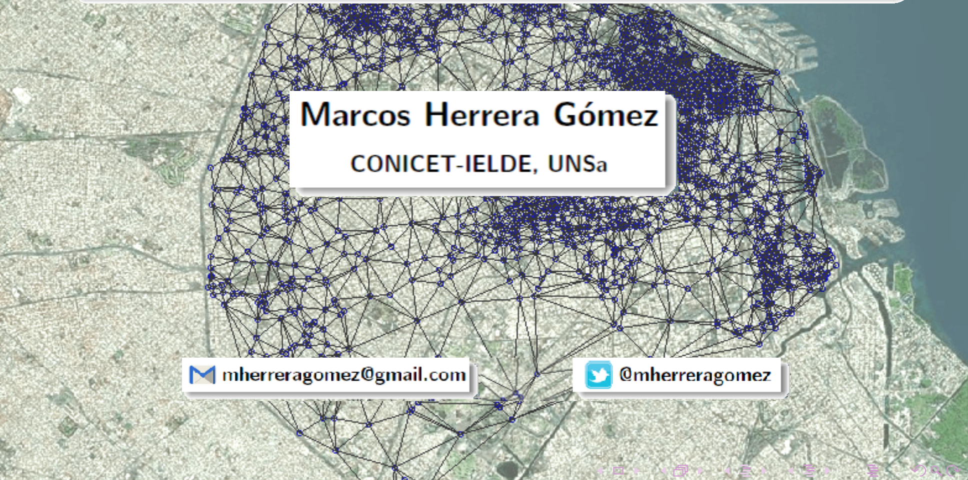
# Econometría Espacial: Estado actual y recientes avances teóricos-aplicados




**Marcos Herrera Gómez**

CONICET-IELDE, UNSa

mherreragomez@gmail.com

@mherreragomez

# Esquema de Presentación

# Esquema de Presentación

# Econometría espacial

- Definición actual: Rama econométrica que analiza las **EFECTOS ESPACIALES** en modelos de regresión en datos de corte transversal y de panel, considerando explícitamente el arreglo espacial.
- **ESPACIAL** entendido en sentido amplio: interacciones de naturaleza geográfica y/o económica-social.
- Desarrollo de la disciplina:
  - Fines de los 70's (hace 40 años!): dentro de ciencia regional, economía regional-urbana.
  - Mainstraim de economía: antes de los 90's, aspectos geográficos considerados irrelevantes.
  - Desde los 90's: (1) desarrollos teóricos (NGE, Social-networks), (2) geo-data.

# Esquema de Presentación

# Primera década: 1979-1987

- Jean Paelinck primer investigador en lanzar el término "econometría espacial"
- Libros de referencia inicial:
  Paelinck y Klaassen (1979): Spatial Econometrics.
  Bennett (1979): Spatial time series.
- Definición:
  "Explícita incorporación del espacio (topología) y de interdependencia/simultaneidad en la especificación econométrica de modelos regionales"

# Segunda década: 1era agenda de investigación

- Libros de referencia:
  Anselin (1988): Spatial econometrics: methods and models.
  Anselin y Florax (eds, 1995): New Directions in Spatial Econometrics.
- Definición:
  "Colección de técnicas que tratan con peculiridades causadas por el espacio en el análisis estadístico en modelos de ciencia regional"
- Case (1992), Fotheringham (1997), Kelejian y Prucha (1997,1998,1999), LeSage (1997).
- Mainstream economía: interés empírico por avances teóricos (Krugman, Fujita, Glaeser, Akerlof, Manski).

# Tercera década (2000's): consolidación

- Libros de referencia:
  Fotheringham, Brunsdon y Charlton (2002): Geographically Weighted Regression.
  Anselin, Florax y Rey (eds,2004): Advances in Spatial Econometrics.
  LeSage y Pace (2009): Introduction to spatial econometrics.
- Spatial Econometrics Association (2006):
  Anselin, Arbia, Baltagi, Bera, Cressie, Fingleton, Kelejian, Mur, Lee, LeSage, Pesaran, Prucha, Rey, Robinson.
- Spatial Econometrics Advanced Institute (Roma, Italia): 1era edición 2008.
- Mainstream economía: Krugman (Nobel 2008) ubica a la localización como central en teoría y econometría.

