Documento No. 61

Análisis de Cruces Fronterizos con
Modelos Lineales Generalizados

por

Lorenzo Moreno Navarro^{*}

ABRIL, 1987

^{*} Las ideas contenidas en el presente documento son responsabilidad exclusiva del autor y no reflejan la posición del Banco de México.

ANÁLISIS DE CRUCES FRONTERIZOS CON MODELOS LINEALES GENERALIZADOS

por

Lorenzo Moreno Navarro*

Sumario: Utilizando dos tipos de modelos lineales generalizados, se analizan diversos aspectos del fenómeno de “cruces fronterizos” obtenidos del proceso de generación del marco de referencia de la Encuesta de Transacciones Fronterizas del Banco de México. Se discute la influencia que varios factores explicativos, como el lugar de residencia habitual o el día de la semana, tienen sobre el comportamiento a través del tiempo de la estructura de fenómeno. Se ensayan modelos con variables ordinales, que son más parsimoniosos que los submodelos saturados de un orden específico

* Subgerencia de Métodos Estadísticos. Dirección de Investigación Económica. Banco de México.

El autor agradece los comentarios y sugerencias del Dr. Jesús Romo y G. para la realización de este documento. Gran apoyo se recibió de la Mat. Lilia Luna López en el manejo de la información y análisis de la misma.

I. INTRODUCCIÓN

Los modelos lineales generalizados (MLG) se han convertido en una importante herramienta para resolver el problema de caracterizar las estructuras subyacentes de las medias poblacionales de una variable tabulada según varios factores clasificatorios, datos cuya variabilidad muestral respecto a la poblacional puede ser descrita con una función de distribución proveniente de la familia exponencial. Entre las preguntas que pueden responderse con este tipo de modelos está la referente a la igualdad entre estructuras subyacentes a clasificaciones de una variable responsiva para diferentes mediciones transversales independientes en el tiempo. Este es el caso de la Encuesta de Transacciones Fronterizas, de la Dirección de Investigación Económica del Banco de México, de la que se analizan en este documento las tabulaciones respecto a “cruces fronterizos”.

La estructura del documento es como sigue. Primero, se señalan los objetivos del análisis para después escribir los datos y variables que se considerarán. En seguida viene una explicación general sobre los modelos lineales generalizados propuestos para caracterizar a los “cruces fronterizos”, para después discutir los resultados de su aplicación y hacer las conclusiones pertinentes.

II. OBJETIVOS

La Encuesta de Transacciones Fronterizas (Banco de México, 1982) fue diseñada para: i) estimar los ingresos y egresos de divisas que generan las transacciones fronterizas, para ser incluidos en las estadísticas de la Balanza de Pagos; ii) obtener elementos suficientes para analizar el comportamiento de este rubro.

Recientemente, como consecuencia de una revisión de la metodología para la medición de los fenómenos señalados, se consideraron algunas modificaciones al diseño muestral de la encuesta. (ver: Banco de México, 1987). Para hacer estos cambios, principalmente en la generación del marco de referencia de “cruces fronterizos”^{*} según residencia, se necesitaba conocer si la distribución de cruces fronterizos (peatones y automovilistas) y la proporción de cruces de mexicanos residentes ambas

^{*} Se considera “cruce fronterizo” a cualquier visita temporal a la faja fronteriza de un país distinto al de residencia habitual, por cualquiera de los siguientes motivos: por negocios; por asistencia a convenciones; por recreo, compras, razones de familia o motivos de salud. Quedan excluidos del concepto de “cruce

variables según horario, día de la semana y fecha del conteo exhaustivo o “censo”, diferían entre sí a través del tiempo, i.e. variación “intercensal”. De igual interés era conocer la influencia que el día de la semana y el horario tenían sobre la distribución de cruces y la proporción que de estos se declararon mexicanos residentes.

El ejercicio se requería en forma global y desagregada para cada una de las plazas donde se levanta continuamente la referida encuesta.

En la siguiente sección se describen los datos analizados y se consideran las variables explicativas a utilizarse en los MLG.

III. DATOS Y VARIABLES

Para estudiar las características de una población tan extensa, tanto en el sentido numérico (se estima que el número anual de cruces fronterizos es de aproximadamente 150 millones), como en el sentido de extensión territorial (se tiene una línea divisoria con los Estados Unidos de aproximadamente 3 mil kms.), se diseñó una encuesta por muestreo probabilístico.

Este estudio contempla dos etapas:

1. La generación del marco de referencia, mismo que consiste en un conteo del número total de cruces fronterizos durante una semana determinada, en las plazas seleccionadas, con el objeto de obtener la estructura diaria de los cruces de personas, según el medio de transporte que utilicen.
2. Levantamiento de una encuesta mensual y continua que, con base en un diseño probabilística, muestre a los grupos de personas sujetas de la misma, clasificados en “peatones” y “automovilistas”; “residentes” y “noresidentes”, con el objeto de obtener la estimación del gasto medio de cada categoría.

fronterizo”: las visitas que rebasen los límites de la faja fronteriza señalados y los que impliquen una permanencia mayor de tres días en la población fronteriza de un país distinto al de residencia habitual.

Esta encuesta está a cargo de la Oficina de Servicios No Factoriales, Subgerencia del Sector Externo, Dirección de Investigación Económica del Banco de México.

La información que aquí se analiza se refiere a la primera etapa del estudio, i.e. la generación del marco de referencia, consistente en los conteos o “censos” de cruces fronterizos levantados en Noviembre de 1983, Julio de 1984, Marzo de 1985 y Enero de 1986.

