así, la autocovarianza de mayor orden que se puede calcular es  $g_{T-1}$ , que tendría un solo sumando. Obviamente entre menos términos intervengan en su cálculo, menor fiabilidad tendrá este estadístico del correspondiente parámetro poblacional.

Las autocovarianzas indican en qué medida están relacionados los valores de la variable con sus propios valores retardados en distintos periodos. El problema que presentan, es que, al igual que la covarianza son medidas de carácter absoluto. Razón por la cual es preferible utilizar los coeficientes de autocorrelación. Así pues, el coeficiente de autocorrelación muestral de orden  $h$  está dado por la siguiente expresión:

$$r_h = \frac{\sum_{t=h+1}^T [(Y_t - \bar{Y})(Y_{t-h} - \bar{Y})]}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

$$= \hat{\rho}_h$$

En un proceso estocástico estacionario la autocovarianza teórica de orden  $h$  está definida por la siguiente expresión:

$$\gamma_h = E [(Y_t - \mu_Y)(Y_{t-h} - \mu_Y)]$$

Asimismo, el coeficiente de autocorrelación teórico de orden  $h$  está dado por:

$$\rho_h = \frac{\gamma_h}{\gamma_0}$$

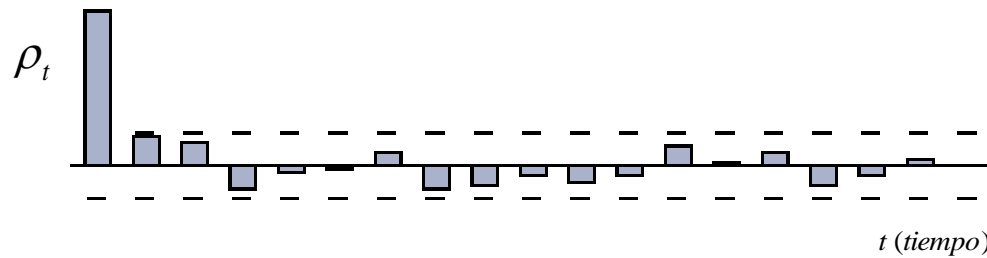
$$= \frac{E [(Y_t - \mu_Y)(Y_{t-h} - \mu_Y)]}{E (Y_t - \mu_Y)^2}$$

Como se ha mencionado los coeficientes de correlación muestral se obtienen a partir de una serie de tiempo. Por otra parte, los coeficientes de correlación teórica se deducen de un modelo. El modelo teórico es el siguiente:

$$Y_t = \varepsilon_t$$

Así, en el modelo anterior la variable  $Y$  es igual a una variable de ruido blanco. Es muy fácil trazar el correlograma teórico de este modelo, ya que, al ser  $E(\varepsilon_t \varepsilon_t) = 0 \quad \forall t$ , se tiene que  $\rho_t = 0 \quad \forall t$ .

Al graficar  $\rho_t$  frente a  $t$  resulta lo que se conoce como correlograma, como la figura 3.6. Esta es una herramienta muy útil para determinar qué tan correlacionado está el término de error respecto a los errores pasados.



**Figura 3.6** Ejemplo de correlograma.

Como una variable de ruido blanco se caracteriza por tener nulos todos sus coeficientes de autocorrelación, entonces su correlograma teórico aparecerá en blanco.

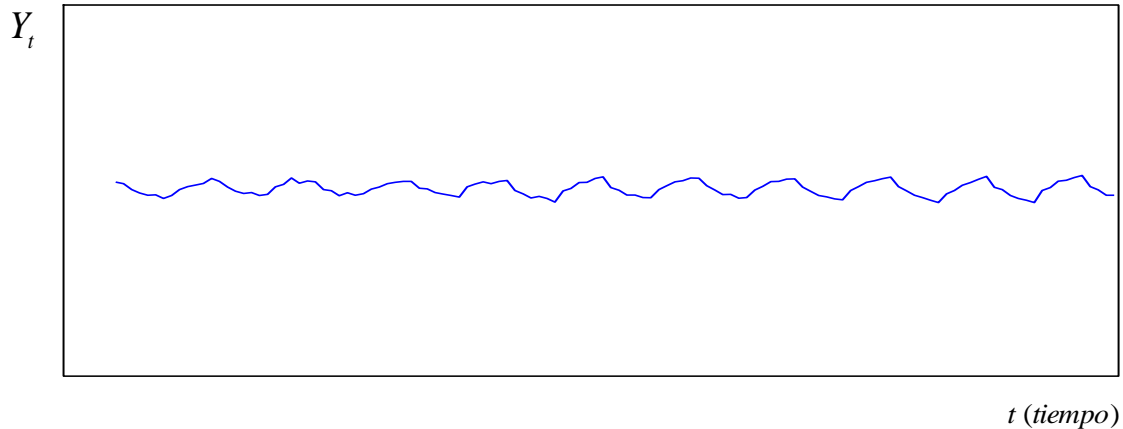
Es importante mencionar que el correlograma es un instrumento que tiene diversas aplicaciones en el análisis de las series de tiempo.

### 3.6 Estacionariedad

Las series de tiempo se pueden generar a través de un proceso estocástico o aleatorio (conjunto de mediciones de algún fenómeno en el tiempo) y un conjunto concreto de información se puede considerar como una realización del proceso estocástico subyacente. (Gujarati, 2003)

Así como se utiliza información muestral para inferir acerca de una población, para las series de tiempo se utiliza la realización para inferir acerca de un proceso estocástico subyacente.

Se dice que una serie de tiempo es estacionaria cuando su media y su varianza permanecen constantes a lo largo del tiempo.

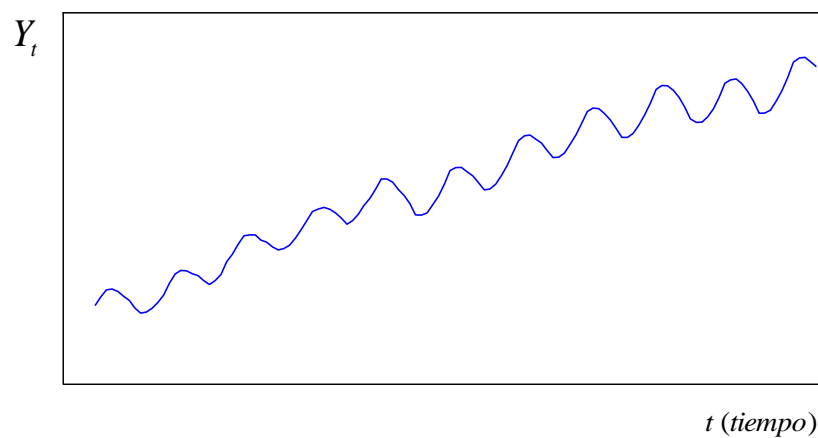


**Figura 3.7** Ejemplo de serie ( $Y_t$ ) estacionaria.

Si las características del proceso estocástico cambian con el tiempo se dice que el proceso es no estacionario. En la práctica, es común encontrar series de tiempo no estacionarias, pero que cumplen la condición básica de que su estructura de comportamiento sea semejante en periodos diferentes de tiempo, es decir, no estacionarias homogéneas. Estas series poseen la característica particular de que tras las adecuadas transformaciones de sus datos, se convierten en estacionarias.

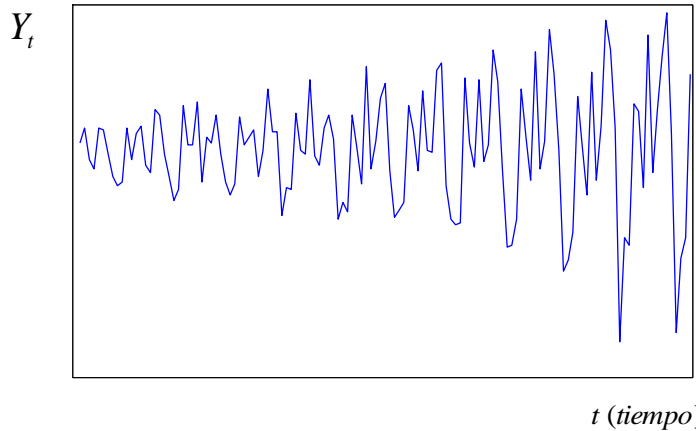
La no estacionariedad homogénea puede presentarse de tres formas:

1. Cuando la media no es constante en el tiempo, lo cual se corrige aplicando el método de diferenciación.



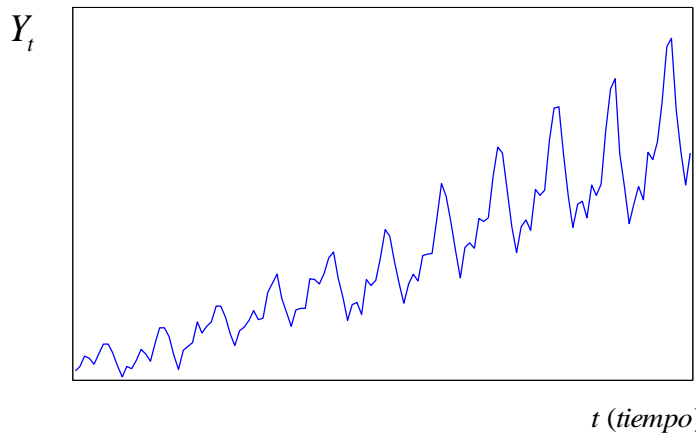
**Figura 3.8** Ejemplo de serie ( $Y_t$ ) con media no constante.

2. Cuando la varianza no permanece constante en el tiempo, lo cual se corrige mediante la estabilización de la varianza. La transformación más utilizada es la de Box-Cox.



**Figura 3.9** Ejemplo de serie ( $Y_t$ ) con varianza no constante.

3. Cuando la media y la varianza no son constantes en el tiempo, entonces lo primero que se corrige es la varianza.



**Figura 3.10** Ejemplo de serie ( $Y_t$ ) con media y varianza no constantes.

Cabe señalar que cuando la serie es no estacionaria y no homogénea, entonces el problema no tiene solución.

Existen dos tipos de estacionariedad: la débil y la estricta. (Guerrero, 2003)

Se dice que un proceso estocástico es **estacionario débil** si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende



solamente de la distancia o rezago entre estos dos periodos de tiempo y no del tiempo en el cual se ha calculado la covarianza.

Entonces si la serie de tiempo  $Y_t$  es estacionaria, tendrá las siguientes propiedades:

$$\text{Media: } E(Y_t) = \mu_Y \quad \forall t$$

$$\text{Varianza: } \text{Var}(Y_t) = E[(Y_t - \mu_Y)^2] = \sigma_Y^2 \quad \forall t$$

$$\text{Covarianza: } \gamma_h = \text{Cov}(Y_t, Y_{t+h}) = E[(Y_t - \mu_Y)(Y_{t+h} - \mu_Y)]$$

La covarianza (o autocovarianza) al rezago  $h$ , es la covarianza entre los valores  $Y_t$  y  $Y_{t+h}$ , es decir, entre dos valores de  $Y$  que están separados  $h$  periodos. Si  $h$  es igual a 0, se obtiene  $\gamma_0$  que es simplemente la varianza de  $Y$ ; si  $h=1$ ,  $\gamma_1$  es la covarianza entre dos valores adyacentes de  $Y$ .

La estacionariedad puede ser calculada por medio de la función de autocorrelación definida como:

$$\rho_h = \frac{\gamma_h}{\gamma_0}$$

Nótese que si  $h$  es igual a 0, entonces  $\rho_0$  será 1.

La estacionariedad débil no garantiza la estabilidad completa del proceso.

Una serie de tiempo es **estrictamente estacionaria** cuando la probabilidad conjunta de  $Y_1, Y_2, \dots, Y_T$  es la misma que la distribución conjunta de:  $Y_{(1+h)}, Y_{(2+h)}, \dots, Y_{(T+h)}$ . En otras palabras, un proceso estocástico estacionario estricto es aquel cuya distribución conjunta es invariante con respecto a un desplazamiento en el tiempo:

$$f_{Y_1 Y_2 \dots Y_T} = f_{Y_{(1+h)} Y_{(2+h)} \dots Y_{(T+h)}}$$

donde  $h$  es el desfase o rezago en el tiempo.

Si se parte de la suposición de que las variables aleatorias tienen distribución normal, entonces la estacionariedad débil coincide con la estacionariedad estricta.

Para el caso de la estacionariedad estricta se deben cumplir tres condiciones:

1. La media es constante y el valor observado para cada periodo puede representarse por dicha media.

$$E(Y_t) = E(Y_{t+h}) = \mu \qquad \bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t$$

2. La varianza es constante:

$$Var(Y_t) = Var(Y_{t+h}) = \sigma_Y^2$$

3. El coeficiente de correlación  $\rho_h$  debe depender únicamente de  $h$ , es decir, de la amplitud del intervalo que los separa.

Una de las propiedades más importantes de los procesos estocásticos es tener incrementos estacionarios, es decir, si  $Z_t$  es un proceso estacionario, entonces el proceso  $W_t = Z_t - Z_{t-1}$  es también estacionario.

Algebraicamente esto puede demostrarse de la siguiente manera:

$$\begin{aligned} E(W_t) &= E(Z_t - Z_{t-1}) \\ &= E(Z_t) - E(Z_{t-1}) \\ &= \mu - \mu \\ &= 0 \end{aligned}$$

$$\begin{aligned} Var(W_t) &= Var(Z_t - Z_{t-1}) \\ &= Var(Z_t) + Var(Z_{t-1}) - 2Cov(Z_t, Z_{t-1}) \\ &= \sigma^2 + \sigma^2 - 2\gamma_1 \\ &= 2\sigma^2 - 2\gamma_1 \\ &= 2(\sigma^2 - \gamma_1) \end{aligned}$$

$$\begin{aligned} Cov(W_t, W_{t+h}) &= E[(W_t - \mu)(W_{t+h} - \mu)] \\ &= E[W_t W_{t+h}] \\ &= E[(Z_t - Z_{t-1})(Z_{t+h} - Z_{t+h-1})] \\ &= E(Z_t Z_{t+h} - Z_t Z_{t+h-1} - Z_{t-1} Z_{t+h} + Z_{t-1} Z_{t+h-1}) \\ &= E(Z_t Z_{t+h}) - E(Z_t Z_{t+h-1}) - E(Z_{t-1} Z_{t+h}) + E(Z_{t-1} Z_{t+h-1}) \\ &= \gamma_h - \gamma_{h-1} - \gamma_{h+1} + \gamma_h \\ &= 2\gamma_h - \gamma_{h-1} - \gamma_{h+1} \end{aligned}$$

$Cov(W_t, W_{t+h})$  depende solo de  $h$  y la media es constante, por lo tanto  $W_t$  es un proceso estacionario.

Si se supone que el origen de  $Y$  es desplazado de  $Y_t$  hasta  $Y_{t+m}$ , entonces si se espera que  $Y_t$  sea estacionaria, la media, la varianza y la covarianza de  $Y_{t+m}$  serán las mismas que las de  $Y_t$ . Por lo tanto, si una serie de tiempo es estacionaria, su media, su varianza y su autocovarianza permanecen iguales sin importar el momento en el cual se midan.

Dado que en la práctica sólo se tiene una realización muestral, entonces se podrá calcular únicamente la función de autocorrelación muestral denominada:  $\hat{\rho}_h$ . Pero para esto antes se debe calcular la autocovarianza muestral al rezago  $h$ , definida como:  $\hat{\gamma}_h$ , y la varianza muestral definida como:  $\hat{\gamma}_0$ ; que están definidas como sigue:

$$\hat{\gamma}_h = \frac{\sum_{t=h+1}^T (Y_t - \bar{Y})(Y_{t-h} - \bar{Y})}{T - h} \quad \text{y} \quad \hat{\gamma}_0 = \frac{\sum_{t=1}^T (Y_t - \bar{Y})^2}{T}$$

donde  $T$  es el tamaño de la muestra y  $\bar{Y}$  es la media muestral.

Por lo tanto, la función de autocorrelación muestral al rezago  $h$  es:

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0}$$

Que representa la razón entre la autocovarianza y la varianza muestrales. La gráfica de  $\hat{\rho}_h$  frente a  $h$  se conoce como correlograma muestral.

### 3.7 Diferenciación

La forma más común de encontrar una serie de tiempo no estacionaria homogénea, es cuando su media no es constante a lo largo del tiempo. Puede mostrar existencia de tendencias crecientes y decrecientes e incluso de cambio de pendiente dentro de estas tendencias. Estos problemas pueden corregirse diferenciando una o más veces la serie, obteniendo una serie estacionaria. Si se diferencia una vez, la serie resultante se conoce como “la de primeras diferencias”. (Guerrero, 2003)

Si la serie de tiempo es no estacionaria en la media, es decir, que muestre un crecimiento constante a lo largo del tiempo o tendencia lineal, se puede transformar en estacionaria obteniendo sus primeras diferencias. Se aplica el método de diferenciación

que consiste en restar los valores de las observaciones uno de otro en un cierto orden, hasta obtener una serie estacionaria. Donde la nueva serie  $\{X_1, X_2, \dots, X_{t-1}\}$  está formada por la serie original  $\{Y_1, Y_2, \dots, Y_n\}$ . La primera diferencia está dada por:

$$\nabla Y_t = Y_t - Y_{t-1}$$

El proceso puede repetirse tantas veces como sea necesario para asegurar la estacionariedad. Sin embargo, son raras las ocasiones en que son necesarias dos o más diferenciaciones. De hecho, la primera diferenciación corrige el desnivel de la media, mientras que la segunda corrige los cambios de pendiente de la tendencia de la serie de tiempo. Cabe señalar que la segunda diferencia no se obtiene de restar  $Y_t - Y_{t-2}$  sino que está dada por:

$$\begin{aligned}\nabla^2 Y_t &= \nabla Y_t - \nabla Y_{t-1} \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2}\end{aligned}$$

Como se puede notar, cada vez que se realiza una diferencia se pierde una observación, una vez que se ha eliminado la tendencia, si se sigue diferenciando se obtendrán series sin tendencia; sin embargo, se pierden observaciones y el modelo puede volverse complicado. Cuando se elimina una tendencia se quiere obtener una serie que pueda ser analizada fácilmente, para después introducir nuevamente la tendencia, por ejemplo, si se calcularon las primeras dos diferenciaciones obtiene:

$$X_t = Y_t - Y_{t-1}$$

Entonces, puede reconstruirse la serie de tiempo en forma recursiva de la siguiente manera:

$$\begin{aligned}Y_t &= X_t + Y_{t-1} \\ &= X_t + X_{t-1} + Y_{t-2} \\ &\quad \vdots \\ &= X_t + X_{t-1} + X_{t-2} + X_{t-3} + \dots\end{aligned}$$

Una vez aplicando lo anterior, la media será constante y por lo tanto la serie será estacionaria.

A continuación se enunciarán tres formas fundamentales para identificar el hecho de que una serie sea no estacionaria:

- Observando la gráfica de la serie, tratando de ver si existen distintos niveles medios, tendencias, cambios de pendiente, etc.
- Observando el gráfico de la función de autocorrelación (correlograma muestral). Cuando una serie de tiempo es no estacionaria en su media tiene una función de autocorrelación en la que los rezagos se amortiguan muy lentamente, además de un primer coeficiente muy cercano a la unidad, lo cual indica la necesidad de una diferenciación. Cuando se consigue hacer estacionaria la serie de tiempo, el correlograma cambia totalmente de aspecto, en el sentido de que se produce una amortiguación de los rezagos en una forma mucho más rápida, incluso exponencial en el caso de los procesos autorregresivos.
- Analizando si los coeficientes del modelo ya estimado cumplen las condiciones de estacionariedad.

La expresión final para la nueva serie  $Z_t$ , que será protagonista de las etapas de la metodología de Box-Jenkins se puede expresar en términos del operador de retraso  $B$  o del operador de adelanto  $F$ .

El operador de retraso  $B$  está definido por la siguiente relación:

$$BZ_t = Z_{t-1} \quad \forall t$$

Cuando se aplica sucesivamente este operador se obtiene el siguiente resultado:

$$\begin{aligned} B^2 Z_t &= B(BZ_t) = Z_{t-2} \\ B^3 Z_t &= B(B^2 Z_t) = Z_{t-3} \\ &\vdots \\ B^h Z_t &= B(B^{h-1} Z_t) = Z_{t-h} \end{aligned}$$

Por lo tanto, en general se tiene:

$$B^h Z_t = Z_{t-h} \quad \text{para } h = 0, 1, 2, \dots, \text{ y } \forall t$$

Cuando se opera  $B^h$  a  $Z_t$  el resultado que se obtiene es la variable retrasada  $h$  periodos.

Partiendo de lo anterior, el operador diferencia puede escribirse en función de la operación retraso como:

$$\begin{aligned}\nabla Y_t &= Y_t - Y_{t-1} \\ &= Y_t - BY_t \\ &= (1 - B)Y_t\end{aligned}$$

Por lo tanto  $\nabla = 1 - B$ .

