



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**EFFECTO DE LA REFORMA DE PENSIONES DEL
IMSS DE 1997 EN EL TAMAÑO DEL SECTOR
FORMAL**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A CTUARIO

P R E S E N T A :

GENARO CORTEZ AGUILAR



**DIRECTOR DE TESIS:
DRA. MARTHA MIRANDA MUÑOZ
2013**



Universidad Nacional
Autónoma de México



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.

Hoja de datos del Jurado

1. Datos del Alumno
Cortez
Aguilar
Genaro
47 61 27 57
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
303070937
2. Datos del tutor
Dra.
Martha
Miranda
Muñoz
3. Datos del sinodal 1
Dra.
María del Pilar
Alonso
Reyes
4. Datos del sinodal 2
Act.
Carlos Fernando
Lozano
Nathal
5. Datos del sinodal 3
Act.
Francisco Fernando
Morales
Castro
6. Datos del sinodal 4
Act.
Ernesto
Rosas
García
7. Datos del trabajo escrito
Efecto de la reforma de pensiones del IMSS de 1997 en el tamaño del sector formal.
125 p
2013

Agradecimientos

Antes que nada, quiero dar las gracias a Dios por haberme permitido llegar a éste punto de la vida y haberme dejado compartir momentos de felicidad con todos mis seres queridos.

Manifiesto mi total y eterna gratitud a mis padres:

***Gisela Aguilar Díaz** y **Genaro Cortez Prado**, porque ante las adversidades siguieron adelante, porque sus limitaciones fueron su impulso para construirme un mejor presente; porque me enseñaron apreciar lo poco que se tiene, y que el trabajo diario y constante es la base del éxito. Gracias por haberme inculcado principios y valores que son la base de mi conducción diaria.*

Agradezco a mis tres hermanos:

***Rafael:** Quien me contagia su fortaleza y convicción para conseguir objetivos, es fuente de mi inspiración para seguir creciendo en lo profesional. Te dedico este trabajo por lo mucho que le has dado a mí persona y porque es un pequeño reflejo de tu gran esfuerzo.*

***Juan Carlos:** Que a pesar de los malos momentos vividos muestra una sonrisa para seguir adelante, mostrándome con su protección lo mucho que me ama y porque nos ha presentado a su primer ángel, Leonel.*

***Gisela:** Mi compañera de juegos y diversiones de infancia, que a pesar de ser la menor de los cuatro muestra mucha madurez para realiza actividades que aún no le corresponden y por ser la que reconforta con su cariño en todo momento.*

Gracias a todos los integrantes de mi familia, porque aún nos seguimos divirtiendo como niños.

*A **Carina Báez** que como parte de mi familiar, es una excelente amiga, gracias por alentarme a dar el siguiente paso sin importar que pudiera pasar, por inyectarme esa alegría que te caracteriza, por tus valiosos consejos, y porque hemos sido cómplice de algunos proyectos humanos.*

*A **Alicia** y **Ernesto** quienes en su hogar hemos compartido momentos de alegría, contribuyendo a que mi familia se pueda unir. Y que muy a su manera dispusieron de su tiempo para colaborar en nuestra educación.*

*A la **Dra. Martha Miranda** que gracias a su gran apoyo y a su confianza, pude poner punto final a éste importante y grandioso ciclo, además de encaminarme en la senda de la investigación con su ejemplo.*

La conclusión de este trabajo se ha llevado a cabo gracias a la participación activa de mis amigos, que me han extendido la mano para apoyarme, de mis profesores que son el motor para la apertura y transmisión de su conocimiento que tanto hace falta en este México, y de aquellas personas que aportaron con sus opiniones y me ayudaron en la resolución de dudas.

Agradezco a todos los que formaron parte de mi jurado a la Dra. María del Pilar Alonso Reyes, al Act. Carlos Fernando Lozano Nathal, al Act. Francisco Fernando Morales Castro y al Act. Ernesto Rosas García, por haber dispuesto de su tiempo en la revisión de este trabajo, que con sus observaciones tan precisas, no sólo ayudaron a mejorarlo sino ampliarme aun más el panorama.

Finalmente quiero agradecer a la prestigiada Universidad Nacional Autónoma de México y a la Facultad de Ciencias, porque mi periodo de estudiante fue enriquecedor con profesores tan excelentes, que hacen que nos enamoremos de lo difícil.

Índice de Contenido

Introducción	1
I. Reforma de 1997 al programa de pensiones de la Ley del Seguro Social	4
1.1 Ley del Seguro Social de 1973	6
1.2 Ley del Seguro Social de 1997	7
1.2.1 Análisis de los cambios más relevantes al esquema de pensiones en la Ley del seguro social de 1997	10
1.3 Afiliación	16
II. Mercado Laboral	26
Introducción	26
2.1 Concepto de sector informal	26
2.2 Medición del mercado laboral	29
2.3 Sector informal y protección social	30
III. Metodología	36
3.1 Motivación y aplicación del modelo de selección binaria	36
3.1.1 Modelo de probabilidad lineal	37
3.2 Modelo de regresión logística.....	41
3.2.1 Ajuste del modelo de regresión logística.....	43
3.2.2 Pruebas de significancia de los coeficientes estimados	46
3.2.3 Interpretación del modelo	49
3.2.4 Bondad de Ajuste	53
3.2.5 Selección del modelo.....	57
IV. Las encuestas nacionales de empleo en México	59
4.1 Descripción de la base de datos de la Encuesta Nacional de Empleo (ENE) y la Encuesta Nacional Ocupación de Empleo (ENOE)	62
4.2 Sector formal e informal de la ENE y ENOE	65
4.3 Variables que explican al sector formal e informal en la ENE de 1996 y ENOE del 2006.....	65
4.3.1 Depuración de las bases de datos	72

V. Construcción del modelo que explica el efecto de la reforma al esquema de pensiones en el tamaño del sector formal	74
5.1 Implementación del modelo logit en la ENE96	74
5.1.1 Ajuste del modelo logit	74
5.1.2 Prueba de significancia de los coeficientes	78
5.1.3 Comparando modelos	80
5.1.4 Bondad de Ajuste	82
5.2 Implementación del modelo logit en la ENOE06	85
5.2.1 Ajuste del modelo logit	85
5.2.2 Prueba de significancia de los coeficientes	88
5.2.3 Comparando modelos	90
5.2.4 Bondad de Ajuste	91
5.3 Resultados: Interpretación del modelo logit de 1996 Y 2006	94
5.3.1 Cociente de momios	95
5.3.2 Efectos marginales	103
Conclusiones	113
Referencias Bibliográficas	118
Anexo 1	121
Anexo 2	123

Introducción

Mesa-Lago (2004) menciona que la Organización Mundial del Trabajo (OIT) y la Asociación Internacional de Seguridad Social (AISS), definen la seguridad social como “la protección que la sociedad proporciona a sus miembros, mediante una serie de medidas públicas, contra las privaciones económicas y sociales que de no ser así ocasionarían la desaparición o una fuerte reducción de los ingresos por causa de enfermedad, maternidad, accidente de trabajo o enfermedad profesional, desempleo, invalidez y vejez, y también la protección en forma de asistencia médica y de ayudas familiares con hijos”. El concepto es muy amplio ya que engloba a los programas contributivos y los no contributivos.

Entonces resulta muy importante brindar y extender la cobertura de seguridad social, siendo éste uno de los mayores retos que enfrenta el país, pero no sólo es un reto propio, si no que es un reto que enfrentan muchos países, y el logro va tener una fuerte dependencia con la industrialización, el tamaño del sector formal y de la antigüedad del sistema de cada uno (Mesa-Lago, 2004).

Dentro de los programas contributivos los institutos encargados de dar cobertura de seguridad social en México son el Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE), Petróleos Mexicanos (PEMEX), Fuerzas Armadas Mexicanas (FNM), Secretaría de la Defensa Nacional (SEDENA), Secretaría de Marina (SEMAR), Estatales y el Instituto Mexicano del Seguro Social (IMSS). Pero el IMSS es quien más contribuye en cobertura de seguridad social que emana del tamaño del sector formal y de la afiliación voluntaria. Por este motivo, en la presente tesis se usara el término seguridad social y protección social refiriéndose sólo a los beneficios y prestaciones que ofrece la Ley del Seguro Social de 1997, actualmente vigente.

En la última reforma a la Ley del Seguro Social, que sustancialmente tuvo más modificaciones a las que se refieren a las pensiones de vejez, se diseñó para mejorar el bienestar del trabajador y el de su familia, aunque se decía que iba a venir consigo efectos positivos en el incremento nacional en los niveles de ahorro, la *formalización del mercado laboral* y el desarrollo de los mercados financieros; además que se buscaba darle viabilidad financiera al IMSS en el largo plazo (Villagómez 2009, Conferencia Interamericana de Seguridad Social 2003 y 2005).

Como se mencionó que la formalización del mercado laboral está estrechamente relacionada con el incremento en la cobertura de protección social, la base donde se centra el objetivo de la tesis, es probar la hipótesis del incremento en la *formalización del mercado laboral*. Por lo dicho, se analizará ¿Sí la reforma de pensiones causo un efecto en el tamaño del sector formal? Además de ello se estudia la relación de los factores que influyen en el trabajador para que se encuentre empleado en el sector formal.

Intentar responder esta pregunta es algo muy complejo dado que no se cuenta con la información ideal para responder cuál es el efecto de una reforma al esquema de pensiones del IMSS. Esto implicaría, idealmente contar con un experimento para evaluar una política

pública (contar con un grupo de tratamiento y un grupo de control Aedo (2005)) y contar con datos longitudinales que permitieran saber cómo cambia en el tiempo la historia laboral de un trabajador para identificar transiciones entre el sector formal e informal; estos datos aun no existen en México (el panel sintético de la Encuesta Nacional de Empleo y Ocupaciones solo permite conocer información cinco trimestres). En esta tesis, con el propósito de examinar el posible efecto de la reforma al esquema de pensiones de vejez del IMSS de julio de 1997, se examinan datos de encuestas de empleo en dos momentos en el tiempo: antes y después de la reforma.

En base a la probabilidad de que un individuo pertenezca al sector formal (o informal) de la economía, se ajusta un modelo de regresión logística sobre la Encuesta Nacional de Empleo (ENE) de 1996 (antes de la reforma) y la Encuesta Nacional de Ocupación y Empleo (ENOE) del 2006 (después de la reforma), y se compararan los resultados obtenidos de un periodo a otro en cuanto a la variación en el tamaño del sector formal en función de características individuales de cada trabajador: género, la edad, el nivel de instrucción, su ocupación, el tamaño de la unidad económica, su jornada laboral y la actividad económica a la que se dedica.

Los resultados del modelo estadístico de regresión con una variable dependiente dicotómica, sugieren que la probabilidad de pertenecer al sector formal de la economía no ha variado significativamente entre 1996 y 2006, es decir, después de 9 años de ocurrida la reforma al esquema de pensiones de vejez de la Ley del Seguro Social, contrario a lo que se esperaba en 1997 cuando se planteaban las posibles ventajas de llevar a cabo una reforma en la que se sustituyera el sistema de reparto por el de cuentas individuales. El modelo estadístico da evidencia de que: i) el tamaño del sector formal es mayor en la región norte (en donde se han instalado empresas extranjeras dedicadas principalmente a la maquila); ii) hay un ligero decremento en la propensión a pertenecer al sector formal conforme aumenta la edad (en el rango 14-60 años de edad); iii) la brecha respecto al género de pertenecer al sector formal ha ido disminuyendo; iv) la probabilidad de pertenecer al sector formal es mayor para los individuos cuyo estado civil es “casado”; v) el nivel de instrucción sigue siendo una característica importante para estar en el sector formal, pero probabilísticamente esta ha descendido; vi) los trabajadores subordinados y remunerados sigue contribuyendo importantemente en la formalidad, pero probabilísticamente esta también ha descendido; vii) el tamaño de la empresa es un determinante importante de la cobertura de la seguridad social: a mayor tamaño de la empresa, mayor probabilidad de ser trabajador formal; viii) la actividad económica que más genera empleo formal es el sector de servicios.

La presente tesis se divide en cinco capítulos. En el Capítulo I de manera general se describe el proceso histórico del nacimiento de la Ley de Seguro Social en México y de los cambios se ha sufrido, pero se detiene a dar un análisis más detallado sobre los cambios al esquema de pensiones que se dio en la reforma de 1997. Seguido de ello se presenta la evolución de la cobertura en seguridad social, haciendo especial énfasis en la contribución que tiene el IMSS en cobertura de protección social.

El Capítulo II comienza definiendo el concepto del empleo informal la cual tiene varias acepciones y se llega a integrar una definición con los diversos enfoques. A partir de ahí se pasa al campo de la estadística para exponer de que maneras se mide el sector informal y se

explica cual es la influencia que tiene sobre a cobertura de protección social. Es en este capítulo donde se tiene una idea clara de lo que es el sector informal y en consecuencia el formal.

Sobre el Capítulo III se describe la base metodológica que se sigue para modelar un fenómeno cuya variable dependiente es de naturaleza dicotómica. Para tratar este tipo de fenómenos se introduce primero a través de un modelo lineal de probabilidad y se plantean los problemas a los que se enfrenta. Para evitar estos problemas se recurre a una transformación de la variable por medio de la función logística, lo que lleva a un planteamiento de un modelo de regresión logístico. Del modelo logístico se da detalle del ajuste de modelo y se establecen las estadísticas necesarias para determinar la significancia de los coeficientes estimados, su bondad de ajuste, la selección de modelo y su interpretación.

En el Capítulo IV se describen las encuestas nacionales de empleo en México, que son la fuente de información para analizar el sector formal e informal. Se inicia con un contexto histórico y se mencionan las actualizaciones que ha sufrido con el propósito de que refleje la situación en la que vive el país. En este capítulo se especifica cómo a partir de las encuestas se calcula el sector formal e informal, y cuales variables se seleccionan para explicar ambos sectores.

En el Capítulo V se ajusta un modelo logístico para la Encuesta Nacional de Empleo (ENE) de 1996 y para la Encuesta Nacional de Empleo y Ocupación (ENOE) del segundo trimestre del 2006. Para ambos modelos se evalúa la significancia de los parámetros estimados, se prueba la bondad de ajuste, y se selecciona el mejor modelo para cada encuesta. Para finalizar este capítulo se da una interpretación del modelo ajustado para la ENE de 1996 y la ENOE del 2006 mediante el cociente de momios y los efectos marginales; se lleva un análisis de los factores que inciden sobre el trabajador para que se encuentre en el sector formal, asimismo se compara el cociente de momios y los efectos marginales de cada periodo para encontrar evidencia del efecto de la reforma de pensiones del IMSS en el tamaño del sector formal.

Por último, se presentan las conclusiones del efecto de la reforma de pensiones del IMSS de 1997 en el tamaño del sector formal, así como las referencias bibliográficas utilizadas en el capítulo I, II, III y IV. En la parte de los anexos se muestra gráficas del Capítulo V y el código que se utilizo en STATA (Statistical Software for Profesional) versión 12, en donde se evalúan los modelos.

Capítulo I

Reforma de 1997 al programa de pensiones de la Ley del Seguro Social

A principios del siglo XIX comenzaron a surgir las primeras ideas para crear una legislación en materia de seguridad social. Dado su momento histórico que vivía la República Mexicana en ese periodo, algunos sectores sociales comenzaban a ejercer una presión a la clase patronal y política para la creación de un esquema de protección para la clase trabajadora ante contingencias laborales. Inicialmente se trabajó en un esquema de protección que otorgará pensiones por vejez y accidentes de trabajo.

La creación de una legislación en materia de seguridad social fue un esfuerzo de diferentes actores de la sociedad y de la clase política. En aquella época existieron iniciativas de ley que fueron rechazadas, por falta de bases técnicas y actuariales, y algunas otras que por diferentes causas no pasaron. Cada nuevo proyecto presentado incluía más y nuevas prestaciones, no solo se consideraba dar cobertura a los trabajadores si no también a sus familias, por tanto se planteó otorgar pensiones de invalidez, vida y prestaciones familiares, además de las pensiones de vejez y accidentes de trabajo¹.

Bajo la administración del presidente Manuel Ávila Camacho, con la creación de la Secretaría del Trabajo y Previsión Social – cuyo objetivo era vigilar el cumplimiento de la Ley Federal de Trabajo-, dentro de la misma dependencia se creó el Departamento de Seguros Sociales y la Comisión Técnica de Seguros Sociales, en conjunto se encargaron de presentar una iniciativa de Ley de Seguridad Social.

La Ley del Seguro Social se publicó en el Diario Oficial de Nación el 19 de enero de 1943, pero el 1 de enero de 1944 comenzó a operar, con ella también nace el Instituto Mexicano del Seguro Social (IMSS). Los riesgos cubiertos bajo la ley fueron accidentes de trabajo y enfermedades profesionales, enfermedades generales y maternidad, invalidez, vejez y muerte, y desocupación en edad avanzada.

Uno de los objetivos de la Ley de Seguridad Social desde su nacimiento, fue procurar el bienestar económico y de salud, y por ende el bienestar social del trabajador, un argumento que se derivó de esto fue como lo apunta el IMSS (Evidencias para el debate, 1997) que el trabajador aumentaría la productividad, ya que estaría tranquilo de saber que cuenta con la protección ante diferentes contingencias, lo que motivaría y daba incentivos a realizar sus tareas laborales.

¹ Para conocer al detalle el proceso histórico véase el Libro “Seguridad Social- Una visión latinoamericana” Capítulo IV de Antonio Ruezga Barba.

Al comenzar a operar el IMSS contó con niveles importantes de inversión, para el desarrollo de su infraestructura y la apertura de nuevas clínicas en varios estados de la República Mexicana, con el propósito de poner al alcance las prestaciones y servicios al afiliado, y contribuir a una ampliación en la cobertura de la población. Desde sus inicios (1943) el IMSS reportó un crecimiento constante en la afiliación y por ende en la cobertura, por las cifras que registra el instituto, comenzó con una población asegurada de 355,527 y llegando a los setenta a 9 772, 492. Los primeros asegurados correspondía a las zonas urbanas, al transcurrir de los años se fue afiliando población rural (IMSS Aportaciones al debate, 1997).

La ampliación en los servicios y cobertura han sido objetivos desde la creación de la Ley del Seguro Social. En la pretensión de cumplir con los objetivos, la ley ha sido modificada en diversas ocasiones, que ha dependido del contexto demográfico, y socioeconómico, así como de la situación del IMSS. Aunque la mayoría de estas reformas ha sido para reparar un mal funcionamiento en la parte administrativa o fiscal, a excepción de Ley de 1973 y 1995 (IMSS, 1997).

Derivado de la creciente demanda de los servicios y prestaciones que otorgaba el IMSS a sus asegurados y de la necesidad de contar con los fondos suficientes, fue necesario actualizar para cada periodo requerido, los porcentajes de contribución del trabajador, patrono y Estado. En el Cuadro 1.1, muestra para cada periodo las contribuciones a la seguridad social para los diferentes ramos. Inicialmente para financiar el seguro de enfermedades y maternidad, invalidez, vejez, cesantía y muerte, la aportación total del patrón era del 6 por ciento, en tanto que el empleado y el estado contribuían con el 3 por ciento respectivamente. Ahora comparando cada periodo respecto al anterior se observa que: en el segundo (II) periodo se dio un incremento en las contribuciones de 33 por ciento en el seguro de enfermedades y maternidades, y para el tercer (III) periodo tuvo un incremento de 13 por ciento sobre el mismo ramo. El cuarto (IV) periodo la contribución para el seguro de enfermedades y maternidades, tuvo un aumento sólo para el patrón de un 25 por ciento y una disminución para el estado del 50 por ciento. La misma variación se dio sobre el seguro de invalidez, vejez, cesantía y muerte. Y para el quinto (V) periodo nuevamente se dio un aumento en la contribución del patrón de 12 por ciento y una disminución del Estado de 60 por ciento, esto fue para ambos seguros. Es decir en el (IV) y quinto (V) periodo el incremento solo fue para el patrón, mientras que el Estado disminuyó su participación para los mismos periodos.

Una contribución creciente y elevada a la seguridad social, donde el manejo de los recursos es poco transparente, y además donde se afecta el salario del trabajador, el ingreso de la empresa y el gasto del Gobierno tiene implicaciones negativas en la formalidad del empleo. Ya que provocan que el trabajador o la empresa incurran en la evasión y subdeclaración del salario reportado, además que para la empresa se aumenta el costo laboral; y para el Gobierno se vuelve una carga fiscal los recursos transferidos al Instituto -por concepto de pensiones-, más cuando los pasivos del IMSS se incrementan.

Cuadro 1. 1
Estructura de las Aportaciones para los Seguros de Enfermedades y Maternidad e Invalidez,
Vejez, Cesantía y Muerte
(En porcentaje de los salarios de los trabajadores)

<i>Periodo</i>	<i>1943-1949</i>	<i>1949-1959</i>	<i>1960-1965</i>	<i>1966-1986</i>	<i>1986</i>
	(I)	(II)	(III)	(IV)	(V)
<i>- Seguro de enfermedades y maternidad</i>					
<i>Patrón</i>	3.000	4.000	4.500	5.625	6.300
<i>Trabajador</i>	1.500	2.000	2.250	2.250	2.250
<i>Estado</i>	1.500	2.000	2.250	1.125	0.450
<i>Cuota Total</i>	6.000	8.000	9.000	9.000	9.000
<i>- Seguro de invalidez, vejez, cesantía y muerte</i>					
<i>Patrón</i>	3.000	3.000	3.000	3.750	4.200
<i>Trabajador</i>	1.500	1.500	1.500	1.500	1.500
<i>Estado</i>	1.500	1.500	1.500	0.750	0.300
<i>Cuota Total</i>	6.000	6.000	6.000	6.000	6.000

Fuente: Rabasa, Emilio. El IMSS: Salud, seguridad y solidaridad social.

1.1 Ley del Seguro Social de 1973

La Republica Mexicana en los años setenta experimentaba con el proceso de la Industrialización una transformación de su población, pasando de una rural a una más urbanizada, y por ende crecía el trabajo urbano asalariado, existiendo una población más preparada. Dentro del proceso de transformación, comenzaba a incrementarse el mercado laboral y en particular el sector femenino, a pesar de ser escasa su participación laboral y de los problemas de discriminación a los que se enfrentaba. Pero siendo aun muy significativos la cantidad de trabajadores de las comunidades rurales, era necesario incluirlos dentro del sistema de seguridad social. Por ello era primordial extender la cobertura y ampliar los servicios del IMSS (IMSS Aportaciones al debate, 1997).

Dada las distintas condiciones sociales y económicas de las regiones del país, se elabora la primera reforma a la Ley del Seguro Social publicándose el 12 de marzo de 1973 en el Diario Oficial de la Federación (DOF). En la Ley de Seguro Social de 1973 (LSS73) se ampliaron los beneficios del régimen obligatorio y se inició servicios en los municipios en que aún no operaba el IMSS. Ruezga (2009) menciona que se amplía la cobertura, ya que se obligaba a asegurar a los trabajadores de industrias familiares y demás trabajadores no asalariados, así como a ejidatarios, comuneros, colonos y pequeños propietarios organizados. Se estableció el régimen voluntario. Se anexa la prestación del servicio de guardería, esto promovería que más mujeres se vieran motivadas a ingresar al mercado laboral, pudiendo contribuir en el ingreso del hogar, estando tranquilas en el cuidado de sus hijos. Con ello se pretendía mejorar la protección de los trabajadores asegurados y extenderla a los grupos no sujetos a una relación de trabajo.

Al implementar mecanismos para ampliar la cobertura se pretendía que trabajadores que no contaban con protección social, comenzarán progresivamente a afiliarse, sin embargo los mecanismos no cumplieron el objetivo ya que fueron incapaces de conducirlos a su afiliación, y se comenzó a desarrollar y a constituir aceleradamente lo que hoy se conoce como la economía informal (IMSS Evidencias para el debate, 1997).

1.2 Ley del Seguro Social de 1997

Un paso antes de la reforma al esquema de pensiones fue la creación del Sistema de Ahorro para el Retiro (SAR), donde el patrón depositaría el 2% de salario base de cotización que iría a un fondo de retiro, más una cotización del 5% que irían al fondo de vivienda. Con la reforma al esquema de pensiones este método de financiamiento cambiaría. Además que el SAR atravesó por varios cambios.

Bajo la administración del presidente Ernesto Zedillo se realizó un estudio que mostró los problemas financieros del IMSS, así como un deterioro en las prestaciones – particularmente en pensiones -, además se observó que los mecanismos para promover el aseguramiento eran obsoletos. Entre 1990 y 1995 se experimentó un estancamiento en la afiliación por lo que era necesario nuevos mecanismos para reactivarlo. Por otro lado el Instituto tenía serios problemas en la administración de la información y de los recursos. Contaba con bases de datos que contenían las semanas y salarios cotizados, pero estas bases no eran confiables dado que presentaban muchos errores. Estos errores en algunos casos desembocaban en un reconocimiento insuficiente de los derechos del trabajador (IMSS Evidencias para el debate, 1997; IMSS, 1999). Al conjugarse estos elementos, conducían a confundir al trabajador en sus derechos y obligaciones, y tuvieran una perspectiva poco clara y turbia del funcionamiento de nuestro esquema de protección que brindaba el estado a los trabajadores a través del IMSS. Más aún factores demográficos como el crecimiento en la esperanza de vida, la disminución de la tasa de mortalidad y un descenso en la tasa de fertilidad fueron argumentos agregados para reformar el esquema de pensiones, pensando en el sostenimiento en el corto y largo plazo del instituto.

En 1995 se envió al Congreso una propuesta para una nueva Ley de Seguridad Social, dentro de sus principales objetivos el IMSS (1999) menciona que “(a) incentivara la afiliación, disminuirá la “economía informal” y elevará la productividad del trabajo, además de que mejoraría la equidad en la contribución al introducir instrumentos para combatir la subdeclaración de ingresos mediante la disminución de los impuestos marginales al trabajo; [...]” además de buscar darle una viabilidad financiera al instituto, así como promover nuevos mecanismos que impulsaran el desarrollo y crecimiento de la seguridad social.

En México la nueva Ley del Seguro Social entró en vigor el 1 de julio de 1997, fue una reforma estructural al sistema de pensiones, dado que el método de financiamiento pasó de un esquema de Beneficio Definido (BD), donde las aportaciones de todos los trabajadores cotizantes

se constituía un fondo para financiar las pensiones, a uno de Cuentas Individuales (CD), donde cada trabajador mediante cotizaciones obligatorias y aportaciones voluntarias constituye su propio fondo de pensiones.

El seguro de invalidez, vejez, cesantía en edad avanzada y muerte (IVCM) de la LSS73, fue segmentada en la reforma en seguros de retiro, cesantía en edad avanzada y vejez (RCV), en invalidez y vida (IV) y en gastos médicos para pensionados. La Ley del Seguro Social de 1997 (LSS97), indica que cada cuenta individual para el retiro está constituida por dos subcuentas, el seguro de RCV y la subcuenta de vivienda, ambas son administradas por Administradoras de Fondos para el Retiro (AFORE). El trabajador tiene una tercera cuenta destinada para las aportaciones voluntarias. El seguro de IV es administrada por el Instituto Mexicano del Seguro Social (IMSS), pero la prestación es otorgada por una compañía de seguros.

Dada la reforma de pensiones, los trabajadores afiliados al IMSS están agrupados en *nuevos trabajadores* y en *trabajadores en transición*. Los primeros, se afiliaron a partir del 1 de julio de 1997, por tanto sus prestaciones son otorgadas bajo la LSS97, contribuyendo al sistema de cuentas individuales donde su pensión dependerá de su fondo. Los segundos que comenzaron contribuyeron en el esquema de BD bajo la LSS73, desde 1 de julio de 1997 contribuyen en su cuenta individual; sin embargo tienen la opción de elegir entre los beneficios otorgados por la LSS73 o la LSS97, estos trabajadores elegirán retirarse bajo el sistema BD o CD, siendo obligación de Instituto calcular el importe de la pensión para cada uno de los regímenes, a efecto de que el trabajador pueda elegir el que más le convenga². El saldo acumulado de agosto del 1992 a julio de 1997, del INFONAVIT y SAR, serán entregados al retiro a los *trabajadores en transición*, independientemente del plan de pensiones elegido. Una vez que obtienen los derechos para jubilarse, sí eligen retirarse bajo el esquema de CD, podrán elegir entre retiros programados o bien contratar una anualidad con una aseguradora.

La Pensión Mínima Garantizada (PMG) se otorga a los trabajadores nuevos y en transición, cuando sus ahorros sean menores a la Pensión Mínima³ (PM). El Estado contribuirá con la diferencia entre la PM y el fondo acumulado de cuenta individual. Para proteger a las pensiones de la inflación la PM se indexó al INPC, asegurando un ingreso para los trabajadores que haya cotizado con bajos salarios.

En cuanto a la administración de recursos, se sanearon las finanzas del instituto ya que el Gobierno sería el responsable de fondear el pasivo generado bajo la legislación anterior, además se crearon nuevos mecanismos para contabilizar el gasto corriente y los pasivos generados por la reserva de cada ramo de aseguramiento, entre otras acciones⁴.

² Ley de Seguro Social de 1997, Artículo cuarto transitorio.

³ Un salario mínimo mensual para el Distrito Federal.

⁴ Se detalla más el tema en el Libro "Perspectivas para el Debate" Capítulo III. La Política Presupuestal.

En el nuevo régimen, se garantiza la portabilidad de sus recursos y la elección del plan de retiro; se conservan las contribuciones realizadas en su cuenta individual del trabajador si éste deja de laborar en el sector formal de la economía, así como el acceso a las prestaciones medicas (750 contribuciones); se relaciona el monto de la pensión con el historial laboral; y se contribuye a una mayor transparencia en el numero de cotizaciones y recursos acumulados por el trabajador (IMSS Aportaciones para el debate, 1997).

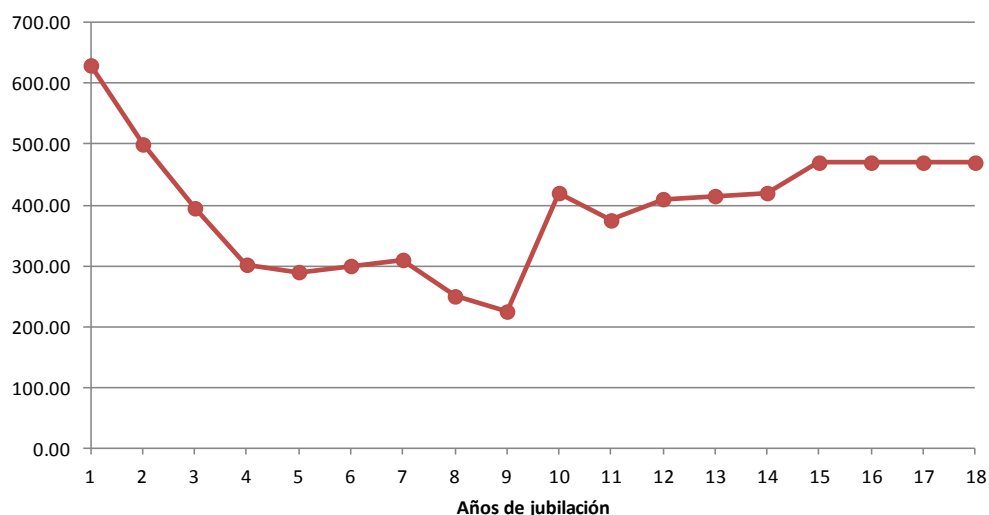
Hasta antes de la reforma de la LSS97, las deducciones a la nomina del trabajador así como las aportaciones del gobierno y patrón para contribuir a la seguridad social se fue modificando la participación de cada uno. Con la nueva reforma se ha incrementado la participación del Gobierno (después de 1960, la contribución del estado estaba decreciendo). Actualmente la participación del Gobierno incluye financiar las pensiones en curso de pago y del reconocimiento de derechos de pensiones generados bajo la ley anterior, los gastos de transición y la cuota social.

La inflación juega un papel importante dado que deprecia el valor de las pensiones y las reservas, el IMSS (Evidencias para el debate, 1997) muestra como a menudo el valor real de la pensión no se mantenía en el tiempo deteriorándolo al paso del tiempo, perjudicándose en mayor proporción a generaciones que han vivido etapas inflacionarias, el ejemplo se da para un trabajador que se pensionó en el periodo de los ochentas (Gráfica 1.1).

El actual esquema de pensiones garantiza los derechos adquiridos de los trabajadores que cotizaron antes del 1 de enero de 1997. Todos los trabajadores del anterior régimen y del actual, obligatoriamente deben de contar con una cuenta individual donde se irán depositando sus aportaciones. Los trabajadores que adquirieron derechos en el anterior régimen se divide en dos: a) pensionados; y b) la Generación de Transición. Los que comenzaron a cotizar a partir del 1 de enero de 1997 es la Generación Nueva.

A través de su fondo de pensiones del trabajador, iniciará su participación en instrumentos de inversión, con la intención de generar rendimientos y con ello incrementar su saldo total de su cuenta. Las AFORES que son las encargadas de administrar las cuentas individuales de los trabajadores transfieren los recursos a las Sociedades de Inversión Especializadas en Fondos para el Retiro (SIEFORES), estas a su vez invierten los fondos de pensiones. El movimiento de los flujos permite el desarrollo y competencia en el mercado financiero. La asignación de una SIEFORE dependerá de la edad del cotizante.

Gráfica 1. 1
Valor real de la pensión promedio para los jubilados
de 1980



Fuente: IMSS, Evidencias para el debate (1997).

1.2.1 Análisis de los cambios más relevantes al esquema de pensiones en la Ley del seguro social de 1997

Seguro de Riesgo de Trabajo (SRT)

Una de las modificaciones que se realizaron en la reforma, es en el cálculo de la prima pagada por las empresas para cubrir el SRT, ya que ahora tendría una relación directa con su siniestralidad. Una disminución en la prima se daría si se implementan medidas en la prevención de riesgos de trabajo e higiene laboral. Este elemento no era considerado en el anterior régimen, ya que la prima estaba en relación a la industria a la que pertenecía la empresa. El patrón se vería beneficiado bajo el actual esquema, ahorrando el diferencial de prima entre el anterior y actual régimen, así mismo el trabajador se sentiría más seguro en su lugar de trabajo.

La administración financiera de SRT es similar a la del seguro de invalidez y vida, como se verá más adelante (IMSS, 1999).

Seguro de invalidez y vida (SIV)

En la LSS73, para otorgar la pensión del seguro de invalidez y muerte (SIM), se necesitaba haber cotizado 150 semanas y tener una pérdida del ingreso del 50% -la cobertura es por riesgo común-, y el pago de la pensión es por invalidez parcial o total.

Con este esquema menciona el IMSS (*Evidencias para el debate*, 1997) que se observó que cerca del 70% de los beneficiarios de esta pensión no alcanzaron ni a cotizar 1000 semanas y el 33% cotizó menos de 400 semanas. En 1984, países miembros de la Organización para la Cooperación y el Desarrollo Económico (OCDE) destinaban un gasto para el pago de pensiones de invalidez de entre el 10 y 30 por ciento del total de pensiones. En México en ese mismo año destinaba –Pensiones directas⁵– alrededor del 40 por ciento del total de las pensiones, este porcentaje se fue incrementando año tras año. El resultado de la forma de financiamiento de este seguro, llevó a que hasta antes de la reforma, más de la mitad del déficit actuarial del IMSS fuese por el pago de pensiones de invalidez.

Una de las modificaciones que se realizaron en la reforma para contar con más recursos para financiar este ramo fue incrementar el número de semanas de cotización a 250, además que se le asignó una mayor participación al trabajador ligando su ingreso con el monto de la pensión, así mismo contarán con una protección económica (mediante el seguro de sobrevivencia) sus beneficiarios⁶ cuando el trabajador fallezca, en adición se incluye la participación del gobierno en caso que el fondo del trabajador sea insuficiente para adquirir la pensión mínima. Sin embargo el IMSS (1997) apunta que la medida, que generaría un mayor impacto para reducir este pasivo, sería la reducción de los siniestros que deriven en la invalidez.

Una de las razones por lo que el IMSS actualmente administra el Seguro de Invalidez y Vida (SIV) es porque la prima que pagarían los trabajadores a una aseguradora serían diferenciadas por la actividad que desarrollan, esto es, un trabajador de altos ingresos, por su nivel de educación comúnmente no realiza trabajos que debiliten su capacidad física o mental, donde su nivel de exposición al riesgo es menor, cosa contraria ocurre con trabajadores de bajos ingresos, por tanto sería más cara la prima del trabajador de bajos ingresos. El SIV será financiado con el saldo de la cuenta individual del trabajador más una suma asegurada entregada por el Instituto, ambas conformarían el monto constitutivo, cuyos recursos serían enviados a una aseguradora para que ésta administre la reserva para el pago de la pensión.

Seguro de cesantía en edad avanzada y vejez (SCV)

En la LSS73 para otorgar la pensión de vejez, el trabajador tenía que cotizar 500 semanas y tener 65 años de edad.

El cálculo de la pensión de vejez del trabajador en la LSS73 es de la siguiente manera:

⁵ No incluye pensiones por sobrevivencia.

⁶ Los beneficiarios son el cónyuge, hijos y padres dependientes económicamente.

$BD^7: \%CB (X) + IA (E)$

Donde:

$\%CB$ (Porcentaje de la cuantía básica) = Promedio de las últimas 250 semanas de cotización sobre SMGDF⁸

X = Promedio de las últimas 250 semanas de cotización

$\%IA$ (Porcentaje de Incrementos Anuales) = Promedio de las últimas 250 semanas de cotización sobre SMGDF

E (Exceso de cotización) = Por cada 52 semanas en exceso de 250

El porcentaje de cuantía básica e incrementos son tomados conforme a la tabla del artículo 167 de la LSS73.

En la LSS97, para tener derecho a una pensión a la edad de retiro, se necesita haber cotizado 1250 semanas y tener 65 años. Si sólo se llega a cotizar 750 semanas únicamente podrá gozar de las prestaciones en especie del seguro de enfermedades y maternidad. Uno de los cambios importantes que se dio en el SCV, fue que el monto de las cotizaciones del trabajador no se perdería, si éste no llegara a juntar las semanas requeridas para pensión de retiro, ya que ahora tiene la opción de retirar su saldo de su cuenta individual. Esta opción no era posible en la LSS73

En la LSS73 para otorgar la pensión de cesantía, se tenía que cotizar 500 semanas, tener 60 años de edad y no emplearse de nuevo. Si el trabajador no contaba con las cotizaciones mínimas a la edad de retiro perdía las contribuciones realizadas, quedándose el trabajador en el desamparo económico y médico, sin reconocimiento alguno por haber contribuido a la seguridad social (IMSS, 1997). La diferencia en el cálculo de la pensión de cesantía respecto de la de vejez, es en la cuantía de la pensión, que se expresa como un porcentaje de la cuantía de la pensión de vejez, que varía de acuerdo a los años cumplidos a partir de los 60.

Considerar sólo las últimas doscientos cincuenta semanas de contribución para el cálculo de la pensión, traía consecuencias para trabajadores que reingresaban al sector formal con un sueldo inferior al último percibido en este mismo sector, como puede ser el caso de trabajadores de edad adulta o mujeres que pudieron separarse de su empleo por atender labores del hogar. Claramente esto reducía su pensión por lo que no veían incentivos para reingresar al sector formal y optarían por continuar o integrarse al sector informal.

Una de las medidas que se implementó para evitar que el trabajador o patrono realizara una subdeclaración o evasión respecto a las cotizaciones, fue ligar el monto de la pensión con base a toda la carrera salarial, además que con ello se redistribuirían las pensiones. La redistribución

⁷ Beneficio definido: Un esquema de pensiones donde las contribuciones que se aportan durante la carrera laboral se depositan en un fondo y los beneficio generados se retiran de acuerdo a formulas basadas en el último periodo de contribución, sin un vinculo directo con las tasas de interés.

⁸ Salario Mínimo General para el Distrito Federal.

indican Grandolini y Cerda (1998) consiste en el tiempo de cotización, ya en el régimen anterior, trabajadores de bajos salarios que contribuían alrededor de 20 años financiaban las pensiones de aquellos que sólo cotizaban 10 años con altos salarios. Así mismo al considerar toda su historia laboral para el cálculo de su pensión se contribuirá a premiar el ahorro.

Bajo estos elementos el IMSS (Evidencias para el debate, 1997) mencionaba que el anterior esquema desincentiva al ahorro y el reingreso al sector formal, puesto que si el trabajador estimaba que no llegaría a cubrir las semanas requeridas para el goce de una pensión y que perdería sus cotizaciones realizadas, o que se vería disminuido el monto de la pensión si ingresa al sector formal con un salario bajo al que venía percibiendo, se vería motivado a seguir o buscar una opción de trabajo en el sector informal.

Aunque visto desde otro ángulo, si el trabajador considera que el beneficio al que será acreedor al momento del retiro, no corresponde con las aportaciones hechas durante su carrera laboral, esto podría actuar como un desincentivo para trabajar en el sector formal, sobre todo si se considera que el tiempo de espera para recibir su pensión al retiro aumento de 10 años (LSS73) a 25 años (LSS97) (Hernández, 2001). Cotizar 25 años puede ser un horizonte lejano considerando la escasa oferta laboral y la alta frecuencia del empleo temporal. El ahorro es cuestionable cuando existe una subcontratación de los trabajadores.

La LSS97 el Estado asegura una pensión mínima garantizada una vez que se hayan cumplido los requisitos de elegibilidad. Este cambio en la ley beneficia a los trabajadores con bajos salarios, ya que les garantiza una pensión de al menos un salario mínimo general. En el momento que entro en vigor esta ley la pensión mínima garantizada se actualiza conforme al Índice Nacional de Precios al Consumidor (INPC), garantizando el poder adquisitivo de dicha pensión.

Para finalizar, el Cuadro 1.2 sintetiza la información de los cambios más importantes que se dieron en la pensión de retiro de la LSS73 respecto de la LSS97. En este se ingresa información adicional sobre el porcentaje de las contribuciones sobre cada seguro.

Cuota Social

El gobierno a través de la “Cuota Social” realiza aportaciones a la cuenta individual de cada trabajador, la cuota está ligada al salario base de cotización del trabajador. Un trabajador que gana el equivalente a diez salarios mínimos de 2012, la cuota representa el 0.52 por ciento de su salario, mientras que un trabajador que gana el equivalente a un salario mínimo representa el 6.21 por ciento de su salario. Sin importar el nivel salarial, todos los trabajadores se ven beneficiados por la cuota que aporta el gobierno, pero se ven proporcionalmente más beneficiados aquellos que cotizan con salarios bajos. La Cuota Social ayuda a elevar el ahorro de la cuenta individual del trabajador; contribuye al alcance de la pensión mínima; y esta indexado al INPC.

Cuadro 1. 2
Cambios en la reforma al esquema de pensiones de vejez

Área	Sistema Anterior IVCM	Sistema Nuevo (RCV + IV)
A. INSTITUCIONES RESPONSABLES		
Cesantía y Vejez (RCV)	IMSS	<ul style="list-style-type: none"> Trabajador selecciona el fondo de Pensiones (AFORE) Si está en el nuevo sistema los fondos provienen de las AFORES, o de IMSS si elige retirarse bajo el régimen de reparto (solo para las generaciones de transición)
Invalidez y Seguro de Vida (IV)	IMSS	IMSS
B. CONTRIBUCIONES (% salario)	15.5	16.5 - 21.0 (promedio 17.5)
Cesantía y Vejez (RCV)	11.5	12.5 - 17.0 (=11.5 + Cuota social): 13.5 promedio por trabajador
Invalidez y Vida (IV)	4.5	4
	IVCM (IMSS): 8.5 Cesantía y Vejez 3.0 Invalidez y Muerte 3.0 Reserva para pensionados en salud 1.5 Gasto de administración 0.6 Asistencia social 0.4 SAR- Retiro (Banco de México) 2.0 SAR - Vivienda (INFONAVIT) 5.0	Cuenta Individual de RCV: AFORE 6.5 + Cuota social Cesantía y Vejez 4.5 Sub-cuenta de retiro 2 Cuota social (por día) 5.5% por un salario mínimo INFONAVIT 5 IV para IMSS: 4 Invalidez y Vida 2.5 Reserva para los pensionados en salud 1.5
C. REQUISITOS DE ELEGIBILIDAD		
Vejez	500 semanas (10 años) y 65 años de edad	1,250 semanas (25 años) y 65 años
Cesantía	500 semanas (10 años) y 60 años de edad	1,250 semanas y 60 años
Seguros de invalidez	150 semanas, pérdida del ingreso del 50%	250 semanas
Seguros de vida	150 semanas	250 semanas
D. BENEFICIOS		
Vejez: Monto de la pensión	(i) Bajo el esquema de reparto (% del promedio de los salarios de los últimos 5 años más una fracción por cada año en exceso de 10, con un máximo del salario promedio; indexado a los cambios en el salario mínimo). + (ii) contribuciones acumuladas de INFONAVIT (iii) retiro de la sub-cuenta 5/92 - 8/97	Nuevos trabajadores: Saldo acumulado de la cuenta individual desde 1/09/1997; Trabajadores en transición: Elige la pensión mas alta entre: (i) beneficio actual (ver columna anterior) y (ii) saldo de la cuenta individual (AFORE + INFONAVIT) desde 1/09/1997 + saldo acumulado de la subcuenta de retiro 5/92 - 8/92 (si aún están en el Banco de México)
Vejez: Retiro anticipado		<ul style="list-style-type: none"> Retiro temporales de la cuenta individual de la AFORE o Compra una anualidad en una compañía de seguros
Seguro de invalidez: Monto de la pensión	Depende del grado de la invalidez. Se otorga por el IMSS: % del promedio de los salarios de los últimos 5 años más una fracción por cada año en exceso de 10, con un máximo del salario promedio.	<ul style="list-style-type: none"> El grado de invalidez es determinado por el IMSS La pensión se otorga por una compañía de seguros: 35% del salario promedio (en términos reales) de las últimas 500 semanas de contribución
Seguro de vida: Monto de la pensión	Viuda: 90% de la pensión de invalidez; Dependientes: 20% de la pensión de invalidez o 30% si ambos padres fallecen	Mismo beneficio que para el seguro de invalidez y vida
Pensión mínima garantizada (PMG)	Equivale a un salario mínimo legal indexado al actual salario mínimo	Equivale a un salario mínimo legal de 1/07/1997 indexado al INPC

Fuente: Grandolini y Cerda. The 1997 pension reform in Mexico: Genesis and design feature, 1998.