# Cuarta década (actual): expansión

- Libros de referencia:
  Elhorst (2014): Spatial econometrics: from cross-sectional data to spatial panels.
  Anselin y Rey (2014): Modern Spatial Econometrics in Practice.
- Era “big data”: Explosión de grandes volúmenes de información georreferenciada.
- Desarrollos de software: R, Python, además de tradicionales Matlab, GeoDa.
- Estadísticas oficiales espaciales.
- Numerosos avances espacio-temporales.

# Esquema de Presentación

# Efectos espaciales

## Dependencia espacial

El valor observado en un punto depende funcionalmente de los valores vecinos. Primera ley de la geografía (Tobler, 1970).

# Efectos espaciales

## Dependencia espacial

El valor observado en un punto depende funcionalmente de los valores vecinos. Primera ley de la geografía (Tobler, 1970).

## Heterogeneidad espacial

Inestabilidad estructural en forma funcional: (1) Varianza de error no-constante (heterocedasticidad espacial); (2) Coeficientes no-constantes (regímenes espaciales, variación continua).

# Características de los modelos espaciales

La econometría espacial reconoce desde el inicio la naturaleza dependiente de los datos:

- supuesto i.i.d. no es útil.

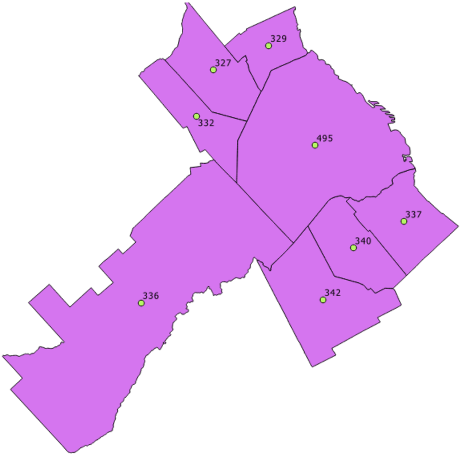
La dependencia espacial es un resultado esperado:

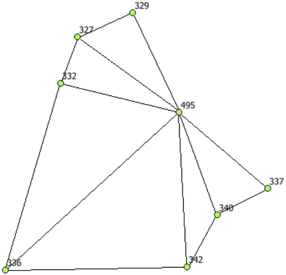
- multidireccionalidad (muy diferente a series de tiempo).
- conductas estratégicas: spillovers-contagio-copycat.
- resultado observado por agregación.

El vecindario (contexto geográfico-social) es de principal relevancia:

- necesidad de definir "quién es vecino de quién": matriz W.

# Definiendo “vecindario”: matriz de contactos W



# Incorporando “vecindario” al modelo econométrico

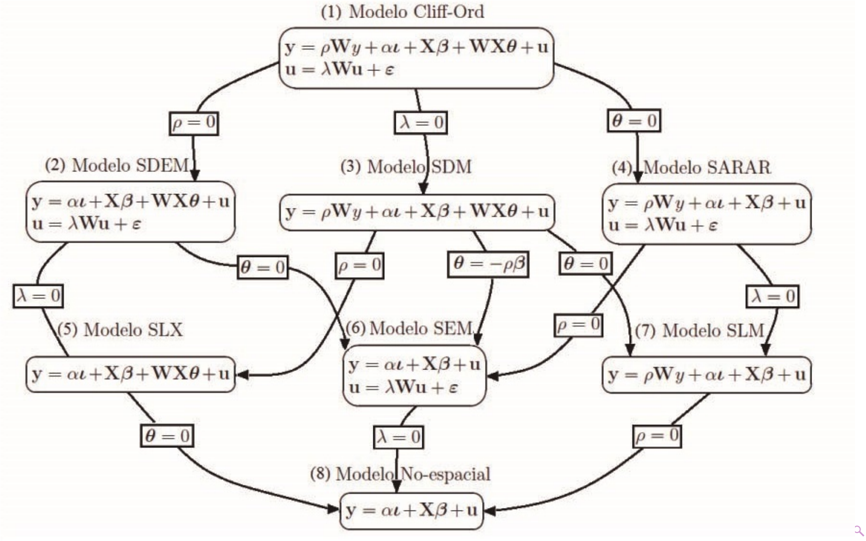
## (1) Modelo No Restringido

$$\left.\begin{array}{l} y_i = \alpha_{ij} y_j + \alpha_{ik} y_l + x_i \beta + \varepsilon_i \\ y_j = \alpha_{ji} y_i + \alpha_{jk} y_l + x_j \beta + \varepsilon_j \\ y_l = \alpha_{li} y_i + \alpha_{lj} y_j + x_l \beta + \varepsilon_l \end{array}\right\} \mathbf{y} = \Gamma \mathbf{y} + X\beta + \varepsilon$$