A continuación se define el operador adelanto en términos del operador de retraso de la siguiente forma:

$$\begin{aligned}F &= B^{-1} \\ FZ_t &= Z_{t+1} \\ F^m Z_t &= Z_{t+m}\end{aligned}$$

### 3.8 Metodología de Box-Jenkins

La metodología de Box-Jenkins trata sobre el análisis de series de tiempo: predicción y control, la cual estableció una nueva serie de herramientas de predicción para el análisis univariante. Cuando se habla de análisis univariante se está considerando que la estimación y predicción de valores futuros de una serie de tiempo se basará únicamente en los valores anteriores de la misma serie y no en los de otras variables exógenas. Las predicciones en este tipo de análisis se efectúan mediante un modelo teórico (expresión algebraica que relacionará el valor actual de la serie con valores precedentes de la misma y con ciertas perturbaciones), que se ajusta a los datos de la serie. Este tipo de modelos se conocen con el nombre de: Modelos Autorregresivos Integrados de Medias Móviles (*ARIMA*). (Bowerman, O'Connell y Koehler, 2007) Estos modelos pueden expresarse en su forma general de la siguiente manera:

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots, Y_{t-n}; \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-n})$$

donde  $Y_t$  es el valor de la variable en el periodo  $t$  y  $\varepsilon_t$  es la perturbación aleatoria en el periodo  $t$ .

Box y Jenkins propusieron una familia amplia de modelos *ARIMA* y desarrollaron una metodología que consiste en elegir el miembro de esta familia que pueda considerarse generador de la serie de tiempo real en estudio. Después utilizan el modelo elegido

para predecir valores futuros de la serie. La identificación del miembro óptimo de la familia *ARIMA* se lleva a cabo a partir de la observación de dos funciones calculadas con los datos de la serie: las funciones de autocorrelación simple y parcial. La función de autocorrelación simple fue anteriormente definida. La función de autocorrelación parcial es importante para la identificación de modelos autorregresivos, ya que sin ella sería prácticamente imposible conocer el orden del proceso (máximo desfase de la variable  $Y$  en el modelo). La función de autocorrelación está formada por los coeficientes de autocorrelación parcial de los diferentes desfases, los cuales se pueden definir de la siguiente forma: el coeficiente de correlación parcial con rezago  $h$  muestra la correlación existente entre las observaciones de  $Y$  alejadas  $h$  periodos, después de eliminar los efectos intermedios del resto de las observaciones. La función de autocorrelación parcial además de identificar los modelos *ARIMA* también da el número requerido de parámetros.

El supuesto básico de la metodología Box-Jenkins es que los valores sucesivos de la serie temporal son dependientes estadísticamente, ya que si esto no ocurriera así, entonces no se podrían inferir valores futuros a partir de valores pasados, por lo que no tendría sentido este tipo de análisis univariante. Pueden destacarse los siguientes aspectos de esta metodología:

- Las predicciones deben tener un horizonte de corto plazo. Los modelos *ARIMA* son idóneos para predicciones a corto plazo debido a que la relación estadística que vincula  $Y_t$  con los valores precedentes de la serie, generalmente contienen pocos términos y siempre los más recientes, por lo que ésta se diluye con el tiempo. Por otra parte los intervalos de confianza que acompañan a las predicciones, van ampliándose conforme se alejan en el tiempo, hasta que llega un momento que dichas predicciones tienen un error probable tan grande que pierden su utilidad.
- Los datos deben tomarse siempre en intervalos iguales de tiempo, de lo contrario se rompería la homogeneidad de la serie. La metodología Box-Jenkins sólo se puede aplicar a variables discretas. Sin embargo las variables continuas siempre

se pueden convertir en discretas eligiendo las observaciones a intervalos iguales de tiempo.

- Para la construcción de un buen modelo de *ARIMA* son necesarios al menos 50 datos, pero si el modelo tiene una variación estacional es deseable un tamaño de muestra mayor, de al menos cinco períodos estacionales completos.
- Las series utilizadas deben ser estacionarias.

### 3.8.1 Proceso de modelación de Box-Jenkins

Box-Jenkins proponen un procedimiento de cuatro etapas cuyo objetivo es encontrar un modelo que pueda considerarse como generador de la serie de tiempo en estudio.

Este proceso se caracteriza sobre todo por ser iterativo, en el sentido de que si el resultado no es satisfactorio en cualquiera de las etapas, debe volverse al principio. En el figura 3.11 se presentan las cuatro etapas.

Para facilitar los cálculos se usa Eviews, paquete estadístico que puede ser empleado para llevar a cabo análisis estadístico general, pero es especialmente útil para realizar análisis econométrico, como modelos de corte transversal, datos en panel y estimación y predicción con modelos de series de tiempo.

#### **Etapa 1: Identificación**

##### **Análisis exploratorio**

Box y Jenkins proponen una familia de modelos *ARIMA* como posibles generadores de la serie de tiempo dada. Primero se grafica la serie objeto de estudio para identificar si es o no estacionaria. Si no lo es se procede a realizar las transformaciones adecuadas para convertirla en estacionaria. Con la serie estacionaria se realiza la elección del miembro de la familia que mejor se adapte a la serie de datos.



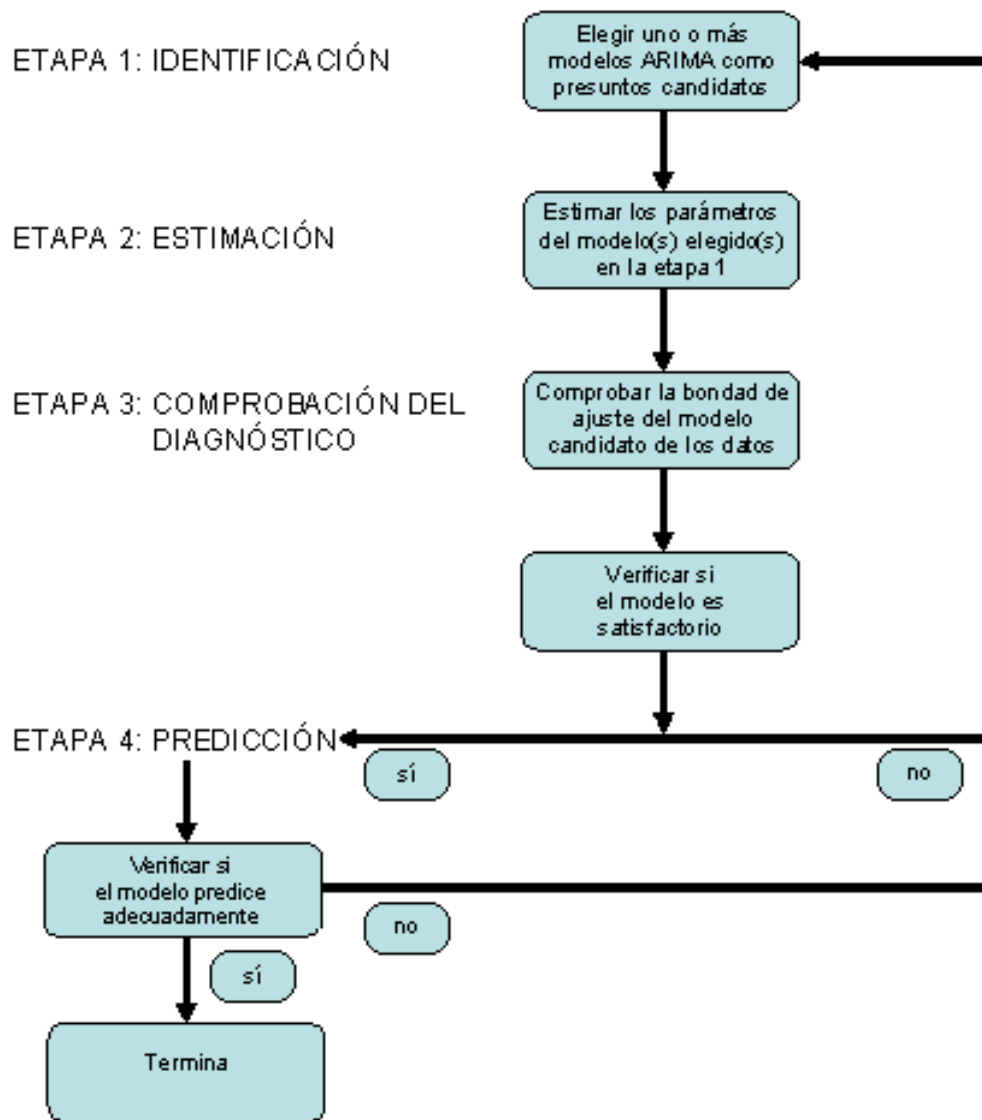


Figura 3.11 .Metodología Box-Jenkins. (Gujarati, 2003)

Uno de los métodos formales para analizar si una serie es estacionaria es el test de Dickey Fuller (DF), test de raíces unitarias más difundido, que presenta una variante denominada test de Dickey Fuller aumentado (ADF). El test de Dickey Fuller aumentado (ADF), que permite identificar la presencia de raíces unitarias, considera un modelo autorregresivo de primer orden: (Guerrero, 2003)

$$Y_t = \mu + \rho Y_{t-1} + a_t$$

donde  $\mu$  y  $\rho$  son parámetros a estimar y  $a_t$  es un término del error que se supone cumple la propiedad de ruido blanco. Dado que  $\rho$  es un coeficiente de autocorrelación que, toma valores  $-1 < \rho < 1$ , si  $\rho = 1$ , la serie  $Y_t$  es no estacionaria, la varianza decrece en el tiempo. Si el valor de  $\rho > 1$ , entonces se dice que la serie es explosiva.

De esta forma, la hipótesis de estacionariedad de una serie puede ser evaluada analizando si el valor absoluto de  $\rho$  es estrictamente menor que uno. Este test plantea como hipótesis nula la existencia de una raíz unitaria:

$$H_0 : \rho = 1$$

$$H_a : \rho > 1$$

La interpretación al test estadístico ADF sería: Se rechaza la hipótesis nula si el estadístico de Dickey-Fuller Aumentado (ADF) es mayor que los valores críticos de MacKinnon<sup>2</sup>. Por lo tanto se afirmarí que la serie es estacionaria en la componente regular; caso contrario hay un problema de raíces unitarias que amerita una primera diferencia para resolver el problema de no estacionariedad.

### Estudio de la estacionariedad

El tratamiento de la estacionariedad en los modelos *ARIMA* se da habitualmente cuando los datos son mensuales, como en el caso de este trabajo, es decir, en un contexto de predicción a corto plazo. Por ello, el tratamiento de la estacionariedad debe ocupar un papel central en el proceso de modelización.

Los modelos *ARIMA* tratan de captar el comportamiento estacional en forma similar al componente regular. En este sentido, el componente estacional puede exigir una o varias diferencias de orden estacional ( $D$ ).

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

---

<sup>2</sup> MacKinnon (1991) estima los valores de respuesta del test utilizando los resultados de la simulación, permitiendo que se calculen los valores críticos de Dickey-Fuller para cualquier tamaño muestral y cualquier número de variables explicativas. El programa Eviews proporciona de manera automática esta salida.

original, por ello, se elegirá entre ellas la que posea menor desviación estándar y se procederá a la identificación del modelo *ARIMA* a través de esta selección.

### Identificación de posibles modelos *ARIMA*

El primer paso en el análisis de una serie es la identificación del modelo. Si la serie no presenta estacionalidad, determinar los órdenes del modelo  $ARIMA(p,d,q)$ , y si la serie presenta estacionalidad de periodo  $E$ , los órdenes del modelo serán de tipo multiplicativo  $ARIMA(p,d,q)(P,D,Q)_E$ . (Bowerman, O'Connell y Koehler, 2007)

Una vez determinados los parámetros  $d$  y  $D$  en el análisis exploratorio, el siguiente paso será identificar los parámetros  $p$  y  $q$  y, si la serie presentara el caso de estacionalidad, los parámetros  $P$  y  $Q$ . El instrumento técnico básico para identificar un modelo *ARIMA* es la denominada función de autocorrelación simple y parcial. Las autocorrelaciones parciales tienen un comportamiento contrario a las autocorrelaciones simples: se cortan para los procesos autorregresivos y decaen exponencialmente o en forma sinusoidal para los modelos de media móvil.

En la práctica cuando se estudia las autocorrelaciones simples y parciales obtenidas a través de un conjunto de datos, se observan las primeras autocorrelaciones que son las más confiables y la identificación se realiza a partir de éstas.

Sin embargo, es importante el análisis de la componente estacional. Por lo tanto, si al calcular las autocorrelaciones se notará que, luego de un corte o decaimiento inicial correspondiente al comportamiento de un *AR* o *MA*, las autocorrelaciones reaparecen nuevamente tomando valores altos alrededor del periodo estacional. Esto significa pues, que hay dos enlaces: uno entre periodos sucesivos como en cualquier modelo *ARIMA* y otro entre periodos distintos. Por lo tanto, se necesita usar un modelo de tipo multiplicativo y se presenta como la combinación de dos modelos  $ARIMA(p,d,q)(P,D,Q)_E$ , donde el primero corresponde a la parte regular y el segundo a la parte estacional.

No resulta fácil seleccionar un modelo *ARIMA*. En muchas ocasiones se tendrán dudas razonables sobre el modelo más adecuado. Es importante tomar en cuenta que

la etapa de identificación del modelo no representa una decisión irreversible, sino como el inicio de un proceso de selección que puede ser revisado en etapas sucesivas.

Es importante señalar que en esta etapa se puede considerar un estadístico que permita continuar con la etapa de la selección del modelo.

En el programa Eviews se incorpora el estadístico  $Q$  de Ljung-Box (Bowerman, O'Connell y Koehler, 2007):

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

para  $n$  datos y  $m$  coeficientes, que se distribuye aproximadamente como una  $\chi^2$  con  $(n-m-p)$  grados de libertad, bajo la hipótesis nula:

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

Es decir, sólo si el valor de  $Q$  calculado supera el teórico de las tablas, podrá rechazarse la hipótesis nula de todos los coeficientes de autocorrelación y tendrá sentido buscar el orden del proceso *ARIMA* correspondiente. En caso contrario, la serie analizada parece corresponder a un ruido blanco.

## **Etapa 2: Estimación**

Una vez identificado el presunto modelo se procede a estimar los parámetros de éste, que pueden ser obtenidos mediante paquetes estadísticos, como Eviews.

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

$$H_0 : \beta_i = 0$$

Dicha hipótesis puede ser interpretada como que la variable asociada al parámetro  $\beta_i$  no mejora el ajuste con respecto al obtenido con las restantes variables incluidas en el modelo. Si el p-value asociado al valor del estadístico de contraste  $t$  es menor que  $\alpha$ ,

se rechazará la hipótesis nula al nivel de significación  $\alpha$ , y se concluirá que el parámetro es significativo y será considerado en el modelo.

El hecho de que el modelo identificado sea adecuado no debe hacer descartar la posibilidad de que otro modelo algo más complejo pueda ajustarse mejor a la serie observada. Para comprobar si algún otro modelo se ajusta mejor conviene utilizar la técnica de sobre ajuste, es decir, estimar cada uno de los modelos obtenidos al aumentar en una unidad cada uno de los órdenes  $p$ ,  $q$ ,  $Q$  y  $P$  en el modelo original.

Una cuestión importante en el análisis de los coeficientes es el examen de estacionariedad e invertibilidad. En las salidas que proporciona el Eviews suele aparecer junto al coeficiente estimado los coeficientes de estacionariedad e invertibilidad. Si el modelo estimado cumple con la estacionariedad, procede de factorizar la parte autorregresiva y se concluye que si el coeficiente de estacionariedad está próximo a 1, es un indicio de que el modelo no es estacionario, siendo aconsejable, en este caso, tomar una diferencia adicional. Por otra parte, para comprobar si el modelo estimado es invertible se factoriza la parte media móvil y se concluye, si algún valor de coeficiente de invertibilidad está próximo a 1, será indicativo de que el modelo estimado es no invertible. En muchos casos, este problema se ha podido producir como consecuencia de una sobre diferenciación. Para completar el análisis de los coeficientes, es conveniente examinar la matriz de correlación entre los coeficientes estimados con el objeto de detectar si existe el problema de multicolinealidad. Cuando la correlación entre dos coeficientes estimados está próximo a 1, será un indicio suficiente para considerar que existe un problema grave de multicolinealidad. Si esto ocurre los coeficientes estimados serían muy inestables con lo que podría cambiar mucho de una realización a otra. Para evitar este problema puede ser conveniente eliminar algún parámetro del modelo con el objeto de conseguir unos estimadores más estables aún a costa de que el grado de ajuste sea pequeño, y en consecuencia, lo que se debe entender en esta primera etapa, que los coeficientes deben ser estadísticamente significativos, que el modelo estimado cumpla con la estacionariedad e invertibilidad y que los coeficientes estimados no tengan el problema de multicolinealidad.

### **Etapa 3: Comprobación del diagnóstico**

Box y Jenkins proponen pruebas estadísticas para comprobar la adecuación del modelo elegido a los datos. Esto es, una vez que un modelo se ha identificado y sus parámetros se han estimado, es necesario detectar si existen violaciones a los supuestos, lo cual se realiza a través del análisis de los residuales. Se considera como residual aquella parte de las observaciones que no es explicada por el modelo, es decir, los residuales miden la discrepancia entre los valores observados y los valores estimados.

La validación consistirá que el residuo tiene media cero, que es estable en varianza, que sus observaciones están incorrelacionadas y que procede de una distribución normal.

Respecto de la nulidad de la media y la estabilidad de la varianza de la serie de los residuos, la primera es consecuencia del método de estimación de los coeficientes del modelo, mientras que la segunda es consecuencia de la estabilidad en la varianza de la serie de los residuos.

Los residuos deben ser incorrelacionados, es decir, no debe existir ningún tipo de correlación en la serie de residuos y para este fin se pueden utilizar las funciones de autocorrelación y las funciones de autocorrelación parcial. Sin embargo, un valor fuera de los límites de confianza en los retardos iniciales debería ser considerado como un indicio claro de que el modelo no es correcto. Un estadístico adecuado para contrastar la hipótesis de independencia, aleatoriedad o ruido blanco de los residuos es el estadístico de Box-Ljung que, para cada coeficiente de la función de autocorrelación simple, permite contrastar la hipótesis nula de que todos los coeficientes anteriores, hasta el correspondiente, son nulos. La interpretación es, si el nivel de significancia crítico o el p-value es menor que  $\alpha$ , se rechazará la hipótesis nula al nivel de significación  $\alpha$ , caso contrario se aceptaría que los residuos tienen un comportamiento independiente, aleatorio o de ruido blanco. (Bowerman, O'Connell y Koehler, 2007)

El estadístico de Jarque-Bera se utiliza para determinar si la serie sigue una distribución normal. Un valor bajo de la probabilidad conduce a rechazar la hipótesis nula de una distribución normal. Si el valor obtenido del estadístico Jarque-Bera es inferior al valor

de referencia de tablas y la probabilidad es distinta de cero, podemos suponer que la serie presenta características normales. (Bowerman, O'Connell y Koehler, 2007)

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

### **Análisis de intervención**

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

- La variable impulso, se trata de un atípico aditivo, el efecto de la observación anómala es inmediato y dura solo un periodo.
- La variable escalón, conocido también como level shift, es el caso opuesto, en el que dicho efecto es permanente.
- Finalmente la situación intermedia en que el efecto de la observación anómala no es permanente pero persiste durante algún tiempo. Este tipo de atípicos recibe el nombre de “transitorio” o “temporary change”.

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

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

### **Etapa 4: Predicción**

Una vez seleccionado el modelo definitivo que cumple satisfactoriamente los criterios de evaluación establecidos, puede pasarse a la etapa de predicción. En la mayoría de

los casos, el problema consiste en obtener estimaciones de las predicciones futuras, aunque también puede resultar útil analizar cómo se habría comportado el modelo si hubiera tenido que hacer una predicción dentro del período histórico ya conocido, que ha servido de base a su estimación y contraste.

Es importante en esta etapa de predicción evaluar la capacidad predictiva del modelo, esta evaluación dependerá que los errores de predicción un período por delante en media, sean menores comparadas con la desviación teórica. El error relativo es en realidad una aproximación al error de predicción un periodo por delante en términos relativos. (Bowerman, O'Connell y Koehler, 2007)

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

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

Para el período histórico pueden realizarse dos tipos de predicción:

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

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

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

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



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

Cuando el modelo queda rechazado en cualquiera de estas etapas, la metodología Box-Jenkins sugiere el modo de mejorarlo, con lo cual se vuelve a la etapa 1 y se repite el proceso.

### 3.8.2 Modelos de promedios móviles (MA)

Este tipo de modelos se introdujeron con el fin de representar a un proceso estocástico  $\{Y_t\}$  como una suma finita ponderada de choques aleatorios, es decir, que las variables del proceso son generadas por una combinación lineal de ruidos blancos actuales y pasados.

Cuando se habla de promedios o medias móviles, pareciera que la forma de obtener el modelo es promediando los choques que intervienen, pero no es así, ya que no es necesario que los parámetros sean positivos y tampoco que al sumarlos el resultado sea la unidad, como sucede en los promedios.

Si se define un proceso estocástico puro como  $\{Z_t\}$ , con media  $\mu = 0$ , varianza  $\sigma_Z^2$  y se asume que las perturbaciones aleatorias están distribuidas independientemente a través del tiempo, entonces el proceso  $\{Y_t\}$  se conoce como proceso de medias móviles de orden  $q$ . Se denota como  $MA(q)$  y está dado por: (Bowerman, O'Connell y Koehler, 2007)

$$Y_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

donde  $\theta_1, \theta_2, \dots, \theta_q$  determinan las ponderaciones y son constantes que pueden ser positivas o negativas. Se supone, por convención, que  $\theta_0 = 1$ , entonces el modelo se expresa de la siguiente forma:

$$Y_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

Todo proceso de medias móviles de orden finito siempre va a ser estacionario y va a tener una memoria que alcanza tantos periodos como orden tiene el proceso, las covarianzas y los coeficientes de autocorrelación distintos de 0, en número igual al orden del proceso.

