



UNIVERSIDAD NACIONAL DE COLOMBIA

Modelación longitudinal de la migración de parejas mexicanas a Estados Unidos y su fecundidad

Maite Michell Gómez Gómez

Universidad Nacional de Colombia
Maestría en Ciencias - Estadística
Bogotá, Colombia
2024

Modelación longitudinal de la migración de parejas mexicanas a Estados Unidos y su fecundidad

Maite Michell Gómez Gómez

Trabajo de grado como requisito parcial para optar al título de:
Magíster en Ciencias - Estadística

Directora:
Ph.D. Luz Mery Gonzalez García

Universidad Nacional de Colombia
Facultad de ciencias
Bogotá, Colombia
2024

Dedicatoria

A mis padres, Dilcen y Leonardo.

A mis hermanos, Harold y Michael.

A mis Abuelas.

A Agustin.

Agradecimientos

A Luz Mery Gonzalez, Ph.D, mi directora, gracias por su tiempo, paciencia, perseverancia y gran apoyo en este trabajo de grado. Mi eterno agradecimiento.

A Adriana Carolina Silva Arias, Ph.D, quien me compartió su pasión por la academia y aportó de forma significativa en mi desarrollo profesional y personal.

A mi familia por su constante apoyo y dedicación en todos los proyectos propuestos, son mi mayor impulso.

A Karen Mejía por su confianza y su ímpetu para apoyarme en cada uno de mis proyectos, por no dejarme desfallecer ante las dificultades.

Resumen

Modelación longitudinal de la migración de parejas mexicanas a Estados Unidos y su fecundidad

La migración y la fecundidad generan alteraciones en las dinámicas demográficas de los países de origen y destino, de ahí la importancia de su estudio. Este trabajo de grado analiza la relación entre migración internacional a Estados Unidos y la fecundidad en las parejas mexicanas. La información proviene del Proyecto Mexicano de Migración (MMP), el cual provee la historia de eventos de jefes de hogar. Con los datos del proyecto, se reconstruye la historia de eventos de las parejas mexicanas debido a que la migración y fecundidad son decisiones familiares. Mediante ecuaciones de estimación generalizada se estima la probabilidad de que la pareja tenga un hijo, dado que uno o los dos cónyuges migró y la probabilidad de que al menos uno de los cónyuges migre dado que tienen hijos. De esta manera, se usa análisis longitudinal en datos demográficos, que han sido estudiados principalmente con datos de corte transversal. En específico, se centra en temporalidad de 1964 al 2013.

Palabras clave: Modelos lineales generalizados, modelos longitudinales, demografía, migración.

Abstract

Longitudinal modeling of the migration of Mexican couples to the United States and their fertility

Migration and fertility generate alterations in the demographic dynamics of the countries of origin and destination, hence the importance of their study. This degree work analyzes the relation between international migration to the United States and fertility in Mexican couples. The information comes from the Mexican Migration Project (MMP), which provides the history of events of household heads. With the project data, the history of events of Mexican couples is reconstructed because migration and fertility are family decisions. This analysis aims to estimate, using generalized estimation equations, the probability that the couple has a child, given that one or both spouses migrated, and the probability that at least one of the spouses migrates, given that they have children. In this way, we enter longitudinal analyzes on demographic data that have been studied mainly with cross-sectional data. Specifically, it focuses on temporality from 1964 to 2013.

Keywords: Generalized linear models, longitudinal models, demography, migration.

Contenido

Agradecimientos	iv
Resumen	v
1. Introducción	2
2. Marco teórico	5
2.1. Notación	5
2.2. Modelos lineales generalizados	6
2.3. Ecuaciones de estimación generalizada (EEG)	8
2.3.1. Estructuras de correlación	8
2.3.2. Estimación de parámetros	10
2.4. Métodos de diagnóstico EEG	12
2.4.1. Selección de una estructura de correlación	12
2.4.2. Análisis de residuales	13
2.5. Datos desbalanceados	15
3. Fuentes de información y análisis descriptivo	16
3.1. Datos	16
3.1.1. Análisis descriptivo	18
4. Modelación longitudinal	25
4.1. Modelo propuesto para migró / no migró	25
4.1.1. Diagnóstico modelo propuesto migró / No migró	27
4.1.2. Resultados modelo propuesto migró / No migró	30
4.2. Modelo propuesto para cantidad de hijos	31
4.2.1. Diagnóstico modelo propuesto cantidad de hijos	34
4.2.2. Resultados modelo cantidad de hijos	36
5. Conclusiones	39
6. Trabajos futuros	41
A. Anexo I: Código QIC	42

B. Anexo II: Dfbetas Modelo Propuesto Migró/ No Migró	44
C. Anexo III: Dfbetas Modelo Propuesto Cantidad de Hijos	48
Bibliografía	51

Lista de Figuras

3-1. Mapa de las comunidades del MMP encuestadas. (MMP, 2022b)	17
3-2. Cantidad de hijos según estado de la pareja (migrantes o no migrantes).	24
4-1. Δ_{ij} vs las observaciones modelo migró /no migró	28
4-2. r_{SD} vs parejas modelo migró / no migró	29
4-3. Distancia de cook modelo migró / no migró.	29
4-4. Residuales vs predicho modelo migró / no migró	30
4-5. Δ_{ij} vs las observaciones modelo número de hijos	35
4-6. r_{SD} vs parejas modelo número de hijos	35
4-7. Distancia de Cook modelo número de hijos.	36
4-8. Residuales vs predicho modelo cantidad de hijos	36
B-1. Edad de los hombres	44
B-2. Años educación de mujeres	44
B-3. Años pareja	45
B-4. Experiencia laboral de hombres	45
B-5. Hombre no en la fuerza laboral	45
B-6. Hombre ocupación agricultura	46
B-7. Hombre ocupación Manufactura	46
B-8. Hombre ocupación Profesional o administrativo	46
B-9. Hombre ocupación servicios	47
B-10 Hombre ocupación Técnicos	47
B-11 Prevalencia migratoria de hombres	47
C-1. Migración hombres	48
C-2. Migración mujeres	48
C-3. Edad mujeres	49
C-4. Años educación hombres	49
C-5. Años educación mujeres	49
C-6. Prevalencia migratoria hombres	50
C-7. Años pareja	50

Lista de Tablas

2-1. Funciones de enlace comunes	7
3-1. Cambios de estado migratorio	18
3-2. Movimientos migratorios	20
3-3. Categorías migratorias.	21
3-4. Características de las parejas mexicanas según estado migratorio.	22
3-5. Cantidad de parejas según máximo de años pareja.	23
4-1. Estructuras de correlación. QIC y CIC.	26
4-2. Variables modelo migró / no migró	27
4-3. Categorías variable ocupación	27
4-4. Estimación de parámetros modelo migró / no migró.	31
4-5. Estructuras de correlación. QIC y CIC.	32
4-6. Variables modelo cantidad de hijos	34
4-7. Estimación de parámetros Número de hijos.	37

1. Introducción

El papel de los migrantes para compensar los patrones demográficos actuales ha sido un tema de agenda pública y académica, con repercusiones en el mercado laboral y en el proceso de integración de los migrantes (Ferrer, 2014; ONU, 2001). Sin embargo, es importante conocer las dinámicas demográficas de décadas pasadas y sus transformaciones.

La población de América Latina en el extranjero, principalmente en Estados Unidos, aumentó de 1 millón de personas en 1960 a 14.5 millones en 2000 (Pellegrino, 2003), siendo principalmente migrantes mexicanos. Además del crecimiento de las migraciones de América Latina, la fecundidad¹ también ha influido en la transformación demográfica de la región. En 1960, América Latina registró una tasa global de fecundidad (TGF) de 6 hijos por mujer, pero en la actualidad, esta tasa es de sólo 2 (CEPAL, 2019). Lo anterior es debajo de la tasa de reemplazo de 2.1 hijos por mujer. Sin embargo, en el caso de México, esta tasa es de 2.29 pero se proyecta a 1.72 en el 2050 (UNFPA, 2018).

Debido a las bajas tasas de fecundidad y las altas tasas de migración presentados por algunos países, especialmente en México, varias investigaciones identificaron el impacto de la migración sobre la fecundidad (Choi, 2014; Fargues, 2011; Carter, 2000; Schultz, 1980; Kulu, 2013; Wolf, 2016; Grande, 2017; Andersson, 2017; Dubuc, 2012; Frank Heuveline, 2005; Lindstrom, 2002), además del impacto de la fecundidad en la migración (Long, 1972; Lindstrom, 2007; Sandell, 1977; Kulu, 2008; Milewski, 2010; Madhavan, 2012; Oliveira, 2016; Cooke, 2016).

En el caso de la fecundidad como variable respuesta, Wolf (2016) evaluó los migrantes de Turquía en Alemania. Para conocer la relación entre un nacimiento con la duración de estadía de los migrantes en Alemania, lo realizó usando un modelo logístico en el que considera una variable tiempo-dependiente como los años de duración de estadía en Alemania y variables constantes en el tiempo, como la inclusión del historial de matrimonio y migración de la pareja, la cohorte de nacimiento, el nivel educativo y la edad de migración del encuestado. Por su parte, Carter (2000) y Andersson (2017) utilizaron las historias de vida y nacimientos disponibles (Carter en México y Anderson en Suecia), para estimar un modelo logit multinomial, donde la variable respuesta fue el primer, segundo y tercer nacimiento utilizando principalmente variables explicativas como el nivel educativo, estatus socio-económico, esta-

¹Número de nacimientos por cada mil mujeres en edad fértil.

tus laboral y edad de las mujeres (siendo todas las variables tiempo-dependientes) en Estados Unidos y Suecia, respectivamente.

Entre los estudios que toman la migración como variable respuesta, Kulu (2008) realizó un modelo logit multinomial con variable respuesta la migración a áreas rurales, pequeñas ciudades o grandes ciudades, y variables explicativas tiempo-dependientes como la edad, años desde el primer nacimiento, años desde el segundo nacimiento, años desde el tercer nacimiento, años previos antes de migrar, nivel educativo, estatus laboral y estado civil y variables constantes como género del nacimiento.

De forma similar, Kulu (2013) usó un modelo logit multinomial para estimar las transiciones de vivienda de las parejas, para esto usaron como variable respuesta las categorías de vivienda: casa unifamiliar, casa adosada (vivienda con tres o más casas en una fila de casas y que comparten una pared con su vecino adyacente) y apartamento. Como variables explicativas la duración de la unión (años), edad de la mujer, idioma hablante, nivel educativo de la mujer, sus ingresos anuales (ninguno, bajo, medio, alto o muy alto), y el tipo de lugar de residencia.

Por su parte Lindstrom (2007) realizó un modelo logit multinomial con variable respuesta el estatus de la migración del hombre y mujer, siendo estos que ambos estén en México, solo la mujer en México, solo el hombre en México o ambos en Estados Unidos y como variables explicativas la edad de ambos, ocupación, nivel educativo, legalidad de la migración, años de escolaridad y datos de la comunidad como variables tiempo-dependientes, mientras que como variable constante usó los años de migración previa.

Algunos autores consideran apropiado el estudio de las parejas en lugar de hombres y mujeres por separado porque las parejas deciden juntos sobre la migración y la fecundidad (Lindstrom, 2007). Por lo tanto, el objetivo de este trabajo es realizar una modelación longitudinal de la migración de parejas mexicanas a Estados Unidos y su fecundidad.

Se hace uso de los datos del Proyecto de Migración de México (MMP), que lleva a cabo encuestas en las comunidades de México. A través de la reconstrucción de la historia de vida de los jefes de hogar y sus cónyuges, se define la muestra de años-pareja con un período de estudio entre 1964 a 2013. MMP es una iniciativa de colaboración entre investigadores de la Universidad de Princeton y la Universidad de Guadalajara, con el apoyo financiero del Instituto Nacional de Salud Infantil y Desarrollo Humano de los Estados Unidos (NICHD). Los detalles sobre el muestreo y los cuestionarios están disponibles en el sitio web de MMP (<http://mmp.opr.princeton.edu/>).

Para analizar el perfil de las parejas de migrantes mexicanas, se identificaron las característi-

cas individuales de cada hombre y mujer en cada pareja como: la edad, años de educación, experiencia laboral acumulada (meses) y ocupación. Además, como característica de la pareja, el número de hijos. También se consideraron como datos de contexto: prevalencia migratoria e índice de capacidad física masculina y femenina², la tasa de fecundidad global³, la inflación de los precios al consumidor⁴ y el crecimiento per cápita⁵. De esta forma, se cuenta para el análisis con 14 variables cuantitativas tiempo-dependientes y una variable cualitativa tiempo-dependiente.

Por el uso de datos longitudinales y el estudio a nivel pareja, este trabajo de grado presenta varios retos, como por ejemplo, el tiempo de inicio de la pareja, duración de la relación e interrelación de las variables respuestas, que se explica a detalle en la sección de metodología y datos.

Así, este trabajo de grado se compone de seis capítulos, siendo el primero de ellos la presente introducción. Seguido por el marco teórico en el cual se ilustra el uso de datos longitudinales, los modelos lineales generalizados y las ecuaciones de estimación generalizada. Luego, se exponen las fuentes de información, recolección de datos, tipo de variables y análisis descriptivos. En el cuarto capítulo se presentan las aplicaciones de los modelos longitudinales. Como quinto, se escriben las conclusiones y finalmente se señala el trabajo futuro y bibliografía.

²Cantidad de propiedades, como número de casas, apartamentos o lotes.

³Suma de las tasas específicas de fecundidad por grupos quinquenales de edades de las mujeres entre 15 y 49 años, multiplicado por 5.

⁴Es una medida del cambio (variación), en el precio de bienes y servicios representativos del consumo de los hogares del país conocido como canasta.

⁵Indicador económico que mide la relación existente entre el nivel de renta de un país y su población.

2. Marco teórico

Los datos longitudinales se obtienen cuando repetidamente se toman mediciones de unas características de interés a un conjunto de unidades o sujetos, en un estudio (Diggle, 2002; Karlsson, 2007; Weiss, 2005). Estos datos pueden ser categóricos, discretos o continuos. Además de la respuesta, las covariables que son observadas en los sujetos (en este caso, parejas) pueden o no depender del tiempo (Zimmerman, 2009).

En los estudios longitudinales, los sujetos o individuos son independientes, pero los datos dentro del mismo individuo se encuentran correlacionados. De esta forma, aunque los objetivos inferenciales de un análisis de datos longitudinales se centran en describir cómo la respuesta se encuentra relacionada en el tiempo con otras covariables (incluyendo el mismo tiempo), es necesario incorporar un modelo apropiado y estimar con precisión la estructura de las correlaciones existentes. Por esto, el análisis de datos longitudinales es más complejo en comparación con el análisis de datos de corte transversal (Bossio, 2017).

