PROBIT MODELS FOR GROUPED-DATA MIGRATION FLOWS: A THEORETICAL NOTE

Coro Chasco

(Autonomous University of Madrid – UAM, Spain)

Luc Anselin

(University of Chicago, USA)

Patricio Aroca

(Adolfo Ibáñez University, Chile)

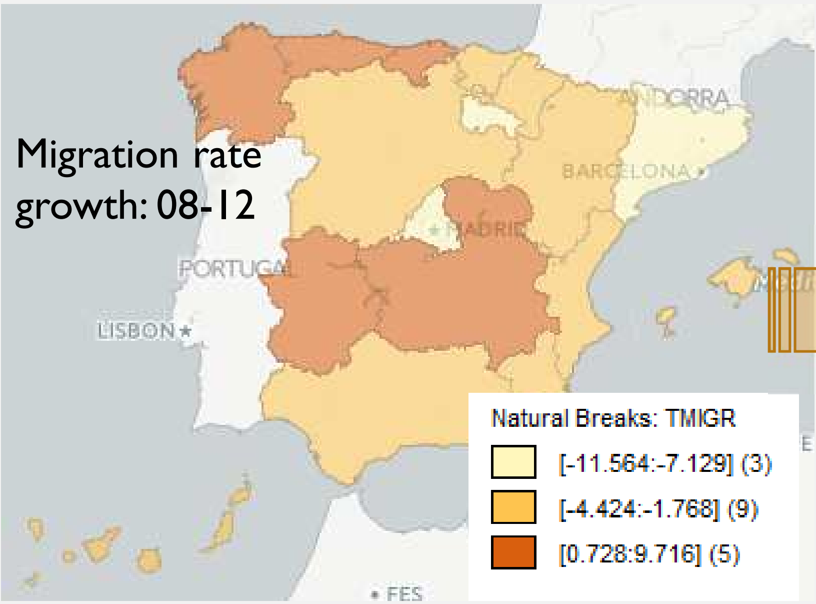
MOTIVATION

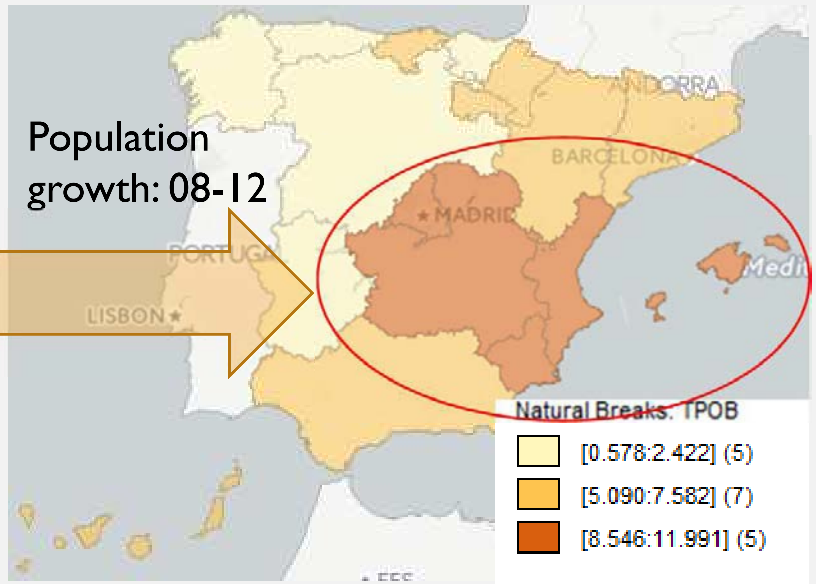
- FIRST MOTIVATION: Interested in internal migration: within one geopolitical entity (i.e. Spain), people travel for improving education and economic conditions
- We are interested in **Spain**:

  + Since the 80's (urban sprawl): lower internal migration rates.

  + Slowly but surely: progressive pop. desertification in the inland & NW.

  + Where are these people going? And why (only education/economy)?





MOTIVATION #ii

- SECOND MOTIVATION: Interested in the formulation of migration models.
- Migration modeling has been applied to both, micro and macro-levels (*Aleshkovski and Iontsev 2006*):

- Micro approach focuses on the migration behavior of individuals or households based on disaggregated data usually delivered by surveys. They are costly to collect or inaccessible. Tool: Discrete choice models.

- Macro approach studies the patterns of migration of certain social groups within a given territory. Choice data is aggregated across groups of individuals in the form of counts or shares. Easier to obtain. Tool: Gravity or interaction models.

- Our database follows a macro approach =

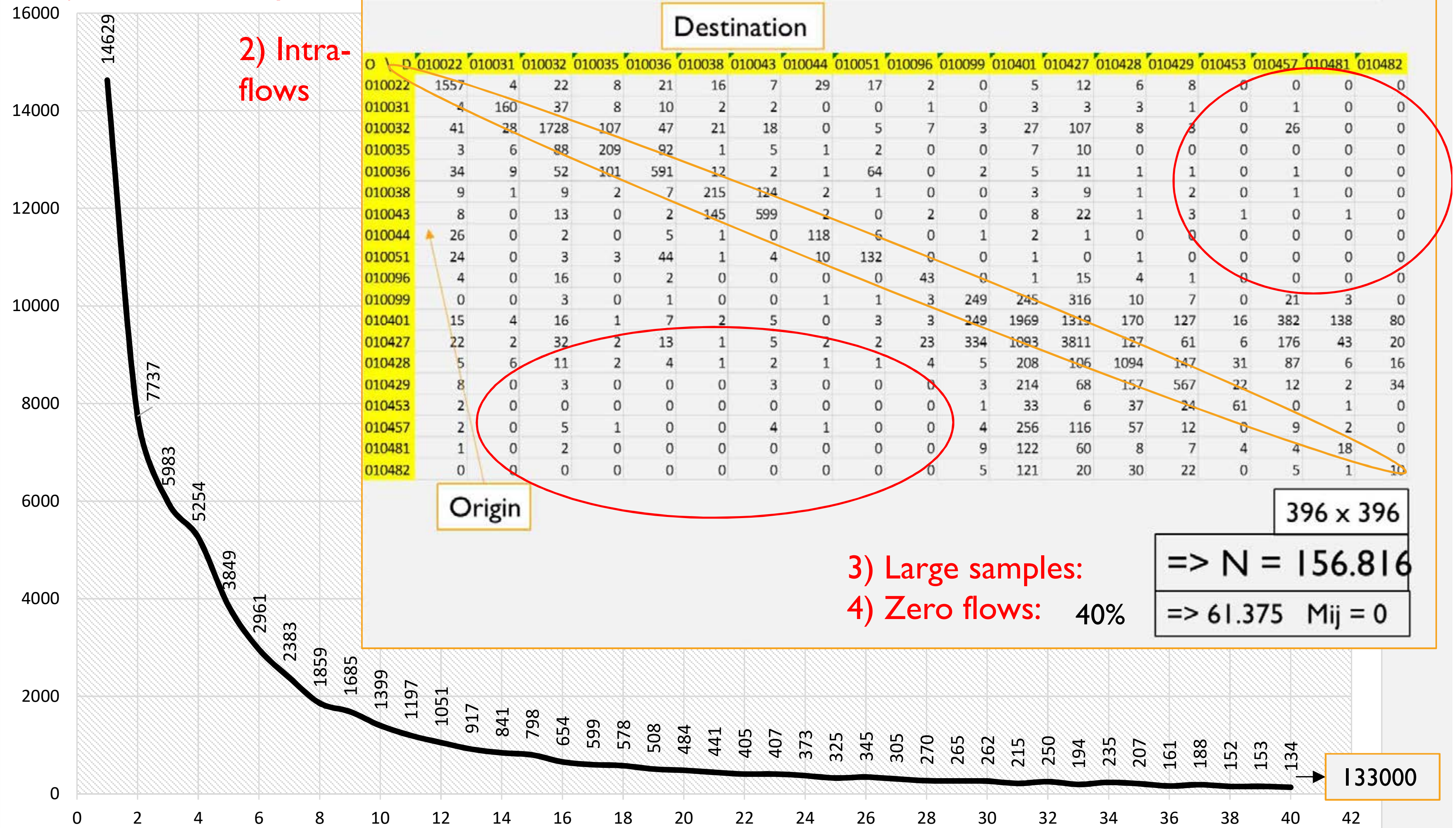
We propose a PROBIT CHOICE MODEL but for GROUPED-DATA flows, due to some important specification problems of the standard spatial interaction models of flows (*LeSage and Fischer 2010*).