Los modelos de medias móviles son utilizados en varios campos de estudio, principalmente en la econometría, debido a que los indicadores económicos son afectados por muchos eventos aleatorios.

Una de las ventajas que se presentan al utilizar los modelos de medias móviles es que se reduce considerablemente el número de parámetros, siendo este uno de los propósitos que se persiguen cuando se está analizando un fenómeno, ya que se busca representar el proceso en el menor número de parámetros posible, lo cual se conoce como principio de parsimonia. Este principio tiene como objetivo obtener el modelo que represente de la mejor manera el fenómeno que se está estudiando, pero evitando cuantos parámetros sea posible.

### 3.8.3 Dualidad entre los modelos de medias móviles (MA) y autorregresión (AR)

El modelo de medias móviles de primer orden  $MA(1)$  se define como: (Bowerman, O'Connell y Koehler, 2007)

$$Y_t = Z_t + \theta_1 Z_{t-1}$$

del cual se deriva:

$$Y_{t-1} = Z_{t-1} + \theta_1 Z_{t-2}$$

$$Y_{t-2} = Z_{t-2} + \theta_1 Z_{t-3}$$

Para facilitar la forma de expresar lo anterior se utiliza comúnmente  $-\theta$  y al despejar  $Z_{t-1}$  y  $Z_{t-2}$  se tiene:

$$Z_{t-1} = Y_{t-1} + \theta_1 Z_{t-2}$$

$$Z_{t-2} = Y_{t-2} + \theta_1 Z_{t-3}$$

Cuando se sustituye la fórmula de  $Z_{t-1}$  en la expresión para  $Y_t$  se obtiene:

$$\begin{aligned} Y_t &= Z_t + \theta_1(Y_{t-1} + \theta_1 Z_{t-2}) \\ &= Z_t + \theta_1 Y_{t-1} + \theta_1^2 Z_{t-2} \end{aligned}$$

Sustituyendo ahora  $Z_{t-2}$  en la expresión anterior:

$$\begin{aligned} Y_t &= Z_t + \theta_1 Y_{t-1} + \theta_1^2 Z_{t-2} \\ &= Z_t + \theta_1 Y_{t-1} + \theta_1^2 (Y_{t-2} + \theta_1 Z_{t-3}) \\ &= Z_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \theta_1^3 Z_{t-3} \end{aligned}$$

Si se continúa sustituyendo se obtiene que:

$$Y_t = \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \theta_1^3 Y_{t-3} + \dots + Z_t$$

Cuando se trabaja con modelos de orden mayor, lo anterior resulta válido. A la expresión anterior se le llama forma invertida de un proceso de medias móviles que, como se verá a continuación, corresponde a la ecuación de un proceso autorregresivo de orden infinito.

### 3.8.4 Modelos autorregresivos (AR)

Los procesos autorregresivos son aquellos que expresan los valores de las variables aleatorias del proceso como una media ponderada de variables anteriores, junto con un ruido blanco actual. También suele interpretarse a este tipo de procesos como modelos de regresión múltiple, pero con la diferencia de que la variable dependiente no es explicada por las variables independientes, sino por valores pasados de la misma, ponderados de acuerdo con los coeficientes de regresión denotados por  $\Phi_1, \dots, \Phi_p$ , por esa razón recibe el nombre de autorregresivo, ya que se necesita a sí misma.

Se define  $\{Z_t\}$  como un proceso puro que tiene media 0 y varianza  $\sigma_Z^2$ , entonces el proceso  $\{Y_t\}$  recibe el nombre de modelo autorregresivo de orden  $p$  y se denota por  $AR(p)$ .

Su expresión está dada por: (Bowerman, O'Connell y Koehler, 2007)

$$Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + Z_t$$

### 3.8.5 Modelos mixtos autorregresivos y de medias móviles (ARMA)

Los modelos mixtos o procesos  $ARMA$  son una generalización de los modelos  $MA$  y  $AR$ , que resultan de añadir una estructura de medias móviles a un proceso autorregresivo o una autorregresiva a un proceso de medias móviles. Además debe preservarse el principio de parsimonia.

Un modelo  $ARMA(p, q)$  está descrito por la siguiente ecuación: (Bowerman, O'Connell y Koehler, 2007)

$$Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

Comúnmente en la práctica, las series de tiempo que se encuentran son no estacionarias, debido a que tienen tendencia, varianza no constante o porque están influenciadas por la estacionalidad. Cuando se presenta tendencia esta puede ser eliminada por medio de la diferenciación, de lo que resultan los modelos  $ARIMA$  que se explican a continuación.

### 3.8.6 Modelos autorregresivos integrados de medias móviles (ARIMA)

En general, las series de tiempo de variables económicas con las que se trabaja normalmente suelen tener una evolución presente en el tiempo, lo que suponen que es un proceso con una media que no es constante y por lo tanto el proceso estocástico generador de la serie de tiempo no cumple las condiciones básicas para poder ser estacionario. Sin embargo, muchas de las series no estacionarias poseen una no

estacionariedad homogénea por lo que se pueden convertir en series estacionarias si se les aplica las transformaciones adecuadas sobre los datos de las series, es decir, si la serie es no estacionaria en la media se aplica la diferenciación y si es no estacionaria en la varianza, se aplica la transformación de Box-Cox. Una vez que se dispone de una serie estacionaria se le puede aplicar la metodología de Box-Jenkins para identificar posibles candidatos de la familia *ARIMA* como generadores de la serie de tiempo objeto de estudio.

Los modelos *ARIMA* nacen de la necesidad de analizar los fenómenos de tipo caminata aleatoria, con el objeto de eliminar los factores no estacionarios de variación que convertirán al modelo en uno estacionario. Si se supone que se tiene una serie no estacionaria debido a que su media no es constante, entonces se procede a aplicar la diferenciación; si  $Y_t$  es reemplazada por la diferencia  $\nabla^d Y_t$  en la ecuación del modelo *ARMA*( $p, q$ ): (Bowerman, O'Connell y Koehler, 2007)

$$Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

Entonces, el modelo resultante es apto para describir ciertos tipos de series de tiempo no estacionarias. Se define  $W_t = \nabla^d Y_t$  como la forma generalizada del modelo *ARIMA*:

$$W_t = \Phi_1 W_{t-1} + \dots + \Phi_p W_{t-p} + Z_t + \dots + \theta_q Z_{t-q}$$

En términos del operador de retraso se tiene que:

$$\begin{aligned} \Phi(B)\nabla^d Y_t &= \theta(B)Z_t, \quad d \geq 1 \\ \Phi(B)W_t &= \theta(B)Z_t \end{aligned}$$

donde el término  $\Phi(B)$  representa un polinomio autorregresivo de orden  $p$  y  $\theta(B)$  representa un polinomio de medias móviles de orden  $q$ .

Cuando se habla del término integrado en el modelo *ARIMA*, significa que  $Y_t$ , resulta de la inclusión del operador  $\nabla^d$ , obteniéndose así una suma infinita o integración de

términos  $W_t$ . Dicho proceso define la  $d$ -ésima diferencia de los datos, razón por la cual el modelo se denomina Autorregresivo Integrado y de Medias Móviles, *ARIMA* de orden  $(p,d,q)$ , donde  $p$  es el número de términos autorregresivos y  $q$  es el número de términos de media móvil. En la práctica, generalmente se tiene que  $d = 1$ .

A pesar de que los modelos *ARIMA* son aplicados a muchos fenómenos, existen series que no se pueden describir por medio de estos modelos, es por ello que resulta necesario ampliar la utilidad de los modelos *ARIMA* para el caso en que se tengan series de tiempo no estacionales.

Una serie de tiempo no estacionaria en la media, debido a la estacionalidad es aquella que tiene la siguiente característica: presenta fluctuaciones que se presentan de forma redundante cada periodo (normalmente un año), aunque pueden existir cambios graduales a lo largo de los años.

El periodo estacional se define por  $E$ , el cual involucra  $E$  observaciones intermedias y  $D$  representa la diferenciación estacional. Por lo tanto se utiliza el operador  $\nabla_E^D$ , donde se eliminan  $E \times D$  observaciones en forma automática, por lo cual se pueden tener expresiones totalmente estacionarias del tipo  $ARIMA(P,D,Q)_E$  como: (Bowerman, O'Connell y Koehler, 2007)

$$\Phi(B^E)\nabla_E^D Y_t = \theta(B^E)Z_t$$

donde:  $\Phi(B^E)$  = polinomio autorregresivo estacional de orden  $P$ .

$\theta(B^E)$  = polinomio de medias móviles estacional de orden  $Q$ .

$\{Z_t\}$  = sucesión de elementos de ruido blanco.

Como se puede notar, es posible aplicar las mismas consideraciones tanto a los modelos estacionales como a los procesos puramente estacionales, lo cual es de gran utilidad, ya que es común encontrar series de tiempo estacionales y no estacionales. Para tener en cuenta ambos efectos, Box-Jenkins sugieren el siguiente modelo: (Bowerman, O'Connell y Koehler, 2007)

$$\Phi(B^E) \nabla_E^D Y_t = \Psi(B^E) Z_t$$

donde  $Z_t$  es generado a través de un proceso  $ARIMA(p, d, q)$  y no es ruido blanco, es decir:

$$\Phi(B) \nabla^d Z_t = \theta(B) a_t$$

donde el ruido blanco está representado por  $a_t$ . A continuación se presenta el modelo estacional multiplicativo:

$$\Phi(B) \Phi(B^E) \nabla^d \nabla_E^D Z_t = \theta(B) \Psi(B^E) a_t$$

Este modelo se denota como  $ARIMA(p, d, q)(P, D, Q)_E$ . (Bowerman, O’Connell y Koehler, 2007)

Como es natural, cuando existe mayor complejidad en el modelo, a este le corresponderá una estructura más compleja.

### 3.8.7 Comportamientos típicos de la FAC Y FACP

Para realizar la primera etapa de la metodología de Box-Jenkins, es necesario tener presentes las características de la función de autocorrelación (FAC) y de la función de autocorrelación parcial (FACP), las cuales se presentan en la tabla 3.1.

**Tabla 3.1** Características de la FAC y la FACP de modelos generales no estacionales

PROCESO	FAC	FACP
$AR(p)$	Convergencia a cero, con comportamiento dictado por la ecuación: $\Phi(B)\rho_h = 0$ , para $h \geq p$ .	Solo las primeras $p$ autocorrelaciones parciales son distintas de cero.
$MA(q)$	Sólo las primeras autocorrelaciones son distintas de cero.	Sucesión infinita convergente a cero.
$ARMA(p, q)$	Comportamiento irregular de las primeras $q$ autocorrelaciones y después convergen a cero de acuerdo con: $\Phi(B)\rho_h = 0$ , para $h \geq p$ .	Sucesión infinita convergente a cero.

**Fuente:** Bowerman, Bruce L.; O’Connell, Richard T.; Koehler, Anne B.; Pronósticos, series de tiempo y regresión; Cengage Learning Editores, S. A. de C. V.; México; 2007, p. p. 436-437.

# 4

## PRONÓSTICOS DE LA DESOCUPACIÓN FEMENINA EN MÉXICO

---

---

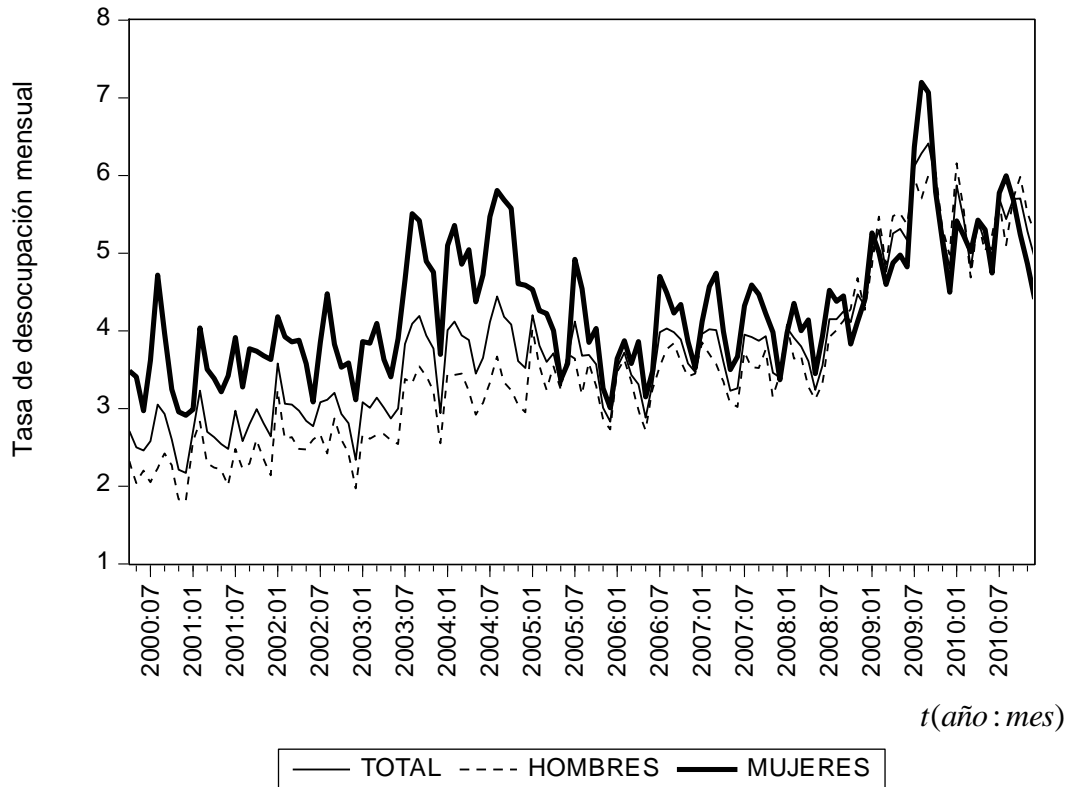
### 4.1. Introducción

La tasa de desocupación en México se eleva ante el hecho de que cada vez más personas se encuentran buscando algún empleo en el país. En septiembre de 2009 la tasa de desocupación nacional total alcanzó el mayor nivel desde que empezó la medición de este fenómeno a escala nacional en abril de 2000, con 6.41 por ciento, esto de acuerdo con lo informado por el INEGI. Por género, la desocupación afectó en ese mismo mes a 7.07 por ciento de las mujeres que forman parte de la PEA y al 6 por ciento de la PEA masculina. Esto se ve reflejado en la figura 4.1, donde se presentan las tasas de desocupación nacional total, de los hombres y de las mujeres, desde abril de 2000 a diciembre de 2010.

Otro detalle observable en la figura 4.1 es que la tasa de desocupación de las mujeres ha sido, en la generalidad de los meses, mayor que la de los hombres, tanto en etapas de crisis como en las de recuperación, aunque con el tiempo la diferencia ha ido disminuyendo. Esto puede deberse al crecimiento de la participación económica

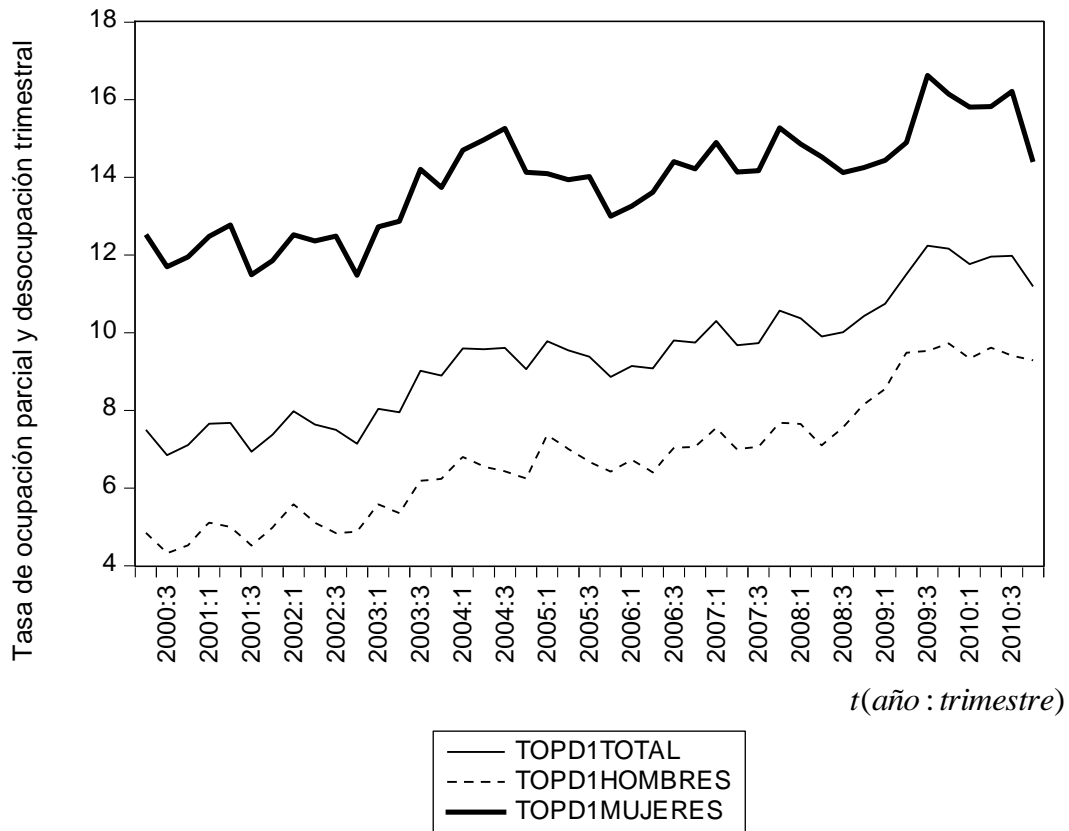


femenina de los últimos decenios, aunque de forma paralela se ha dado el aumento de las dificultades para obtener y conservar el empleo.



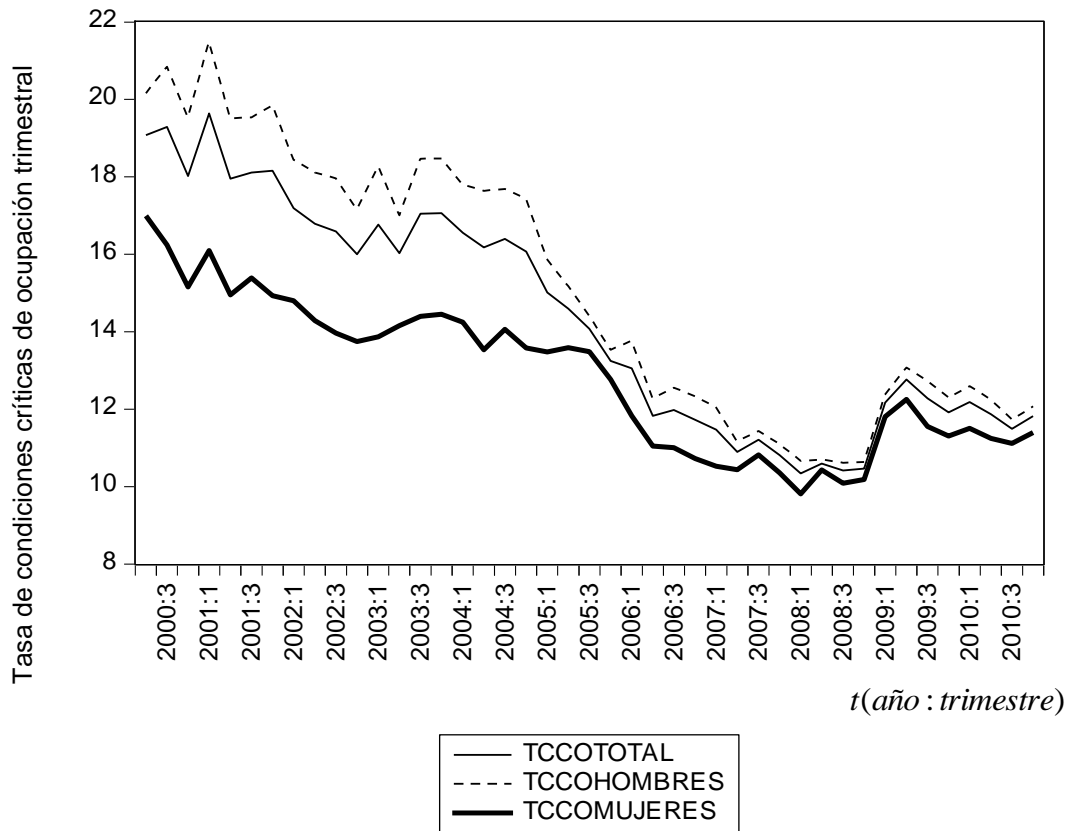
**Figura 4.1** Tasa de desocupación nacional (mensual) total, para hombres y para mujeres de abril de 2000 a diciembre de 2010. (INEGI, 2012)

Aun cuando la diferencia entre las tasas de desocupación es cada vez menor, sigue existiendo una mayor diferencia, a favor de los hombres, en las tasas complementarias de desocupación, como se observa en la figura 4.2 de la tasa de ocupación parcial y desocupación TOPD1 (porcentaje de la población económicamente activa que se encuentra desocupada, más la ocupada que trabajó menos de 15 horas en la semana de referencia); y la correspondiente a la tasa de condiciones críticas de ocupación TCCO (porcentaje de la población ocupada que se encuentra trabajando menos de 35 horas a la semana por razones de mercado, más la que trabaja más de 35 horas semanales con ingresos mensuales inferiores al salario mínimo y la que labora más de 48 horas semanales ganando hasta dos salarios mínimos).