En cada uno de estos casos se obtuvo el conteo de personas según el medio de transporte que utilizaran, el día de la semana y horario de cruce y el lugar de residencia habitual para cada una de las plazas consideradas (ver Cuadro A.1 para la definición de los niveles de los factores intervinientes en el análisis).

Claramente la distribución de cruces según horario y día de la semana, así como la proporción de cruces de residentes mexicanos es fácilmente derivable de las anteriores variables.

En la siguiente sección se describen los modelos lineales generalizados considerados en este ejercicio.

IV. MÉTODO Y MODELOS

Como se mencionó en la Secc. II), la revisión metodológica descrita requiere probar la hipótesis respecto a la igualdad entre los “censos” o conteos de cruces fronterizos de

- i) la distribución de cruces según horario y día de la semana para cada tipo de viajero,
- ii) La proporción de cruces de residentes mexicanos, también según la clasificación de i).

Desde el punto de vista estadísticos, la prueba de la hipótesis mencionada sugiere el uso de modelos lineales generalizados, dado que ambos conjuntos de datos se refieren a variables medidas en un corte transversal y clasificadas según varios factores explicativos. Esta aplicación de los MLG fue propuesta por Mc Cullagh y Nelder (1983) y puede encontrarse una amplia discusión del tema en Moreno (1986). Aquí se discuten únicamente los modelos específicos aplicados a probar las hipótesis respecto a la estructura subyacente a las distribuciones y proporciones señaladas en i) – ii).

La teoría de los MLG señala dos componentes básicos a especificar para modelar datos clasificados (Moreno, Ibid.), a saber: la hipótesis respecto a la estructura de las medias poblacionales de cada celda (componente sistemático), y los supuestos respecto a la distribución, en cada celda de la tabla, de las medias maestras respecto a las poblacionales (estructura de los errores).

Mientras que el primero se refiere a lo que tradicionalmente se conoce como modelar, i.e. seleccionar los factores y el tipo de efectos de estos que más contribuyen a explicar la variabilidad de los datos, el segundo componente debe ser especificado y validado empíricamente (es decir, con los datos maestras mismos) antes de iniciar la discusión respecto a la forma del componente sistemático.

Así, se discuten las estructuras de error (con función de probabilidad de la familia exponencial) consideradas más apropiadas para replicar la estructura de la variabilidad muestral de las distribuciones de cruces según horario y fecha y proporciones de cruces de residentes mexicanos acorde a los mismos criterios antes referidos.

Dado que las variables consideradas en i) – ii) se derivan de un proceso de conteo (de cifras absolutas en su caso, y de eventos “favorables” respecto a un número de ensayos en el otro), fácilmente se puede aceptar que las estructuras de error apropiadas para modelar los datos corresponden a distribuciones Poisson y binomial, respectivamente. En el caso de la primera existe la posibilidad de verificar empíricamente si los datos apoyan la evidencia de que la variabilidad muestral respecto a la poblacional se comporte como generada por un proceso Poisson; en el segundo caso, dado que la relación entre la media y la varianza de la distribución binomial queda definida una vez conocidos el número de ensayos y la proporción de “favorables” que se obtuvo (i.e. que un residente mexicano cruce la frontera), no es necesario verificar la adecuación de la elección de la estructura del error.

En el caso de este ejercicio, dado que no es posible estimar las varianzas muestrales para cada celda de la tabla de la distribución de cruces, pues la variable responsiva corresponde a un conteo sobre la unidad de estudio, se decidió seleccionar a la distribución poisson sobre la base de la argumentación teórica que sustenta la generación de un proceso de este tipo. (Ver Dobson, A., 1983).

Bajo estas condiciones, en la siguiente sección se buscarán los componentes sistemáticos que mejor describan las estructuras de datos referidos y, dentro de este proceso, verificar si la selección de la función de varianza ha sido adecuada para replicar, la variabilidad de las observaciones muestrales.

V. SELECCIÓN DE MODELOS

La discusión se presenta primeramente para las distribuciones de cruces según tipo de viajero, para después analizar lo correspondiente a las proporciones de mexicanos que cruzan la frontera respecto al total.

En ambos casos, el ejercicio se realiza para cada una de las plazas fronterizas en donde se han levantado los cuatro censos aludidos (ver Cuadro 1), así como de manera agregada para estas localidades.

V.I Análisis de las distribuciones de cruces

Dado que el análisis se decidió hacer separadamente para cada tipo de viajero y plaza (ver cuadro 4), las distribuciones porcentuales de cruces están clasificadas según horario, así como el día de la semana y el censo de registro. Esto da una tabla con $9 \times 7 \times 4 = 252$ celdas por modelar. Ante el gran número de posibles modelos a ensayar (sólo 19 modelos jerárquicos para una clasificación de triple entrada), era necesario buscar un método de selección que fuera informativo de los cambios en la precisión del ajuste al incluir o remover términos, pero que fuera económico en términos de computación.

De esta manera, se probaron todos los submodelos completamente saturados del modelo de tres factores, i.e. se estimaron los tres modelos unifactoriales y los tres modelos saturados de dos factores. A partir de estos se buscaron las interacciones que redujeran más significativamente al estadístico del cociente de log-verosimilitudes χ^2 .

El modelo final resultó aquel que incluyera todos los efectos unifactoriales significativos más la interacción de segundo orden que hubiera producido la mayor reducción en χ^2 . Así, por ejemplo, el Cuadro 1 reporta los valores del estadístico χ^2 obtenidos a partir del ajuste de un modelo loglineal con

reestructura de error Poisson a la distribución de frecuencias según horario, día de la semana y número de censo, de los cruces fronterizos de peatones para todas las plazas y considerando todos los modelos unifactoriales, los submodelos saturados de orden 2 y el modelo saturado de tres factores.