Esta sección se centra en explicar el tipo de datos que se usaron y los aspectos teóricos de los modelos estadísticos usados para analizarlos.

2.1. Notación

Suponga que se tiene n sujetos en el estudio longitudinal indexados por $i = 1, \dots, n$. Si todos los sujetos tienen datos completos, y son medidos en las mismas ocasiones, se tiene $j = 1, \dots, t$, observaciones para cada individuo; si el número de mediciones es diferente en cada sujeto, $j = 1, \dots, t_i$.

Debido a que cada una de las parejas entran al estudio en el momento de la unión y se mantienen mientras estén juntos como pareja, se tiene que el ingreso al estudio y la duración de observaciones es diferente para cada sujeto (pareja).

El número total de observaciones es $\mathbf{T} = \sum_{i=1}^n t_i$, las respuestas de los sujetos vienen dadas por $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{it_i})^\top$ y el vector de tiempos de medida para el sujeto i es $t_i = (t_{i1}, t_{i2}, \dots, t_{it_i})^\top$. Los datos son balanceados si los tiempos de medición para todos los

sujetos es el mismo, sin datos perdidos (Bossio, 2017; Gibbons, 2006). En nuestro caso, los tiempos de medición no son los mismos pero los sujetos de estudio no tienen datos perdidos.

2.2. Modelos lineales generalizados

Los modelos lineales generalizados (MLG) son una amplia colección de modelos de regresión e incluyen casos especiales, como por ejemplo la regresión lineal estándar, los modelos de análisis de varianza para una respuesta continua normalmente distribuida, modelos de regresión logística para una respuesta binaria o dicotómica y modelos de regresión log-lineal o regresión de Poisson.

Un MLG consta de tres componentes (Fox, 2016):

1. Un *componente aleatorio*, que especifica la distribución condicional de la variable respuesta, \mathbf{Y}_i , dados los valores de las variables explicativas en el modelo. En la formulación original de Nelder (1972), la distribución de \mathbf{Y}_i es un miembro de una familia exponencial, como las familias de distribuciones gaussiana (normal), binomial, Poisson, gamma o gaussiana inversa. Sin embargo, los MLG se han extendido a familias exponenciales multivariadas (como la distribución multinomial).

La familia exponencial de un parámetro se da si existe una familia de distribuciones de un modelo $P = \{P_\theta, \theta \in \Theta\}$ con:

- $c(\theta)$ y $d(\theta)$ funciones de valor real sobre Θ ,
- $T(x)$ y $S(x)$ funciones de valor real sobre \mathbb{R}^n y
- un conjunto $A \subset \mathbb{R}^n$ que no depende de θ .

Un caso especial es la distribución binomial. Sea X una variable aleatoria con distribución $B(n, \theta)$, $0 < \theta < 1$, luego

$$\begin{aligned} p(x; \theta) &= \binom{n}{x} \theta^x (1 - \theta)^{n-x} I_A(x) \\ &= \binom{n}{x} \left(\frac{\theta}{1 - \theta} \right)^x (1 - \theta)^n I_A(x) \\ &= \exp \left[\ln \left\{ \binom{n}{x} \right\} + x \ln \left(\frac{\theta}{1 - \theta} \right) + n \ln(1 - \theta) \right] I_A(x). \end{aligned}$$

Otro caso especial, es la distribución Poisson, en que:

$$\begin{aligned}
 p(x; \theta) &= \frac{e^{-\theta} \theta^x}{x!} I_A(x) \\
 &= \exp(-\theta + x \ln(\theta) - \ln(x!)) I_A(x).
 \end{aligned}$$

2. Un conjunto de parámetros, $\boldsymbol{\beta}$, y de variables explicativas \mathbf{X} :

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1^\top \\ \mathbf{X}_2^\top \\ \vdots \\ \mathbf{X}_n^\top \end{pmatrix} = \begin{pmatrix} X_{10} & X_{11} & X_{12} & \cdots & X_{1p} \\ X_{20} & X_{21} & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{n0} & X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}. \quad (2-1)$$

3. Una *función de enlace*, expresada como una función del valor de la media, $\mu_i = E(\mathbf{Y}_i)$, en el predictor lineal:

$$g(\mu_i) = \eta_i = \mathbf{X}_i^\top \boldsymbol{\beta}. \quad (2-2)$$

Si la función de enlace es invertible, μ_i se puede escribir como:

$$\mu_i = g^{-1}(\eta_i).$$

Las funciones de enlace comúnmente empleadas y sus funciones inversas se muestran en la tabla **2-1**.

Tabla 2-1.: Funciones de enlace comunes

Algunas funciones de enlace comunes y sus funciones inversas		
<i>Link</i>	$\eta_i = g(\mu_i)$	$\mu_i = g^{-1}(\eta_i)$
Identidad	μ_i	η_i
Log	$\ln(\mu_i)$	e^{η_i}
Inversa	μ_i^{-1}	η_i^{-1}
Cuadrada-inversa	μ_i^{-2}	$\eta_i^{-1/2}$
Raíz cuadrada	$\sqrt{\mu_i}$	η_i^2
Logit	$\ln\left(\frac{\mu_i}{1-\mu_i}\right)$	$\frac{e^{\eta_i}}{1+e^{\eta_i}}$

En este trabajo se usan las funciones de distribución binomial y Poisson; por esto, las funciones enlace usadas, son logit y log, respectivamente.

2.3. Ecuaciones de estimación generalizada (EEG)

Las EEG producen estimaciones de los parámetros del modelo, junto con los errores estándar, sin especificar una función de probabilidad en su totalidad. Se debe tener en cuenta la correlación entre las respuestas múltiples que surgen de un solo sujeto, pero se puede estimar en gran medida estas correlaciones a partir de los datos sin tener que especificar con precisión la forma de la estructura de correlación.

Debido a que las GEE son una ampliación de los MLG, siguen tres especificaciones (Fitzmaurice, 2011):

- La esperanza condicional de cada respuesta, $E(\mathbf{Y}_{ij}|\mathbf{X}_{ij}) = \mu_{ij}$, que depende de las covariables a través de la función de enlace $g(\mu_{ij}) = \eta_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta}$, donde \mathbf{X}_{ij} viene dado por:

$$\mathbf{X}_{ij} = \begin{pmatrix} X_{ij0} \\ X_{ij1} \\ \vdots \\ X_{ijk} \\ \vdots \\ X_{ijp} \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} X_{i10} & X_{i11} & X_{i12} & \cdots & X_{i1k} & \cdots & X_{i1p} \\ X_{i20} & X_{i21} & X_{i21} & \cdots & X_{i1k} & \cdots & X_{i2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{ij0} & X_{ij1} & X_{ij1} & \cdots & X_{ijk} & \cdots & X_{ijp} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{it_0} & X_{it_1} & X_{it_2} & \cdots & X_{it_k} & \cdots & X_{it_p} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{i1}^\top \\ \mathbf{X}_{i2}^\top \\ \vdots \\ \mathbf{X}_{ij}^\top \\ \vdots \\ \mathbf{X}_{it_i}^\top \end{pmatrix}. \quad (2-3)$$

- La varianza de cada \mathbf{Y}_{ij} :

$$\mathbb{V}(\mathbf{Y}_{ij}) = \phi^{-1} \mathbb{V}(\mu_{ij}) = \sigma^2 \mathbb{V}(\mu_{ij}).$$

donde la función de varianza $\mathbb{V}(\mu_{ij})$ es una función conocida de la media, μ_{ij} , ϕ es un parámetro de escala que puede conocerse o que puede ser necesario estimar.

- La correlación entre mediciones repetidas que es una función de las medias, μ_{ij} , y un conjunto de parámetros, $\boldsymbol{\alpha}$, que caracterizan la correlación dentro del sujeto y deben estimarse. La matriz de covarianza viene dada por:

$$\boldsymbol{\Omega}_i = \sigma^2 \mathbb{V}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbb{V}_i^{1/2}, \quad (2-4)$$

donde $\mathbf{R}_i(\boldsymbol{\alpha})$ es la matriz de correlación de “trabajo” y \mathbb{V}_i es una matriz diagonal de tamaño $t_i \times t_i$ con $\mathbb{V}(\mu_{ij})$ en la diagonal.

2.3.1. Estructuras de correlación

Para modelar la asociación entre la respuesta de un mismo individuo en dos instantes diferentes se utiliza una matriz de correlación de “trabajo”, $\mathbf{R}_i(\boldsymbol{\alpha})$, la cual está totalmente especificada excepto por el vector de parámetros desconocidos $\boldsymbol{\alpha}$. La matriz $\mathbf{R}(\boldsymbol{\alpha})$ busca definir

una estructura de correlación que sea aproximadamente la que siguen los datos (Zeger, 1986).

De esta forma, con GEE se analizan las relaciones entre algunas estructuras de correlación del modelo en diferentes puntos de tiempo simultáneamente (Twisk, 2013). Algunas estructuras de correlación son:

Correlación “Independence” donde cada observación de cada sujeto no se correlaciona con otra observación.

$$\mathbf{R}_i = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (2-5)$$

Correlación “Exchangeable”, la cual supone que las respuestas están igualmente correlacionadas dentro de un individuo. $Corr(y_{ij}, y_{ij'}) = \alpha$.

$$\mathbf{R}_i = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{pmatrix}. \quad (2-6)$$

Correlación autoregresiva, donde dos observaciones más cercanas en el tiempo están más correlacionadas que dos más alejadas en el tiempo. $Corr(y_{ij}, y_{ij'}) = \alpha^{|j-j'|}$.

$$\mathbf{R}_i = \begin{pmatrix} 1 & \alpha & \cdots & \alpha^{t_i-1} \\ \alpha & 1 & \cdots & \alpha^{t_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{t_i-1} & \alpha^{t_i-2} & \cdots & 1 \end{pmatrix}. \quad (2-7)$$

En este caso se puede estimar $\hat{\alpha}$ con los residuos de Pearson (Hardin, 2013):

$$\hat{\alpha} = \frac{1}{\hat{\theta}} \left[\sum_{i=1}^n \left(\frac{\sum_{t=1}^{t_i} \hat{r}_{i,t} \hat{r}_{i,t}}{t_i}, \dots, \frac{\sum_{t=1}^{t_i-k} \hat{r}_{i,t} \hat{r}_{i,t+k}}{t_i} \right) \right].$$

donde $\hat{r}_{i,t} = \frac{y_{it} - \hat{\mu}_{it}}{\sqrt{V(\hat{\mu}_{it})}}$.

Correlación no estructurada cuando no se hace ninguna suposición sobre la estructura de correlación entre dos observaciones dentro de un mismo sujeto. Se especifica como:

$$\mathbf{R}_i = \begin{pmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1t_i} \\ \alpha_{21} & 1 & \cdots & \alpha_{2t_i} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{t_i1} & \alpha_{t_i2} & \cdots & 1 \end{pmatrix}. \quad (2-8)$$

No estacionaria M dependiente ($M = 1$), dada por:

$$\mathbf{R}_i = \begin{pmatrix} 1 & \alpha_1 & 0 & \cdots & 0 \\ \alpha_1 & 1 & \alpha_2 & \cdots & 0 \\ 0 & \alpha_2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \alpha_i \\ 0 & 0 & \cdots & \alpha_i & 1 \end{pmatrix}. \quad (2-9)$$

donde los α_i se pueden estimar usando

$$\hat{\alpha}_i = \sum_{j=1}^n \frac{\hat{r}_{ij}\hat{r}_{ij+1}}{n - (p + 1)}.$$

2.3.2. Estimación de parámetros

Para estimar β se considera la ecuación de estimación general (Zeger, 1986):

$$\sum_{i=1}^n \mathbf{D}_i^\top \boldsymbol{\Omega}_i^{-1} \mathbf{S}_i = 0 \quad (2-10)$$

donde $\mathbf{S}_i = \mathbf{y}_i - \hat{\boldsymbol{\mu}}_i$ con $\hat{\boldsymbol{\mu}}_i = (\hat{\mu}_{i1}, \dots, \hat{\mu}_{it_i})^\top$, $\boldsymbol{\Omega}_i = \mathbb{V}(\mathbf{Y}_i)$ ver (2-4), $\mathbf{D}_i = \mathbf{H}_i \mathbf{X}_i$ y $D_{ijk} = \partial \mu_{ij} / \partial \beta_k$.

Si $H_{ij} = \partial \mu_{ij} / \partial \eta_{ij}$ y $\mathbf{H}_i = \text{diag}\{H_{i1}, H_{i2}, \dots, H_{it_i}\}$ entonces $D_{ijk} = H_{ij} X_{ijk}$, y

$$\mathbf{D}_i = \begin{pmatrix} H_{i1} X_{i11} & H_{i1} X_{i12} & \cdots & H_{i1} X_{i1p} \\ H_{i2} X_{i21} & H_{i2} X_{i22} & \cdots & H_{i2} X_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ H_{it_i} X_{it_i1} & H_{it_i} X_{it_i2} & \cdots & H_{it_i} X_{it_i p} \end{pmatrix} = \begin{pmatrix} H_{i1} & 0 & \cdots & 0 \\ 0 & H_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_{it_i} \end{pmatrix} \mathbf{X}_i, \quad (2-11)$$

Por lo tanto, la ecuación (2-10) se puede expresar como:

$$\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{H}_i \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) = 0,$$

Además, $\boldsymbol{\Omega}_i$ se especificó en la ecuación (2-4), por esto $\boldsymbol{\Omega}_i \approx \mathbb{V}(\mathbf{Y}_i)$.

Así, la función de estimación de la ecuación (2-10) es dada por:

$$\psi_n(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{D}_i^\top \boldsymbol{\Omega}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{H}_i \boldsymbol{\Omega}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})). \quad (2-12)$$

Hardin (2013) sugieren la estimación de $\boldsymbol{\beta}$ como:

$$\hat{\boldsymbol{\beta}}^{(r)} = \hat{\boldsymbol{\beta}}^{(r-1)} - \left\{ \sum_{i=1}^n \mathbf{D}_i^\top \mathbb{V}(\boldsymbol{\mu}_i)^{-1} \mathbf{D}_i \right\} \left\{ \sum_{i=1}^n \mathbf{D}_i^\top \mathbb{V}(\boldsymbol{\mu}_i)^{-1} \mathbf{S}_i \right\},$$

con $\mathbf{D}_i = \mathbf{H}_i \mathbf{X}_i$ y $\mathbf{S}_i = \mathbf{y}_i - g^{-1}(\hat{\eta}_i)$.