**Figura 4.2** Tasa de ocupación parcial y desocupación TOPD1 (trimestral) total, para hombres y para mujeres del segundo trimestre de 2000 al cuarto trimestre de 2010. (INEGI, 2012)

Es decir, aunque las tasas de desocupación de hombres y mujeres están tendiendo a igualarse, en la TOPD1, que abarca tanto la ocupación parcial como la desocupación, las diferencias son más acentuadas entre las correspondientes a mujeres y a hombres, esto posiblemente como consecuencia de que muchas mujeres consiguen trabajos de ocupación parcial, por menos de 15 horas a la semana, debido a las actividades que tienen que desempeñar además en el hogar. En cuanto a la TCCO, actualmente no hay diferencias muy marcadas entre mujeres y hombres que están en condiciones críticas de ocupación, de acuerdo con lo que muestra la figura 4.3.



**Figura 4.3** Tasa de condiciones críticas de ocupación TCCO (trimestral) total, para hombres y para mujeres del segundo trimestre de 2000 al cuarto trimestre del 2010. (INEGI, 2012)

**Tabla 4.1** Desocupados por sexo de acuerdo a grupo de edad, para el cuarto trimestre de 2010.

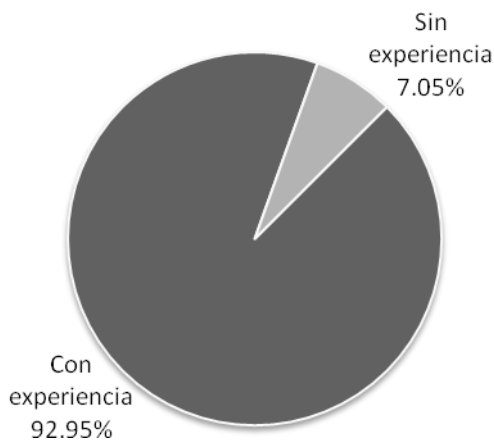
GRUPOS DE EDAD	DESOCUPADOS		
	Total	Hombres	Mujeres
<b>Total</b>	2,482,727	1,639,349	843,378
<b>14 a 19 años</b>	409,578	289,579	119,999
<b>20 a 29 años</b>	924,083	536,162	387,921
<b>30 a 39 años</b>	512,187	335,269	176,918
<b>40 a 49 años</b>	327,560	223,441	104,119
<b>50 a 59 años</b>	217,008	171,057	45,951
<b>60 años y más</b>	91,849	83,379	8,470
<b>No especificado</b>	462	462	0

Fuente: Con datos del BIE, INEGI, 2012.

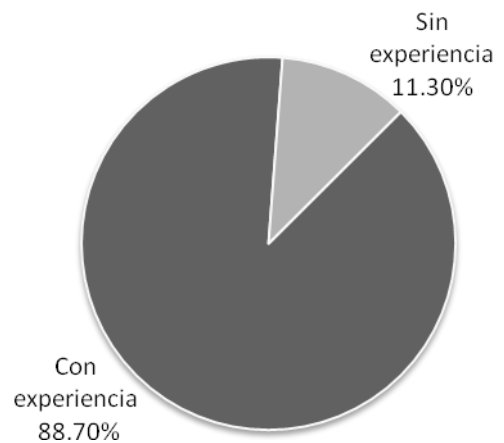
En valores absolutos, para el cuarto trimestre del 2010 la desocupación afectó a más de 2 millones 482 mil personas mayores de 14 años, de las cuales más de 1 millón 639 mil eran hombres, es decir, el 66.03 por ciento, y más de 843 mil eran mujeres, 33.97 por ciento del total. En la tabla 4.1 se muestra el número de desocupados por grupo de edad, tanto para hombres como para mujeres, para el cuarto trimestre de 2010.

La composición del desempleo en México es distinta en hombres y mujeres. Para el último mes de 2010, en el desempleo masculino la proporción de los que buscaban empleo sin experiencia laboral fue del 7.05 por ciento, mientras que en el desempleo femenino fue del 11.3 por ciento, siendo el 88.7 por ciento aquellas desocupadas porque perdieron su empleo.

### HOMBRES DESOCUPADOS



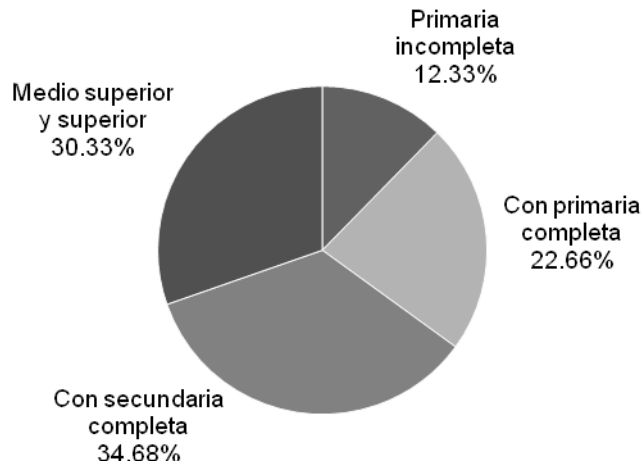
### MUJERES DESOCUPADAS



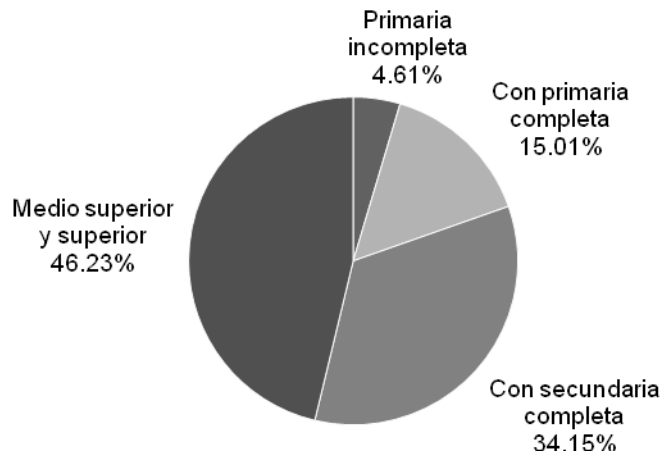
**Figura 4.4** Porcentaje de los desocupados por sexo de acuerdo a su condición de experiencia laboral, en diciembre de 2010. (INEGI, 2012)

En diciembre de 2010, por cada 100 hombres desocupados en México 30 contaban con estudios de nivel medio superior y superior, mientras que para las mujeres había 46 de cada 100, lo cual refleja una desventaja para la mujer.

## HOMBRES DESOCUPADOS



## MUJERES DESOCUPADAS



**Figura 4.5** Porcentaje de los desocupados por sexo de acuerdo a su nivel de instrucción, en diciembre de 2010. (INEGI, 2012)

Lo anterior refleja que también en los sectores en desempleo por sexo, pese a padecer una situación adversa, guardan expectativas, por contar con el capital humano de la educación, y son más selectivos, es decir, no aceptan cualquier ocupación bajo cualquier condición sin haber empleado un periodo de búsqueda previamente. Para el caso de las mujeres este detalle es más notorio, pues se tiene el 46.23% de mujeres desocupadas con nivel de instrucción medio superior y superior, contra el 30.33% de los hombres.

Hay que recordar que dada una situación de carencia de trabajo, a mayor selectividad mayor tiempo de búsqueda y, con ello, mayor probabilidad de ser captado como desempleado. El que un individuo esté dispuesto a asumir el costo de oportunidad de dejar de percibir, aunque sea ingresos marginales, quiere decir que para dicho individuo vale la pena esa inversión de tiempo completo en búsqueda y espera.

En contraste, un individuo no calificado no renuncia nunca a captar ingresos marginales, porque no tiene ni la expectativa ni la garantía de que obtendrá algo mejor a esos ingresos una vez que se presenta la oportunidad de obtenerlos, por carecer de capital humano, ser menos selectivo y más maleable para asumir cualquier rol ocupacional.

En el informe estructural de ocupación de México del INEGI para el cuarto trimestre de 2010 se indicó que 51.2 por ciento de la población desocupada, perdió o terminó su empleo; 29.48 por ciento renunció o dejó su trabajo; 5.51 por ciento dejó o cerró su negocio propio; 5.02 por ciento argumentó otras causas y en 8.78 por ciento de los casos se trató de personas sin experiencia laboral, es decir, de quienes se han incorporado recientemente a la demanda de empleo. Los porcentajes para hombres y mujeres correspondientes a cada una de estas causas aparecen en la tabla 4.2:

**Tabla 4.2** Desocupados por sexo de acuerdo a grupo de edad, para el cuarto trimestre de 2010.

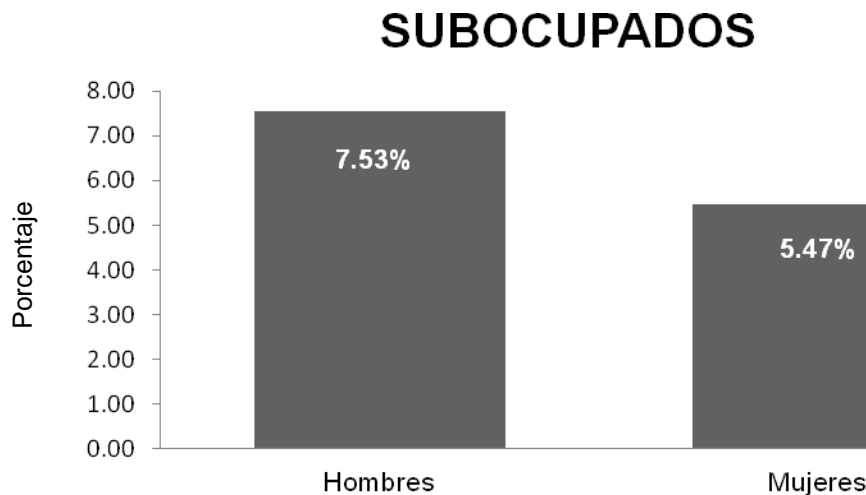
MOTIVOS DE DESOCUPACIÓN	DESOCUPADOS		
	Total	Hombres	Mujeres
<b>Perdió o terminó su empleo</b>	51.20%	57.89%	38.21%
<b>Renunció o dejó su empleo</b>	29.48%	23.50%	41.12%
<b>Dejó o cerró un negocio propio</b>	5.51%	6.14%	4.30%
<b>Otras causas</b>	5.02%	6.89%	3.60%
<b>Sin experiencia laboral</b>	8.78%	7.09%	12.78%

Fuente: Con datos del BIE, INEGI, 2012.

Otro de los fenómenos que ilustran la presión ejercida por las condiciones económicas sobre la población es el de la subocupación, concepto en el que se clasifica a las

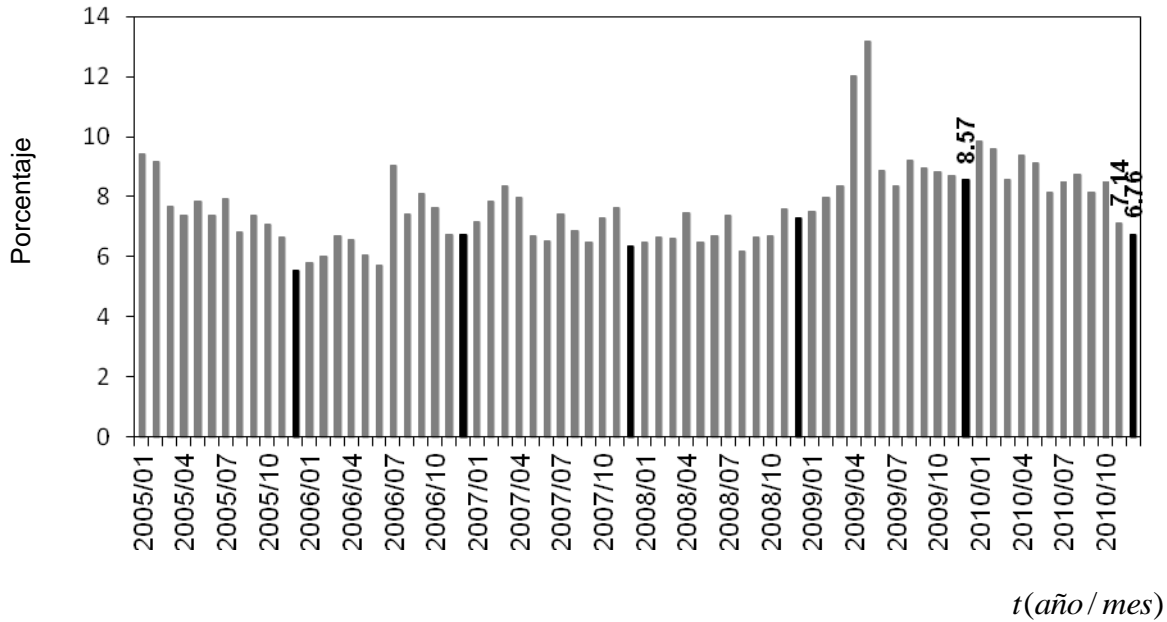
personas de 14 años y más que tienen una ocupación pero han declarado en la ENOE “tener necesidad y disponibilidad para trabajar más horas”.

Así la información de esta encuesta nacional de ocupación y empleo señala que la población subocupada para diciembre de 2010 representó 6.4 por ciento de la población económicamente activa y 6.8 por ciento de la población ocupada. Estos porcentajes comparados con los correspondientes a diciembre de 2009 son menores, pues se ubicaron en 8.2 y 8.6 por ciento respectivamente. En diciembre de 2010, más hombres que mujeres presentaron características de subocupación, correspondiendo a esta categoría el 7.53 por ciento de la población masculina frente al 5.47 por ciento de la femenina, como se muestra en la figura 4.6.



**Figura 4.6** Porcentaje de la población por sexo subocupada, en diciembre de 2010. (INEGI, 2012)

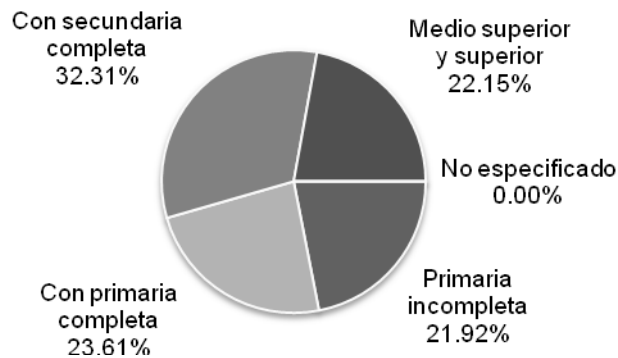
En su comparación mensual, este indicador reportó una disminución de 0.38 puntos porcentuales en el último mes de 2010 respecto al mes inmediato anterior.



**Figura 4.7** Porcentaje mensual de la población ocupada clasificada como subocupada, de enero de 2005 a diciembre de 2010. (INEGI, 2012)

Como ocurre con la relación de que a mayor preparación académica se presentan más dificultades para encontrar trabajo, en el subempleo se presenta una situación similar. Para diciembre de 2010, de la población subocupada 22 de cada 100 contaban con estudios de nivel medio superior y superior.

## POBLACIÓN SUBOCUPADA

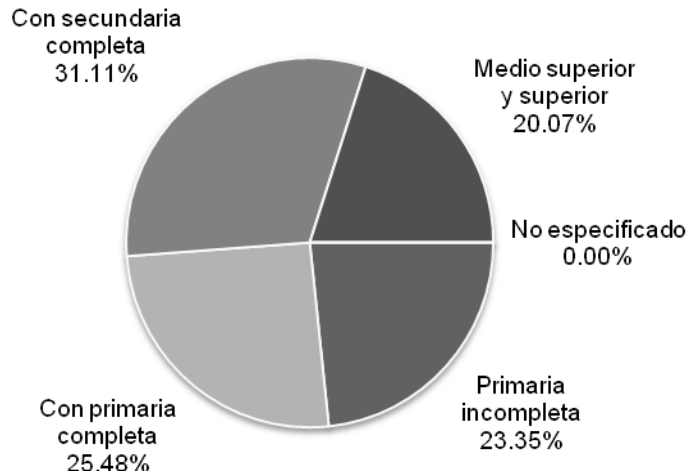


**Figura 4.8** Porcentaje de los subocupados de acuerdo a su nivel de instrucción, en diciembre de 2010. (INEGI, 2012)



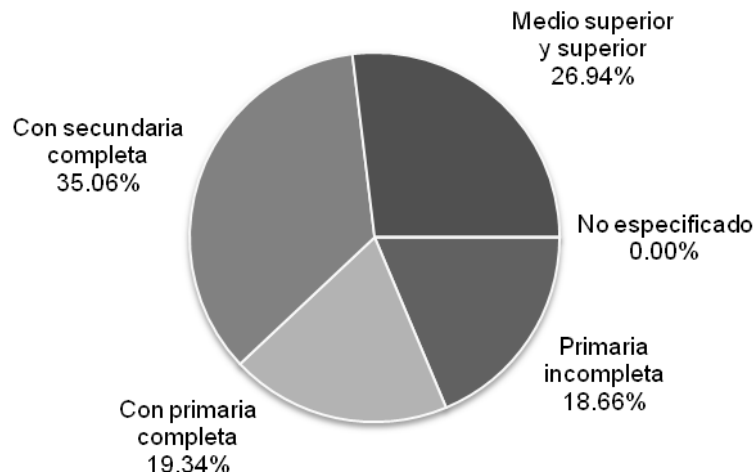
De los hombres subocupados, en diciembre de 2010, 20.07 por ciento contaban con estudios de nivel medio superior y superior, mientras que de la población femenina subocupada fue de 26.94 por ciento.

### HOMBRES SUBOCUPADOS



**Figura 4.9** Porcentaje de los hombres subocupados de acuerdo a su nivel de instrucción, en diciembre de 2010. (INEGI, 2012)

### MUJERES SUBOCUPADAS



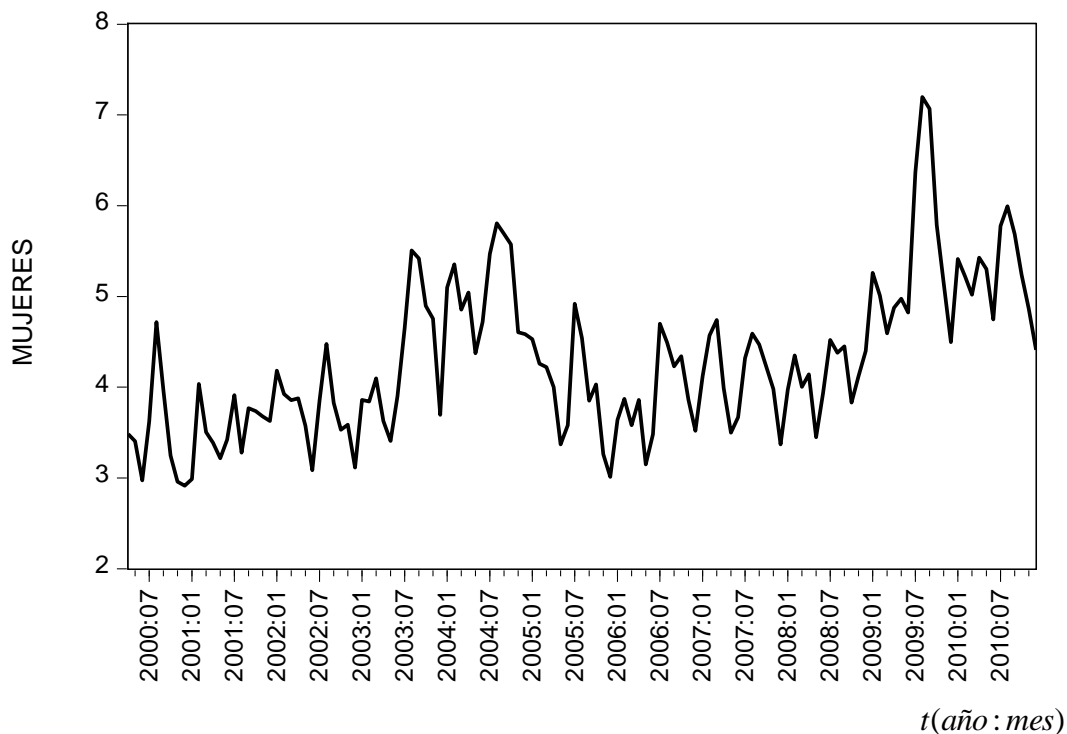
**Figura 4.10** Porcentaje de las mujeres subocupadas de acuerdo a su nivel de instrucción, en diciembre de 2010. (INEGI, 2012)

De todo lo anterior, se observa que la situación de las mujeres es diferente a la de los hombres en cuanto al empleo, subocupación y desempleo. Y aunque últimamente las diferencias entre las tasas de desocupación de hombres y mujeres han sido menores,

al igual que en la tasa de condiciones críticas de ocupación (TCCO), en la tasa de ocupación parcial y desocupación (TOPD1) son significativas las diferencias, así para diciembre de 2010 se tuvo un valor de 14.4 para mujeres contra 9.28 para hombres, lo que sugiere que no es que la situación de las mujeres haya mejorado en el tema que se está tratando, sino que se ha adaptado a lo que el mercado le ofrece y a sus actividades en el hogar. Ante esta inquietud se procede a analizar la tasa de desocupación para mujeres, con la finalidad de encontrar un modelo que refleje su comportamiento y pronostique los valores futuros para anticiparse en lo que puede pasar con el desempleo para las mujeres en México en un futuro, siempre y cuando continúe la misma tendencia.