## (2) Modelo Restringido

$$\left.\Gamma = \begin{bmatrix} 0 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & 0 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & 0 \end{bmatrix} = \rho \begin{bmatrix} 0 & \omega_{12} & \omega_{13} \\ \omega_{21} & 0 & \omega_{23} \\ \omega_{31} & \omega_{32} & 0 \end{bmatrix}\right\} = \rho W$$

$\mathbf{y} = \rho W \mathbf{y} + X\beta + \varepsilon$


(1) Modelo Cliff-Ord
y = ρWy + αι + Xβ + WXθ + u
u = λWu + ε
ρ = 0
λ = 0
θ = 0
(2) Modelo SDEM
y = αι + Xβ + WXθ + u
u = λWu + ε
(3) Modelo SDM
y = ρWy + αι + Xβ + WXθ + u
(4) Modelo SARAR
y = ρWy + αι + Xβ + u
u = λWu + ε
θ = 0
ρ = 0
θ = −ρβ
θ = 0
λ = 0
λ = 0
ρ = 0
(5) Modelo SLX
y = αι + Xβ + WXθ + u
(6) Modelo SEM
y = αι + Xβ + u
u = λWu + ε
(7) Modelo SLM
y = ρWy + αι + Xβ + u
θ = 0
λ = 0
ρ = 0
(8) Modelo No-espacial
y = αι + Xβ + u

# Algunos modelos espacio-tiempo dinámicos

Algunos tipos de modelos de Durbin dinámicos:

1. Simultáneo espacio-temporal

$$y_t = \tau y_{t-1} + \rho W y_t + X_t \beta + W X_t \theta + \mu + \varepsilon_t.$$

2. Recursivo puro

$$y_t = \gamma W y_{t-1} + X_t \beta + W X_t \theta + \mu + \varepsilon_t.$$

3. Recursivo espacio-temporal

$$y_t = \tau y_{t-1} + \gamma W y_{t-1} + \rho W y_t + X_t \beta + W X_t \theta + \mu + \varepsilon_t.$$

# Esquema de Presentación

- Selección de W exógena:
  - Zhang y Yu (2018): Spatial weights matrix selection and model averaging for spatial autoregressive models
  - Herrera et al. (2019): A Comparison Study on Criteria to Select the Most Adequate Weighting Matrix.
- Quiebres estructurales:
  - Mur et al (2018): Testing for breaks in the weighting matrix
- Test endogeneidad:
  - Bera et al (2018): Simple tests for endogeneity of spatial weights matrices
- Estimación bajo W endógena:
  - Shi y Lee (2018): Estimating a spatial autoreg. model with an endogenous spatial W

- Paneles estáticos y dinámicos:
  - Elhorst et al. (2018):
    - Spillovers in space and time: Where spatial econometrics and Global VAR models meet
  - Bera et al (2018):
    - Robust LM tests for spatial dynamic panel data models
- Mix espacio-multinivel:
  - Anselin et al (2018):
    - Spatially-Correlated Multilevel Models.

1. Interacción con "geo-data science":
    - Uso similar de información pero enfoques muy diferentes.
2. Combinación con modelos jerárquicos:
    - Avances escasos en comparación a su potencialidad.
3. Causalidad:
    - Dificultad del enfoque experimental.
    - Complejidad espacio-tiempo bajo datos observacionales.
4. Limitaciones de software:
    - Programas trabajan para muestras "pequeñas" según "big data" (más de 70mil obs, resultados no-confiables)

# Esquema de Presentación

## Análisis simbólico. Conceptos Básicos

- *Análisis Simbólico*: transformación de una serie en una secuencia de símbolos que capturan información útil no directamente observable.
  - Dado un proceso $\{x_s\}_{s \in S}$, definimos una función $f : \{x_s\}_{s \in S} \to \Gamma_n$, donde $f$ es una *función simbolizadora y* $\Gamma_n$ *es un conjunto de símbolos* ($\sigma^x s$).
- Conceptos Claves:
  - $m$ es la *dimensión de encaje*: brinda la longitud del vecindario o entorno.
  - $m - entorno$: conjunto formado con la información de cada observación y sus $(m-1)$ vecinos. *El* $m - entorno$ *es la información que se utiliza para construir el* $\sigma^x$.