La “Cuota Social” que estimula el ahorro del trabajador, es canalizada a su cuenta individual implicando un incremento del ahorro interno del país, y como ya se mencionó los recursos del fondo de pensiones son invertidos, generando un desarrollo de productos de inversión de largo plazo.

De acuerdo con el IMSS (1997) la Cuota Social está enfocada a resolver dos problemas fundamentales, independientes de la propuesta. Fiscalmente, ¿cuál es la mejor manera de financiar el ingreso de personas relativamente pobres? Laboralmente, ¿cuál es el diseño de programas que minimizan el efecto negativo sobre el ahorro de las familias? Todos los trabajadores jóvenes y viejos se ven beneficiados con la aportación del gobierno mediante la Cuota social. Ayuda a evitar que haya una subdeclaración de salarios ya que esto juega en contra del trabajador al recibir una cuota inferior a la que realmente es acreedor.

Desempleo

El trabajador que cumpla con ciertas condiciones, puede acceder a una ayuda por desempleo. Está medida implementada en la reforma donde el trabajador puede disponer de recursos de su cuenta individual, si desafortunadamente se queda temporalmente sin empleo. Podrá disponer del mínimo que resulte entre el 10% del monto total y 75 días de salario de cotización cada 5 años.

No obstante este mecanismo de financiamiento afecta a todos los trabajadores pero más aún a los que cotizan con bajos salario, la afectación recae en una disminución en el ahorro de su cuenta individual y en sus semanas de cotización. Lo último es interesante a destacar ya que si el trabajador no cumple con las 1250 semanas no tendrán derecho a una pensión mínima, destinando al trabajador a la pobreza. A pesar de que la ley contempla la posibilidad de reintegrar las aportaciones que haya tomado de su cuenta individual para restableces sus semanas de cotización, posiblemente muy pocos trabajadores lo harán.

Guardería

Si bien es cierto que el servicio de guardería ya se contemplaba a partir de la Ley del Seguros Social de 1973, el IMSS (Evidencias para el debate, 1997) mencionaba que los recursos que de este ramo eran destinados para otros cubrir pasivos de otros ramos. Es por ello que en la LSS97 se incluyeron mejoras contables para que los recursos que se generan en este ramo sean destinados propiamente al servicio de guardería y no para otros fines.

En una sociedad donde se vuelve más común que el padre y madre trabajen, es necesario que se mejore e incremente el servicio de guardería, es así como esta mejora en la administración de recursos puede ser un incentivo para trabajadores del sector formal, y para que trabajadores del sector informal vean los beneficios a los que pueden tener derecho con LSS97. Esta prestación

sería la más socorrida por mujeres jefas de familia, que tengan hijos en edad de asistir a una guardería, sintiéndose seguras del cuidado de sus hijos mientras ellas se van a trabajar.

1.3 Afiliación

En esta sección se analizará la evolución de la cobertura de seguridad social, tocando primeramente la cobertura en general de todas las instituciones que la brindan, posteriormente se revisará la cobertura que prevé el IMSS. En las gráficas y cuadros que se presentan, se utilizan datos oficiales de las *Memorias Estadísticas del IMSS*, así como datos del sitio web del INEGI, en su sección de *Consulta Interactiva de datos* o en su caso se toman datos de *documentos publicados* por el mismo instituto.

En la Gráfica 1.2 se muestra un panorama general de la evolución en la cobertura de la seguridad social considerando todos los esquemas de protección -con excepción del seguro popular⁹-, donde se incluyen trabajadores asegurados y pensionados, así como a sus familiares dependientes. Muestra la contribución del IMSS (la institución de salud más importante en términos de cobertura) en la protección social de la población, donde tuvo un crecimiento sostenido hasta década de los noventas, a partir de ahí a la fecha, ha tenido variaciones negativas y positivas. Lo ideal es que mantuviera un crecimiento continuo en cuanto a la afiliación, puesto que la cobertura está estrechamente ligada con la situación del empleo en el país. Pero como se menciono anteriormente, el incremento de la cobertura hasta 1990 fue debido a la transformación de la sociedad. Por un lado fue el proceso de pasar de una población rural a una más urbanizada y por otro un proceso de industrialización del país.

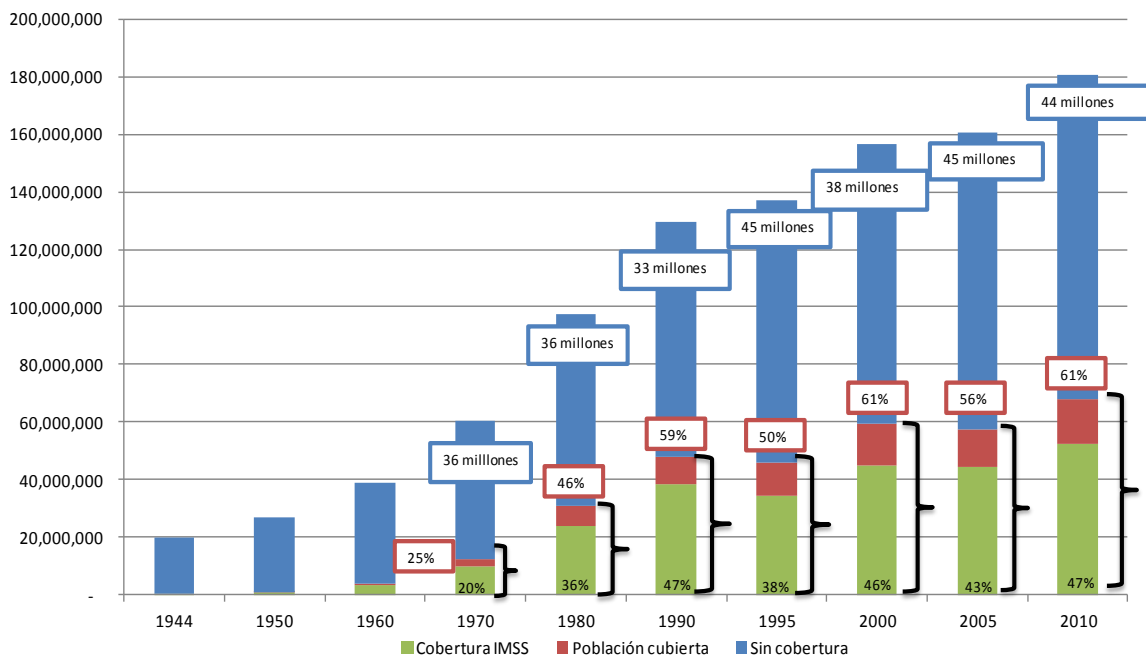
Pero no hay que dejar de señalar que los mecanismos para promover que el trabajador cuente con un esquema de protección han sido insuficientes, dado que desde 1970 a la fecha se ha excluido de la seguridad social una población importante. La población que no cuenta con protección social de 1970 a la fecha es del orden de los 30 millones, dicha cifra no solo se ha sostenido si no que ha crecido.

La contribución en cobertura de otras instituciones de salud (ISSSTE, PEMEX, FNM, SEDENA, SEMAR y Estatales), en conjunto siempre ha sido menor respecto a la del IMSS. Además que éstas son las que más beneficios otorgan a sus trabajadores y dependientes.

En 2010 México contaba con una población de 112 millones de la cual sólo el 61 por ciento tenía protección social, el IMSS contribuía con el 47 por ciento, sin embargo 44 millones seguían excluidos de sistema de seguridad social.

⁹ El Seguro popular de salud se creó en 2001, derivado del Plan Nacional de Salud 2001-2006

Gráfica 1. 2
Evolución de la cobertura de la seguridad social respecto a la población total
(Periodo 1944 a 2010)

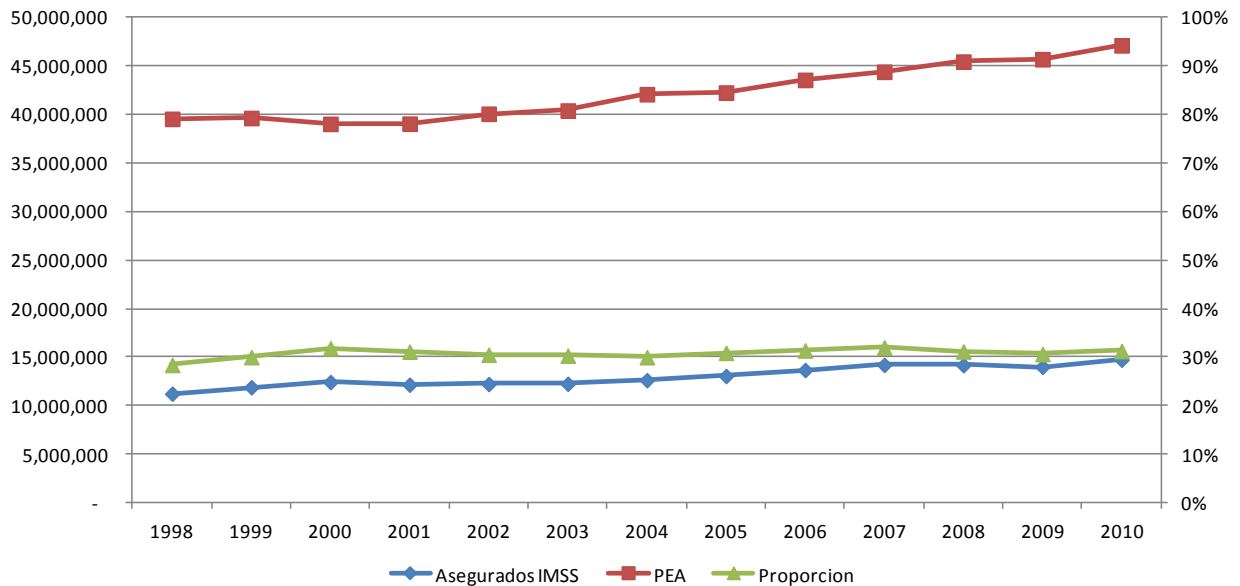


Fuente: Elaboración propia con datos de la Memoria Estadística (IMSS), Salud (INEGI), Población (INEGI).

Tomando en cuenta la proporción de trabajadores asegurados respecto de la población económicamente activa¹⁰ (PEA), de 1998 a 2010, sólo ha crecido 3 por ciento, pasando de 28 por ciento a 31 por ciento. A partir de 1999 la afiliación se ha ubicado alrededor de 31 por ciento con variaciones más menos uno. La máxima afiliación fue de 32 por ciento en 2000 y 2007. El IMSS (Evidencias para el debate, 1997) había mencionado que de 1990 a 1995 se experimentó un estancamiento en la afiliación, y la reforma consideraba dentro de sus objetivos nuevos mecanismos para reactivarlo, sin embargo la afiliación de 1999 a 2010 en general ha presentado periodos de estancamiento (Gráfica 1.3, Cuadro 1.3).

¹⁰ Como lo indica el INEGI (2002) la PEA “es el grupo de proveedores u oferentes de servicios laborables, algunos de los cuales han logrado que sus servicios sean utilizados (ocupados) mientras que otros buscan que lo sean (desocupados); es decir algunos han encontrado quien demande sus servicios mientras que otros no, aun cuando –esto es fundamental– todavía lo sigan ofreciendo. Es importante señalar aquí que un ocupado puede ser un trabajador por su cuenta o independiente y no necesariamente un trabajador subordinado y remunerado.”

Gráfica 1.3
Cobertura del IMSS de trabajadores asegurados respecto a la PEA
(1998-2010)



Fuente: Elaboración propia con datos de la Memoria Estadística 2011 (IMSS), Capítulo I Evolución del Empleo, y con datos de INEGI. Nota: Gráfica a dos escalas, la serie *Proporción* pertenece a la escala de porcentajes; la de *Asegurados IMSS* y *PEA* pertenece a la escala de cifras en millones.

Cuadro 1.3
Cobertura de seguridad social

Año	Población Total	PEA	IMSS (trabajadores afiliados)	IMSS (Cobertura total) ^a	% PEA vs Población total	% IMSS (trabajadores afiliados) vs PEA
2000	97,483,412	39,043,393	12,432,377.00	45,053,710	40.1%	31.8%
2005	103,263,388	42,274,306	13,059,129.00	44,531,666	40.9%	30.9%
2010	112,336,538	47,137,757	14,737,571.00	52,310,086	42.0%	31.3%

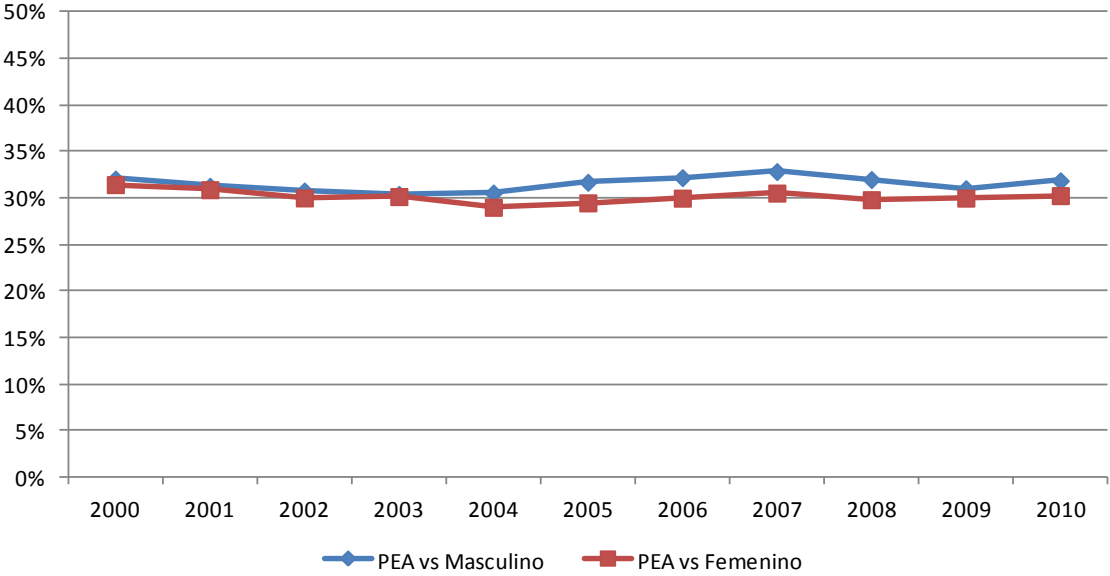
Fuente: Elaboración propia con datos de la Memoria Estadística 2011 (IMSS), capítulo I Evolución del Empleo, y con datos de INEGI

La cobertura de aseguramiento para trabajadores activos para ambos géneros, en 10 años se ha mantenido alrededor del 30 por ciento (Gráfica 1.4). Para el género masculino la máxima se dio en e 2007 con 33 por ciento y la mínima en 2003 con 30 por ciento. Para el femenino la máxima se dio en 2001, 2002 y en 2007 con 31 por ciento y la mínima en 2004 y 2005 con 29 por ciento.

La afiliación de mujeres (trabajadoras activas), pasó de 31 por ciento (año 2000) a 30 por ciento (año 2010), el decremento sucedió en un periodo de 10 años, resulta preocupante, porque si bien es cierto que la participación de la mujer dentro del sector productivo es creciente, la cobertura está teniendo un retroceso. Tomando en cuenta que la reforma dentro de los objetivos,

se encontraba promover una mayor afiliación de parte de las mujeres y otros grupos vulnerables, se observa que se ha tenido un retroceso temporal de 10 años, por lo cual está resultando insuficiente las mejoras realizadas al esquema de seguridad social. De seguir con un estancamiento o un retroceso en la afiliación, en el largo plazo se podría tener una proporción importante de mujeres destinadas a la pobreza, más se agrava el problema cuando en la actualidad más mujeres son las jefa de familia.

Gráfica 1. 4
Cobertura del IMSS por género de los trabajadores asegurados
(2000-2010)

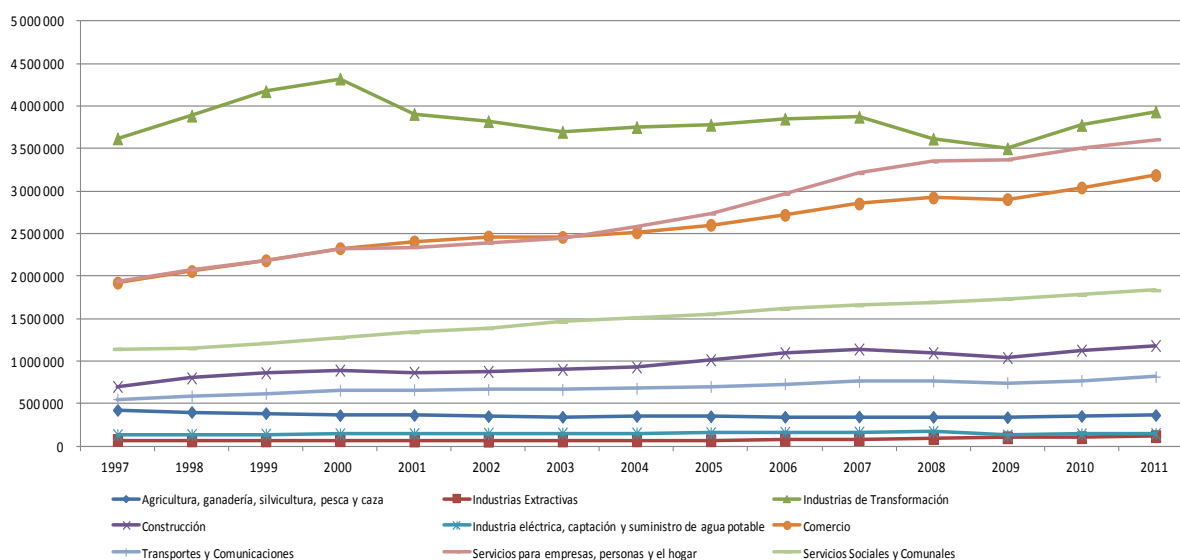


Fuente: Elaboración propia con datos de la Memoria Estadista 2011, Capítulo I Evolución del Empleo.

La Gráfica 1.5 muestra que ha pasado con la afiliación por Actividad Económica. El sector primario (Agricultura, ganadería, silvicultura, pesca y caza) ha tenido una variación negativa de 14% en la afiliación. La población que trabaja en este rubro se conoce por percibir bajos salarios y tener menor esperanza de vida (es un grupo vulnerable). La nueva LSS97 se ha visto limitada en incrementar la cobertura en este sector, dado que está experimentando una variación negativa. Ante ello Hernández (2004) señala que gran parte del déficit de la cobertura de seguridad social es de la población que vive en comunidades rurales, que en su mayoría se dedican a actividades primarias y no pueden acceder a un empleo formal, por es necesario crear políticas más amigables con estos grupos vulnerables. La proporción mínima de trabajadores de este sector, que logra realizar todas sus aportaciones, una vez que cumplan los requisitos legales, el Estado les garantiza una pensión mínima. De no cumplir con los requisitos el afiliado podrá retirar el saldo de su cuenta individual en una sola exhibición o por medio de retiros programados.

El mayor crecimiento se presenta en un rubro (Servicios para empresas, personas y el hogar) del sector terciario la cual es del 86%. La literatura indica que México está experimentando un aumento en el empleo en el rubro de Servicios, por lo que se puede decir que aquí se está generando más oferta de trabajo la cual brinda la prestación de seguridad social y por ende está contenido dentro del sector formal, mientras que otras actividades en el sector primario y secundario han experimentado caídas en el empleo (Alcaraz y García, 2006).

Gráfica 1. 5
Cobertura del IMSS por actividad económica
(1997-2011)



Fuente: Elaboración propia con datos de la Memoria Estadística 2011, Capítulo II Población Derechohabiente.

Apoyándose en la información que presenta el IMSS en su *Memoria Estadística* del Capítulo I se examina la afiliación por región¹¹, para ello se toman los datos a 2011 (Gráfica 1.6). La mayor cobertura está en la región 6 con 22 por ciento, seguido de la región 4 con 22 por ciento. La mayor concentración de afiliados corresponde a Jalisco con 8.5 por ciento y el menor es en Colima y Nayarit con 0.7 por ciento.

La región 7 que se compone sólo del Distrito Federal presenta una mayor cobertura (17 por ciento) que el total de la región 1 y 2 (16 por ciento), aunque estas últimas son equiparables con la cobertura de la región 5. Ello nos indica que regiones (1 y 2) que están compuestas por

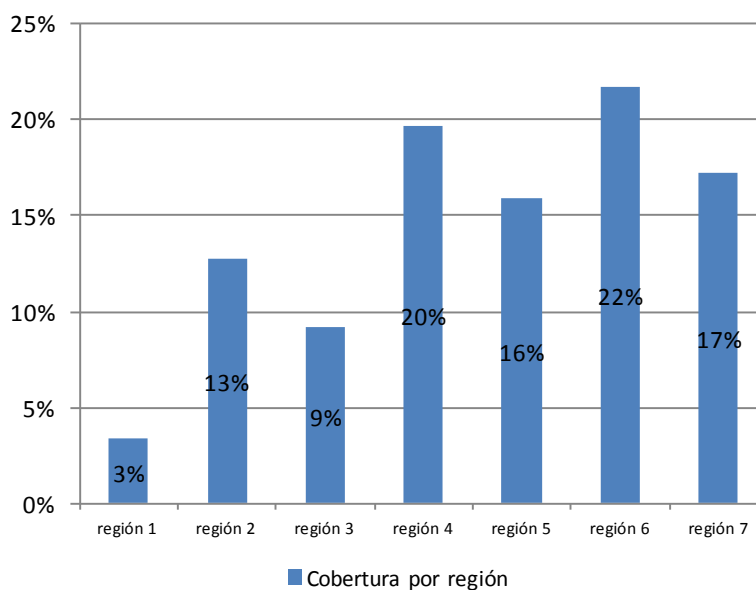
¹¹ La clasificación se apoya en la información que presenta el INEGI en el producto "Regiones socioeconómicas de México", que se puede consultar en <http://www.inegi.org.mx/est/contenidos/espanol/sistemas/regsoc/default.asp?s=est&c=11723>

estados que contienen una alta población vulnerable, tienen un déficit importante en cuanto a protección social.

En la región 3 tiene al estado que presenta la menor afiliación al seguro social con 0.5 por ciento (Tlaxcala).

En general para las siete regiones, el rango de edad de 20 a 39 años concentra la mayor afiliación siendo ésta superior al 13 por ciento sobre el total, esto es evidente ya que México es actualmente una población joven.

Gráfica 1. 6
Cobertura del IMSS por región
(2011)



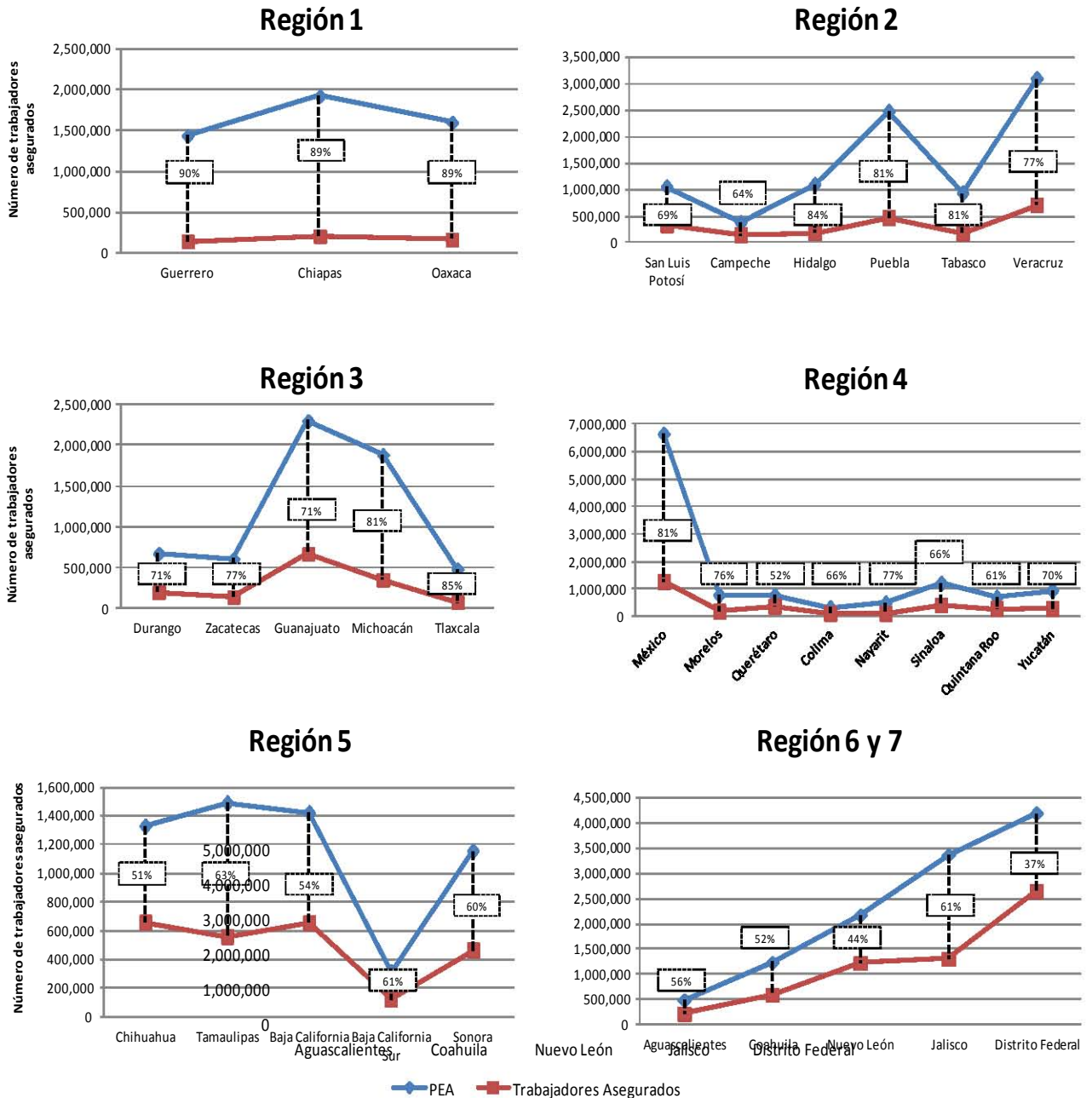
Nota: Región 1 se compone de Chiapas, Guerrero y Oaxaca. Región 2 se compone de Campeche, Hidalgo, Puebla, San Luis Potosí y Veracruz. Región 3 se compone de Durango, Guanajuato, Michoacán, Tlaxcala, Zacatecas. Región 4 se compone de Colima, México, Morelos, Nayarit, Querétaro, Quintana Roo, Sinaloa y Yucatán. Región 5 se compone de Baja California, Baja California Sur, Chihuahua, Sonora, Tamaulipas. Región 6 se compone de Aguascalientes, Coahuila, Jalisco y Nuevo León. Región 7 se compone del Distrito Federal.
Fuente: Elaboración propia con datos de la Memoria Estadista 2011, Capítulo I Evolución del Empleo.

Desagregando la información de la cobertura de las regiones por estados y comparando sólo los trabajadores asegurados que pertenece al estado respecto a la PEA del mismo, se visualiza la brecha en cuanto a cobertura (Gráfica 1.7). La PEA se toma de la Encuesta Nacional de Empleo del segundo trimestre del 2011.

En general se observa que en cada uno de los estados, más del 50 por ciento de la población que pertenece a la PEA no se encuentra afiliado al IMSS, siendo la excepción Nuevo

León (44%) y El Distrito Federal (37%). Po tanto el Distrito Federal es quien alcanza la mejor cobertura con 63 por ciento de la PEA, seguido de Nuevo León con 56 por ciento.

Gráfica 1. 7
Desagregación de la cobertura por Estado
(2011)



Fuente: Elaboración propia con datos de la Memoria Estadista 2011, Capítulo I Evolución del Empleo del IMSS y de la ENOE del segundo trimestre del 2011 del INEGI.

En los estados de la región 1 se observa que en promedio el 90 por ciento de la población que pertenece a la PEA se encuentra fuera de la protección social. Estos estados presentan el mayor déficit en la cobertura de toda la República Mexicana. El promedio de la brecha en cuanto a cobertura va disminuyendo al pasar de una región a otra.

Tomando en cuenta el promedio diario del salario base de cotización (PSBC) del Instituto (Cuadro 1.4) y la distribución de cotizantes correspondientes al salario mínimo (Gráfica 1.8) de 2011. Se tiene que en general el PSBC de la República Mexicana es de 237 pesos (Cuadro 1.4), donde el 70 por ciento de la población cotiza por debajo de este salario¹² (Gráfica 1.8). Esto quiere decir que el 70 por ciento de los trabajadores recibe un salario mensual promedio inferior a 6,895 pesos¹³, lo cual nos habla de un salario que derivará en una baja pensión. Ahora la masa de trabajadores que cotizan al instituto con un salario mínimo a 2011 y cuya tasa de remplazo será de 100 por ciento una vez que se retiren, es solo del 2.5 por ciento.

Para la región 1 que presenta el mayor déficit en cuanto a cobertura (90%), se tiene que el 77 por ciento de su población cotiza por debajo del PSBC de la República Mexicana (237 pesos). Los trabajadores de la región 7, son los que cotizan con el PSBC más alto (338 pesos) y tienen el menor déficit en cobertura (37%), esta población percibe un sueldo mensual promedio de alrededor 10, 128 pesos.

Cuadro 1. 4
Promedio diario del Salario Base de Cotización por región 2011
(Cifra en pesos)

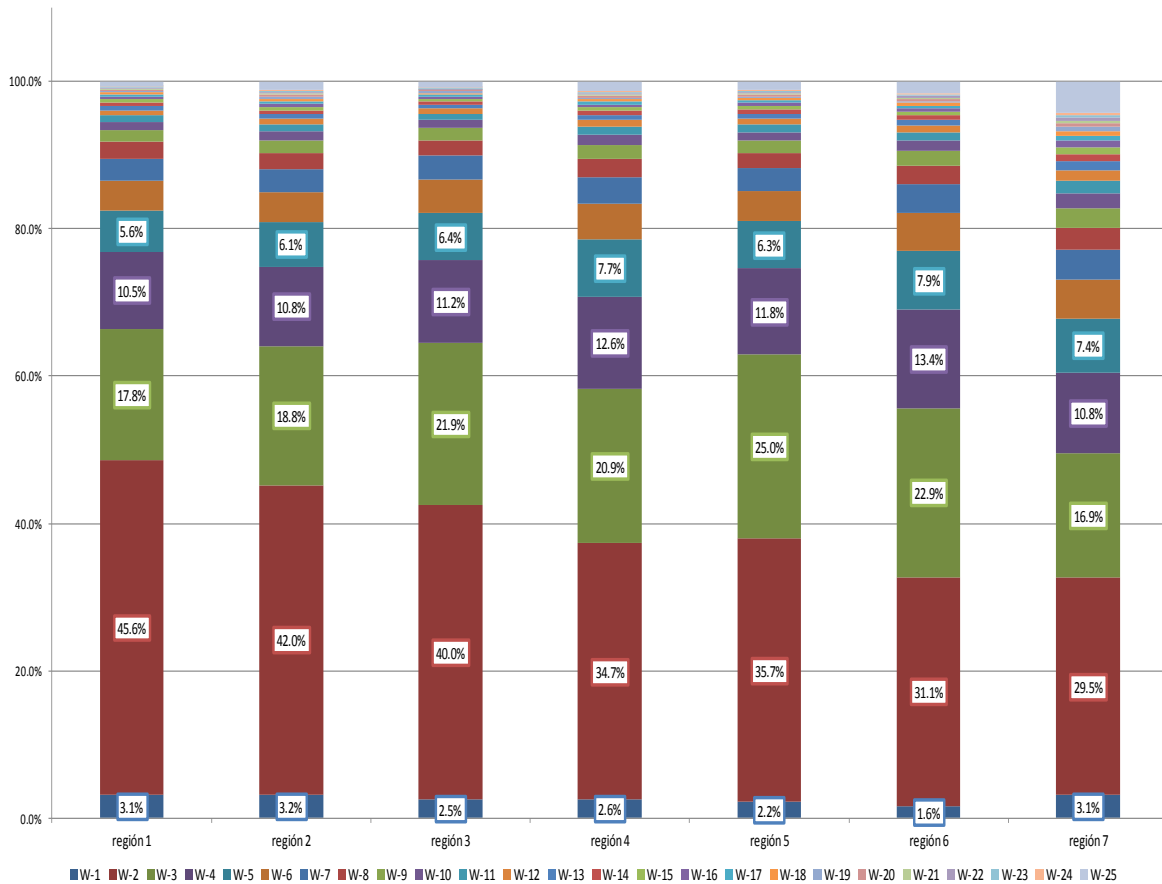
<i>Periodo</i>	<i>Promedio diario del salario base de cotización al Instituto Mexicano del Seguro Social por región Total</i>	<i>Región 1</i>	<i>Región 2</i>	<i>Región 3</i>	<i>Región 4</i>	<i>Región 5</i>	<i>Región 6</i>	<i>Región 7</i>
2011/01	237.29	203.09	230.04	201.99	222.16	227.68	238.72	337.32
2011/02	236.97	202.55	229.52	202.43	221.85	226.97	238.44	337.06
2011/03	235.76	199.66	227.97	199.74	220.78	227.61	237.43	337.10
2011/04	235.45	199.12	227.83	199.20	220.42	227.24	237.51	336.83
2011/05	238.01	202.48	230.51	201.52	223.65	228.88	239.92	339.07
2011/06	237.41	202.18	229.23	200.90	222.97	228.55	239.77	338.26
2011/07	240.59	207.07	231.60	206.53	225.55	230.44	242.21	340.73
2011/08	239.76	205.95	230.36	206.06	224.80	229.87	241.42	339.88
2011/09	236.47	200.46	227.61	201.40	221.24	227.94	239.07	337.54
2011/10	235.51	200.07	226.41	200.69	220.22	226.93	238.20	336.06
2011/11	236.46	201.43	228.91	201.23	220.77	228.12	239.27	335.49
2011/12	236.08	199.29	228.28	201.65	219.78	228.31	239.51	335.72

Fuente: Consulta interactiva de datos del INEGI.

¹² Se realiza la comparación del promedio diario del salario base de cotización respecto al salario mínimo del Distrito Federal para el 2011 (\$57.46 pesos)

¹³ Se calcula multiplicando el SMDF respecto del número de salario mínimo que recibe por mes (30 días).

Gráfica 1. 8
Distribución de cotizantes correspondiente al salario mínimo
(1 a 25)



Fuente: Elaboración propia con datos de la Memoria Estadista 2011, Población derechohabiente.

Ante la baja cobertura de protección social y un PSBC que a nivel población es bajo, surge el cuestionamiento y causa incertidumbre, de sí este nivel de cotización, donde además la capacidad de ahorro de los trabajadores es limitada, cuáles serán las tasas de remplazo que se alcanzarán a nivel población. Esta pregunta no se contestará en la presente tesis, sólo es un cuestionamiento que podría ser el principio del camino para que en futuros estudios¹⁴ se pueda dilucidar la respuesta.

En estos últimos párrafos se habla tanto del estancamiento en la afiliación, así como de una población importante de asegurados que cotiza por debajo del promedio diario del salario base de cotización, de lo último se puede decir que existe una mayoría de la población que cotiza con bajos salarios. Esto en el largo plazo puede tener consecuencias para el trabajador, ya que

¹⁴ Una aproximación en la materia se pueden consultar en “El cálculo de la tasas de remplazo bajo el modelo de aportación definida” de Haza 2007.

bajo el actual esquema de pensiones, las tasas de reemplazo son inferiores en comparación a las que se alcanzaban bajo la LSS73 así lo indica Hernández (2001), así mismo Hazas (2007) menciona que las tasas de remplazo correspondientes al ahorro individual nunca podrán superar los esquemas de financiamiento colectivos basados en el principio de solidad distributiva. Para el gobierno también tiene consecuencias, ya que es el responsable de las transferencias de recursos que tenga que realizar a favor de los trabajadores que cotizaron y no alcanzaron el derecho a un Pensión Mínima, siempre y cuando cumplan las condiciones de elegibilidad.

Se observa que a partir del último medio siglo el crecimiento acelerado de la población fue debido a diversos factores, como son las mejoras en el entorno económico, en lo social, en el acceso a las instituciones de salud, avances en la medicina. Hechos que han influido en el incremento en la esperanza de vida, en la tasa de fecundidad (después de la revolución se dio un crecimiento, actualmente ha descendido). Actualmente la población mexicana es una población joven, y cuando la población alcance la etapa de vejez existirá un riesgo social si el trabajador no cuenta con recursos que le permitan mantener su nivel de consumo, agravándose más si este no cuenta con los medios (IMSS o recursos propios) de protección para el cuidado y prevención de su salud. Con base a ello se vuelve relevante que cada vez más se vaya incrementando la cobertura en seguridad social.

CAPÍTULO II

Mercado Laboral

Introducción

Hasta este momento se ha dado una descripción de los cambios más importantes en la reforma al esquema de pensiones de la Ley del Seguro Social. Además se ha dado un panorama general de la cobertura de protección social proporcionada por el IMSS. Pero una de las vías para tener acceso a los beneficios de la Ley del Seguro Social es a través de un empleo formal, la otra vía es la afiliación voluntaria. Entonces se pasará a describir lo contrapuesto de un empleo formal, que es el empleo informal, los que conducirá a la forma en que se mide y del papel que juega frente a la protección social.

2.1 Concepto de sector informal

Definiendo primero el concepto de informalidad la cual tiene varias acepciones, los investigadores que han estudiado al mercado laboral, no han llegado a un consenso en la definición del sector informal¹⁵, ya que tres escuelas del pensamiento marcan la línea de la informalidad de manera distinta y lo definen como sigue a continuación:

“Dualistas: consideran que el sector informal es el segmento inferior de un mercado de trabajo dual, sin vínculos directos con la economía formal.

Estructuralista: consideran que está integrado por empresas pequeñas y trabajadores no registrados, subordinados a grandes empresas.

Legalistas: consideran que el sector informal comprende microempresarios que prefieren actuar de manera informal para evitar los costos asociados con el registro”. (OIT y OMC 2009, p.43).

La diferentes concepciones de informalidad se deben a distintas experiencias de cada país o región, cada quien adopta un término de informalidad que más se adapte o más explique su mercado informal.

Actualmente una nueva escuela ha consolidado una nueva definición integrando los diferentes enfoques del sector informal, la idea se basa en mercado de trabajo multisegmentado

¹⁵ Sector informal es sinónimo de economía informal, por lo que en el desarrollo de este capítulo se utilizara indistintamente.

(OIT y OMC, 2009). El nuevo enfoque utiliza como base la visión de las tres escuelas (Dual, Estructuralista y Legalista), construye el concepto de sector informal en niveles de segmento – nivel inferior, nivel intermedio y nivel superior-, cada uno integrados por diferentes tipos de unidades económicas. El segmento de nivel inferior está constituido por hogares que realizan actividades de subsistencia, con un delgado nexo con el sector formal (dualista); el segmento a nivel superior está constituido por microempresas que evaden responsabilidades tributarias (legalistas); el segmento intermedio se compone de microempresas y trabajadores subordinados de grandes empresas, que no cuentan con prestaciones laborales.

Con el nuevo enfoque que integra las diferentes visiones, ahora la discusión se centra en conocer la magnitud de cada segmento (OIT y OMC, 2009).

Ahora pasando al campo de la estadística la Conferencia Internacional de Estadísticas Laborales (CIEL) define al sector informal como:

“las empresas informales son: unidades dedicadas a la producción de bienes y servicios con el objetivo principal de generar empleo e ingreso para las personas interesadas. Estas unidades funcionan con un bajo nivel de organización, con poca o ninguna división de la mano de obra y el capital como factores de producción y en pequeña escala. Las relaciones laborales –cuando existen-se basan principalmente en el empleo ocasional, el parentesco o relaciones personales y sociales y no en arreglos contractuales con garantías formales.” (OIT y OMC 2009, p.57)

Dado que no se especifica el tamaño de la unidad económica en la definición anterior, dice Flodman Becker, que deja el concepto ambiguo y la puerta abierta para que cada país considere al empleo agrícola y a trabajadores domésticos dentro del sector informal (OIT y OMC, 2009).

El Grupo internacional de expertos en estadísticas del sector informal (Grupo Delhi) da una definición más exacta, descartando la imprecisión anterior, entonces define al sector informal como:

“las empresas privadas no constituidas en sociedad (o cuasisociedades), que producen por lo menos algunos de sus bienes y servicios para la venta o el trueque, tienen menos de cinco empleados remunerados, no están registradas y se dedican a actividades no agrícolas.” (OIT y OMC 2009, p.58).

Por su parte el Instituto Nacional de Estadística y Geografía (INEGI) (2002) adopta el concepto de economía informal o sector informal “a todas aquellas actividades económicas de mercado que operan a partir de los recursos de los hogares, pero sin constituirse como empresas con una personalidad jurídica independiente de esos hogares.” (p.52). Según lo argumentado por Flodman Becker en párrafos anteriores, en México el INEGI no considera dentro del sector informal, a los empleos cuyo modo de producción sea de subsistencia (agrícola) y al trabajador doméstico asalariado.

Como se ha visto hasta aquí las definiciones de informalidad son apoyadas sobre la idea de unidades de producción, pero también se puede definir la informalidad a nivel de empleo del trabajador, respecto a la relación laboral que tiene con su empresa o empleador, es decir, si cuentan con seguridad social o prestaciones laborales (OIT y OMC 2009).

En su caso el INEGI (2002) hace una distinción entre el empleo en el sector informal y las relaciones laborales informales respecto de sus unidades empleadoras. La primera asocia al sector informal con la conformación de unidades económicas a partir de los recursos del hogar (bajo ciertas condiciones), en tanto la segunda, que hay una relación informal cuando las unidades empleadoras (pueden ser formales) no otorgan prestaciones laborales a sus trabajadores, o bien, no tienen contrato, pero que en algún momento pueden tener todas las prestaciones de ley y ser formales.

Aquellos trabajadores que tienen una relación informal con su empleador y no están protegidos por la seguridad social se dice que es un empleo no protegido. En México éste se conforma por algunos trabajadores bajo la modalidad de confianza (trabajo por tiempo determinado, para obra determinada, o por inversión de capital determinado), de aquellos que están subcontratados por una empresa diferente para la cual prestan sus servicios u otras modalidades donde la empresa se desconecta de sus compromisos con el trabajadores, los empleados domésticos y los productores agrícolas a pequeña escala. Por tanto el empleo no protegido se agrupa en tres subconjuntos el empleo no agrícola en el sector informal, empleo no agrícola y no protegido que opera fuera del sector informal y empleo no protegido agrícola¹⁶.

No se debe dejar pasar por alto que un trabajador puede pasar de un empleo formal a un empleo informal o viceversa. Esta transición del trabajador, de un sector a otro, depende de las características institucionales de cada país, características individuales y físicas del trabajador, características específicas de las empresas y de las condiciones del mercado laboral, este mecanismo de transición se ilustra y explica en el Cuadro 2.1 .

Para finalizar se puede entender como sector formal, el complemento del sector informal, sin incluir a la economía delictiva ya que no forman parte de ninguna de éstas.