CUADRO 1. CAMBIOS EN CHI-CUADRADA PARA AJUSTES DE SUBMODELOS SATURADOS DE ORDEN 1, 2 Y 3

Modelo	gl	$D(\chi^2)$	$\bar{D}$
NULO	251	733	2.9203
NC	248	733	2.9556
HR	243	61	0.2510*
DS	245	733	2.9918
NC*HR	216	49	0.2269
NC*DS	224	733	3.2723
NC*HR*DS	0	0	--

* Mejor ajuste en términos de desviación media.

Nota: Ver nombres de las variables en el Cuadro 1. Dado el carácter exploratorio de este ejercicio, se utiliza el factor HR con nueve niveles. Ver nota al Cuadro A.1 en el Apéndice.

Es claro de los resultados que el único factor cuyo efecto principal explica variabilidad en los datos es el horario del día en el que se cruza, y ni el día de la semana ni la fecha en que se hizo el conteo explican de manera alguna la distribución de cruces.

Por el lado de los modelos saturados de orden 2, se observa que tanto el que incluye el efecto conjunto censo-horario, como el del horario-día de la semana reducen el valor del estadístico χ^2 . Las correspondientes pruebas de significancia de las interacciones se presentan en el Cuadro 2.

Dado que la única interacción estadísticamente significativa es la del efecto horario-día de la semana, el modelo final correspondería a uno ni jerárquico ni saturado con términos

$$1 + \text{HR} + \text{HR.DS}$$

CUADRO 2. PRUEBA DE SIGNIFICANCIA DE LOS EFECTOS BI-FACTORIALES SOBRE LA DISTRIBUCIÓN DE CRUCES FRONTERIZOS

Modelo	gl	$D(\chi^2)$	
NC+HR	240	61	$\chi^2 (24) = 12$
NC*HR	216	49	La interacción NC.HR <u>no</u> es significativa
NC+DS	242	733	$\chi^2 (18) = 0$
NC*DS	224	733	La interacción NC.DS <u>no</u> es significativa
HR+DS	237	62	$\chi^2 (48) = 42$
HR*DS	189	19	La interacción HR.DS es significativa

Donde 1 corresponde al término constante. El valor de χ^2 es de 19 con 186 grados de libertad. Este modelo explica aproximadamente el 97% de la variación total de los datos, como se puede observar en el Cuadro 3.

CUADRO 3. ANÁLISIS DE DESVIACIONES PARA LAS DISTRIBUCIONES DE VIAJEROS FRONTERIZOS.
MODELO LOG-LINEAL CON ERROR PISSON

DESVIACIÓN:	D	%D ^{TOT}	%D ^{EXP}
- Explicada por el modelo:			
HR+HR.DS	714	97.41	100.00
HR*	672	91.68	94.12
HR.DS*	42	5.73	5.88
- No explicada por el modelo	19	2.59	-
TOTAL	733	100.00	-

* Significativamente diferente de cero al 5%

De estos resultados se puede concluir que la distribución porcentual de cruces fronterizos de peatones no difiere significativamente a través de las fechas en que se levantaron los censos de estos movimientos, ni respecto al día de la semana en que se registró el cruce (excepto por el efecto conjunto de esta variable con el horario). La fuente de variabilidad principal corresponde a la hora del día en que se efectúa el movimiento.

V.2 Análisis de las distribuciones de cruces por plazo y tipo de viajero

A continuación se presenta un resumen de los resultados para el análisis específico de cada plaza según el tipo de viajero (peatones y automovilistas) . 10.

Para el mejor modelo ajustado*, se reportan cuadros similares al análisis de varianza jerárquico con datos no balanceados, señalando la proporción de la desviación total explicada por modelo y la atribuible a errores aleatorios y/o de medición. Igualmente, se cuantifica la contribución de cada término del modelo a la desviación total explicada

Destacan de estos resultados los siguientes puntos:

- a) Los efectos principales de los factores CENSO (NC) Y DÍA DE LA SEMANA (DS) no son significativamente diferentes de cero para ninguna de las plazas consideradas. De hecho, el HORARIO (HR) de cruce es el único factor cuyos efectos principales explica significativamente la variación de las observaciones (84% de los peatones y 60% de los automovilistas).
- b) La variabilidad de los datos sobre automovilistas representa, en promedio, el 30% de la variabilidad de los cruces de peatones. Como consecuencia de esto, los modelos ajustados a los primeros datos son más simples (parsimoniosos) que los ajustados al segundo conjunto de datos.
- c) La variación no explicada por los modelos es pequeña (9.2% para peatones y 22.2% para automovilistas). Cabe destacar que la variabilidad pudo haber sido reducida a costa de incrementar el número de parámetros en el modelo. Esto no es deseable sobre todo cuando los parámetros en el modelo. Esto no es deseable sobre todo cuando los parámetros se refieren a efectos conjuntos de factores, los cuales no son necesariamente fáciles de interpretar y cuantificar su impacto sobre la media total de los cruces fronterizos.
- d) Para ambos conjuntos de datos se detecta efectos simultáneos significativos de los factores HORARIO (HR) Y DÍA DE LA SEMANA (DS) en la mayoría de las plazas. Este efecto explica aproximadamente el 7% de la varianza de los datos de peatones y del 23% de la de los automovilistas.

Estos resultados implican diferentes formas relacionales de los efectos de los horarios sobre el promedio de cruces fronterizos para los días de la semana.

* Siguiendo los criterios señalados en V.1., el que minimiza la desviación media del ajuste.

Se presentan a continuación los cuadros 4 y 5 con el análisis de desviaciones para peatones y automovilistas respectivamente.