## Procedimiento de Simbolización: Caso Particular

(1ro) Se propone la mediana, $M_e^x$, como una medida de inicio de la simbolización.

(2do) Sea la función indicadora $\tau_s = \begin{cases} 1 & si \quad x_s \geq M_e^x \\ 0 & otro \quad caso \end{cases}$

(3ro) Elegir $m \geq 2$ y luego construir el $m - entorno$ para cada observación $s$: $x_m(s) = \left(\tau_s, \tau_{s_1}, \ldots, \tau_{s_{m-1}}\right)$.

(4to) Aplicar la función simbolizadora $f : \{x_s\}_{s \in S} \rightarrow \Gamma_m$, definida como:

$$f(x_s) = \sum_{i=1}^{m-1} 1\left(\tau_s = \tau_{s_i}\right) = \sigma_s^x$$

## Mapa Regular 3x3: Ejemplo

| | | |
|---|---|---|
| $X_{s_1} = 4$ | $X_{s_2} = 1$ | $X_{s_3} = 3$ |
| $X_{s_4} = 6$ | $X_{s_5} = 2$ | $X_{s_6} = 5$ |
| $X_{s_7} = 1$ | $X_{s_8} = 2$ | $X_{s_9} = 4$ |

| | | |
|---|---|---|
| $Y_{s_1} = 5$ | $Y_{s_2} = 2$ | $Y_{s_3} = 4$ |
| $Y_{s_4} = 0$ | $Y_{s_5} = 2$ | $Y_{s_6} = 3$ |
| $Y_{s_7} = 7$ | $Y_{s_8} = 9$ | $Y_{s_9} = 3$ |

# Estimar la Mediana de X y de Y: $X < M_e^x$, $Y < M_e^y$ en gris

| | | |
|---|---|---|
| $X_{s_1} = 4$ | $X_{s_2} = 1$ | $X_{s_3} = 3$ |
| $X_{s_4} = 6$ | $X_{s_5} = 2$ | $X_{s_6} = 5$ |
| $X_{s_7} = 1$ | $X_{s_8} = 2$ | $X_{s_9} = 4$ |

| | | |
|---|---|---|
| $Y_{s_1} = 5$ | $Y_{s_2} = 2$ | $Y_{s_3} = 4$ |
| $Y_{s_4} = 0$ | $Y_{s_5} = 2$ | $Y_{s_6} = 3$ |
| $Y_{s_7} = 7$ | $Y_{s_8} = 9$ | $Y_{s_9} = 3$ |

## Función Indicadora: Si $X \geq M_e^x \rightarrow 1$, $Y \geq M_e^y \rightarrow 1$, en otro caso$\rightarrow 0$

| | | |
|---|---|---|
| $\tau_{s_1} = 1$ | $\tau_{s_2} = 0$ | $\tau_{s_3} = 1$ |
| $\tau_{s_4} = 1$ | $\tau_{s_5} = 0$ | $\tau_{s_6} = 1$ |
| $\tau_{s_7} = 0$ | $\tau_{s_8} = 0$ | $\tau_{s_9} = 1$ |

| | | |
|---|---|---|
| $\tau_{s_1} = 1$ | $\tau_{s_2} = 0$ | $\tau_{s_3} = 1$ |
| $\tau_{s_4} = 0$ | $\tau_{s_5} = 0$ | $\tau_{s_6} = 1$ |
| $\tau_{s_7} = 1$ | $\tau_{s_8} = 1$ | $\tau_{s_9} = 1$ |

## Supongamos $m = 4$, entonces el $4-$*entorno* de X e Y para $s_1$ es:

## $x_4(s_1) = (\tau_{s_1}, \tau_{s_2}, \tau_{s_4}, \tau_{s_5})$ y $y_4(s_1) = (\tau_{s_1}, \tau_{s_2}, \tau_{s_4}, \tau_{s_5})$ respectivamente

$x_4(s_1) = (1, 0, 1, 0)$

| | | |
|---|---|---|
| $\tau_{s_1} = 1$ | $\tau_{s_2} = 0$ | $\tau_{s_3} = 1$ |
| $\tau_{s_4} = 1$ | $\tau_{s_5} = 0$ | $\tau_{s_6} = 1$ |
| $\tau_{s_7} = 0$ | $\tau_{s_8} = 0$ | $\tau_{s_9} = 1$ |