MOTIVATION #iii

Example: Spanish NUTs4 flows

1) Non-normality

2) Intra-flows



Distribution of the Spanish NUTs4 flows

Destination

O \ D	010022	010031	010032	010035	010036	010038	010043	010044	010051	010096	010099	010401	010427	010428	010429	010453	010457	010481	010482
010022	1557	4	22	8	21	16	7	29	17	2	0	5	12	6	8	0	0	0	0
010031	4	160	37	8	10	2	2	0	0	1	0	3	3	3	1	0	1	0	0
010032	41	28	1728	107	47	21	18	0	5	7	3	27	107	8	3	0	26	0	0
010035	3	6	88	209	92	1	5	1	2	0	0	7	10	0	0	0	0	0	0
010036	34	9	52	101	591	12	2	1	64	0	2	5	11	1	1	0	1	0	0
010038	9	1	9	2	7	215	124	2	1	0	0	3	9	1	2	0	1	0	0
010043	8	0	13	0	2	145	599	2	0	2	0	8	22	1	3	1	0	1	0
010044	26	0	2	0	5	1	0	118	6	0	1	2	1	0	0	0	0	0	0
010051	24	0	3	3	44	1	4	10	132	0	0	1	0	1	0	0	0	0	0
010096	4	0	16	0	2	0	0	0	0	43	0	1	15	4	1	0	0	0	0
010099	0	0	3	0	1	0	0	1	1	3	249	245	316	10	7	0	21	3	0
010401	15	4	16	1	7	2	5	0	3	3	249	1969	1319	170	127	16	382	138	80
010427	22	2	32	2	13	1	5	2	2	23	334	1093	3811	127	61	6	176	43	20
010428	5	6	11	2	4	1	2	1	1	4	5	208	106	1094	147	31	87	6	16
010429	8	0	3	0	0	0	3	0	0	0	3	214	68	157	567	22	12	2	34
010453	2	0	0	0	0	0	0	0	0	0	1	33	6	37	24	61	0	1	0
010457	2	0	5	1	0	0	4	1	0	0	4	256	116	57	12	0	9	2	0
010481	1	0	2	0	0	0	0	0	0	0	9	122	60	8	7	4	4	18	0
010482	0	0	0	0	0	0	0	0	0	0	5	121	20	30	22	0	5	1	10

Origin

396 x 396

3) Large samples: => N = 156.816

4) Zero flows: 40% => 61.375 Mij = 0

CONTENTS

- Spatial interaction models
  - Data configuration
  - Model specification and estimation
  - Model problems
- GProbit: an alternative approach for OD flow models
  - Model specifications: for independent and spatially correlated flows
  - Solutions for the spatial interaction model problems
- Application for migration interregional flows across NUTS 2 in Spain (2008 – 2012)
- Conclusiones

I. SPATIAL INTERACTION MODELS

I.I. Data configuration # i

JOURNAL OF REGIONAL SCIENCE

SPATIAL ECONOMETRIC MODELING OF ORIGIN-DESTINATION FLOWS

James P. LeSage, R. Kelley Pace

Volume 48, Issue 5
December 2008
Pages 941–967

Origin-Destination Notation and Ordering

Let $\mathbf{Y}$ denote an n by n square matrix of interregional flows from each of the n origin regions to each of the n destination regions where the n columns represent different origins and the n rows represent different destinations as shown in (1). The flows considered here reflect a closed system.

$$
\begin{array}{cc}
 & \begin{array}{cccc} o_1 & o_2 & \cdots & o_n \end{array} \\
\begin{array}{c} d_1 \\ d_2 \\ \vdots \\ d_n \end{array} &
\begin{pmatrix}
o_1 \to d_1 & o_2 \to d_1 & \cdots & o_n \to d_1 \\
o_1 \to d_2 & o_2 \to d_2 & \cdots & o_n \to d_2 \\
\vdots & \vdots & \ddots & \vdots \\
o_1 \to d_n & o_2 \to d_n & \cdots & o_n \to d_n
\end{pmatrix}
\end{array}.
$$

I. SPATIAL INTERACTION MODELS

I.I. Data configuration # ii

Ann Reg Sci (2002) 36:197–218

The Annals of Regional Science

Migration and regional labor market adjustment: Chile 1977–1982 and 1987–1992

Patricio Aroca[1], **Geoffrey J.D. Hewings**[2]

[1] IDEAR, Universidad Católica del Norte, Angamos 0610, Antofagasta, CHILE
[2] REAL, University of Illinois, South Mathews Av. 607, Urbana 61801-3671, IL, USA

Received: January 1999/Accepted: August 2001

Regional Science and Urban Economics 23 (1993) 211–233. North-Holland

Does migration arbitrage regional labor market differentials?

Stuart A. Gabriel
University of Southern California, Los Angeles, CA 90089, USA

Janice Shack-Marquez and William L. Wascher*
Board of Governors of the Federal Reserve System, Washington, DC 20551, USA

Table A.2. Data for migrant between 1987–1992

i/j	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	RM	
I	184217	4640	1404	2078	3552	487	568	1680	420	528	79	122	8904	208679
II	5460	229630	3258	5081	2770	529	571	1604	387	513	75	143	8229	258250
III	1301	3703	123291	5078	1626	351	217	379	126	179	42	43	3408	139744
IV	2788	6337	5995	282952	4055	586	372	601	226	391	74	96	7947	312420
V	3919	2726	2342	3423	797336	2591	1663	5878	1134	2580	378	2887	28439	855296
VI	494	616	543	765	3318	403591	3169	2182	800	927	103	617	19727	436852
VII	616	528	439	492	2546	4205	491351	5383	1489	1223	129	1264	25704	535369
VIII	1705	1490	786	834	7294	3611	4598	1032149	7988	4171	694	2182	40074	1107576
IX	521	449	318	297	1596	2447	1410	7289	435078	5742	349	367	23703	479566
X	679	544	358	468	2912	1321	1441	4134	6992	548567	2223	3177	21043	593859
XI	90	52	45	89	319	128	213	574	620	1984	40939	252	1535	46840
XII	263	220	81	147	3270	395	435	2040	755	2504	174	76945	4455	91684
RM	10302	7392	4180	7480	27011	14209	12588	19064	12028	13877	1763	3096	3079528	3212518
	212355	258327	143040	309184	857605	434451	518596	1082957	468043	583186	47022	91191	3272696	8278653