CUADRO 4. ANÁLISIS DE DESVIACIONES PARA LAS DISTRIBUCIONES DE CRUCES DE PEATONES SEGÚN PLAZA

NUEVO LAREDO	D	%D ^{TOT}	%D ^{EXP}
- Explicada por el modelo:			
NC+HR+DS+HR.DS	935.8	90.2	100.0
NC	0	0	0
DS	0	0	0
HR*	851.7	82.1	91.0
HR.DS*	84.1	8.1	9.0
- No explicada por el modelo	102.0	100.0	-
TOTAL	1,037.8	100.0	-
TIJUANA			
- Explicada por el modelo:			
NC+HR+DS+HR.DS	695.3	85.2	100.0
NC	0	0	0
DS	0	0	0
HR*	637.1	78.0	91.6
HR.DS*	58.2	7.1	8.4
- No explicada por el modelo	121.0	14.8	-
TOTAL	816.3	100.0	-
MEXICALI			
- Explicada por el modelo:			
NC+HR+NC+HR	1,085.3	93.9	100.0
NC	0	0	0
HR*	1,024.4	88.6	94.4
NC.HR*	60.9	5.3	5.6
- No explicada por el modelo	70.6	6.1	-
TOTAL	1,155.9	100.0	-
CIUDAD JUÁREZ			
- Explicada por el modelo:			
NC+DS+HR.DS	1,463.1	92.8	100.0
NC	0	0	0
DS	0	0	0
HR*	1,310.3	83.1	89.6
HR.DS*	152.8	9.7	10.4
- No explicada por el modelo	114.6	7.2	-
TOTAL	1,577.7	100.0	-
MATAMOROS			
- Explicada por el modelo:			
NC+HR+DS+HR.DS	1,059.6	91.9	100.0

NC	0	0	0
DS	0	0	0
HR*	996.8	86.4	94.7
HR.DS*	62.8	5.5	5.9
- No explicada por el modelo	93.9	8.1	-
TOTAL	1,153.5	100.0	-

* Significativamente diferente de cero al 5%.

CUADRO 5. ANÁLISIS DE DESVIACIONES PARA LAS DISTRIBUCIONES DE CRUCES DE AUTOMOVILISTAS SEGÚN PLAZA

NUEVO LAREDO	D	%D ^{TOT}	%D ^{EXP}
- Explicada por el modelo:			
NC+HR+DS+HR.DS	935.8	90.2	100.0
NC	0	0	0
DS	0	0	0
HR*	851.7	82.1	91.0
HR.DS*	84.1	8.1	9.0
- No explicada por el modelo	102.0	100.0	-
TOTAL	1,037.8	100.0	-
TIJUANA			
- Explicada por el modelo:			
NC+HR+DS+HR.DS	695.3	85.2	100.0
NC	0	0	0
DS	0	0	0
HR*	637.1	78.0	91.6
HR.DS*	58.2	7.1	8.4
- No explicada por el modelo	121.0	14.8	-
TOTAL	816.3	100.0	-
MEXICALI			
- Explicada por el modelo:			
NC+HR+DS	1,085.3	93.9	100.0
NC	0	0	0
DS	1,024.4	88.6	94.4
HR*	60.9	5.3	5.6
- No explicada por el modelo	70.6	6.1	-
TOTAL	1,155.9	100.0	-
CIUDAD JUÁREZ			
- Explicada por el modelo:			
NC+DS+HR.DS	1,463.1	92.8	100.0
NC	0	0	0
DS	0	0	0
HR*	1,310.3	83.1	89.6
HR.DS*	152.8	9.7	10.4
- No explicada por el modelo	114.6	7.2	-
TOTAL	1,577.7	100.0	-
MATAMOROS			
- Explicada por el modelo:			
NC+HR+DS+HR.DS	1,059.6	91.9	100.0
NC	0	0	0
DS	0	0	0
HR*	996.8	86.4	94.7
HR.DS*	62.8	5.5	5.9
- No explicada por el modelo	93.9	8.1	-
TOTAL	1,153.5	100.0	-

* Significativamente diferente de cero al 5%.

Dado que NC, el efecto a través del tiempo, resultó ser no significativo, no ofrecen mayor información los coeficientes del factor dado que el nivel y tendencia que implicarían no serían significativos estadísticamente hablando. Lo mismo se puede decir de DS. Para HR, horario del día, los resultados pueden solicitarse al autor.

En resumen, el análisis de la distribución de cruces fronterizos según horario, día de la semana, censo y localidad, indica que a través del tiempo, i.e. para los cuatro levantamientos señalados, no existen diferencias significativas para las distribuciones de peatones y automovilistas según horario de cruce. Esta regularidad de la estructura sugiere que las estimaciones derivadas del modelo seleccionado pueden ser utilizadas, ceteris paribus, para calcular los expansores de los estimadores de razón que se utilizan para inferir el número de cruces de residentes en México y, a partir de este, el gasto medio de los viajeros fronterizos residentes en México, mismo que se incluye en la Balanza de Pagos del país. (Banco de México, 1982).

El modelo se reestimaría cuando a partir de la evidencia obtenida de los levantamientos mensuales se detectaran discrepancias significativas de la estructura común a los cuatro censos de "cruces fronterizos".

V.3 Análisis de las proporciones de cruces fronterizos de residentes en México

Persiguiendo objetivos similares a los de la anterior sección, se presenta aquí el análisis del número de cruces fronterizos de los residentes en México (o su contraparte, en el extranjero) respecto al total de cruces fronterizos.