¹⁶ En la Guía de conceptos, uso e interpretación de la Estadística sobre la Fuerza Laboral en México, capítulo IV describe esta segmentación del empleo no protegido.

Cuadro 2. 1
 Mercados de trabajo multisegmentado



Nota: En el gráfico se presentan los diferentes flujos en el mercado laboral en presencia de una economía informal integrada por segmentos múltiples. Se representan los flujos siguientes: 1) transición entre el empleo formal y el nivel superior del empleo informal para evitar impuestos y reglamentaciones; 2) transición entre el empleo formal y el nivel inferior del empleo informal; 3) transición entre el nivel inferior del empleo informal y el desempleo en espera de obtener puestos en el sector formal; 4) transición entre el empleo formal y el desempleo cuando hay sistemas de prestaciones apropiados; 5) transición entre el nivel superior y el nivel inferior del empleo informal, por ejemplo debido a conocimientos adicionales; 6) transición entre diferentes segmentos del nivel inferior de la economía informal debido a un cambio de redes; 7) transición entre la inactividad y el mercado de trabajo formal; 8) transición entre la inactividad y la economía informal.

Fuente: OIT y OMC, La globalización y el empleo informal en los países en desarrollo (2009), adaptado por Gagnon (2008)

2.2 Medición del mercado laboral

Así como ha resultado difícil llegar a un consenso en la definición de sector informal, su medición se torna también complicada. Por un lado los países la cuantifican a partir de encuestas¹⁷ por periodos determinados, otros prefieren hacerlo de formas indirectas preguntando acerca de su consumo de electricidad.

Las encuestas miden al sector informal por el tamaño y producción de la unidad económica, sin alguna relación contractual con el empleado. Pero como ya se ha venido mencionando algunos países puede dejar afuera a trabajadores domésticos, a empresas formales que también tienen participación en el mercado informal y al empleo precario. Otra forma de medirlo es por las características del empleo, aquí también se presentan exclusión de grupos que no son considerados dentro de la informalidad. También pueden aportar información de la informalidad las encuestas que incluyen tópicos que miden la cobertura de los sistemas de la seguridad social. En este último método de medir el sector informal, se puede encontrar mercados

¹⁷ En México se mide el empleo a través de la Encuesta Nacional de Ocupación y Empleo (ENOE) que se levanta trimestralmente.

emergentes que no tengan un sistema de seguridad social bien desarrollado¹⁸, o si sólo se proporcionan algunas prestaciones básicas de la seguridad social; que el trabajador no tenga conocimiento de si empresa se encuentre registrada, o si el mismo esté registrado en su sistema de seguridad social. Y así como estos existen otros métodos para evaluar y medir la economía informal (OIT y OMC 2009).

Entonces dada la existencia de la diversidad de perspectivas, los países adoptan una definición de sector informal y una metodología para medirla que más describa su situación o que más se adapte dado sus recursos y limitaciones como país. Por tanto, el argumento que utiliza cada autor u organismo para definir y medir la informalidad, son igualmente aceptadas y validas para describir dicho fenómeno.

2.3 Sector informal y protección social

De principio hasta mediados del siglo XX se da el proceso y formalización del sistema de seguridad social en México y con ello la creación de empleos que contaban con la protección de la seguridad social, en aquella época la mayoría de la población era una sociedad rural, que al paso del tiempo comenzó a darse la transición, para constituir una sociedad urbana, esto llevaba a pensar que en el largo plazo la mayoría de la población estaría cubierta por la seguridad social, dejando sólo a una minoría sin protección.

Sin embargo esto no ha ocurrido, porque dentro del mercado laboral, la población se ha conducido a emplearse en el sector informal -careciendo de seguridad social-, hecho que se ha reflejado notoriamente en el último cuarto del siglo XX. En época reciente se ha mostrado una preocupación por el crecimiento del sector informal, ya que la brecha de la protección social se va ampliando cada vez más, dado que los trabajadores (que pertenecen al sector informal) y sus dependientes están siendo excluidos de algún tipo de protección social. (CISS, 2005)

Pero el buscar una oportunidad de empleo en el sector informal tiene que ver con la interacción de varios ejes, como son el sistema fiscal, la aplicación y ejecución de las leyes tributarias de las autoridades correspondientes, de la situación económica del país, de las relaciones contractuales en materia laboral y de las expectativas laborales de cada trabajador.

Aquellos individuos que deciden integrarse al sector informal, cuya percepción de ingresos sea baja por sus servicios que ofrece, en el largo plazo están condenados a la pobreza, afectando el bienestar social, es entonces cuando el Estado estará motivado para crear programas sociales

¹⁸ Por ejemplo en Guatemala se implementa una mejora en las herramientas informáticas para administrar la información y registro de los derechohabientes, se realizan inversiones en el recurso humano y en la infraestructura de los hospitales, así como la gestión de recursos para otorgar las prestaciones en dinero y salud.

complementarios dirigidos a esta clase de población, cuyo objetivo será brindarles una protección económica o de salud. Pero no se debe perder de vista que a su vez estos programas sociales (pensiones no contributivas, pagos asistenciales, programas de salud o programas alternativos) en algunos trabajadores cuya percepción de ingresos les permita contribuir al sistema de seguridad social, puede causarles incentivos para no contribuir.

Entonces por un lado se tiene a los trabajadores informales que por sus bajos ingresos no les alcanza o están limitados para aportar a la seguridad social, pero por otro lado hay trabajadores que obtienen los ingresos necesarios para contribuir y afiliarse, pero esto no entra dentro de sus necesidades prioritarias. Por esta parte la afiliación no solo se puede ver como la falta de recursos para contribuir a la seguridad social, si no que el diseño del esquema de seguridad social no es tan atractivo, como para que la afiliación sea en forma voluntaria.

Parte del poco o nulo interés de la afiliación voluntaria del trabajador, se debe a que no percibe una función de equilibrio entre lo que tendría que cotizar –en caso de estar afiliado- y lo que recibiría de beneficios, siendo éste un punto de partida para que el trabajador vea las cotizaciones como un gasto innecesario en vez de una inversión en el corto y largo plazo. En tanto no se pueda cambiar esta perspectiva del trabajador, el sector informal seguirá creciendo y la cobertura –afiliación al IMSS- seguirá decreciendo o se estancará (Hernández, 2001).

La mayoría de los trabajadores del sector formal, por la relación contractual que tienen con su patrón, en el que se incluye un salario “fijo”, se encuentra en la posición de destinar una parte de sus ingresos para contribuir a la seguridad social, y mejor aún tiene la posibilidad de recibir una pensión al retiro. Por el contrario los trabajadores del sector informal no están en la disposición para contribuir a la seguridad social y peor aún, ahorrar para su retiro. Usualmente éstos le dan prioridad a las necesidades más inmediatas como vivienda, servicios de salud, educación y alimentación para ellos y sus dependientes, destinando el mayor porcentaje de su presupuesto para cubrir estas necesidades (Martínez, Miranda, Garro, 2005).

Por lo anterior se puede presentar un grado de inequidad, de los trabajadores y empleadores del sector formal que tienen que soportar la carga de la seguridad social, a través de contribuciones o de algún otro tipo de impuestos, respecto de los del sector informal que no contribuyen a la seguridad social y directamente no pagan impuestos¹⁹, especialmente si éstos lo hacen deliberadamente. (Martínez, Miranda y Garro 2005). En México, ante esta situación, en los últimos 30 años, ha existido una disminución del gravamen en el ingreso del trabajador y de las empresas, lo cual busca beneficiarlos, evitando en la medida de lo posible una menor afectación a su ingreso²⁰. Así mismo las contribuciones obligatorias como las voluntarias del trabajador –hasta un tope de 5 salarios mínimos- están exentas de impuestos, en tanto el monto acumulado al momento del retiro se grava, ciertamente el propósito de esto es incentivar al ahorro.

¹⁹ Los que no se pagan directamente sino que lo trasladan o cobran a una tercera persona, y quien lo absorbe es el consumidor final del bien o servicio, por ejemplo el IVA.

²⁰ Otros países de América también iniciaron políticas para disminuir el impuesto al ingreso del trabajador

Martínez, Miranda y Garro (2005) indican que “los grupos más grandes de trabajadores informales en la mayoría de los países son: (i) hombres que son autoempleados o dueños de microempresas no registradas con menos de cinco trabajadores (frecuentemente estos últimos involucran a miembros de la familia del autoempleado), (ii) el asalariado informal, que trabaja en empresas sin seguro social, y (iii) mujeres, que aparecen en varias modalidades del trabajo informal como servicio domestico.” (p.47).

Los autoempleados, pueden tener un trabajo de baja productividad y bajos salarios, o bien, pueden ser trabajadores que habiendo acumulado los suficientes recursos financieros y físicos, derivado de haber trabajado en el sector formal, deciden experimentar trabajar en el sector informal inaugurando un negocio propio, por todas las ventajas o beneficios que esto le pueda traer. Este segundo tipo de autoempleado, habla de una decisión voluntaria del trabajador para moverse de sector, que ocurre alrededor de los cincuenta años; según Martínez, Miranda y Garro (2005) mencionan que hay evidencia que muestra que si existieran mecanismos flexibles de financiamiento más trabajadores serían autoempleados. Además de ellos también hay trabajadores mayores que fueron separados de su trabajo en el sector formal, que por sus características físicas y aptitudes, el sector formal ya es incapaz de acogerlos, por lo que se ven obligados a autoemplearse dentro del sector informal.

También no hay que olvidarse de los autoempleados que por sus aptitudes y características, no podrán conseguir un “buen empleo” en el sector formal que les pueda proporcionar el nivel de ingresos sobre el cual consideren que se encuentra valorado su trabajo.

Se puede decir que hay dos clases de auto-empleado, uno que es voluntario y otro involuntario. El involuntario como ante sus características busca “cualquier trabajo” que le pueda generar un nivel de ingresos suficientes como para cubrir sus gastos del día a día. Éstos trabajadores usualmente tienen un nivel de instrucción bajo y viven en condiciones de pobreza. Por las actividades que desempeñan se exponen a altos riesgos en lo salubre y en lo laboral. Por otro lado los voluntarios son los que acumulan los suficientes ingresos, y experiencia en alguna actividad, que les son recursos para abrir un negocio y ser su “propio jefe”. Éstos pueden ser al mismo tiempo, pequeños empresarios que generan empleo informal. Tomando como referencias la experiencia de otros países en el Informe de sobre la seguridad social en América Latina de la CISS (2005) mencionan que “más del 62 por ciento de los hombres autoempleados en Brasil no quieren unirse al sector formal, y en Buenos Aires sólo el 26 por ciento de los autoempleados buscan un empleo formal”. (p.51)

El autoempleado voluntario e involuntario así como sus dependientes económicos, están expuestos ante contingencias en la salud o económicas en el corto o largo plazo. En este sentido existe una posibilidad que ante estos riesgos el autoempleado voluntario, cuente con planes en salud y de financiamiento en el sector privado, en tanto los involuntarios difícilmente cuentan con este tipo de coberturas.

Las microempresas se caracterizan por tener niveles bajos de productividad y condicionan su expansión, ya que si pasan a ser grande o mediana empresa se enfrentarían con la obligación de pagar los impuestos que por ley corresponden. Por la actividad que desempeñan, posiblemente presenten una menor exposición a riesgos relacionados a la salud y a lo laboral.

Tanto microempresas como grandes empresas también contribuyen en alojar e incrementar la economía informal. La primera que por su propia naturaleza emplean a individuos cercanos a su círculo social (parentesco, relaciones personales y sociales), en tanto las segunda emplea a individuos que no les ofrece ninguna prestación social o laboral, tal vez por evitar impuestos y reglamentaciones (regulaciones excesivas) o para compensar con un mejor salario al trabajador, ante la falta de seguridad social.

Los trabajadores asalariados informales están conformados principalmente de jóvenes, particularmente de baja instrucción y pocas habilidades, donde su única opción es acogerse o acoplarse a lo que ofrece el sector informal, donde generalmente se les paga un salario bajo. Asimismo se puede encontrar mujeres con hijos donde tienen que equilibrar su tiempo para atender labores del hogar y actividades que le generen un ingreso, cosa que el trabajo formal no proporciona. Existen también jóvenes que terminan la escuela y deciden ingresar a trabajar, pero su poca o nula experiencia los conduce hacia la informalidad donde “si les va bien” reciben un salario bajo. Otro nicho de trabajo para este tipo de asalariados es laborar en un negocio familiar.

Al observar que jóvenes y mujeres que por sus bajas habilidades, o bien su disponibilidad de tiempo les impide cubrir la jornada laboral que demanda un empleo del sector formal, ven al sector informal como un refugio o una ventana de “oportunidad” para generar ingresos; así mismo los trabajadores de mayor edad que son despedidos del sector formal y no encuentran una oportunidad de empleo en este mismo sector, encontrando solo la puerta abierta a la informalidad. Ante esta situación existe la urgencia de crear programas especiales que atiendan las necesidades de estos grupos y fomenten una inclinación por el trabajo formal, para evitar que se integren al sector informal.

Debe señalarse que el Artículo 13 de la Ley de 1997 del Seguro Social abre la posibilidad para que trabajadores del sector informal se afilien voluntariamente y tengan acceso a servicios de salud y de pensiones, sin embargo la falta de promoción e información muestra que hay un bajo porcentaje de trabajadores que se interesan por esta opción. Hernández (2001) menciona que al realizar una investigación de campo, personal del IMSS mostraba desconocimientos de los requisitos para la afiliación voluntaria, por tanto el camino es difícil para la afiliación, ya que comienza desde la solicitud de información.

En el IMSS existe la posibilidad que el afiliado titular pueda incorporar a miembros de su grupo familiar²¹ para que tengan acceso al cuidado de la salud, prestaciones sociales y el goce de

²¹ Puede incorporar a su cónyuge e hijos o en su caso a sus padres siempre y cuando dependan económicamente del afiliado titular.

una pensión en caso de alguna contingencia estipulada en la ley²², teniendo un incentivo individual y familiar para el trabajador, ahora si un segundo individuo del mismo grupo familiar recibe seguro social de parte su empleo –afiliado titular-, solo tendría el incentivo individual en el mediano y largo plazo al gozar de una pensión.

Entonces las obligaciones presentes (contribuciones al sistema) del primero y segundo individuo del mismo grupo familiar son las mismas (ambos afiliados titulares), sin embargo los beneficios causan un impacto diferente, entonces en el algún momento el segundo podría considerar incorporarse al sector informal si recibe mayores ingresos que el sector formal, que podría ser el caso de jóvenes que recién se incorporan al mercado laboral, o madres de familia, o trabajadores de mayor edad. La incorporación de jóvenes -que están en edad productiva- al sector informal tiene un impacto negativo en la cobertura de la seguridad social.

Ante esta coexistencia de la protección social y el sector informal Martínez, et. al. (2005) mencionan “que se tienen que diseñar políticas específicas en al menos 4 tipos de trabajadores: (i) los que están autoempleados, quienes ganan relativamente bien (por arriba de las oportunidades que les brinda el sector formal), pero están desconcertados ante las dificultades planteadas por las regulaciones y consideran que los beneficios brindados por el sistema de salud financiado por la seguridad social no es satisfactorio; (ii) los trabajadores mayores que son desplazados (usualmente en las edades de entre 40 y 50 años), que no pueden encontrar un trabajo en la economía formal y terminan como autoempleados en una microempresa; (iii) mujeres con bajo acceso al mercado laboral en el pasado o actualmente, quienes necesitan apoyo en la forma de políticas amigables hacia la familia y de políticas que faciliten la transición hacia un empleo de tiempo completo ; y (iv) trabajadores jóvenes poco capacitados y que no poseen capital para funcionar como autoempleados.

Todos los trabajadores y sus familias están expuestos a contingencias o riesgos catastróficos que los pueden llevar a un desequilibrio en el ingreso familiar, pudiéndolos condenar a un estado de endeudamiento permanente. Estas contingencias pueden ser accidentes que debiliten su capacidad productiva o enfermedades que los deje deshabilitados para trabajar, así como problemas de salud de un integrante de la familia. La situación se vuelve gravosa para trabajadores del sector informal que no cuentan con alguna protección social que los cubra contra este tipo de riesgos. Llegar a la vejez ya se vuelve un riesgo, si no se tienen los recursos para mantener su nivel de vida, así como para solventar los gastos que son provocados por la debilidad de su salud.

Ante esta situación, los trabajadores del sector formal están protegidos por la seguridad social, tal como lo menciona el artículo 2 de la Ley del Seguro Social de 1997 que la seguridad social tiene por finalidad garantizar el derecho a la salud, la asistencia médica, la protección de los medios de subsistencia y los servicios sociales necesarios para el bienestar individual y colectivo,

²² Art.11 de la Ley del Seguro Social de 1997.

así como el otorgamiento de una pensión que en su caso y previo cumplimiento de los requisitos legales, será garantizado por el estado.

La decisión –obligada o voluntaria- de un trabajador de ingresar al sector informal, afecta sin duda uno de los objetivos de la seguridad social, que es ampliar la cobertura para llegar a todos los sectores de la sociedad. Ante el crecimiento que está experimentando el sector informal, la protección social se ha impactado negativamente, llevándola a un estancamiento y debilitamiento, este fenómeno es un serio problema para el estado y el trabajador. Para el trabajador porque se encuentra en una grave situación, ya que al no constituir un ahorro a futuro que los pueda proveer de un ingreso estable, una vez que sus capacidades físicas se han visto disminuidas para continuar trabajando, están destinados a la pobreza; se vuelve aún más grave, si se agrega que la salud de las personas adultas se va deteriorando, y estos no cuentan con protección de su salud, su futuro se torna muy grave. Y para el Estado porque de los recursos públicos debe financiar programas asistenciales, que disminuyan la situación precaria y de vulnerabilidad de los individuos que quedaron excluidos de los esquemas contributivos. En este punto queda una pregunta ¿Cómo se financiarán en el corto y largo plazo los programas sociales, dirigida a una población que no cuenta con protección y que va en crecimiento?, algunos países de América Latina y de la OCDE apuntan a señalar que la fuente de financiamiento se pueden obtener del incremento en el IVA (CISS 2005), no obstante ésta puede no ser la única opción.

Capítulo III

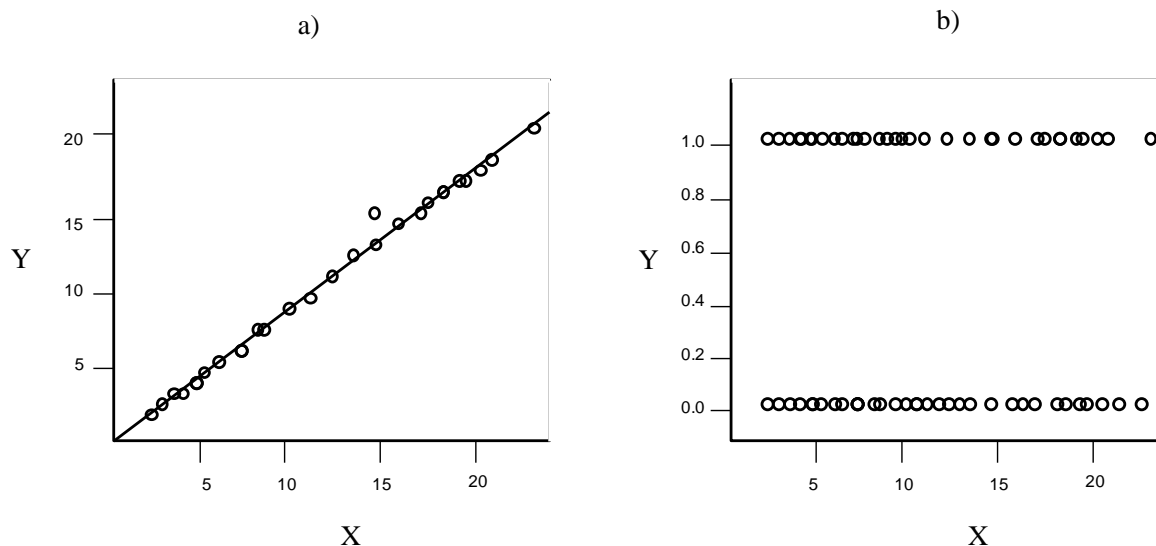
Metodología

3.1 Motivación y aplicación del modelo de selección binaria

La variable dependiente Y en los modelos de regresión lineal múltiple se supone que son de naturaleza cuantitativa, pero ahora se estudiará una variable de naturaleza *dicotómica o binaria*, que tomará los valores 0 ó 1. Algunos fenómenos que tienen esta propiedad dicotómica en la variable dependiente es cuando un individuo presenta o no algún atributo, como es el ser pobre, si cuenta con un seguro de vida, sí está empleado, etc.

Al observar en la siguiente gráfica la representación de las parejas apareadas (X_i, Y_i) , de un modelo de regresión donde la variable dependiente es cuantitativa (Gráficas 3.1a) y otro donde es dicotómica (Gráfica 3.1b), se puede notar que la primera pareja ordenada de valores puede situarse en cualquier cuadrante del plano cartesiano, en tanto la segunda sólo toma valores en el primero y segundo cuadrante. Por tanto para modelos de regresión donde la variable dependiente es una *dicotómica o binarias* se debe de buscar una función de densidad que transforme las variables.

Gráfica 3. 1
De variable dependiente cuantitativa y dicotómica



Fuente: Hosmer and Lemeshow en Applied Logistic Regression (2000)

En principio no siempre es posible aplicar directamente un modelo de regresión lineal cuando se trata de estudiar este tipo de fenómenos. Pero lo que se puede hacer es tratar de explicar la conexión que existe entre la probabilidad que adopta de pertenecer a uno de dos

grupos disponibles y los factores o valores de los atributos que influyen en su determinación. Para la construcción del modelo alternativo se seguirá un análisis similar que cuando se construye un modelo de regresión lineal.

Los tres modelos que más se utilizan para estudiar este tipo de fenómenos son:

- a) El modelo de probabilidad lineal (MPL)
- b) El modelo Logit
- c) El modelo Probit

3.1.1 Modelo de probabilidad lineal

Primero se usará un modelo de regresión que describa y caracterice al fenómeno, posteriormente se identificarán los problemas que se presentan cuando se modela de ésta manera y por último se pondrá una transformación del modelo de regresión que permita describir mejor el fenómeno.

Para introducirse en el problema se define Y_i como la variable dependiente dicotómica:

$$Y_i = \begin{cases} 1 & \text{si "El individuo se encuentra en el sector formal"} \\ 0 & \text{si "El individuo se encuentra en el sector informal"} \end{cases}$$

Con $i=1, \dots, n$ que denota al individuo

Y un vector $X = \{x_{i1}, x_{i2}, \dots, x_{ik}\}$ de variables explicativas o factores que describen fenómeno, cada elemento del vector puede ser la edad, el género, el nivel de instrucción, la actividad a la que se dedica, etc.

El modelo de regresión lineal que se presenta a continuación es llamado *modelo de probabilidad lineal* (MLP), la diferencia con respecto al modelo de regresión lineal múltiple, es que la variable de respuesta es dicotómica.

Entonces mediante el MLP se puede describir el fenómeno, donde Y_i es:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (3.1)$$

De la ecuación anterior se tienen k variables independientes, de la cual se estimarán $k+1$ parámetros. Los coeficientes β_j con $j = 1 \dots k$ muestran el aporte que X (vector de variables independientes) tiene sobre el MLP. Es decir, a modo de ejemplo, si $x =$ nivel de instrucción²³, se puede preguntar lo siguiente: qué tan influyente es el nivel de instrucción del trabajador para que labore en el sector formal o informal. Además término ε_i es el error aleatorio del modelo.

²³ Nivel de instrucción es primaria incompleta, primaria completa, secundaria completa, medio superior y superior.

En el MLP la esperanza de los errores ε_i no tienen distribución normal, la varianza no es constante y la \mathcal{Y}_i sigue una distribución de probabilidad Bernoulli, como se muestra a continuación.

El modelo que explica a la variable dependiente, dado que el i –ésimo individuo presenta ciertos atributos, es el siguiente:

$$\mathbb{E}(\mathcal{Y}_i|\mathcal{X}_i) = \beta_0 + \beta_1x_{i1} + \dots + \beta_nx_{ik} \quad (3.2)$$

Incluyendo el término de error aleatorio, se puede reescribir la ecuación (3.1) como:

$$\mathcal{Y}_i = \mathbb{E}(\mathcal{Y}_i|\mathcal{X}_i) + \varepsilon_i \quad (3.3)$$

Uno de los cuestionamientos que se puede formular es qué tanta influencia en lo particular y en su conjunto contribuyen las variables independientes para dar una explicación del fenómeno. Lo anterior predispone a encontrar la probabilidad condicional, que cuantifique, la probabilidad que un trabajador pertenezca al sector formal dado sus características. Se le asignará π_i a la probabilidad que se desconoce y de desea encontrar

$$\mathcal{P}(\mathcal{Y}_i = 1|\mathcal{X}_i = x_{ij}) = \pi_i, \quad (3.4)$$

la probabilidad de pertenecer al sector informal sería el complemento

$$\mathcal{P}(\mathcal{Y}_i = 0|\mathcal{X}_i = x_{ij}) = 1 - \pi_i. \quad (3.5)$$

Ahora si se toma la esperanza condicional se tiene que:

$$\begin{aligned} \mathbb{E}(\mathcal{Y}_i|\mathcal{X}_i) &= 0 * \mathcal{P}(\mathcal{Y}_i = 0|\mathcal{X}_i = x_{ij}) + 1 * \mathcal{P}(\mathcal{Y}_i = 1|\mathcal{X}_i = x_{ij}) = \mathcal{P}(\mathcal{Y}_i = 1|\mathcal{X}_i = x_{ij}) \\ \Rightarrow \mathbb{E}(\mathcal{Y}_i|\mathcal{X}_i) &= \pi_i = \beta_0 + \beta_1x_{i1} + \dots + \beta_nx_{ik} \end{aligned} \quad (3.6)$$

De lo anterior se tiene que la esperanza condicional es una probabilidad, por tanto debe de estar entre cero y uno.

$$0 \leq \mathbb{E}(\mathcal{Y}_i|\mathcal{X}_i) \leq 1$$

Intervalo de probabilidad

El hecho que el modelo tenga que cumplir que $0 \leq \mathbb{E}(\mathcal{Y}_i|\mathcal{X}_i) \leq 1$ no se puede garantizar que los parámetros estimados y en general el modelo $\hat{\mathcal{Y}}_i$ se encuentre en el intervalo [0,1], la primera alternativa es, una vez que se estimaron los parámetros del modelo, si se obtienen valores de $\hat{\mathcal{Y}}_i$ mayores a uno se supone que todos serán iguales a uno, y si los valores son menores a 0, se supone que serán cero, entonces la expresión está dada por:

$$\hat{\mathcal{Y}}_i = \begin{cases} 1 & \text{si } \beta_0 + \beta_1x_{i1} + \dots + \beta_nx_{ik} \geq 1 \\ 0 & \text{si } \beta_0 + \beta_1x_{i1} + \dots + \beta_nx_{ik} \leq 0 \end{cases}$$

La segunda opción es asegurar que mediante la aplicación de una transformación al modelo de regresión, se obligue a que las probabilidades condicionadas estén entre 0 y 1, esto lo hará el modelo logit.

Normalidad

Asignándole los valores que toma la variable dependiente en la ecuación (3.1) se tiene lo siguiente:

$$\text{Si } Y_i = 1 \Rightarrow 1 = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_k x_{ik} + \varepsilon_i$$

$$\Rightarrow 1 - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_j x_{ij} - \dots - \beta_k x_{ik} = \varepsilon_i$$

$$\text{Si } Y_i = 0 \Rightarrow 0 = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_k x_{ik} + \varepsilon_i$$

$$\Rightarrow \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_j x_{ij} - \dots - \beta_k x_{ik} = \varepsilon_i$$

$$\Rightarrow \mathbb{E}(Y_i | X_i) = \mathcal{P}(Y_i = 1 | X_i = x_{ij}) = \mathcal{P}(\varepsilon_i = 1 - \beta_0 - \beta_1 x_{i1} - \dots - \beta_j x_{ij} - \dots - \beta_k x_{ik} | X_i = x_{ij})$$

de la ecuación (3.4) y (3.5) se llega a que:

$$\mathcal{P}(\varepsilon_i = 1 - \beta_0 - \beta_1 x_{i1} - \dots - \beta_j x_{ij} - \dots - \beta_k x_{ik} | X_i = x_{ij}) = \pi_i \quad (3.7)$$

$$\mathcal{P}(\varepsilon_i = -\beta_0 - \beta_1 x_{i1} - \dots - \beta_j x_{ij} - \dots - \beta_k x_{ik} | X_i = x_{ij}) = 1 - \pi_i \quad (3.8)$$

Entonces la distribución de ε_i es:

$$\varepsilon_i = \begin{cases} \pi, & \text{si } \varepsilon_i = 1 - \beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{ik} \\ 1 - \pi, & \text{si } \varepsilon_i = -\beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{ik} \end{cases} \quad (3.9)$$

Se observa que los errores no tienen un comportamiento de distribución normal. El supuesto de normalidad de los errores ε_i se utiliza con el fin de realizar pruebas de hipótesis y determinar intervalos de confianza, éste supuesto no es una condición necesaria para estimar los parámetros por el método de mínimos cuadrados ordinarios (MCO), pero mientras más grande sea la muestra por el teorema de *Limite Central*, los errores ε_i tenderán a distribuirse como una normal estándar. Una vez que se calcularon estimadores por MCO, a partir del supuesto de normalidad de los errores ε_i se encontraría la distribución que siguen los parámetros y con ello se generarían los intervalos de confianza.

Varianza

Si se deja pasar por alto que $\mathbb{E}(\varepsilon_i) = 0$ y son no correlacionados, calculando la varianza de los errores se obtiene lo siguiente:

$$V(\varepsilon_i) = \mathbb{E}(\varepsilon_i^2) - \mathbb{E}^2(\varepsilon_i) = \mathbb{E}(\varepsilon_i^2)$$

$$\begin{aligned} \mathbb{V}(\varepsilon_i) &= (1 - \beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{ik})^2 \mathcal{P}(\varepsilon_i = 1 - \beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{ik}) \\ &+ (-\beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{ik})^2 \mathcal{P}(\varepsilon_i = -\beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{ik}) \end{aligned}$$

de 3.9 se sigue que:

$$\begin{aligned} \mathbb{V}(\varepsilon_i) &= (1 - \beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{ik})^2 \pi_i + (-\beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{ik})^2 (1 - \pi_i) \\ &= (1 - \pi_i)^2 \pi_i - (\pi_i)^2 (1 - \pi_i) = (1 - \pi_i) [(1 - \pi_i) \pi_i + (\pi_i)^2] \end{aligned}$$

entonces se llega a que:

$$\mathbb{V}(\varepsilon_i) = \pi_i (1 - \pi_i) \quad (3.10)$$

El hecho que la esperanza condicionada esté fuera del rango antes mencionado provocaría varianzas negativas, lo cual se debe evitar. Dos propiedades importantes que presentan los parámetros β_j cuando se estiman por el MCO, es que son insesgados y presentan la mínima varianza. Entonces al estimar los parámetros por el MCO se obtiene que los estimadores $\hat{\beta}_j$ son insesgados, $\mathbb{E}(\hat{\beta}) = \beta$, pero se llega a que no son los mejores, es decir, que no tienen la mínima varianza. Pero aun así la heterocedasticidad se puede eliminar con otras técnicas estadísticas y proseguir a estimar los parámetros de la regresión por el método de MCO.

3.1.2 Síntesis de los principales problemas de la aplicación de un modelo lineal de probabilidad con variable dependiente dicotómica

En resumen los problemas que presenta el MLP son: (i) Los errores no se distribuyen normales, (ii) y no presentan varianza constante (heterocedasticidad), (iii) \hat{Y}_i puede tomar valores afuera del intervalo cero uno y (iv) los valores de R^2 usualmente son bajos (Gujarati, 1995).

Pero estos problemas se pueden resolver parcialmente, de los cuales no se entraran a detalle y sólo se limitara a mencionar algunas técnicas. La heterocedasticidad se puede corregir por el método de cuadrados ponderados (MCP) o alguna transformación estabilizadora de varianza²⁴, para forzar que los errores se distribuyan normales al aumentar la muestra y con el método de mínimos cuadrados restringidos o técnicas computacionales hacer que \hat{Y}_i este en el intervalo cero uno.

Pero el problema consistente es que la probabilidad π_i está como función lineal de X_i , es decir, a medida que varíe X_i será en la misma proporción que variará π_i , esto se desapega de la realidad porque a valores muy altos de X_i aportará una nula probabilidad ya que el valor máximo de π_i es uno y lo mismo ocurre para valores negativos de X_i (Gujarati, 1995).

²⁴ El MCP y las transformaciones estabilizadoras de varianza se muestran en el Capítulo V del libro de "Introducción al análisis de regresión".

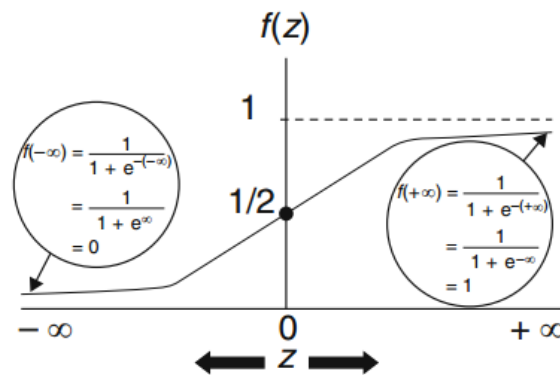
3.2 Modelo de regresión logística²⁵

Introducción

La utilización de modelos de regresión lineal plantea una serie de obstáculos cuando se trata de modelar un fenómeno donde la variable dependiente es dicotómica, pero dentro de los Modelos Lineales Generalizados (MLG) hay modelos que permiten librarlos.

Entonces se utilizará una función que transforme al MLP tal que sea una función monótona creciente (decreciente) en forma de S alargada como se muestra en la Gráfica 3.2 y que de manera natural restrinja las probabilidades de $E(Y = 1|X = x)$ y elimine el mayor obstáculo el cual es evitar que la probabilidad π_i esté como función lineal de X_i (Hosmer y Lemeshow 2000).

Gráfica 3. 2
Función logística



Rango: $0 \leq f(z) \leq 1$

$0 \leq \text{probabilidad} \leq 1$ (riesgo individual)

Fuente: Kleinbaum D. G., Klein M.. *Logistic Regression*.

La función que se utilizará y que cumple lo anterior es la regresión logit que se define como:

$$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}} = \frac{1}{1+e^{-g(x)}} \quad (3. 11)$$

donde $g(x) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ es una función lineal, entonces la probabilidad que Y_i tome el valor uno dado X_i está dada por la siguiente ecuación:

$$P(Y_i = 1|X_i = x_{ij}) = \pi(x_i) = \frac{1}{1+e^{-g(x)}} \quad (3. 12)$$

²⁵ El Modelo de regresión logística también es llamado modelo logit.

A lo largo del capítulo se entenderá como $\pi_i = \pi(x_i)$, por tanto el modelo estadístico que modela la probabilidad de ocurrencia π_i de un evento, por medio de variables que explican el fenómeno quedará descrito de la siguiente manera:

$$Y_i = \pi(x_i) + \varepsilon_i \quad (3.13)$$

Nótese que $Y_i \sim \mathbb{B}(\pi_i)$ (Y_i se distribuye como una Bernoulli con parámetro π_i). De la ecuación 3.11 se observa que $g(x) \in [-\infty, +\infty]$, mientras que $\pi_i \in [0,1]$. Por otro lado π_i no es lineal respecto a β_i y X_{ij} , motivo para no usar el método de mínimos cuadrados para estimar los parámetros. Para que exista una relación lineal se divide entre $1 - \pi_i$ y posterior a ello se toma el logaritmo, tal como se muestra a continuación:

$$\frac{\pi_i}{1-\pi_i} = e^{g(x)} \Rightarrow \ln\left(\frac{\pi_i}{1-\pi_i}\right) = g(x) \quad (3.14)$$

Resumiendo las características del modelo

- a) Mientras $\pi_i \in [0,1]$ la variación de $g(x) \in [-\infty, +\infty]$
- b) El logaritmo de probabilidades es lineal respecto de $g(x)$, pero las probabilidades en si mismo no lo son.

El modelo de regresión $g(x) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ usualmente suele tener variables independientes cuya escala es nominal, ordinal de intervalo o de razón²⁶, se pone especial atención en aquellas variables que son nominales y ordinales, un ejemplo es el estado civil y nivel de instrucción respectivamente. Ya que ingresar este tipo de variables directamente en el modelo de regresión logit es incorrecto, la solución consistirá en crear tantas variables dicotómicas necesarias para que sean incluidas en el modelo. El siguiente ejemplo ilustra una forma para crear las variables dicotómicas.

Nivel de Instrucción	Diseño de variables		
	N1	N2	N3
Primaria incompleta	0	0	0
Primaria completa	1	0	0
Secundaria completa	0	1	0
Medio superior y superior	0	0	1

En la tabla anterior se ejemplifica que para una variable independiente ordinal que tiene cuatro opciones, el número de variables dicotómicas a crear son 3, entonces en general para una variable independiente x_{ij} que tiene p categorías, el número de variables dicotómicas a crear son $p - 1$ (Hosmer y Lemeshow 2000), entonces las regresión quedaría expresada como:

²⁶ En el modelo de regresión lineal múltiple también puede presentar variables independientes cuya escala sea nominal, ordinal de intervalo o de razón.

$$\varphi(x) = \beta_0 + \beta_1 x_{i1} + \dots + \sum_{l=1}^{p-1} \beta_{jl} D_{jl} + \beta_k x_{ik}$$

3.2.1 Ajuste del modelo de regresión logística

Se estimarán los parámetros de la ecuación (3.13) por el método de máxima verosimilitud, como su nombre lo indica, este método arroja el valor de los parámetros en los que se maximiza la función de distribución conjunta de \mathcal{Y}_i .

Suponiendo que \mathcal{Y}_i es una variable aleatoria independiente y como se distribuye Bernoulli entonces la función de distribución conjunta es:

$$\ell(\beta) = f(\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_n) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

aplicando el logaritmo a la función de verosimilitud:

$$L(\beta) = \ln(\ell(\beta)) = \sum_{i=1}^n y_i \ln(\pi_i) + \sum_{i=1}^n (1 - y_i) \ln(1 - \pi_i)$$

$$L(\beta) = \sum_{i=1}^n y_i \ln(\pi_i) - \sum_{i=1}^n y_i \ln(1 - \pi_i) + \sum_{i=1}^n \ln(1 - \pi_i)$$

$$L(\beta) = \sum_{i=1}^n y_i (\ln(\pi_i) - \ln(1 - \pi_i)) + \sum_{i=1}^n \ln(1 - \pi_i)$$

$$L(\beta) = \sum_{i=1}^n y_i \left(\ln \left(\frac{\pi_i}{1 - \pi_i} \right) \right) + \sum_{i=1}^n \ln(1 - \pi_i)$$

de la ecuación (3.14) se sigue que:

$$L(\beta) = \sum_{i=1}^n y_i(\varphi(x)) + \sum_{i=1}^n \ln(1 - \pi_i) = \sum_{i=1}^n y_i(\varphi(x)) - \sum_{i=1}^n (\ln(1) - \ln(1 - \pi_i))$$

$$L(\beta) = \sum_{i=1}^n y_i(\varphi(x)) - \sum_{i=1}^n \ln \left(\frac{1}{1 - \pi_i} \right)$$

sumando un cero $0 = \pi_i - \pi_i$ en el numerador del logaritmo se llega a:

$$L(\beta) = \sum_{i=1}^n y_i(\varphi(x)) - \sum_{i=1}^n \ln\left(\frac{1 + \pi_i - \pi_i}{1 - \pi_i}\right) = \sum_{i=1}^n y_i(\varphi(x)) - \sum_{i=1}^n \ln\left(\frac{1 - \pi_i}{1 - \pi_i} + \frac{\pi_i}{1 - \pi_i}\right)$$

$$L(\beta) = \sum_{i=1}^n y_i(\varphi(x)) - \sum_{i=1}^n \ln(1 + \exp(\varphi(x)))$$

como $\varphi(x) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$, tomando la derivada parcial respecto de β_0 e igualando a cero, se tiene que:

$$\frac{\partial(L(\beta))}{\partial\beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n \left(\frac{\exp(\varphi(x))}{1 + \exp(\varphi(x))}\right) = 0$$

de la ecuación (3.11) se sigue que:

$$\Rightarrow \sum_{i=1}^n y_i = \sum_{i=1}^n (\pi_i)$$

ahora tomando la derivada parcial respecto de β_j con $j \neq 0$ e igualando a cero:

$$\frac{\partial(L(\beta))}{\partial\beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n \left(\frac{\exp(\varphi(x)) x_{ij}}{1 + \exp(\varphi(x))}\right) = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n (\pi_i) x_{ij} = 0$$

$$\Rightarrow \sum_{i=1}^n x_{ij} (y_i - (\pi_i)) = 0 \text{ para } j = 1, \dots, k$$

se llega a un sistema de ecuaciones dadas por:

$$\sum_{i=1}^n y_i = \sum_{i=1}^n (\pi_i) \quad (3.15)$$

$$\sum_{i=1}^n x_{ij} (y_i - (\pi_i)) = 0 \text{ para } j = 1, \dots, k \quad (3.16)$$

de las cuales se desea encontrar los valores de los parámetros que maximizan la función.

Dado que $\pi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}$ no es lineal, por tanto no se puede despejar los parámetros β_j , por ello existen métodos numéricos que permiten estimar los parámetros del modelo. Mostrar estos métodos no será objeto de estudio de la tesis. Por lo que se utilizará el paquete estadístico STATA cuando se estimen los parámetros en nuestro caso de estudio.

Matriz de varianza y covarianzas

Tomando la segunda derivada parcial de la función de máxima verosimilitud $L(\beta)$; de la ecuación (3.16) se sigue que:

$$\begin{aligned}\frac{\partial^2(L(\beta))}{\partial \beta_j^2} &= -\sum_{i=1}^n x_{ij} \left[\frac{(1 + \exp(\varphi(x))) \exp(\varphi(x)) - \exp(\varphi(x)) \exp(\varphi(x))}{(1 + \exp(\varphi(x)))^2} \right] \\ &= -\sum_{i=1}^n x_{ij} \left[\frac{\exp(\varphi(x))}{(1 + \exp(\varphi(x)))} - \left\{ \frac{\exp(\varphi(x))}{1 + \exp(\varphi(x))} \right\}^2 \right] = -\sum_{i=1}^n x_{ij} [\pi_i - \pi_i^2]\end{aligned}$$

se llega:

$$\frac{\partial^2(L(\beta))}{\partial \beta_j^2} = -\sum_{i=1}^n x_{ij} \pi_i [1 - \pi_i]. \quad (3.17)$$

Siguiendo un procedimiento análogo se obtiene que:

$$\frac{\partial^2(L(\beta))}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^n x_{ij} x_{il} \pi_i [1 - \pi_i] \quad \text{donde } j = 1 \dots k \quad (3.18)$$

Se denota por $\mathbb{I}(B)$ una matriz de $(k+1)$ por $(k+1)$, la cual contiene los términos negativos de la ecuación (3.17) y (3.18). La matriz $\mathbb{I}(B)$ se le llama matriz de información. La varianza y covarianza de los coeficientes estimados, resultan de la inversa de la matriz de información que se denota como $Var(B) = \mathbb{I}^{-1}(B)$, el estimador de la varianza y covarianza se denotará como $\widehat{Var}(\widehat{B})$. La varianza del j -ésimo elemento de la matriz es denotada como $Var(B_j)$ y su estimador como $\widehat{Var}(\widehat{B}_j)$, en tanto la covarianza de B_j, B_l con $l \neq j$ es denotada como $Cov(B_j, B_l)$ y su estimador como $\widehat{Cov}(\widehat{B}_j, \widehat{B}_l)$.

Así que la desviación estándar estimada de B_j , se denota y calcula como:

$$\widehat{SE}(\widehat{B}_j) = [\widehat{Var}(\widehat{B}_j)]^{1/2}; j = 1 \dots k$$

La manera de representar la estimación de la varianza y covarianza en la matriz de información será $\widehat{\mathbb{I}}(\widehat{B}) = X^T V X$, donde:

$$\widehat{\mathbb{I}}(\widehat{B}) = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ x_{1k} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \widehat{\pi}_1(1 - \widehat{\pi}_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \widehat{\pi}_n(1 - \widehat{\pi}_n) \end{pmatrix} \begin{pmatrix} 1 & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ 1 & \dots & x_{nk} \end{pmatrix} = X^T V X \quad (3.19)$$

La matriz será de utilidad a la hora de mostrar en forma matricial los resultados de las pruebas de significancia y los intervalos de confianza (Hosmer y Lemeshow 2000).