II. SPATIAL INTERACTION MODELS

I.I. Data configuration # iii

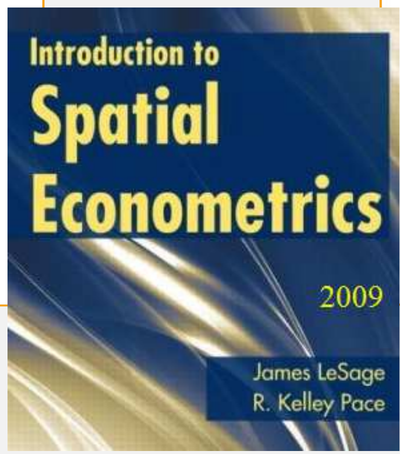
origin-centric ordering

$l^{(o)}$	$o^{(o)}$	$d^{(o)}$
1	1	1
$\vdots$	$\vdots$	$\vdots$
n	1	n
$\vdots$	$\vdots$	$\vdots$
$N-n+1$	n	1
$\vdots$	$\vdots$	$\vdots$
N	n	$n,$

$N=n^2$

destination-centric ordering

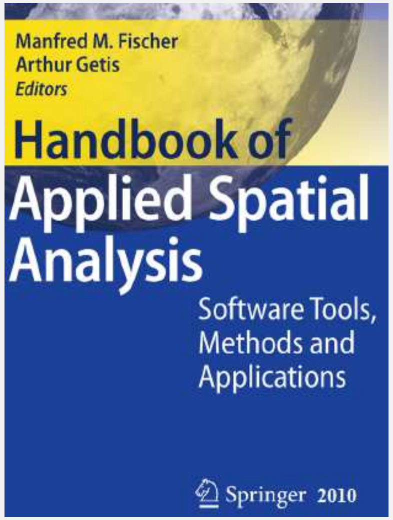
$l^{(d)}$	$o^{(d)}$	$d^{(d)}$
1	1	1
$\vdots$	$\vdots$	$\vdots$
n	n	1
$\vdots$	$\vdots$	$\vdots$
$N-n+1$	1	n
$\vdots$	$\vdots$	$\vdots$
N	n	$n.$



Chapter 8

Spatial Econometric Interaction Models

1. SPATIAL INTERACTION MODELS
1.2. Model specification and estimation



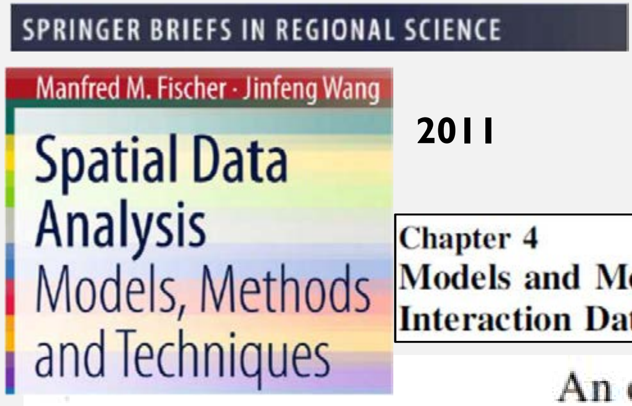
C.3 Spatial Econometric Methods for Modeling Origin-Destination Flows

James P. LeSage and *Manfred M. Fischer*

$$y = \alpha \iota_n + X_o \beta + X_d \gamma + \theta d + \varepsilon$$

log(counts)

y	N-by-1 vector of origin-destination flows,
X_o	N-by-Q matrix of Q origin-specific variables that characterize the ability of the origin zones to produce flows,
β	the associated Q-by-1 parameter vector that reflects the origin effects,
X_d	N-by-R matrix of R destination-specific variables that represent the attractiveness of the destination zones,
γ	the associated R-by-1 parameter vector that reflects the destination effects,
d	N-by-1 vector of distances between origin and destination zones,
θ	scalar distance sensitivity parameter that comes from the power deterrence function and reflects the distance effects,
ι_n	N-by-1 vector of ones,
α	constant term parameter on ι_n,
ε	N-by-1 vector of disturbances with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$.



2011

Chapter 4
Models and Methods for Spatial Interaction Data

1. SPATIAL INTERACTION MODELS

1.2. Model specification and estimation #ii

An origin-centric scheme for origin–destination flow arrangements

Dyad label	ID origin	ID destination	Flows	Origin variables X_o	Destination variables X_d	Distance variable
1	1	1	$y(1,1)$	$A_1(1)\ldots A_Q(1)$	$B_1(1)\ldots B_R(1)$	$D(1,1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots$	$\vdots \quad \vdots$	$\vdots$
n	1	n	$y(1,n)$	$A_1(1)\ldots A_Q(1)$	$B_1(n)\ldots B_R(n)$	$D(1,n)$
$n+1$	2	1	$y(2,1)$	$A_1(2)\ldots A_Q(2)$	$B_1(1)\ldots B_R(1)$	$D(2,1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots$	$\vdots \quad \vdots$	$\vdots$
$2n$	2	n	$y(2,n)$	$A_1(2)\ldots A_Q(2)$	$B_1(n)\ldots B_R(n)$	$D(2,n)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots$	$\vdots \quad \vdots$	$\vdots$
$N-n+1$	n	1	$y(n,1)$	$A_1(n)\ldots A_Q(n)$	$B_1(1)\ldots B_R(1)$	$D(n,1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots$	$\vdots \quad \vdots$	$\vdots$
N	n	n	$y(n,n)$	$A_1(n)\ldots A_Q(n)$	$B_1(n)\ldots B_R(n)$	$D(n,n)$

1. SPATIAL INTERACTION MODELS

1.2. Model specification and estimation #iii

Using LeSage & Pace (2008)

Spatial Econometric OD-Flow Models 83

Christine Thomas-Agnan and James P. LeSage

$$y = \alpha \iota_{n^2} + X_o \beta_o + X_d \beta_d + \gamma g + \varepsilon$$

$$Z = \begin{pmatrix} \iota_{n^2} & X_o & X_d & g \end{pmatrix}$$

$$Z'Z = \begin{pmatrix} n^2 & (1,1) & 0_k & (1,k) & 0_k & (1,k) & \iota_n' G \iota_n & (1,1) \\ 0_k' & (k,1) & nX'X & (k,k) & 0_k' 0_k & (k,k) & X' G \iota_n & (k,1) \\ 0_k' & (k,1) & 0_k' 0_k & (k,k) & nX'X & (k,k) & X' G \iota_n & (k,1) \\ \iota_n' G \iota_n & (1,1) & \iota_n' G' X & (1,k) & \iota_n' G' X & (1,k) & tr(G^2) & (1,1) \end{pmatrix}$$

It avoids the "N"

$$Z'y = \begin{pmatrix} \iota_n' Y \iota_n \\ X' Y \iota_n \\ X' Y' \iota_n \\ \text{tr}(GY) \end{pmatrix}$$

$$\delta = (\alpha \; \beta_o \; \beta_d \; \gamma)'$$

OLS

$$\boxed{\hat{\delta} = \left[(1/n^2) Z'Z\right]^{-1} (1/n^2) Z'y}$$

vector of logged distances g.

G represents the $n \times n$ matrix of logged distances.

1. ORIGIN-DESTINATION FLOW MODELS

1.3. Model problems

1) Non-normality of count-data of migration flows:

Basic assumption for OLS inference and ML estimation

- Count data of flows → Poisson distribution. **Log(counts) ≠ log-normal** (*Fischer & Wang 2011*).
- Large number of zero flows. In log-normal models: **log(0)=?**.

  log(y+1) introduce extra flows in the model=downward bias in the OLS estimators (*LeSage & Fischer 2010*).