De igual forma, Zeger (1986) usan una variación del algoritmo Fisher-Scoring:

$$\boldsymbol{\beta}^{m+1} = \hat{\boldsymbol{\beta}}^m + (\mathbf{D}^\top(m) \boldsymbol{\Omega}^{-1(m)} \mathbf{D}^{(m)})^{-1} \mathbf{D}^\top(m) \boldsymbol{\Omega}^{-1(m)} (\mathbf{Y} - \boldsymbol{\mu}^{(m)}) \quad (2-13)$$

con:

$$\begin{aligned} \mathbf{D}^{(m)} &= (\mathbf{D}_1^\top(m), \dots, \mathbf{D}_n^\top(m))^\top \\ \boldsymbol{\Omega}^{(m)} &= \text{diag}\{\boldsymbol{\Omega}_i^{(m)}\} \\ \mathbf{Y} &= (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_n^\top)^\top \\ \boldsymbol{\mu} &= (\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_n^\top)^\top \end{aligned}$$

si

$$\mathbf{H}_i^{(m)} = \mathbf{H}_i|_{\{\eta_{ij} = \eta_{ij}^{(m)}\}}$$

$$\mathbf{H}_i = \text{diag}\{\mathbf{H}_i^{(m)}\} = \begin{pmatrix} \mathbf{H}_1^{(m)} & 0 & \dots & 0 \\ 0 & \mathbf{H}_2^{(m)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{H}_n^{(m)} \end{pmatrix}.$$

Debido a que $\mathbf{D}^{(m)} = \mathbf{H}^{(m)}\mathbf{X}$, (2-13) se puede escribir como

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathcal{A}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathcal{A}^{(m)} \mathbf{Z}^{(m)}, \quad (2-14)$$

donde

$$\mathcal{A}^{(m)} = \mathbf{H}^{(m)} \boldsymbol{\Omega}^{-1(m)} \mathbf{H}^{(m)} \quad y \quad \mathbf{Z}^{(m)} = \boldsymbol{\eta}^{(m)} + [\mathbf{H}^{(m)}]^{-1} (\mathbf{Y} - \boldsymbol{\mu}^{(m)}).$$

2.4. Métodos de diagnóstico EEG

Se presentan medidas para evaluar la bondad del ajuste del modelo, seleccionar una estructura de correlación, elegir una adecuada colección de covariables y, si es el caso, detectar observaciones de apalancamiento o influyentes.

2.4.1. Selección de una estructura de correlación

El **criterio de información de Akaike (CIA)** es una estadística de bondad de ajuste establecida para la comparación de modelos basada en la probabilidad. Hardin y Hilbe (2013) introdujeron dos extensiones de esta medida que se ilustran a continuación.

Para seleccionar entre diferentes estructuras de correlación se puede usar el QIC dado por Pan (2001):

$$QIC(\mathbf{R}) = -2\mathcal{Q}(y; \beta_R) + 2traza(\mathbf{A}\mathbf{B}), \quad (2-15)$$

con

$$\mathbf{A} = \left[\sum_{i=1}^n [\hat{\mathbf{D}}_i^\top \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i] \right]^{-1} \left[\sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\boldsymbol{\Omega}}_i^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i) (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)^\top \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i \right] \left[\sum_{i=1}^n [\hat{\mathbf{D}}_i^\top \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i] \right]^{-1}. \quad (2-16)$$

y

$$\mathbf{B} = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\mathbf{V}}_i^{-1} \mathbf{D}_i. \quad (2-17)$$

Además, $\mathcal{Q}(y; \beta_R)$ es el valor de la cuasiverosimilitud calculado usando los coeficientes del modelo con la estructura de correlación $\mathbf{R}(\boldsymbol{\alpha})$.

Dado que la ecuación (2-15) esta en términos de la estructura de correlación $\mathbf{R}(\boldsymbol{\alpha})$, se puede usar esta medida para elegir entre varias estructuras de correlación en “competencia”. Así

como en el valor de AIC, el mejor modelo es el de menor medida (Hardin, 2013).

Otra alternativa es el uso del “correlation information criteria” (CIC) el cual puede ser decisivo al momento de escoger la mejor estructura (Pan, 2001).

$$CIC = \text{traza}(\mathbf{AB}). \quad (2-18)$$

Además, de forma similar como se selecciona la mejor estructura de correlación, se puede seleccionar el mejor conjunto de covariables usando

$$QIC_u = -2Q(y; \beta_R) + 2(p + 1). \quad (2-19)$$

Para la selección se iteran con las combinaciones de variables posibles y se selecciona el QIC_u con menor valor. Adicional, en el anexo A se incluye el código creado para su evaluación en bucle.

2.4.2. Análisis de residuales

Examinar los residuos es un paso crucial en el análisis estadístico para identificar las discrepancias entre los modelos y los datos, y evaluar las bondades de ajuste general del modelo. A menudo se utilizan residuos desvío y de Pearson, que tienen una distribución aproximadamente normal estándar cuando el modelo se ajusta a los datos adecuadamente. Sin embargo, cuando la variable respuesta es discreta, estos residuos se distribuyen lejos de la normalidad y tienen curvas casi paralelas de acuerdo con los distintos valores de respuesta discreta (Li, 1998).

Un análisis de resultados incluye una verificación importante donde se debe observar que el modelo seleccionado se ajuste a los datos de forma adecuada. De acuerdo a Gonzalez (2022), se tiene el siguiente análisis de residuales, teniendo en cuenta que β^{m+1} , dada en la ecuación (2-14), utiliza una descomposición de valores propios de $\mathcal{A}^{(m)}$ para obtener los residuales no correlacionados y con varianza constante.

$$\mathcal{A} = \mathbf{L}^*(\mathbf{L}^*) = (\mathbf{L}^*)^2. \quad (2-20)$$

De esta forma, $\hat{\beta}$ se interpreta como solución de mínimos cuadrados de la regresión $\mathbf{L}^* \hat{\mathbf{Z}}$ sobre $\mathbf{L}^* \hat{\mathbf{X}}$, así el residuo ordinario según Botter (2005) es:

$$\mathbf{r}^* = \mathbf{L}^* \hat{\mathbf{Z}} - \hat{\boldsymbol{\eta}}.$$

Con $\hat{\mathbf{Z}}$ dado en (2-14), se tiene el residuo ordinario:

$$\mathbf{r}^* = \mathbf{L}^*[\hat{\mathbf{H}}]^{-1}(\mathbf{Y} - \hat{\boldsymbol{\mu}}). \quad (2-21)$$

Además, de acuerdo con Venezuela (2007), el residuo estandarizado esta dado por:

$$(r_{SD})_{ij} = \frac{\mathbf{e}_{ij}^\top \mathbf{L}_i^* \hat{\mathbf{H}}_i^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)}{\sqrt{1 - \Delta_{ij}}}. \quad (2-22)$$

donde \mathbf{e}_j es un vector de tamaño t con 1 en la posición asociada a j y 0 en otro caso, y Δ_{ij} el j -ésimo elemento de la diagonal de $\boldsymbol{\Delta}_i$, con $\boldsymbol{\Delta}_i = \mathbf{L}_i^* \mathbf{X}_i (\mathbf{X}_i^\top \hat{\mathcal{A}} \mathbf{X}_i)^{-1} \mathbf{X}_i^\top \mathbf{L}_i^*$.

Mediante el uso de los residuos estandarizados se puede detectar **puntos de palanca (leverage points)**, **puntos influyentes (influential points)** y **valores atípicos (outliers)**. A continuación se hará una breve descripción de cada uno de estos puntos.

- **Puntos palanca.** Un punto palanca, aparece cuando se tiene un perfil diferente a los demás en cuanto a las variables explicativas continuas. Se puede investigar la existencia de una observación (punto) palanca si

$$\Delta_{ij} \geq \frac{2\text{Traza}(\boldsymbol{\Delta}_i)}{T} = \frac{2(p+1)}{T}. \quad (2-23)$$

Con $T = \sum_{i=1}^n t_i$ representando el total de observaciones. Es posible que una unidad experimental sea de palanca si

$$\Delta_{i*} \geq \frac{2(p+1)}{T}$$

Los valores de Δ_{ij} dependen de la matriz de pesos $\hat{\mathcal{A}}$, ver (2-20) observaciones con valores altos de Δ_{ij} no son necesariamente puntos de palanca.

- **Puntos influyentes.** Un punto influyente es un punto que tiene un perfil diferente de las demás observaciones ya que la variable de respuesta tiene un valor alto en Δ_{ij} . Un punto influyente tiene demasiada relevancia o valor en la estimación de los parámetros del modelo. Sea:

$$\begin{aligned} (DC)_{ij} &= \frac{1}{p+1} \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(ij)} \right)^\top \mathbf{X}^\top \hat{\mathcal{A}} \mathbf{X} \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(ij)} \right) \\ &= (r_{SD})_{ij}^2 \frac{\Delta_{ij}}{(p+1)(1-\Delta_{ij})}. \end{aligned} \quad (2-24)$$

$(DC)_{ij}$ será un punto influyente si su valor sobresale de las demás observaciones.

- **Puntos atípicos(outliers).** Un outlier es una observación cuyo perfil de respuesta es diferente a las demás y tiene un valor bajo en Δ_{ij} , es decir, que tiene bajo nivel de palanca, pero el valor de residuo estandarizado es alto $(r_{SD})_{ij}$, ver ecuación (2-22).

Para la detección de estos puntos, se pueden tener en cuenta los DFBETAS y la distancia de Cook (ver ecuación (2-24)). Las medidas diagnósticas DFBETA evalúan la diferencia entre el estimador de muestra completa y el vector de coeficientes ajustado basado en la eliminación de observaciones, ver Hardin (2013).

2.5. Datos desbalanceados

Existen diversos patrones en los datos que no son usuales. En caso de desigualdad en la distribución entre las clases de los datos, estos puede considerarse desbalanceados. Esto se produce cuando hay más muestras en una clase que de otra clase. En un conjunto de datos desbalanceado, la clase mayoritaria tiene un gran porcentaje de todas las muestras, mientras que las instancias de la clase minoritaria sólo ocupan una pequeña parte de las observaciones (Maalouf, 2015). Por otra parte, también puede ser un diseño desbalanceado con relación al tiempo, esto se puede dar debido a que los datos se recolectan en tiempos diferentes o a pérdida de información.

En este caso, se tienen datos donde su mayor población no es migrante y tienen 0 hijos. Sin embargo, también se presenta el reto de entrada y salida de parejas, ya que las parejas entran en el estudio en el momento de la unión y sale al momento de muerte de alguno, cuando la mujer tiene más de 49 años o finaliza la unión. Es decir, la pareja 1 puede entrar en 1990 y salir en 2010 pero la pareja 2 puede entrar en 2001 y terminar en 2008.

De esta forma, no se presentan datos con faltantes ni censurados, ya que la encuesta realiza un seguimiento a la población Mexicana generando continuidad de información. Sin embargo, algunas parejas presentan censuras a la derecha, ya que se tiene información hasta 2013 debido a que en ese momento se realizaron las últimas encuestas completas.

3. Fuentes de información y análisis descriptivo

3.1. Datos

El Proyecto Sobre Migración Mexicana (MMP) fue creado en 1982 con el fin de recabar información social, económica y demográfica sobre el proceso migratorio entre México y los Estados Unidos. Este proyecto fue creado por investigadores de México y Estados Unidos (MMP, 2022a).

El proyecto utilizó la encuesta etnográfica, que combina las técnicas del trabajo de campo antropológico y los métodos de muestreo por encuesta. De esta forma, logra un estudio exhaustivo de las comunidades migratorias. Además, recaba información de los migrantes en ambos lados de la frontera, lo que permite un entendimiento más amplio de los cambios migratorios (MMP, 2022a).

MMP realiza encuestas en viviendas seleccionadas de manera aleatoria en comunidades localizadas por todo México. Recaban información demográfica, social y económica sobre los miembros de cada vivienda. Luego, la encuesta se enfoca en recabar información básica sobre el primer y el último viaje a los Estados Unidos de cada miembro de la familia. Del jefe de familia se compila una historia migratoria anual, enfocándose en información sobre empleo, ingresos y uso de servicios sociales en los Estados Unidos (MMP, 2022a).

Como el MMP no se enfoca en cierta población, el factor expansión no varía según los hogares o personas, sino que son específicos por cada comunidad según los distintos grupos que se encuesten (MMP, 2012).

Los datos recopilados contienen información de 174 comunidades de México como se observa en la figura **3-1**:



Figura 3-1.: Mapa de las comunidades del MMP encuestadas. (MMP, 2022b)

Aunque las encuestas las realizan desde 1982, se tendrán en cuenta las encuestas realizadas a partir de 1998. Debido a que el análisis se centró en las parejas, considerándose las uniones conyugales “estables” (matrimonio, unión consensual) como la principal variable que explica la fecundidad (Bongaarts, 1978), se tomaron aquellas parejas heterosexuales en las que las mujeres (cabeza de familia o cónyuge) estaban en edad de procrear (14 a 49 años). Además, solo se consideró a los jefes de hogar mayores de 15 años.

La muestra contiene datos desde el año de unión hasta el año en que finaliza la encuesta, cuando la mujer cumple 49 años o un año antes del segundo viaje migratorio. Se reconstruye la historia de vida del jefe de familia y su cónyuge, y la historia de los eventos de la pareja. Por lo tanto, el período de estudio se limita a los años de 1964 a 2013.

Se calculó la razón de prevalencia de la migración (RPM) anual definida como la proporción de personas mayores de 15 años en un año determinado (t) que han realizado al menos un viaje internacional respecto del total de personas vivas mayores a 15 años (Massey, 1994).

$$RPM_{comunidad_t} = \frac{\#_Personas_ > 15_anos_con_experiencia_migratoria_{comunidad_t}}{Total_personas_vivas_ > 15_anos_{comunidad_t}}. \quad (3-1)$$

Debido a que el análisis se centra en las parejas, tanto la migración internacional de uno de los miembros de la pareja como el nacimiento de un hijo tienen un impacto directo en

la estructura y decisiones del hogar. Por tanto, las principales variables consideradas para analizar la asociación entre migración y fecundidad son el número de hijos y la experiencia migratoria internacional de la pareja.

Se consideraron las características de la pareja, como el número de hijos por hogar, la edad, el total de experiencia laboral, el total de años de educación y ocupación de cada miembro de la pareja así como el índice de capital físico (que incluye la posesión de bienes y negocios), la experiencia migratoria de la pareja y el país de origen. Entre las variables de contexto, como indicador de la fecundidad, se tomó la TGF, es decir, el promedio de nacimientos de mujeres entre 14 y 49 años.

$$TGF_{comunidad_t} = \frac{Nacimientos_{comunidad_t}}{Mujeres_{15_49}_{comunidad_t}} * 1000 \quad (3-2)$$

Dentro del contexto macroeconómico, se consideró el crecimiento del PIB per cápita en pesos y la tasa de inflación nacional a precios al consumidor en cada comunidad y año.