3.2.2 Pruebas de significancia de los coeficientes estimados

Una vez que se estimaron los parámetros del modelo logit ahora se necesita saber si dichos parámetros son significativos estadísticamente, es decir, si existe una variación en la variable de respuesta \mathcal{Y}_i dadas la variables explicativas.

Se mencionan dos de las pruebas que muestran la significancia del parámetro en el modelo ajustado. La primera es la prueba del cociente de máxima verosimilitud y la segunda es la prueba de *Wald*. Nótese a continuación que para probar la significancia de un coeficiente, se utilizará la función de máxima verosimilitud.

Prueba de cociente de verosimilitud

La hipótesis nula consiste en que los parámetros estimados son iguales a cero versus la hipótesis alternativa que es distinta de cero:

$$\mathcal{H}_0: \beta_1 = \beta_2 \dots = \beta_k = 0 \quad vs \quad \mathcal{H}_a: \beta_j \neq 0 \quad j \in \{1, \dots, k\}$$

Esta misma hipótesis se prueba para los modelos de regresión lineal múltiple a través de la tabla de análisis de varianza, cuya construcción se da a partir del análisis de los residuos $\sum_1^n e_i^2 = \sum_1^n (y_i - \widehat{y}_i)^2$. Un hecho similar utiliza el cociente de verosimilitudes, en el cual se comparan los datos observados respecto de los datos estimados. La prueba consiste en tomar menos dos veces el logaritmo natural de la función de verosimilitud ajustada entre la verosimilitud del modelo saturado, cuya expresión es:

$$\mathcal{D} = -2 \ln \left[\frac{\text{verosimilitud de modelo ajustado}}{\text{verosimilitud del modelo saturado}} \right]$$

Rescribiendo la expresión anterior, se tiene que:

$$\mathcal{D} = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right]$$

A la estadística \mathcal{D} se le conoce como la devianza. Cuando \mathcal{D} es cercana a cero, indica que la función de verosimilitud del modelo ajustado es similar a la del modelo saturado $L(\beta) \approx L(s)$. Por tanto \mathcal{D} mide de la desviación de ambos modelos.

El interés se centra al evaluar si el conjunto de variables independientes $\{x_{i1}, x_{i2} \dots \dots x_{ik}\}$ son significativas estadísticamente en el modelo, es decir, si contribuyen a explicar una variación en $\mathcal{P}(Y = 1)$, para ello se toma la diferencia de la prueba del cociente de verosimilitud no incluyendo e incluyendo la variable explicativa en el modelo, quedando la expresión como sigue:

$$G = \mathcal{D}(\text{modelo sin la variable independiente}) - \mathcal{D}(\text{modelo con la variable independiente})$$

Sea

vms = verosimilitud de modelo ajustado sin la variable independiente

vmc = verosimilitud de modelo ajustado con la variable independiente

vm = verosimilitud del modelo saturado ,

entonces:

$$\Rightarrow G = -2\ln\left[\frac{vms}{vm}\right] - \left\{-2\ln\left[\frac{vmc}{vm}\right]\right\} = -2\left[\ln\left(\frac{vms}{vm}\right) - \ln\left(\frac{vmc}{vm}\right)\right]$$

$$\Rightarrow G = -2\ln\left[\frac{vms}{vmc}\right]$$

Para probar la hipótesis planteada se utiliza el hecho que la estadística G se distribuye aproximadamente como una Ji-cuadrada con $k - v$ grados de libertad. Se rechaza \mathcal{H}_0 con un nivel de significancia α si

$$G > \chi^2(k - v)$$

Prueba de Wald

La prueba de Wald, plantea realizar la siguiente prueba de hipótesis para cada uno de los parámetros estimados.

$$\mathcal{H}_0: \beta_j = 0 \quad vs \quad \mathcal{H}_0: \beta_j \neq 0 \quad j = 1, \dots, k$$

La prueba consiste en comparar el parámetro estimado $\hat{\beta}_j$ sobre su error estándar $\widehat{SE}(\hat{\beta}_j)$ (desviación estándar asintótica de $\hat{\beta}_j$), donde el cociente (W) se distribuye como una normal estándar. Se denota a la estadística W como:

$$W_j = \frac{\hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)}$$

donde el p -valor para dos colas es $P(|z| > W_j)$, z sigue una distribución normal estándar, entonces si $p\text{-valor} < \alpha$, se rechaza la hipótesis nula. No se discutirá a mayor detalle de los motivos que llevan a establecer la estadística W para la prueba de hipótesis²⁷.

La estadística de Wald en el caso multivariado, se obtiene de la matriz de información mencionada anteriormente, entonces la estadística queda como sigue:

$$W = \hat{B}^T [\widehat{Var}(\hat{B})]^{-1} \hat{B}$$

sustituyendo $\hat{I}(\hat{B})$, se tiene que:

²⁷ Hosmer y Lemeshow (2005) mencionan que Hauck y Donner (1977) estudian a profundidad la estadística W .

$$W = \hat{\beta}^T [X^T V X]^{-1} \hat{\beta}$$

W se distribuye como una Ji-cuadrada con $k+1$ grados de libertad.

La prueba del *cociente de verosimilitud* y la de *Wald*, indican aquellos estimadores $\hat{\beta}_j$ que resultaron ser significativos en el modelo, un cálculo adecuado sería ajustar el modelo sólo con una de las variables y compararlo con el modelo que contiene toda las variables regresoras (Hosmer y Lemeshow 2000).

Estimación de intervalo de confianza

El intervalo de confianza para los coeficientes estimados se obtiene apoyándose en la distribución de W_j , por lo tanto β_j se distribuye como:

$$\beta_j \sim N(\hat{\beta}_j, \widehat{SE}(\hat{\beta}_j)).$$

Restando beta estimada y dividiendo sobre la desviación estándar estimada, se obtiene una normal estándar:

$$\begin{aligned} \frac{\beta_j - \hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)} \sim N(0,1) &\Rightarrow P\left(-Z_{\frac{\alpha}{2}} < \frac{\beta_j - \hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)} < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha \\ \Rightarrow P\left(-Z_{\frac{\alpha}{2}} * \widehat{SE}(\hat{\beta}_j) + \hat{\beta}_j < \beta_j < Z_{\frac{\alpha}{2}} * \widehat{SE}(\hat{\beta}_j) + \hat{\beta}_j\right) &= 1 - \alpha. \end{aligned}$$

El intervalo de confianza de $100 \times (1 - \alpha)\%$ para los coeficientes es:

$$\hat{\beta}_j \pm Z_{\frac{\alpha}{2}} * \widehat{SE}(\hat{\beta}_j).$$

Ahora para estimar el intervalo de confianza para el modelo de regresión lineal en general $\widehat{g}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \dots + \widehat{\beta}_k x_{ik}$, primero se calcula la varianza:

$$\begin{aligned} Var(\widehat{g}(x)) &= Var(\widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \dots + \widehat{\beta}_k x_{ik}) \\ &= \sum_{j=0}^k x_{ij}^2 Var(\hat{\beta}_j) - 2 \sum_{j=0}^k \sum_{l=j+1}^k x_{ij} x_{il} Cov(\hat{\beta}_j, \hat{\beta}_l) \end{aligned}$$

El intervalo de confianza de $\widehat{g}(x)$ al $100 \times (1 - \alpha)\%$, es:

$$\widehat{g}(x) \pm Z_{\alpha/2} * \widehat{SE}(\widehat{g}(x))$$

(Hosmer y Lemeshow 2000).

3.2.3 Interpretación del modelo

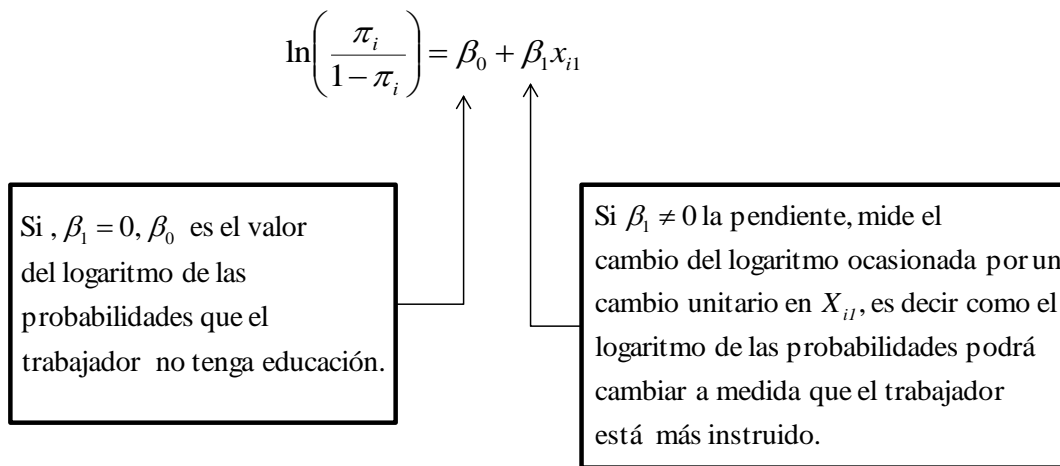
Una vez que se ajusta el modelo y se evalúan los coeficientes estimados, se procede a dar una interpretación de éstos, para conocer si dichos parámetros proporcionan información relevante del fenómeno en estudio. De la ecuación (3.14) se tiene que:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = g(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik}$$

De la ecuación anterior, resulta complicado interpretar el modelo tomando directamente los coeficientes estimados. Si se observa, la interpretación del fenómeno se expresa como un logaritmo de una probabilidad, cosa que resulta complicado de entender. Lo que se puede decir es que el cambio en el logaritmo es ocasionado por un cambio x_{ij} , quedando expresado como $g(x_{ij} + 1) - g(x_{ij}) = \beta_j$. Véase el siguiente ejemplo:

- Sea $g(x_i) = \beta_0 + \beta_1 x_{i1}$ donde
 $x_{i1} = \{\text{primaria incompleta, primaria completa, secundaria completa, medio superior y superior}\}$,

El logatimo se interpreta como:



Otro método, consiste en tomar la razón de probabilidad de ocurrencia como sigue:

$$\frac{\pi(x_i)}{1-\pi(x_i)} = e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik}}$$

La variación del coeficiente estimado $\hat{\beta}_j$ dado un incremento en x_{ij} , queda expresado en términos de un cociente de probabilidad (la razón de probabilidad que el suceso ocurra entre la probabilidad de que no ocurra, una vez que x_{ij} presenta la condición \mathcal{A}), esta expresión es llamada en el lenguaje anglosajón *odd (momio)*.

- Si el momio es igual a uno, indica que la probabilidad de éxito para la condición \mathcal{A} es tan probable como la de un fracaso.

$$P(\mathcal{Y} = 1|X = A) = P(\mathcal{Y} = 0|X = A)$$

- Si el momio es mayor que uno, indica una mayor probabilidad de éxito dada la condición \mathcal{A} , sobre la una probabilidad de fracaso.

$$P(\mathcal{Y} = 1|X = A) > P(\mathcal{Y} = 0|X = A)$$

- Si el momio es menor que uno, indica una mayor probabilidad de fracaso dada la condición \mathcal{A} , sobre la una probabilidad de éxito.

$$P(\mathcal{Y} = 1|X = A) < P(\mathcal{Y} = 0|X = A)$$

El siguiente método puede resultar más intuitivo para el investigador en la interpretación del modelo. Los coeficientes quedarán expresados en términos de la probabilidad de ocurrencia del suceso, tal como se escribe a continuación:

$$P(\mathcal{Y} = 1|X_i = x) = \pi(x_i) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})}$$

La ecuación anterior no es lineal, por lo que la probabilidad de ocurrencia del suceso, no sólo dependerá de la variación de una variable independiente si no también dependerá de los valores que pueda tomar el resto de las variables independientes. Las combinaciones posibles que se pueden realizar dado los posibles valores que pueden tomar las variables independientes, expresan distintas probabilidades de éxito.

Por último se debe de mencionar, que para interpretar el coeficiente estimado del modelo logit, usualmente se realiza a través del *cociente de momios*. El *cociente de momios* muestra el efecto del coeficiente estimado $\hat{\beta}_j$ en la variable dependiente. El *cociente de momios*, Silva y Barroso (2004) mencionan que se define como *la razón de momios correspondiente a un suceso bajo cierta condición entre el momio que corresponde al mismo suceso bajo otra condición*, y se escribe como:

$$OR = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} \quad (3.20)$$

Si $g(x_i) = \beta_0 + \beta_1 x_{i1}$, el Cuadro 3.1 muestra el cociente de momios para cada tipo de variable independiente.

Cuadro 3. 1
Cociente de momios en función del tipo de variable independiente

Valores que toma la variable independiente		$y_i = 1$	$y_i = 0$	$OR = \frac{\pi(1)}{1 - \pi(1)} \cdot \frac{\pi(0)}{1 - \pi(0)}$
x_{i1} (dicotómica)	0	$\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$	$1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$	$OR = \exp(\beta_1)$
	1	$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$	
x_{i1} (niveles)	a	$\pi(a) = \frac{\exp(\beta_0 + \beta_1 a)}{1 + \exp(\beta_0 + \beta_1 a)}$	$1 - \pi(a) = \frac{1}{1 + \exp(\beta_0 + \beta_1 a)}$	$OR = \exp(\beta_1(a - b))$ Si a-b=1, el resultado es similar al caso anterior.
	b	$\pi(b) = \frac{\exp(\beta_0 + \beta_1 b)}{1 + \exp(\beta_0 + \beta_1 b)}$	$1 - \pi(b) = \frac{1}{1 + \exp(\beta_0 + \beta_1 b)}$	
x_{i1} (continua) Sea $\Delta c = c_1 - c_2$	Δc	$\pi(a) = \frac{\exp(\beta_0 + \beta_1 c_1)}{1 + \exp(\beta_0 + \beta_1 c_1)}$	$1 - \pi(a) = \frac{1}{1 + \exp(\beta_0 + \beta_1 c_1)}$	$OR = \exp(\beta_1 \Delta c)$
		$\pi(b) = \frac{\exp(\beta_0 + \beta_1 c_2)}{1 + \exp(\beta_0 + \beta_1 c_2)}$	$1 - \pi(b) = \frac{1}{1 + \exp(\beta_0 + \beta_1 c_2)}$	

Fuente: Construido con información de Hosmer y Lemeshow. *Applied Logistic Regression*.

El *cociente de momios* experimentará un cambio, cuando la variable independiente aumente en una unidad. Si \mathcal{A} y \mathcal{B} son dos condiciones, OR toma los siguientes valores.

- Si $OR = 1$, indica que el *momio* para la condición \mathcal{A} y la condición \mathcal{B} es la misma.
- Si $OR > 1$, indica que el *momio* de la condición \mathcal{A} es mayor que para la condición \mathcal{B} .
- Si $OR < 1$, indica que el *momio* de la condición \mathcal{B} es mayor que para la condición \mathcal{A} .

Efectos marginales

El cambio de la probabilidad debida a un incremento en la variable independiente es llamado efecto marginal, y se obtiene calculando la primera derivada de la función de distribución del modelo logit, resultando la función de densidad.

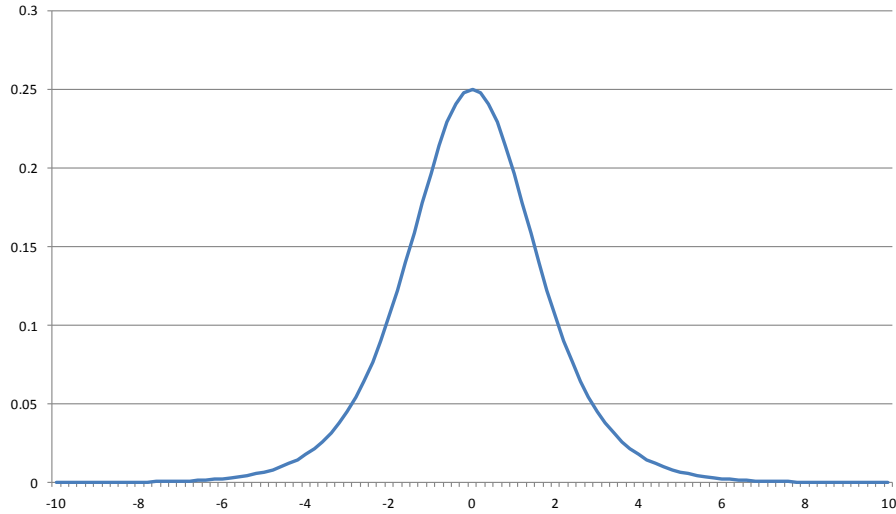
Como $\pi(x)$ es una composición de función de la forma $\mathcal{F}(g(x))$, calculando la primera se obtiene la función de densidad:

$$\frac{\partial \mathcal{F}(g(x))}{\partial g(x)} = \frac{e^{-g(x)}}{(1 + e^{-g(x)})^2} = f(g(x))$$

Además esta última ecuación expresa el efecto marginal de $g(x)$. Observando la Gráfica 3.3, la función $f(g(x))$ alcanza su máximo cuando $g(x)$ es igual a cero. Ahora el efecto marginal en un punto \mathcal{X}_j está dado por la siguiente expresión:

$$\frac{\partial \mathcal{F}(g(x))}{\partial x_j} = \left[\frac{e^{-g(x)}}{(1+e^{-g(x)})^2} \right] \beta_j = f(g(x)) \beta_j \quad (3.21)$$

Gráfica 3.3
Función de densidad de modelo logit



Fuente: Elaboración propia

El efecto marginal de \mathcal{X}_i estudia el cambio sobre la probabilidad de que ocurra el evento. La ecuación (3.21) muestra que el efecto no sólo depende del parámetro β_i , sino también del conjunto de valores que pueda tomar la función $g(x)$, por tanto el efecto marginal no es constante (para los modelos lineales si es constante). Para este caso se puede estudiar el efecto marginal, tomando la media muestral de las variables independientes $\bar{\mathcal{X}}$ y evaluarlo en $g(x)$; tomado el efecto marginal promedio de conjunto de la observaciones; o bien se elige un punto específico \mathcal{X}_0 y se evalúa en $g(x)$. Para cada caso los efectos marginales de \mathcal{X} está dado por:

$$f(g(\bar{\mathcal{X}})) \beta_j = \left[\frac{e^{-g(\bar{\mathcal{X}})}}{(1+e^{-g(\bar{\mathcal{X}})})^2} \right] \hat{\beta}_i \quad (3.22)$$

$$\frac{f(g(\mathcal{X})) \beta_j}{N} = \frac{\left[\frac{e^{-g(\mathcal{X})}}{(1+e^{-g(\mathcal{X})})^2} \right] \hat{\beta}_i}{N} \quad (3.23)$$

$$f(g(\mathcal{X}_0)) \beta_j = \left[\frac{e^{-g(\mathcal{X}_0)}}{(1+e^{-g(\mathcal{X}_0)})^2} \right] \hat{\beta}_i \quad (3.24)$$

3.2.4 Bondad de Ajuste

Sean valores observados (x_i, y_i) , con $i = 1, \dots, n$ y los valores estimados \hat{y}_i . Se dice que un modelo está bien ajustado si la suma de las distancias entre los valores observados y_i y los estimados \hat{y}_i es pequeña, y en general la suma de todas las distancias es pequeña, una manera de realizar esta comparación es con la prueba de Hosmer-Lemeshow que se enuncia a continuación.

Prueba de Hosmer-Lemeshow

Una vez que se estimó el valor de π_i para $i = 1, \dots, n$, los valores se van a ordenar de menor a mayor y se agruparán en n_g columnas $g = 1, \dots, G$.

La hipótesis nula es:

\mathcal{H}_0 : El modelo se ajusta bien a los valores observados vs \mathcal{H}_a : No se ajusta bien

Se conocen dos métodos para construir las n_g columnas. El primero consiste en dividir la región de probabilidad en deciles. El segundo agrupa a las probabilidades estimadas $\hat{\pi}_i$ en intervalos de 0.1 en forma ascendente. La idea es construir una tabla de $2 \times G$, el 2 porque y_i toma el valor cero o uno y G el número de columnas que se crearon.

La prueba de Hosmer-Lemeshow es la siguiente:

$$\hat{C} = \sum_{g=1}^G \frac{(o_g - n_g \bar{\pi}_g)^2}{n_g \bar{\pi}_g (1 - \bar{\pi}_g)}$$

donde

n_g = es el número de individuos agrupados en la columna g^{th} ,

o_g = es el número de valores observados para los que Y toma el valor de 1 ó 0, además que c_g es el número de apareamientos observados respecto de las variables independientes, entonces la expresión es:

$$o_g = \sum_{i=1}^{c_g} y_i$$

$\bar{\pi}_g$ = es el promedio de la suma de los valores de la probabilidad estimada $\hat{\pi}_i$ que están en cada uno de los grupos, y la expresión es:

$$\bar{\pi}_g = \sum_{i=1}^{c_g} \frac{m_i \hat{\pi}_i}{n_g},$$

nótese que $\sum_i m_i = n$. Según Hosmer-Lemeshow la estadística \hat{C} se distribuye aproximadamente como una $\chi^2_{(G-2)}$. Lo que se busca en esta prueba es no rechazar la hipótesis nula.

Probar la bondad de ajuste del modelo mediante la prueba de Hosmer y Lemshow, tiene algunas desventajas, ya que la estadística \hat{C} no siempre se distribuye como una ji-cuadrada (Tsiatis (1980) citado por Silva Aycaguer y Barroso Utra (2004)). Por tal razón, es válido inspeccionar visualmente, la diferencia entre grupos de los valores observados respecto del esperado²⁸, concluyendo que el ajuste no sea “desastroso”.

Curva ROC

Dado un punto de corte (c_p) en un intervalo de cero a uno, se puede evaluar en qué proporción las probabilidades estimadas reflejan los datos observados de la variable dependiente. Al número total de predicciones, que reflejaron verdaderamente la ocurrencia del evento de interés ($Y = 1$) respecto del total del número de eventos ocurridos es llamado *Sensibilidad (Se)*; y los que reflejaron verdaderamente la no ocurrencia del evento de interés ($Y = 0$) respecto del total del número de eventos no ocurridos es llamado *Especificidad (Sp)*. La forma de cálculo de cada uno, dado el c_p es la siguiente:

Sea

$$D = \text{ocurrencia del evento de interés } (Y = 1)$$

$$\sim D = \text{no ocurrencia del evento de interés } (Y = 0)$$

y

$$\hat{\pi} > c_p \text{ la predicción del individuo } x \text{ será un evento ocurrido}$$

$$\hat{\pi} \leq c_p \text{ la predicción del individuo } x \text{ será un evento no ocurrido}$$

Se calcula *Se* y *Sp* como:

$$Se = \frac{\# \text{ de predicciones de la } D | \text{ que verdaderamente es un evento ocurrido}}{\text{total de } D}$$

$$Sp = \frac{\# \text{ de predicciones de la } \sim D | \text{ que verdaderamente es un evento no ocurrido}}{\text{total de la } \sim D}$$

Se puede acomodar lo mencionado anteriormente dentro de una *tabla de clasificación* (Cuadro 3.2) (Kleinbaum y Klein 2010), como se muestra a continuación.

²⁸ Véase el libro Regresión Logística sección 4.4 Evaluación de la Bondad del Ajuste de Silva Aycaguer y Barroso Utra (2004).

Cuadro 3. 2
Clasificación de eventos

Probabilidad de Corte	Clasificados Positivos		Clasificados Negativos		Porcentajes			
	Casos positivos (D)	Casos negativos (~D)	Casos positivos (D)	Casos negativos (~D)	Se	Sp	1-Sp (falsos positivos)	Clasificados Correctamente
0.1	46,916	26,877	1,099	61,431	97.71%	69.56%	30.44%	79.48%
...								
...								
...								
...								
1	-	-	48,015	88,308	0.00%	100.00%	0.00%	64.78%

↓
Clasificados incorrectamente

Fuente: Cuadro adaptado de Kleinbaum y Klein. *Logistic Regression*.

Para el caso general, donde se desea realizar una clasificación se utiliza la *Curva de Receiver Operating Characteristic (ROC)* (Gráfica 3.4). La curva ROC se construye graficando *sensibilidad (Se)* respecto de *uno menos su especificidad (1-Sp)*, dado una infinidad de $c_p \in (0,1)$; comúnmente a esto se le llama, la capacidad que el modelo tiene para *discriminar*, donde

$$1 - Sp = \frac{\# \text{ de predicciones de la } D \text{ que falsamente es un evento ocurrido}}{\text{Total de la } \sim D}$$

Se espera idealmente que *Se* y *Sp* sean muy próximos a 1, mientras que $1 - Sp$, debiera ser muy próximo a cero.

El área bajo la curva (AUC) es una medida de discriminación. Un modelo que proporciona una buena distinción de casos observados de los no observados, tienen la propiedad de que los casos que se clasificaron como una ocurrencia del evento tienen una mayor probabilidad de ser clasificados correctamente, que aquellos eventos no ocurridos que se clasificaron como casos ocurridos; cumpliendo la siguiente relación $Se > 1 - Sp$ (Kleinbaum y Klein 2010).

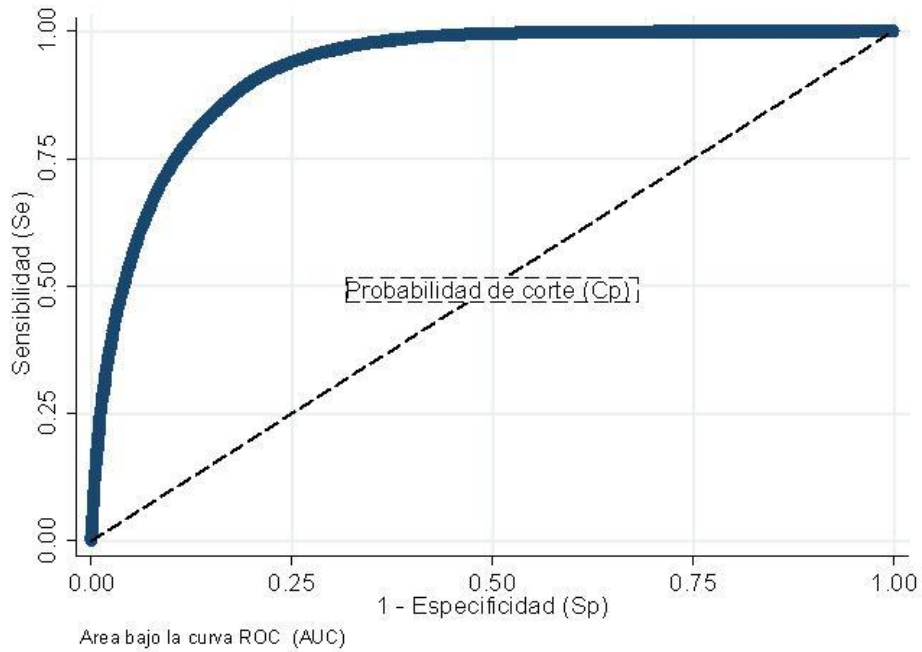
El Cuadro 3.3 muestra la guía de clasificación para valores de AUC, que son tomadas como referencia para evaluar la capacidad de discriminación de un modelo.

Cuadro 3. 3
Guía de clasificación para los valores de AUC

- 0.90 – 1.0 = Excelente discriminación (A)
- 0.80 – 0.90 = Buena discriminación (B)
- 0.70 – 0.80 = Razonable discriminación (C)
- 0.60 – 0.70 = Pobre discriminación (D)
- 0.50 – 0.60 = Fracasada discriminación (F)

Fuente: Kleinbaum y Klein. *Logistic Regression*.

Gráfica 3. 4
Curva ROC



Fuente: Elaboración propia con base de datos de la ENE de 1996

Estadístico \mathcal{R}^2

El estadístico \mathcal{R}^2 evalúa en general la capacidad de predicción de modelo²⁹. Si se tienen J configuraciones posibles y $J = n$, el estadístico se define como:

$$\mathcal{R}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{\pi}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Donde $\bar{\pi}$ es el porcentaje de las observaciones para las cuales ocurre el evento ($Y = 1$). \mathcal{R}^2 está entre 0 y 1; cuando es 1, indica que el modelo está asignado la probabilidad igual a 1 a los individuos para cuando el evento ocurre ($Y = 1$) y de misma manera asignaría la probabilidad de 0 para cuando el evento no ocurre, y si \mathcal{R}^2 se aproxima a cero, indica que las probabilidades estimadas por el modelo distan mucho de los valores observados de la variable dependiente (Y).

Por otra parte si se tiene J configuraciones posibles donde $J < n$, el coeficiente se define como:

$$\mathcal{R}^2 = 1 - \frac{\sum_{i=1}^n (y_i - m_j \hat{\pi}_i)^2}{\sum_{i=1}^n (y_i - m_j \bar{y})^2}$$

²⁹ Esta estadística no evalúa la bondad de ajuste de un modelo.

donde m_j es el número de individuos agrupados en la configuración $j = 1, \dots, J$.

Si en lugar de usar la suma de los cuadrados se utiliza la función de verosimilitud para el modelo que sólo contiene $\widehat{\beta}_0$ y se compara con el modelo completo que contiene todos los parámetros estimados $\widehat{\beta}_i$ ($i = 1, \dots, k$) además de $\widehat{\beta}_0$, \mathcal{R}^2 se expresa como:

$$R_L^2 = 1 - \frac{L_k}{L_0}$$

3.2.5 Selección del modelo

A continuación se enunciarán brevemente algunas técnicas o algoritmos para seleccionar un subconjunto de variables regresoras que “mejor” se ajustan y que mayor aportan en su descripción al modelo.

El problema de la selección de variables es analizado en un escenario donde la forma funcional de las variables independientes es totalmente conocida y que no se presentan outliers o datos que tienen mucho peso en el modelo. Evidentemente lo anterior en la práctica no ocurre, es por ello que el análisis de residuos ayudará a eliminar o reducir este tipo de problemas.

Debe entenderse que los algoritmos para la selección de variables no pueden garantizar que se obtendrá el modelo que “mejor” se ajusta a las observaciones, ya que se pueden encontrar varios modelos que pueden ser igual de buenos. El “mejor” modelo ajustado se encontrara con base a varias iteraciones.

En general se desea describir el modelo con el menor número de variables regresoras posibles, que expliquen a su vez la mayor parte de la variabilidad de \mathcal{Y} . Los modelos con menos variables se pueden interpretar mejor y a su vez son más estables y presentan menos sesgo.

Hay métodos que adicionan o eliminan un subconjunto de variables independientes del modelo completo, estimando nuevamente los coeficientes del modelo en cada iteración. El método que parte de un modelo que únicamente contiene $\widehat{\beta}_0$, y va adicionando una variable a la vez en cada iteración, con el objetivo de encontrar el subconjunto óptimo de variables independientes del cual se derive el “mejor” modelo, se llama *selección hacia adelante (forward)*. El método que va quitando variables independientes del modelo completo se llama *eliminación hacia atrás (backward)*. La combinación de ambos procedimientos es una *regresión por segmentos (stepwise)* (Montgomery, Peck y Vining 2002). Los paquetes estadísticos pueden ser una herramienta de mucha utilidad en la orientación de la elección del “mejor” modelo.

Una segunda técnica para evaluar el “mejor” modelo ajustado es a través de dos *medidas de información*. El criterio de Akaike’s Information Criterion, o An Information Criterion (AIC) y Bayesian Information Criterion (BIC), que consisten en elegir al modelo que minimice o tenga el valor más pequeño de:

$$AIC(p) = -2 \log(L) + 2p ,$$
$$BIC(p) = -2 \log(L) + p \log (N) ,$$

donde L es la función de verosimilitud del modelo; p = número de parámetros; N = tamaño de la muestra.

Capítulo IV

Las encuestas nacionales de empleo en México

Las encuestas de empleo han constituido la fuente principal de información sobre las características ocupacionales de la población, demográficas y económicas que son la base para el análisis de los aspectos laborales, además de que proveen de información relacionada con la seguridad social. Las encuestas de empleo datan desde los años setentas, donde el levantamientos de las muestras se realizaban en áreas metropolitanas, posterior a ello, llegaron a cubrir a las ciudades más importantes y por último llegaron a cubrir a todo el país.

Los cambios en las encuestas se fueron dando con base a la transformación que estaba experimentado el país, razón por la que las encuestas se han tenido que modificar, para que muestren una realidad más cercana en la que vive el país. Es así como van surgiendo varias encuestas de empleo, entre las que destacan la Encuesta Nacional de Empleo Urbano (ENEU), posteriormente se complementa con la Encuesta Nacional de Empleo (ENE), ambas se actualizan y se conjuntan para conforman lo que en la actualidad se conoce como la Encuesta Nacional de Empleo y Ocupación (ENOE)³⁰ (Cuadro 4.1).

Cuadro 4. 1
Encuestas de Empleo

Nombre	Siglas	Periodo
Encuesta Continua de Mano de Obra	ECMO	1973-1974
Encuesta Continua Sobre Ocupación	ECSO	1974-1984
Encuesta Nacional de Empleo Urbano	ENEU	1983-2003
Encuesta Nacional de Empleo	ENE	1988-2004
Encuesta Nacional de Empleo y Ocupación	ENOE	2004-actualidad

Fuente: Elaboración propia con información del INEGI de la Encuesta Nacional de Ocupación y Empleo 2005: Una nueva encuesta para México.

En general el levantamiento de campo de la muestra es como lo indica La Encuesta Nacional de Ocupación y Empleo (2005) donde menciona que la ENOE, al igual que su antecesora la ENEU y la ENE, se aplica, sobre la base de una selección de viviendas, a los miembros de un hogar. Esto requiere que el personal de campo del INEGI (entrevistadores y supervisores) acuda a cada una de las viviendas que aparecen en la muestra, identifique a los hogares y solicite una entrevista a sus residentes.

³⁰ Encuesta Nacional de Ocupación de Empleo 2005: Una nueva encuesta para México da más detalle sobre el proceso histórico.

Uno de los motivos por que es necesario la creación de una encuesta de empleo, es porque instituciones como el IMSS muestran un panorama incompleto –pero que cubre las necesidades de información del mismo instituto- del empleo, dado que no proporcionan información de la situación de la población desocupada, así como de la población no económicamente activa (PNEA), entre otras (INEGI 2002).

La encuesta muestra aspectos de la población en general de su conjunto así como sus elementos particulares de un censo de población. Su periodo de actualización es cada tres meses, el hecho que se levante en periodos cortos, es porque los elementos del empleo estudiados cambian y son muy dinámicos en el largo, mediano y corto plazo.

En el levantamiento trimestral de la muestra se aplica el cuestionario básico o ampliado. El cuestionario básico se usa en el trimestre I, III y IV, y ampliado en el II. Los temas de estudio se dividen en 11 bloques llamadas “baterías” como se muestra en Cuadro 4.2

Cuadro 4. 2
Batería

Batería	CANTIDAD DE PREGUNTAS	
	Versión ampliada	Versión básica (preliminar)
I. Condición de ocupación	6	6
II. No ocupados	9	12
III. Contexto laboral	21	12
IV. Características de la unidad económica	10	8
V. Jornada y regularidad laboral	9	7
VI. Ingresos y atención médica	5	5
VII. Trabajo secundario	5	4
VIII. Búsqueda de otro trabajo	3	2
IX. Antecedentes laborales	15	0
X. Apoyos económicos	2	0
XI. Otras actividades	1	0
TOTAL	88	36

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

La actualización de las encuestas de ocupación (ENEU-ENE), que llegó a la fecha a conformar la ENOE, se fundamenta en la versión propuesta por la OCDE, apoyándose en un marco general que la OIT sobre los estudios de empleo. Las actualizaciones propuestas sobre la ENEU-ENE, se muestran en el Cuadro 4.3, el cual ilustra la diferencia de criterios que utilizaba la OCDE respecto de las que usaba la ENEU-ENE.

La actualización de las encuestas de ocupación (ENEU-ENE), se da sobre los siguientes rubros: en el rango de edad que se considera para determinar la *desocupación abierta*, en la número de horas laboradas (para considerarse ocupado o desocupado abierto), en la percepción de ingresos por la actividad que realiza (en el sentido de oferentes-demandantes), así como aquellos que han sido contratados pero no aun no empiezan a trabajar (para considerarse

ocupado o desocupado abierto); los cambios impactan en el cálculo de la *tasa de desocupación*. Además hay un cambio en el marco muestral para mejorar la selección de la muestra

Cuadro 4. 3
Importancia de nuevos marcos conceptuales

Sub Universos	SUR – OCDE	Otras modalidades adoptadas por la OIT	ENEU – ENE México
Iniciadores	<p>No ocupados y de manera sumaria bajo la modalidad de desempleados abiertos</p> <p>↓</p> <p>Todo forman parte de la PEA</p>	<p>No ocupados</p> <p>↓</p> <p>Con búsqueda Sin búsqueda</p> <p>↓ ↓</p> <p>Desocupados Económicamente</p> <p>Abiertos inactivos</p> <p>↓ ↓</p> <p>Formados No forma parte</p> <p>parte de la PEA de la PEA</p>	<p>De manera sumaria se les considera como ocupados</p> <p>↓</p> <p>Todos forman parte de la PEA</p>
Ausentes de un trabajo sin ingresos ni nexo laboral	<p>No ocupados</p> <p>↓</p> <p>Con búsqueda Sin búsqueda</p> <p>↓ ↓</p> <p>Desocupados Económicamente</p> <p>Abiertos inactivos</p> <p>↓ ↓</p> <p>Formados No forma parte</p> <p>parte de la PEA de la PEA</p>	<p>No ocupados</p> <p>↓</p> <p>Con búsqueda Sin búsqueda</p> <p>↓ ↓</p> <p>Desocupados Económicamente</p> <p>Abiertos inactivos</p> <p>↓ ↓</p> <p>Formados No forma parte</p> <p>parte de la PEA de la PEA</p>	<p>De manera sumaria se les considera como ocupados</p> <p>↓</p> <p>Todos forman parte de la PEA</p>

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

Por mencionar algunos elementos nuevos en la ENOE, ahora capta la información de empleo relacionándola con la migración, cuantificando la edad de antigüedad del empleado, recepción de remesas, identifica la rigidez o flexibilidad laboral, cuantifica la población sindicalizada y caracteriza de manera más explícita los mercados laborales como ve en el Cuadro 4.4.

Cuadro 4. 4
Caracterización de los mercados laborales

<p>Gobierno</p> <ul style="list-style-type: none"> •Administración pública •Sector paraestatal •Escuelas, hospitales y servicios asistenciales
<p>Asociaciones civiles e instituciones sin fines de lucro</p>
<p>Empresas constituidas en sociedad y corporaciones</p> <ul style="list-style-type: none"> •Multinacionales •Cadenas •Empresas locales con personalidad moral
<p>Negocios familiares formales y empresas no constituidas en sociedad</p>
<p>Sector formal</p>
<p>Unidades domesticas (trabajo empleado en los hogares)</p>
<p>Agricultura de subsistencia</p>

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

Si bien la información de la ENE y la ENOE no es comparable, debido a la actualización y clasificación de las variables de la ENE, esto no afecta en el diseño del problema, ya que se equiparo la encuesta ENE con la ENOE, como se verá más adelante. Aun dado estos hechos sobre

la encuesta ENE, sus variables se pueden adaptar para fines de nuestro estudio a unas variables con el criterio ENOE tal cual como se indica en el manual Descripción de archivos ENE con criterio ENOE (2009).

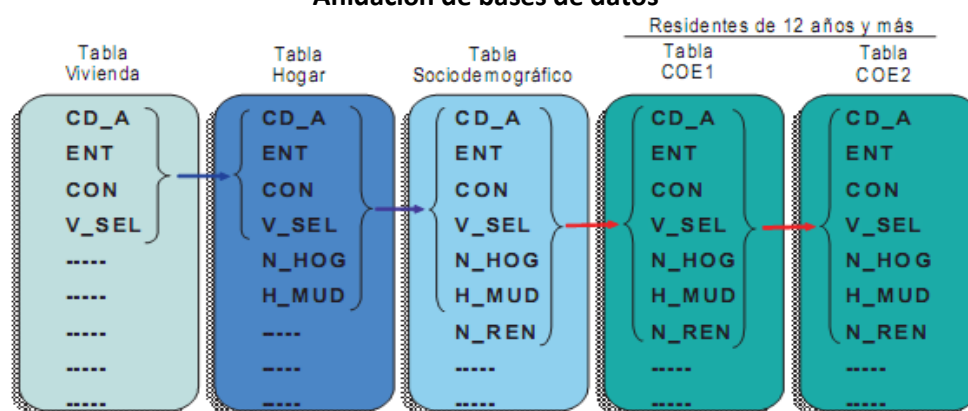
4.1 Descripción de la base de datos de la Encuesta Nacional de Empleo (ENE) y la Encuesta Nacional Ocupación de Empleo (ENOE)

A continuación se da una breve descripción de la base de datos de la ENE y ENOE, que se utilizó para nuestro estudio. Tanto la base de datos de la ENE como de ENOE se descargan de la página del INEGI, seleccionando la sección de *microdatos* de cada encuesta. El manejador de datos que se utiliza es STATA versión 12.

La base de datos de la ENE está conformada por dos tablas las cuales son: la tabla de Hogares (HOG) y la tabla de residentes³¹ (ENE). Ambas tablas se encuentran por separada, si se desea agruparlas se deben de agrupar por los campos siguientes: Área metropolitana (A_MET), Entidad (ENT), Control (CON), Vivienda seleccionada (V_SEL), Hogar (HOG) y Hogar mudado (H_MUD). La base de datos a trabajar corresponde al año de 1996³².

La base de datos de la ENOE está conformada por cuatro tablas las cuales son: la tabla de vivienda (VIV), la tabla de hogares (HOG), la tabla de sociodemográfico (SDEM) y la tabla de cuestionario y ocupación empleo I y la de cuestionario de ocupación empleo II³³. Las tablas se encuentran por separada, si se desea agruparlas se toman las variables de la base que se muestran en el Cuadro 4.5. La base de datos a trabajar corresponde al segundo trimestre del 2006

Cuadro 4.5
Anidación de bases de datos



Fuente: INEGI, Conociendo la base de datos de la ENOE (2007).

³¹ Véase para mayor detalle "Conociendo la base de datos de la ENE con criterio ENOE".

³² La ENE se levanta de forma anual en 1996 y a partir del segundo trimestre del 2000 se levanta trimestralmente.

³³ Véase "Conociendo la base de datos de la ENOE" para mayor información.

El cuadro anterior INEGI (2007) lo interpreta como: para cada vivienda existe uno o más hogares, para cada hogar existe uno o más residentes con sus características sociodemográficas, para cada residente con edad de 12 años cumplidos o más existe un cuestionario de ocupación y empleo.

La tabla VIVIENDA está compuesta por 73 campos

La tabla HOGAR está compuesta por 108 campos

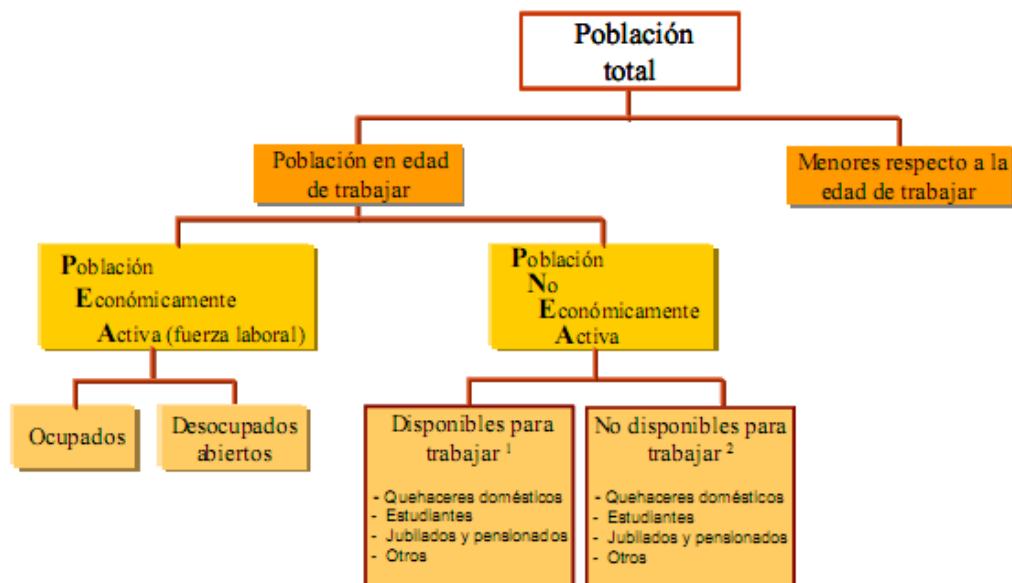
La tabla SOCIODEMOGRÁFICO está compuesta por 115 campos

La tabla CUES. OCUPACIÓN Y EMPLEO está compuesta por 188 campos

La tabla CUES. OCUPACIÓN Y EMPLEO está compuesta por 115 campos

Para tener una idea de la composición interna de la base de datos de la ENE y ENOE así como de su contenido, téngase en mente que en una encuesta de ocupación, la población total se divide en los individuos en edad de trabajar (según la OCDE mayor e igual a 14 años) y en los que no tiene la edad para ingresar a trabajar (menores de 14 años). La población en edad de trabajar se compone de la PEA y de la PNEA y cada una se compone de grupos con determinadas características, tal como se ejemplifica en el cuadro 4.6.