2) Intra-regional flows:

Of different nature than interregional flows.

Intra-flows quite larger and determined by different explanatory variables (*LeSage & Pace 2008*).

- When known: specification of a (less-parsimonious) nested model for intra-flows.
- When unknown: estimation (e.g. *Tsutsumi and Tamesue 2012*).

THIS PAPER PROPOSAL: IMPROVING THE SPATIAL INTERACTION MODEL

II. GPROBIT: AN ALTERNATIVE
II.1. Specification

GProbit = Probit choice model for grouped-data flows.

Theoretical foundation: Random utility theory for aggregations of decisions (probabilities) made by individuals who share a similar characteristic; e.g. living in a same region.

Individual:

$$P(y=1) = P(y^* \geq 0) = P(U_{od} \geq U_{oo})$$
$$y^* = U_{od} - U_{oo} = x'\theta + u$$

Adding up the independent probabilities for all the individuals who move from *o* to *d*.

$$P_{od} = \frac{M_{od}}{R_o} \xrightarrow{\text{Each of the group components } \to \infty} P_{od} = \pi_{od} + u_{od}$$

$R_o = M_{od} + M_{oo}$
M_{oo}='stayers'+intra-flows

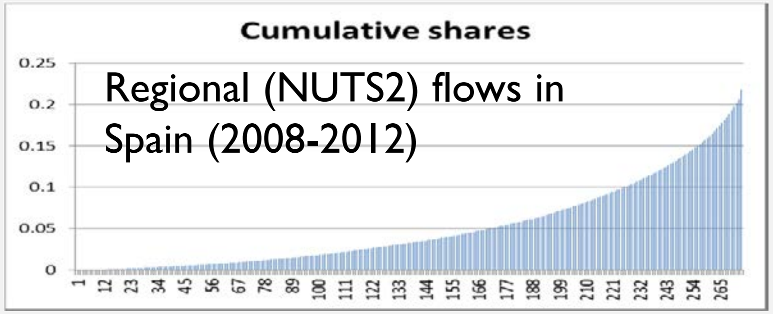
$$\pi_{od} = F\left(x'_{od}\beta\right)$$

Theoretical proportion

Share, proportion (relative frequency) of people who migrate from *o* to *d* during a certain period (M_{od}) over the total resident population living in o *'at risk' of migrating* during this same period (R_o).
'Meaningful estimates of interaction probabilities between OD pairs' (*Sen and Smith 1993*)

II. GPROBIT: AN ALTERNATIVE

II.1. Specification #ii



$$P_{od} = \pi_{od} + u_{od}$$

$$\pi_{od} = F\left(x'_{od}\beta\right)$$

$$\boxed{P_{od} = F\left(x'_{od}\beta\right) + u_{od}}$$

Non-linear GProbit model of flows

$$\boxed{P_{od} = \Phi\left(x'_{od}\beta\right) + u_{od}}$$

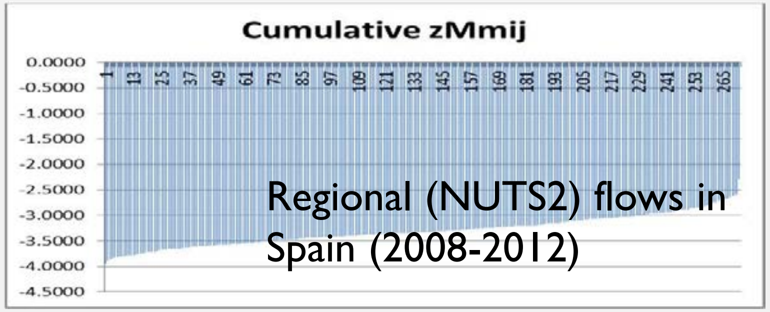
Can be linearized:

(*Gourrieroux 2000*, section 4.2):
Slutsky's theorem on convergence in probability + Large number of group shares

$$\boxed{Z_{od} = \Phi^{-1}(P_{od}) = \alpha\iota_N + X_d\beta_d + X_o\beta_o + \lambda D + \varepsilon_{od}}$$

Linear function GProbit model of flows

Dependent variable:
Inverse of the cumulative standard normal distribution of P_{od}



II. GPROBIT: AN ALTERNATIVE
II.1. Specification #iii

$$\boxed{Z_{od} = \Phi^{-1}(P_{od}) = \alpha\iota_N + X_d\beta_d + X_o\beta_o + \lambda D + \varepsilon_{od}}$$

Linear function Gprobit model of flows

Equal to the spatial interaction model except the dependent variable.

$$N(0,\Omega)$$

Linear estimation models can be applied: OLS, ML, 2SLS, GMM…

$$\Omega \rightarrow \begin{cases} \sigma_\varepsilon^2 = \dfrac{P_{od}(1-P_{od})}{R_o \cdot \left[\varphi\left[\Phi^{-1}(P_{od})\right]\right]^2} \longrightarrow \text{Heteroskedasticity} \\ \operatorname{cov}(\varepsilon_{od}) \neq 0 \longrightarrow \text{Spatial autocorrelation} \end{cases}$$

Heteroskedasticity: Varies with each OD flow pair. Other spatial causes

Spatial autocorrelation

Solutions:
1) Berkson's Min-χ^2 for this structure.
2) Unmodeled for unknown structure: robust inference

Spatial lag:

$$Z_{od} = \boxed{\rho_d W_d Z_{od} + \rho_o W_o Z_{od} + \rho_\omega W_\omega Z_{od}} + \alpha\iota_N + X_d\beta_d + X_o\beta_o + \lambda D + \varepsilon_{od}$$

Spatial error:

$$\begin{cases} Z_{od} = \alpha\iota_N + X_d\beta_d + X_o\beta_o + \lambda D + u_{od} \\ u_{od} = \boxed{\rho_d W_d u_{od} + \rho_o W_o u_{od} + \rho_\omega W_\omega u_{od}} + \varepsilon_{od} \end{cases}$$