3.1.1. Análisis descriptivo

Se analizó la transición entre estados migratorios de las parejas, esto es, como las parejas pasan de un estado migratorio a otro, ver tabla 3-1.

Tabla 3-1.: Cambios de estado migratorio

ID	Tipo migración
OA	Origen ambos (Ambos en México)
MH	Migró hombre
MM	Migró mujer
MA	Migraron ambos

La tabla 3-2 proporciona información sobre los movimientos entre los estados migratorios de las parejas. Se distingue entre movimientos migratorios cuando la pareja estaba en el país de origen (OA), y luego migraba el hombre (MH), la mujer (MM) o ambos (MA). Otro movimiento fue cuando el hombre migró (MH) y luego regresó a México (OA). También cuando migró el hombre (MH) y luego migró la mujer, es decir, ambos al exterior (MA). Otro, cuando la mujer migró (MM) y luego volvió a México (OA). Finalmente, cuando la pareja migró (MA) y regresaron juntos a México (OA).

No se tuvo en cuenta la transición cuando el hombre migró, volvió, y ese mismo año migró la mujer porque ninguna pareja hizo esa transición. Tampoco se tuvo en cuenta cuando migró la

mujer, volvió, y ese mismo año migró el hombre, o cuando migró la pareja junta, y luego volvió solo el hombre o solo la mujer porque fueron solo 12 parejas las que tuvieron dichas transiciones. Estas parejas solo se retiraron del estudio descriptivo pero no del modelaje de datos.

Se observó que las parejas de migrantes realizan un movimiento migratorio a una edad más temprana cuando están en México y luego migran juntos. En cambio, cuando la pareja estaba en México, y entonces migraba solo la mujer, era el movimiento migratorio en el que presentaban mayor edad promedio tanto el hombre como la mujer en comparación con los otros movimientos. Strzelecka (2012) muestra que las mujeres migran más jóvenes que los hombres. De la misma forma lo evidencia la muestra, porque cuando migra solo el hombre, tienen en promedio 28 años; mientras que cuando solo migran mujeres, tienen en promedio 26 años (ver tabla **3-2**).

En cuanto a los años de escolaridad, se observó que en el año en que el hombre estuvo en el exterior y retornó, la pareja presentó bajos niveles educativos (en promedio 5 años). Por otro lado, cuando la pareja estaba en México, y luego la mujer migró, las mujeres tenían los años de educación más altos (en promedio 8 años) en comparación con los otros movimientos migratorios.

Al observar el mercado laboral, se evidenció que los hombres tienen 181 meses y las mujeres 168 meses de experiencia laboral en el año en que, dado que la mujer había migrado, regresa, lo cual es en promedio mayor a los otros movimientos. Además, en los años de ese movimiento, se encontró mayor inserción laboral de hombres y mujeres debido a que el 15 % de los hombres se encontraban empleados como profesionales, el 50 % de las mujeres en el sector servicios y solo el 15 % de ellas no se encontraban en la fuerza laboral.

El número de hijos es una variable determinante de la decisión de migrar solo, acompañado, regresar o no migrar (Madhavan, 2012; Milewski, 2010). Se reveló que cuando las parejas estaban en México y luego migraban juntas, en promedio tenían un solo hijo. Lo anterior, anidado con las características previas, puede sugerir el acercamiento de la hipótesis de selección porque las parejas mexicanas migraron juntas desde un inicio cuando eran jóvenes, con buen nivel educativo, poca experiencia laboral y pocos hijos.

La RPM muestra que a pesar de la feminización de la migración (Tapia, 2011), en todas las transiciones migratorias, los hombres tuvieron una mayor RPM que las mujeres. Sin embargo, se encontró que cuando la pareja estaba en México, y luego migró el hombre, el hombre tuvo un 1100 % más de RPM que la mujer. En cambio, cuando la pareja estaba en el origen y luego migró la mujer, el hombre solo tuvo un 285 % más de RPM que la mujer. Lo anterior implica que la participación de las mujeres reduce la brecha en la RPM.

Tabla 3-2.: Movimientos migratorios

	Total Media / %	Se mueve a					
		OA - MH Media / %	OA - MM Media / %	OA - MA Media / %	MH - OA Media / %	MH - MA Media / %	MA - OA Media / %
Edad							
Hombres	32.68	28.50	29.07	25.42	30.36	30.75	30.53
Mujeres	30.11	26.10	26.74	24.12	27.94	28.87	28.41
Años de escolaridad							
Hombres	5.09	5.61	8.07	7.68	5.54	6.17	7.23
Mujeres	5.23	5.53	8.69	7.35	5.26	6.06	6.66
Experiencia laboral (M)							
Hombres	150.16	111.24	89.89	64.17	128.20	130.35	119.05
Mujeres	139.42	93.36	86.63	55.09	122.63	124.90	107.97
Hombres ocupación							
Desempleo	0.01 %	0.1 %	0 %	0 %	0 %	2.3 %	0.4 %
No está en la fuerza laboral	0.23 %	0.1 %	2.9 %	0.8 %	0 %	0 %	0 %
Agricultura	85.6 %	96.2 %	7.6 %	20.6 %	62.4 %	40.2 %	26.7 %
Profesional o administrativo	4 %	0.4 %	17.1 %	6.6 %	0.1 %	2.3 %	0.4 %
Manufactura	7 %	2.4 %	48.6 %	42.9 %	22.5 %	37.9 %	57.9 %
Servicios	2.9 %	0.8 %	23.8 %	29.1 %	15 %	17.2 %	21.2 %
Técnicos	0.04 %	0 %	0 %	0 %	0 %	0 %	0 %
Mujeres ocupación							
Desempleo	0.1 %	0 %	0 %	0 %	0 %	0 %	0.4 %
No está en la fuerza laboral	89.9 %	92.7 %	40 %	76.2 %	92.5 %	79.3 %	40.8 %
Agricultura	6.24 %	6.4 %	2.9 %	0.4 %	0.1 %	1.1 %	8.7 %
Profesional o administrativo	1.55 %	0.2 %	35.2 %	1.8 %	0.1 %	2.3 %	0 %
Manufactura	1.67 %	0.4 %	10.5 %	7.4 %	7.1 %	16.1 %	32.1 %
Servicios	0.56 %	0.3 %	11.4 %	13.6 %	0.2 %	1.1 %	16.7 %
Técnicos	0.03 %	0 %	0 %	0.6 %	0 %	0 %	0 %
Número total de niños nacidos	3.07	2.13	1.23	1.07	2.36	2.56	1.73
Índice de capital físico	- 0.10	-0.38	-0.47	-0.70	-0.25	-0.36	-0.40
Prevalencia migratoria (RPM)							
Hombres	0.11	0.11	0.20	0.23	0.12	0.21	0.24
Mujeres	0.01	0.01	0.07	0.09	0.01	0.07	0.09
Datos de contexto							
TGF	3.38	3.43	3.23	3.94	3.35	3.44	3.36
Inflación	25.09 %	22.70 %	21.98 %	32.76 %	22.52 %	42.49 %	22.63 %
Tasa de crec per cápita	0.97	1.59	2.56	1.52	2.61	1.42	1.73
N		159,402	315	1,497	147,132	261	180

Fuente: MMP. Cálculos propios

Al evidenciar el índice de capital físico, se encuentra una menor posesión de bienes en el año que la pareja se encuentra en México y decide migrar(-0.7) seguido por el año en el que la pareja se encontraba en México y la mujer migró (-0.47). Sin embargo, en el momento donde la mujer se encontraba en el exterior y decidió retornar, el índice de capital físico es mayor (-0.08).

En cuanto a los datos de contexto, tanto la TGF, la inflación y la tasa de crecimiento per capita, son datos de México a nivel comunidad, por lo que la información de la tabla **3-2** son el promedio de esta información en los años relacionados.

Por otra parte, el análisis de las categorías migratorias (ver tabla **3-3**) en los años que duró el estado de estas, se realizó con 2 categorías, los años pareja en los que estuvieron en México y los años pareja en los que alguno de los dos migró, las estadísticas descriptivas se representan en la tabla **3-4**, donde los números en paréntesis indican la desviación estándar.

Tabla 3-3.: Categorías migratorias.

Categoría	Descripción
No migrantes	Años en los que el hombre y mujer de la pareja se encontraban en México
Migrantes	Años en los que al menos una persona de la pareja se encontraba en EEUU

El promedio de la edad de las parejas del estudio es menor cuando son migrantes, los hombres migrantes tenían en promedio 31 años y las mujeres migrantes 28, mientras que los años en los que no eran migrantes los hombres tenían 33 años y las mujeres 30 años, esto debido a que deciden hacerlo en edades tempranas. A pesar de ser jóvenes, los años en los que se encuentran en EEUU presentan en promedio 6 años de escolaridad, mientras los que se encuentran en México en promedio 5 años.

En cuanto a la experiencia laboral, es mayor (150 meses en México y 129 meses en EEUU) en los años que se mantienen en México. Sin embargo, en los años donde al menos uno de la pareja migra genera más diversidad de ocupación, ya que cuando se encuentran en México la mayoría de los hombres se dedican a la agricultura pero al migrar se incorporan a sectores como manufactura y servicios. Mientras que las mujeres en México no se encuentran en la fuerza laboral, pero al migrar se incorporan en sectores de manufactura y servicios.

En términos de fecundidad se encuentran que en los años donde se mantuvo en EEUU al menos una persona de la pareja, el número de hijos es menor pasando de 3 hijos (los que se encuentran en México) a 2 hijos (los de EEUU).

Tabla 3-4.: Características de las parejas mexicanas según estado migratorio.

	No migrantes	Migrantes
Edad (prom)		
Hombres	32,68 (7,77)	30,88 (6,46)
Mujeres	30,11 (7,44)	27,94 (5,52)
Años de escolaridad		
Hombres	5,09 (2,22)	5,79 (1,78)
Mujeres	5,23 (1,69)	6,11 (1,33)
Experiencia laboral (Meses)		
Hombres	150,16 (87,23)	128,65 (59,39)
Mujeres	139,42 (87,56)	99,80 (60,47)
Hombres ocupación		
Desempleo	0,01 %	0,01 %
No está en la fuerza laboral	0,23 %	0,15 %
Agricultura	85,6 %	59,17 %
Profesional o administrativo	4,01 %	0,38 %
Manufactura	7,01 %	18,51 %
Servicios	3,1 %	21,5 9 %
Técnicos	0,04 %	0,19 %
Mujeres ocupación		
Desempleo	0,01 %	0 %
No está en la fuerza laboral	89,94 %	65,46 %
Agricultura	6,24 %	2,87 %
Profesional o administrativo	1,55 %	0,32 %
Manufactura	1,67 %	20,05 %
Servicios	0,56 %	11,21 %
Técnicos	0,03 %	0 %
Número total de niños nacidos	3,07 (1,47)	2,35 (0,74)

Índice de capital físico	-0,10 (0,09)	-0,35 (0,06)
Prevalencia migratoria		
Hombres	0,11 (0,03)	0,22 (0,08)
Mujeres	0,01 (0,001)	0,08 (0,006)
N	16,116,291	374,679

En la tabla **3-5**, se observa que el 87 % de las parejas en la muestra tienen más de 7 años juntos. Sin embargo, solo el 19 % tiene más de 25 años, sin embargo, si el estudio se centrará solo en las parejas migrantes, solo el 25 % tienen más de 25 años, es posible que los picos de disminución de hijos de las parejas migrantes se deba a salidas en la muestra.

Tabla 3-5.: Cantidad de parejas según máximo de años pareja.

Max Años pareja	% Parejas
1-3	3,78 %
4-6	9,06 %
7-13	27,37 %
14-17	13,58 %
18-25	26,58 %
26 años y más	19,64 %

De manera similar, en la figura **3-2** se encuentra que las parejas migrantes tienen en su mayoría entre 2 y 3 hijos, sin embargo, proporcionalmente los migrantes tienen menos hijos en comparación con los no migrantes.

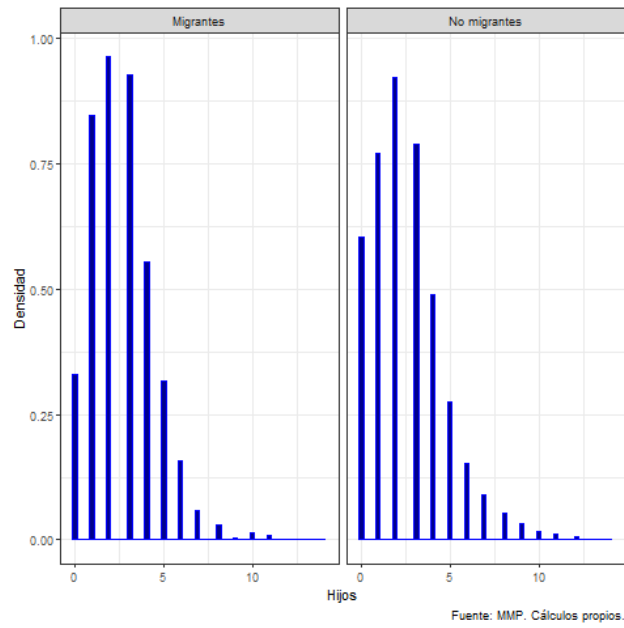


Figura 3-2.: Cantidad de hijos según estado de la pareja (migrantes o no migrantes).

4. Modelación longitudinal

En esta sección, se propone un modelo para la variable respuesta “si la pareja migró o no migró” y se propone otro modelo para la variable respuesta “cantidad de hijos”. Se tiene en cuenta que las parejas entran al estudio en el momento de unirse y salen del estudio en el momento que la mujer tenga más de 49 años, muerte de alguno o la separación de la pareja, por esto, tienen tiempos de entrada, duración y salida distintos. Como se expresó en la sección 2.5, los datos se encuentran desbalanceados, sin embargo, en este estudio se trabajan como datos balanceados y sin censuras.

4.1. Modelo propuesto para migró / no migró

Se ajustó el modelo (4-1),

$$\eta = g(\mu_{ij}) = \ln \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \gamma + \beta^{hj} Hijos_{ij} + \beta^{se} SE_i + \beta^{co} CO_i \quad (4-1)$$

con i el número de parejas y j el número de años pareja con

$$i = 1, 2, 3, \dots, 8794$$

$$j = 1, 2, 3, \dots, 36,$$

donde:

- $g(\mu_{ij})$ es la función de enlace logit.
- γ es el efecto medio en la escala logartimo.
- β^{hj} los efectos poblacionales asociados al número de hijos de la pareja i en el periodo j .
- β^{se} , los efectos poblacionales asociados a las variables socioeconómicas.
- β^{co} , los efectos poblacionales asociados a las variables de contexto.
- $Hijos$ es el número de hijos de la i –ésima pareja en el año j .