## 4.2. Modelo ARIMA para la tasa de desocupación femenina

### 4.2.1 Identificación

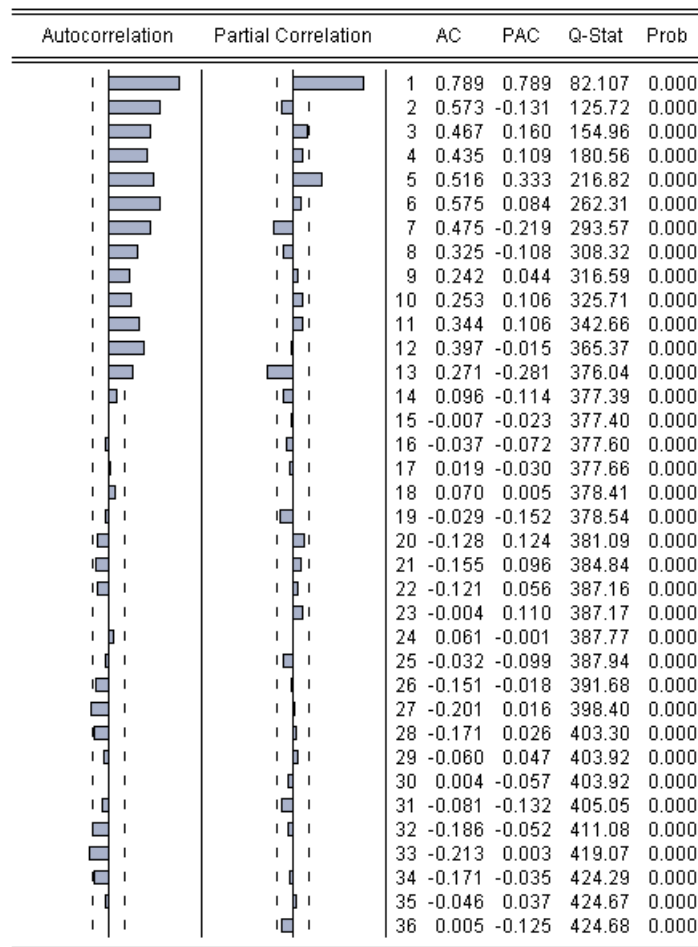


**Figura 4.11** Serie de la tasa de desocupación mensual para mujeres (*MUJERES*) de abril de 2000 a diciembre de 2010. (INEGI, 2012)

### Análisis exploratorio

Como primer paso del análisis se grafica la serie de tiempo de interés, así pues, se presenta la figura 4.11 de la serie de la tasa de desocupación para mujeres, señalada en las diferentes gráficas como *MUJERES* .

Para el análisis, la estimación de los modelos ARIMA y los pronósticos se utilizan las tasas de desocupación femenina desde abril de 2000 a diciembre de 2010, y los datos de los años 2011 y 2012 para evaluar la capacidad predictiva de los modelos.



**Figura 4.12** Autocorrelación de la serie de la tasa de desocupación mensual para mujeres (*MUJERES* ), generada con Eviews. (INEGI, 2012)

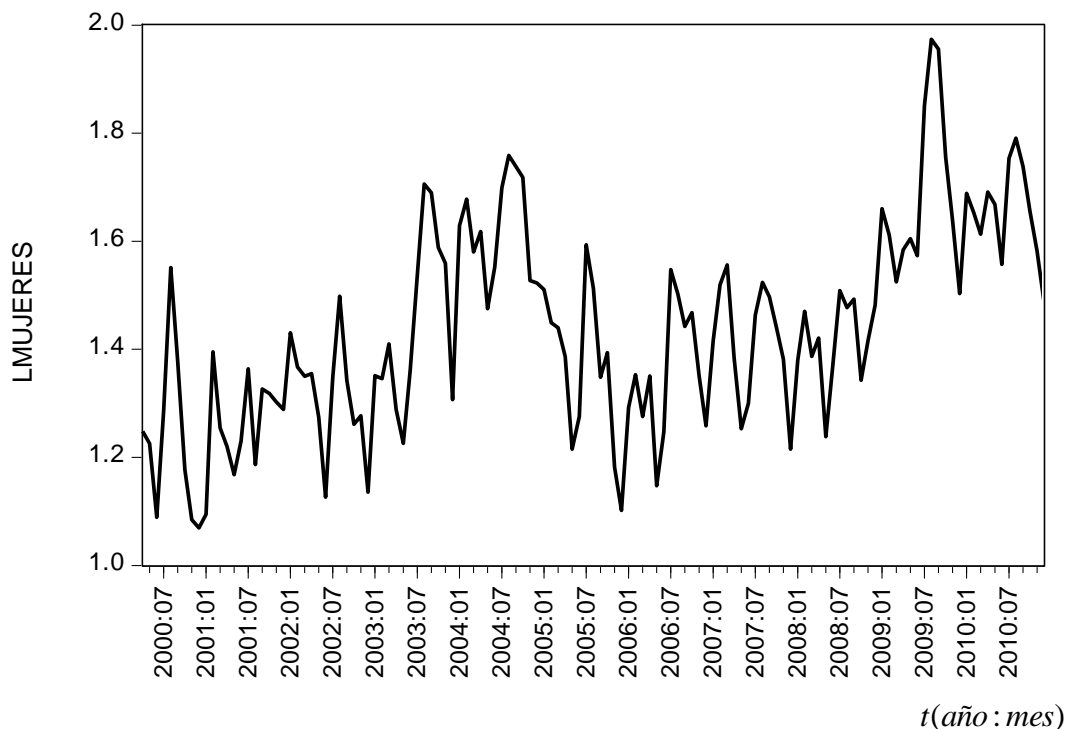
La serie *MUJERES* muestra una evolución creciente en la serie, con el paso del tiempo, especialmente a partir del año 2009. Esto significa que la serie no presenta un valor medio constante en todo el periodo muestral, es decir, no oscila en torno al mismo

valor. Por lo tanto, se puede suponer ya a priori que, probablemente no será estacionaria.

En la figura 4.12 se puede apreciar que la función de autocorrelación decrece exponencialmente y de forma lenta, mientras que la función de autocorrelación parcial presenta un valor significativo en el retardo uno, con un coeficiente de autocorrelación de 0.789. Este gráfico puede considerarse como indicativo de la no estacionariedad de la serie.

### Estudio de la estacionariedad

En la figura 4.11 de la serie *MUJERES* se observa cierto comportamiento no homogéneo en la varianza, por lo que se decidió trabajar con la transformación logarítmica que se ha denominado *LMUJERES*, figura 4.13.



**Figura 4.13** Serie de la transformación logarítmica de la tasa de desocupación mensual para mujeres (*LMUJERES*) de abril de 2000 a diciembre de 2010, generado con Eviews. (INEGI, 2012)

Null Hypothesis: MUJERES has a unit root		
Exogenous: Constant		
Lag Length: 4 (Automatic - based on SIC, maxlag=12)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.668481	0.4447
Test critical values:		
1% level	-3.483751	
5% level	-2.884856	
10% level	-2.579282	
*MacKinnon (1996) one-sided p-values.		

Null Hypothesis: LMUJERES has a unit root		
Exogenous: Constant		
Lag Length: 4 (Automatic - based on SIC, maxlag=12)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.586523	0.4864
Test critical values:		
1% level	-3.483751	
5% level	-2.884856	
10% level	-2.579282	
*MacKinnon (1996) one-sided p-values.		

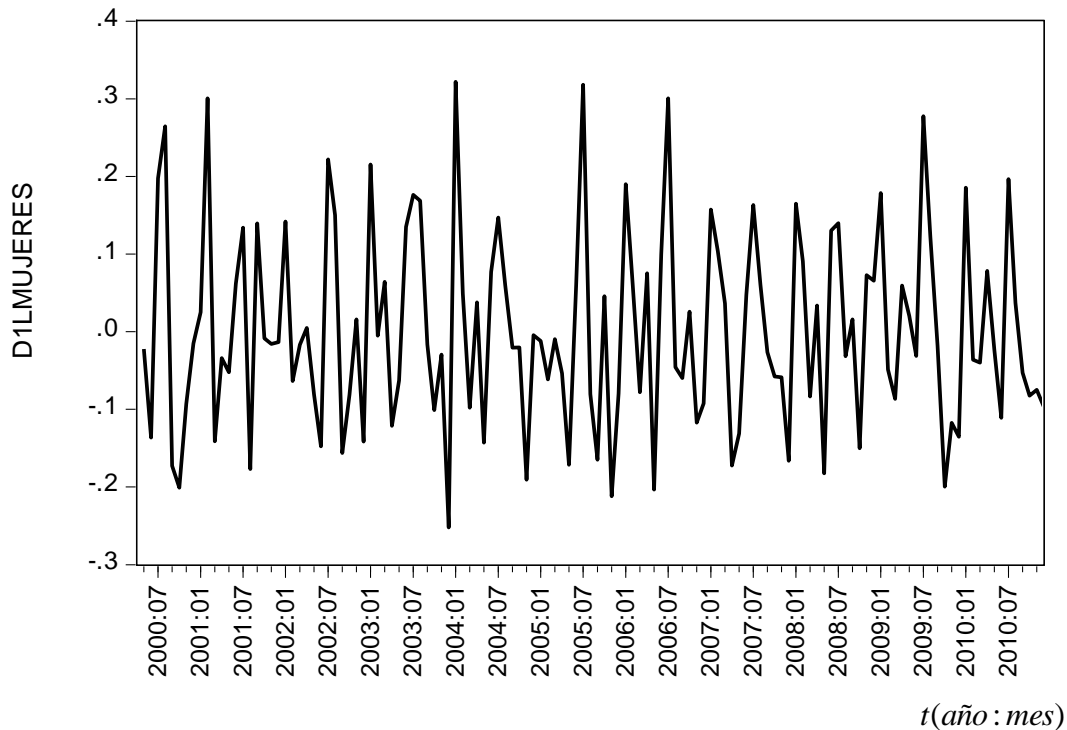
Null Hypothesis: D1LMUJERES has a unit root		
Exogenous: Constant		
Lag Length: 3 (Automatic - based on SIC, maxlag=12)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-10.97499	0.0000
Test critical values:		
1% level	-3.483751	
5% level	-2.884856	
10% level	-2.579282	
*MacKinnon (1996) one-sided p-values.		

**Figura 4.14** Resultados del estadístico de Dickey-Fuller aumentado (ADF) para las series *MUJERES*, *LMUJERES* y *D1LMUJERES*, generados con Eviews. (INEGI, 2012)

En la figura 4.14 se muestran los resultados del estadístico de Dickey-Fuller aumentado (ADF), los cuales permiten identificar la presencia de raíces unitarias. Como se puede apreciar, en los casos de los niveles de *MUJERES* y el logaritmo de *MUJERES* (*LMUJERES*) no se puede rechazar la hipótesis nula ( $H_0$ ) de que la serie es no estacionaria. Al tomar las primeras diferencias del logaritmo de *MUJERES* (*D1LMUJERES*), si se puede rechazar la  $H_0$ , lo cual indica que estas son

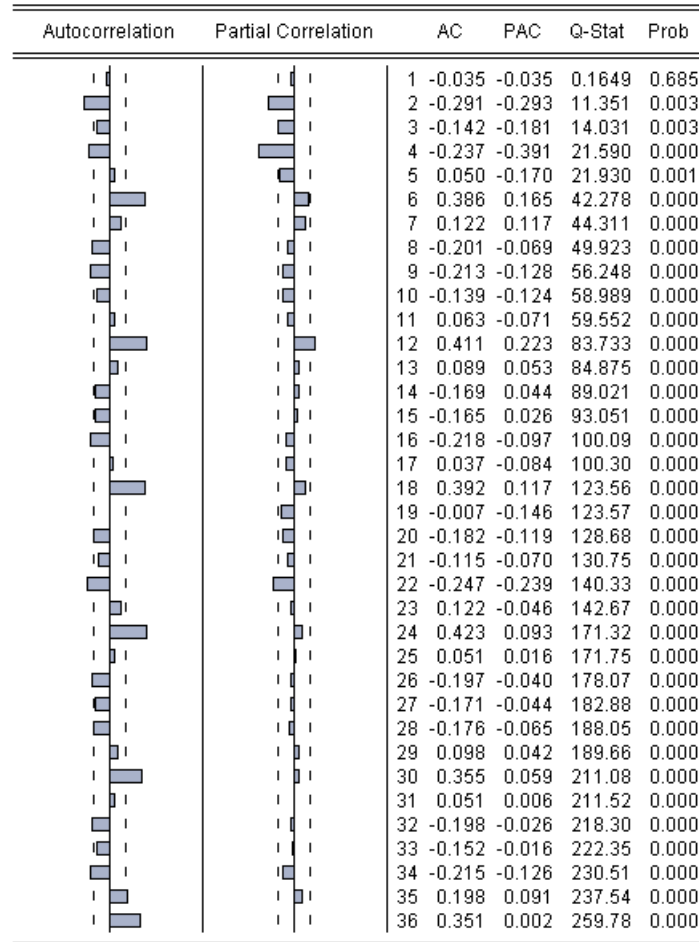
efectivamente estacionarias, confirmando que la serie en primera diferencia de *MUJERES* con transformación logarítmica es integrada de orden  $I(1)$  o  $d = 1$ , por lo que se decidió utilizar las primeras diferencias de los logaritmos y no de la serie original *MUJERES*.

En la figura 4.15 de la serie  $D1LMUJERES = (1 - B)LMUJERES$  se puede observar que es estacionaria en medias, es decir no hay presencia de tendencia.



**Figura 4.15** Serie de las primeras diferencias de la transformación logarítmica de la tasa de desocupación mensual para mujeres (*D1LMUJERES*), generada con Eviews. (INEGI, 2012)

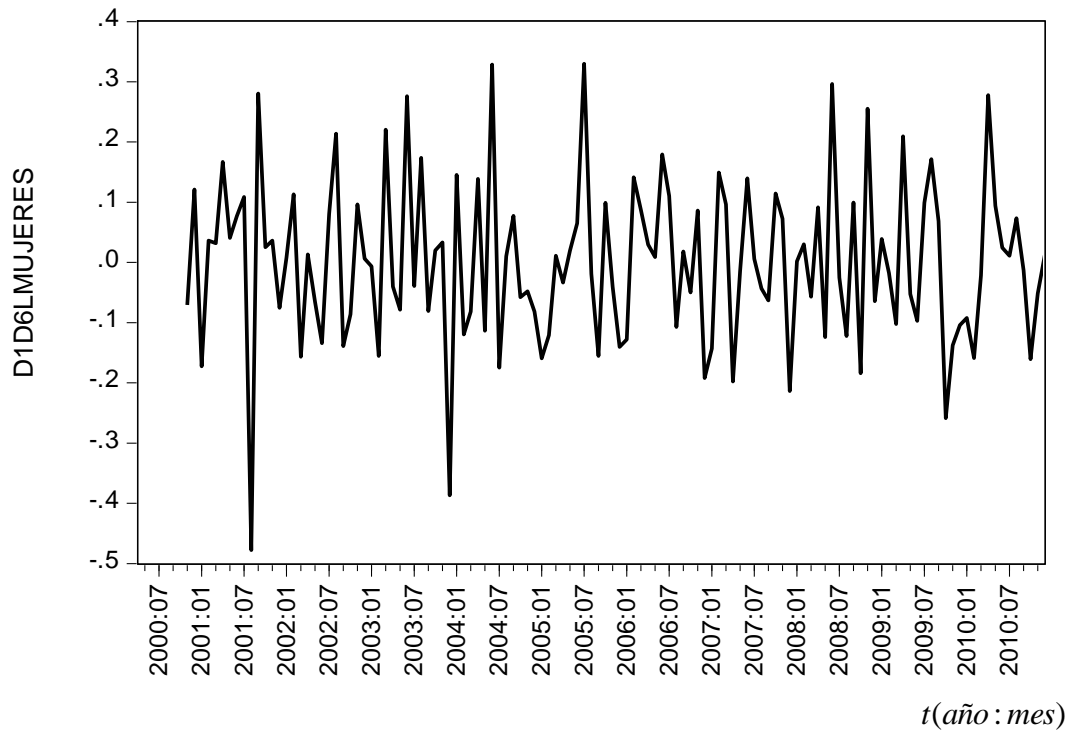
En la figura 4.16 se incluyen los valores de los treinta y seis primeros coeficientes de autocorrelación y autocorrelación parcial, así como su representación gráfica, correspondiente a la serie *D1LMUJERES*. Se aprecia en el gráfico de la función de autocorrelación, un coeficiente sistemáticamente significativo para los retardos estacionales 6, 12, 18, 24, 30 y 36, indicando la existencia de tendencia en la parte estacional, que habrá que corregir para proceder a la correcta identificación de los procesos *ARIMA* presentes en esta serie.



**Figura 4.16** Autocorrelación de la serie de las primeras diferencias de la transformación logarítmica de la tasa de desocupación mensual para mujeres (*D1LMUJERES*), generada con Eviews. (INEGI, 2012)

Por lo anterior es necesario extraer el comportamiento estacional con una diferencia  $D=1$  en la parte estacional, entonces la transformación será de la siguiente manera  $(1-B)(1-B^6)LMUJERES = D1D6LMUJERES$  (figura 4.17), el resultado de la transformación es confirmado por el gráfico de la autocorrelación (figura 4.18). En esta gráfica, se observa que el efecto sistemático en los periodos de estacionalidad han sido extraídos; por lo tanto se puede decir que se ha corregido la existencia de tendencia en la parte estacional, pues los retardos estacionales 12, 18, 24, 30 y 36 han sido corregidos.

En la figura 4.17 de *D1D6LMUJERES* se destaca un valor anormalmente alto para el periodo 2001:08 (agosto de 2001) y otro como el de 2003:12 (diciembre de 2003) más leve que sería importante darle algún tipo de tratamiento más adelante.



**Figura 4.17** Serie de la transformación  $(1 - B)(1 - B^6)LMUJERES = D1D6LMUJERES$ , generada con Eviews. (INEGI, 2012)

**Tabla 4.3** Resultados de las series de diferentes transformaciones de la serie  $LMUJERES$ , generados con Eviews.

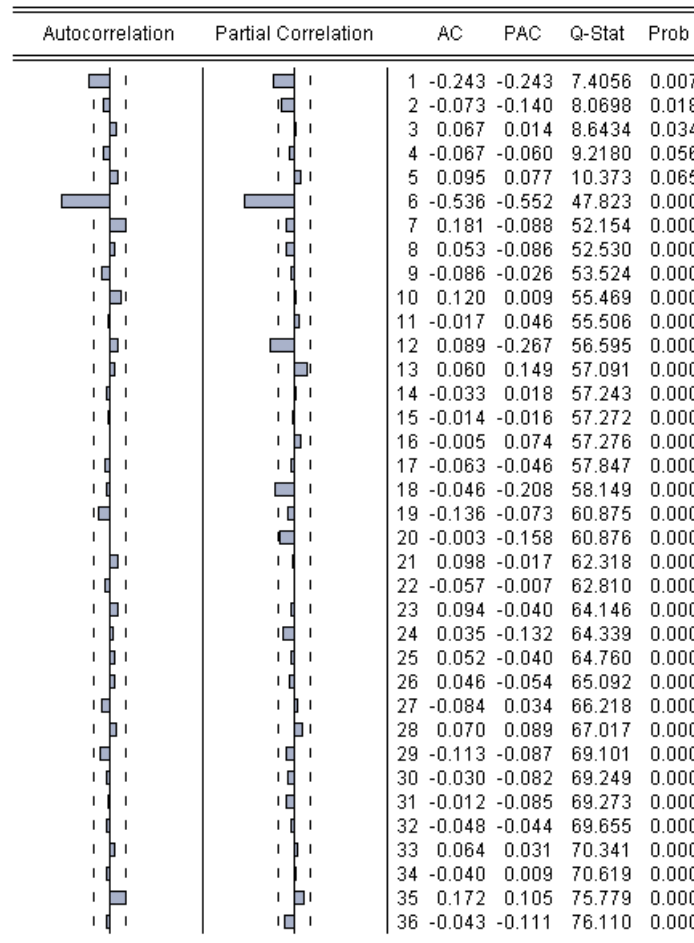
	D1LMUJERES	D2LMUJERES	D1D6LMUJERES	D2D6LMUJERES
<b>Mean</b>	0.001845	-0.000589	-0.000021	-0.009313
<b>Median</b>	-0.016629	-0.007780	0.005957	0.031712
<b>Maximum</b>	0.321868	0.573874	0.330014	1.339398
<b>Minimum</b>	-0.252006	-0.441713	-0.477515	-1.370997
<b>Std. Dev.</b>	0.127888	0.184545	0.137460	0.534763
<b>Skewness</b>	0.453790	0.217745	-0.068626	-0.105177
<b>Kurtosis</b>	2.668086	3.036939	3.780429	2.946268

Fuente: Con datos del BIE, INEGI, 2012.

En la tabla 4.3, se observan diferentes transformaciones que convierten a  $LMUJERES$  en estacionaria. Los criterios de desviación estándar mínima y de correlograma más



simple llevan a elegir a la transformación  $D1D6LMUJERES = (1 - B)(1 - B^6)LMUJERES$  frente a las transformaciones  $D2D6LMUJERES = (1 - B)^2(1 - B^6)LMUJERES$ ,  $D1LMUJERES = (1 - B)LMUJERES$ , y  $D2LMUJERES = (1 - B)^2 LMUJERES$ , como la mejor transformación para  $LMUJERES$ .