Dado que los factores explicativos considerados son los mismos que para el análisis de las distribuciones, i.e. el componente sistemático se seleccionará de entre el mismo conjunto de modelos subsaturados de 1 y 2 factores, sólo habrá que modificar la distribución de probabilidad que represente la estructura del error, así como la función conectiva que garantice el que las proporciones estimadas por el modelo siempre estén entre cero y uno. Para la primera se utilizará la distribución binomial, que es precisamente la que genera el proceso estocástico que se intenta modelar; para la segunda, se empleará la transformación logito (o logit). (ver Moreno, 1986).

En cuanto al plan de análisis, se siguen las mismas ideas que en el ejercicio por plazas de V.2.

Los cuadros 6 y 7 reportan el análisis de variabilidad de las proporciones observadas de peatones y automovilistas residentes en México según plaza de cruce

CUADRO 6. ANÁLISIS DE VARIABILIDAD PARA LAS PROPORCIONES DE CRUCES DE PEATONES RESIDENTES EN MÉXICO SEGÚN PLAZA*

NUEVO LAREDO	D	%D^{TOT}
- Explicada por el modelo:		
NC+HR+DS+HR.DS	6,642	72.9
HR.DS	291	3.2
- No explicada por el modelo	2,465	27.1
TOTAL	9,107	100.0
TIJUANA		
- Explicada por el modelo:		
NC+HR+DS+NC.HR	6,903	71.3
NC.HR	1,003	10.4
- No explicada por el modelo	2,784	28.7
TOTAL	9,687	100.0
MEXICALI		
- Explicada por el modelo:		
NC+HR+DS+HR.DS	3,613	68.3
HR.DS	738	14.0
- No explicada por el modelo	1,676	31.7
TOTAL	5,289	100.0
CIUDAD JUÁREZ		
- Explicada por el modelo:		
NC+HR+DS+NC.HR	2,337	43.7
HR.DS	753	14.1
- No explicada por el modelo	3,007	56.3
TOTAL	5,344	100.0
MATAMOROS		
- Explicada por el modelo:		
NC+HR+DS+NC.DS	4,753	67.1
NC.DS	798	11.3
- No explicada por el modelo	2,326	32.9
TOTAL	7,079	100.0

* La contribución de cada efecto principal a la variabilidad total se reporta por separado.

CUADRO 7. ANÁLISIS DE VARIABILIDAD PARA LAS PROPORCIONES DE

CRUCES DE AUTOMOVILISTAS RESIDENTES EN MÉXICO SEGÚN PLAZA*

NUEVO LAREDO	D	%D^{TOT}
- Explicada por el modelo:		
NC+HR+DS+HR.DS	1,580	64.5
HR.DS	273	11.1
- No explicada por el modelo	870	35.5
TOTAL	2,450	100.0
TIJUANA		
- Explicada por el modelo:		
NC+HR+DS+HR.DS	9,077	74.7
HR.DS	2,232	18.4
- No explicada por el modelo	3,069	25.3
TOTAL	12,146	
MEXICALI		
- Explicada por el modelo:		
NC+HR+DS+HR.DS	4,552.0	75.7
HR.DS	437.0	7.3
- No explicada por el modelo	1,461.0	24.3
TOTAL	6,013.0	100.0
CIUDAD JUÁREZ		
- Explicada por el modelo:		
NC+HR+DS+HR.DS	2,923	76.1
HR.DS	182	4.7
- No explicada por el modelo	916	23.9
TOTAL	3,839	100.0
MATAMOROS		
- Explicada por el modelo:		
NC+HR+DS+HR.DS	2,224	70.6
HR.DS	310	9.8
- No explicada por el modelo	925	29.4
TOTAL	3,149	100.0

*La contribución de cada efecto principal a la variabilidad total se reporta por separado.

Las siguientes conclusiones se derivan de los siguientes cuadros:

- Para ambos tipos de viajeros (peatones y automovilistas), los efectos principales de los tres factores son significativos, por lo que se decidió incluir a los tres en el modelo final. Respecto a los efectos bifactoriales, se prefirió el que más contribuyera a reducir el valor de χ^2 , que para los automovilistas en todas las plazas resultó ser el efecto horario-día de la semana. En el caso de los peatones también se destacó este factor, si bien apareció como predominante en dos ciudades el efecto horario-número de censo.

- El modelo seleccionado (los tres efectos principales más la interacción bifactorial que en mayor porcentaje reduce la variabilidad de los datos) explican aproximadamente el 75 por ciento de la varianza total. Vuelve a destacarse que este porcentaje pudo haberse incrementado a costa de perder grados de libertad por la inclusión de más parámetros en el componente sistemático, lo cual se aleja del objetivo de detectar patrones temporales en el comportamiento de las proporciones de cruces.
- La importante variabilidad de estos datos (comparada con la de la distribución de cruces) requiere de un número considerablemente mayor de parámetros para explicar su estructura, cuyos efectos deberán de ser interpretados de manera sencilla y clara. Para esto, se discuten a continuación el concepto de “riesgo relativo”. Este se refiere al cociente de las probabilidades de experimentar el evento “favorable” para dos grupos de individuos según diferentes niveles de los factores explicativos. Estas probabilidades se obtienen de dividir el número de cruces de residentes mexicanos estimados por el modelo entre el total de cruces fronterizos de la celda correspondiente.

Una medida aproximada al “riesgo relativo” es el de “cociente de momios”, que equivale a la cuantificación del efecto de encontrarse en un nivel de un factor respecto al nivel que se emplea para contrastes. Se obtiene precisamente de aplicar el antilogaritmo al coeficiente obtenido para el nivel del factor que se discute en un modelo logia-lineal.