- **SE** asociada a las de variables socioeconómicas para la i – *ésima* pareja en el año j , siendo estas la edad, años de educación, experiencia laboral en meses y ocupación para el hombre y la mujer.
- **CO** asociada a las variables de contexto para la i – *ésima* pareja en el año j , tales como TGF, PIB per capita, inflación, prevalencia migratoria por género e índice de capital físico.

Se evaluó el modelo (4-1) usando diferentes estructuras de correlación y se evaluó el QIC (2-15) y el CIC (2-18), ver tabla 4-1, encontrando la más adecuada, siendo esta “AR2”, esto indica que las observaciones de cada pareja son más correlacionadas si son más cercanas.

Tabla 4-1.: Estructuras de correlación. QIC y CIC.

Estructura de correlación	QIC	CIC
Exchangeable	42.359,9	277,2
Non-Stationary-M-dependent(1)	44.582,1	235,9
Non-Stationary-M-dependent(2)	43-816,2	386,2
Stationary-M-dependent(1)	47.899,5	496,3
Independence	41.511,1	175,3
AR2	37.487,1	163,2

La matriz de correlación estimada correspondiente a la estructura de correlación seleccionada fue

$$\mathbf{R}_i = \begin{pmatrix} 1 & 0,6985 & \cdots & 0,0003 \\ 0,6985 & 1 & \cdots & 0,0004 \\ \vdots & \vdots & \ddots & \vdots \\ 0,0003 & 0,0004 & \cdots & 1 \end{pmatrix}. \quad (4-2)$$

A su vez, seleccionada la estructura de correlación, se realizó una selección de variables mediante el criterio “ QIC_u ” especificado en (2-19), el modelo seleccionado es

$$\begin{aligned} \ln \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = & \gamma + \beta_1 Edadm_{ij} + \beta_2 Edadh_{ij} + \beta_3 (Edadh_{ij} * Edadh_{ij}) + \beta_4 Premigh_{ij} \\ & + \beta_5 ExpLabh_{ij} + \beta_6 ExpLabm_{ij} + \beta_7 (ExpLabh_{ij} * Edadh_{ij}) \\ & + \beta_8 (ExpLabm_{ij} * Edadm_{ij}) + \beta_9 Ocupa_{h1ij} + \beta_{10} Ocupa_{h2ij} \\ & + \beta_{11} Ocupa_{h3ij} + \beta_{12} Ocupa_{h4ij} + \beta_{13} Ocupa_{h5ij} + \beta_{14} Ocupa_{h6ij} \end{aligned}$$

$$\begin{aligned}
& +\beta_{15}Educ_{ij} + \beta_{16}(Educ_{ij} * Edad_{ij}) + \beta_{17}Hijos_{ij} \\
& +\beta_{18}TGF_j + \beta_{19}Añospareja_{ij}.
\end{aligned}$$

en las que las variables son las de la tabla 4-2

Tabla 4-2.: Variables modelo migró / no migró

Variable	Descripción
Edadm	Edad de la mujer
Edadh	Edad del hombre
Premigh	Prevalencia migratoria del hombre
ExpLabh	Experiencia laboral en meses del hombre
ExpLabm	Experiencia laboral en meses de la mujer
Educ _m	Años de educación de la mujer
Ocupah	Tipo de ocupación del hombre
Hijos	Número de hijos por pareja
TGF	Tasa Global de Fecundidad
Añospareja	Años que lleva la pareja junta

Como la variable ocupación es categórica, se tiene la tabla 4-3

Tabla 4-3.: Categorías variable ocupación

Categoría	Descripción
Ocupah1	No en la fuerza laboral
Ocupah2	Agricultura
Ocupah3	Profesional o administrativo
Ocupah4	Manufactura
Ocupah5	Servicios
Ocupah6	Técnicos

4.1.1. Diagnóstico modelo propuesto migró / No migró

La matriz de varianzas de los parámetros ajustados “naive” para el modelo propuesto es:

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_n)_{Naive} = \left[\sum_{i=1}^n \left(\hat{\mathbf{D}}_i^\top \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i \right) \right]^{-1} = \begin{pmatrix} 1,58e^{-6} & \dots & \dots & \dots \\ -0,00 & 4,16 & \dots & \dots \\ -3,56e^{-7} & 1,34e^{-4} & \dots & \dots \\ \vdots & \ddots & \ddots & \dots \\ -0,00 & -1,26 & 4,18e^{-4} & \dots & 0,87 \end{pmatrix}.$$

Mediante este se encuentra \mathbf{A} y \mathbf{B} especificado en (2-15), (2-19), (2-18), con

$$CIC(\mathbf{R}) = \text{traza}(\mathbf{AB}) = 9,61.$$

Se realizó el análisis diagnóstico del modelo, iniciando con el análisis de puntos palanca (2-23), se puede observar en la figura 4-1 que no hay altos valores de Δ_{ij} , lo que nos indica la baja posibilidad de puntos palanca.

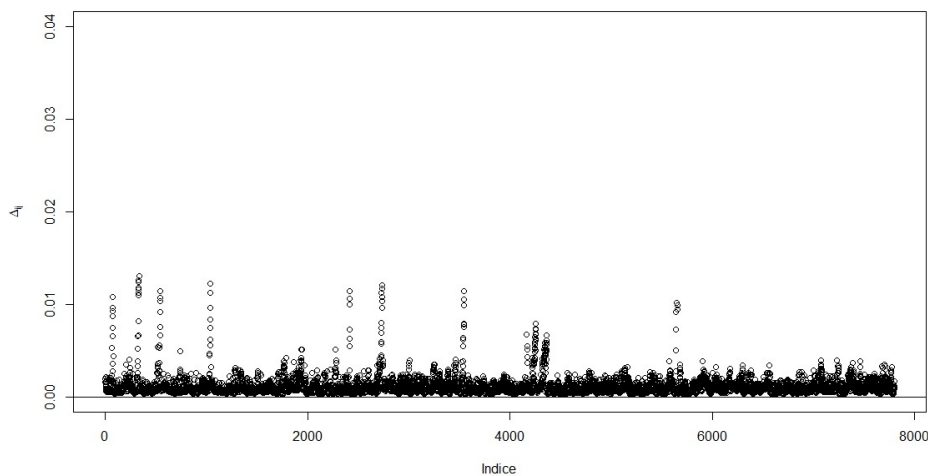


Figura 4-1.: Δ_{ij} vs las observaciones modelo migró /no migró

Al revisar los residuales ordinarios (2-21) y evaluar los residuos estandarizados (2-22) de la figura 4-2, se observan parejas que tienen un residual (r_{SD}) alto pero no un apalancamiento (Δ_{ij}).

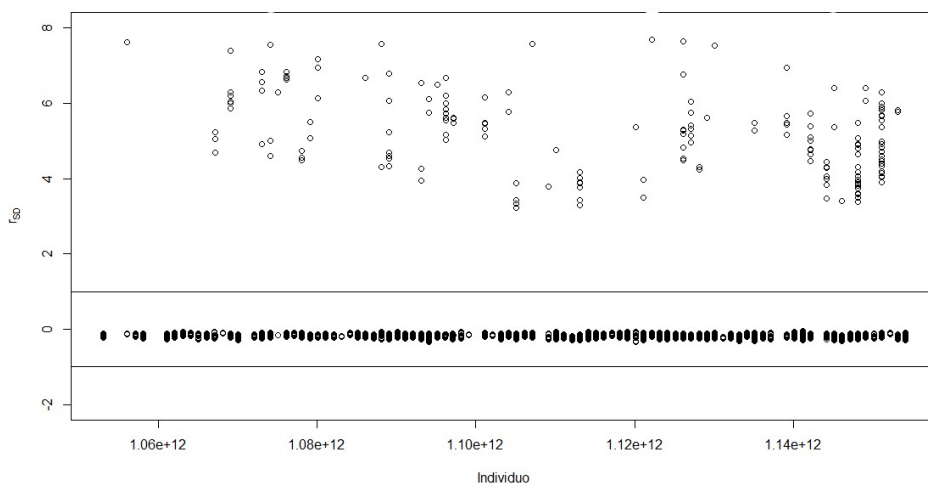


Figura 4-2.: r_{SD} vs parejas modelo migró / no migró

Para observar el efecto de quitar una observación, sobre la estimación de los parámetros en el modelo, se usa Dfbeta, los resultados se muestran en los anexos. Al evaluar las distancias de Cook $(DC)_{ij}$ vs las parejas. Se observa que hay cuatro observaciones influyentes, siendo estas parejas que en todos sus años pareja se han encontrado en el exterior, por lo que se decide retirarlas. Adicionalmente, para observar el efecto, sobre la estimación de los parámetros en el modelo, de quitar no una observación, sino un individuo, se analizó los DFbetas de cada variable que se encuentran en los anexos en estos no se encuentra algún punto influyente.

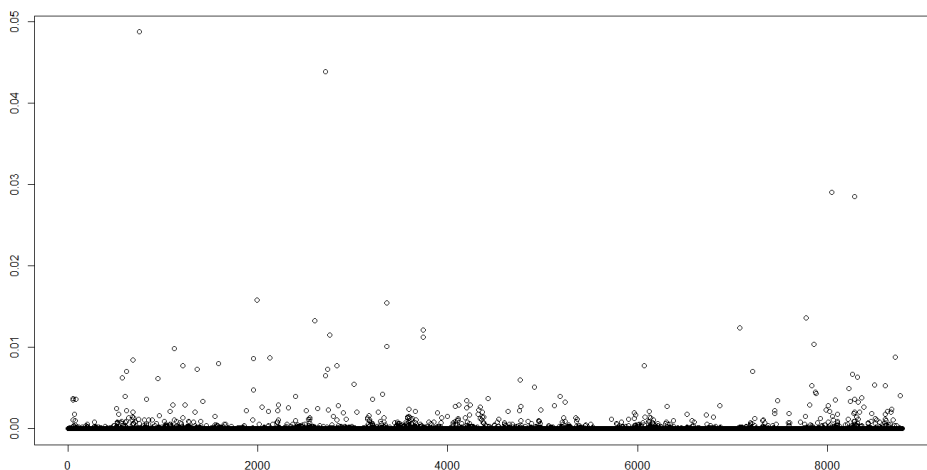


Figura 4-3.: Distancia de cook modelo migró / no migró.

En cuanto a los residuos estandarizados y los valores ajustados, se observan algunos residuos con valores altos.

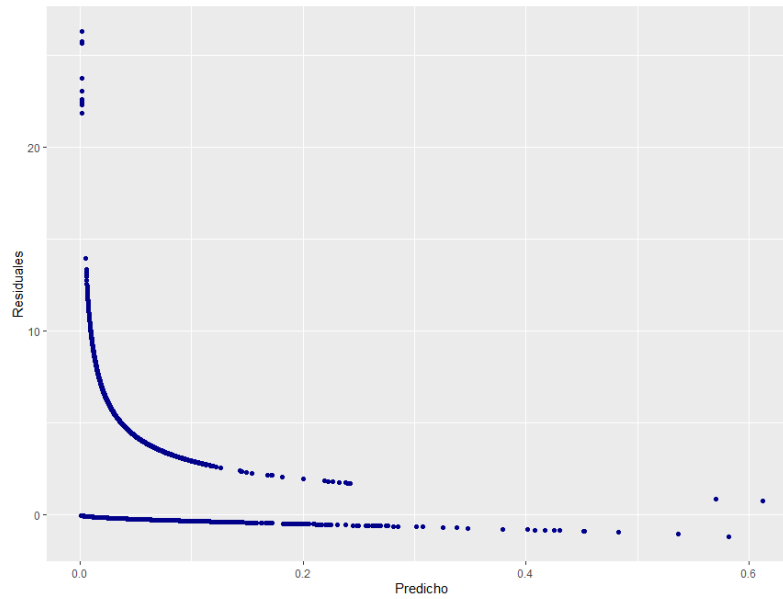


Figura 4-4.: Residuales vs predicho modelo migró / no migró

4.1.2. Resultados modelo propuesto migró / No migró

Según los resultados encontrados en la tabla (4-4), $Edad_h * Edad_h$ al ser positivo y encontrar $Edad_h$ negativo, representa que en edades tempranas es mayor la probabilidad de migrar, luego disminuye dicha probabilidad hasta llegar a un punto de inflexión donde en edades más maduras es más alta la probabilidad de migrar. Como $e^{\hat{\beta}_3} = 1,00045$, se espera que el riesgo (odds ratio) de migrar, al aumentar en una unidad la edad de los hombres de la pareja, aumenta 1,0004 veces, es decir, el riesgo de migrar es 1,0004 veces mayor para una edad $X_i + 1$ que para una edad X_i .

Teniendo en cuenta la ocupación, $Ocupa_{h1}$ que son los hombres que en ese tiempo se encontraban fuera del mercado laboral ($e^{\hat{\beta}_9} = 0,01$), $Ocupa_{h3}$ los profesionales ($e^{\hat{\beta}_{11}} = 0,003$) y $Ocupa_{h6}$ técnicos ($e^{\hat{\beta}_{14}} = 0,03$), tienen menos riesgo de migrar que un desempleado (categoría base), ya que de cada 1.000 parejas el riesgo de migrar al hombre no encontrarse en el mercado laboral es de 10 parejas, si el hombre es profesional es de 3 parejas y si es técnico de 30 parejas. Mientras que los agricultores ($e^{\hat{\beta}_{10}} = 17,88$), técnicos de servicios ($e^{\hat{\beta}_{13}} = 12,7$) y manufactura ($e^{\hat{\beta}_{12}} = 9,31$) tienen más riesgo y son los cargos en los que más se desempeñan los migrantes.

A nivel pareja, el riesgo de migrar es 1,2272 ($e^{\hat{\beta}_{19}}$) veces mayor para unos años pareja $X_i + 1$ que para unos años pareja X_i . Sin embargo, esta migración suele ser únicamente del hombre. Así mismo, las migraciones suceden en mayor medida cuando las parejas tienen más hijos

($e^{\hat{\beta}_{17}} = 1,0575$), la literatura indica que esto se debe a búsqueda de mejores oportunidades por parte del jefe del hogar (usualmente hombre) para su familia, además de su entorno, teniendo en cuenta la TGF de la comunidad en la que se encuentra, influye, siendo parte del aumento de la probabilidad de migrar posiblemente debido a las conexiones que se tengan al ser una comunidad migrante.

Tabla 4-4.: Estimación de parámetros modelo migró / no migró.