Cuadro 4. 6
Esquema de población por condición de actividad



¹ No han buscado trabajo en el periodo de referencia inmediato, pero aceptarían uno al momento que se les ofreciera.

² No han buscado trabajo en el periodo de referencia inmediato y no aceptarían uno aunque se les ofreciera.

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

El cuadro siguiente representa la base de datos, donde cada columna es una variable. Cada variable puede tomar valores dentro de un conjunto de elementos que se llaman códigos, cada código representa una propiedad que puede adquirir el individuo, por ejemplo en la base de datos

la edad es una variable discreta por lo que código clasificación es la edad. Para clasificar la PEA de la PNEA se asigna el código 1 y 2 respectivamente. Ahora el código de clasificación para ocupados es 1 y para desocupados es 2 (división de la PEA); el código de clasificación para los que están disponibles para trabajar es 3 y los no disponibles es 4 (división de la PNEA). Como se puede observar cada desagregación de una variable tiene un código que clasifica la información como se muestra en el cuadro siguiente.

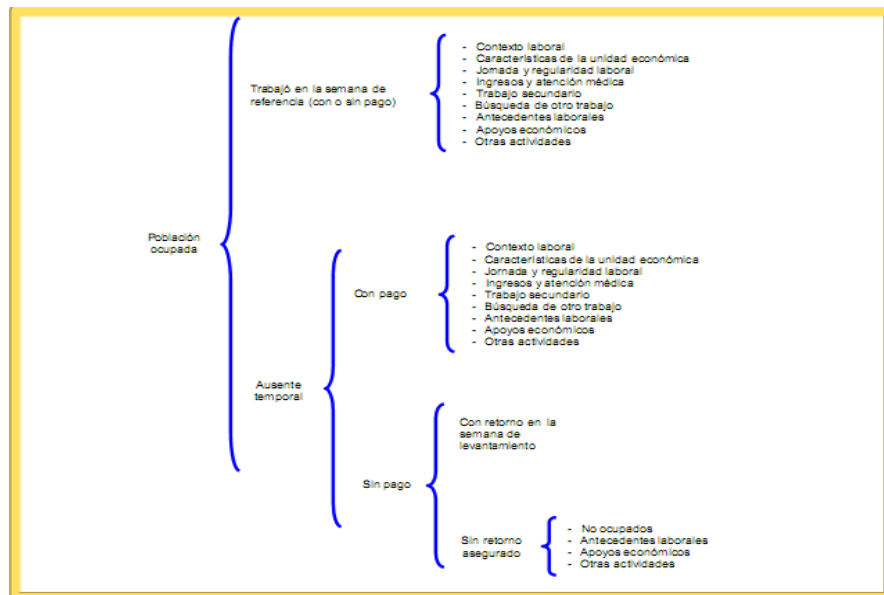
variables

	clase1	clase2	eda	l_nac	e_civ	t_hij	esc	c_res	mig	tnt
1	1	1	37	09	1	01	42292	1	000	
2	1	2	18	09	1	00	33000	1	000	
3	1	1	35	09	2	98	23000	1	000	
5	2	1	31	09	2	02	23000	1	000	
6	2	4	14	09	1	00	21000	1	000	
7	1	1	17	01	1	00	32000	1	000	
8	2	4	41	01	4	03	16000	1	000	
9	2	4	15	01	1	98	23000	1	000	
10	2	4	20	01	2	01	23000	3	000	
11	1	1	14	01	1	98	14000	1	000	
12	1	1	23	14	2	98	23000	1	000	
13	0	0	12	01	1	00	14000	1	000	
14	2	4	28	01	2	03	16000	1	000	
15	1	1	31	01	2	98	14000	1	000	
16	1	1	45	01	2	98	16000	1	000	

códigos

El cuadro 4.7 muestra la información que es recolectada de la población ocupada y es representada en las bases de datos de la encuestas de empleo. Mismos conceptos serán utilizados para la construcción del modelo.

Cuadro 4. 7
Desagregación temática de la población ocupada



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

4.2 Sector formal e informal de la ENE y ENOE

El concepto de sector informal que el INEGI adopta y se menciona en el Capítulo II, es el mismo que el instituto utiliza para el cálculo de dicho sector. En la base de datos de la ENE y ENOE esta información es captada por la variable *TUE2*, en la categoría de *Sector Informal*. Como ya se ha mencionado esta categoría deja fuera del sector informal, actividades agrícolas de autosubsistencia y el trabajo domestico asalariado. Para el objeto de nuestro estudio esta forma de medir al sector informal no será tomada en cuenta y en consecuencia se adoptará lo descrito en el Capítulo II por la IOT y OMC (2009) donde menciona que una forma de definir y medir el sector formal e informal está relacionada con la protección social.

En nuestro caso la protección social solo considera a los individuos afiliados al IMSS. En el cuestionario básico de la ENE esta información es captada por la pregunta 7d. *En el TRABAJO PRINCIPAL de la SEMANA PASADA, ¿cuáles de las siguientes prestaciones le dan a...?*. En el cuestionario ampliado de la ENOE la información es captada por la pregunta 6d. *Por parte de este trabajo ¿... tiene acceso a atención médica en?*. En ambas encuestas estas preguntas se usarán para construir una nueva variable llamada *sector*, que identificará al *sector formal* y al *sector informal*. Más adelante se realizara una descripción más a detalle sobre la construcción de la variable *sector*.

4.3 Variables que explican al sector formal e informal en la ENE de 1996 y ENOE del 2006

Se hace la precisión, que la forma de cuantificar y construir la nueva variable *sector* -nuestra variable dependiente en el modelo-, al igual que la selección de las variables independientes que explican al sector formal e informal, se toman en base a las evidencias halladas en la literatura del *mercado laboral*³⁴; este diseño se sigue para la ENE de 1996 (ENE96) y la ENOE del segundo trimestre del 2006 (ENOE06). Una vez hecho esto se procederá a dar la descripción de la variable dependiente y las variables independientes, la cuales se utilizarán en el ajuste del modelo.

La razón de haber utilizado como periodos de referencia la ENE96 y la ENOE06, el primero fue porque se necesitaba analizar el comportamiento del mercado laboral antes de la reforma y el segundo porque se determino un periodo de espera razonable –aproximadamente 9 años-, para observar el asentamiento y efectos de la reforma. Como se menciono anteriormente la ENOE se levanta trimestralmente y el motivo de porque se utiliza el segundo trimestre para 2006, es porque hay más estabilidad laboral. No se toma un periodo posterior a 2006 dado que se intenta aislar la inestabilidad económica que experimenta México en un periodo aproximado de 2007 a

³⁴ Véase el capítulo II.

2010, que evidentemente afectan el mercado laboral. De modo que con estas dos encuestas se determinara el incremento en la formalización del mercado laboral.

Variable dependiente de la ENE

La variable *sector*, se construirá a través de la pregunta 7d del cuestionario básico de la ENE96. Esta pregunta dice “En el TRABAJO PRINCIPAL de la SEMANA PASADA, ¿cuáles de las siguientes prestaciones le dan a...?”; cuyas opciones son:

- 1) Aguinaldo
- 2) Vacaciones con goce de sueldo
- 3) Participación de utilidades
- 4) IMSS
- 5) ISSSTE
- 6) SAR (Sistema de Ahorro para el Retiro)
- 7) Crédito para vivienda
- 8) Seguro Médico particular o Seguro de Salud
- 9) Otros:_____

En la base de datos la información se refleja en la variable *7d_4*, donde el número *7d* señala la pregunta y el número *_4* señala la opción, está corresponde al *IMSS*. La cual está codificada con 1, que significa que tiene protección social por parte del *IMSS* y 2 que significa que no tiene protección social. Además de está, se toma la variable *7d_5* que corresponde al *ISSSTE*, codificada con 1 y 2. Entonces se construye una nueva variable llamada *sector*, con la información que se dispone de la variable *7d_4* y *7d_5*. De la variable *7d_4* los códigos se renombraran y ahora se identificara con el *código 1* a los trabajadores que están en el sector formal y el *código 2* a los trabajadores que están en el sector informal. Entonces la variable dependiente del modelo logit se llamara *sector*. A continuación se escribe el nombre de la variable, el significado, la escala y la codificación.

Sector (variable dependiente).-Esta asegurado en el *IMSS* o no cuentan con protección social

Escala nominal dicotómica. Codificación: 0 Informal, 1 formal

Variables independientes de la ENE

En seguida se enuncia el nombre de cada variable dentro de la base de datos, una breve descripción, la escala, la codificación y en caso de ser necesaria la recodificación. Cuando se requiera se enunciará la pregunta base del cuestionario de la ENE96, que capta la información.

EDA.- Corresponde a la edad de encuestado (debe tener de 14 años en adelante)

Escala de intervalo. Codificación: El número asignado será la edad de individuo.

Recodificación *edad*.

Sexo.- Corresponde al género del trabajador.

Escala nominal. Codificación: 1 Masculino, 2 Femenino.

Recodificación *genero*.

Escala nominal. Codificación: 1 Masculino, 0 Femenino.

e_con.- Estado conyugal del trabajador.

Escala nominal politómica. Codificación: 1 Unión Libre, 2 Separado, 3 Divorciado, 4 Viudo, 5 Casado, 6 Soltero, 9 No especificado.

Recodificación *edo_conyugal*.

ent.- Hace referencia al estado donde vive en trabajador.

Escala nominal politómica. Codificación: 1 Aguascalientes, 2 Baja California, 3 Baja California Sur, 4 Campeche, 5 Coahuila, 6 Colima, 7 Chiapas, 8 Chihuahua, 9 Distrito Federal, 10 Durango, 11 Guanajuato, 12 Guerrero, 13 Hidalgo, 14 Jalisco, 15 México, 16 Michoacán, 17 Morelos, 18 Nayarit, 19 Nuevo León, 20 Oaxaca, 21 Puebla, 22 Querétaro, 23 Quintana Roo, 24 San Luis Potosí, 25 Sinaloa, 26 Sonora, 27 Tabasco, 28 Tamaulipas, 29 Tlaxcala, 30 Veracruz, 31 Yucatán, 32 Zacatecas.

Recodificación *region_e*. La nueva clasificación se apoya en la información que presenta el INEGI en el producto "Regiones socioeconómicas de México"³⁵, este producto agrupa regiones homogéneas relacionando aspectos como el bienestar, la educación, el empleo, la ocupación, la vivienda y la salud.

Escala nominal politómica. Codificación: 1 Chiapas, Guerrero y Oaxaca; 2 Campeche, Hidalgo, Puebla, San Luis Potosí, Tabasco y Veracruz; 3 Durango, Guanajuato, Michoacán, Tlaxcala, Zacatecas; 4 Colima, México, Morelos, Nayarit, Querétaro, Quintana Roo, Sinaloa y Sinaloa; 5 Baja California, Baja California Sur, Chihuahua, Sonora y Tamaulipas; 6 Aguascalientes, Coahuila, Jalisco, Nuevo León; 7 Distrito Federal.

O_NIV_INS.- Nivel de instrucción o educación del encuestado medido en grado escolar.

Escala ordinal. Codificación: 1 primaria, 2 primaria completa, 3 secundaria completa, 4 medio superior y superior, 5 No especificado.

Recodificación *nivel_instruccion*.

Pregunta: "3a. En su TRABAJO PRINCIPAL de la SEMANA PASADA...era:".

p3a.- Rol del trabajador en el lugar donde labora.

³⁵ El producto Regiones socioeconómicas de México se puede consultar en <http://www.inegi.org.mx/est/contenidos/espanol/sistemas/regsoc/default.asp?s=est&c=11723>

Escala nominal politómica. Codificación: 1 Patrón, 2 Trabajador por su cuenta, 3 Trabajador a destajo, comisión o porcentaje, 4 Trabajador a sueldo fijo, salario o jornal, 5 Miembro de una cooperativa, 6 Trabajador familiar sin pago, 7 Trabajador no familiar sin pago, 8 Otro.

Recodificación *tipo_ocupacion*.

Escala nominal politómica. Codificación: 1 Patrón, 2 Trabajador por su cuenta, 4 Trabajador a destajo, comisión o porcentaje y Trabajador a sueldo fijo, salario o jornal, 5 Miembro de una cooperativa, 6 Trabajador familiar sin pago y Trabajador no familiar sin pago, 8 Otro.

Pregunta: “3b. ¿Cuántas personas, incluyendo al dueño, trabajaron en el negocio, empresa o establecimiento de...la SEMANA PASADA?”

p3b.- Número de trabajadores que laboran en la misma empresa, aplica para los código de 1 a 4 de la variable *p3a*.

Escala ordinal. Codificación: 1 persona, 2 de 2 a 5 personas, 3 de 6 a 10 personas, 4 de 11 a 15 personas, 5 de 16 a 50 personas, 6 de 51 a 100 personas, 7 de 101 a 250 personas, 8 de 251 y más personas, 9 Ocupó personas pero no sabe el número, 99 no sabe.

Pregunta: “3e. ¿Cuántas personas en total, incluyendo al dueño, trabajan para la empresa, cooperativa, negocio o patrón con el cuál... trabajó la SEMANA PASADA?”

p3e.- Número de trabajadores que laboran en la misma empresa, aplica para los código de 5 a 8 de la variable *p3a*.

Escala ordinal. Codificación: 1 persona, 2 de 2 a 5 personas, 3 de 6 a 10 personas, 4 de 11 a 15 personas, 5 de 16 a 50 personas, 6 de 51 a 100 personas, 7 de 101 a 250 personas, 8 de 251 y más personas, 99 no sabe.

Recodificación *c_trabajadores2* (agrupa códigos de la pregunta *p3b* y *p3e*).

Escala ordinal. Codificación: 1 persona, 2 de 2 a 5 personas, 3 de 6 a 10 personas, 4 de 11 a 15 personas, 5 de 16 a 50 personas, 6 de 51 y más personas, 9 Ocupó personas pero no sabe le número, 99 no sabe.

Pregunta: “6a. CLASIFIQUE de acuerdo a las horas registradas en la SEMANA DE REFERENCIA”

p6a.- Número de horas que trabajo en la semana.

Escala ordinal. Codificación: 1 Menos de 35 horas, 2 De 35 a 48 horas, 3 Más de 48 horas, 4 No trabajó la semana pasada, 9 No sabe.

Recodificación *hrs_trabajo*.

Pregunta: “5b. ¿A qué se dedica la empresa, negocio o patrón donde trabajó o ayudó... la SEMANA PASADA?”

p5b.- Indica la actividad de la empresa de acuerdo al clasificador oficial de actividades económicas es el Sistema de Clasificación Industrial de América del Norte³⁶ (SCIAN) en donde labora el individuo.

Escala nominal politómica. La variable toma valores de 101 a 9999, cada valor indica una actividad en específico.

Recodificación *scian_ene*.

Escala nominal politómica. Codificación: 11 Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza, 21 Minería, 22 Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final, 23 Construcción, 3133 Industrias manufactureras, 43 Comercio al por mayor, 46 Comercio al por menor, 4849 Transportes, correos y almacenamiento, 51 Información en medios masivos, 52 Servicios financieros y de seguros, 53 Servicios inmobiliarios y de alquiler de bienes muebles e intangibles, 54 Servicios profesionales, científicos y técnicos, 56 Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación, 61 Servicios educativos, 62 Servicios de salud y de asistencia social, 71 Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos, 72 Servicios de alojamiento temporal y de preparación de alimentos y bebidas, 81 Otros servicios excepto actividades gubernamentales, 93 Actividades legislativas, gubernamentales, de impartición de justicia y de organismos internacionales y extraterritoriales, 99 otros.

Variable dependiente de la ENOE

Para construir la variable *sector*, se utilizó el *cuestionario ampliado de ocupación y empleo de la ENOE06*. Se utilizó la pregunta “*6d. Por parte de este trabajo ¿ ... tiene acceso a atención médica en?*”; cuyas opciones son:

- 1) *el Seguro Social (IMSS)?*
- 2) *el hospital o clínica naval, militar o de Pemex?*
- 3) *el ISSSTE?*
- 4) *el ISSSTE estatal (ISSSTELEÓN, ISSEMYM)?*
- 5) *otra institución médica?*
- 6) *No recibe atención médica*
- 10) *NS*

En la base de datos de la ENOE06, la variable que concentra esta información es la *p6d*. Su codificación es la misma que contiene la pregunta, nuestro interés se enfocara en el código 1 que señala que recibe atención médica a través del IMSS y el 6 que señala que no recibe atención médica. De la variable *p6d* se recodificaran el código 1 y 6, ahora identificarán al sector formal e

³⁶ La clasificación de actividades económicas de la Encuesta Nacional de Empleo (CAE-ENE), era usada para la categorización de actividades. Para realizar una comparación de la ENE y ENOE se llevo la categorización con base en la SCIAN.

informal respectivamente. Entonces la variable de respuesta del modelo logit se llamara *sector*. A continuación se escribe el nombre de la variable, el significado, la escala y la codificación.

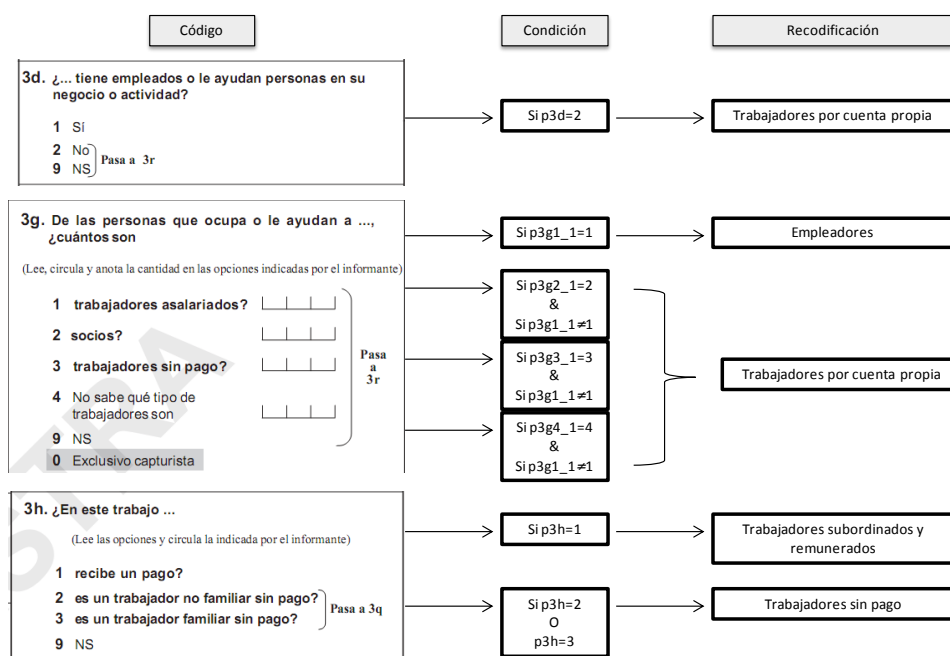
Sector (variable de respuesta).-Esta asegurado en el IMSS o no cuentan con protección social
Escala nominal dicotómica. Codificación: 0 Informal, 1 formal

Variables independientes de la ENOE

Al igual que la ENE96, para la ENOE06, se enuncia el nombre de cada variable dentro de la base de datos, una breve descripción, la escala, la codificación y en caso de ser necesaria la recodificación. Cuando se requiera se enuncia la pregunta base del cuestionario de la ENOE06, que capta la información. Con la intención de no redundar en el tratamiento que se les dio a las variables explicativas seleccionadas de la ENOE06, se asumirá que las no presentadas, tuvieron el mismo tratamiento el cual tuvo la ENE96. Por tanto solo se mencionaran las variables que hayan tenido un tratamiento diferente.

Pregunta: “3d. ¿... tiene empleados o le ayudan personas en su negocio o actividad?”, “3g. De las personas que ocupa o le ayudan a ..., ¿cuántos son” y “3h. ¿En este trabajo ...”.
p3d, p3h, p3g1_1, p3g2_1, p3g3_1, p3g4_1.- Del conjunto de las preguntas enunciadas se determinara el rol del trabajador en el lugar donde labora.
El Cuadro 4.8 muestra como a partir de su codificación se recodifican.

Cuadro 4. 8
Algoritmo de recodificación para la ocupación de trabajador



Fuente: Elaboración propia con información del Cuestionario de ampliado.

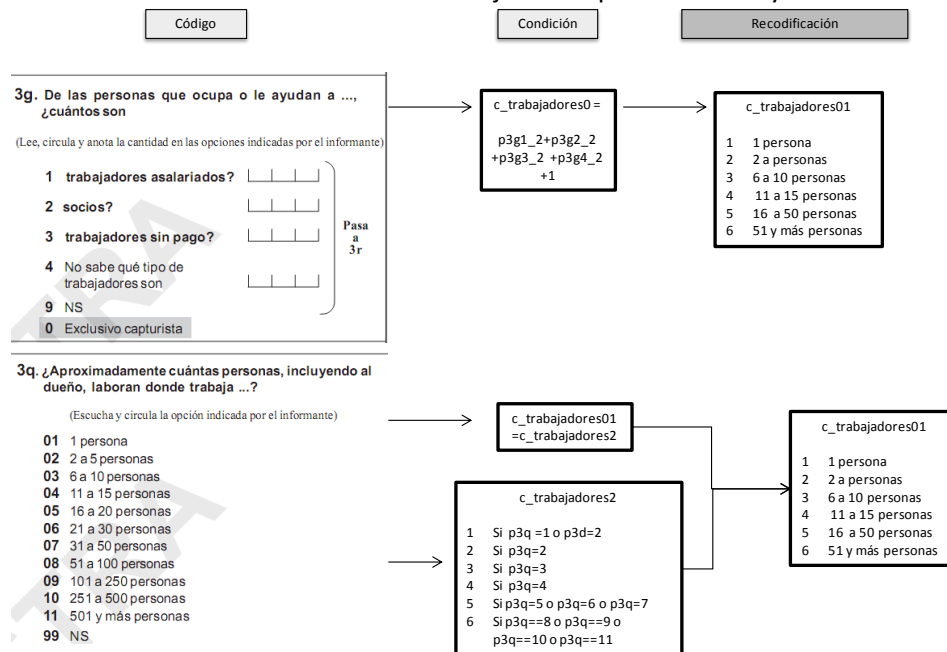
Recodificación *tipo_ocupacion*.

Escala nominal politómica. Codificación: 1 Trabajador subordinado y remunerado, 2 Empleadores, 3 Trabajadores por cuenta propia, 4 Trabajador sin pago.

Pregunta: “3d. ¿... tiene empleados o le ayudan personas en su negocio o actividad?”, “3g. De las personas que ocupa o le ayudan a ..., ¿cuántos son” y “3q. ¿Aproximadamente cuántas personas, incluyendo al dueño, laboran donde trabaja ...?”.

p3g1_2, p3g2_2, p3g3_2, p3g4_2, p3q, p3d.- Del este conjunto de las preguntas se determinara el número de trabajador de la empresa. El conteo de la pregunta *p3d*, aplica para *trabajadores por cuenta propia*; para la pregunta *p3g* aplica para *trabajadores por cuenta propia y empleadores*; para la pregunta *p3q* aplica para *trabajadores subordinados, remunerados y sin pago*, el algoritmo para la construcción de la nueva variable se muestra en el Cuadro 4.9.

Cuadro 4. 9
Algoritmo de recodificación del número de trabajadores que se constituye la unidad económica



Fuente: Elaboración propia con información del *Cuestionario ampliado*.

Recodificación *c_trabajadores2*.

Escala ordinal. Codificación: 1 persona, 2 de 2 a 5 personas, 3 de 6 a 10 personas, 4 de 11 a 15 personas, 5 de 16 a 50 personas, 6 de 51 y más personas.

Pregunta: “5c. ¿Qué días y cuántas horas le dedicó ... a su trabajo la semana pasada?”.

p5c_thrs.- Número de horas que trabajo en la semana.

Escala de intervalo. Codificación: Registra el número de horas y/o minutos por día.

Recodificación *hrs_trabajo*. Codificación: 1 Menos de 35 horas, 2 De 35 a 48 horas, 3 Más de 48 horas, 999 No sabe.

Pregunta: “4a. ¿A qué se dedica esta empresa, negocio o institución?”.

p4a.- Clasificación de la actividad económica donde trabaja el individuo. La ENOE está clasificada de acuerdo al SCIAN.

Escala nominal politómica. La variable toma valores de 1110 a 9999, cada valor indica una actividad en específico.

Recodificación *scian*.

Escala nominal politómica. Codificación: La única diferencia respecto a la ENE, radica en que se incluye el código 55 Corporativo.

4.3.1 Depuración de las bases de datos

Depuración ENE

Como el interés se centra en conservar solo los individuos que presente información completa y clara, se eliminarán las variables que no presenten esta condición.

.

Variable dependiente

Es preciso señalar que en la intersección del código 1 en la variables *7d_4* y *7d_5*, se precederá a eliminar esta información.

Variable independiente

edad.- El intervalo de edad seleccionado es de 14 a 60 años.

tipo de ocupación.- Se elimina el código 8 otro y 9 No especificado.

c_trabajadores2.- Se elimina el código 9 Ocupó personas pero no sabe le número, 99 no sabe.

hrs_trabajo.- Se elimina el código 4 No trabajó la semana pasada, 9 No sabe.

scian_ene.- Se elimina el código 99 otros.

edo_conyugal.- Se elimina el código 9 No especificado.

Depuración ENOE

Variable dependiente

Se eliminan de la variable *p6d* los códigos del 2 al 5 y el 9. Así como elementos vacíos.

Variable independiente

edad.- El intervalo de edad seleccionado es de 14 a 60 años.

tipo de ocupación.- Se elimina el código 9 No especificado de la variable *p3d*.

c_trabajadores2.- Se elimina el código 998 y 999 No especificado de las variables *p3g1_2* *p3g2_2* *p3g3_2* *p3g4_2*, *p3g9*.

hrs_trabajo.- Se elimina el código 999 No especificado y los elementos vacíos de la variable *p5c_thrs*.

scian.- Se elimina el código 99 otros.

edo_conyugal.- Se elimina el código 9 No especificado.

nivel de instrucción.- Se elimina el código 5 No especificado.

El Cuadro 4.10 resume las variables que se tomaron directamente de la base de datos de la ENE96 y la ENOE06, que serán utilizadas para la construcción del modelo.

Cuadro 4. 10
Variables seleccionadas de la ENE96 y la ENOE06

Base de Datos ENE 1996		Base de Datos ENOE 2006	
Variables	Renombre	Variables Cualitativas	Variables Cuantitativas
Variable dependiente		Variable dependiente	
7d_4 y 7d_5	sector	p6d	sector
Variable independiente		Variable independiente	
Variables Población total		Variables Población total	
eda	edad	eda	edad
sexo	genero	sexo	genero
ent	region_e	ent	region_e
Población Ocupada		Población Ocupada	
e_con	edo_conyugal	e_con	edo_conyugal
o_niv_inst	nivel_instruccion	o_niv_inst	nivel_instruccion
p3a	tipo_ocupacion	p3d,p3h,p3g1_1,p3g2_1, p3g3_1,p3g4_1	tipo_ocupacion
p3b	c_trabajadores2	p3g1_2,p3g2_2,p3g3_2,p3g4_2,p3q,p3d	c_trabajadores2
p6a	hrs_trabajo	p5c_thrs	hrs_trabajo
p5b	scian_ene	p4a	scian
Otras Variables (Apoyo)		Otras Variables (Apoyo)	
clase1	PEA	clase1	PEA
clase2	Población ocupada	clase2	Población ocupada
fac	Población total estimada	fac	Población total estimada

Capítulo V

Construcción del modelo que explica el efecto de la reforma al esquema de pensiones en el tamaño del sector formal

Tanto en la base de datos de la ENE96 como en la ENOE06, se utilizan las mismas variables como factores explicativos. La variable dependiente que diferencia a los trabajadores del sector formal del informal, se construyó en relación a su afiliación o no por medio de su trabajo al IMSS. Al tomar el mismo criterio para construir la variable dependiente y seleccionar las mismas variables independientes para ambas encuestas, una vez ajustado el modelo se podrán comparar los resultados obtenidos de la ENE96 con respecto de la ENOE06, mostrando como se ha afectado el tamaño del sector formal o informal por las características del individuo.

5.1 Implementación del modelo logit en la ENE96

Inicialmente la base de datos de la ENE96 es una muestra de N=499,737. Seleccionado las variables que se especifican en el Capítulo IV y tomando sólo la muestra que es de interés analizar, se trabajó con una población de n=143,971. El estudio determinó la decisión de un individuo de laborar en el sector formal o informal bajo nueve variables independientes (región, género, edad, nivel de instrucción, tipo de ocupación, actividad económica (SCIAN), jornada laboral, número de trabajadores en la unidad de trabajo y su estado conyugal).

Para siete de nueve variables independientes se diseñó el número de variables dicotómicas dependiendo del número de categorías en que esté compuesta. En el modelo para la variable, *región* se diseñó seis variables dicotómicas; *nivel de instrucción* se diseñó tres variables dicotómicas; *tipo de ocupación* cuatro variables dicotómicas; *actividad económica* dieciocho variables dicotómicas; *jornada laboral* dos variables dicotómicas; *número de trabajadores* cinco variables dicotómicas; *estado conyugal* cinco variables dicotómicas. La variable género por si misma está determinada como una variable dicotómica (masculino, femenino), la única variable que no se descompone en dicotómica es la edad.

5.1.1 Ajuste del modelo logit

Modelando el fenómeno por medio de una regresión logit múltiple, la ecuación queda como sigue:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \hat{g}(x_i) = i.\text{región} * x_1 + i.\text{genero} * x_2 + i.\text{edad} * x_3 + i.\text{nivel_instruccion} * x_4 +$$

$$i.\text{tipo_ocupacion} * x_5 + i.\text{c_trabajadores2} * x_6 + i.\text{hrs_trabajo} * x_7 \\ + i.\text{scian_ene} * x_9 + i.\text{edo_conyugal} * x_{10}$$

Aplicando en STATA, el comando *logit* con la sintaxis abajo especificada, se obtienen los coeficientes estimados $\hat{\beta}_i$ del modelo de regresión logit. Nótese que si se antepone el índice "i" sobre cada variable explicativa es con la intención de crear las variables dicotómicas necesarias, tal como se mencionó en el Capítulo III.

```
. xi:logit sector i.region_e genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2 i.hrs_trabajo i.scian_ene i
> .edo_conyugal
```

Los resultados del ajuste se muestran en el Cuadro 5.1, donde los coeficientes estimados aparecen en la columna 1, su error estándar en la columna 2, la estadística \mathcal{W} en la columna 3, su p-valor de la estadística en la columna 4 y los intervalos de confianza de los coeficientes estimados en la columna 5 y 6.

Quedando la ecuación como:

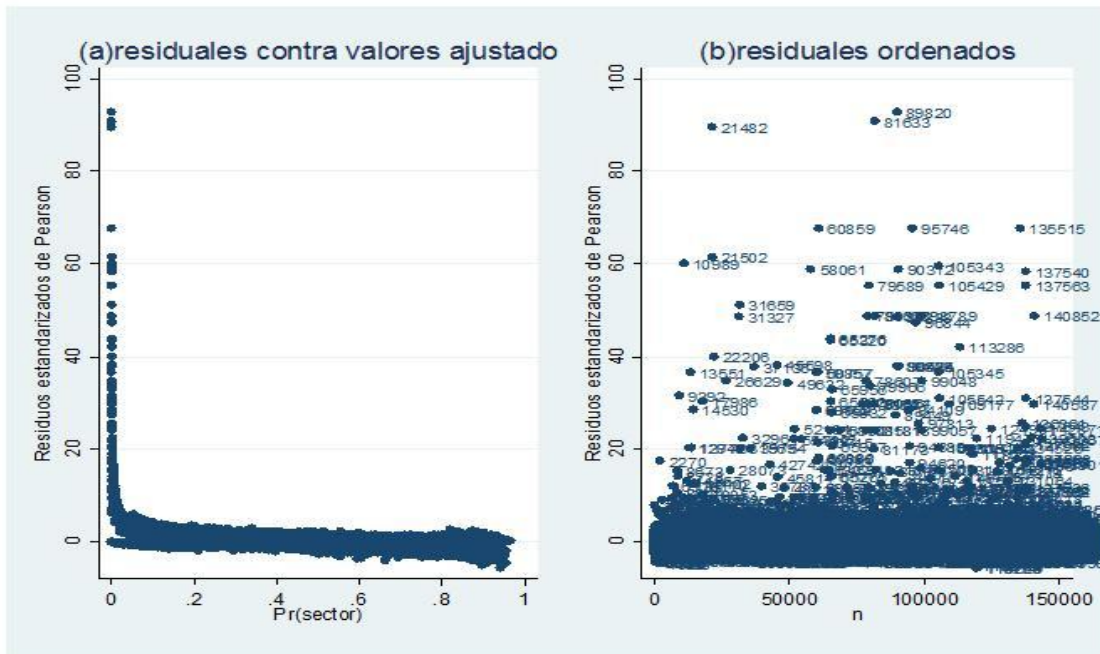
$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \hat{g}(x_i) = -10.80452 + 0.07319x_{i1} + 0.39850x_{i2} + 0.45586x_{i3} + 0.65663x_{i4} + 0.75891x_{i5} + \\ 0.22555x_{i6} + 0.14100x_{i7} + 0.01622x_{i8} + 0.29061x_{i9} + 0.70904x_{i10} + 0.71701x_{i11} + 0.22607x_{i12} + 3.79821x_{i13} + \\ 3.78202x_{i14} + 1.63982x_{i15} + 2.95679x_{i16} + 3.45553x_{i17} + 4.16529x_{i18} + 4.69669x_{i19} + 1.21405x_{i20} + \\ 0.87558x_{i21} - 0.17099x_{i22} + 1.54978x_{i23} + 0.62918x_{i24} + 0.92600x_{i25} + 1.55497x_{i26} + 1.47365x_{i27} - \\ 0.57366x_{i28} + 1.42574x_{i29} + 1.48787x_{i30} + 1.75952x_{i31} + 0.05923x_{i32} + 0.70047x_{i33} + 0.41253x_{i34} + \\ 1.34513x_{i35} + 0.73570x_{i36} - 0.74163x_{i37} + 1.36924x_{i38} + 0.99193x_{i39} + 0.12248x_{i40} + 0.02500x_{i41} + \\ 0.30829x_{i42} + 0.04281x_{i43} + 0.35903x_{i44}$$

Cuadro 5.1
Resultados del ajuste

sector	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Iregion_e_2	.0731905	.038218	1.92	0.055	-.0017155	.1480964
_Iregion_e_3	.3985095	.0384684	10.36	0.000	.3231129	.4739061
_Iregion_e_4	.4558666	.036999	12.32	0.000	.38335	.5283833
_Iregion_e_5	.6566355	.0382123	17.18	0.000	.5817408	.7315301
_Iregion_e_6	.7589106	.0390809	19.42	0.000	.6823134	.8355078
_Iregion_e_7	.2255505	.0602388	3.74	0.000	.1074846	.3436163
genero	.1410079	.0193479	7.29	0.000	.1030867	.1789292
edad	.0162203	.0009526	17.03	0.000	.0143533	.0180874
_Inivel_ins_2	.2906176	.0279161	10.41	0.000	.235903	.3453321
_Inivel_ins_3	.7090488	.0285939	24.80	0.000	.6530057	.7650918
_Inivel_ins_4	.7170195	.0296538	24.18	0.000	.6588991	.7751398
_Itipo_ocup_1	.2260767	.1544661	1.46	0.143	-.0766714	.5288247
_Itipo_ocup_4	3.79821	.1315103	28.88	0.000	3.540455	4.055966
_Itipo_ocup_5	3.782021	.2485794	15.21	0.000	3.294815	4.269228
_Ic_trabaja_2	1.639821	.1060945	15.46	0.000	1.431879	1.847762
_Ic_trabaja_3	2.956794	.1083383	27.29	0.000	2.744455	3.169133
_Ic_trabaja_4	3.455536	.1104063	31.30	0.000	3.239144	3.671928
_Ic_trabaja_5	4.165293	.108164	38.51	0.000	3.953295	4.37729
_Ic_trabaja_6	4.696698	.1072823	43.78	0.000	4.486429	4.906968
_Ihrs_traba_2	1.214056	.0249761	48.61	0.000	1.165104	1.263008
_Ihrs_traba_3	.8755835	.0279428	31.33	0.000	.8208167	.9303503
_Iscian_ene_21	-.1709916	.0886293	-1.93	0.054	-.3447017	.0027186
_Iscian_ene_22	1.549783	.1049223	14.77	0.000	1.344139	1.755427
_Iscian_ene_23	.6291817	.0494109	12.73	0.000	.532338	.7260253
_Iscian_ene_43	.9260016	.0504869	18.34	0.000	.8270491	1.024954
_Iscian_ene_46	1.554975	.0446574	34.82	0.000	1.467448	1.642502
_Iscian_ene_51	1.473658	.0994612	14.82	0.000	1.278718	1.668598
_Iscian_ene_52	-.5736698	.065331	-8.78	0.000	-.7017163	-.4456233
_Iscian_ene_53	1.42574	.1032837	13.80	0.000	1.223308	1.628173
_Iscian_ene_54	1.48787	.0609737	24.40	0.000	1.368363	1.607376
_Iscian_ene_56	1.759523	.1927997	9.13	0.000	1.381643	2.137403
_Iscian_ene_61	.0592375	.0554021	1.07	0.285	-.0493486	.1678236
_Iscian_ene_62	.7004733	.0568367	12.32	0.000	.5890755	.8118711
_Iscian_ene_71	.4125333	.0870228	4.74	0.000	.2419718	.5830947
_Iscian_ene_72	1.345139	.0492603	27.31	0.000	1.248591	1.441687
_Iscian_ene_81	.7357021	.0476385	15.44	0.000	.6423323	.8290719
_Iscian_ene_93	-.7416368	.0479567	-15.46	0.000	-.8356302	-.6476434
_Iscian_ene_3133	1.36924	.0411667	33.26	0.000	1.288555	1.449925
_Iscian_ene_4849	.9919313	.050838	19.51	0.000	.8922907	1.091572
_Iedo_conyu_1	.1224835	.0355887	3.44	0.001	.0527309	.192236
_Iedo_conyu_2	.0250053	.0519176	0.48	0.630	-.0767514	.1267619
_Iedo_conyu_3	.3082965	.0753313	4.09	0.000	.1606499	.4559431
_Iedo_conyu_4	.0428173	.0757164	0.57	0.572	-.105584	.1912187
_Iedo_conyu_5	.3590304	.0208319	17.23	0.000	.3182006	.3998603
_cons	-10.80452	.1621488	-66.63	0.000	-11.12233	-10.48671

Una vez que se estimaron los coeficientes del modelo logit se procede a verificar los residuos. En la Gráfica 5.1a se observa que no hay un buen ajuste del modelo, porque la dispersión de los valores de los residuos va de -5.63 a 93. Lo ideal es que los residuos pudieran estar contenidos en un intervalo de longitud pequeña.

Gráfica 5. 1
Residuos



Al ordenar los residuos y mostrarlos en la Gráfica 5.1b, fue de apoyo para inspeccionar los datos que presenten un valor alto en el residuos y verificar si existe algún tipo de inconsistencia en la información. Recurriendo a lo anterior, se puede decir que hay una inconsistencia en la información en algunos registros; en su mayoría son de individuos que dicen estar afiliados al IMSS por medio de su trabajo siendo que son trabajadores independientes que laboran en el sector primario³⁷ y su nivel de instrucción es hasta la secundaria terminada.

Para evaluar la influencia que tienen estos casos sobre el modelo logit, se miden a través del cambio que éstos puedan producir sobre el modelo. Si al eliminar estos datos cambian significativamente los coeficientes, el intervalo o la constante se puede decir que son casos de gran influencia. Para evitar eliminar cada dato influyente, y volver a calcular los parámetros del modelo, lo que se puede usar es la predicción de Pregibon³⁸. Los casos de gran influencia se obtienen a partir del siguiente comando:

³⁷ Agricultura, explotación forestal, ganadería, minería y pesca.

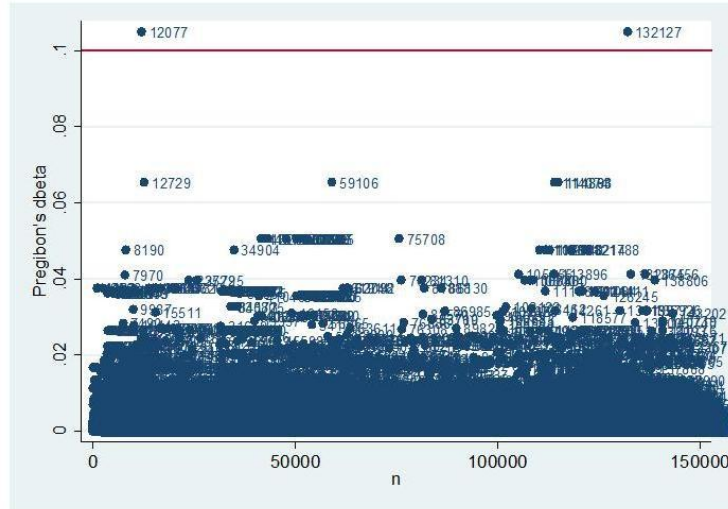
³⁸ La formula de Pregibon es $P_i = \frac{r_i h_{ii}}{(1-h_{ii})^2}$, la estadística mide la desviación horizontal y vertical de los residuos alrededor de función ajustada, la medida determina un caso influyente.

```
. predict cook, dbeta
```

```
. scatter cook n if cook<., mlabel(n) yline(.1)
```

La Gráfica 5.2 muestra los casos que ejercen una influencia importante en el modelo. Se presentan de manera ordenada asignándoles un número para identificarlos en la base de datos.

Gráfica 5. 2
Datos que ejercen influencia en el modelo



Al eliminar la observación 12,077 y 132,127, no se observa un cambio significativo en los coeficientes estimados, ni en la significancia de éstos (Anexo 1). Por tanto se mantiene la base de datos con todas las observaciones, así como el modelo ajustado.

```
. drop if n==132127  
(1 observation deleted)
```

```
. drop if n==12077  
(1 observation deleted)
```

5.1.2 Prueba de significancia de los coeficientes

Al ajustar el modelo y obtener los coeficientes estimados, se realizan la prueba de significancia de los coeficientes utilizando la estadística del cociente de verosimilitud y de Wald con un nivel de significancia $\alpha = .05$.

Prueba de cociente de verosimilitud

La hipótesis nula indica que los coeficientes estimados para el modelo completo $\hat{\beta}_j$, con $j = 1, \dots, 44$ son iguales a cero. De ser cierto esto el modelo estaría perfectamente explicado sólo por la constante $\hat{\beta}_0$.

Una vez que se ajusta el modelo completo $\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_{44}$ y se obtienen los coeficientes estimados, el valor del logaritmo de verosimilitud que se presenta en la parte inferior del siguiente cuadro, es -47,891.133 (*Log likelihood*).

Regresión logit (Modelo completo) $\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_{44}$

Logistic regression	Number of obs	=	143971
	LR chi2(44)	=	91390.08
	Prob > chi2	=	0.0000
Log likelihood = -47891.133	Pseudo R2	=	0.4883

Ahora el valor del logaritmo de verosimilitud para el modelo que tiene sólo la constante ($\hat{g}(x) = \hat{\beta}_0$), mostrado en el cuadro siguiente, es -93,586.175.

Regresión logit $\hat{g}(x) = \hat{\beta}_0$

Logistic regression	Number of obs	=	143971
	LR chi2(0)	=	-0.00
	Prob > chi2	=	.
Log likelihood = -93586.175	Pseudo R2	=	-0.0000

El valor del cociente de verosimilitud es:

$$G = -2(-93,586.175 + 47,891.133) = 91,390.084$$

El *p* – *valor* de la prueba es de $P[\chi^2(43) > 91,390.084] = 0.0000$, entonces se rechaza la hipótesis nula, y se puede decir, que las variables independientes aportan información sobre la decisión del trabajador de pertenecer al sector formal, o que al menos uno de los 44 coeficientes es distinto de cero.