Precisamente a partir de los riesgos relativos se puede cuantificar, en forma resumida y sencilla, la propensión relativa a cruzar la frontera para residentes en México que lo han hecho en diferentes horarios o días de la semana e incluso cuyos movimientos fueron registrados en distintos censos. Sea, por ejemplo, 0.5703 la probabilidad estimada de un residente mexicano de cruzar la frontera como peatón en Matamoros un lunes entre las seis y ocho de la mañana para el período del primer censo (Noviembre 1983). Para el mismo horario y día de la semana, pero para el censo 3 (Marzo de 1985), la probabilidad es de 0.3676. Luego pues, el riesgo relativo de un residente mexicano a cruzar la frontera entre el censo uno y el tres es igual a $0.5703/0.3676 = 1.551$, esto es, cincuenta y cinco por ciento mayor en Noviembre de 1983 que en Marzo de 1985.

Las tablas no se reportan en este documento por razones de espacio, pudiéndolas solicitar directamente el autor.

- Los coeficientes de los efectos principales de los factores no muestran patrón regular a través del tiempo (censos) para el caso de los peatones. Para el otro tipo de viajero-automovilistas- hay evidencia de alguna tendencia a través del tiempo; así, la proporción de cruces fronterizos de residentes mexicanos se redujo entre el primero y último censo en Mexicali y Tijuana. En las plazas de Matamoros y Nuevo Laredo, la referida proporción, se incrementó. No hay regularidad destacable en Cd. Juárez. La explicación a este comportamiento deberá de buscarse en fenómenos económicos que hayan afectado considerablemente el tráfico fronterizo de automovilistas y, menos claramente, el de peatones.

En resumen, las proporciones de cruces fronterizos de mexicanos residentes en México, presentan una considerable variabilidad tanto para peatones como para automovilistas. Los tres factores explicativos utilizados en este ejercicio: censo, horario y día de la semana, resultaron significativos en explicar el comportamiento de las referidas proporciones. Las interacciones entre estos factores (de orden 2) también resultaron significantes para la mayoría de las combinaciones y se retuvo la de mayor poder explicativo en el modelo final. Por consecuencia, este resultó con un importante número de parámetros y, por lo mismo, pérdida de grados de libertad.

Finalmente, se utilizaron los “riesgos relativos” para comparar respecto a los niveles de alguno de los factores analizados la propensión relativa a cruzar la frontera entre individuos de diferentes celdas de la tabla.

VI. UTILIZACIÓN DE FACTORES CON ESCALAS ORDINALES PARA MODELOS DE ASOCIACIÓN DE CRUCES FRONTERIZOS

Como se ha visto hasta aquí, los modelos lineales aditivos para probar independencia usualmente no se ajustan adecuadamente a los datos y, en el esquema jerárquico usual, el siguiente

modelo de mayor complejidad es el saturado, que consiste de un número substancialmente mayor de parámetros adicionales asociados a efectos conjuntos de factores (interacciones).

Sin embargo, utilizando as información ordinal de las variables (cuando es el caso), existen modelos más simples que son menos complejos y más realistas que el modelo de independencia, si bien no son saturados. Por lo anterior, los modelos con variables ordinales son más parsimoniosos que los modelos estándar. Este punto es particularmente útil cuando se requieren interpretar un número importante de interacciones estadísticamente significativas, como es el caso del modelo logitlineal de proporciones discutido en la Secc. V.3.

Una descripción detallada de los modelos lineales con variables ordinales se ofrece en Agreste (1983). Aquí se presentan algunos ejercicios ilustrativos de estos conceptos de interés en el contexto del modelaje de los cruces fronterizos.

Con respecto a la distribución relativa de cruces según horario, analizada en V.1 y V.2, fácilmente pueden destacarse las bondades de utilizar variables ordinales en el correspondiente modelo loglineal seleccionado.

Considérense las nuevas variables $ORNC=NC-MNC$, $ORHR=HR-MHR$ y $ORDS=DS-MDS$, donde MNC, MHR y MDS corresponden a las medias de los valores asignados a las categorías ordinales de los factores censo, horario y día de la semana, respectivamente. A partir de éstas, los términos $OR1=ORNC.ORHT$, $OR2=ORN.ORDS$ y $OR3=ORDS.ORHT$, representan las desviaciones del modelo lineal aditivo (independencia) o asociaciones condicionales.

Así, el modelo

$$\text{Log}(VF)=1+NC+HR+DS+OR1+OR2+OR3$$

sólo tiene tres parámetros más que el modelo aditivo pero es mucho más parsimonioso y fácil de interpretar que el submodelo jerárquico saturado de orden 2, i.e. el que incluye las tres interacciones bifactoriales.

El coeficiente para cada par de variables ordinales refleja el alejamiento de $\log(VF)$ de la independencia, donde la asociación condicional es homogénea para cada par de variables.

Para ilustrar los anterior, se discute el caso de los peatones registrados en Ciudad Juárez.

Como era de esperarse, según puede verse en el Cuadro 8, el único efecto ordinal de dos factores que resulta significativo es OR3, es decir, el de horario y día de la semana.

CUADRO 8. CAMBIOS EN G^2 ATRIBUIBLES A LA INCLUSIÓN DE EFECTOS PRINCIPALES Y ORDINALES BIFACTORIALES

TÉRMINO INCLUIDO	G^2	gl.	Diferencia en G^2	Diferencia en gl
1	1577.7	233	0	3
NC	1577.7	220	0	6
DS	1577	214	1 310.3*	7
Hr	267.4	207	0	1
OR1	267.4	206	1.4	1
OR2	266.0	205	10.3*	1
OR3	255.7	204		

* Significativo al 0.01 por ciento.

Restringiéndose al modelo de términos significativos, $1+HR+OR3$, con sólo 9 parámetros y que explica el 84% de la varianza total, sin duda es mucho más fácil de interpretar que el similar de la sección V.1, con 62 parámetros y explicando el 93% de la variabilidad.