$y_4(s_1) = (1, 0, 0, 0)$

| | | |
|---|---|---|
| $\tau_{s_1} = 1$ | $\tau_{s_2} = 0$ | $\tau_{s_3} = 1$ |
| $\tau_{s_4} = 0$ | $\tau_{s_5} = 0$ | $\tau_{s_6} = 1$ |
| $\tau_{s_7} = 1$ | $\tau_{s_8} = 1$ | $\tau_{s_9} = 1$ |

## Función Simbolizadora: El símbolo ($\sigma_{s_1}$) captura el número de vecinos que tienen igual valor que la localización $s_1$

$$x_4(s_1) = (1,0,1,0) \rightarrow \sigma^x_{s_1} = 1$$

| | | |
|---|---|---|
| $\tau_{s_1} = 1$ | $\tau_{s_2} = 0$ | $\tau_{s_3} = 1$ |
| $\tau_{s_4} = 1$ | $\tau_{s_5} = 0$ | $\tau_{s_6} = 1$ |
| $\tau_{s_7} = 0$ | $\tau_{s_8} = 0$ | $\tau_{s_9} = 1$ |

$$y_4(s_1) = (1,0,0,0) \rightarrow \sigma^y_{s_1} = 0$$

| | | |
|---|---|---|
| $\tau_{s_1} = 1$ | $\tau_{s_2} = 0$ | $\tau_{s_3} = 1$ |
| $\tau_{s_4} = 0$ | $\tau_{s_5} = 0$ | $\tau_{s_6} = 1$ |
| $\tau_{s_7} = 1$ | $\tau_{s_8} = 1$ | $\tau_{s_9} = 1$ |

# Función Simbolizadora: El símbolo ($\sigma_{s_2}$) captura el número de vecinos que tienen igual valor que la localización $s_2$

$x_4(s_2) = (0,1,1,0) \rightarrow \sigma^x_{s_2} = 1$

| | | |
|---|---|---|
| $\tau_{s_1} = 1$ | $\tau_{s_2} = 0$ | $\tau_{s_3} = 1$ |
| $\tau_{s_4} = 1$ | $\tau_{s_5} = 0$ | $\tau_{s_6} = 1$ |
| $\tau_{s_7} = 0$ | $\tau_{s_8} = 0$ | $\tau_{s_9} = 1$ |

$y_4(s_1) = (0,1,1,0) \rightarrow \sigma^y_{s_2} = 1$

| | | |
|---|---|---|
| $\tau_{s_1} = 1$ | $\tau_{s_2} = 0$ | $\tau_{s_3} = 1$ |
| $\tau_{s_4} = 0$ | $\tau_{s_5} = 0$ | $\tau_{s_6} = 1$ |
| $\tau_{s_7} = 1$ | $\tau_{s_8} = 1$ | $\tau_{s_9} = 1$ |

## Etapa Final: Distribución de Símbolos

| | | |
|---|---|---|
| $\sigma^x_{s_1}=1$ | $\sigma^x_{s_2}=1$ | $\sigma^x_{s_3}=1$ |
| $\sigma^x_{s_4}=1$ | $\sigma^x_{s_5}=1$ | $\sigma^x_{s_6}=2$ |
| $\sigma^x_{s_7}=2$ | $\sigma^x_{s_8}=2$ | $\sigma^x_{s_9}=1$ |

| | | |
|---|---|---|
| $\sigma^y_{s_1}=0$ | $\sigma^y_{s_2}=1$ | $\sigma^y_{s_3}=1$ |
| $\sigma^y_{s_4}=1$ | $\sigma^y_{s_5}=2$ | $\sigma^y_{s_6}=2$ |
| $\sigma^y_{s_7}=1$ | $\sigma^y_{s_8}=2$ | $\sigma^y_{s_9}=2$ |

Finalmente, no se necesita la distribución en el espacio de los símbolos. Cada símbolo captura la información espacial para cada localización.

## Conceptos Básicos de Entropía

La **Entropía** es una medida de incertidumbre asociada con la distribución de una variable aleatoria.

Definición 1: La entropía de Shannon, $h(x)$, de una variable aleatoria discreta $x$ es: $h(x) = -\sum_{i=1}^{n} p(x_i)\, ln(p(x_i))$.