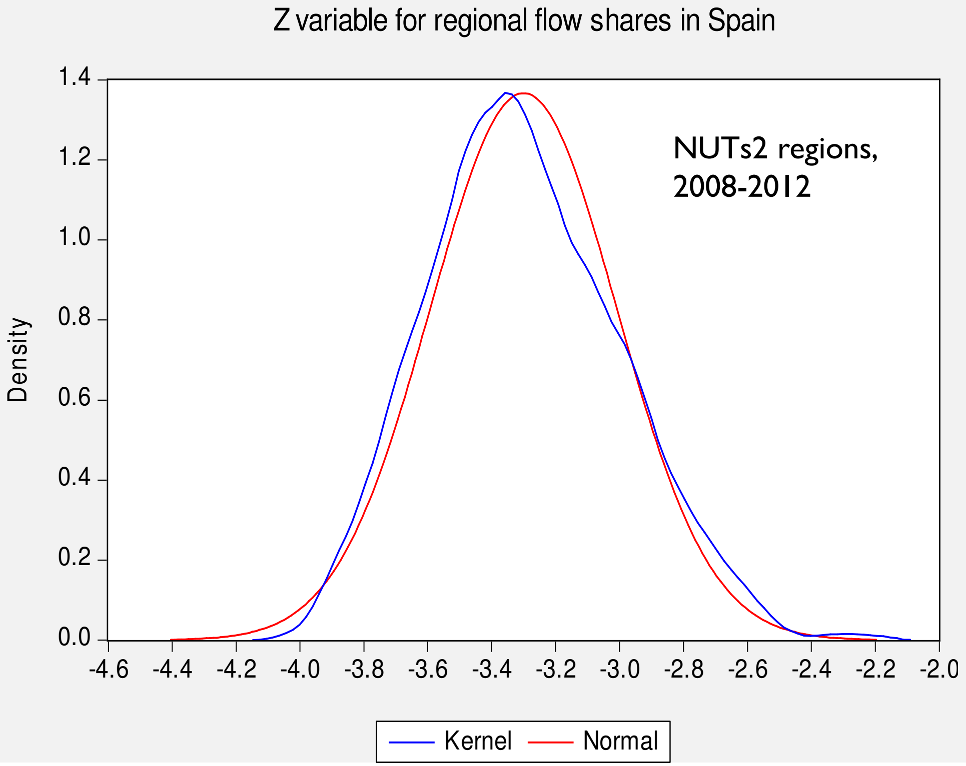
Spatial cross-regressive

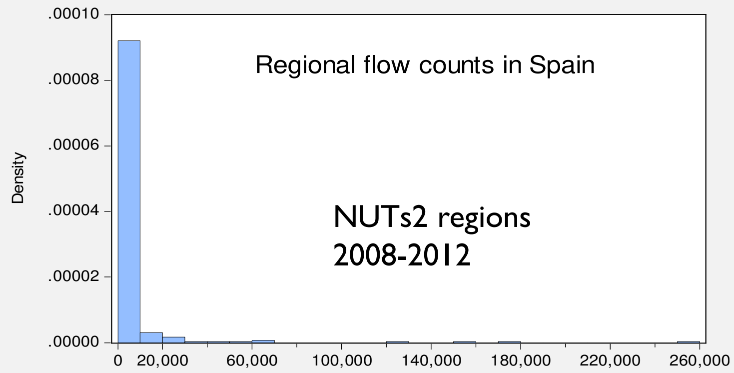
$$Z_{od} = \alpha\iota_N + X_d\beta_d + X_o\beta_o + \lambda D + \boxed{W_d X_d\theta_d + W_o X_o\theta} + \varepsilon_{od}$$

II. GPROBIT: AN ALTERNATIVE
II.2. Solutions for model problems #ii

Problem	Spatial interaction model	Gprobit model for OD flows
Non-normality of count-data	Instead of counts, log(counts) (*very frequent in the literature*)	*y*=z: inverse cumulative standard normal distribution of flow shares ¡

Dependent variable (Z_{od}): inverse cumulative standard **normal** distribution of flow shares. Normality is assumed.

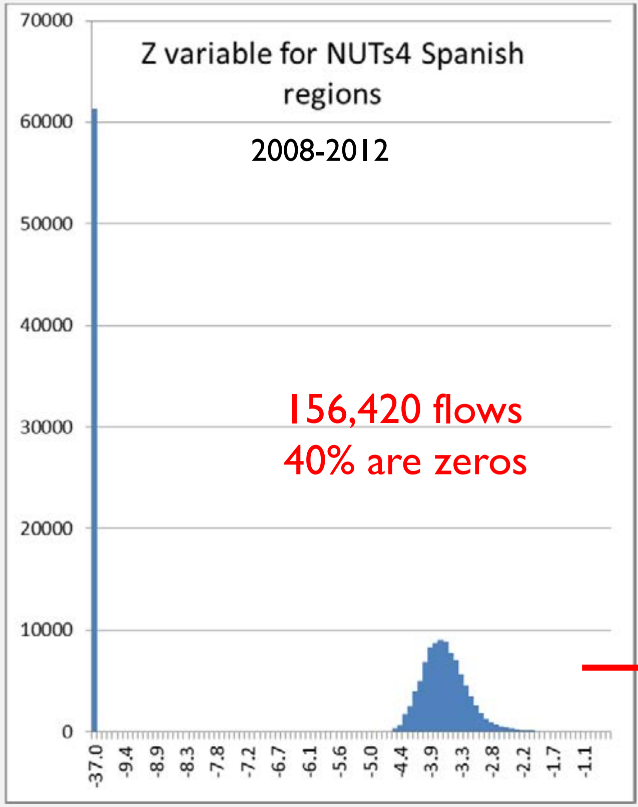




Shares are preferable (relative): Flows (e.g. 10) can be originated from populations of different sizes (e.g. 100 and 1000, implying that people at those regions have different 'propensity to migrate' (0.1 and 0.01, respectively

II. GPROBIT: AN ALTERNATIVE
II.2. Solutions for model problems #iii

Problem	Spatial interaction model	Gprobit model for OD flows
Zero flows	Instead of log(0), log(1+0) (*LeSage & Pace 2008*)	z's domain is 0-1 (zero is included) i



Zero is *theoretically* part of the domain Z values, because it is part of the shares (P_{od}):

$$\boldsymbol{Z = \Phi^{-1}(P_{od})}$$
Domain = [0, 1]

In empirical apps. (*STATA 2017*), in order to linearize the model, the extreme values are:

$Z = \Phi^{-1}(P_{od}) \rightarrow$ Domain = $[10^{-323}, (1-2^{-53})]$.

Hence, the values of the dependent variable Z range from –38.449394 to 8.2095362.

Zeros are possible values for Z, but always problematic when presented largely in a variable.

II. GPROBIT: AN ALTERNATIVE

II.2. Solutions for model problems #iv

Problem	Spatial interaction model	Gprobit model for OD flows
Intra-regional flows ≠ Interregional flows	Estimation, nested model for intra-flows (*LeSage & Pace 2008*)	Z=function(Probability): **i** Intra-flows=1 – Sum(Inter-flows)

$$P_{od} = \frac{M_{od}}{R_o}$$

$R_o = M_{od} + M_{oo}$

M_{oo}='stayers'+intra-flows

P_{od}: Share, proportion (relative frequency) of M_{od} over R_o.

→ This model allows estimating the intra-flow proportions: the shares for each region *o* must sum up to one:

$$\sum_{d=1}^{n} \frac{m_{od}}{m_o} + \frac{m_{oo}}{m_o} = 1$$

Regions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	**170,118**	1,516	807	4,706	4,672	520	2,142	3,667	11,271	6,865	2,379	1,934	13,489	3,981	835	1,962	341
2	1,514	**21,230**	217	422	573	176	918	650	4,677	2,816	215	431	2,350	390	929	660	376
3	726	197	**17,486**	467	696	506	1,465	296	944	816	136	1,138	2,309	172	166	457	111
4	4,511	396	431	**29,659**	1,211	185	749	748	5,102	3,293	619	1,057	2,958	610	145	473	85
5	5,105	545	810	1,347	**62,873**	351	1,228	833	3,188	2,235	424	3,392	4,804	560	302	965	135
6	480	157	521	188	365	**14,936**	910	173	562	393	72	321	1,332	105	150	2,187	105
7	2,325	1,044	1,729	845	1,266	1,103	**52,062**	1,528	2,892	2,505	954	2,060	12,010	619	633	2,904	628
8	3,899	713	377	832	842	208	1,434	**36,210**	2,550	6,173	1,017	654	22,910	1,738	267	619	160
9	12,237	4,636	982	5,363	2,931	579	2,798	2,460	**257,534**	8,955	1,822	3,164	7,881	1,928	945	2,239	518
10	7,463	2,755	900	3,652	2,113	452	2,388	5,593	9,830	**128,348**	810	1,662	9,898	5,603	744	1,920	397
11	2,619	217	143	529	460	77	958	1,059	1,467	747	**13,911**	235	3,788	210	108	463	62
12	1,746	367	1,061	1,035	2,340	308	1,659	578	2,707	1,467	203	**63,437**	4,334	418	239	913	136
13	11,672	1,870	2,029	3,301	4,375	1,152	9,633	28,363	7,383	9,307	3,575	3,788	**154,569**	2,557	971	2,634	521
14	3,660	372	204	625	494	97	560	1,412	2,011	5,407	204	432	2,842	**22,266**	300	539	128
15	660	766	152	138	208	124	418	176	811	531	107	228	1,027	193	**17,821**	1,360	800
16	1,404	461	413	392	687	2,388	2,656	369	1,702	1,269	501	1,055	2,661	254	1,454	**44,968**	1,051
17	334	366	114	72	134	128	526	129	538	396	66	168	625	109	938	1,071	**5,577**

We can eliminate de n intra-flows and specify the GProbit model for the interregional flows only: N – n flows.