**Figura 4.18** Autocorrelación de la serie de la transformación  $(1 - B)(1 - B^6)LMUJERES = D1D6LMUJERES$ , generada con Eviews. (INEGI, 2012)

### Identificación de posibles modelos ARIMA

Se ha comprobado que la serie  $LMUJERES$  presenta tendencia creciente ( $d = 1$ ). Además, al presentar estacionalidad de periodo  $E = 6$ , ha sido necesario aplicar una diferencia en la parte estacional para eliminarla ( $D = 1$ ). En consecuencia, el modelo que se ajustará será un  $ARIMA(p,1,q)(P,1,Q)_6$  en logaritmos.

La identificación de los parámetros  $p$  y  $q$ , órdenes de los polinomios autorregresivos y de medias móviles de la parte regular del modelo, se realizará a partir de las funciones de autocorrelación simple y parcial de la serie diferenciada regular y estacionalmente, las cuales aparecen en la figura 4.18. En ésta se observa que el primer coeficiente de la autocorrelación (AC) es no nulo de -0.243 y que decrece al aumentar los retardos. Por otro lado, en la autocorrelación parcial (PAC) el primer retardo está fuera de los límites de la banda de confianza, con valor de -0.243, mientras que el segundo retardo está muy cercano a la banda. Una primera tentativa sería suponer que la serie *LMUJERES* diferenciada regular y estacionalmente presenta, en la parte regular, la estructura de un modelo de medias móviles de orden  $q$  igual a 1. Sin embargo, el hecho de que el coeficiente del segundo retardo de la PAC esté muy cercano a la banda, puede hacer pensar en la posibilidad de un modelo autorregresivo de medias móviles de órdenes  $p$  y  $q$  distintos de cero. Se supondrá entonces que la estructura es la de este segundo modelo y, en particular, que  $p = q = 1$ . Si, al ajustarlo, el coeficiente asociado al parámetro autorregresivo no fuera significativamente distinto de cero se recurriría al modelo de medias móviles de orden  $q = 1$ .

La identificación de los parámetros  $P$  y  $Q$ , órdenes de autorregresión y de medias móviles de la parte estacional del modelo, se realizará a partir de las funciones de autocorrelación simple y parcial para la serie diferenciada regular y estacionalmente, es decir también en la figura 4.18, sobre los tres primeros retardos estacionales, de la serie *D1D6LMUJERES*, tanto en la parte de la autocorrelación simple como en la autocorrelación parcial.

Así se observa que el primer coeficiente estacional de la autocorrelación (AC) es no nulo de -0.536 y que decrece al aumentar los retardos. Por otro lado, en la autocorrelación parcial (PAC) aunque los tres primeros retardos estacionales están fuera de los límites de la banda de confianza, el único significativamente grande es el primero con -0.552. Por lo que una primera tentativa sería suponer que la serie *LMUJERES* diferenciada regular y estacionalmente presenta, en la parte estacional, la estructura de un modelo de medias móviles de orden  $q$  igual a 1. Sin embargo, el

hecho de que los coeficientes del segundo y tercer retardos estacionales de la PAC no puedan ser considerados nulos hace pensar en la posibilidad de un modelo autorregresivo de medias móviles para la parte estacional de órdenes  $P$  y  $Q$  distintos de cero.

Se procederá a ajustar un modelo considerando la parte regular, una vez ajustada, se analizará las AC y las PAC de los residuos en busca de alguna estructura en la parte estacional.

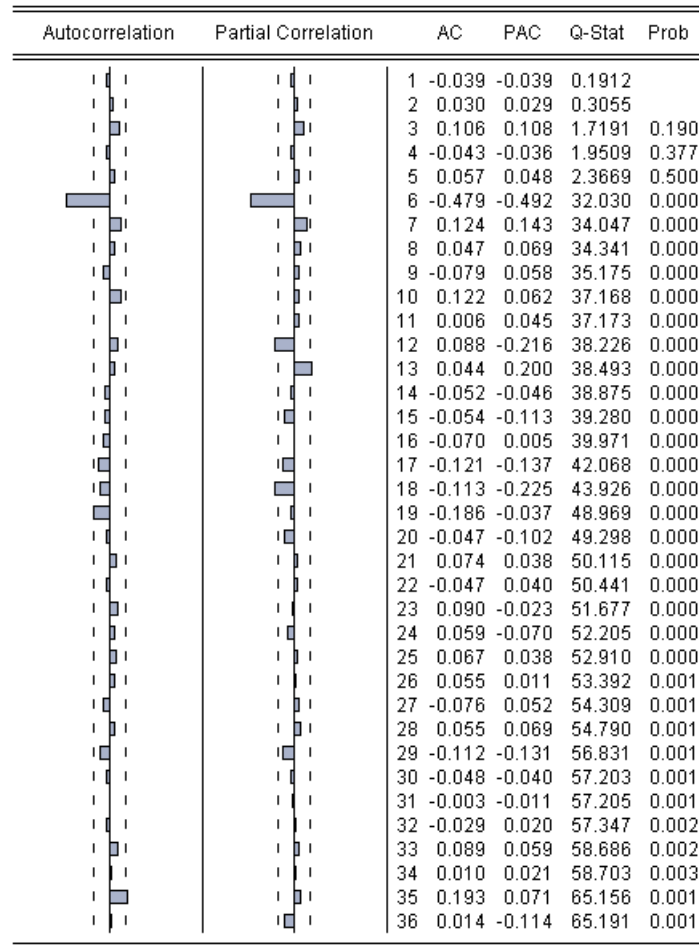
### 4.2.2 Estimación de los posibles modelos ARIMA

Dependent Variable: D(LOG(MUJERES),1,6)  
 Method: Least Squares  
 Date: 02/03/12 Time: 14:28  
 Sample (adjusted): 2000M12 2010M12  
 Included observations: 121 after adjustments  
 Convergence achieved after 11 iterations  
 MA Backcast: 2000M11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.613487	0.074983	8.181689	0.0000
MA(1)	-0.991317	0.011489	-86.28459	0.0000
R-squared	0.188621	Mean dependent var		0.000562
Adjusted R-squared	0.181803	S.D. dependent var		0.137880
S.E. of regression	0.124718	Akaike info criterion		-1.309131
Sum squared resid	1.850997	Schwarz criterion		-1.262920
Log likelihood	81.20242	Hannan-Quinn criter.		-1.290363
Durbin-Watson stat	2.069457			
Inverted AR Roots	.61			
Inverted MA Roots	.99			

**Figura 4.19** Ajuste del modelo  $ARIMA(1,1,1)(0,1,0)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

Una vez identificados los posibles órdenes se empezará a estimar el modelo  $ARIMA(1,1,1)(0,1,0)_6$  en logaritmos. La figura 4.19 proporciona la estimación de los parámetros del modelo, en la que el parámetro tanto de la componente autorregresivo como el de medias móviles es significativo.



**Figura 4.20** Autocorrelación de los residuos del modelo  $ARIMA(1,1,1)(0,1,0)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

Al analizar las autocorrelaciones AC y la PAC de la serie residual del modelo  $ARIMA(1,1,1)(0,1,0)_6$  en logaritmos sobre los primeros retardos estacionales en las autocorrelaciones simple y parcial, proporcionadas en la figura 4.20, se observa que, tanto en la parte AC como en la PAC, únicamente el primer coeficiente está un tanto fuera de los límites de la banda de confianza, siendo más significativo en el caso de la autocorrelación parcial (PAC), comportamiento similar al que se había observado en el correlograma de  $D1D6LMUJERES$ , por lo que se procederá a ajustar un modelo autorregresivo de medias móviles para la parte estacional de órdenes  $P$  y  $Q$  distintos de cero.

Dependent Variable: D(LOG(MUJERES),1,6)  
 Method: Least Squares  
 Date: 02/03/12 Time: 15:03  
 Sample (adjusted): 2001M06 2010M12  
 Included observations: 115 after adjustments  
 Convergence achieved after 12 iterations  
 MA Backcast: 2000M11 2001M05

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.246866	0.308547	0.800094	0.4254
SAR(6)	-0.211062	0.090224	-2.339321	0.0211
MA(1)	-0.511042	0.274402	-1.862382	0.0652
SMA(6)	-0.960030	0.020324	-47.23640	0.0000
R-squared	0.639466	Mean dependent var		-0.001363
Adjusted R-squared	0.629722	S.D. dependent var		0.139070
S.E. of regression	0.084625	Akaike info criterion		-2.067010
Sum squared resid	0.794915	Schwarz criterion		-1.971534
Log likelihood	122.8531	Hannan-Quinn criter.		-2.028257
Durbin-Watson stat	1.966790			
Inverted AR Roots	.67+.39i	.67-.39i	.25	.00-.77i
	-.00+.77i	-.67+.39i	-.67-.39i	
Inverted MA Roots	.99	.51	.50+.86i	.50-.86i
	-.50+.86i	-.50-.86i	-.99	

**Figura 4.21** Ajuste del modelo  $ARIMA(1,1,1)(1,1)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

La figura 4.21 proporciona la estimación de los parámetros del modelo  $ARIMA(1,1,1)(1,1)_6$  en logaritmos, en la que el parámetro de la componente autorregresivo regular no es significativo, entonces el modelo se reduce a un modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos.

Se procede a estimar el modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos, cuyos parámetros se presentan enseguida en la figura 4.22. En ella se observa que todos los parámetros del modelo son significativos.

Dependent Variable: D(LOG(MUJERES),1,6)  
 Method: Least Squares  
 Date: 02/03/12 Time: 15:09  
 Sample (adjusted): 2001M05 2010M12  
 Included observations: 116 after adjustments  
 Convergence achieved after 16 iterations  
 MA Backcast: 2000M10 2001M04

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(6)	-0.215190	0.088855	-2.421801	0.0170
MA(1)	-0.287354	0.090858	-3.162663	0.0020
SMA(6)	-0.960695	0.020590	-46.65794	0.0000
R-squared	0.637390	Mean dependent var		-0.000999
Adjusted R-squared	0.630973	S.D. dependent var		0.138520
S.E. of regression	0.084147	Akaike info criterion		-2.086969
Sum squared resid	0.800130	Schwarz criterion		-2.015756
Log likelihood	124.0442	Hannan-Quinn criter.		-2.058061
Durbin-Watson stat	1.944924			
Inverted AR Roots	.67+.39i	.67-.39i	.00-.77i	-.00+.77i
	-.67+.39i	-.67-.39i		
Inverted MA Roots	.99	.50+.86i	.50-.86i	.29
	-.50+.86i	-.50-.86i	-.99	

**Figura 4.22** Ajuste del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

Al analizar las autocorrelaciones AC y la PAC de la serie residual del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos sobre los primeros retardos estacionales en las autocorrelaciones simple y parcial, proporcionadas en la figura 4.23, se observa que, tanto en la parte AC como en la PAC, ningún coeficiente está fuera de los límites de la banda de confianza. Por lo tanto los residuos a través del análisis de la función de autocorrelación simple y parcial tienen un comportamiento de ruido blanco (puramente aleatorios, comprobando que todos cumplen con la hipótesis nula de la no presencia de autocorrelación). Entonces este podría ser el posible modelo.

El hecho de que el modelo identificado sea adecuado no debe hacer descartar la posibilidad de que otro modelo algo más complejo pueda ajustarse mejor a la serie observada. Entonces se procede a realizar la técnica del sobre ajuste:

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.020	0.020	0.0461	
		2	-0.100	-0.100	1.2420	
		3	0.077	0.082	1.9610	
		4	-0.088	-0.104	2.9043	0.088
		5	-0.098	-0.077	4.0848	0.130
		6	0.041	0.021	4.2952	0.231
		7	0.041	0.037	4.5111	0.341
		8	0.073	0.085	5.1910	0.393
		9	-0.046	-0.067	5.4666	0.486
		10	0.072	0.088	6.1363	0.524
		11	-0.062	-0.083	6.6382	0.576
		12	0.003	0.058	6.6397	0.675
		13	-0.008	-0.040	6.6474	0.758
		14	0.003	0.023	6.6487	0.827
		15	-0.007	-0.020	6.6558	0.879
		16	-0.020	-0.032	6.7113	0.916
		17	-0.126	-0.123	8.8916	0.838
		18	-0.138	-0.161	11.556	0.712
		19	-0.190	-0.201	16.675	0.407
		20	-0.038	-0.086	16.879	0.463
		21	0.082	0.054	17.856	0.465
		22	0.044	-0.000	18.142	0.513
		23	0.045	0.039	18.445	0.558
		24	-0.033	-0.077	18.603	0.611
		25	-0.062	-0.006	19.174	0.635
		26	0.070	0.109	19.919	0.647
		27	-0.040	0.011	20.167	0.687
		28	0.002	0.029	20.167	0.738
		29	0.106	0.084	21.944	0.692
		30	-0.125	-0.138	24.421	0.607
		31	-0.003	0.015	24.423	0.659
		32	0.050	-0.006	24.833	0.687
		33	-0.014	-0.003	24.866	0.732
		34	0.025	-0.014	24.971	0.769
		35	0.206	0.162	32.132	0.460
		36	0.005	-0.096	32.136	0.510

**Figura 4.23** Autocorrelación de los residuos del modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

De los sobre ajustes al modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos, en la tabla 4.4 solo aparecen aquellos que sus parámetros al ser estimados son significativos, dejando de lado aquellos que no lo son.

Como se puede observar en la tabla 4.4 y la tabla 4.5, los cuatro modelos que se presentan como posibles seleccionados poseen coeficientes significativamente diferentes a cero, y cumplen con la estacionariedad e invertibilidad. Además se puede observar que el modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos posee una menor suma de residuos al cuadrado y un mayor  $R^2$  ajustado, por lo que este modelo presenta un mejor grado de ajuste. También se puede analizar el estadístico de Durbin-Watson, los cuatro modelos ARIMA están indicando que sus residuos tienen un comportamiento de ruido

blanco y por lo tanto no presentan autocorrelación de ningún orden, debido a que el valor ideal del estadístico Durbin-Watson es 2, sin embargo este estadístico solo se puede utilizar de forma aproximada, es más conveniente realizar una identificación de los residuos a través del estadístico de Q de Ljung-Box.

**Tabla 4.4** Sobre ajustes al modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generados con Eviews.

Criterios de Validez	MODELOS			
	$ARIMA(1,1,1)(0,1,0)_6$	$ARIMA(0,1,1)(1,1,0)_6$	$ARIMA(0,1,1)(0,1,1)_6$	$ARIMA(0,1,1)(1,1,1)_6$
Adjusted R-squared	0.181803	0.353767	0.586849	0.630973
Sum squared resid	1.850997	1.413572	0.936787	0.800130
Log likelihood	81.20242	91.03641	123.9181	124.0442
Akaike info criterion (AIC)	-1.309131	-1.535110	-1.998657	-2.086969
Schwarz criterion(SBC)	-1.262920	-1.487635	-1.952689	-2.015756
Durbin-Watson stat	2.069457	1.908488	1.983859	1.944924
Q-7 (p-value)	34.047 (0.000)	9.2290 (0.100)	7.7071 (0.173)	4.5111 (0.341)
Q-14 (p-value)	38.875 (0.000)	24.169 (0.019)	9.4918 (0.660)	6.6487 (0.827)
Q-21 (p-value)	50.115 (0.000)	33.422 (0.021)	21.041 (0.335)	17.856 (0.465)

Fuente: Con datos del BIE, INEGI, 2012.

En lo que respecta a los indicadores de parsimonia, el AIC y SBC favorecen también al modelo  $ARIMA(0,1,1)(1,1)_6$ . Finalmente, en el análisis del estadístico desarrollado por Ljung-Box en el que presenta valores menores para este modelo, se puede decir que no hay dependencia entre el residuo y su valor en el periodo anterior. De esta manera los valores de Q(7), Q(14) y Q(21) no rechazan la hipótesis nula de la no presencia de autocorrelación. Por lo tanto se seguirá el análisis con este modelo y se usará en la etapa de predicción.



**Tabla 4.5** Parámetros de los modelos en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generados con Eviews.

Dependent Variable: D(LOG(MUJERES),1,6)				
Method: Least Squares				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.613487	0.074983	8.181689	0.0000
MA(1)	-0.991317	0.011489	-86.28459	0.0000
Inverted AR Roots	0.61			
Inverted MA Roots	0.99			
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SAR(6)	-0.550329	0.077844	-7.069628	0.0000
MA(1)	-0.348453	0.088424	-3.940719	0.0001
Inverted AR Roots	0.78+0.45i	0.78-0.45i	0.00-0.91i	-0.00+0.91i
	-0.78+0.45i	-0.78-0.45i		
Inverted MA Roots	0.35			
Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.314176	0.085781	-3.662524	0.0004
SMA(6)	-0.963438	0.023393	-41.18473	0.0000
Inverted MA Roots	0.99	0.50+0.86i	0.50-0.86i	0.31
	-0.50+0.86i	-0.50-0.86i	-0.99	
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SAR(6)	-0.215190	0.088855	-2.421801	0.0170
MA(1)	-0.287354	0.090858	-3.162663	0.0020
SMA(6)	-0.960695	0.020590	-46.65794	0.0000
Inverted AR Roots	0.67+0.39i	0.67-0.39i	0.00-0.77i	-0.00+0.77i
	-0.067+0.39i	-0.67-0.39i		
Inverted MA Roots	0.99	0.50+0.86i	0.50-0.86i	0.29
	-0.50+0.86i	-0.50-0.86i	-0.99	

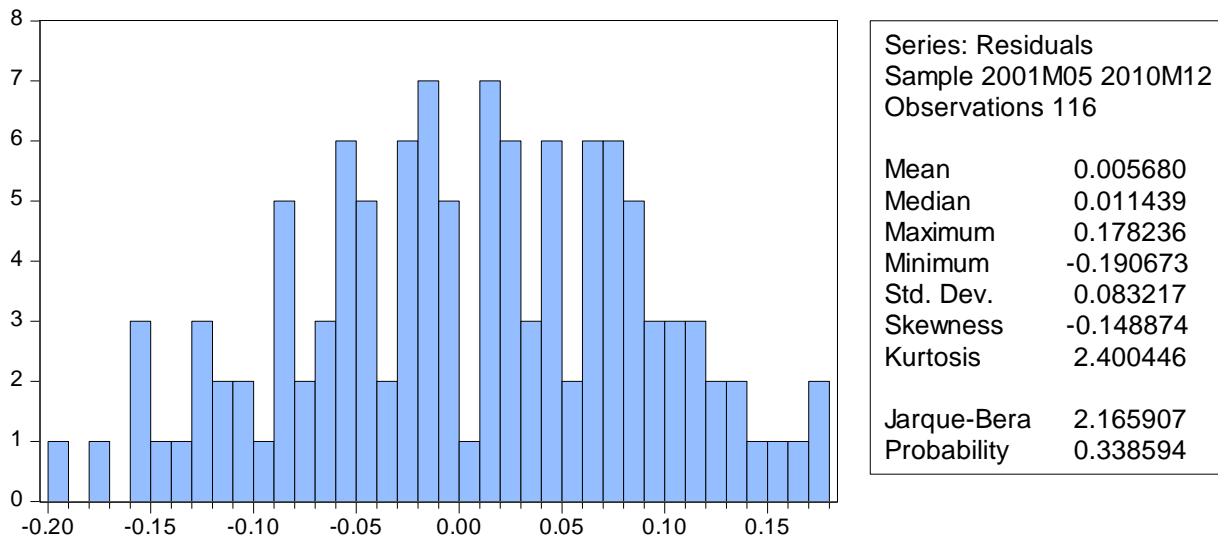
Fuente: Con datos del BIE, INEGI, 2012.

### 4.2.3 Diagnóstico al modelo elegido $ARIMA(0,1,1)(1,1,1)_6$ en logaritmos

#### • Incorrelación

En el correlograma del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos de la figura 4.23, el estadístico de Q de Ljung-Box determina que hay ausencia de autocorrelación en los residuos, es decir el comportamiento se asemeja al de un ruido blanco. Se observa también que todos los coeficientes caen dentro de la banda de confianza o en casos extremos están en el límite de la banda al 95% de confianza, además todos los p-valores asociados al estadístico de Ljung-Box para cada retardo (p-value) son lo suficientemente grandes como para no rechazar la hipótesis nula de que, en cada caso, todos los coeficientes anteriores, hasta el correspondiente, son nulos. En consecuencia se puede concluir que los residuos del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos están incorrelacionados.