Coefficientes	Estimación	95 % intervalo	de confianza	
Intercepto	-4,74503	-6,3978	-3,0922	***
Edad _m	-0,0042	-0,02708	-0,00185	***
Edad _h	-0,00216	-0,0782	-0,0007	***
Edad _h * Edad _h	0,00045	0,00021	0,00071	***
Premig _h	16,9099	12,805	21,0143	***
ExpLab _h	-0,01249	-0,0214	-0,00351	***
ExpLab _m	0,00355	0,001414	0,00952	*
ExpLab _h * Edad _h	-0,0036	-0,00454	-0,00278	***
ExpLab _m * Edad _m	-0,00043	-0,00103	-0,00057	*
Ocupa _{h1}	-4,5833	-6,685	-2,4815	***
Ocupa _{h2}	2,8839	1,991	3,776	***
Ocupa _{h3}	-5,804	-8,6521	-2,955	***
Ocupa _{h4}	2,5416	1,6637	3,4194	***
Ocupa _{h5}	2,2316	1,3562	3,0761	***
Ocupa _{h6}	-3,4785	-6,2328	-0,7242	***
Educ _m	-0,05054	-0,18498	-0,00838	***
Educ _m * Edad _m	0,00087	0,00003	0,00368	***
Hijos	0,056	0,0136	0,0985	***
TGF	0,1383	0,095	0,371	*
Añospareja	0,2048	0,1595	0,25	***

4.2. Modelo propuesto para cantidad de hijos

El siguiente modelo pretende explicar el impacto de la migración en la fecundidad, con variable respuesta Poisson que realiza el conteo de hijos de la pareja en cada año. Además, entre las variables explicativas se presenta **Mig** con variables que indican si migró el hombre o no y si migró la mujer o no.

$$\eta = g(\mu_{ij}) = \ln(\mu_{ij}) = \gamma + \beta^{mig} \mathbf{Mig}_i + \beta^{se} \mathbf{SE}_i + \beta^{co} \mathbf{CO}_i + v_i. \quad (4-3)$$

Con i el número de parejas y j el número de años pareja,

$$i = 1, 2, 3, \dots, 8794$$

$$j = 1, 2, 3, \dots, 36,$$

En que

- $\ln(\mu_{ij})$ es la función de enlace log.
- γ es el efecto medio.
- β^{mig} los efectos poblacionales asociados a la migración.
- β^{se} , los efectos poblacionales asociados a las variables socioeconómicas.
- β^{co} , los efectos poblacionales asociados a las variables de contexto.
- **Mig** asociada a las de variables migratorias.
- **SE** asociada a las de variables socioeconómicas para la $i - \text{ésima}$ pareja en el año t , siendo estas la edad, años de educación, experiencia laboral en meses y ocupación para el hombre y la mujer.
- **CO** asociada a las variables de contexto para la $i - \text{ésima}$ pareja en el año t , tales como TGF, PIB per capita, inflación, prevalencia migratoria por género e índice de capital físico.

Se evaluó el modelo (4-3) mediante diferentes estructuras de correlación por QIC (2-15) y CIC (2-18), se obtuvieron los resultados de la tabla 4-5, siendo la más adecuada “Stationary-M-dependent(1)”, en esta, se supone que existen correlaciones para un pequeño número de unidades de tiempo (1).

Tabla 4-5.: Estructuras de correlación. QIC y CIC.

Estructura de correlación	QIC	CIC
Exchangeable	26.435	25.799
Stationary-M-dependent(1)	22.856	18.685
Stationary-M-dependent(2)	29.457	21.963
Non-Stationary-M-dependent(1)	33.547	30.574
Independence	118.894	118.495
AR2	38.391	36.819

La matriz de correlación estimada correspondiente a la estructura de correlación seleccionada se muestra en la ecuación (4-4).

$$R_i = \begin{pmatrix} 1 & 0,9292 & 0 & \dots & 0 \\ 0,9292 & 1 & 0,9292 & \dots & 0 \\ 0 & 0,9292 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \dots & 1 & 0,9292 \\ 0 & 0 & \dots & 0,9292 & 1 \end{pmatrix}. \quad (4-4)$$

Al igual que con el modelo “Migró / No Migró”, se seleccionan las variables mediante el criterio “ QIC_u ” (2-19), de esta forma, el modelo seleccionado fue:

$$\begin{aligned} \ln(\mu_{ij}) = & \gamma + \beta_1 Migh_{ij} + \beta_2 Migm_{ij} + \beta_3 Premigh_{ij} + \beta_4 Edadm_{ij} + \beta_5 Edadh_{ij} \\ & + \beta_6 (Edadm_{ij} * Edadm_{ij}) + \beta_7 ExpLabh_{ij} + \beta_8 ExpLabm_{ij} \\ & + \beta_9 (ExpLabh_{ij} * Edadh_{ij}) + \beta_{10} (ExpLabm_{ij} * Edadm_{ij}) \\ & + \beta_{11} Ocupa_{h2_{ij}} + \beta_{12} Ocupa_{h3_{ij}} + \beta_{13} Ocupa_{h4_{ij}} + \beta_{14} Ocupa_{h5_{ij}} \\ & + \beta_{15} Ocupa_{h6_{ij}} + \beta_{16} Educh_{ij} + \beta_{17} Educ_{ij} \\ & + \beta_{18} (Educh_{ij} * Edadh_{ij}) + \beta_{19} (Educ_{ij} * Edadm_{ij}) \\ & + \beta_{20} PIBper_j + \beta_{21} Inflación_j + \beta_{22} Añospareja_{ij} \end{aligned}$$

en las que las variables se presentan a continuación

Tabla 4-6.: Variables modelo cantidad de hijos

Variable	Descripción
Migh	¿Migró el hombre?
Migm	¿Migró la mujer?
Edadh	Edad del hombre
Edadm	Edad de la mujer
Premigh	Prevalencia migratoria del hombre
ExpLabh	Experiencia laboral en meses del hombre
ExpLabm	Experiencia laboral en meses de la mujer
Ocupah	Tipo de ocupación del hombre
Educh	Años de educación del hombre
Educum	Años de educación de la mujer
PIBper	Producto Interno Bruto Per Capita
Inflación	Tasa de inflación
Añospareja	Años que lleva la pareja junta

En la tabla **4-3** se encuentran las categorías de ocupación.

4.2.1. Diagnóstico modelo propuesto cantidad de hijos

La matriz de varianzas de los parámetros ajustados “naive” para el modelo propuesto es

$$\hat{V}(\hat{\beta}_n)_{Naive} = \left[\sum_{i=1}^n \left(\hat{D}_i^\top \hat{\Omega}_i^{-1} \hat{D}_i \right) \right]^{-1} = \begin{pmatrix} 0,0738 & \cdots & \cdots & \cdots \\ 0,00798 & 0,0021 & \cdots & \cdots \\ -0,00113 & -0,00197 & \cdots & \cdots \\ \vdots & \ddots & \ddots & \cdots \\ 0,000399 & 0,00077 & 0,00009 & \cdots & 0,0004 \end{pmatrix}.$$

Mediante este se encuentra **A** y **B**, con $CIC(\mathbf{R}) = traza(\mathbf{AB}) = 3,53$.

Se realizó el análisis de puntos palanca (2-23), se puede observar en la figura **4-5** que no hay altos valores de Δ_{ij} .

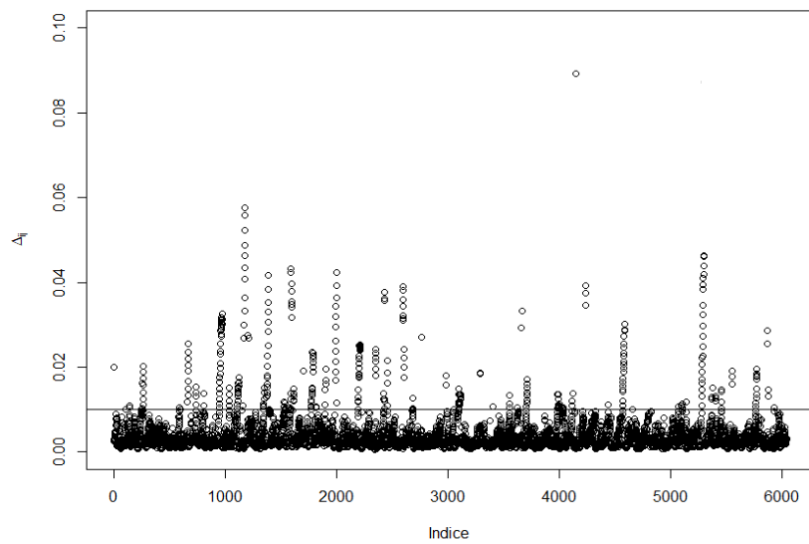


Figura 4-5.: Δ_{ij} vs las observaciones modelo número de hijos

Al revisar los residuales ordinarios (2-21) y evaluar los residuos estandarizados (2-22) de la figura 4-6, no se observan parejas que tengan un residual (r_{SD}) ni un apalancamiento (Δ_{ij}).

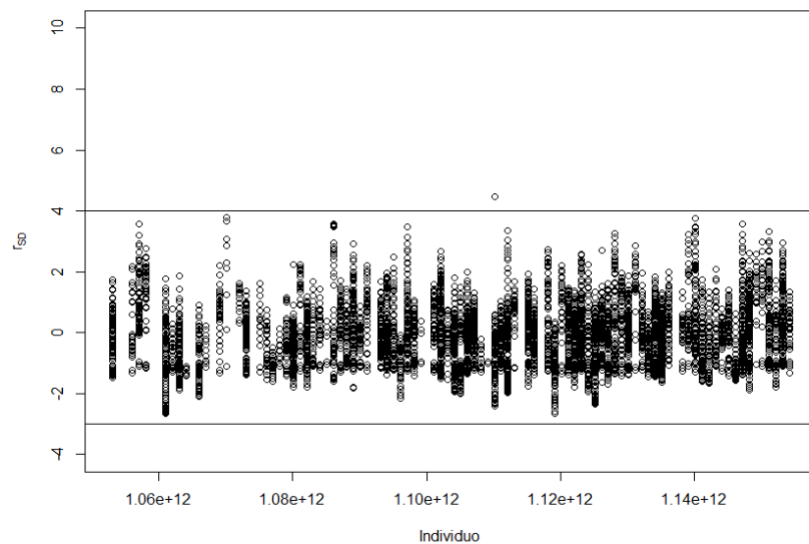


Figura 4-6.: r_{SD} vs parejas modelo número de hijos

Para observar el efecto de quitar un individuo, sobre la estimación de los parámetros en el modelo, se usa Dfbeta, los resultados se muestran en los anexos. Al evaluar las distancias de Cook $(DC)_{ij}$ vs las parejas, se observa que no hay parejas influyentes.

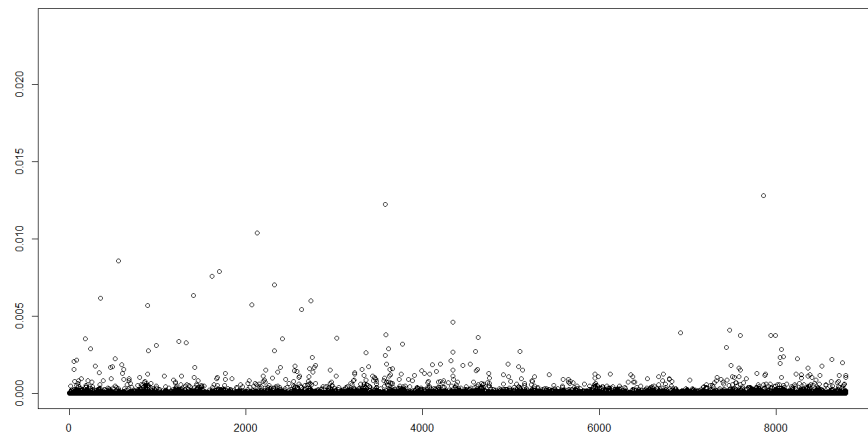


Figura 4-7.: Distancia de Cook modelo número de hijos.

En cuanto a los residuos estandarizados y los valores ajustados, se observa en la figura 4-8 algunos residuos con valores altos.

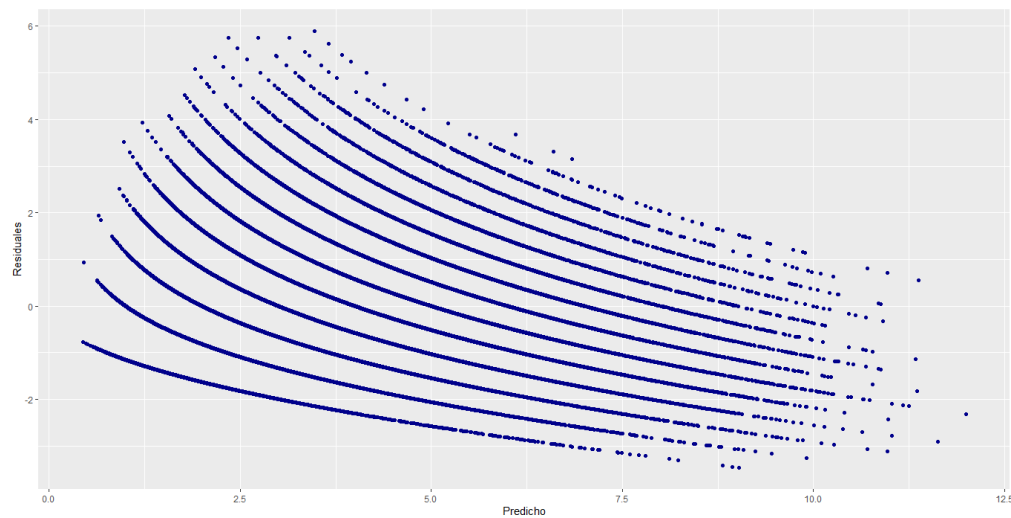


Figura 4-8.: Residuales vs predicho modelo cantidad de hijos

4.2.2. Resultados modelo cantidad de hijos

Los resultados presentados en la tabla (4-7) muestran en cuanto a variables migratorias, que cuando migra el hombre hay mayor probabilidad de tener un hijo adicional, esto es, $e^{0,02727} = 1,027$, es decir, se espera que la razón de aumento del número de hijos con el hombre migrante con respecto al no migrante es 102,7%, esto es, se espera aumente. Puede darse debido a que la pareja queda embarazada y en búsqueda de mejores oportunidades el hombre decide migrar, pero específicamente migra solo, ya que se encontró que si una mujer migra,

dicha probabilidad disminuye ($e^{-0,1183} = 0,8884$). Es decir, las parejas que reportan “a” hijos con respecto a las que reportan “a+1” hijos sea 89 %, esto es, se espera que disminuya.

Por otra parte, a mayor prevalencia migratoria de los hombres, menor es la cantidad de hijos porque las parejas que reportan “a” prevalencia migratoria con respecto a las parejas que reportan “a+1” es del 4,5 % ($e^{-3,082} = 0,0458$), es decir, disminuye.