4. EMPIRICAL ILLUSTRATION FOR MIGRATION INTERREGIONAL FLOWS ACROSS NUTS 2 IN SPAIN (2008 – 2012) #i

The NUTs2 Spanish regions are only 17, generally large and rather self-sufficient.

Natural barriers: Central Mountain Chain (Madrid from Castile-Leon), Cantabrian Mountain Chain (Northern regions).

Natural connectors: Ebro Valley (from Bask C. to Catalonia) and Segura Basin (from Madrid to Valencian C.)

4. EMPIRICAL ILLUSTRATION FOR MIGRATION INTERREGIONAL FLOWS ACROSS NUTS 2 IN SPAIN (2008 – 2012) #ii

- We illustrate the performance of a GProbit model to estimate internal migration flows for the 17 NUTS 2 regions in Spain taken from the EVR register, INE.
- Flows: (emigrants from *o* to *d*) / total *o*'s in/out-emigrants).
- We compare the performance and results of this model with the gravitational model using the conventional log transformation of flows for the dependent variable.

4. EMPIRICAL ILLUSTRATION FOR MIGRATION INTERREGIONAL FLOWS ACROSS NUTS 2 IN SPAIN (2008 – 2012) #iii

Variable		Units	Source	Period
	Dependent variable:			
Mod	Migrant flow (5-year sum)	Persons	Spanish National Statistics Office	2008-2012
	Independent variables:			
Income and quality of life				
pibpc	GDP per capita	Euros	National Statistics Office (INE)	2003-2007
incpc	Household disposable income per capita	Euros	National Statistics Office (INE)	2003-2007
wage	Salaries and wages per capita	Euros	National Statistics Office (INE)	2003-2007
act	Activity rate growth	Percentage	National Statistics Office (INE)	2003-2007
Labor and housing markets				
emp	Population	Percentage	National Statistics Office (INE)	2003-2007
unem	Population	Percentage	National Statistics Office (INE)	2003-2007
pviv	Housing price	Euros	Ministry of Development of Spain	2003-2007
delin	People declaring having delinquency problems	Percentage	National Statistics Office (INE)	2003-2007
Agglomeration economies				
Pop	Population	Persons	National Statistics Office (INE)	2003-2007
dens	Population density	Persons per km^2	National Statistics Office (INE)	2003-2007
PPu	Urban population share*	Percentage	National Statistics Office (INE) and self-elaboration	2003-2007
pd3g	Population aged 25-64 with university degree	Percentage	National Statistics Office (INE)	2003-2007
rad	R&D expenditure per capita	Thou. euros	National Statistics Office (INE)	2003-2007
Natural endowments				
tmed	Annual average temperature	Degrees	State Meteorological Agency	2003-2007
tmax	Annual maximum temperature	Degrees	State Meteorological Agency	2003-2007
tmin	Annual minimum temperature	Degrees	State Meteorological Agency	2003-2007
sun	Sun hours	Hours	State Meteorological Agency	2003-2007
rain	Atmospheric precipitation	Millimeters	State Meteorological Agency	2003-2007
marit	Length of coastline (destination)	Km	National Geographic Institute	2003-2007
Distance:				
Dod	Origin – destination distance	Km	Self-elaboration with GIS	-
Tod	Origin – destination travel time	Minutes	Self-elaboration with Google Maps	-

. Data has been ordered according to the origin-centric scheme.

. Flows: emigrants from o to d / total people of o who have changed their residence during this period (including intra-regional movements).

. X: 'push' and 'pull' factors (ratio D/O values).

. D: log-transformed distance between the capital cities.

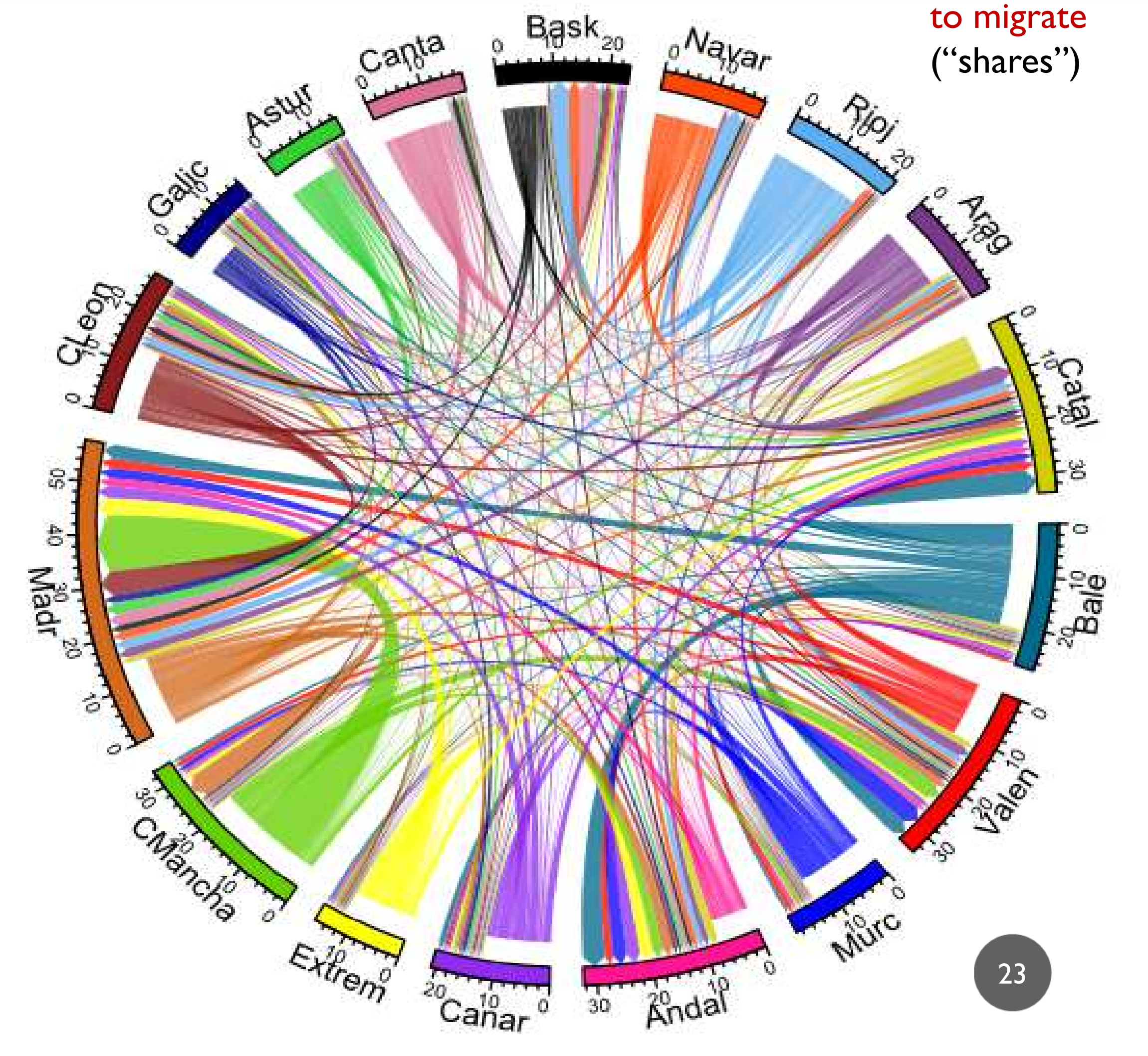
. In gravity model: log transformation of flows.