#### • Normalidad

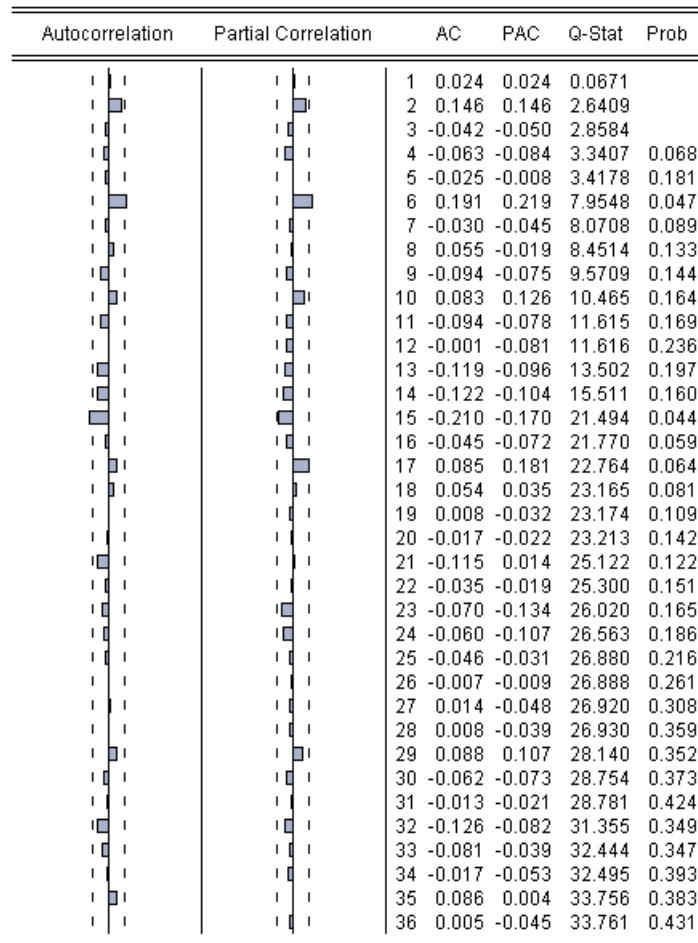


**Figura 4.24** Histograma de los residuos del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

El estadístico de Jarque-Bera se utiliza para determinar si la serie de residuos sigue una distribución normal; en la figura 4.24 se muestra el resultado del histograma residual, en este caso el valor de la probabilidad conduce a no rechazar la hipótesis nula de una distribución normal, como el valor obtenido del estadístico del Jarque-Bera

es menor al valor de referencia de tablas (aproximadamente, 6) y la probabilidad es mayor a  $\alpha = 5\%$ , se puede suponer que la serie residual presenta características que se asemejan a una normal.

• **Heteroscedasticidad residual**



**Figura 4.25** Correlograma de los residuos al cuadrado del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

Una variante al estadístico Q de Ljung-Box, es el test de los residuos al cuadrado, que se utiliza, principalmente, para analizar la presencia de una heterocedasticidad condicional autorregresiva (ARCH) en los residuos. En la figura 4.25 se presenta el resultado del correlograma de los residuos al cuadrado del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos, al analizar el estadístico Q, se puede aceptar la hipótesis nula que establece que los residuos no siguen un comportamiento ARCH, también se observa

que los residuos al cuadrado son incorrelacionados. Por lo tanto no hay rezagos iniciales significativos que recoger y no se tiene un comportamiento ARCH en los residuos de este modelo.

#### 4.2.4 Análisis de intervención

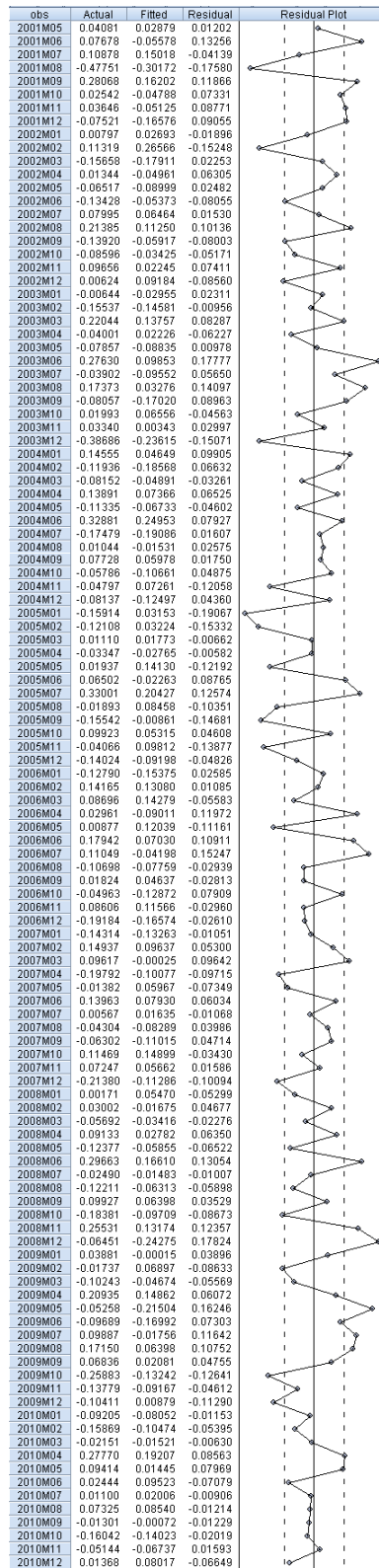
En la figura 4.23 del modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos, se comprobó que los residuos siguen un comportamiento de ruido blanco; es decir, los residuos son incorrelacionados.

En la figura 4.26 los residuos del modelo en análisis presentan movimientos bruscos que no será posible ser captados por una dependencia sistemática de su pasado, con respecto a los residuos del modelo  $ARIMA(0,1,1)(1,1)_6$ . Resulta de especial interés centrarse en los puntos raros o valores atípicos.

En este caso, los valores atípicos del modelo se sitúan en los datos de agosto de 2001, junio de 2003, enero de 2005 y diciembre de 2008, siendo enero de 2005 el mayor de ellos, por una diferencia muy pequeña. En algunos de estos casos, el dato ajustado es superior al real, mientras que para otros es menor, dando lugar a errores cuantificados de -0.17586, 0.17777, -0.19067 y 0.17824, respectivamente.

Es de suma importancia estudiar la causa de estas posibles diferencias, saber exactamente que ocurrió, pues esto podría ayudar a mejorar o incluso a replantear el modelo de predicción. Así, el caso de la mayor diferencia entre el dato ajustado y el real, que se da en enero de 2005, puede deberse a que el INEGI comenzó a aplicar en ese año una nueva metodología para igualarla con los sistemas de la OCDE.

El tratamiento de este problema se hará mediante el análisis de intervención que consiste en introducir variables artificiales impulso. Se crearán cuatro variables artificiales para las cuatro fechas antes señaladas, donde cada una de las variables tomará el valor de uno en el mes para el cual fueron creadas y cero en todos los demás meses:



**Figura 4.26** Residuos del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

D0108	agosto de 2001
D0306	junio de 2003
D0501	enero de 2005
D0812	diciembre de 2008

Estas variables artificiales se introducirán en el modelo  $ARIMA(0,1,1)(1,1)_6$ , tanto de forma conjunta, como de manera individual. De los ajustes sólo se presenta aquel que sus parámetros al ser estimados son significativos, dejando de lado aquellos que no lo son.

Dependent Variable: D(LOG(MUJERES),1,6)  
 Method: Least Squares  
 Date: 02/07/12 Time: 15:23  
 Sample (adjusted): 2001M05 2010M12  
 Included observations: 116 after adjustments  
 Convergence achieved after 34 iterations  
 MA Backcast: 2000M10 2001M04

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D0812	-0.053096	0.025812	-2.056988	0.0420
AR(6)	-0.219204	0.083885	-2.613159	0.0102
MA(1)	-0.251299	0.093197	-2.696422	0.0081
SMA(6)	-0.982076	0.020738	-47.35625	0.0000

R-squared	0.642655	Mean dependent var	-0.000999
Adjusted R-squared	0.633084	S.D. dependent var	0.138520
S.E. of regression	0.083906	Akaike info criterion	-2.084354
Sum squared resid	0.788513	Schwarz criterion	-1.989403
Log likelihood	124.8925	Hannan-Quinn criter.	-2.045809
Durbin-Watson stat	1.951340		

Inverted AR Roots	.67+.39i	.67-.39i	.00-.78i	-.00+.78i
	-.67+.39i	-.67-.39i		
Inverted MA Roots	1.00	.50+.86i	.50-.86i	.25
	-.50+.86i	-.50-.86i	-1.00	

**Figura 4.27** Ajuste del modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos con análisis de intervención con variable artificial de impulso D0812, de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

Los resultados de este ajuste se pueden observar en la figura 4.27, donde la variable de intervención impulso D0812 es significativa con parámetros D0812 igual a -0.053096, que puede ser interpretada como en una reducción de la tasa de desempleo para

mujeres en el nivel de ese mes del 5.3096%, siendo la disminución únicamente en ese mes, sin afectar a los meses sucesivos. Hasta el momento se puede afirmar que el modelo estimado es adecuado, en el sentido de que todos sus coeficientes son significativamente distintos de cero.

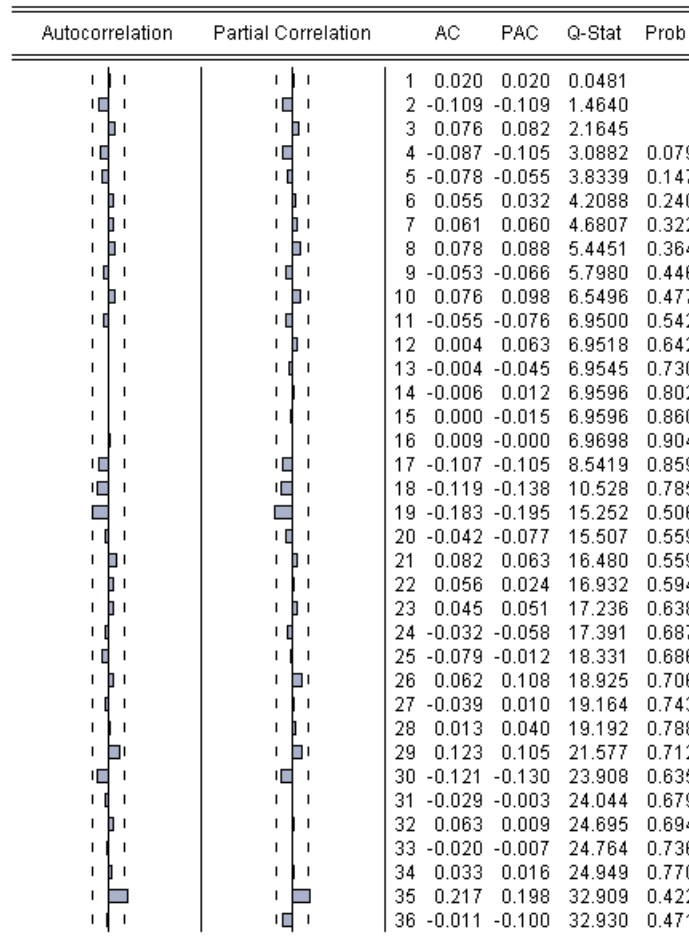
**Tabla 4.6** Resultados del modelo  $ARIMA(0,1,1) (1,1,1)_6$  en logaritmos con y sin intervención, de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generados con Eviews.

Criterios de Validez	MODELOS	
	$ARIMA(0,1,1) (1,1,1)_6$ con intervención D0812	$ARIMA(0,1,1) (1,1,1)_6$
Adjusted R-squared	0.633084	0.630973
Sum squared resid	0.788513	0.800130
Log likelihood	124.8925	124.0442
Akaike info criterion (AIC)	-2.084354	-2.086969
Schwarz criterion (SBC)	-1.989403	-2.015756
Durbin-Watson stat	1.951340	1.944924
Q-7 (p-value)	4.6807 (0.322)	4.5111 (0.341)
Q-14 (p-value)	6.9596 (0.802)	6.6487 (0.827)
Q-21(p-value)	16.480 (0.559)	17.856 (0.465)

Fuente: Con datos del BIE, INEGI, 2012.

En este caso si se compara el modelo  $ARIMA(0,1,1) (1,1,1)_6$  con y sin intervención mostrados en la tabla 4.6 se puede comentar que no necesariamente se ha logrado un mejor ajuste en el modelo con intervención, pues aunque la desviación típica residual es un tanto menor que el modelo sin intervención y el  $R^2$  ajustado es mayor; sin embargo, los estadísticos de parsimonia el AIC y SC no son menores. Aun cuando no todo apunta a haber logrado un mejor ajuste, se hará un diagnóstico a los residuos del modelo con intervención.

Al analizar los diferentes rezagos de la figura 4.28 todos aceptan la hipótesis de aleatoriedad, es decir, los residuos son incorrelacionados.



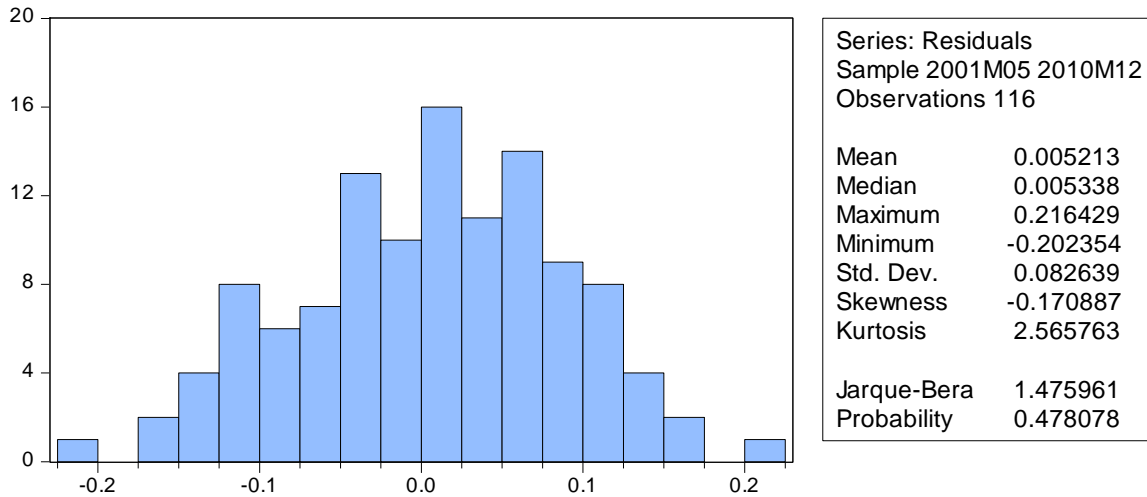
**Figura 4.28** Correlograma de los residuos del modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

Si se analiza la figura 4.29 se puede observar que la probabilidad asociada al estadístico de Jarque-Bera es mayor que el valor  $\alpha$  establecido, por lo que no se rechaza la hipótesis nula, además se puede afirmar que se asemeja a una normal y el estadístico ARCH de la figura 4.30, prueba que la relación cuadrática entre el residuo y su valor retardado no tienen una relación de dependencia. Por lo tanto, no es conveniente construir un modelo ARCH.

Los residuos de la estimación del modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos con intervención D0812 se dan en la figura 4.31, obsérvese que el error atípico de enero de



1993 no ha desaparecido, lo que apoya la idea de que este modelo con intervención no necesariamente mejora el ajuste.



**Figura 4.29** Histograma de los residuos del modelo  $ARIMA(0,1,1) (1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

Heteroskedasticity Test: ARCH

F-statistic	0.506448	Prob. F(1,113)	0.4781
Obs*R-squared	0.513112	Prob. Chi-Square(1)	0.4738

Test Equation:

Dependent Variable: RESID^2  
 Method: Least Squares  
 Date: 02/07/12 Time: 15:27  
 Sample (adjusted): 2001M06 2010M12  
 Included observations: 115 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006398	0.001018	6.283245	0.0000
RESID^2(-1)	0.066785	0.093846	0.711652	0.4781

R-squared	0.004462	Mean dependent var	0.006856
Adjusted R-squared	-0.004348	S.D. dependent var	0.008447
S.E. of regression	0.008466	Akaike info criterion	-6.688348
Sum squared resid	0.008099	Schwarz criterion	-6.640610
Log likelihood	386.5800	Hannan-Quinn criter.	-6.668971
F-statistic	0.506448	Durbin-Watson stat	1.994184
Prob(F-statistic)	0.478147		

**Figura 4.30** Prueba ARCH al modelo  $ARIMA(0,1,1) (1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generada con Eviews. (INEGI, 2012)

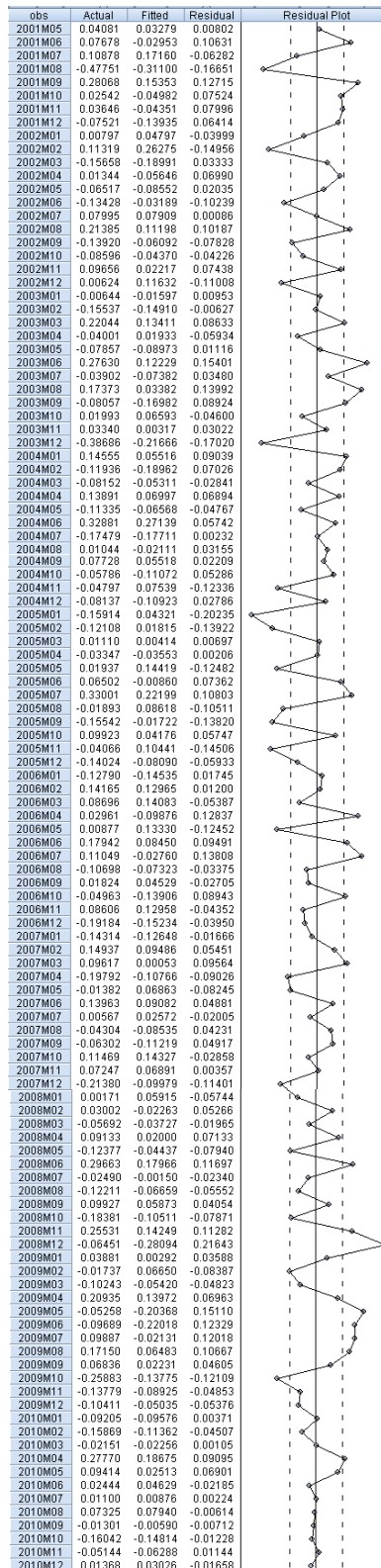
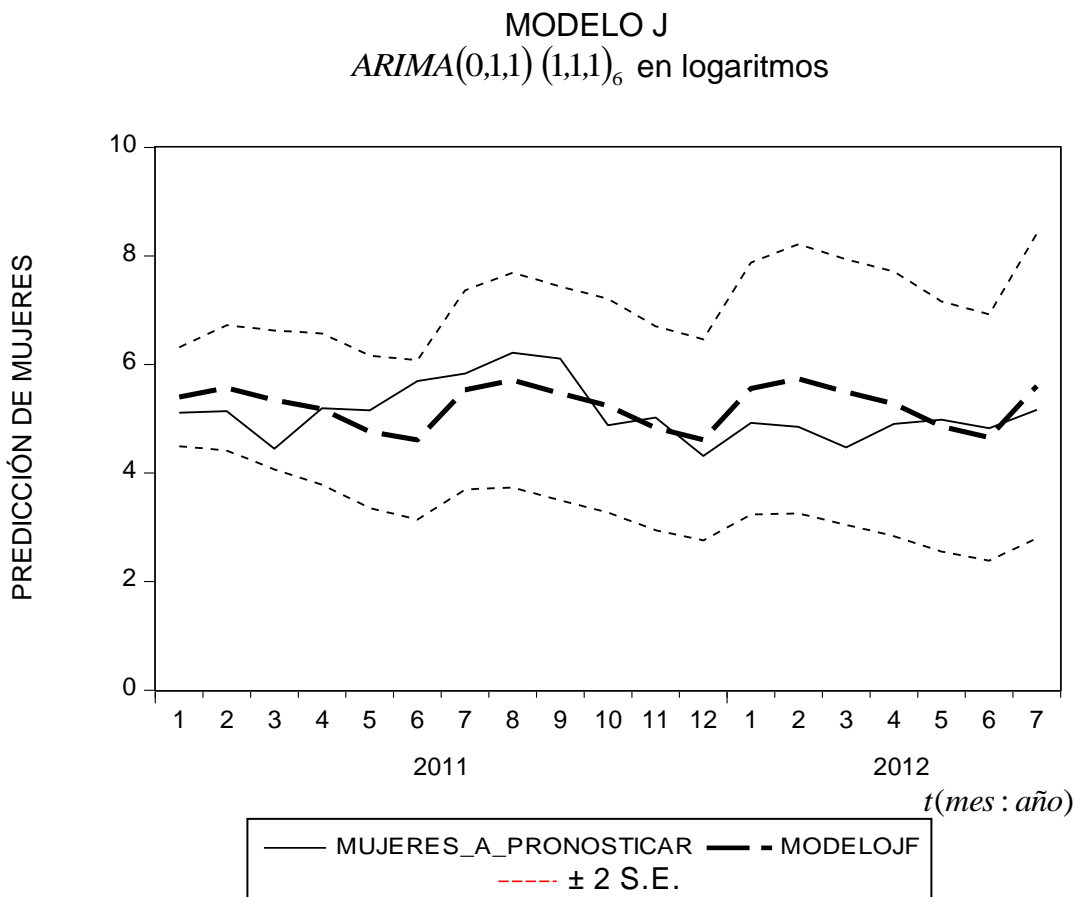


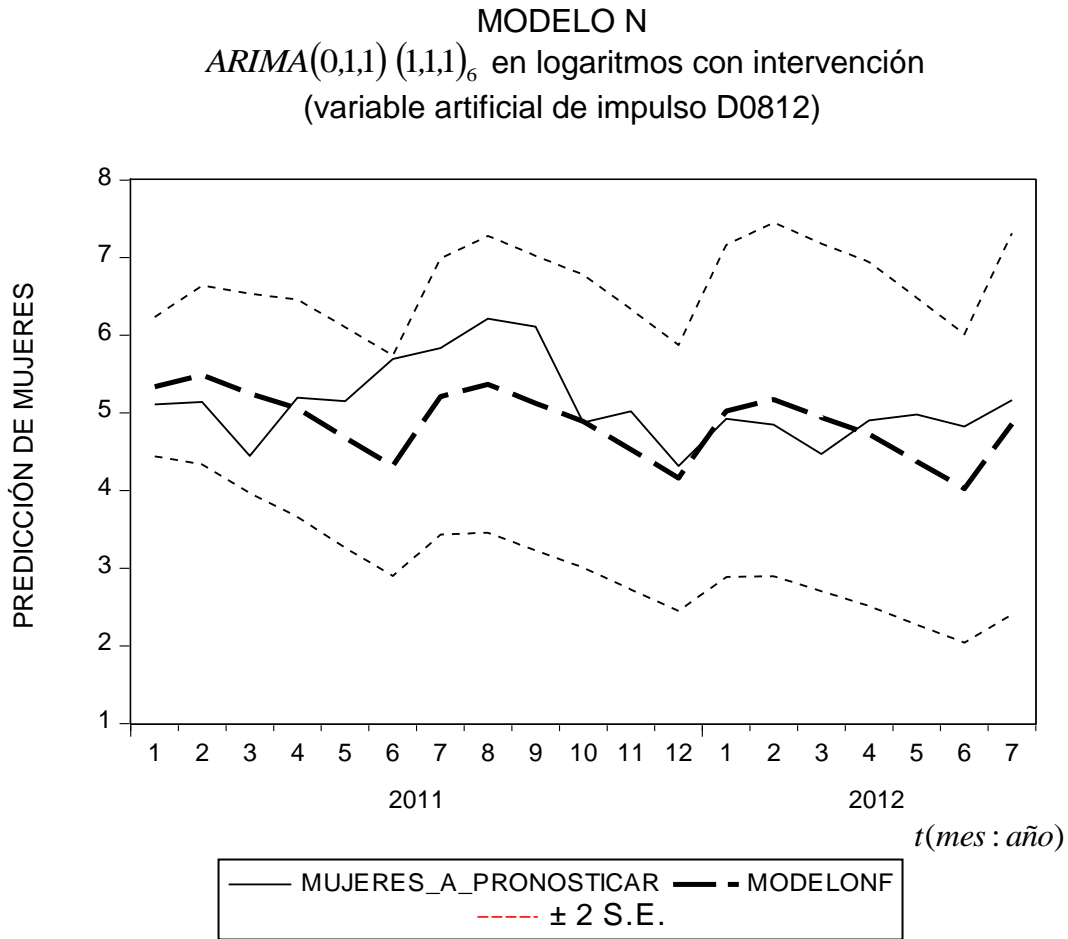
Figura 4.31 Residuos del modelo  $ARIMA(0,1,1) (1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres, con datos a diciembre de 2010, generado con Eviews. (INEGI, 2012)

### 4.2.5 Predicción

Una vez desarrollado un modelo basado en las observaciones en el período histórico (2000:04 al 2010:12), se genera a partir de él, predicciones para el período de validación (2011:01 al 2012:07). La comparación entre las predicciones y las observaciones en el periodo de validación permitirá evaluar la capacidad predictiva del modelo. Finalmente se tienen dos modelos elegidos para una misma serie (tasa de desocupación mensual para mujeres) a la que se evaluará su capacidad predictiva; los modelos son  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos y con intervención en diciembre de 2008, se elegirá aquel modelo cuya capacidad predictiva sea mayor.