Definición 2: La entropía condicional $h(x|y)$ con distribución $p(x,y)$ es definida como:
$h(x|y) = -\sum_{x}\sum_{y} p(x,y)\, ln(p(x|y))$.

# Medida de Información Simbólica

En el caso de var. discretas se define la **entropía simbólica**:

$$h_x(m) = -\sum_{\sigma \in \Gamma} p_\sigma ln(p_\sigma).$$

con $p_\sigma$ probab. del símbolo $\sigma$. aproximada por la frec. relativa de los símbolos ($\hat{p}_\sigma \approx p_\sigma$).
La **entropía simbólica condicional** de $x_s$ dada $y_s$:

$$h_{x|y}(m) = \sum_{\sigma^y \in \Gamma} p(\sigma^y) h_{x|\sigma^y}(m)$$

Esta medida cuantifica la entropía remanente de $x_s$ dados los valores $y_s$ (en téminos simbólicos).

# Test de Causalidad Espacial

Causalidad en el espacio es entendida como "unidireccionalidad en la información" (Herrera et al., 2016).
Diremos que $\{x_s\}_{s\in S}$ *no causa* a $\{y_s\}_{s\in S}$ *bajo una estructura espacial* $X_{\mathcal{W}}$ *e* $Y_{\mathcal{W}}$ *si*

$$h_{y|Y_{\mathcal{W}}}(m) = h_{y|Y_{\mathcal{W}},X_{\mathcal{W}}}(m).$$

con el siguiente test unilateral:

$$\hat{\delta}(Y_{\mathcal{W}}, X_{\mathcal{W}}) = \hat{h}_{y|Y_{\mathcal{W}}}(m) - \hat{h}_{y|Y_{\mathcal{W}},X_{\mathcal{W}}}(m)$$

Esto es, si $X_{\mathcal{W}}$ no contiene información adicional sobre $Y$:
$\hat{\delta}(Y_{\mathcal{W}}, X_{\mathcal{W}})=0$, en otro caso, $\hat{\delta}(Y_{\mathcal{W}}, X_{\mathcal{W}}) > 0$.
Los valores críticos son obtenidos por bootstrap en bloques.

# Ejemplo: desempleo regional y migración neta

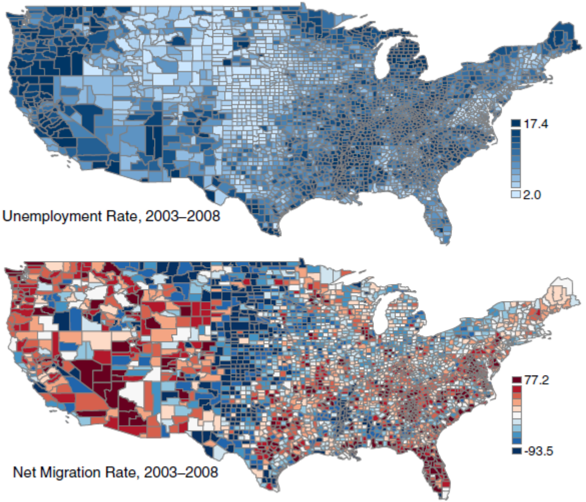



**Fig. 3.** Spatial distribution of unemployment and net migration by county. Average 2003–2008

# Resultados de cortes transversales (Papers in Regional Science, 2016)

**Table 6.** Results of spatial causality test

| $H_0$ | Unemploment $\nRightarrow$ Migration | Migration $\nRightarrow$ Unemploment | Conclusion |
|---|---|---|---|
| Periods | $p$ – value | $p$ – value | |
| 2003–2005 | 0.100 | 0.045 | *Migr.* $\Rightarrow$ *Unemp.* |
| 2004–2006 | 0.882 | 0.048 | *Migr.* $\Rightarrow$ *Unemp.* |
| 2005–2007 | 0.827 | 0.049 | *Migr.* $\Rightarrow$ *Unemp.* |
| 2006–2008 | 0.812 | 0.047 | *Migr.* $\Rightarrow$ *Unemp.* |
| 2003–2008 | 0.652 | 0.050 | *Migr.* $\Rightarrow$ *Unemp.* |

*Notes*: '$\nRightarrow$' means does not cause and '$\Rightarrow$' means causes. $m = 9$.
Boots: 399, Blocks: 42 (74 observations by block).

*GRACIAS POR SU ATENCIÓN!*