4. EMPIRICAL ILLUSTRATION FOR MIGRATION INTERREGIONAL FLOWS ACROSS NUTS 2 IN SPAIN (2008 – 2012) #iv

- Circular plot

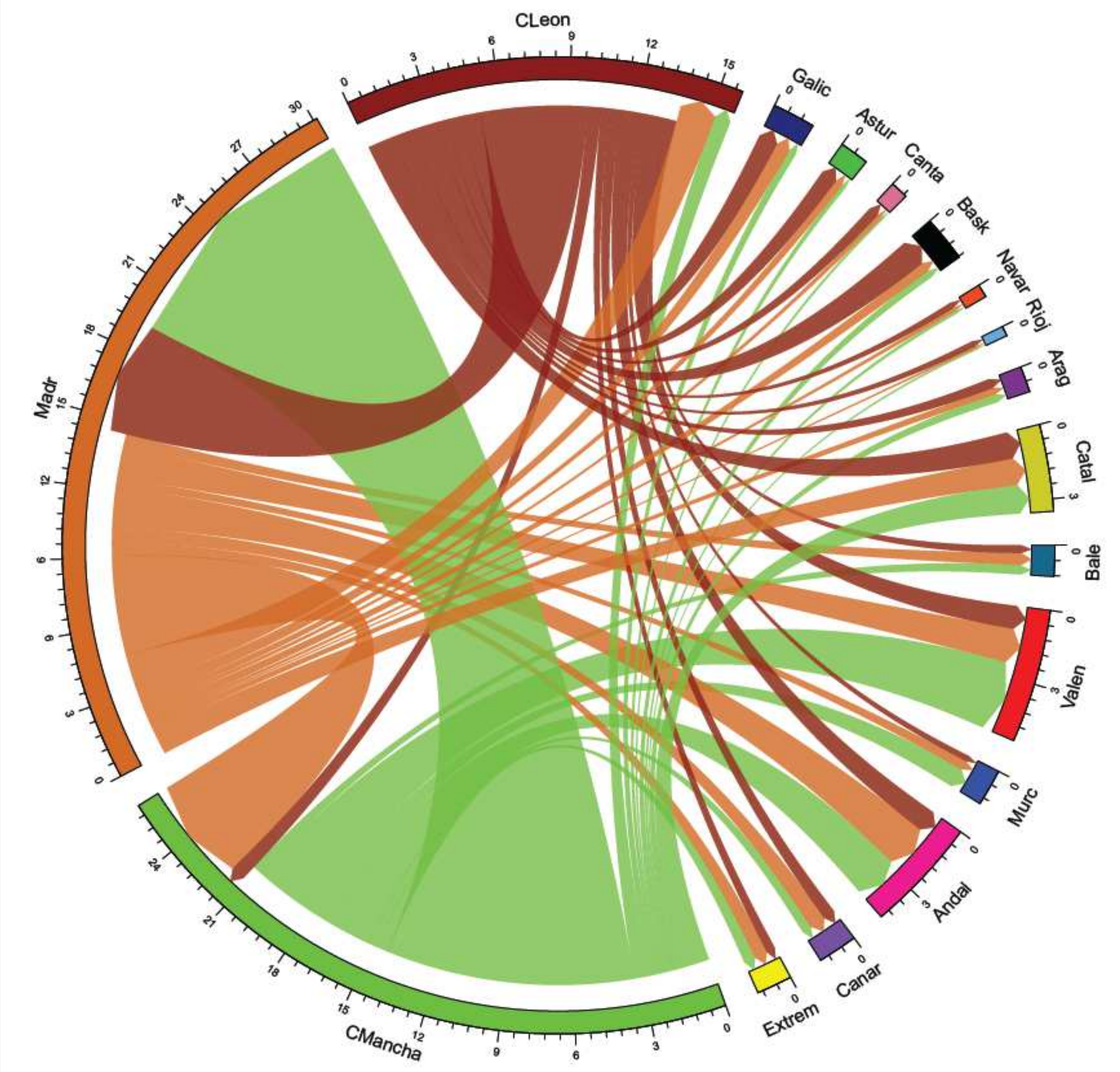
. It visualizes migrants' flows.
. Origins and destinations are each assigned a color.
.The volume of migration rate = width of the arrow.
.Tick marks on the circle segments show the migrant rate figures (inflows and outflows).

Propensity to migrate ("shares")