Pruebas de Wald

La prueba de Wald evalúa la hipótesis nula si el coeficiente estimado ($\hat{\beta}_j$) es igual a cero. Como primer ejemplo, se propone como hipótesis nula, que la variable dicotómica *_lregion_e_3* es igual a cero. Apoyándonos en el Cuadro 5.1, la estadística de Wald se obtiene al dividir el coeficiente estimado (columna 1) entre el error estándar del coeficiente (columna 2), el valor obtenido es $W_j = \frac{0.3985095}{0.0384684} = 10.36$ (columna 3), y por último el nivel de significancia es $P(|z| > 10.36) = 0.000$ (columna 4), por tanto el coeficiente estimado $\hat{\beta}_2$ es significativo, o que es lo mismo, que es distinto de cero, entonces se rechaza la hipótesis nula. Realizando un segundo ejemplo para la

variable *_ledo_conyu_4* (*Viudo*) y se obtiene que no es significativo ya que él *p – valor* es 0.572, entonces no se rechaza la hipótesis nula, sin embargo, estas variables dicotómicas no se pueden sacar del modelo, dado que está contenida dentro de la variable *edo_conyugal*. Solo se pueden eliminar a variables completas, como puede ser todas las categorías de la variable edad, entidad, nivel de instrucción, etc., y no un subconjunto de ellas (Hosmer y Lemeshow, 2000). Este proceso se sigue para todos los coeficientes estimados para revisar si son significativos o no.

En general la mayoría de los coeficientes son distintos de cero, con excepción de las siguientes variables dicotómicas creadas: *_lregion_e_2*, *_ltipo_ocup_1*, *_lscian_ene_21*, *_lscian_ene_61* *_ledo_conyu_2*, y *_ledo_conyu_4*, con un *p – valor* de 0.55, 0.144, 0.054, 0.285, 0.630 y 0.572 respectivamente.

Los intervalos de confianza son muy útiles para saber el rango de variación de los parámetros estimados $\hat{\beta}_j$. Si éstos se encuentran en un intervalo cuyo rango va de un número negativo a uno positivo, indica que en algún momento éstos pueden tomar el valor de cero (columna 6 y 7). Las variables antes mencionadas son las que se encuentran en este caso, y por tanto podrían no ser significativas.

5.1.3 Comparando modelos

En este paso se eliminan las variables que resultaron no ser significativas, para ello se construyen cuatro modelos reducidos y se comparan con el modelo completo a través de la estadística *G*. Entonces se elimina del Modelo02 la *región_e*, del Modelo03 el *tipo_ocupacion*, del Modelo04 el *scian_ene* y del Modelo05 el *edo_conyugal*, el modelo completo será el Modelo01. El proceso de eliminación será una a la vez y se compara con el modelo completo.

Se compara el modelo completo (Modelo01) respecto de los modelos reducidos (Modelo02, Modelo03, Modelo04 y Modelo05), para evaluar si la variable independiente que se excluye del modelo completo, contribuye a explicar la variación de la variable dependiente. Para esto se utilizan los valores de la estadística *G* (*LR chi2*) y su *p – valor* (*Prob>chi2*) del Cuadro 5.2.

El *p – valor* para cada estadística *G*, es menor al $\alpha = .05$, con lo que se puede determinar que las variables *region_e*, *tipo_ocupacion*, *scian_ene* y *edo_conyugal* aportan información al modelo. Entonces se trabajará con el modelo completo (Modelo01).

Cuadro 5. 2 Comparación de Modelo

```

. lrtest Modelo01 Modelo02

Likelihood-ratio test                                LR chi2(6) =    918.01
(Assumption: Modelo02 nested in Modelo01)         Prob > chi2 =    0.0000

. lrtest Modelo01 Modelo03

Likelihood-ratio test                                LR chi2(3) =   6020.57
(Assumption: Modelo03 nested in Modelo01)         Prob > chi2 =    0.0000

. lrtest Modelo01 Modelo04

Likelihood-ratio test                                LR chi2(18) =  7332.16
(Assumption: Modelo04 nested in Modelo01)         Prob > chi2 =    0.0000

. lrtest Modelo01 Modelo05

Likelihood-ratio test                                LR chi2(5) =    329.84
(Assumption: Modelo05 nested in Modelo01)         Prob > chi2 =    0.0000

```

Ahora para seleccionar el modelo ajustado con el menor número de variables que mejor expliquen el modelo, se comparan las estadísticas que arroja el cuadro 5.3.

Cuadro 5. 3 Selección de Modelo

Variable	Modelo01	Modelo02	Modelo03	Modelo04	Modelo05
N	143971	143971	143971	143971	143971
ll_0	-93586.175	-93586.175	-93586.175	-93586.175	-93586.175
ll	-47891.133	-48350.136	-50901.419	-51557.214	-48056.054
chi2	91390.084	90472.077	85369.512	84057.921	91060.241
r2_p	.48826701	.4833624	.45610109	.44909369	.48650477
aic	95872.266	96778.273	101886.84	103168.43	96192.109
bic	96316.747	97163.49	102301.69	103435.12	96587.203
rank	45	39	42	27	40

En la primera fila se muestra el tamaño de la población ($N = 143,971$), siendo idéntica para todos los modelos. Abajo se muestra el valor del logaritmo de la función de verosimilitud del modelo sólo con el coeficiente estimado $\widehat{\beta}_0$ (fila dos), este valor (-93,586) debe ser el mismo para los cinco modelos; el logaritmo de la función de verosimilitud para cada modelo que contiene un subconjunto de variables independientes del modelo completo se muestra en la fila tres, de la comparación de ambos logaritmos (fila dos y tres) resulta la estadística G (cociente de

verosimilitud). El cociente de verosimilitud (fila cuatro) indica si el subconjunto tomado de las variables independientes, respecto del modelo que contiene sólo la constante, ejercen influencia alguna en la variable dependiente. La capacidad de predicción en general para todos los modelos varía del 45% al 48% (fila cinco). El renglón a destacar es el que contiene las *medidas de información (AIC y BIC)*, que serán de utilidad para determinar el mejor modelo de los cinco que se proponen. Como se mencionó anteriormente se toma el modelo que presente el valor mínimo tanto del *AIC* como del *BIC* (fila seis y siete), esta condición la cumple el Modelo01, dado que el valor que toma AIC es 95,872 y BIC es 96,317, que corresponden al modelo completo.

5.1.4 Bondad de Ajuste

Cuadro 5. 4
Bondad de ajuste del modelo logit por sector

(Table collapsed on quantiles of estimated probabilities)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.0007	31	4.9	14367	14393.1	14398
2	0.0028	42	20.1	14355	14376.9	14397
3	0.0189	94	121.7	14313	14285.3	14407
4	0.0938	613	758.4	13774	13628.6	14387
5	0.2181	2214	2150.1	12183	12246.9	14397
6	0.4411	5145	4694.8	9252	9702.2	14397
7	0.6363	7458	7790.8	6939	6606.2	14397
8	0.7932	10013	10400.1	4385	3997.9	14398
9	0.8732	12093	12043.7	2303	2352.3	14396
10	0.9649	13298	13016.3	1099	1380.7	14397

```

number of observations = 143971
number of groups = 10
Hosmer-Lemeshow chi2(8) = 413.76
Prob > chi2 = 0.0000

```

El cuadro anterior indica y brinda una idea, de que tan próximo, está el valor estimado \hat{Y} del valor observado Y . El cuadro es construido dividiendo la región de probabilidad en deciles.

Se observa que el p – *valor* es menor al α .05, esto implica que se acepta la hipótesis nula, es decir, se acepta que hay una falta de ajuste en el modelo. Sin embargo, realizando una inspección visual³⁹, se observa que los valores estimados \hat{Y} no “distan mucho” de los valores observados Y .

³⁹ Esta es una forma permitida que se puede usar para validar los resultados, hecho que se menciona en el Capítulo III.

Curva ROC

La base de datos de la ENE96 con que se está trabajando contiene información de 143,971 individuos, que mediante el modelo logit se pretende establecer si un sujeto trabaja dentro del sector formal o informal (variable dependiente), una vez determinaron las características de cada uno (variables independientes).

Una vez que se ajustó el modelo y se evaluó la significancia de los parámetros estimados ($\hat{\beta}$), se procede a evaluar la capacidad que éste tiene, para diferenciar y clasificar a los individuos que laboran en el sector formal de aquellos que laboran en el sector informal, dado diferentes puntos de corte (c_p) con incrementos de 0.1 (Cuadro 5.5).

La tasa de sensibilidad (Se) muestra los casos donde se dio una correcta predicción (\hat{Y}) de los individuos que verdaderamente laboran en el sector formal ($Y = 1$). La tasa de especificidad (Sp) muestra los casos donde se dio una predicción correcta (\hat{Y}) de aquellos individuos que verdaderamente están en el sector informal ($Y = 0$). Derivado de esto se mostrara la curva de ROC.

Cuadro 5. 5

Tabla de clasificación del sector formal e informal con cortes de probabilidad de 0.1

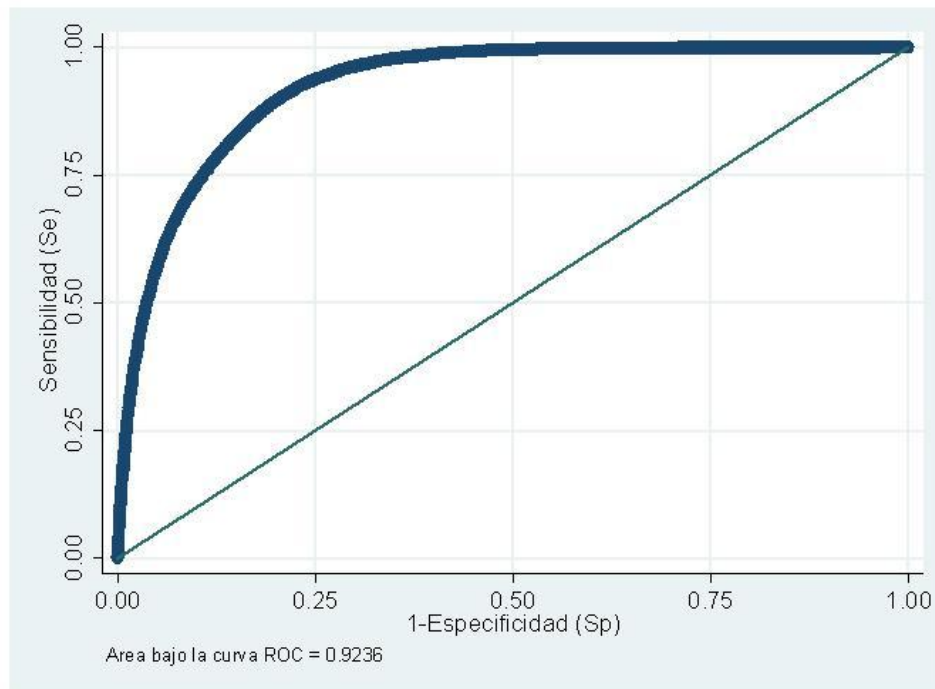
Punto de Corte	Casos Positivos		Casos Negativos		Porcentajes			
	D	~D	D	~D	Se	Sp	1-Sp (falsos positivos)	Clasificados Correctamente
0.1	50,143	35,274	858	57,696	98.32%	62.06%	37.94%	74.90%
0.2	48,393	25,176	2,608	67,794	94.89%	72.92%	27.08%	80.70%
0.3	46,333	19,879	4,668	73,091	90.85%	78.62%	21.38%	82.95%
0.4	44,045	16,203	6,956	76,767	86.36%	82.57%	17.43%	83.91%
0.5	40,945	12,632	10,056	80,338	80.28%	86.41%	13.59%	84.24%
0.6	37,007	9,013	13,994	83,957	72.56%	90.31%	9.69%	84.02%
0.7	32,283	5,992	18,718	86,978	63.30%	93.55%	6.45%	82.84%
0.8	24,635	3,193	26,366	89,777	48.30%	96.57%	3.43%	79.47%
0.9	7,084	489	43,917	43,917	13.89%	98.90%	1.10%	53.46%
1	-	-	51,001	92,970	0.00%	100.00%	0.00%	64.58%

Analizando los datos de la quinta fila cuando c_p es 0.5, del Cuadro 5.5, para aquellos cuya probabilidad estimada sea mayor que 0.5 el evento ocurrirá (casos positivos), por tanto el individuo estará en el sector formal; de la misma manera para aquellos cuya probabilidad sea menor a 0.5 el evento no ocurrirá (casos negativos), por tanto el individuo estará en el sector informal. Nótese que para el punto de corte 0.5, 40,945 casos de 53,577 se clasificaron correctamente como individuos que trabajan en el sector formal y 12,632 casos fueron clasificados incorrectamente, dado que estos individuos trabajan en sector informal; ahora 80,338 casos de

90,394 fueron clasificados correctamente como individuos que trabajan en el sector informal, en tanto que 10,056 casos fueron clasificados incorrectamente, dado que estos individuos trabajan en el sector formal. Del modelos se obtiene que la sensibilidad (Se) es de 80.28%, la especificidad (Sp) es de 86.41% y los falsos positivos fueron de 13.59%. Así que la sensibilidad es mayor que los falsos positivos. En general la tasa de los eventos que se clasifican correctamente (ocurrencia y no ocurrencia del evento) es del 84.24%.

En la Gráfica 5.3 se muestra qué el modelo tiene una capacidad excelente para discriminar entre los individuos que trabajan el sector formal de los que trabajan el informal⁴⁰ (AUC= 0.9236), es decir que hay una mayor proporción de predicciones correctas ($\hat{\pi}$) que reflejan verdaderamente los casos donde el individuo labora en el sector formal ($Y = 1$). Por tanto los datos se ajustan bien al modelo.

Gráfica 5.3
Curva ROC



⁴⁰ Véase la Guía de clasificación para valores de AUC, capítulo III.

5.2 Implementación del modelo logit en la ENOE06

La base de datos de la ENOE06 contiene una muestra de 319,332 individuos. Tomando las variables que se indican el Capítulo IV, y realizando los filtros correspondientes, ahora la nueva muestra está compuesta de 136,323 individuos.

De la misma manera que en la ENE96, para siete de nueve variables independientes se diseña el número de variables dicotómicas dependiendo del número de categorías en que este compuesta. Con la única excepción que en la variable SCIAN ahora se compone de diecinueve variables dicotómicas. Esto último se deriva del hecho que en la ENOE06 se incluye una nueva actividad económica que es *Corporativo*.

5.2.1 Ajuste del modelo logit

El modelo de regresión logit que modela el fenómeno del sector formal o informal es:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \hat{g}(x_i) = \text{i.region} * x_1 + \text{i.genero} * x_2 + \text{i.edad} * x_3 + \text{i.nivel_instruccion} * x_4 + \\ \text{i.tipo_ocupacion} * x_5 + \text{i.c_trabajadores2} * x_6 + \text{i.hrs_trabajo} * x_7 \\ + \text{i.scian} * x_9 + \text{i.edo_conyugal} * x_{10}$$

Aplicando en STATA, el comando *logit* con la sintaxis abajo especificada, se obtiene el coeficiente estimado $\hat{\beta}_i$ del modelo de regresión logit. Se antepone el índice "i" sobre cada variable explicativa para crear las variables dicotómicas necesarias.

```
. xi:logit sector i.region_e genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores  
> 2 i.hrs_trabajo i.scian i.edo_conyugal
```

Los resultados del ajuste se muestran en el cuadro 5.6, los coeficientes estimados, el error estándar, la estadística \mathcal{W} , el *p – valor* y los intervalos de confianza de los coeficientes estimados aparecen en las mismas columnas que se mencionaron para la ENE96.

Quedando la ecuación como:

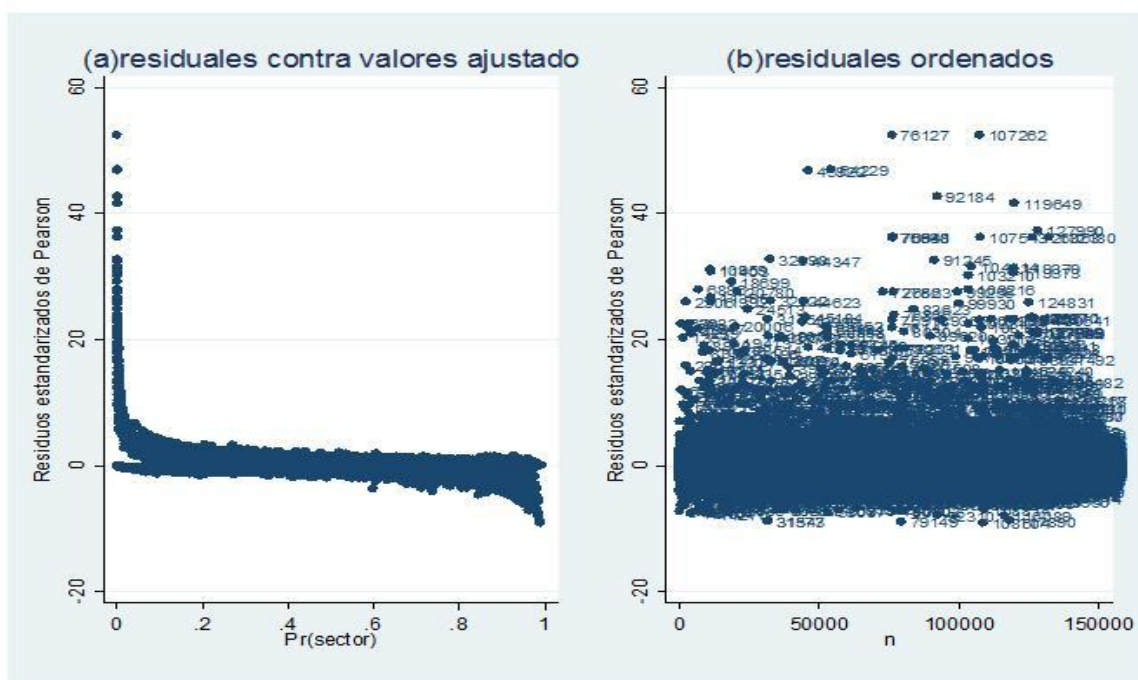
$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \hat{g}(x_i) = -10.48555 + 0.16044x_{i1} + 0.26984x_{i2} + 0.42614x_{i3} + 1.06365x_{i4} + 0.93134x_{i5} - \\ 0.23992x_{i6} + 0.05224x_{i7} + 0.02567x_{i8} + 0.37089x_{i9} + 0.89845x_{i10} + 1.0956x_{i11} + 2.99834x_{i12} - 0.13536x_{i13} - \\ 1.07762x_{i14} + 1.07880x_{i15} + 2.42020x_{i16} + 3.06922x_{i17} + 3.61296x_{i18} + 4.56788x_{i19} + 1.20101x_{i20} + \\ 0.85946x_{i21} + 1.99739x_{i22} + 2.44424x_{i23} + 1.14089x_{i24} + 2.39677x_{i25} + 2.15755x_{i26} + 1.94272x_{i27} + \\ 2.34050x_{i28} + 1.77160x_{i29} + 1.91910x_{i30} + 2.30331x_{i31} + 2.20327x_{i32} + 1.76113x_{i33} + 1.81806x_{i34} + \\ 0.91096x_{i35} + 1.76352x_{i36} + 1.16058x_{i37} + 0.63232x_{i38} + 1.89918x_{i39} + 1.42737x_{i40} + 0.13838x_{i41} + \\ 0.11606x_{i42} + 0.25187x_{i43} + 0.08504x_{i44} + 0.40870x_{i45}$$

Cuadro 5. 6
Resultados del ajuste

sector	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Iregion_e_2	.1604428	.039634	4.05	0.000	.0827615	.238124
_Iregion_e_3	.2698429	.0406832	6.63	0.000	.1901052	.3495805
_Iregion_e_4	.4261496	.0377434	11.29	0.000	.3521738	.5001254
_Iregion_e_5	1.063652	.041507	25.63	0.000	.9823002	1.145005
_Iregion_e_6	.9313478	.0414861	22.45	0.000	.8500366	1.012659
_Iregion_e_7	-.2399282	.0634312	-3.78	0.000	-.3642512	-.1156052
genero	.0522456	.021285	2.45	0.014	.0105278	.0939634
edad	.0256742	.0010347	24.81	0.000	.0236462	.0277022
_Inivel_ins_2	.3708997	.0355836	10.42	0.000	.3011571	.4406423
_Inivel_ins_3	.8984599	.0345243	26.02	0.000	.8307936	.9661262
_Inivel_ins_4	1.095699	.0365106	30.01	0.000	1.02414	1.167259
_Itipo_ocup_1	2.998341	.0853142	35.14	0.000	2.831128	3.165554
_Itipo_ocup_2	-.1353695	.1060189	-1.28	0.202	-.3431627	.0724236
_Itipo_ocup_4	-1.077625	.1785965	-6.03	0.000	-1.427668	-.7275821
_Ic_trabaja_2	1.078803	.0868512	12.42	0.000	.9085776	1.249028
_Ic_trabaja_3	2.420206	.0897858	26.96	0.000	2.244229	2.596183
_Ic_trabaja_4	3.06922	.0930167	33.00	0.000	2.886911	3.25153
_Ic_trabaja_5	3.612964	.0895901	40.33	0.000	3.437371	3.788558
_Ic_trabaja_6	4.567885	.0902613	50.61	0.000	4.390976	4.744794
_Ihrs_traba_2	1.201019	.0271174	44.29	0.000	1.147869	1.254168
_Ihrs_traba_3	.859468	.0298664	28.78	0.000	.8009309	.9180051
_Iscian_21	1.997392	.1597385	12.50	0.000	1.684311	2.310474
_Iscian_22	2.444248	.1791335	13.64	0.000	2.093153	2.795344
_Iscian_23	1.140894	.0518486	22.00	0.000	1.039272	1.242515
_Iscian_43	2.396775	.0634143	37.80	0.000	2.272485	2.521065
_Iscian_46	2.15755	.0504584	42.76	0.000	2.058654	2.256447
_Iscian_51	1.942727	.0959144	20.25	0.000	1.754739	2.130716
_Iscian_52	2.340504	.0965341	24.25	0.000	2.1513	2.529707
_Iscian_53	1.771603	.1067056	16.60	0.000	1.562464	1.980742
_Iscian_54	1.919103	.0715687	26.81	0.000	1.778831	2.059375
_Iscian_55	2.303315	.72474	3.18	0.001	.8828507	3.723779
_Iscian_56	2.203278	.0737968	29.86	0.000	2.058639	2.347917
_Iscian_61	1.761135	.0668721	26.34	0.000	1.630068	1.892201
_Iscian_62	1.818062	.0702624	25.88	0.000	1.68035	1.955774
_Iscian_71	.9109653	.0973371	9.36	0.000	.720188	1.101743
_Iscian_72	1.763526	.0544667	32.38	0.000	1.656774	1.870279
_Iscian_81	1.160586	.0562647	20.63	0.000	1.050309	1.270863
_Iscian_93	.6323219	.059764	10.58	0.000	.5151866	.7494572
_Iscian_3133	1.899188	.0485558	39.11	0.000	1.804021	1.994356
_Iscian_4849	1.427375	.0588516	24.25	0.000	1.312029	1.542722
_Iedo_conyu_1	.138384	.0329028	4.21	0.000	.0738958	.2028723
_Iedo_conyu_2	.1160632	.0517907	2.24	0.025	.0145553	.2175712
_Iedo_conyu_3	.2518769	.0720802	3.49	0.000	.1106023	.3931516
_Iedo_conyu_4	.085044	.0861686	0.99	0.324	-.0838433	.2539313
_Iedo_conyu_5	.4087042	.0241577	16.92	0.000	.3613559	.4560524
_cons	-10.48555	.1163543	-90.12	0.000	-10.7136	-10.2575

Ya que se estimaron los coeficientes del modelo logit se procede a verificar los residuos. Al igual que en la ENE96 se observa que en la Gráfica 5.7a no hay un buen ajuste del modelo, ya que la dispersión de los valores de los residuos va de -9.07 a 52, aunque esta dispersión es menor que la que presenta la ENE96, sigue siendo alta.

Cuadro 5. 7
Residuos



Se ordenan los residuos y se asigna un número que los identifique (Gráfica 5.7b). Se visualizan los valores más altos de los residuos y se procede a inspeccionarlos dentro de la base de datos. La inconsistencia de la información se presenta en individuos que dicen: estar afiliados al IMSS por medio de su trabajo siendo que son trabajadores independientes, con diversos niveles de instrucción y quienes en su mayoría laboran en el sector primario (misma inconsistencia ENE96 y ENOE06) o en la minería.

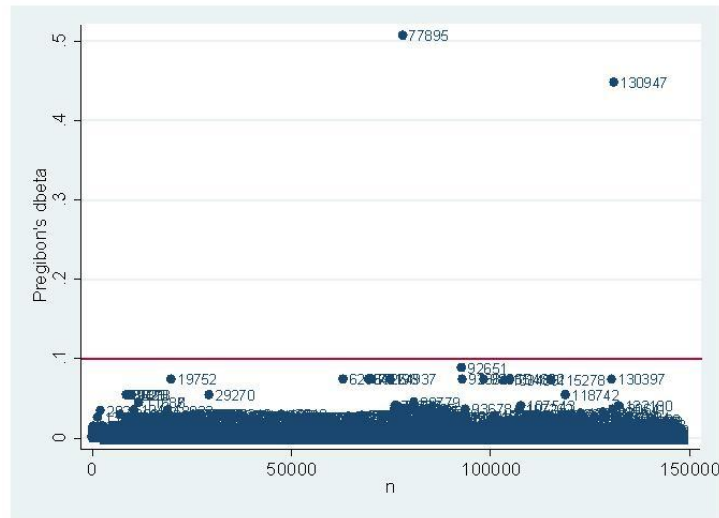
Para evaluar la influencia que tienen estos casos sobre el modelo logit, se mide a través del cambio que puedan producir en el modelo. Si al eliminar estos datos cambian significativamente los coeficientes, el intervalo o la constante se puede decir que son casos de gran influencia. Para evitar eliminar cada dato influyente y volver a calcular los parámetros del modelo, se usa la predicción de Pregibon. Los casos de gran influencia los obtenemos a partir del siguiente comando:

```
. predict cook, dbeta
```

```
. scatter cook n if cook<., mlabel(n) yline(.1)
```

La gráfica 5.8 muestra los casos que ejercen una influencia importante en el modelo, se les asigna un número para identificarlos en la base de datos.

Cuadro 5. 8
Datos que ejercen influencia en el modelo



Se eliminan las dos observaciones que se muestran el grafico (77,895 y 130,947) que ejercen influencia en los parámetros, o intervalos de confianza de estos. Se calculan nuevamente los parámetros del modelo, y se obtiene que no hay cambio significativo en alguno de ellos, con excepción del coeficiente de la variable independiente `_Iscian_55` que pasó de 2.303315 a 3.866768 (Anexo 1). Pero por no existir un cambio significativo en el modelo en general, mantenemos los datos en la base.

5.2.2 Prueba de significancia de los coeficientes

Una vez que se ajustó el modelo y obtuvieron los coeficientes estimados, se realizan la prueba de significancia de los coeficientes utilizando la estadística del cociente de verosimilitud y de Wald con un nivel de significancia $\alpha = .05$.

Prueba de cociente de verosimilitud

La hipótesis nula indica que los coeficientes estimados para el modelo completo $\hat{\beta}_j$, con $j = 1, \dots, 45$ son iguales a cero. De ser cierto esto nos diría que el modelo está perfectamente explicado sólo por la constante $\hat{\beta}_0$.

Una vez que se ajusta el modelo completo $\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1x_1 + \dots + \hat{\beta}_kx_{45}$ y se obtienen los coeficientes estimados, el valor del logaritmo de verosimilitud se presenta en la parte inferior del siguiente cuadro, es -38,487.141 (*Log likelihood*).

Logistic regression	Number of obs	=	136323
	LR chi2(45)	=	99920.36
	Prob > chi2	=	0.0000
Log likelihood = -38487.141	Pseudo R2	=	0.5649

El valor del logaritmo de verosimilitud para el modelo que tiene sólo la constante ($\hat{g}(x) = \hat{\beta}_0$), que se muestra abajo y es -88,447.319.

Logistic regression	Number of obs	=	136323
	LR chi2(0)	=	0.00
	Prob > chi2	=	.
Log likelihood = -88447.319	Pseudo R2	=	0.0000

Calculando el valor de cociente de verosimilitud de es:

$$G = -2(-88447.319 - (-38,487.141)) = 99,920.356$$

El *p – valor* de la prueba es $P[\chi^2(44) > 99,920.356] = 0.0000$, por tanto se rechaza la hipótesis nula, y se puede decir, que las variables independientes aportan información sobre la decisión del trabajador de pertenecer al sector formal, o que al menos uno de los 45 coeficientes es distinto de cero.

Pruebas de Wald

La prueba de Wald evalúa la hipótesis nula si el coeficiente estimado ($\hat{\beta}_j$) es igual a cero. Se propone como hipótesis nula que la variable dicotómica *_lregion_e_2* es igual a cero. Sobre los resultados del Cuadro 5.6, la estadística de Wald se obtiene al dividir el coeficiente estimado (columna 1) entre el error estándar del coeficiente (columna 2), el valor obtenido es $W_j = \frac{0.1604428}{0.039634} = 4.05$ (columna 3), donde el *p – valor* es 0.0 (columna 4), entonces el coeficiente estimado $\hat{\beta}_2$ es significativo, o que es lo mismo, es distinto de cero, por lo tanto se rechaza la hipótesis nula. El proceso anterior se sigue para todas las coeficientes estimados para revisar si son significativos o no.

En general, la mayoría de los coeficientes son distintos de cero, con excepción de las siguientes variables dicotómicas creadas: *_ltipo_ocup_2* y *_ledo_conyu_4*. El *p – valor* para *_ltipo_ocup_2* y *_ledo_conyu_4* es 0.202 y 0.324 respectivamente, los cuales resultan no ser significativas. Para evaluar la hipótesis nula de estas variables se eliminar todas las categorías de la

variable *tipo_ocup* y *edo_conyugal*, tal como se realizó en la ENE96. En la ENOE06 son 4 menos, las variables dicotómicas que resultaron ser no significativas respecto la ENE96.

Para saber el rango de variación de los parámetros estimados $\hat{\beta}_j$, se observa su intervalo de confianza (columna 6 y 7). Si éstos se encuentran en un intervalo cuyo rango va de un número negativo a uno positivo, indica que en algún momento éstos pueden tomar el valor de cero. Las variables antes mencionadas son las que se encuentran en este caso, y por tanto podrían no ser significativas.

5.2.3 Comparando modelos

En este paso se eliminan las variables que resultaron no ser significativas, para ello se construyen dos modelos reducidos y se comparan con el modelo completo a través de la estadística *G*. Entonces se elimina del Modelo02 la variable *tipo_ocup* y del Modelo03 el *edo_conyugal*. Se elimina una a la vez y se compara con el modelo completo (Modelo01).

Se compara el modelo completo (Modelo01) respecto de los modelos reducidos (Modelo02 y Modelo03), para evaluar si la variable independiente que se excluye del modelo completo, contribuye a explicar la variación de la variable dependiente. Para esto se utilizan los valores de la estadística *G* (*LR chi2*) y su *p – valor* (*Prob>chi2*) del Cuadro 5.9.

Cuadro 5. 9
Comparación de Modelo

```
. lrtest Modelo01 Modelo02

Likelihood-ratio test                                LR chi2(3) =    7168.74
(Assumption: Modelo02 nested in Modelo01)         Prob > chi2 =     0.0000

. lrtest Modelo01 Modelo03

Likelihood-ratio test                                LR chi2(5) =     309.82
(Assumption: Modelo03 nested in Modelo01)         Prob > chi2 =     0.0000
```

El *p – valor* para cada estadística *G*, resulta ser menor a 0.05, con lo que se puede determinar que las variables *tipo_ocupacion* y *edo_conyugal* aportan información al modelo. Entonces se deja el modelo completo (Modelo01).

Para seleccionar el modelo ajustado con el menor número parámetros que mejor expliquen el modelo, se comparan las estadísticas que arroja el cuadro 5.10.

Cuadro 5. 10
Selección de Modelo

Variable	Modelo01	Modelo02	Modelo03
N	136323	136323	136323
ll_0	-88447.319	-88447.319	-88447.319
ll	-38487.141	-42071.509	-38642.048
chi2	99920.356	92751.62	99610.541
r2_p	.56485803	.52433257	.56310662
aic	77066.281	84229.018	77366.096
bic	77518.129	84651.397	77768.831
rank	46	43	41

En la primera fila se muestra el tamaño de la población ($N = 136,323$), siendo idéntica para los tres modelos. En el cuadro se muestra el valor del logaritmo de la función de verosimilitud del modelo sólo con el coeficiente estimado $\widehat{\beta}_0$ (fila dos), este valor (-88,447.319) debe ser el mismo para los tres modelos; el logaritmo de la función de verosimilitud de los modelos reducidos (modelo dos y tres) del modelo completo se muestra en la fila siguiente, de la comparación de ambos logaritmos se obtiene la estadística G (cociente de verosimilitud). El cociente de verosimilitud (fila cuatro) indica si el subconjunto tomado de las variables independientes, respecto del modelo que contiene solo la constante, ejercen influencia alguna en la variable dependiente. La capacidad de predicción en general para todos los modelo varía del 52% al 56% (fila cinco). El renglón a destacar es el que contiene las *medidas de información* (AIC y BIC), que serán de utilidad para determinar el mejor modelo, de los tres que se proponen. Se toma el modelo que presente el valor mínimo tanto del AIC como del BIC (fila seis y siete), esta condición la cumple el Modelo01, dado que el valor que toma AIC es 77,066 y BIC es 77,516, que corresponden al modelo completo.

5.2.4 Bondad de Ajuste

El Cuadro 5.11 indica y brinda una idea, de la proximidad del valor estimado \widehat{Y} respecto del valor observado Y . El cuadro es construido dividiendo la región de probabilidad en deciles.

Se obtiene que el valor del p – *valor* es 0.0, esto implica que no se rechaza la hipótesis nula, por tanto hay una falta de ajuste en el modelo. Sin embargo, realizando una inspección visual, se observa que los valores estimados \widehat{Y} están no “distan mucho” de los valores observados Y .

Cuadro 5. 11
Bondad de ajuste del modelo logit por sector

(Table collapsed on quantiles of estimated probabilities)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.0017	42	10.6	13591	13622.4	13633
2	0.0050	95	42.3	13538	13590.7	13633
3	0.0150	135	124.0	13496	13507.0	13631
4	0.0531	320	408.2	13313	13224.8	13633
5	0.1490	1166	1291.5	12466	12340.5	13632
6	0.3758	3374	3338.3	10258	10293.7	13632
7	0.6784	7405	7314.9	6228	6318.1	13633
8	0.8392	10446	10453.9	3186	3178.1	13632
9	0.9235	12065	12074.3	1567	1557.7	13632
10	0.9915	12967	12957.0	665	675.0	13632

```

number of observations = 136323
number of groups = 10
Hosmer-Lemeshow chi2(8) = 196.40
Prob > chi2 = 0.0000

```

Curva de ROC

En la base de datos de la ENOE06 que contiene 136,323 individuos, en la cual se pretende establecer si un sujeto trabaja dentro del sector formal o informal (variable dependiente), una vez que se determinaron las características de cada uno (variables independientes) se calcula a través de otro método su bondad de ajuste.

Una vez que se ajustó el modelo y se evaluó la significancia de los parámetros estimados ($\hat{\beta}$), ahora se evalúa la capacidad que éste tiene para diferenciar y clasificar a los individuos que laboran en el sector formal de aquellos que laboran en el sector informal, dado diferentes puntos de corte (c_p) con incrementos de 0.1 (cuadro 5.12).

Al igual que en la ENE96 en la ENOE06, la tasa de sensibilidad muestra los casos donde se dio una correcta predicción (\hat{Y}) de los individuos que verdaderamente laboran en el sector formal ($Y = 1$). La tasa de Sp muestra los casos donde se dio una predicción correcta (\hat{Y}) de aquellos individuos que verdaderamente están en el sector informal ($Y = 0$). Derivado de esto se muestra la curva de ROC.

Cuadro 5. 12

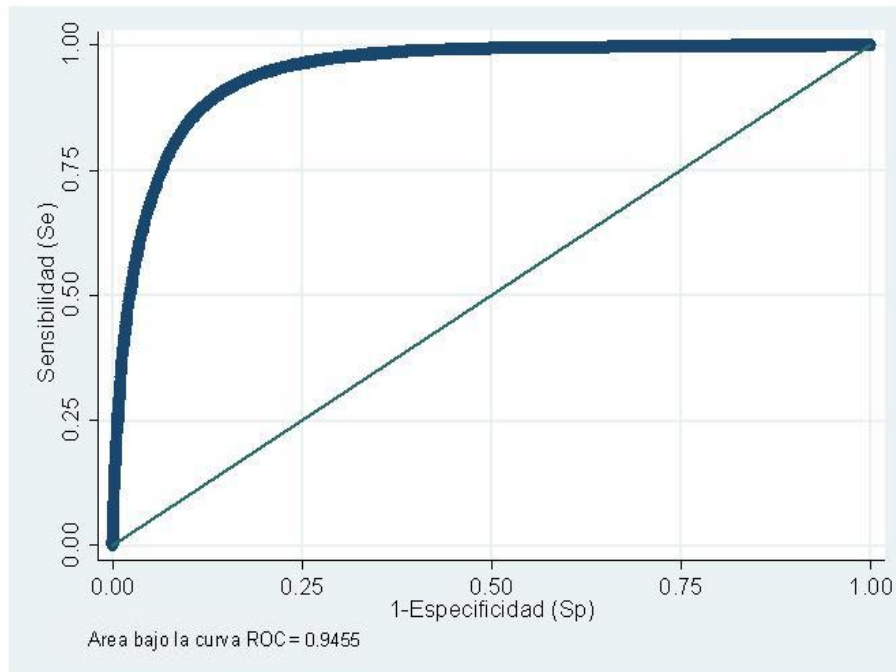
Tabla de clasificación del sector formal e informal con cortes de probabilidad de 0.1

Punto de Corte	Clasificados Positivos		Clasificados Negativos		Porcentajes			
	Casos positivos (D)	Casos negativos (~D)	Casos positivos (D)	Casos negativos (~D)	Se	Sp	1-Sp (falsos positivos)	Clasificados Correctamente
0.1	46,916	26,877	1,099	61,431	97.71%	69.56%	30.44%	79.48%
0.2	45,547	18,189	2,468	70,119	94.86%	79.40%	20.60%	84.85%
0.3	44,043	13,814	3,972	74,494	91.73%	84.36%	15.64%	86.95%
0.4	42,472	11,068	5,543	77,240	88.46%	87.47%	12.53%	87.81%
0.5	40,638	8,797	7,377	79,511	84.64%	90.04%	9.96%	88.14%
0.6	38,161	6,910	9,854	81,398	79.48%	92.18%	7.82%	87.70%
0.7	34,483	5,003	13,532	83,305	71.82%	94.33%	5.67%	86.40%
0.8	28,450	2,987	19,565	85,321	59.25%	96.62%	3.38%	83.46%
0.9	17,261	1,057	30,754	87,251	35.95%	98.80%	1.20%	76.66%
1	-	-	48,015	88,308	0.00%	100.00%	0.00%	64.78%

Analizando los datos de la quinta fila cuando c_p es 0.5, para aquellos cuya probabilidad estimada sea mayor que 0.5 el evento ocurrirá (clasificados positivos), por tanto el individuo estará en el sector formal; de la misma manera para aquellos cuya probabilidad sea menor a 0.5 el evento no ocurrirá (clasificados negativos), por tanto el individuo estará en el sector informal. Para el punto de corte 0.5, 40,638 casos de 49,435 se clasificaron correctamente como individuos que trabajan en el sector formal y 8,797 casos fueron clasificados incorrectamente, dado que estos individuos trabajan en el sector informal; ahora 79,511 casos de 86,888 fueron clasificados correctamente como individuos que trabajan en el sector informal, en tanto que 7,377 casos fueron clasificados incorrectamente, dado que estos individuos trabajan en el sector formal. Del modelos se obtiene que la sensibilidad (Se) es de 91.25%, la especificidad (Sp) es de 89.94% y los casos donde no hay una buen predicción de trabajadores, que se clasificaron como del sector formal, pero los datos observados reflejan que son trabajadores del sector informal (falsos positivos) fueron de 10.06%. Así que la sensibilidad es mayor que los falsos positivos. En general la tasa de los eventos que se clasifican correctamente (ocurrencia y no ocurrencia del evento) es del 88.14%.

En la Gráfica 5.4 se muestra qué modelo tiene una capacidad excelente ($AUC= 0.9455$) para discriminar entre los individuos que trabajan el sector formal de los que trabajan el informal, es decir que hay una mayor proporción de predicciones correctas ($\hat{\pi}$) que reflejan verdaderamente los casos donde el individuo labora en el sector formal ($\mathcal{Y} = 1$). La curva ROC para ambas encuestas (ENE96 y ENOE06) presentó una excelente capacidad para discriminar. Por tanto los datos se ajustan bien al modelo.

Gráfica 5. 4
Curva ROC



5.3 Resultados: Interpretación del modelo logit de 1996 Y 2006

Para interpretar los resultados del modelo logit y analizar el efecto de la reforma del pensiones del IMSS en el tamaño del sector formal, se hará uso del cociente de momios y los efectos marginales. Una vez que se calculen los cocientes de momios y los efectos marginales para ambas encuestas de empleo, se compararan entre si los resultados, con el fin de evaluar si algunas características influyeron más que otras para que el trabajador se formalizara.

Recordando que la variable dicotómica del sector formal e informal fue construida en base a la información individual del trabajado, que indica si su afiliación al IMSS es por parte de su empleo. Y como la cobertura del IMSS está muy relacionada con la situación del empleo en México, se puede determinar cómo se ha ido moviendo la probabilidad de ser formal de un año a otro con base a las características del individuo.

En éste caso la mayoría de las variables explicativas son dicotómicas, por consiguiente el cociente de momios muestra la tasa de probabilidad y el efecto marginal el cambio de probabilidad sobre las características del individuo para que este sea formal.

5.3.1 Cociente de momios

Los valores del cociente de momios (OR) expresan cuánto varía la razón de ocurrencia del suceso en función del cambio en las variables independientes, es decir, cuando la variable independiente aumenta en una unidad, cuánto varía la razón de pertenecer al sector formal.

Recordando que la ecuación del *cociente de momios* es:

$$OR = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}}$$

explica cuan más es probable que ocurra que el individuo pertenezca al sector formal ($Y = 1$) dado que presenta una serie de características.

Región

Ya se ha mencionado que el *Estrato* denota la región donde habita el individuo, de tal modo que se diseñan seis variables dicotómicas $_Iregion_e_i$; i toma los valores de 2 a 7. Se obtiene el cociente de momios para cada *Estrato*, usando a *Estrato_1* como categoría de referencia (Cuadro 5.13).

Calculando el cociente de probabilidad de ser formal por región respecto del *Estrato_1* se tiene que:

$$\begin{aligned} OR_{1996}(_Iregion_e_6, Estrato_1) &= e^{g(1)-g(0)} = \\ & e^{[\hat{\beta}_0 + \hat{\beta}_1(_Iregion_e_2=0) + \hat{\beta}_2(_Iregion_e_3=0) + \hat{\beta}_3(_Iregion_e_4=0) + \hat{\beta}_4(_Iregion_e_5=0) + \hat{\beta}_5(_Iregion_e_6=1) + \hat{\beta}_6(_Iregion_e_7=0)]} \\ & e^{-[\hat{\beta}_0 + \hat{\beta}_1(_Iregion_e_2=0) + \hat{\beta}_2(_Iregion_e_3=0) + \hat{\beta}_3(_Iregion_e_4=0) + \hat{\beta}_4(_Iregion_e_5=0) + \hat{\beta}_5(_Iregion_e_6=0) + \hat{\beta}_6(_Iregion_e_7=0)]} \\ & = e^{\hat{\beta}_5(_Iregion_e_6=1)} = e^{0.7589106(_Iregion_e_6=1)} = 2.1359 \end{aligned}$$

Entonces para 1996 el $OR_{1996}(_Iregion_e_6, Estrato_1)=2.1359$, para el 2006 el $OR_{2006}(_Iregion_e_6, Estrato_1)=2.5379$, lo que indica que laborar en el sector formal en la $_Iregion_e_6$ (Aguascalientes, Coahuila, Jalisco y Nuevo León) paso a ser de dos veces a dos veces y medio más probable.

Este incremento en la probabilidad de trabajar en el sector formal, para individuos que viven en la $_Iregion_e_6$ respecto del *Estrato_1*, se puede describir desde dos enfoques. Primero que la protección social aumentó en su conjunto en los estados de Aguascalientes, Coahuila, Jalisco y Nuevo León, o que disminuyó en los estados de Chiapas, Guerrero y Oaxaca. Siempre será favorable para el bienestar social, que la cobertura se incremente en estos estados, pero también se debe de señalar que sí disminuye la protección social para individuos que viven en estados de

mayor marginación social es preocupante, dado que la reforma de 1997 a la ley de seguridad social, buscaba que grupos vulnerables tuvieran incentivos para emplearse en un trabajo formal que les proporcionara seguridad social. La *_Iregion_e_5* también experimenta este incremento en el cociente de probabilidad, pasando de dos veces ($OR_{1996}=1.9283$) a tres veces más probable ($OR_{2006}=2.8969$) emplearse en el sector formal.