En cuanto a las mujeres se tiene $Edad_m * Edad_m$ al ser negativo y encontrar $Edad_m$ positivo, representa que en edades tempranas es menor la probabilidad de hijos adicionales, luego aumenta dicha probabilidad hasta llegar a un punto de inflexión donde en edades más maduras es menor la probabilidad de migrar. Como $e^{-0,0006} = 0,999$, se espera que disminuya la razón de aumento de número de hijos de la pareja. Esto debido a la biología de la mujer, a pesar de filtrar la base hasta 49 años el cual es la edad reproductiva de la mujer.

De forma similar al modelo Migró / No migro, se encontró la ocupación del hombre como variable determinante en el número de hijos, esto es, cuando el hombre tiene alguna ocupación aumenta la probabilidad de hijos, mientras que si no se encontraban en la fuerza laboral su impacto es negativo. Esto se debe principalmente a causas económicas, donde si no hay un sustento, específicamente del hombre, no habría incentivos a tener hijos.

En cuanto a datos de contexto, se encontró que en los años donde el PIB per cápita o la inflación era mayor, asimismo fue la cantidad de hijos. De igual forma, entre más años dure unida la pareja, mayor es la razón de aumento de hijos.

Tabla 4-7.: Estimación de parámetros Número de hijos.

Coefficientes	Estimación	95 % intervalo	de confianza	
Intercepto	-1,2178	-1,383	-1,0523	***
Mig _h	0,02727	0,0029	0,0575	***
Mig _m	-0,1183	-0,1931	-0,044	**
Premig _h	-3,0823	-3,4141	-2,7506	***
Edad _h	0,01614	0,01257	0,0197	***
Edad _m	0,05434	0,04327	0,06540	***
Edad _m * Edad _m	-0,0006	-0,0008	-0,0004	***
ExpLab _h	0,00346	0,00228	0,00464	***
ExpLab _m	0,008307	0,00718	0,00943	***
ExpLab _h * Edad _h	-8 e ⁻⁵	-1 e ⁻⁴	-6 e ⁻⁵	***
ExpLab _m * Edad _m	-0,0002	-0,00022	-0,00017	***

Coefficientes	Estimación	95 % intervalo	de confianza	
Ocupa _{h1}	-0,0803	-0,1868	-0,035	***
Ocupa _{h2}	0,0679	0,0029	0,1369	**
Ocupa _{h3}	0,0131	0,009	0,0584	*
Ocupa _{h4}	0,0394	0,0259	0,106	*
Ocupa _{h5}	0,0102	0,0052	0,059	*
Ocupa _{h6}	0,0094	0,001	0,0126	**
Educ _h	0,0158	0,0062	0,0255	***
Educ _m	-0,00538	-0,0149	-0,0007	*
Educ _h * Edad _h	-0,00062	-0,00088	-0,00036	***
Educ _m * Edad _m	-0,00024	-0,0005	-0,0001	**
PIBper	0,0015	0,00083	0,0022	***
Inflación	0,0004	0,00035	0,0005	***
Añospareja	0,0769	0,0658	0,08806	***

5. Conclusiones

El presente trabajo de grado se desarrolló con consideraciones teóricas sobre las relaciones entre migración y fecundidad por medio de la aplicación de los Modelos lineales generalizados, ecuaciones de estimación generalizada, específicamente modelos logit y Poisson con datos longitudinales. En estos, la variable de interés corresponde a la opción de estar en Estados Unidos o en México y la cantidad hijos de las parejas en cada año que permanecen juntas las parejas en el estudio.

Tanto el modelo de migró o no migró y de número de hijos presentaron buenos ajustes para explicar la variabilidad de los datos y tuvieron algunos estadísticos de ajuste similares lo que evidencia ser una opción para modelar datos longitudinales de tipo demográfico que buscan realizar inferencia a nivel individual, ya que permiten construir modelos específicos para individuos (en este caso parejas) teniendo en cuenta las relaciones internas.

En cuanto a la probabilidad de migrar se encontró evidencia empírica sobre un efecto positivo de la ocupación del hombre siendo agricultores, obreros o manufactureros y efectos negativos a mayor edad del hombre de la pareja o experiencia laboral de ambos, de manera importante, se evidencia un efecto positivo a mayor cantidad de hijos. Esto puede explicar la separación de las familias Mexicanas con la migración del hombre generando luego redes migratorias para su nueva unión, de esta forma, la tenencia de hijos impulsa la migración en busca de mejores condiciones de vida para la familia.

Por otra parte, el modelo Poisson de número de hijos encontró evidencia de un efecto positivo cuando el hombre se encuentra en la fuerza laboral o la pareja tiene mayor cantidad de años junta. Mientras presenta un efecto negativo a más experiencia laboral, educación, si la mujer migra o si residen en una comunidad donde la prevalencia migratoria del hombre es alta. Esto indica que la separación de la familia impulsada por la migración de uno de los miembros de la pareja, especialmente la mujer, genera un detenimiento de la fecundidad de la pareja.

Los anteriores resultados sirven para tener un panorama general de algunas características que generan interacción entre la migración y fecundidad en un país limítrofe con uno de los mayores importadores de inmigrantes del mundo, la importancia se debe a que son dos variables centrales en los estudios demográficos. Adicional, es una ventana para la aplicación

del uso de datos longitudinales mediante diferentes metodologías como las mencionadas en este trabajo de grado.

6. Trabajos futuros

Se proponen algunos trabajos futuros para seguir explorando la metodología del uso de datos longitudinales en temas demográficos de migración y fecundidad:

- Los modelos propuestos se construyeron sin tener en cuenta la censura de los datos, se puede realizar un análisis teniendo en cuenta dicha censura de los datos por la entrada y salidas de parejas en diferentes años y diferente duración de la relación.
- No se tuvo en cuenta el factor expansión que nos proporciona MMP, por esto uno de los trabajos futuros sería explorar las metodologías con este uso.
- Se pueden considerar modelos de transiciones (**3-2**) y modelos lineales generalizados mixtos y aplicaciones del factor expansión de la muestra que nos proporciona MMP y LAMP.
- Es posible expandir el análisis a otros países mediante LAMP (misma encuesta de MMP aplicada a otros países de Sur América y Centro América) y así conocer los cambios de dinámicas según el país de origen del migrante y aplicación de otras metodologías.
- Una opción es modelar conjuntamente el número de hijos con el estado migratorio de las parejas.
- Estimar cuando las covariables varían en el tiempo y su correlación entre ellas.
- Inclusión de metodologías para establecer si el modelo es adecuado, así como otras medidas alternativas de CIC y QIC.
- Es posible incluir la correlación espacial en el estudio de migración.

A. Anexo I: Código QIC

```
capture program drop qicp
program define qicp

args corr quietly {
capture xtgee dept X, i(llave) t(year) /*
*/ fam(bin) corr('corr') nmp robust

matrix V = e(V)
matrix b = e(b)
tempvar mu ql
predict double 'mu', mu
generate double 'ql' = (abroad1*log('mu'/(1-'mu')) + log(1-'mu'))
summarize 'ql' if e(sample), meanonly
local qlik = r(sum)
local chi2 = e(chi2_dis)
capture xtgee dept X, i(llave) t(year) ///
fam(bin) link(logit) corr(ar 2) robust ///
iter(1) nowarning
matrix A = e(V)
matrix Ai = syminv(A)

matrix T = Ai*V
matrix t = trace(T)
scalar off = t[1,1]
}
display as txt "QIC(P) = .as res %9.0g 2*(off-'qlik')/'chi2'
end

capture program drop qichh
program define qichh

args corr
```

```
quietly
capture xtgee dept X, i(llave) t(year)
fam(bin) corr('corr') nmp robust
if (_rc != 0) {
exit
}

matrix V = e(V)
matrix T = Ai*V
matrix t = trace(T)
scalar off = t[1,1]

tempvar mu ql
predict double 'mu', mu
generate double 'ql' = (y*log('mu'/(1-'mu')) + log(1-'mu'))
summarize 'ql' if e(sample), meanonly
}
display as txt "QIC(HH) = .as res %9.0g 2*(off-r(sum)/e(chi2dis))
end
```