**Figura 4.32** Valores reales, predicciones y límites de intervalo de confianza ( $\pm 2S.E.$ ) del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos, de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews. (INEGI, 2012)



**Figura 4.33** Valores reales, predicciones y límites de intervalo de confianza ( $\pm 2S.E.$ ) del modelo  $ARIMA(0,1,1) (1,1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews. (INEGI, 2012)

En las figuras 4.32 y 4.33 se presentan los valores reales, las predicciones y los límites inferior y superior de los intervalos de confianza al 95% de los modelos  $ARIMA(0,1,1) (1,1,1)_6$  en logaritmos y con intervención en diciembre de 2008

A continuación se presentan los resultados para evaluar la capacidad predictiva de los modelos elegidos, en el análisis de las tablas 4.7 y 4.8 se puede apreciar a primera vista que, en forma bastante general, los dos modelos generan predicciones relativamente buenas para la tasa de desocupación mensual para mujeres.

Utilizando un indicador que permite comparar fácilmente resultados entre diferentes modelos se observa que los errores relativos de predicción con un período adelante son

pequeños en media.

**Tabla 4.7** Predicciones del modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos, de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews.

<b>Predicciones</b>			
<b>Periodo</b>	<b>Valor real</b>	<b>Predicción</b>	<b>Error relativo</b>
ene-11	5.108821	5.400462	0.057086
feb-11	5.141112	5.568421	0.083116
mar-11	4.447691	5.342178	0.201113
abr-11	5.195681	5.175396	0.003904
may-11	5.152711	4.756072	0.076977
jun-11	5.693211	4.607247	0.190747
jul-11	5.834920	5.531169	0.052058
ago-11	6.215433	5.711302	0.081109
sep-11	6.110000	5.465998	0.105401
oct-11	4.877926	5.238083	0.073834
nov-11	5.019738	4.823851	0.039023
dic-11	4.316916	4.607668	0.067352
ene-12	4.925266	5.555087	0.127876
feb-12	4.849047	5.734246	0.182551
mar-12	4.471236	5.490817	0.228031
abr-12	4.904536	5.274194	0.075371
may-12	4.980556	4.854900	0.025229
jun-12	4.823586	4.651377	0.035701
jul-12	5.166720	5.602689	0.084380
Medias de errores (valor absoluto)			0.094256
Desviación teórica			0.084147

**Fuente:** Con datos del BIE, INEGI, 2012.

**Tabla 4.8** Predicciones del modelo  $ARIMA(0,1,1)(1,1)_6$  en logaritmos con intervención (variable artificial de impulso D0812), de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews.

<b>Predicciones</b>			
<b>Periodo</b>	<b>Valor real</b>	<b>Predicción</b>	<b>Error relativo</b>
ene-11	5.108821	5.337418	0.044745
feb-11	5.141112	5.487648	0.067405
mar-11	4.447691	5.249289	0.180228
abr-11	5.195681	5.059734	0.026165
may-11	5.152711	4.681618	0.091426
jun-11	5.693211	4.317288	0.241678
jul-11	5.834920	5.209415	0.107200
ago-11	6.215433	5.367175	0.136476
sep-11	6.110000	5.124685	0.161263
oct-11	4.877926	4.890642	0.002607
nov-11	5.019738	4.528124	0.097936
dic-11	4.316916	4.161020	0.036113
ene-12	4.925266	5.023952	0.020037
feb-12	4.849047	5.173740	0.066960
mar-12	4.471236	4.941967	0.105280
abr-12	4.904536	4.726584	0.036283
may-12	4.980556	4.375599	0.121464
jun-12	4.823586	4.023974	0.165771
jul-12	5.166720	4.857829	0.059785
Medias de errores (valor absoluto)			0.093096
Desviación teórica			0.083906

**Fuente:** Con datos del BIE, INEGI, 2012.

El modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos, tiene la media de errores en valor absoluto igual a 0.094256, valor muy cercano a lo esperado, la desviación teórica de 0.084147, y esta a su vez es igual a la desviación típica residual; mientras que el modelo con intervención (variable artificial de impulso D0812) tiene la media de errores en valor absoluto igual a 0.093096, también cercano a lo esperado 0.083906.

Por lo tanto, considerando el criterio de parsimonia, el modelo que mejor comportamiento tiene para predicción es el modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos.

# CONCLUSIONES

---

---

A lo largo de este trabajo se buscaron diferentes indicadores que miden la desocupación en México.

Haciendo un contraste entre los niveles de instrucción se encontró que los escolarizados predominan entre los buscadores activos de empleo, pues para dichas personas vale la pena esa inversión de tiempo completo en la búsqueda y la espera, con tal de conseguir mejores oportunidades y mayores ingresos. Así, en diciembre de 2010, 35.63% de los desempleados contaban con educación media superior o superior, contra 20.11% que a lo sumo había cursado primaria completa o 9.75% que no llegó a completar la primaria.

En términos regionales también se encontraron diferencias en el tema de la desocupación, pues en áreas urbanas se identifican mayores tasas de desempleo que en zonas rurales, esto debido a la cultura asalariada y a las expectativas que se dan en zonas más estrechamente vinculadas al proceso de modernización. Así en las poblaciones de los estados de Nuevo León, Estado de México o el Distrito Federal, se presentan mayores tasas de desocupación abierta que las observadas para el caso de Chiapas, Guerrero u Oaxaca.

Desde el punto de vista de la edad, se encontró que es mayor el problema de la desocupación entre los jóvenes. El alto desempleo juvenil puede deberse a que estos grupos de la población se encuentran moviéndose de la escuela al mundo laboral y de la residencia en el seno familiar a su emancipación, es decir, a una independencia tanto en el aspecto económico y residencial, como en el de la toma de sus decisiones, razón que lo caracteriza como un grupo con frecuentes entradas y salidas del mercado de trabajo. Además su entrada al mercado laboral depende de otros factores como los requerimientos mínimos de especialización, la capacidad que tienen los empleadores para introducir y/o remplazar trabajadores, entre otros. Así se comparó tasas de desocupación para diferentes grupos por edades y diversos periodos, y se encontró que



el grupo de 14 a 19 años presentaba en su mayoría tasas de tres o más veces las tasas correspondientes a los grupos de 40 a 49 años, 50 a 59 años o de 60 años y más

De acuerdo al análisis por sexo se pudo corroborar que durante los últimos años ha mejorado el panorama laboral de las mujeres en México, con diferencias que han disminuido con el tiempo entre las tasas de desocupación de los hombres y de las mujeres, pero con diferencias más acentuadas entre las tasas complementarias de desocupación, posiblemente debido a que muchas mujeres consiguen trabajos de ocupación parcial por las actividades que tienen que desempeñar además en el hogar.

Los cambios que se han dado en el desempleo para las mujeres son resultado del crecimiento de la participación económica femenina de los últimos decenios, aunque de forma paralela también han aumentado las dificultades para obtener y conservar el empleo para las mujeres, aun cuando ellas tienen más facilidad para adaptarse a las innovaciones, se caracterizan por ser disciplinadas y responsables, son más efectivas durante la jornada académica al administrar mejor el tiempo y hacen su trabajo con una alta calidad y clara orientación al detalle. Es por todas estas razones, que se decidió contar con una visión de lo que va a suceder en el futuro con la desocupación femenina en México y por lo que surgió la idea de predecir el comportamiento de la tasa de desocupación femenina.

Predecir es una tarea compleja, pero la metodología de Box-Jenkins resulta una propuesta metodológica que utiliza de forma eficiente la información completa de la serie para proporcionar información futura que normalmente no se tiene con antelación. Las predicciones que proporciona esta metodología se hacen a través de los modelos ARIMA.

Para el análisis, la estimación de los modelos ARIMA y los pronósticos se utilizaron las tasas de desocupación femenina desde abril de 2000 a diciembre de 2010, y los datos de enero de 2011 a julio de 2012 sirvieron para evaluar la capacidad predictiva de los modelos.

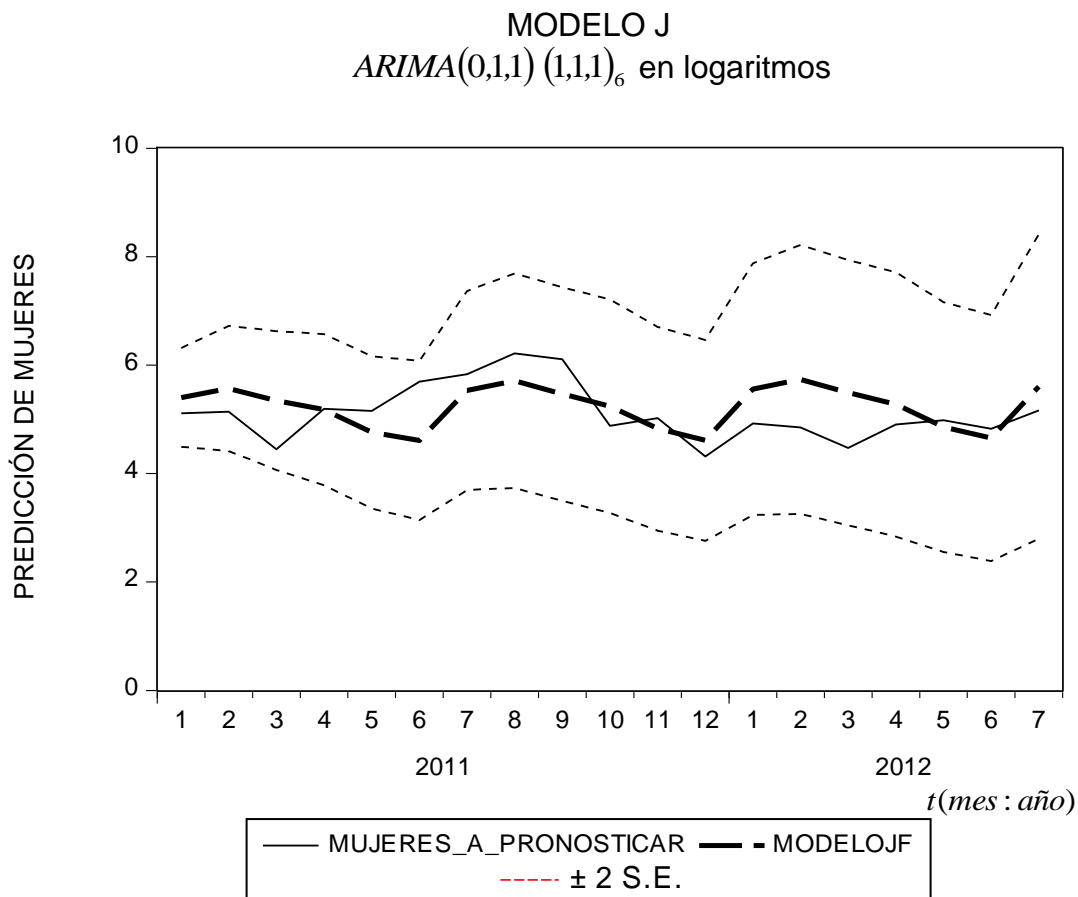
Para la aplicación de la metodología de Box-Jenkins se siguieron los siguientes pasos:

- Se graficó la serie para conocer su comportamiento a lo largo del tiempo.

- Se observó la gráfica para identificar si se trataba de una serie estacionaria. Esto debido a que la metodología de Box-Jenkins así lo requiere, es decir, que la media y la varianza de la serie de tiempo permanezcan constantes a lo largo del tiempo.
- Al presentar evidencias de que no lo era, se obtuvieron y analizaron las autocorrelaciones de la misma serie y se corroboró la no estacionariedad.
- Se procedió a aplicarle diferentes transformaciones para volverla estacionaria.
- Se realizó un estudio de la estacionariedad para verificar que transformación era la más adecuada a través de pruebas estadísticas y se decidió utilizar la serie de primeras diferencias en logaritmos.
- Una vez que se tuvo una serie estacionaria, se obtuvieron y revisaron las gráficas de la función de autocorrelación simple y parcial, con el fin de descubrir en que rezagos se presentaba mayor correlación e incluso se identificó que había un comportamiento estacional.
- Se procedió a aplicarle otra transformación para extraer el comportamiento estacional, en este caso primeras diferencias en la parte estacional, que se identificó en periodos de cada seis meses.
- Se obtuvieron y analizaron las autocorrelaciones a la serie transformada con la finalidad de identificar posibles modelos ARIMA para la tasa de desocupación femenina mediante el proceso de modelación de Box-Jenkins.
- Se estimaron los parámetros de los posibles modelos y se identificó cual era el más adecuado, para este caso un modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos.
- Una vez elegido el modelo ARIMA se procedió a realizar un diagnóstico de éste, donde se realizaron pruebas estadísticas para verificar la incorrelación, normalidad y heterocedasticidad residual, encontrando resultados favorables y concluyendo la validez del modelo.
- Aún cuando ya se tenía el modelo ARIMA para llevar a cabo los pronósticos, se realizó un análisis de intervención al localizar cuatro valores atípicos,

encontrando que sólo el modelo con variable artificial en diciembre de 2008 presentaba parámetros significativos, dejando de lado aquellos que no tenían parámetros significativos en las variables correspondientes a las otras tres fechas. Con este nuevo modelo no se logró un mejor ajuste, sin embargo también fue utilizado para pronosticar y comparar con el primer modelo elegido.

- Se realizaron los pronósticos para enero de 2011 a julio de 2012, y al compararlos gráficamente (figura 4.32) con los datos originales se comprobó que los modelos elegidos brindaron predicciones confiables de la tasa de desocupación femenina en México.



**Figura 4.32** Valores reales, predicciones y límites de intervalo de confianza ( $\pm 2S.E.$ ) del modelo  $ARIMA(0,1,1) (1,1,1)_6$  en logaritmos, de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews. (INEGI, 2012)

- Se evaluó la capacidad predictiva de los modelos elegidos (tabla 4.7) y se concluyó que el modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos, tiene la media de errores en valor absoluto más cercana a la desviación teórica.

**Tabla 4.7** Predicciones del modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos, de la tasa de desocupación mensual para mujeres para enero de 2011 a julio de 2012, generado con Eviews.

<b>Predicciones</b>			
<b>Periodo</b>	<b>Valor real</b>	<b>Predicción</b>	<b>Error relativo</b>
ene-11	5.108821	5.400462	0.057086
feb-11	5.141112	5.568421	0.083116
mar-11	4.447691	5.342178	0.201113
abr-11	5.195681	5.175396	0.003904
may-11	5.152711	4.756072	0.076977
jun-11	5.693211	4.607247	0.190747
jul-11	5.834920	5.531169	0.052058
ago-11	6.215433	5.711302	0.081109
sep-11	6.110000	5.465998	0.105401
oct-11	4.877926	5.238083	0.073834
nov-11	5.019738	4.823851	0.039023
dic-11	4.316916	4.607668	0.067352
ene-12	4.925266	5.555087	0.127876
feb-12	4.849047	5.734246	0.182551
mar-12	4.471236	5.490817	0.228031
abr-12	4.904536	5.274194	0.075371
may-12	4.980556	4.854900	0.025229
jun-12	4.823586	4.651377	0.035701
jul-12	5.166720	5.602689	0.084380
Medias de errores (valor absoluto)			0.094256
Desviación teórica			0.084147

**Fuente:** Con datos del BIE, INEGI, 2012.

Por lo tanto, considerando el criterio de parsimonia, el modelo que mejor comportamiento tiene para predicción de la tasa de desocupación femenina en México es el modelo  $ARIMA(0,1,1)(1,1,1)_6$  en logaritmos.

Con lo que se concluye que la metodología Box-Jenkins es una herramienta útil para describir la evolución futura de la desocupación femenina en México y con ello una forma para adelantarse a lo que puede ocurrir, con la finalidad de atender la participación laboral femenina.

Cabe resaltar que este resultado no tiene la fuerza para eliminar la desocupación, sólo es una herramienta más que se puede utilizar para buscar comprender el problema, complementándolo con diversos análisis considerando los tipos de desempleo, contrastando el desempleo por niveles de escolaridad, regiones, edad, etc., para así crear mejores propuestas para combatirlo.

# REFERENCIAS

---

---

Bowerman, Bruce L.; O'Connell, Richard T.; Koehler, Anne B. (2007). Pronósticos, Series de Tiempo y Regresión. México: Cengage Learning Editores, S. A. de C. V.

Dacorogna, Michel. (2001). An introduction to high-frequency finance. USA: San Diego Academic.

Enders, Walter. (1995). Applied econometric time series. USA: John Wiley.

Farnum, Nicholas. (1989). Quantitative forecasting methods. USA: PWS-KENT.

Guerrero, Víctor Manuel. (2003). Análisis Estadístico de Series de Tiempo Económicas. México: Thompson Editores, S. A. de C. V.

Gujarati, Damodar N. (2003) Econometría. México: Mc Graw Hill.

Hernández, José; Herrador Morales, María del Mar. (2000). Econometría de Series Temporales. España: Ed. Universitas, S. A.

INEGI. (2002). Guía de conceptos, uso e interpretación de la Estadística sobre Fuerza Laboral en México. México: INEGI.

INEGI. (2005). Encuesta Nacional de Ocupación y Empleo. Una nueva encuesta para México. México: INEGI.

INEGI. (2009). Glosarios de estadísticas demográficas. México: INEGI.

Kendall, Maurice G.; Buckland, William. (1971). A dictionary of Statistical Terms. New York: Hafner Publishing Company, Inc.

Loría, Eduardo; Ramos, Manuel. (2007). La ley de Okun: Una relectura para México, 1970-2004. México: Estudios Económicos. El Colegio de México.

Melgar, Ivonne. (2009). Calderón abre paso al seguro de desempleo. México: Periódico Excélsior, S. A. de C. V.

Mills, Terence. (1999). The econometric modelling of financial time series. USA: Cambridge University.

Nava, Alejandro. (2002). Procesamiento de series de tiempo. México: F. C. E.

Reyes, Francisco. (1995). Macroeconomía. México: Adisson-Wesley Iberoamericana, S. A.

Serret Moreno-Gil, Jaime. (1998). Procedimientos Estadísticos. España: ESIC Editorial.

Tsay, Ruey S. (2010). Analysis of financial time series. USA: Wiley.

Umphrey, Gary. (1987). Time series and econometric modelling. USA: D. Reidel.

Velastegui, L. A. (2006). Una estimación de la Ley de Okun para el Ecuador. Ecuador: Observatorio de la Economía Latinoamericana.