Tanto el *Estrato_1* como la *_Iregion_e_2* (Campeche, Hidalgo, Puebla, San Luis Potosí, Tabasco y Veracruz) tienen estados que se caracterizan por tener comunidades con alta marginación social. Un reflejo es que su cociente de probabilidad para emplearse en el sector formal es alrededor de uno, lo quiere decir que es igualmente probable estar en el sector formal en el *Estrato_1* como la *_Iregion_e_2*. No se observa un cambio significativo en el cociente de probabilidad de 1996 a 2006, por lo que puede ser señal de un estancamiento en la cobertura.

Donde se observa un ligero retroceso en la cobertura es en la *_Iregion_e_3* (Durango, Guanajuato, Michoacán, Tlaxcala y Zacatecas) y *_Iregion_e_4* (Colima, Estado de México, Morelos, Nayarit, Querétaro, Quintana Roo, Sinaloa y Yucatán) porque el cociente de probabilidad sobre la formalidad, es muy similar para ambos años.

En la *_Iregion_e_7* (Distrito Federal) se dio un retroceso importante en la decisión que toma un trabajador al buscar un empleo formal. Si bien es cierto que en 1996 había ligeramente una mayor probabilidad de estar en el sector formal ($OR_{1996}(_Iregion_e_7, Estrato_1)=1.2530$) que de no estarlo, a 2006 esto cambió, ya que un individuo que vive en los estados de Chiapas, Oaxaca, Guerrero muestra una mayor probabilidad de estar en el sector formal $OR_{2006}(_Iregion_e_7, Estrato_1)=0.7867$ que en el Distrito Federal.

Cuadro 5. 13
Cociente de probabilidad por región

		<i>Diseño de variables</i>					
<i>Región</i>		<i>_Iregion_e_2</i>	<i>_Iregion_e_3</i>	<i>_Iregion_e_4</i>	<i>_Iregion_e_5</i>	<i>_Iregion_e_6</i>	<i>_Iregion_e_7</i>
<i>Estrato_1 a/</i>		0	0	0	0	0	0
<i>Estrato_2 b/</i>		1	0	0	0	0	0
<i>Estrato_3 c/</i>		0	1	0	0	0	0
<i>Estrato_4 d/</i>		0	0	1	0	0	0
<i>Estrato_5 e/</i>		0	0	0	1	0	0
<i>Estrato_6 f/</i>		0	0	0	0	1	0
<i>Estrato_7 g/</i>		0	0	0	0	0	1
<i>OR</i>	<i>1996</i>	1.0759	1.4896	1.5775	1.9283	2.1359	1.2530
	<i>2006</i>	1.1740	1.3098	1.5314	2.8969	2.5379	0.7867

Nota: Cada Estrato se comprende de los siguientes estados; a/ Chiapas, Guerrero y Oaxaca; b/ Campeche, Hidalgo, Puebla, San Luis Potosí, Tabasco y Veracruz; c/ Durango, Guanajuato, Michoacán, Tlaxcala y Zacatecas; d/ Colima, Estado de México, Morelos, Nayarit, Querétaro, Quintana Roo, Sinaloa y Yucatán; e/ Baja California, Baja California Sur, Chihuahua, Sonora y Tamaulipas; f/ Aguascalientes, Coahuila, Jalisco y Nuevo León; g/ Distrito Federal.

Género

Hacia 1996 ser hombre o mujer, influía “ligeramente” en el cociente de probabilidad de estar empleado en el sector formal ($OR_{1996}(\text{Femenino, Masculino})=1.1514$), este hecho parece que

tiende acercarse a un cociente de probabilidad de uno, lo cual indica que es casi indistintamente ser hombre o mujer para emplearse en el sector formal $OR_{2006}(\text{Femenino, Masculino})=1.0536$.

Edad

La edad es una variable explicativa discreta, como resultado se obtuvo que el cociente de probabilidad para 1996 fue de 1.01635 y para 2006 fue de 1.02601, compara dicho cociente es arbitrario, porque dependerá de la edad que se elija. Por ejemplo comparando el cociente de probabilidad en la formalidad para un individuo que tiene 18 años respecto de otro que tiene 40 es de 1.33 y 1.139 respectivamente, cambia esta probabilidad cuando cambias las edades, por ejemplo si se compara la edad de 25 respecto de 52 estas es de 1.11 y 1.5 respectivamente.

Estado Conyugal

El núcleo familiar del trabajador pudiera jugar un papel importante en la formalidad del trabajador, pero en realidad no se alcanza a notar una significativa diferencia sobre el cociente de probabilidad de cada estatus conyugal respecto a los *solteros* (Cuadro 5.14). Se percibe que de 1996 a 2006, la probabilidad de formalidad descendió en cada estado conyugal respecto de los *solteros*, con excepción de los separados.

Sí el trabajador se encuentra *casado*, presentará el mayor cociente de probabilidad de aproximadamente 1.5 veces de ser formal respecto a los *solteros*, para ambos años. Sin embargo de 1996 a 2006 este cociente decrece mínimamente. Detrás de esta categoría siguen los divorciados.

Cuadro 5. 14
Cociente de probabilidad por estado conyugal

Estado Conyugal		Diseño de variables				
		_ledo_conyu_1	_ledo_conyu_2	_ledo_conyu_3	_ledo_conyu_4	_ledo_conyu_5
Unión Libre		1	0	0	0	0
Separado		0	1	0	0	0
Divorciado		0	0	1	0	0
Viuado		0	0	0	1	0
Casado		0	0	0	0	1
Soltero		0	0	0	0	0
OR	1996	1.204124	1.06256	1.454307	1.114269	1.520169
	2006	1.148416	1.123067	1.286438	1.088765	1.504867

Nivel de Educación

En la medida que el individuo tenga mayor preparación académica el cociente de probabilidad de laborar en el sector formal se va incrementando. En el Cuadro 5.15 muestra que si se cuenta con un grado académico, de al menos la primaria completa, respecto de tener la primaria incompleta, el individuo posiblemente se encuentre laborando en el sector formal. Por ejemplo el

OR_{1996} (Medio superior y sup, primaria incompleta)=2.0483 indica que se estima que laborar en el sector formal es dos veces más probable para los individuos que tienen un grado académico superior al nivel bachillerato que de sólo tener la primaria incompleta, está relación a 2006, pasó a ser tres veces ($OR_{2006} = 2.99$). Lo favorable sobre esta característica del trabajador es que el nivel académico ha influido positivamente en el cociente de probabilidad de la formalidad, dado que esta se ha incrementado para todas las categorías.

Cuadro 5. 15
Cociente de probabilidad por nivel de instrucción

Nivel de Instrucción	Diseño de variables			
	_Inivel_ins_2	_Inivel_ins_3	_Inivel_ins_4	
Primaria incompleta	0	0	0	
Primaria completa	1	0	0	
Secundaria completa	0	1	0	
Medio superior y superior	0	0	1	
OR	1996	1.3373	2.0321	2.0483
	2006	1.4490	2.4558	2.9913

Por tanto es de esperar que los individuos que tienen una mayor preparación académica, estén afiliados al IMSS. Con ello no se quiere decir que si se tiene un mayor grado académico seguramente contara con seguridad social (IMSS), sino lo que se intenta interpretar, que es más probable que si cuente con seguridad social.

Lo anterior muestra que la población que no tiene preparación académica que usualmente son grupos frágiles, no se están inclinando por buscar un trabajo que le brinde seguridad social, o no han encontrado uno que se las proporcione. Estos grupos seguirán estando excluidos de la seguridad social y en consecuencia no podrán acceder a una pensión. Aun más al ser trabajadores que tienen salarios bajos por la actividad que desempeñan, no tendrán un ahorro que los pueda proveer de ingresos para su vejez destinados a caer un mas en la pobreza. A pesar que la tasa de reemplazo es del 100 por ciento para trabajadores que cotizan con un salario mínimo, éste no ha sido incentivo claro para trabajadores que tienen una menor preparación académica y que se emplean en trabajos precarios con ausencias de protección social.

Tipo de ocupación⁴¹

La categoría de referencia que se toma para comparar es *Trabajador por cuenta propia*, ya que estos trabajadores comúnmente están en el sector informal (Cuadro 5.16). Los individuos que desempeñaban el puesto de Patrón o Empleador en 1996 tenían un cociente de probabilidad

⁴¹ Los conceptos que se presentan a continuación son comparables, dado que dentro de los cuestionarios de la ENE (1996) y ENOE (2006), en la encuesta de empleo se pueden usar indistintamente. *Patrón con Empleador. Trabajador por su cuenta con Trabajadores por cuenta propia. Trabajador a destajo, comisión, porcentaje; a sueldo fijo, salario o jornal con Trabajadores subordinados y remunerados.*

ligeramente más alta de ser formales respecto de los que trabajaban por su cuenta, mostrando un cociente de momio de $OR_{1996}(\text{Patrón, Trabajador por su cuenta})=1.2537$. Sin embargo para 2006 este hecho se ha invertido, por que los *Trabajadores por cuenta propia* mostraron una mayor preferencia por emplearse en el sector formal, ya que el $OR_{2006}(\text{Empleador, Trabajador por su cuenta})=0.8734$. Este no es un buen síntoma acerca del comportamiento que los empleadores están teniendo frente a la protección social, porque ello muestra que los mencionados incentivos generados bajo la reforma le son insuficientes o poco atractivos para que se afilien al IMSS.

Los *Trabajadores subordinados y remunerados* son los que muestran un mayor cociente de probabilidad de estar en el sector formal. La tasa en 1996 ($OR_{1996}=44.6212$) es aproximadamente dos veces mayor que la de 2006 ($OR_{2006}=20.0522$); a pesar de seguir siendo alta, comparando ambos periodos, los *Trabajadores subordinados y remunerados* aparentemente cada vez, se emplean menos en una actividad en el sector formal que los provea de seguridad social. Esto puede ser un indicador sobre el comportamiento del trabajador, donde puede existir una inclinación a pensar que la cuota a la seguridad social es más mayor a los beneficios recibidos. Es por ello que acepten trabajar bajo una relación contractual, en la que pudieran obtener mayores ingresos o algún otro beneficio, sacrificando con ello su inclusión en la seguridad social. Aun más puede estar sucediendo que, los empleadores que están generando y ofreciendo este tipo de empleos sin protección, vean que contribuir a la seguridad social, es más una carga fiscal que un beneficio social.

Cuadro 5. 16
Cociente de probabilidad por tipo de ocupación

<i>tipo_ocupacion (1996)</i>	<i>Diseño de variables</i>			
	<i>_ltipo_ocup_1</i>	<i>_ltipo_ocup_4</i>	<i>_ltipo_ocup_5</i>	
<i>Patrón</i>	1	0	0	
<i>Trabajador por su cuenta</i>	0	0	0	
<i>Trabajador a de stajo, comisión, porcentaje; a sueldo fijo, salario o jornal</i>	0	1	0	
<i>Miembro de una cooperativa</i>	0	0	1	
OR	1996	1.2537	44.6212	43.9047

<i>tipo_ocupacion (2006)</i>	<i>Diseño de variables</i>			
	<i>_ltipo_ocup_1</i>	<i>_ltipo_ocup_2</i>	<i>_ltipo_ocup_4</i>	
<i>Trabajadores subordinados y remunerados</i>	1	0	0	
<i>Empleadores</i>	0	1	0	
<i>Trabajadores por cuenta propia</i>	0	0	0	
<i>Trabajadores sin pago</i>	0	0	1	
OR	2006	20.0522	0.8734	0.3404

Jornada y Número de empleados

Otras de las características del empleo en el sector formal, tiene que ver con la jornada laboral y el número de trabajadores que laboran en la misma empresa (Cuadro 5.17).

Cuadro 5. 17
Cociente de probabilidad por jornada laboral

<i>jornada_laboral</i>	<i>Diseño de variables</i>		
	<i>_lhrs_traba_2</i>	<i>_lhrs_traba_3</i>	
<i>Menos de 35 horas</i>	0	0	
<i>De 35 a 48 horas</i>	1	0	
<i>Más de 48 horas</i>	0	1	
<i>OR</i>	<i>1996</i>	<i>3.3671</i>	<i>2.4003</i>
	<i>2006</i>	<i>3.3235</i>	<i>2.3619</i>

Los trabajadores que contractualmente tienen una jornada laboral de 8 hrs. diarias (estipulado por ley), son aquellos que aproximadamente son 3 veces más probable de estar en un trabajo formal que de no estarlo, esto sucede en ambos periodos (1996 y 2006), lo anterior es respecto de aquellos que trabajan menos de 35 hrs a la semana. En tanto aquellos que trabajan contractualmente más de la jornada laboral, es 2 veces más probable que trabajen en el sector formal, de aquellos que trabajan menos horas de las establecidas por ley.

Por otro lado obsérvese que el número de trabajadores que laboran dentro de una misma institución o empresa es importante, ya que esto determina una probabilidad muy alta, para que su empleo del trabajador sea formal (Cuadro 5.18). A medida que exista un mayor número de trabajadores en la empresa, esta tendrá más trabajadores que tienen un empleo formal.

Cuadro 5. 18
Cociente de probabilidad por número de trabajadores

<i>Número de Trabajadores</i>	<i>Diseño de variables</i>					
	<i>_lc_trabaja_2</i>	<i>_lc_trabaja_3</i>	<i>_lc_trabaja_4</i>	<i>_lc_trabaja_5</i>	<i>_lc_trabaja_6</i>	
<i>1 persona</i>	0	0	0	0	0	
<i>2 a 5 personas</i>	1	0	0	0	0	
<i>6 a 10 personas</i>	0	1	0	0	0	
<i>11 a 15 personas</i>	0	0	1	0	0	
<i>16 a 50 personas</i>	0	0	0	1	0	
<i>51 y mas</i>	0	0	0	0	1	
<i>OR</i>	<i>1996</i>	<i>5.1542</i>	<i>19.2362</i>	<i>31.6753</i>	<i>64.4115</i>	<i>109.5847</i>
	<i>2006</i>	<i>2.9412</i>	<i>11.2482</i>	<i>21.5251</i>	<i>37.0758</i>	<i>96.3401</i>

Sin embargo el cociente de momios sobre la formalidad en el número de trabajadores que laboran en una misma empresa mayor a dos, ha descendido significativamente de 1996 a 2006, respecto de los que el número de trabajadores es uno. Para interpretar el cociente de momios, la variable de comparación son las empresas que contienen un solo trabajador, que es el caso de

trabajadores independientes. Por ejemplo en 1996 los individuos que laboraban en empresas cuya platilla de trabajadores era mayor a cincuenta y uno, presentaban un cociente de probabilidad de $OR_{1996}(1 \text{ persona, } 51 \text{ y mas})=109.4857$, a 2006 este cociente de probabilidad paso a ser de $OR_{2006}(1 \text{ persona, } 51 \text{ y mas})=96.3401$.

El cambio en el cociente de momios puede señalar que empresas medianas y grandes están ofreciendo empleos sin seguridad social provocando un retraso en la cobertura. Un tipo de contratación bajo este esquema sin seguridad social son los trabajadores por honorarios. Indudablemente estos trabajadores al no estar bajo el esquema de seguridad social, no tienen cotizaciones de parte de estado y patrón que contribuyan al aseguramiento de un ingreso para su vejez.

Sistema de Clasificación Industrial de América del Norte (SCIAN)

Ahora se describe el cociente de probabilidad de estar en el sector formal, dado la actividad económica donde trabaja el individuo. Las actividades económicas que presentan la mayor probabilidad en la informalidad es en las actividades de *Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza*, es por ello se tomara a esta actividad como la categoría de referencia para compararla respecto de las demás actividades, para así determinar el cociente de probabilidad. El Cuadro 5.19 muestra el cociente de momios por actividad económica para 1996 y 2006.

En general de 1996 a 2006, el modelo estimó que el cociente de probabilidad de un periodo a otro muestra un cambio. El cambio consistió que en 2006 emplearse en cualquier actividad resultaría más probable que fuese formal, que si se empleaba en la Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza, cosa que no ocurría en 1996, porque en este año había otras actividades que indicaban que era menos probable de tener un empleo formal.

El mayor crecimiento (1996 a 2006) en el cociente de probabilidad fue en la actividad de Servicios financieros y de seguros, ya que paso de la mitad a diez veces más probable que el trabajador fuese formal en esta actividad; seguida de la actividad de Comercio al por mayor pasando de dos y media veces a once veces más probable de ser formal. Las dos actividades con menor crecimiento en el cociente de probabilidad sobre la formalidad fueron los Servicios de esparcimiento cultural y deportivo, y otros servicios recreativos y Otros servicios excepto actividades gubernamentales.

Cuadro 5. 19
Cociente de probabilidad por actividad económica (CAEM – SCIAN)

Actividad	_Iscian_ene_21	_Iscian_ene_22	_Iscian_ene_23	_Iscian_ene_3133	_Iscian_ene_43	_Iscian_ene_46	_Iscian_ene_4849	_Iscian_ene_51	_Iscian_ene_52	_Iscian_ene_53	_Iscian_ene_54	_Iscian_ene_56	_Iscian_ene_61	_Iscian_ene_62	_Iscian_ene_71	_Iscian_ene_72	_Iscian_ene_81	_Iscian_ene_93	_Iscian_55	
Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Minería	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Construcción	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Industrias manufactureras	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Comercio al por mayor	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Comercio al por menor	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
Transportes, correos y almacenamiento	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
Información en medios masivos	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
Servicios financieros y de seguros	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
Servicios inmobiliarios y de alquiler de bienes muebles e intangibles	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
Servicios profesionales, científicos y técnicos	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
Servicios educativos	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
Servicios de salud y de asistencia social	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
Servicios de alojamiento temporal y de preparación de alimentos y bebidas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
Otros servicios excepto actividades gubernamentales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
Actividades legislativas, gubernamentales, de impartición de justicia y de organismos internacionales y extraterritoriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
Corporativos a/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
OR	1996	0.8428	4.7104	1.8761	3.9324	2.5244	4.7350	2.6964	4.3652	0.5635	4.1609	4.4277	5.8097	1.0610	2.0147	1.5106	3.8387	2.0869	0.4763	
	2006	7.3698	11.5219	3.1296	6.6805	10.9877	8.6499	4.1677	6.9778	10.3865	5.8803	6.8148	9.0546	5.8190	6.1599	2.4867	5.8330	3.1918	1.8820	10.0073

a/ Aplica solo para la ENOE06

5.3.2 Efectos marginales

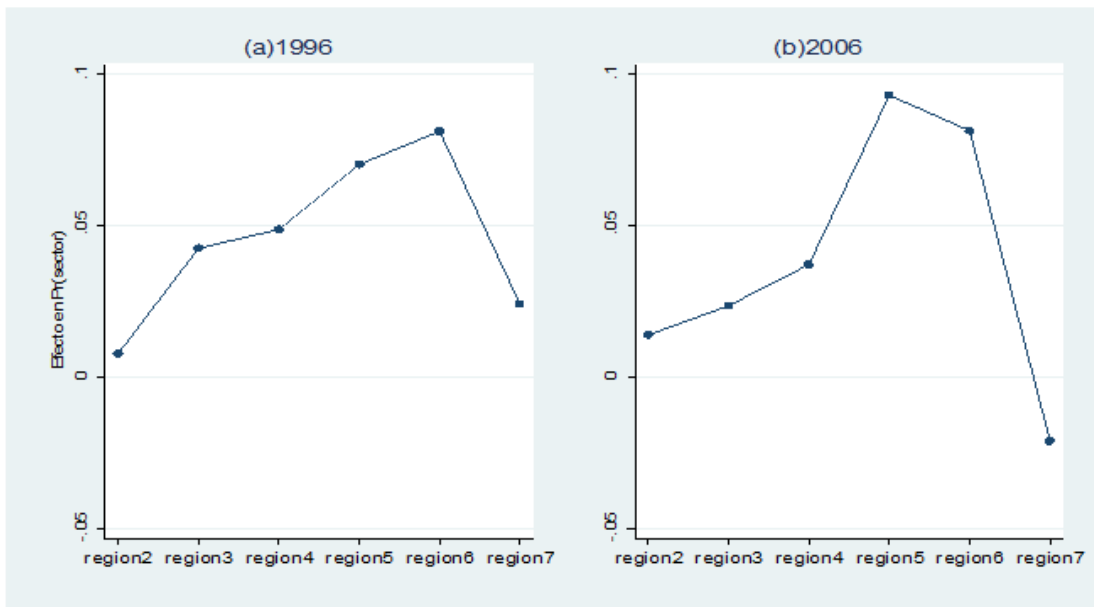
Para calcular los efectos marginales, se toma el promedio de los efectos marginales, cuya ecuación es la siguiente:

$$\frac{f(g(x))\hat{\beta}}{N} = \frac{\left[\frac{e^{-g(x)}}{(1 + e^{-g(x)})^2} \right] \hat{\beta}_i}{N}$$

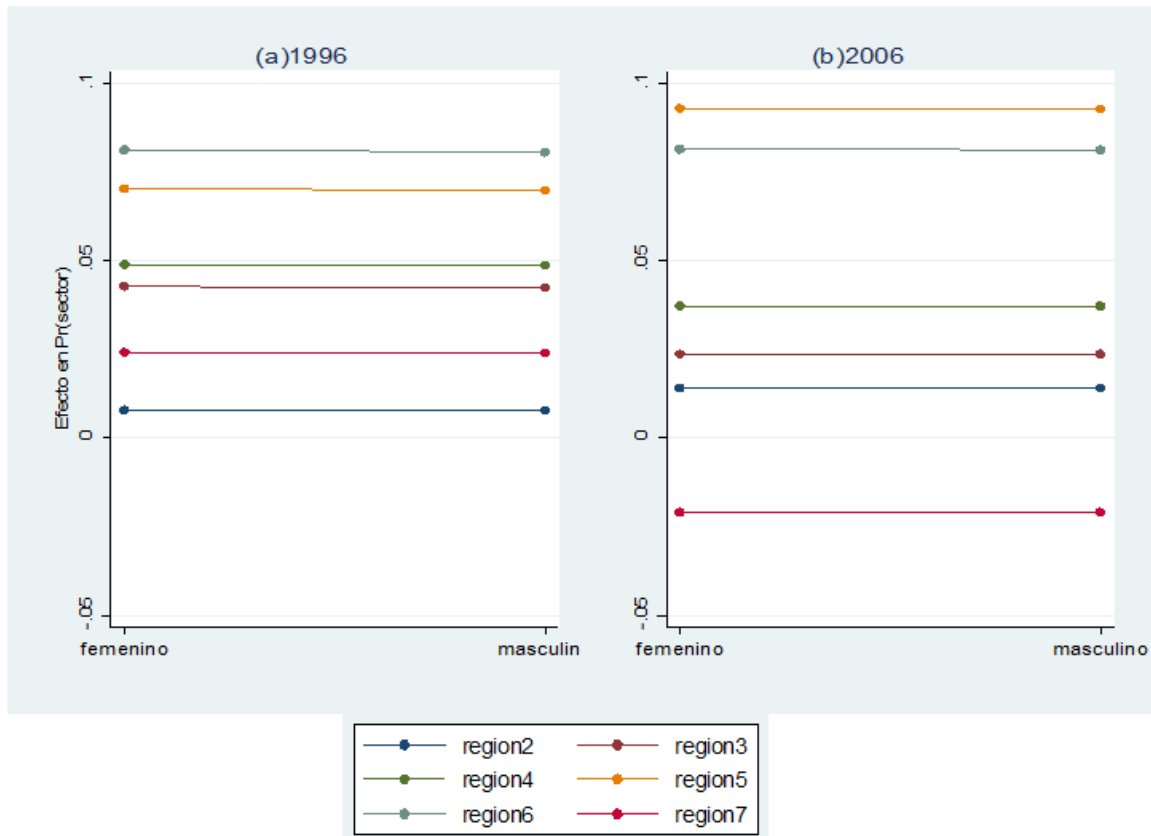
Como las variables independientes en su mayoría son categóricas, el efecto marginal mostrará un cambio en la probabilidad que el individuo este en el sector formal. Y como en el caso del cociente de momios, la probabilidad afecta directamente sobre la protección social del trabajador por medio del IMSS.

En la Gráfica 5.5 muestra el efecto marginal por región de estar en un empleo formal. En 1996 de la región 3 a 6 guardaba una probabilidad superior al 4 por ciento respecto de la región 1 de estar en el sector formar que de no estarlo. En 2006 el escenario cambio, dado que la región 5 y 6 mantuvieron un crecimiento, pasando a una probabilidad de 9 y un poco más del 8 por ciento respectivamente, en tanto que la probabilidad de las demás regiones disminuyó, con excepción de la región 2. Además las probabilidades por región no se ven afectadas significativamente por el género para ambos años, decir es casi igualmente probable estar en el sector formal en cualquier región y sin importar si se es hombre o mujer (Gráfica 5.6).

Gráfica 5. 5
Efectos marginales por región



Gráfica 5. 6
Efectos marginales por género y región



Cabe mencionar que la región 5 y 6 integrados en su mayoría por estados del norte –con excepción de Aguascalientes-, mantuvieron un crecimiento en la probabilidad de trabajadores que cuenta con protección social. Así mismo la región 2 experimento levemente lo anterior. Pero la región 1 que es una de las más vulnerables –si no es que el más- siguen careciendo de tener trabajadores con protección social, lo cual los limita a tener acceso en un plazo no muy largo, a una pensión. En la región 2 como se sabe también existen estados como una alta población desprotegida –por ejemplo Veracruz-, dicha región presenta el efecto marginal positivo más bajo respecto a todas las demás regiones para ambos años. Esto nos dice que los trabajadores de la región 1 tendrán una probabilidad de tener seguridad social muy próxima a la región 2. Entonces la reforma al esquema de pensiones en estas regiones (1 y 2) con población frágil no ha contribuido significativamente a formalizar el empleo. Y en consecuencia no cuentan con protección social que les proporcione una pensión además de otros beneficios.

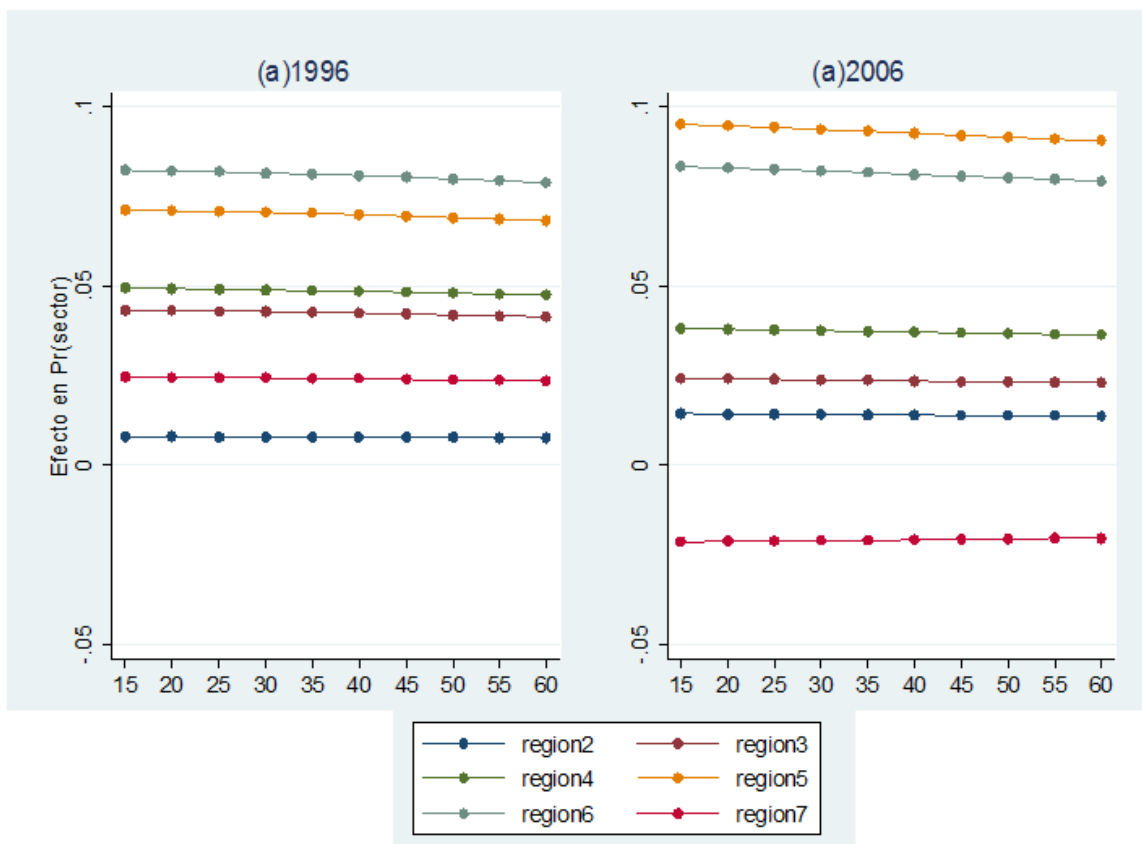
De 1996 a 2006, la región 5 (Baja California Sur, Chihuahua, Sonora y Tamaulipas) creció un poco más del 2 por ciento en la probabilidad de ser formal, siendo ésta la que más creció. En 1996 tener un empleo formal en la región 7 (Distrito Federal) representaba una probabilidad positiva mayor al 2 por ciento, en comparación del grupo de trabajadores de la región 1 (Chiapas,

Guerrero y Oaxaca), sin embargo para 2006 esto se ha invertido y ahora es menos probable emplearse en el sector formal en la región 7 que en la región 1. Aun más el efecto positivo para ser formal disminuyó un poco más del 1 por ciento para las regiones 3 y 4.

Entonces como disminuyó el efecto de los trabajadores sobre la decisión de emplearse en un trabajo formal, que les provea de seguridad social, esto incurre en un retroceso en la cobertura de seguridad social (región 3, 4 y 7). Pero resulta interesante destacar que a 2006 es más probable ser formal si el trabajador radica en la región 1 que en la región 7.

Para todas las regiones, mientras mayor edad tenga el trabajador, la probabilidad de ser formal va disminuyendo de manera similar para ambos años (Gráfica 5.7). Al tener el mismo efecto para ambos años se observa que siguen los trabajadores de edad adulta teniendo una dificultad de reintegrarse al sector formal, afectando su pensión de vejez y la tasa de reemplazo. Siendo que con la reforma de pensiones se esperaba que trabajadores de edad adulta se reintegrarán al sector formal.

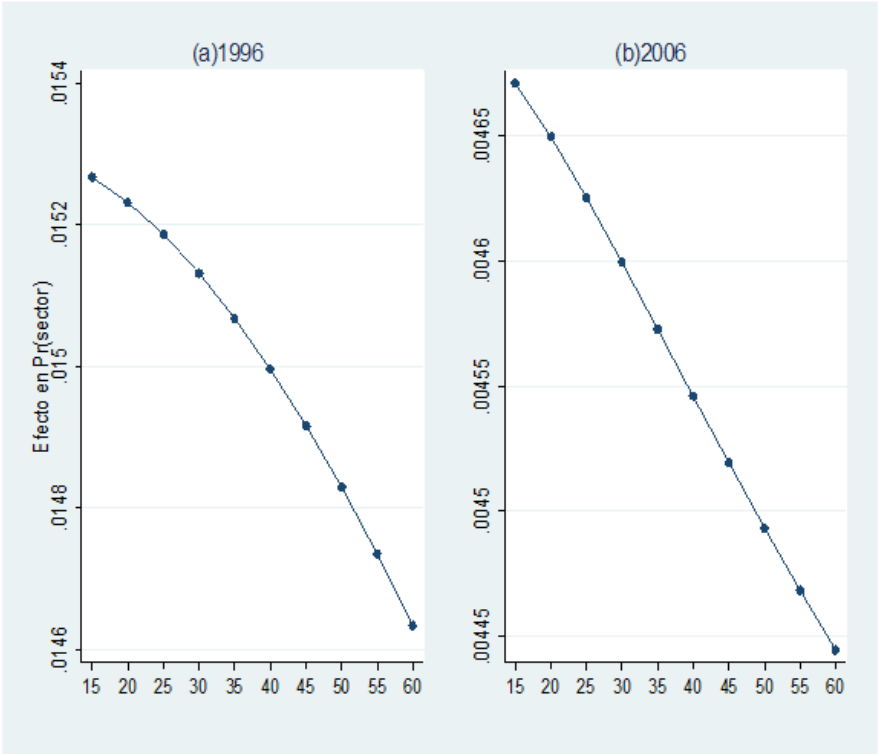
Gráfica 5. 7
Efectos marginales por edad y región



Observando cómo se comporta el efecto del género respecto la edad, hay dos situaciones sobre la probabilidad de ser formal (Gráfica 5.8). La primera es que a medida que avanza la edad del trabajador masculino, la probabilidad de estar en el sector formal va decreciendo ligeramente (hecho que se había tocado anteriormente). La segunda nos dice que en 1996 los hombres respecto de las mujeres, tenían una probabilidad superior que giraba alrededor del 1.5 por ciento de estar empleados en el sector formal (Gráfica 5.8a). Este efecto disminuye en 2006, por que la brecha de probabilidad de los hombres respecto a la mujer, se reduce a menos del 0.5 por ciento (Gráfica 5.8b).

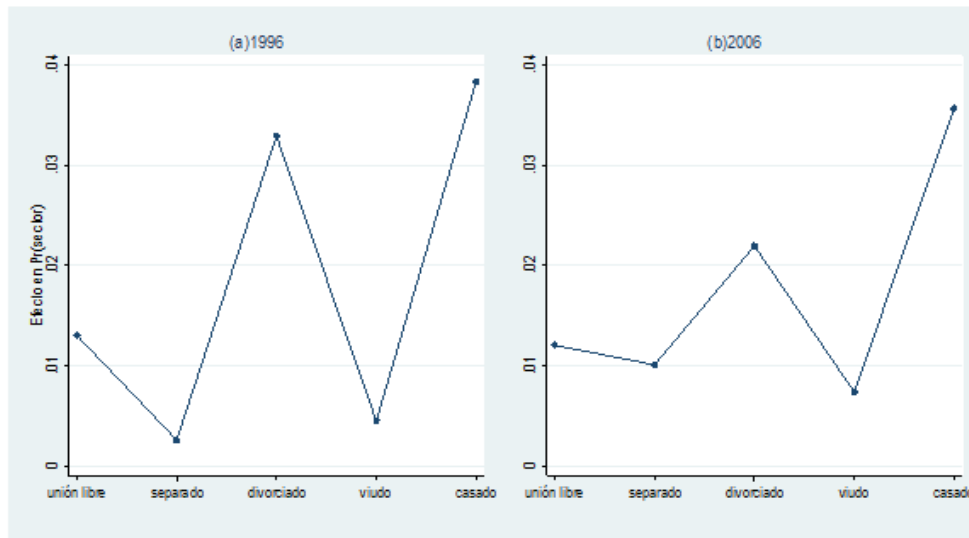
Al ver que el efecto positivo de ser formal se reduce de 1996 a 2006 dado el género y la edad, se tiene que es casi muy similar en probabilidad que el hombre o una mujer estén en el sector formal.

Gráfica 5. 8
Efectos marginales por género y edad



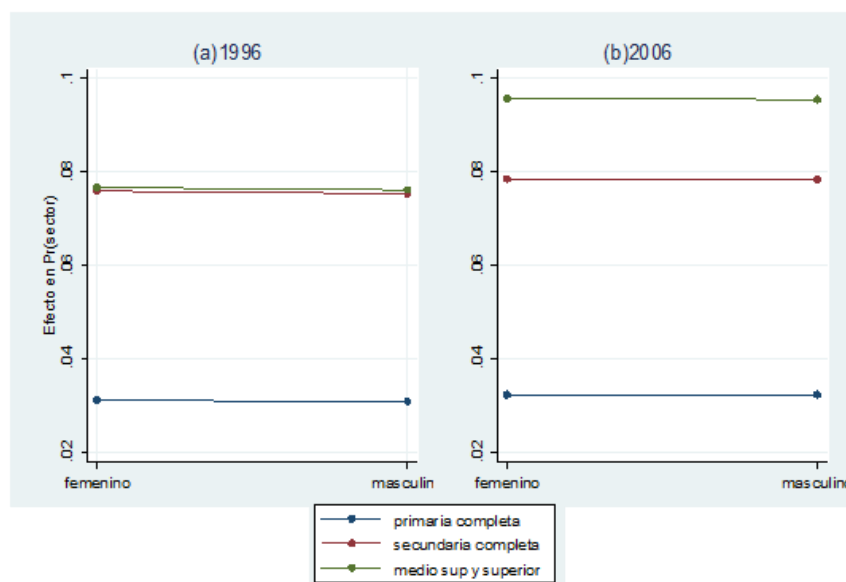
La condición de estado conyugal de trabajador respecto a los solteros, puede influir positiva o negativamente en la probabilidad de ser formal. De 1996 a 2006 se ha dado un crecimiento en el efecto marginal positivo en la formalidad cuando el individuo es separado o viudo, y un descenso cuando se está en unión libre, divorciado o casado. El mayor descenso en la probabilidad fue para el caso cuando se es divorciado con 1 por ciento, en tanto todos los demás están por debajo de este porcentaje (Gráfica 5.9).

Gráfica 5.9
Efectos marginales por estado conyugal



El efecto de la educación tiende a aumentar la probabilidad de ser formal de manera similar para ambos géneros (Gráfica 5.10). La probabilidad de ser formal en 1996, aumenta a 3 por ciento, si se cuenta con la primaria completa; y aumenta oscilando alrededor del 7.5 por ciento si cuenta con al menos la secundaria concluida. En el 2006 el efecto marginal se mantiene casi muy similar a lo que se venía experimentando en 1996, la excepción sucede para los individuos que cuentan con al menos el nivel medio superior, porque aumenta un poco más del 9.5 por ciento la probabilidad de ser formales, todo esto respecto de quienes tienen el nivel más bajo de educación (Primaria incompleta).

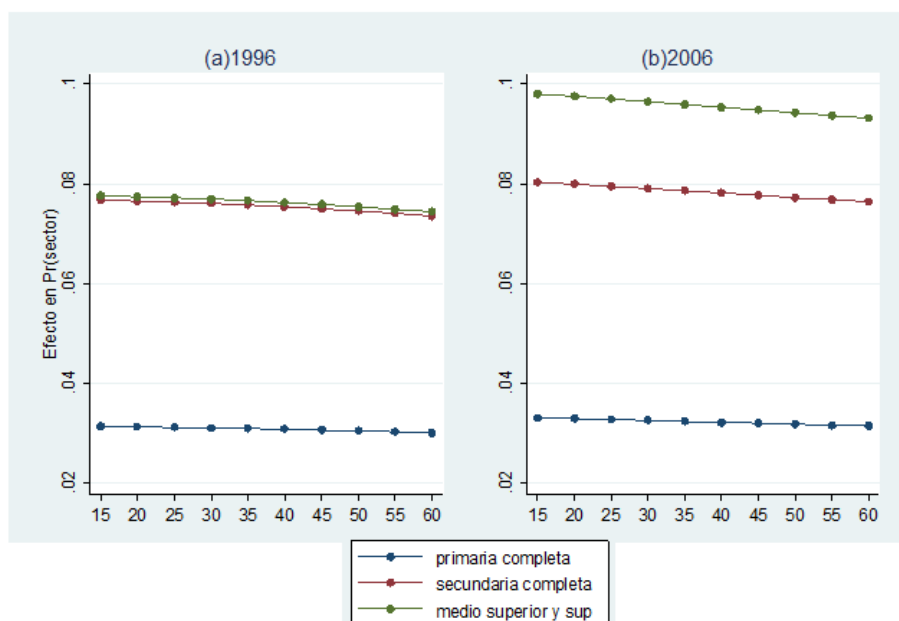
Gráfica 5.10
Efectos marginales por género y nivel de instrucción



Entonces el nivel de educación incrementa la probabilidad de ser un empleado formal, siendo muy similar la probabilidad para ambos géneros. Aunque el efecto marginal para ambos periodos es muy similar, con excepción de aquellos que cuentan con al menos el nivel medio superior –donde en el segundo año hay un mayor incremento-, no hay movimientos significativos a favor de ser un empleado formal de un periodo (1996) a otro (2006).

A pesar de que el trabajador tenga algún nivel de instrucción, la edad sigue ejerciendo influencia en la disminución del efecto marginal positivo para que el individuo sea formal (Gráfica 5.11). Aunque esta disminución en el efecto sobre la probabilidad se refleja más, para los trabajadores cuyo nivel de educación es al menos el nivel medio superior. Un aumento en la edad y en el nivel de instrucción, es una conjunción de factores que juegan en contra de ser un empleado formal, una de las tantas opciones sea deba a que algunos trabajadores al haber acumulado capital humano y financiero decidan trabajar como independientes o sean empleadores, dejando a un lado los beneficios que pudieran recibir de la protección social, que sólo te lo da un empleo formal

Gráfica 5. 11
Efectos marginales por edad y nivel de instrucción

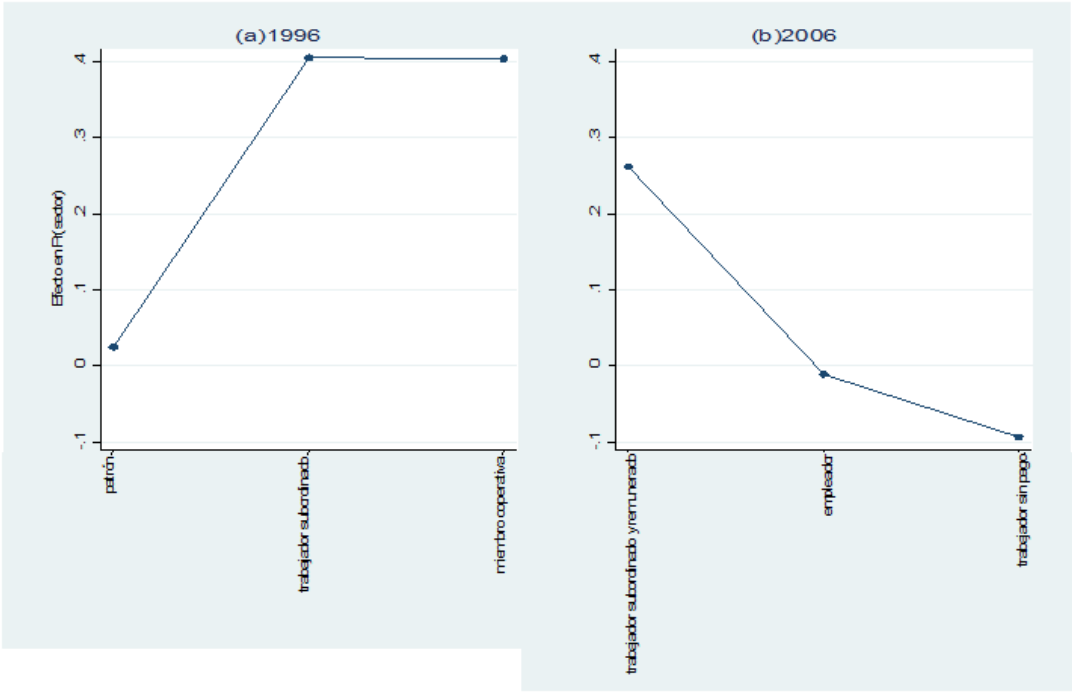


La Gráfica 5.12 expone el efecto marginal de ser formal dado el tipo de ocupación, tomando como referente a los trabajadores por cuenta propia. La condición de ser patrón causa un efecto de 2.5 por ciento en la probabilidad de ser formal; sí el individuo es un trabajador subordinado, éste tendrá 40 por ciento de probabilidad de ser formal, representando el efecto más alto para el primer periodo (1996). Sobre el segundo periodo (2006) sí el individuo tiene la condición de ser un trabajador subordinado, la probabilidad de ser formal (26.15 por ciento) sigue

siendo alta, pero esta disminuye en un 14.37 por ciento; sí presenta la condición de ser empleador el efecto es negativo, es decir es más probable que un trabajador por cuenta propia sea formal a uno que es empleador, la misma situación ocurre para los trabajadores sin pago.

Hasta antes de la reforma al esquema de pensiones la condición de ser empleador mostraba aunque fuese mínimo un efecto positivo sobre la formalidad, pero en el segundo periodo se ve un efecto negativo sobre la probabilidad de la formalidad. Esto es reflejo de un retroceso en la formalización del empleo en los diferentes tipos de ocupación. Tal vez el empleador no caiga en estado de pobreza en el futuro, ante las diversas contingencias -sobre las cual brinda protección él IMSS-, ello al contar con la solvencia económica para afrontar dichas adversidades, pero si es importante que se afilien al IMSS para que en conjunto se pueda alcanzar un estado de bienestar. La situación es más relevante cuando los trabajadores subordinados de un periodo (1996) a otro (2006) -que son la mayoría en el país-, cada vez menos busquen o encuentre un empleo que los pueda afiliar al IMSS y pueda gozar de una pensión o alguna otra prestación estipulada en la ley del seguro social.