B. Anexo II: Dfbetas Modelo Propuesto Migró/ No Migró

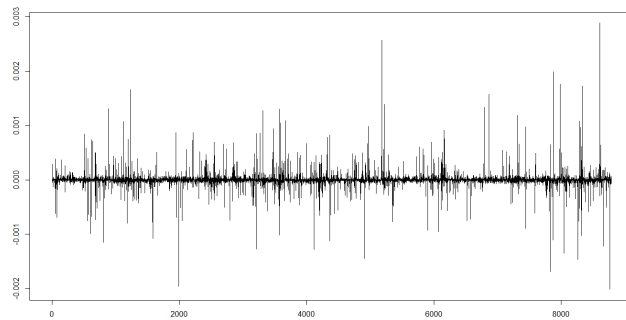


Figura B-1.: Edad de los hombres

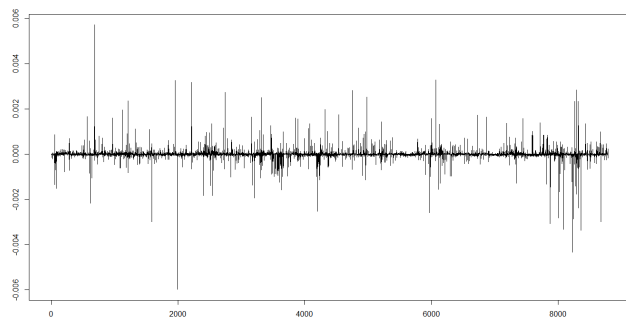


Figura B-2.: Años educación de mujeres

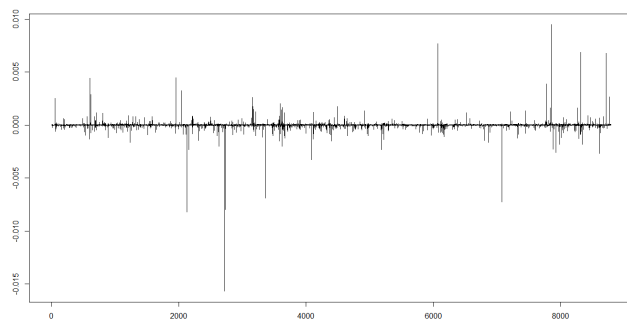


Figura B-3.: Años pareja

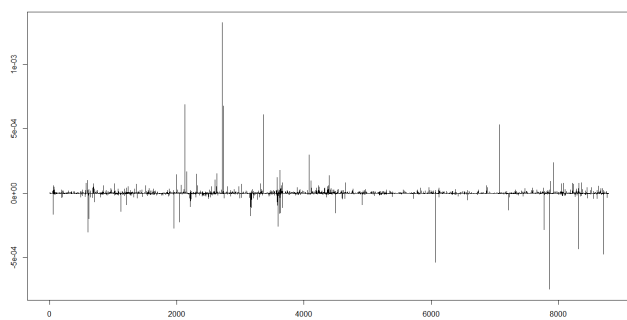


Figura B-4.: Experiencia laboral de hombres

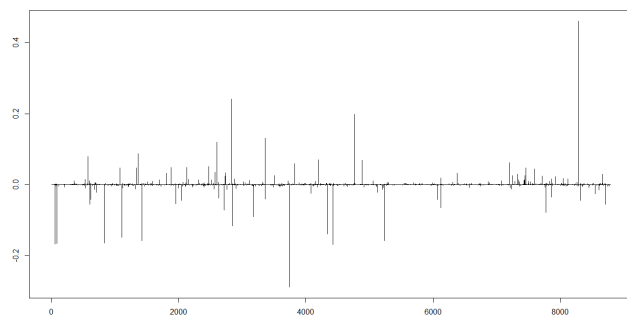


Figura B-5.: Hombre no en la fuerza laboral

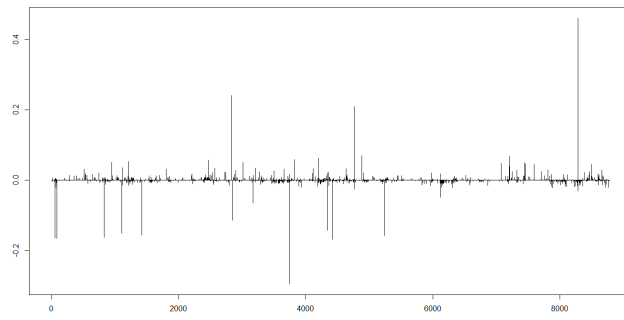


Figura B-6.: Hombre ocupación agricultura

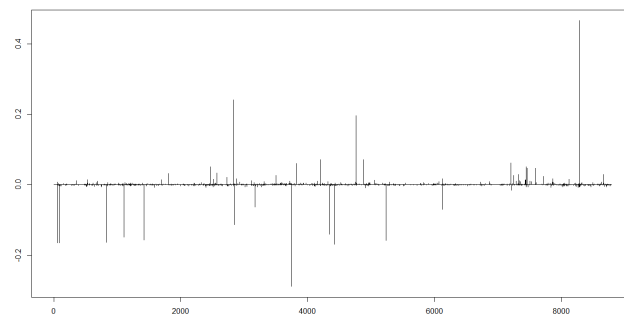


Figura B-7.: Hombre ocupación Manufactura

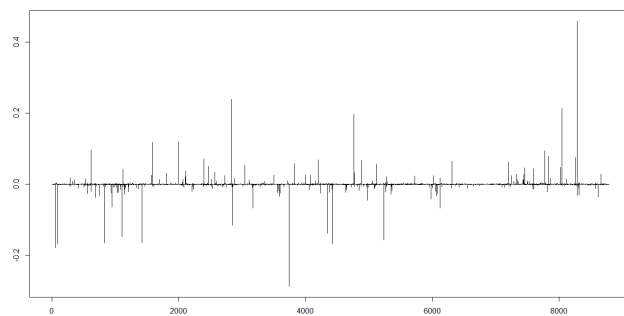


Figura B-8.: Hombre ocupación Profesional o administrativo

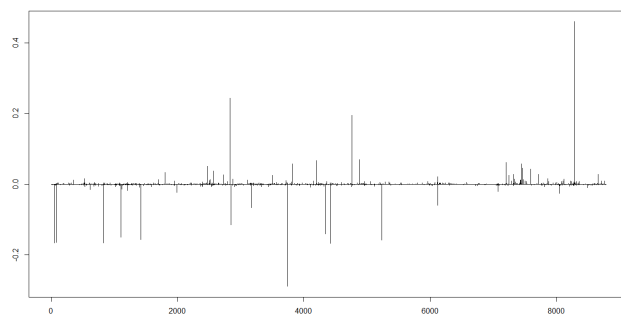


Figura B-9.: Hombre ocupación servicios

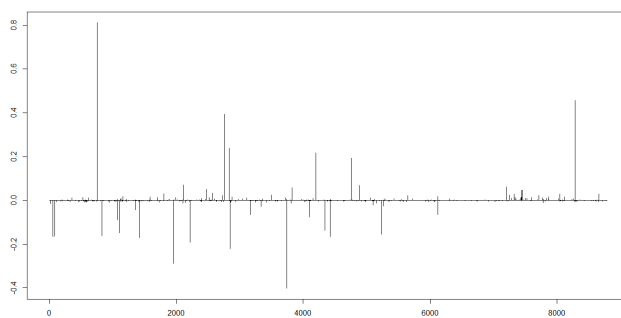


Figura B-10.: Hombre ocupación Técnicos

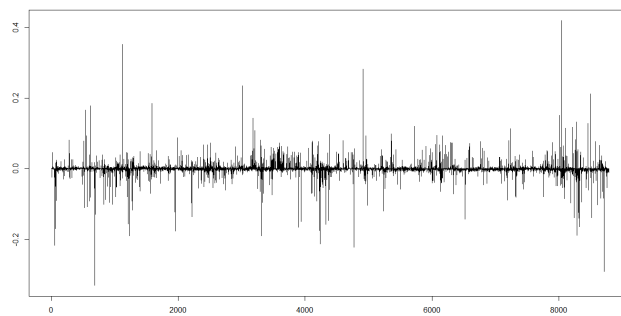


Figura B-11.: Prevalencia migratoria de hombres

C. Anexo III: Dfbetas Modelo Propuesto Cantidad de Hijos

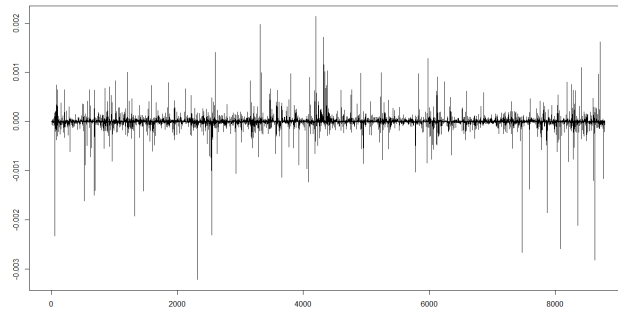


Figura C-1.: Migración hombres

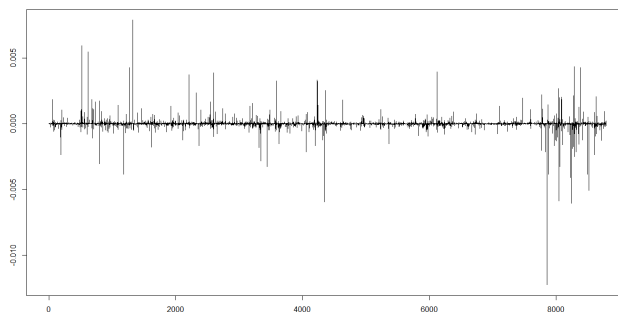


Figura C-2.: Migración mujeres

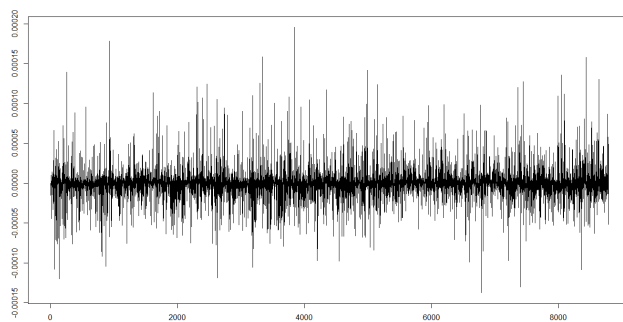


Figura C-3.: Edad mujeres

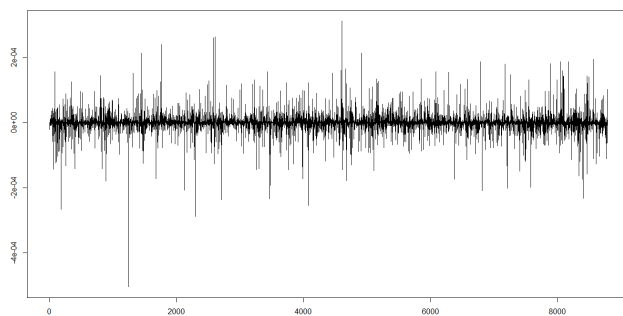


Figura C-4.: Años educación hombres

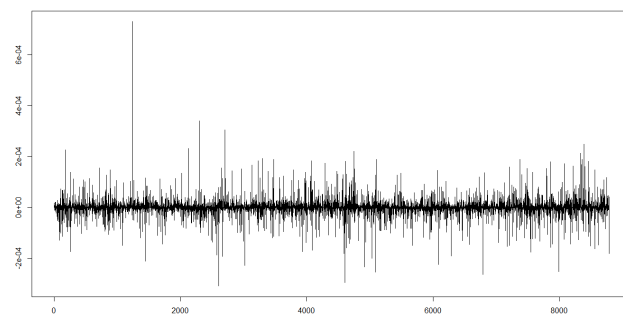


Figura C-5.: Años educación mujeres

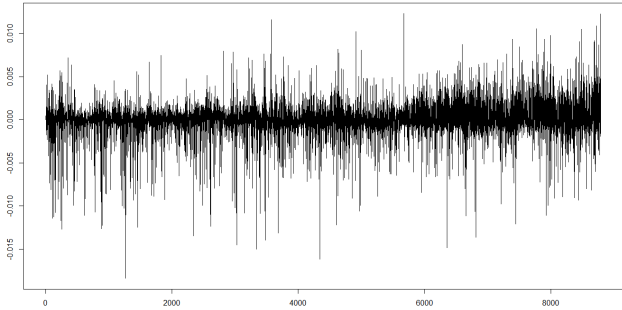


Figura C-6.: Prevalencia migratoria hombres

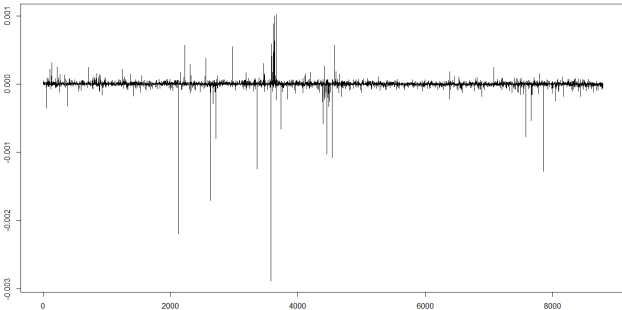


Figura C-7.: Años pareja

Bibliografía

- Andersson, Gunnar Persson, L. . O. O. (2017). Depressed fertility among descendants of immigrants in Sweden. *Demographic Research*, S23(39):1149–1184. <https://www.demographic-research.org/special/23/39/>.
- Bongaarts, J. (1978). A framework for analyzing the proximate determinants of fertility. *Population and Development Review*, 4:105–132.
- Bossio, M. L. C. (2017). Modelos bayesianos para datos longitudinales: extensiones teóricas y metodológicas. *Universidad Nacional de Colombia*.
- Botter, A. . (2005). Funções de estimação em modelos de regressão. *Escola de Modelos de Regressão*, 9. <https://docplayer.com.br/87716665-Rinaldo-artes-denise-aparecida-botter.html>.
- Carter, M. (2000). Fertility of mexican immigrant women in the u.s.: A closer look. *Social Science Quarterly*, 81(4):1073–1086. <http://www.jstor.org/stable/42864041>.
- CEPAL, C. (2019). Tendencias recientes de la población de américa latina y el caribe. https://www.cepal.org/sites/default/files/static/files/dia_mundial_de_la_poblacion_2019.pdf.
- Choi, K. (2014). Fertility in the context of Mexican migration to the United States: A case for incorporating the pre-migration fertility of immigrants. *Demographic Research*, 30(24):703–738. <https://www.demographic-research.org/volumes/vol30/24/>.
- Cooke, Thomas Mulder, C. . T. M. (2016). Union dissolution and migration. *Demographic research*, 34:741–760.
- Diggle, Peter Heagerty, P. . L. K.-Y. . Z. S. (2002). *The Analysis of Longitudinal Data*, volume 90.
- Dubuc, S. (2012). Immigration to the uk from high-fertility countries: Intergenerational adaptation and fertility convergence. *Population and Development Review*, 38(2):353–368. <http://www.jstor.org/stable/41857383>.
- Fargues, P. (2011). International migration and the demographic transition: A two-way interaction. *International Migration Review*, 45(3):588–614. <http://www.jstor.org/stable/23016205>.

- Ferrer, A. A. . A. (2014). Immigrants and demography: Marriage, divorce, and fertility. IZA Discussion Papers 7982. <http://hdl.handle.net/10419/93302>.
- Fitzmaurice, G. M. Laird, N. M. . W.-J. H. (2011). *Applied Longitudinal Analysis*. Wiley.
- Fox, J. (2016). *Applied regression analysis and generalized linear models*. SAGE Publication.
- Frank Heuveline, P. (2005). A crossover in mexican and mexican-american fertility rates: Evidence and explanations for an emerging paradox. *Demographic research*, 12(2):588–614. <https://doi.org/10.4054/DemRes.2005.12.4>.
- Gibbons, R.D. Hedeker, D. (2006). *Longitudinal Data Analysis*. Wiley Series in Probability and Statistics. Wiley. <https://books.google.com.co/books?id=f9p9iIgzQSQC>.
- Gonzalez, L. M. (2022). *Notas de clase: Datos longitudinales*.
- Grande, A. D. R. . R. (2017). El efecto de las condiciones de llegada en la reproducción de las mujeres migrantes: The effect of arrival conditions on the reproduction of migrant women. (159):113–133. <http://www.jstor.org/stable/26382374>.
- Hardin, J.W., . H. J. (2013). *Generalized Estimating Equations*. Chapman and Hall/CRC.
- Karlsson, A. (2007). Longitudinal data analysis, by d. hedeker and r. d. gibbons. *Journal of Biopharmaceutical Statistics*, 17(3):511–513. <https://doi.org/10.1080/10543400601001592>.
- Kulu, Hill Steele, F. (2013). Interrelationship between childbearing and housing transitions in the family life course. *Population Association of America*, 50:1687–1714.
- Kulu, H. (2008). Fertility and spatial mobility in the life course: evidence from austria. *Environment and Planning*, 40:632–652.
- Li, T. P. . C. S. D. . N. (1998). Alternative gee estimation procedures for discrete longitudinal data. *Computational Statistics Data Analysis*, 28(3):243–256. <https://www.sciencedirect.com/science/article/pii/S0167947398000395>.
- Lindstrom, David P. Saucedo, S. G. (2002). The Short- and Long-Term Effects of U.S. Migration Experience on Mexican Women’s Fertility. *Social Forces*, 80(4):1341–1368. <https://doi.org/10.1353/sof.2002.0030>.
- Lindstrom, David P. Saucedo, S. G. (2007). The interrelationship between fertility, family maintenance, and mexico-u.s. migration. *Demographic Research*, 17:821–858.
- Long, L. H. (1972). The influence of number and ages of children on residential mobility. *Demography*, 9(3):371–382. <https://doi.org/10.2307/2060860>.

- Maalouf, M. (2015). Rare events and imbalanced datasets: an overview.
- Madhavan, Sangeetha Schatz, E. . C. S. . C. M. (2012). Child mobility, maternal status, and household composition in rural south africa. *Demography*, 49:699–718.
- Massey, D. (1994). Continuities in transnational migration: An analysis of nineteen mexican communities. *American Journal of Sociology*, 99(6):1492–1533. <http://www.jstor.org/stable/2782582>.
- Milewski, N. (2010). Immigrant fertility in west germany: Is there a socialization effect in transitions to second and third births? *European Journal of Population / Revue Européenne de Démographie*, 26(3):297–323. <http://www.jstor.org/stable/40784331>.
- MMP (2012). Ponderadores en el proyecto sobre migración mexicana.
- MMP (2022a). Proyecto de migración mexicana. Technical report. <https://mmp.opr.princeton.edu/home-es.aspx>.
- MMP (2022b). Proyecto de migración mexicana: Mapa y animaciones. Technical report. <https://mmp.opr.princeton.edu/research/maps-es.aspx>.
- Nelder, J. A. Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–374.
- Oliveira, J. (2016). Fertility, Migration, and Maternal Wages: Evidence from Brazil. *Journal of Human Capital*, 10(3):377–398. <https://ideas.repec.org/a/ucp/jhucap/doi10.1086-687416.html>.
- ONU, N. U. (2001). Migraciones de reemplazo: ¿una solución ante la disminución y el envejecimiento de las poblaciones?
- Pan, W. (2001). Akaike’s information criterion in generalized estimating equations. <https://doi.org/10.1111/j.0006-341x.2001.00120.x>.
- Pellegrino, A. (2003). La migración internacional en América Latina y el Caribe: tendencias y perfiles de los migrantes. *Población y Desarrollo*, 35.
- Sandell, S. H. (1977). Women and the economics of family migration. *The Review of Economics and Statistics*, 59(4):406–414. <http://www.jstor.org/stable/1928705>.
- Schultz, H. R. . T. P. (1980). Migrant and native fertility in colombia in 1973: Migrants selected according to their reproductive preferences?
- Strzelecka, E. K. (2012). Estudios de género y desarrollo: teoría, política y práctica. *Desarrollo Humano: Teoría y Aplicaciones*.

- Tapia, M. (2011). Género y migración: trayectorias investigativas en iberoamérica. *Encrucijada Americana*, 4. <https://encrucijadaamericana.uahurtado.cl/index.php/ea/article/view/99>.
- Twisk, J. W. R. (2013). *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide*. Cambridge University Press, 2 edition.
- UNFPA (2018). Los desafíos de la baja fecundidad en américa latina y el caribe. <https://lac.unfpa.org/sites/default/files/pub-pdf/Baja%20fecundidad%20en%20ALC%20-%20version%20web%20espa%C3%B1ol.pdf>.
- Venezuela, Maria Kelly; Botter, D. A. (2007). Diagnostic techniques in generalized estimating equations. *Journal of Statistical Computation and Simulation*.
- Weiss, R. (2005). *Introduction to Longitudinal Data*, pages 1–26. Springer New York, New York, NY. https://doi.org/10.1007/0-387-28314-5_1.
- Wolf, K. (2016). Marriage migration versus family reunification: How does the marriage and migration history affect the timing of first and second childbirth among turkish immigrants in germany? *European journal of population = Revue europeenne de demographie*, 32(5):731—759. <https://europepmc.org/articles/PMC5126209>.
- Zeger, K.-Y. L. . S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22. <http://www.jstor.org/stable/2336267>.
- Zimmerman, D.L. Núñez-Antón, V. (2009). *Antedependence Models for Longitudinal Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press. <https://books.google.com.co/books?id=l8jxSvWNjAQC>.