Los valores de los correspondientes parámetros se reportan en el Cuadro 9. El parámetro de OR3 (-0.0613) sugiere una asociación negativa entre horario y día de la semana, dada una fecha de levantamiento censal, como fue establecido en el modelo inicial. Los parámetros correspondientes al factor HORARIO se interpretan de la manera habitual.

CUADRO 9. PARÁMETROS DE MODELO ASOCIATIVO HOMOGÉNEO PARA MODELAR LA DISTRIBUCIÓN DE CRUCES FRONTERIZOS DE PEATONES EN CIUDAD JUÁREZ.

Parámetro	Valor Estimado	Error Estándar
1	1.291	0.1074
HR(2)	-0.554	0.1749
HR(3)	0.978	0.1245
HR(4)	1.445	0.1190
HR(5)	1.522	0.1194
HR(6)	1.722	0.1187
HR(7)	2.055	0.1173
HR(8)	0.3125	0.1465
OR3	-0.0163	0.0054

Ahora bien, respecto al análisis de las proporciones de residentes mexicanos que cruzan la frontera, ya considerado en V.3, el ejercicio reviste un interés particular dado el gran número de parámetros requerido para interpretar los efectos bifactoriales: censo-horario y día de la semana-horario, significativos en todos los modelos ensayados. El modelo logit-lineal se recalcula utilizando los efectos principales de las tres variables: NC,HR y DS, más las variables ordinales asociativas OR1, OR2 y OR3 ya definidas con anterioridad.

Por la gran parsimonia que se obtiene con el uso del modelo asociativo homogéneo, los resultados se discuten tanto para peatones como automovilistas en las cinco plazas de interés y señala su superioridad interpretativa respecto a los modelos anteriores.

En el cuadro 10 se presenta un análisis de variación para el modelo asociativo homogéneo. El cuadro reporta, para cada tipo de viajero y plaza, el modelo final seleccionado, el número de parámetros incluidos y el porcentaje de la variación total explicada por el modelo. Estas últimas cifras son comparables con las reportadas en los cuadro 6 y 7.

Varios puntos se destacan de estos resultados:

- Respecto a las variables asociativas incluidas en el modelo, que resultaron las estadísticamente significativas, destaca que no existe una estructura común para todas las plazas, tanto en el caso de peatones como de automovilistas. En el primer caso requieren las tres variables asociativas para replicar las proporciones observadas, excepto en Mexicali y Matamoros. Se señala particularmente el caso de Mexicali, pues en el análisis del Cuadro 6, resultó significativo el efecto OR1: horario-día de la semana en lugar del OR2: horario-censo de este ejercicio. La explicación habrá que sustentarla en que el supuesto de homogeneidad del factor asociativo OR2 no está suficientemente respaldado por la evidencia empírica.

Respecto a los automovilistas, la estructura paramétrica es más sencilla pues sólo se requieren dos términos asociativos en la mayor parte de las plazas, siendo la excepción Cd. Juárez, donde se presenta una situación igual a la ya señalada para Mexicali.

CUADRO 10. ANÁLISIS DE VARIACIÓN PARA EL MODELO ASOCIATIVO HOMOGENEO DE
LAS PROPORCIONES DE CRUCES FRONTERIZOS DE LOS RESIDENTES EN MÉXICO

	MODELO FINAL	PORCENTAJE DE LA VARIACIÓN TOTAL EXPLICADA	PARÁMETROS EN EL MODELO
PEATONES			
PLAZA			
NUEVO LAREDO	OR1+OR2+OR3	63	20
TIJUANA	OR1+OR2+OR3	49	20
MEXICALI	OR2	45	18
CD. JUÁREZ	OR1+OR2+OR3	36	20
MATAMOROS	OR1+OR3	45	19
AUTOMOVILISTAS			
PLAZA			
NUEVO LAREDO	OR1+ OR3	41	19
TIJUANA	OR1+OR3	54	19
MEXICALI	OR2+OR3	62	19
CD. JUÁREZ	OR2	54	18
MATAMOROS	OR1+OR3	40	19

NOTA: Las variables asociativas tienen las siguientes definiciones: OR1=ORHR.ORDS; OR2=ORHR.ORNC; OR3=ORNC.ords donde ORHR=HR-4.5; ORNC=NC-2 Y ORDS=DS-4

El fenómeno podría deberse tanto a lo ya señalado, como igualmente a celdas atípicas debidos a errores no aleatorios.

- La proporción de variación explicada respecto a la total promedia casi 48 porciento para los peatones y 50 porciento para los automovilistas. Si bien es menor en un 20 porciento a la resultante en los ajustes estándar, debe considerarse el importante número de parámetros ahorrados con el modelo asociativo homogéneo: 54 parámetros para la asociación horario-día de la semana, 30 para horario-censo y 26 para la interacción censo-día de la semana.

La facilidad de interpretación es obvia.

En el Cuadro 11 se reportan los valores estimados de los coeficientes de las variables ordinales asociativas incluidas en el modelo final. El signo del coeficiente señala la dirección de la asociación; así, para los peatones de Tijuana, el valor de -0.03037 para el término ORHR.ORNC indica una asociación negativa entre el factor horario y número de censo. Esto es, dado un horario específico, la proporción de residentes en México que cruzan la frontera se reduce en 3% ($1-e^{-0.03037}$) de un censo a otro.

Es importante destacar que la parsimonia ganada por estos modelos se obtiene a costa del supuesto de homogeneidad del efecto para todos los niveles de los factores involucrados, lo cual no siempre es realista al modelar estructuras complejas de datos. En este caso, donde el ejercicio pretendía detectar.