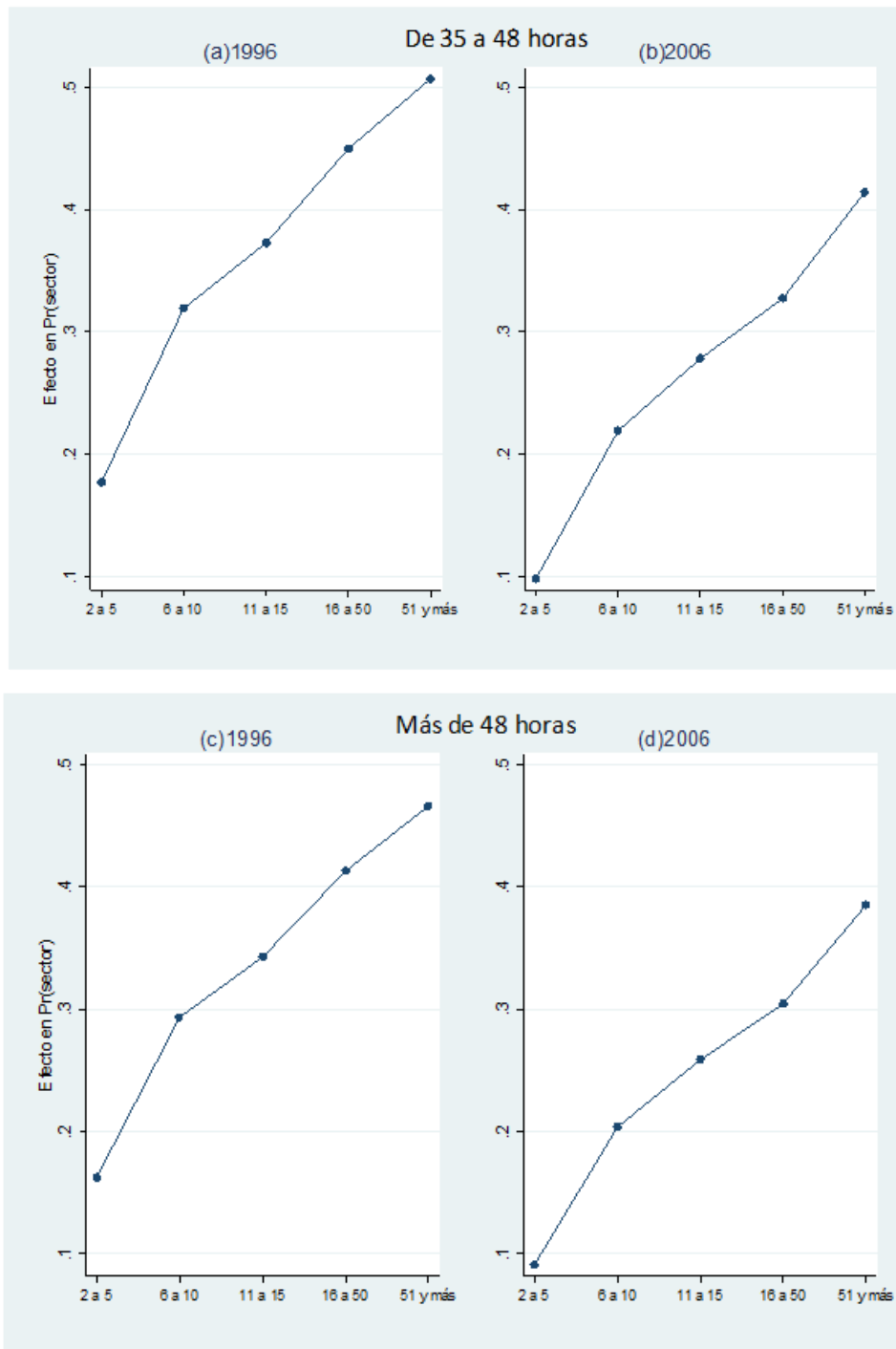
Gráfica 5. 12
Efectos marginales por tipo de ocupación



Analizando el efecto marginal del lugar donde labora el trabajador y la jornada laboral, se deduce lo siguiente (Gráfica 5.13). Hay un efecto positivo creciente en la probabilidad de ser formal, en tanto el trabajador labore en una empresa cuya plantilla laboral sea cada vez mayor, o dicho de otra manera la empresa tendrá un mayor probabilidad de ser formal si su plantilla de

trabajadores es cada vez mayor, aunque la probabilidad será afectada por el número de horas que labora en la empresa.

Gráfica 5. 13
Efectos marginales por número de trabajadores y jornada laboral



Se puede observar en 1996 un efecto desfavorable en el sector formal aproximadamente mayor a 1.5 por ciento –dependiendo del grupo comparado-, sí la jornada laboral excede el intervalo de 35 a 48 horas semanales (alrededor de 8 horas diarias) (Gráfica 5.13 a y c). En 2006 existe también este efecto desfavorable para individuos que trabajan más horas del intervalo antes señalado, pero en un porcentaje más variado (Gráfica 5.13 b y d).

Por tanto si el individuo trabaja dentro de una jornada laboral de 35 a 48 horas semanales tendrá una mayor probabilidad de ser formal. Pero en 2006 se ha presentado un descenso en el efecto marginal respecto de 1996 sobre la jornada laboral y el número de trabajadores que constituyen la empresa. Ello indica que a 2006 las empresas que tiene más de dos trabajadores está bajando la probabilidad de que el empleo que genera y ofrecen sea formal, y en consecuencia no los adhiera al IMSS, este descenso es mayor cuando la empresa está constituida de 16 a 50 trabajadores. Esto puede dar indicios, de un retroceso en la generación y ofrecimiento del los empleos formales, dejando desprotegidos a la clase trabajadora ante contingencias que cubre la ley del seguro social y excluyéndoles del derecho en el largo plazo a una pensión para su vejez.

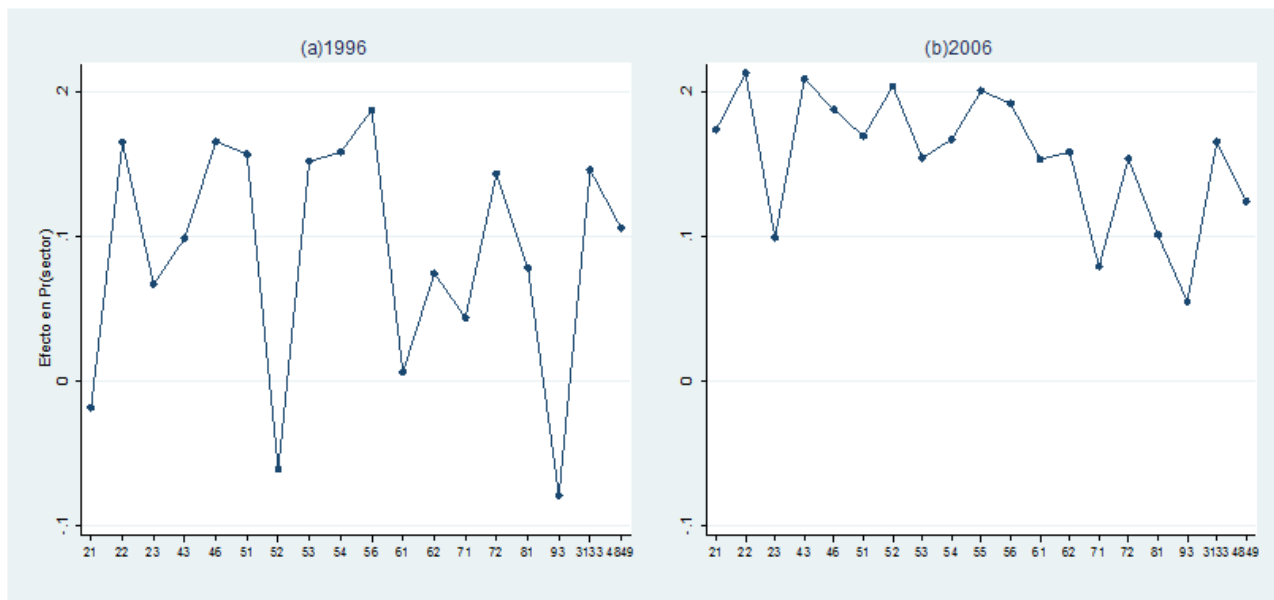
De 1996 a 2006 se dio un crecimiento en el efecto sobre la probabilidad de ser formal en todas las actividades respecto la actividad 11. En 1996 dependiendo del sector de actividad donde se encuentre el trabajador, existirá un efecto positivo o negativo en la probabilidad de ser formal (Gráfica 5.14). Para las actividades 21, 52 y 93 muestran un efecto marginal negativo en la formalidad del trabajador, en tanto el resto de las actividades muestran un efecto positivo en la formalidad. En 2006 todas las actividades, muestran un efecto positivo en la probabilidad de ser formal. La actividad que representó el mayor crecimiento en la formalidad del empleo fue el de 52 con 27 por ciento, seguida de la 21 con 19 por ciento. En tanto el menor crecimiento se dio en la actividad 53 y 56.

Con los hallazgos encontrados, se puede decir que el estancamiento y el retroceso en la probabilidad sobre la formalidad del empleo es promovida por una responsabilidad compartida entre el estado, patrón y trabajador. Al patrón (independientemente de tamaño de la empresa) por que se le ha permitido generar y ofrecer empleo sin protección social. Al trabajador por que en algunas ocasiones su decisión lo lleva a emplearse en el sector informal por los ingresos o beneficios que este pueda obtener si sacrifica su afiliación al IMSS, quedando expuestos ante riesgos que difícilmente individualmente podrá afrontar. Y al Estado por que está dotado de mecanismo para garantizar, promover y extender la seguridad social a toda la población.

Otras causas que se suman a los factores de la carencia en la formalidad de empleo y en consecuencia de la cobertura de seguridad social; son las crisis económicas y escaso crecimiento del país que han impactado negativamente el empleo formal; es la falta de confianza de la población, en las instituciones públicas para garantizar que otras políticas públicas no afecten negativamente el fondo de pensiones o la solidez del sistema financiero; además en una población donde el 60 por ciento tiene un promedio del salario base de cotización menor o igual a tres

salarios mínimos, que son destinados a cubrir los gastos más inmediatos como son alimentación, hogar, vestido y educación, resulta muy difícil para las familias pensar en el ahorro en el largo plazo.

Gráfica 5. 14
Efectos marginales por actividad económica



Nota: 11 Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza. 21 Minería. 22 Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final. 23 Construcción. 3133 Industrias manufactureras. 43 Comercio al por mayor. 46 Comercio al por menor. 4849 Transportes, correos y almacenamiento . 51 Información en medios masivos. 52 Servicios financieros y de seguros. 53 Servicios inmobiliarios y de alquiler de bienes muebles e intangibles. 54 Servicios profesionales, científicos y técnicos. 56 Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación . 61 Servicios educativos. 62 Servicios de salud y de asistencia social. 71 Servicios de esparcimiento cultural y deportivo, y otros servicios recreativos . 72 Servicios de alojamiento temporal y de preparación de alimentos y bebidas. 81 Otros servicios excepto actividades gubernamentales. 93 Actividades legislativas, gubernamentales, de impartición de justicia y de organismos internacionales y extraterritoriales.

Conclusiones

La presente tesis con título “Efecto de la reforma de pensiones de IMSS de 1997 en el tamaño del sector formal”, representa un primer esfuerzo con este enfoque, que intenta probar una de las varias hipótesis planteadas por la reforma, ya que no hay documentación con un análisis teórico y técnico que estudie los efectos que ha traído consigo la reforma.

La tesis examinó la reforma al programa de pensiones del IMSS, centrando su análisis en verificar ¿Sí la reforma de pensiones causó un efecto en el tamaño del sector formal?, esta pregunta no resulta fácil de responder, ya que un ejercicio como éstos es del tipo de *evaluación de impacto*. En la actualidad muchos países realizan evaluaciones de impacto en sus programas sociales, destinado para ello gran capital financiero, humano y tecnológico.

La literatura señala que para realizar una *evaluación de impacto* rigurosa es necesario, idealmente, contar con un experimento, un grupo de tratamiento y uno de control, para así medir el efecto de alguna intervención (Aedo 2005, CONEVAL 2012). Desafortunadamente en este caso no se cuenta con datos experimentales para examinar el efecto de la reforma a las pensiones del IMSS en el tamaño de la cobertura, por lo que la metodología consistió en la construcción de un modelo de regresión logit para analizar las variables que afectan el tamaño del sector formal antes y después de la reforma de 1997. Para medir si había incrementado la formalización del mercado laboral -una vez hecha la reforma de pensiones de vejez del IMSS de 1997-, fue tomar dos periodos, uno previo a la reforma (1996) y el otro posterior (2006). Los insumos utilizados para probar la hipótesis fueron las encuestas de empleo para ambos periodos; la ENE de 1996 y la ENOE del segundo trimestre para 2006, éstas son encuestas transversales, es decir, son fotografías del momento que no captan la información histórica, por lo que se limita a analizar el fenómeno en dos puntos precisos.

Tanto para la ENE96 como para la ENOE06, la construcción de la variable dependiente y las variables independientes se apegaron al marco teórico del Mercado Laboral⁴². La variable dependiente dicotómica captó la información del sector informal y formal, en tanto las variables independientes fueron los factores explicativos que determinaron la decisión del individuo para emplearse en el sector informal o formal.

Por la construcción del fenómeno la metodología a utilizar para medir, estimar y comparar fue el modelo logit. Del modelo se estimaron los parámetros, se realizaron las pruebas de hipótesis y de bondad de ajuste; se compararon varios modelos para un mismo periodo y para interpretar el modelo se utilizó tanto el cociente de momios como los efectos marginales.

⁴² Capítulo II.

Para determinar si existió un *efecto de la reforma de pensiones de vejez del IMSS de 1997 en el tamaño del sector formal* se compara el cociente de momios y los efectos marginales de 1996 respecto del 2006. De un periodo a otro se pudo observar que se ha dado en su mayoría sobre las condiciones o características del trabajador un decrecimiento o estancamiento en la probabilidad de ser formal, sólo en muy pocas variables se ha experimentado un crecimiento. A continuación se describe, dónde se dieron dichos casos.

- La cobertura de la seguridad de 1996 a 2006 social en cuanto a probabilidades se ha visto diferenciada por regiones, siendo las regiones que se componen de los estados del sur y centro del país las más afectada, que en su mayoría estos estados se componen de grupos desprotegidos. El cociente de probabilidad o el efecto marginal en probabilidad de ser formal aumenta (región 5 y 6) o disminuye (región 2, 3, 4 y 7) dependiendo en la región donde resida el trabajador. Pero se puede decir que en la mayoría de los estados del norte es más probable que un trabajador sea formal que los que radican en la parte sur y centro del país. Es preciso señalar que el Distrito Federal que es la entidad federativa donde se encuentra la mayor parte de la población ocupada después del Estado de México, hay un retroceso en la formalización del empleo, dado que es más probable que un individuo de los estados del sur sea formal.
- El IMSS (Aportaciones para el debate, 1997) y Grandolini y Cerda (1997), señalaban que la reforma al esquema de pensiones daba incentivos para que trabajadores de edad avanzada (después de los 50 años) y mujeres tuvieran un empleo formal. Sin embargo esto ha significado una mínima influencia en la formalidad, ya que la condición de ser hombre representaba una ligera probabilidad mayor sobre el ser mujer para ser formal, es decir después de la reforma casi son igualmente probable que un hombre o una mujer se emplearan en el sector formal. Y la edad para ambos géneros siguió influyendo negativamente en la formalidad del trabajador, es decir a mayor edad la probabilidad de ser formal va disminuyendo.
- No hay suficiente evidencia para mencionar que la composición familiar o estado conyugal del trabajador sea un eje fundamental para que el trabajador sea formal, sin embargo los trabajadores cuyo estatus conyugal es “casado”, para ambos periodos representan la mayor probabilidad de ser formal. Aunque Hernández (2001) indica que en un grupo familiar donde dos o más tienen un empleo formal que les de acceso al seguro social, sólo tendrá incentivos en lo familiar e individual aquel que es el jefe de familia, en tanto para el otro u otros este incentivo familiar no existirá, por lo que podría haber una inclinación a emplearse en el sector informal.

- Otro punto a mencionar que también afectan la cobertura, es cuando el trabajador afiliado no percibe la existencia de una adecuada relación entre la contribución y el beneficio ofrecido por parte de Instituto, ello genera un hueco para el incremento de la cobertura de seguridad social. Esto se puede evidencia con trabajadores que cuentan con niveles de instrucción altos. Porque se observa que en términos generales el nivel de instrucción contribuye a que trabajadores se empleen en el sector formal, pero esto es parcialmente cierto dado que después de la reforma los individuos que contaban con una instrucción más alta al avanzar su edad, baja su probabilidad de ser formales, una de las posibilidades es que una vez que ganaron capital humano y financiero decidan dejar el sector formal para ser trabajadores independientes o ser empleadores, otra causa puede ser que el mercado laboral sea incapaz de ofrecerle un empleo formal.
- Pasando a la ocupación que desempeñan los trabajadores, lo que ha sucedido posterior a la reforma es que si bien siguen siendo los *trabajadores subordinados y remunerados* aquellos que tienen altas probabilidades de ser formales, ésta ha descendido (este mismo hecho se presenta en el nivel de instrucción). El motivo de ello puede tener dos posibles canales, por un lado sea la propia decisión del trabajador evaluando todos los beneficios que pueda obtener de un empleo formal sobre el informal, y por otro sobre el tipo de trabajos que las empresas están generando, de ello se hablara más adelante.
- Resulta interesante ver el comportamiento que están teniendo los empleadores ante el empleo formal. Porque el empleador mediante su afiliación al IMSS posiblemente avalaría incentivos que se mencionaban que la reforma traería, no obstante el ser empleador incide negativamente en la probabilidad de la formalidad. Pero tal vez estos últimos difícilmente caigan en estado de pobreza en el futuro, ya que podrían tener los recursos suficientes para asegurarse de mantener sus flujos de consumo en la etapa de la vejez, además de que ante cualquier contingencia podrían contar con otros financiamientos o planes privados.
- Las estimaciones de probabilidad sobre el número de trabajadores que constituyen la empresa y su jornada laboral, son un apoyo que muestra síntomas, de que las empresas perciben la contribución a la seguridad social como un impuesto al empleo y no como un ahorro en beneficio del trabajador, dado que una parte de éstas tienen subcontratados a sus trabajadores, o subdeclara el salarios. Bajo esto último, el trabajador está de alguna forma obligado a aceptar el trabajo, si esas son las reglas de contratación. Por tanto el riesgo moral de parte del patrón juega un papel principal en el fondeo de la pensiones. En este caso, los trabajadores que no acumulan los recursos necesarios para alcanzar la pensión mínima, por la subdeclaración de salarios que realiza su patrón, el Gobierno Federal tendría que aportar la diferencia resultante para que alcance de la pensión mínima. Bajo este tipo de prácticas, los recursos que el Estado aporta para cubrir la diferencia, serían afectados negativamente.

- Para estimar las probabilidades sobre los sectores de actividades, se tomó como grupo de referencia a la actividad agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza. Posterior a la reforma, en 2006 emplearse en cualquier otra actividad es más probable de que esta sea formal. Y resalta en particular la actividad de servicios financieros y de seguros, por ser la que más creció en cuanto la probabilidad de tener un empleo formal que brinde protección social, después le siguen los que se dedican al comercio al por mayor.

Se ha analizado cómo las características individuales del trabajador determinan su posición frente al empleo formal, aunque hay características que ejercen una mayor probabilidad que otras, en su conjunto muestran que posterior a reforma se ha dado un estancamiento y un retroceso en la formalización del empleo. Apoyándose en la metodología que se utilizó para estimar cada sector y vislumbrar el tamaño del sector formal antes y después de la reforma, se tiene que del conjunto de la población que se encontraba ocupada, en 1996 los trabajadores formales eran 7.6 millones y los informales eran 17.6 millones, a 2006 pasaron a ser 10.6 millones de formales y a 23.4 millones de informales. Por lo tanto en un periodo de 10 años hay un crecimiento de 3 millones de trabajadores que se integraron al sector formal y de 6 millones al informal. En términos porcentuales, la distribución del tamaño del sector formal pasó de 30.1 por ciento en 1996 a 31.0 en 2006, lo cual indica un estancamiento en la formalización del empleo. Entonces se puede decir que la distribución de la población ocupada que se compone del sector formal e informal después de 10 años se ha mantenido casi igual, por lo que la reforma no ha logrado uno de los cometidos que es formalizar el empleo.

Estos resultados no son muy alentadores, porque mientras siga bajando la probabilidad en la formalidad del empleo, esto querrá decir que la cobertura o se estancara o seguirá bajando. Entonces posiblemente los incentivos que se mencionaron en la reforma, no son los que se necesitan para que se incremente la cobertura. Entre otras cosas existirá la preocupación de que los trabajadores durante su periodo de vida laboral, no realizaran la acumulación de recursos en su cuenta individual que le pudieran mantener sus flujos de consumo durante la vejez. Además que dejara de percibir beneficios como el crédito a la vivienda y a servicios médicos para el trabajador y su familia, así como de otras prestaciones médicas.

El estudio se podría hacer más robusto si se contara con bases de datos longitudinales, es decir, que captaran la información histórica. Pero a la fecha no se cuenta con dicha información, para que se estudie el incremento del sector formal y los movimientos de la cobertura de seguridad social de manera histórica. Con las encuestas oficiales que se encuentra hoy en día sólo se puede hacer un estudio de manera puntual.

Evidentemente las variables explicativas que modelan el fenómeno no son únicas, ya que de lector pudiera considerar que faltan o sobran algunas. En la presente tesis no se considera

como variable explicativa la interacción de éstas, es decir, si el nivel de instrucción multiplicada por la edad tiene algún efecto en modelo, pero para futuras evaluaciones esto podría incluirse.

Por último es de esperarse que la formalización del empleo y el incremento en la cobertura de seguridad social –IMSS-, no sólo se va a dar con la aplicación de una reforma al esquema de pensiones, si no que se necesita ser complementada con la aplicación de otras políticas públicas, como pueden ser laborales o fiscales, como lo señala el documento *“La interacción entre los sistemas de jubilación, la seguridad social, las políticas del mercado laboral y la fiscalidad en la configuración de los patrones de jubilación”* de la Asociación Internacional de la Seguridad Social (AISS) de 2007, además de que debe haber crecimiento económico en el país para que se generen empleos formales. Aunque como lo señala Hernández (2001) que es necesario poner especial atención en el diseño de políticas a favor de las mujeres trabajadoras, trabajadores de zonas rurales o con baja educación, que son grupos con menos acceso a empleos formales y estables, en la cual se concentra la mayor población del déficit de la cobertura.

Es tema debe de continuar y tener posteriores evaluaciones, porque resulta relevante tanto para Gobierno Federal como para la sociedad en general. Para el Gobierno Federal porque la cobertura del IMSS refleja la situación del mercado laboral del país; permite estimar o prever los gastos que se destinan a programas de asistencia social, en el rubro de pensiones no contributivas y salud; así mismo podría crear medidas de contención de políticas públicas, que sean completarias para la formalización del empleo. Y para la sociedad en general porque es un eje fundamental para alcanzar el estado de bienestar además de que por ley es un derecho contar con seguridad social.

Referencias Bibliográficas

Aedo Cristián. “Evaluación de Impacto”. *Manuales*, no 47 (noviembre 2005): 1-75.

Alcaraz, C., y García Verdú, R. “Cambios en la Composición del Empleo y Evolución de la Productividad del trabajo en el Sector Formal de la Economía Mexicana: 2000-2005”. Documento de investigación No 2006-3. México: Banco de México, 2006.

Asociación Internacional de la Seguridad Social (AISS). “La interacción entre los sistemas de jubilación, la seguridad social, las políticas del mercado laboral y la fiscalidad en la configuración de los patrones de jubilación”. Informe Técnico 11. Moscú: AISS, 2007.

Centro de Investigación y Desarrollo (CIDE). Los modelos Logit y Probit en la investigación social. El caso de la pobreza del Perú en el año 2001. Lima Agosto 2002 .Disponible en:
<http://www.inei.gob.pe/biblioineipub/bancopub/Est/Lib0515/Libro.pdf>

Conferencia Interamericana de Seguridad Social (CISS). “El Mercado Laboral: Efecto de la Reforma de la Seguridad Social”. Capítulo IV en *Informe sobre la Seguridad Social en América 2003*, CISS. México, 2003.

Conferencia Interamericana de Seguridad Social (CISS). “La Informalidad en el Mercado Laboral y la Seguridad Social”. Capítulo III en *Informe sobre la Seguridad Social en América 2005*, CISS. México, 2005.

Grandolini, Gloria y Luis Cerda. “The 1997 pension reform in Mexico: Genesis and design feature”. México: Banco Mundial y Secretaria de Hacienda y Crédito Público, 1998.

Gujarati, Damodar. *Econometría, segunda edición*. México: McGraw-Hill, 1995.

Hernández Licona, Gonzalo. “Políticas para promover una ampliación de la cobertura de los sistemas de pensiones: el caso de México”. *Financiamiento del Desarrollo*, no 107 (enero 2001): 1-56.

Hosmer, David W., y Stanley Inc. Lemeshow. *Applied Logistic Regression, Second Edition*. Estados Unidos: John Wiley & Sons, 2000.

Hazas , Alejandro. *El cálculo de las tasas de remplazo bajo el modelo de aportación definida*. En *Diez años de reforma a la seguridad social en México: Balance, perspectivas y propuestas.*, 61-72. México, 2007.

Instituto Nacional de Estadística y Geografía (INEGI). *Clasificación de Actividades Económicas de la Encuesta Nacional de Empleo (CAE_ENE)*. México, 1995. Disponible en:
<http://www3.inegi.org.mx/sistemas/biblioteca/detalle.aspx?c=17609&upc=702825001941&s=est&tg=0&f=2&pf=ench>

_____. *Conociendo la base de datos de la ENOE*. México: INEGI, 2007. Disponible en:
http://www.inegi.org.mx/est/contenidos/espanol/metodologias/ptos_enoe/descripcion_bd/conociendoenoe.pdf

_____. *Conociendo la base de datos de la ENE con criterio ENOE*. México: INEGI, 2009. Disponible en:
<http://www3.inegi.org.mx/sistemas/biblioteca/detalle.aspx?c=17612&upc=702825001944&s=est&tg=0&f=2&pf=ench>

_____. *Encuesta Nacional de Ocupación y Empleo 2005: Una nueva encuesta para México*. México: INEGI, 2005. Disponible en: <http://www.inegi.org.mx/est/contenidos/proyectos/encuestas/hogares/regulares/enoe/default.aspx>

_____. ENE: Aspectos metodológicos (cuestionarios). 1 de Junio de 2012. Disponible en <http://www.inegi.org.mx/est/contenidos/proyectos/encuestas/hogares/historicas/ene/default.aspx>

_____. ENOE: Aspectos metodológicos (cuestionario ampliado). 1 de Junio de 2012. Disponible en <http://www.inegi.org.mx/est/contenidos/proyectos/encuestas/hogares/regulares/enoe/default.aspx>

_____. ENOE: Descripción de Archivos (cuestionario básico). 1 de Junio de 2012. Disponible en <http://www.inegi.org.mx/est/contenidos/proyectos/encuestas/hogares/regulares/enoe/default.aspx>

_____. Empleo. México: INEGI, 2009. Disponible en: http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/integracion/pais/historicas10/Tema5_Empleo.pdf

_____. *Guía de conceptos, usos e interpretación de la Estadística sobre la Fuerza Laboral en México*. México: INEGI, 2002. Disponible en: <http://www.inegi.org.mx/est/contenidos/espanol/metodologias/otras/guia.pdf>

_____. *Población*. México: INEGI, 2009. Disponible en: http://www.inegi.gob.mx/prod_serv/contenidos/espanol/bvinegi/productos/integracion/pais/historicas10/Tema1_Poblacion.pdf

_____. *Salud*. México: INEGI, 2009. Disponible en: http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/integracion/pais/historicas10/Tema4_Salud.pdf

_____. *Sistema de Clasificación Industrial de América del Norte, México (SCIAN)*. México, 2007. Disponible en: <http://www.inegi.org.mx/sistemas/scian/>

_____. *Regiones socioeconómicas de México*. Disponible en: <http://www.inegi.org.mx/est/contenidos/espanol/sistemas/regsoc/default.asp?s=est&c=11723> (último acceso: 1 de Mayo de 2013).

Instituto Mexicano del Seguro Social (IMSS). *Aportaciones al debate*. México: IMSS, 1997.

_____. *Evidencias para el debate*. México, 1997.

_____. *Perspectivas para el debate*, 1999.

_____. Memoria Estadística 2012. Capítulo I Evolución del Empleo. <http://www.imss.gob.mx/estadisticas/financieras/Pages/memoriaestadistica.aspx> (último acceso: 31 de Noviembre del 2012).

_____. Memoria Estadística 2012. Capítulo II Población Derechohabiente. <http://www.imss.gob.mx/estadisticas/financieras/Pages/memoriaestadistica.aspx> (último acceso: 31 de Noviembre del 2012).

Kleinbaum, David G., y Mitchel Klein. *Logistic Regression*. New York: Springer, 2010.

Mesa-Lago, Carmelo. "Las Reformas de pensiones en América Latina y su impacto en los principios de seguridad social". *Financiamiento del Desarrollo*, no 144 (marzo 2004):132 páginas.

Montgomery, Douglas C., Elizabeth A. Peck, y G. Geoffrey Vining. *Introducción al análisis de regresión.* México: Compañía Editorial Continental, 2002.

Organización Internacional de Trabajo (OIT) , Organización Mundial del Comercio (OMC). *La globalización y el empleo informal en los países en desarrollo.* Estudio conjunto de la Oficina Internacional del Trabajo y la Secretaría de la Organización Mundial de Comercio, Suiza: Secretaría de la OMC, 2009.

Rabasa, Emilio. "El IMSS: Salud, seguridad y solidaridad social". Revista de administración pública. 157-170

Ruezga, Antonio. "Los Seguros Sociales en América Latina". Capítulo IV en *Seguridad Social: Una visión latinoamericana*, editado por el Centro Interamericano de Estudios de Seguridad Social (CIESS). México: CIESS, 2009.

Silva Aycaguer, Luis Carlos, y Isabel María Barroso Utra. *Cuadernos de Estadística: Regresión Logística.* Madrid: Muralla, 2004.

Villagómez, F. Alejandro y Juan Ignacio Hernández. "Impacto de la Reforma de Sistema de pensiones en México sobre el Ahorro". *Documento de Trabajo del CIDE*, número 453 (marzo 2009): 1-31.

Wooldridge, Jeffrey, M. *Econometric Analysis of Cross Section and Panel Data.* Michigan: MIT Press, 2002.

Leyes

Ley del Seguro Social 1973.

Ley del Seguro Social 1997.

Anexo 1

Resultados del Ajuste del modelo logit sobre los datos de la ENE96 sin la observación 12,077 y 132, 127.

sector	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Iregion_e_2	.0736634	.0382193	1.93	0.054	-.0012451	.1485719
_Iregion_e_3	.3985594	.0384691	10.36	0.000	.3231613	.4739576
_Iregion_e_4	.455844	.0369997	12.32	0.000	.3833259	.5283622
_Iregion_e_5	.6564739	.038213	17.18	0.000	.5815777	.7313701
_Iregion_e_6	.7589786	.0390819	19.42	0.000	.6823796	.8355777
_Iregion_e_7	.2254577	.0602397	3.74	0.000	.1073901	.3435253
genero	.1410828	.0193482	7.29	0.000	.1031611	.1790045
edad	.0162378	.0009527	17.04	0.000	.0143706	.018105
_Inivel_ins_2	.2907068	.027917	10.41	0.000	.2359905	.3454231
_Inivel_ins_3	.7092507	.0285949	24.80	0.000	.6532057	.7652957
_Inivel_ins_4	.7176137	.0296559	24.20	0.000	.6594893	.7757381
_Itipo_ocup_1	.2256509	.1544678	1.46	0.144	-.0771005	.5284023
_Itipo_ocup_4	3.798261	.1315114	28.88	0.000	3.540503	4.056018
_Itipo_ocup_5	3.848968	.2508744	15.34	0.000	3.357263	4.340672
_Ic_trabaja_2	1.639713	.1060952	15.46	0.000	1.43177	1.847655
_Ic_trabaja_3	2.956782	.108339	27.29	0.000	2.744442	3.169123
_Ic_trabaja_4	3.45552	.1104072	31.30	0.000	3.239126	3.671914
_Ic_trabaja_5	4.165613	.108165	38.51	0.000	3.953614	4.377613
_Ic_trabaja_6	4.696417	.107283	43.78	0.000	4.486146	4.906688
_Iscian_ene_21	-.1698826	.0886315	-1.92	0.055	-.3435973	.003832
_Iscian_ene_22	1.551073	.104927	14.78	0.000	1.345419	1.756726
_Iscian_ene_23	.630393	.0494208	12.76	0.000	.5335299	.727256
_Iscian_ene_43	.9271655	.0504959	18.36	0.000	.8281953	1.026136
_Iscian_ene_46	1.556174	.0446683	34.84	0.000	1.468626	1.643722
_Iscian_ene_51	1.474807	.0994653	14.83	0.000	1.279858	1.669755
_Iscian_ene_52	-.572532	.0653374	-8.76	0.000	-.7005909	-.4444732
_Iscian_ene_53	1.426876	.1032904	13.81	0.000	1.22443	1.629321
_Iscian_ene_54	1.488943	.0609811	24.42	0.000	1.369422	1.608464
_Iscian_ene_56	1.760497	.1928034	9.13	0.000	1.382609	2.138384
_Iscian_ene_61	.0603311	.0554097	1.09	0.276	-.0482698	.168932
_Iscian_ene_62	.7016313	.0568442	12.34	0.000	.5902187	.8130439
_Iscian_ene_71	.4137425	.0870285	4.75	0.000	.2431698	.5843151
_Iscian_ene_72	1.346367	.0492701	27.33	0.000	1.249799	1.442934
_Iscian_ene_81	.7368947	.0476493	15.46	0.000	.6435039	.8302856
_Iscian_ene_93	-.7403689	.0479662	-15.44	0.000	-.8343809	-.6463569
_Iscian_ene_3133	1.370528	.0411785	33.28	0.000	1.289819	1.451236
_Iscian_ene_4849	.9946655	.0508595	19.56	0.000	.8949827	1.094348
_Ihrs_traba_2	1.214201	.0249766	48.61	0.000	1.165248	1.263154
_Ihrs_traba_3	.8758666	.0279435	31.34	0.000	.8210984	.9306347
_Iedo_conyu_1	.1223027	.0355893	3.44	0.001	.052549	.1920564
_Iedo_conyu_2	.0248861	.0519173	0.48	0.632	-.0768699	.1266422
_Iedo_conyu_3	.3081043	.0753313	4.09	0.000	.1604577	.4557508
_Iedo_conyu_4	.0426492	.0757172	0.56	0.573	-.1057538	.1910521
_Iedo_conyu_5	.3589149	.0208323	17.23	0.000	.3180845	.3997454
_cons	-10.80667	.1621541	-66.64	0.000	-11.12449	-10.48886

Resultados del ajuste del modelo logit sobre los datos de la ENOE06 sin la observación 77,895 y 130,947.

sector	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Iregion_e_2	.1603399	.0396348	4.05	0.000	.0826572	.2380226
_Iregion_e_3	.2696498	.0406851	6.63	0.000	.1899084	.3493911
_Iregion_e_4	.4260936	.0377448	11.29	0.000	.3521153	.500072
_Iregion_e_5	1.063675	.0415081	25.63	0.000	.98232	1.145029
_Iregion_e_6	.9317248	.0414888	22.46	0.000	.8504083	1.013041
_Iregion_e_7	-.2399871	.0634328	-3.78	0.000	-.3643132	-.1156611
genero	.0520226	.0212867	2.44	0.015	.0103014	.0937438
edad	.0256842	.0010348	24.82	0.000	.023656	.0277124
_Inivel_ins_2	.3710648	.0355859	10.43	0.000	.3013177	.440812
_Inivel_ins_3	.8988171	.0345271	26.03	0.000	.8311451	.966489
_Inivel_ins_4	1.095765	.0365142	30.01	0.000	1.024198	1.167332
_Itipo_ocup_1	2.998639	.085315	35.15	0.000	2.831425	3.165853
_Itipo_ocup_2	-.1352807	.1060193	-1.28	0.202	-.3430748	.0725133
_Itipo_ocup_4	-1.086162	.179003	-6.07	0.000	-1.437002	-.7353228
_Ic_trabaja_2	1.078712	.0868517	12.42	0.000	.9084855	1.248938
_Ic_trabaja_3	2.420076	.0897866	26.95	0.000	2.244097	2.596054
_Ic_trabaja_4	3.069091	.0930175	32.99	0.000	2.88678	3.251402
_Ic_trabaja_5	3.612622	.0895909	40.32	0.000	3.437027	3.788217
_Ic_trabaja_6	4.568085	.0902626	50.61	0.000	4.391173	4.744996
_Ihrs_traba_2	1.200987	.0271214	44.28	0.000	1.14783	1.254144
_Ihrs_traba_3	.8596514	.029871	28.78	0.000	.8011054	.9181975
_Iscian_21	1.997354	.1597483	12.50	0.000	1.684253	2.310455
_Iscian_22	2.444196	.1791514	13.64	0.000	2.093066	2.795326
_Iscian_23	1.140869	.05185	22.00	0.000	1.039245	1.242493
_Iscian_43	2.396837	.0634163	37.80	0.000	2.272543	2.521131
_Iscian_46	2.157714	.0504605	42.76	0.000	2.058814	2.256615
_Iscian_51	1.942849	.095918	20.26	0.000	1.754853	2.130844
_Iscian_52	2.340722	.0965359	24.25	0.000	2.151515	2.529929
_Iscian_53	1.771735	.1067077	16.60	0.000	1.562592	1.980878
_Iscian_54	1.919325	.0715709	26.82	0.000	1.779049	2.059602
_Iscian_55	3.866768	1.111389	3.48	0.001	1.688486	6.045051
_Iscian_56	2.203347	.073799	29.86	0.000	2.058703	2.34799
_Iscian_61	1.761401	.0668754	26.34	0.000	1.630328	1.892474
_Iscian_62	1.81823	.0702661	25.88	0.000	1.680511	1.955949
_Iscian_71	.9110638	.0973403	9.36	0.000	.7202804	1.101847
_Iscian_72	1.763682	.0544688	32.38	0.000	1.656925	1.870439
_Iscian_81	1.160607	.0562663	20.63	0.000	1.050327	1.270887
_Iscian_93	.6321817	.0597662	10.58	0.000	.5150421	.7493213
_Iscian_3133	1.899183	.0485575	39.11	0.000	1.804012	1.994354
_Iscian_4849	1.427249	.0588529	24.25	0.000	1.3119	1.542599
_Iedo_conyu_1	.1383966	.0329038	4.21	0.000	.0739062	.2028869
_Iedo_conyu_2	.1160889	.0517918	2.24	0.025	.0145788	.2175989
_Iedo_conyu_3	.2518684	.0720815	3.49	0.000	.1105912	.3931455
_Iedo_conyu_4	.0850578	.0861704	0.99	0.324	-.0838331	.2539488
_Iedo_conyu_5	.4087899	.0241595	16.92	0.000	.3614382	.4561417
_cons	-10.48629	.1163594	-90.12	0.000	-10.71435	-10.25823

Anexo 2

Código empleado en STATA

```
/////Modelo Logit
```

```
***Ajuste del Modelo y prueba de significancia
```

```
xi:logit sector i.region_e genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2 i.hrs_trabajo  
i.scian_ene i.edo_conyugal
```

```
predict YL1
```

```
predict Res1, rstandard
```

```
label var Res1 "Residuos estandarizados de Pearson"
```

```
scatter Res1 edad
```

```
scatter Res1 YL1, title((a)residuales contra valores ajustado)
```

```
gen n=_n
```

```
sort n
```

```
scatter Res1 n if Res1<., mlabel(n) title((b)residuales ordenados)
```

```
*Etiqueta para identificar valores extremos de los residuos
```

```
recode Res1 (. -7/7=0) (else=1), into(extremo)
```

```
xi:logit sector i.region_e genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2 i.hrs_trabajo  
i.scian_ene i.edo_conyugal
```

```
*Datos influyentes
```

```
predict cook, dbeta
```

```
scatter cook n if cook<., mlabel(n) yline(.1)
```

```
*Prueba de razon de verosimilitud
```

```
logit sector
```

```
xi:logit sector i.region_e genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2 i.scian_ene  
i.hrs_trabajo i.edo_conyugal
```

```
quietly logit sector
```

```
est store M00
```

```
xi: quietly logit sector i.region genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2  
i.scian_ene i.hrs_trabajo i.edo_conyugal
```

```
est store M01
```

```
lrtest M00 M01
```

```
***Comparación de modelos
```

```
xi: quietly logit sector i.region_e genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2  
i.scian_ene i.hrs_trabajo i.edo_conyugal
```

```

estimates store Modelo01
xi: quietly logit sector genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2 i.scian_ene
i.hrs_trabajo i.edo_conyugal
estimates store Modelo02
xi: quietly logit sector i.region_e genero edad i.nivel_instruccion i.c_trabajadores2 i.scian_ene i.hrs_trabajo
i.edo_conyugal
estimates store Modelo03
xi: quietly logit sector i.region_e genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2
i.hrs_trabajo i.edo_conyugal
estimates store Modelo04
xi: quietly logit sector i.region_e genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2
i.scian_ene i.hrs_trabajo
estimates store Modelo05
estimates table Modelo01 Modelo02 Modelo03 Modelo04 Modelo05, star stats(N ll_0 ll_chi2 r2_p aic bic
rank)

```

```

lrtest Modelo01 Modelo02
lrtest Modelo01 Modelo03
lrtest Modelo01 Modelo04
lrtest Modelo01 Modelo05

```

***Bondad de Ajuste

*Prueba de Hosmer and Lemeshow
estat gof, group (10) table

*Curva ROC
lstat, cutoff(.1)
lstat, cutoff(.2)
lstat, cutoff(.3)
lstat, cutoff(.4)
lstat, cutoff(.5)
lstat, cutoff(.6)
lstat, cutoff(.7)
lstat, cutoff(.8)
lstat, cutoff(.9)
lstat, cutoff(1)

lroc, xtitle (1-Especificidad (Sp)) ytitle(Sensibilidad (Se))

***Interpretación del modelo

*Cociente de momios
xi:logit sector i.region genero edad i.nivel_instruccion i.tipo_ocupacion i.c_trabajadores2 i.scian_ene
i.hrs_trabajo i.edo_conyugal, or

*Efecto marginal promedio

margins, dydx(*)

margins, dydx(_lregion_e_2 _lregion_e_3 _lregion_e_4 _lregion_e_5 _lregion_e_6 _lregion_e_7)

margins, dydx(_lregion_e_2 _lregion_e_3 _lregion_e_4 _lregion_e_5 _lregion_e_6 _lregion_e_7)
at(genero=(0 1))

margins, dydx(_lregion_e_2 _lregion_e_3 _lregion_e_4 _lregion_e_5 _lregion_e_6 _lregion_e_7)
at(edad=(15(5)60))

margins, dydx(genero) at(edad==(15(5)60))

margins, dydx(_lnivel_ins_2 _lnivel_ins_3 _lnivel_ins_4) at(genero=(0 1))

margins, dydx(_lnivel_ins_2 _lnivel_ins_3 _lnivel_ins_4)at(edad=(15(5)60))

margins, dydx(_ltipo_ocup_1 _ltipo_ocup_4 _ltipo_ocup_5)

margins, dydx(_lc_trabaja_2 _lc_trabaja_3 _lc_trabaja_4 _lc_trabaja_5 _lc_trabaja_6) at(_lhrs_traba_2==1)

margins, dydx(_lc_trabaja_2 _lc_trabaja_3 _lc_trabaja_4 _lc_trabaja_5 _lc_trabaja_6) at(_lhrs_traba_3==1)

margin, dydx(_lscian_ene_21 _lscian_ene_22 _lscian_ene_23 _lscian_ene_43 _lscian_ene_46
_lscian_ene_51 _lscian_ene_52 _lscian_ene_53 _lscian_ene_54 _lscian_ene_56 _lscian_ene_61
_lscian_ene_62 _lscian_ene_71 _lscian_ene_72 _lscian_ene_81 _lscian_ene_93 _lscian_ene_3133
_lscian_ene_4849)

margins, dydx(_ledo_conyu_1 _ledo_conyu_2 _ledo_conyu_3 _ledo_conyu_4 _ledo_conyu_5)