CUADRO 11. PARÁMETROS DE LAS VARIABLES ORDINALES ASOCIATIVAS INCLUIDAS EN EL MODELO FINAL

PEATONES					
VARIABLE	PLAZA NUEVO LAREDO	TIJUANA	MEXICALI	CD. JUÁREZ	MATAMOROS
ORHR.ORNC	0.02691 (0.01028)	0.00943 (0.00894)	-	-0.01312 (0.01983)	-0.00972 (0.01308)
ORHR.ORNC	-0.01754 (0.01795)	-0.03037 (0.01702)	0.00363 (0.01686)	0.02573 (0.03579)	-
ORNC.ORDS	0.02929	0.01376	-	-0.02622	0.07976

	(0.01627)	(0.01531)		(0.03112)	(0.02049)
ATOMOVILISTAS					
VARIABLE	PLAZA NUEVO LAREDO	TIJUANA	MEXICALI	CD. JUÁREZ	MATAMOROS
ORHR.ORDS	0.00968 (0.00571)	0.01411 (0.00581)	-	-	0.01484 (0.00627)
ORHR.ORNC	-	-	0.01507 (0.00581)	0.01821 (0.01173)	-
ORNC.ORDS	0.01126 (0.01091)	-0.01871 (0.01143)	-0.03230 (0.01029)	-	0.03066 (0.01465)

NOTAS

- Las cifras en paréntesis corresponden a los errores estándar.
- El efecto es exponencial, por lo que los coeficientes deben ser interpretados acorde a ello.

y describir tendencias en los efectos de algunos factores sobre una medición de los cruces fronterizos más que cuantificarlos o pronosticarlos, el recurrir a algún modelo claro y sencillo para replicar los datos observados facilita enormemente el problema.

CONCLUSIONES

Utilizando diversos modelos lineales generalizados, a saber: log-lineal y logia-lineal, se han explorado las estructuras tanto de la distribución porcentual según horario de los cruces fronterizos, como de las proporciones de residentes en México para el mismo evento. Mientras que en el primer ejercicio se detectó una estructura bastante estable y regular, en el caso de las proporciones la variabilidad es considerable. De los tres factores explicativos considerados: censo, horario y día de la semana, únicamente el efecto principal del horario resultó significativo en explicar las variaciones de las de distribuciones de cruces, además de conjunto del horario mismo con el día de la semana. Para las proporciones, los efectos principales de los tres factores resultaron ser de importancia en describir la variabilidad de la estructura de datos, además de requerirse al menos la interacción horario-día de la semana y/o horario-censo para reproducir los datos con precisión.

Finalmente, se ensayaron modelos de efectos asociativos homogéneos, i.e. utilizando la información de los factores ordenados, tanto para los modelos log-lineales como los loti-lineales, mismos que sobresalen por su parsimonia y sencillez interpretativa. Estos modelos resultan muy convenientes para detectar y describir tendencias en los efectos de factores, como es el caso del análisis del marco de referencia de la Encuesta de Transacciones Fronterizas.

REFERENCIAS

- AGRESTI, A. (1983), "A survey of Strategies for Modeling Cross-Classifications having ordinal variables".
Journal of the American Statistics Association 78-381, pp. 184-198
- BANCO DE MÉXICO (1982), Encuesta de Transacciones Fronterizas, Metodología.
 Serie de Documentos de Investigación. Dirección de Investigación Económica, México, D.F.
- _____ (1987), Revisión Metodológica de la Encuesta de Transacciones Fronterizas.
 Documento de Trabajo. Dirección de Investigación Económica. México, D.F.
- DOBSON, A. (1983), Introduction to Statistical Modelling, Chapman and Hall, London.
- MC CULLAGH, P. and J.A. NELDER (1983), Generalized Linear Models, Monographs on Statistics and Applied Statistics, Chapman and Hall, London.
- MORENO, L. (1986) Modelos Lineales Generalizados para Datos Clasificados,
 Mimeografiado. Documento de Trabajo. Dirección de Investigación Económica, Banco de México, México, D.F.

APÉNDICE

CUADRO A.1 DEFINICIÓN DE LOS NIVELES DE LOS FACTORES
INTERVINIENTES EN EL ANÁLISIS

VF = Cruces Fronterizos (Absolutos)

LR = Lugar de Residencia Habitual

1 = Residente en México

2 = Residente en el Extranjero

NC = Número de Censo

1 = Noviembre 1983

2 = Julio 1984

3 = Marzo 1985

4 = Enero 1986

HR = Horario de Conteo de Viajes*

1 = 22:00 – 6:00

2 = 6:00 – 8:00

3 = 8:00 – 10:00

4 = 10:00 – 12:00

5 = 12:00 - 14:00

6 = 14:00 – 16:00

7 = 16:00 – 18:00

8 = 18:00 – 20:00

9 = 20:00 – 22:00

DS = Día de la Semana

1 = Lunes

2 = Martes

3 = Miércoles

4 = Jueves

5 = Viernes

6 = Sábado
7 = Domingo

MT = Medio de Transporte**
1 = Peatón
2 = Automovilista

PL = Plaza**
1 = Cd. Juárez
2 = Matamoros
3 = Mexicali
4 = Tijuana
5 = Nuevo Laredo

* En el análisis por plaza tanto de la distribución de cruces como de proporciones de cruces atribuibles a residentes mexicanos se utilizan sólo ocho niveles para el factor HR, siendo el primer nivel el correspondiente al horario de 22:00 a 8:00 horas, el segundo para las horas de 8:00 a 10:00 y así sucesivamente.

** Variables no incluidas como factores explicativos en el análisis.