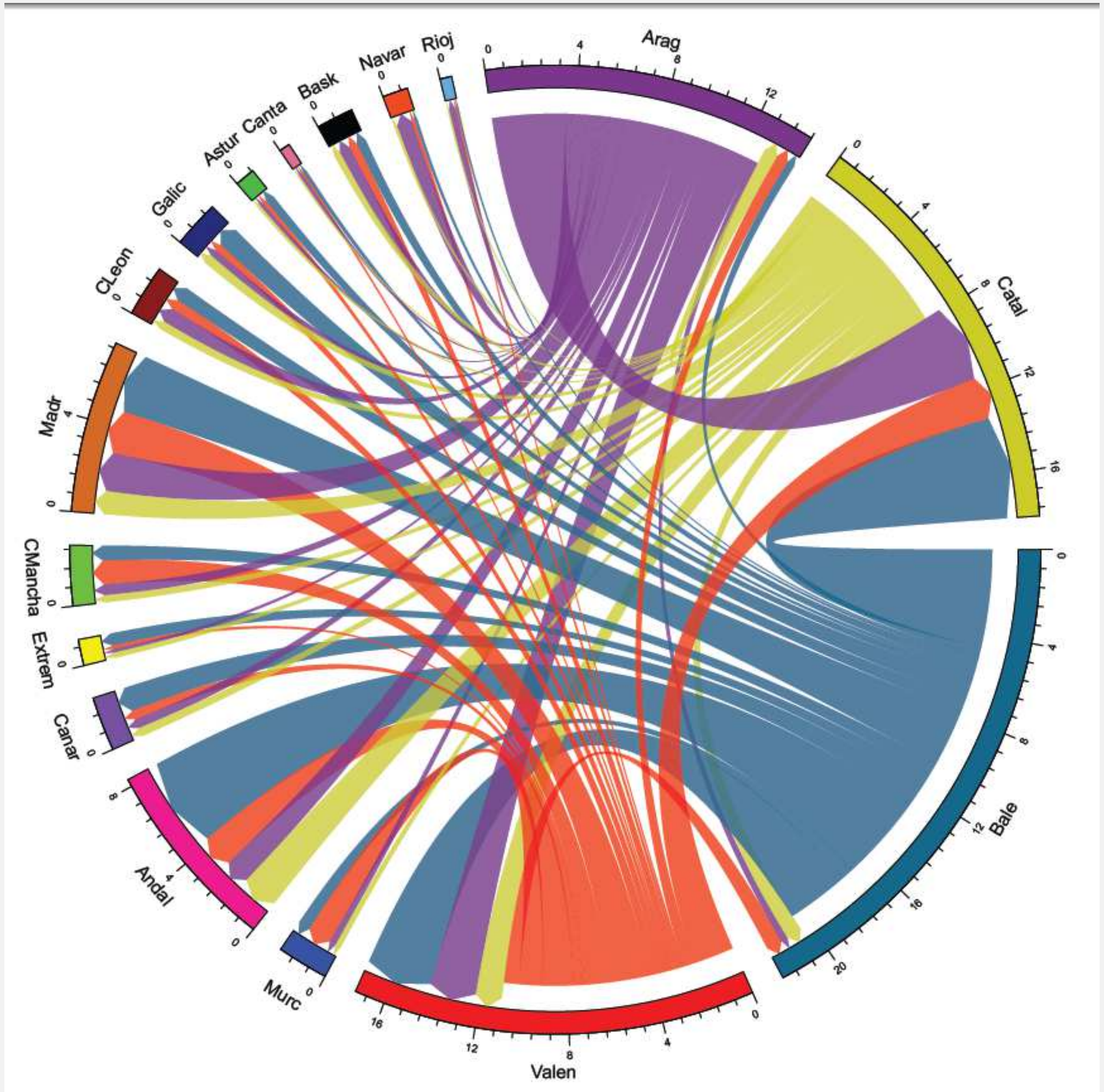


4. EMPIRICAL ILLUSTRATION FOR MIGRATION INTERREGIONAL FLOWS ACROSS NUTS 2 IN SPAIN (2008 – 2012) #v

From Madrid and its neighbors (Castile-La Mancha, and Castile and Leon) TO elsewhere.



4. EMPIRICAL ILLUSTRATION FOR MIGRATION INTERREGIONAL FLOWS ACROSS NUTS 2 IN SPAIN (2008 – 2012) #v



From Catalonia and its neighbors (Aragón, Balearic Islands and Valencian Community) TO elsewhere.

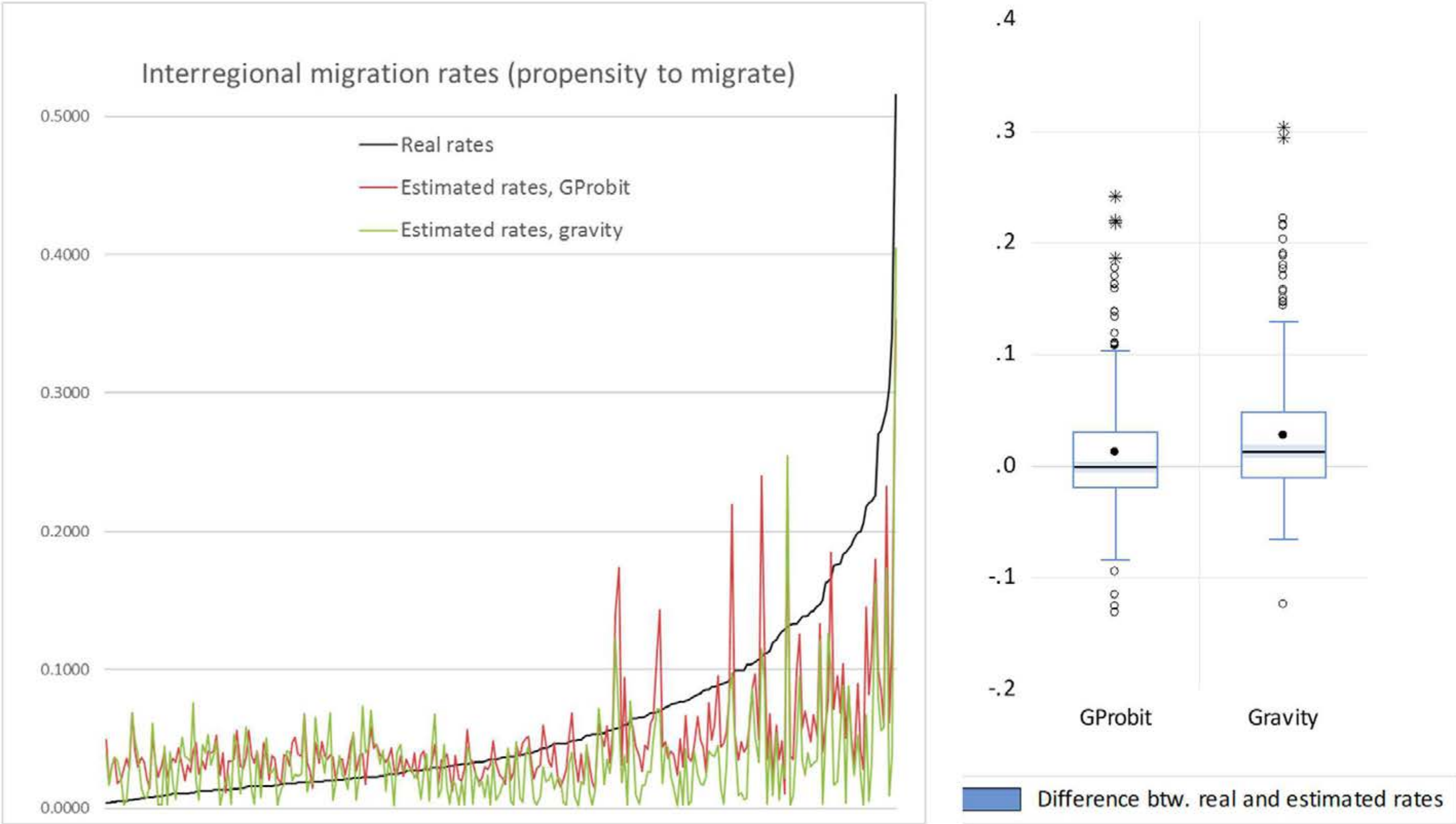
GPROBIT: OLS

Table 1: Estimation results for the interregional migration models

	GProbit model	Gravity model	
Dependent variable	$Z_{od} = \Phi^{-1}(M_{od}/M_o)$ (1)	$\ln(M_{od})$ (2)	$\ln(M_{oo})$ (3)
Constant	-1.820^{***}	7.088^{***}	13.044^{***}
Population D/O ratio	0.036^{***}	-	$0.4 \cdot e^{-7\ ***}$
Housing price D/O ratio	-	-0.481^{**}	-
R&D expenditure p.c. D/O ratio	0.073^{***}	0.137^{***}	-
Average altitude D/O ratio	-0.083^{***}	-0.245^{***}	-
Annual max. temperature D/O ratio	-	-	-0.088^{*}
Atmospheric precipitation D/O ratio	-0.081^{***}	-	-
O-D distance (log)	-0.158^{***}	-0.244^{**}	-
Adj. R-squared	0.312	0.094	0.847
Prediction accuracy measures for the propensity to migrate: $\hat{P}_{od} = \hat{M}_{od}/\hat{M}_o$:			
Bias indicator (RBIAS)	0.79	4.04	
Coefficient of variation (CV)	1.16	311.03	
Relative root mean sq. error (RRMSE)	0.16	0.35	

Note: A robust inference of the GProbit model estimators have been computed.

Fig. 1. Real, estimated and residual interregional flows, GProbit and gravity models



CONCLUSIONS

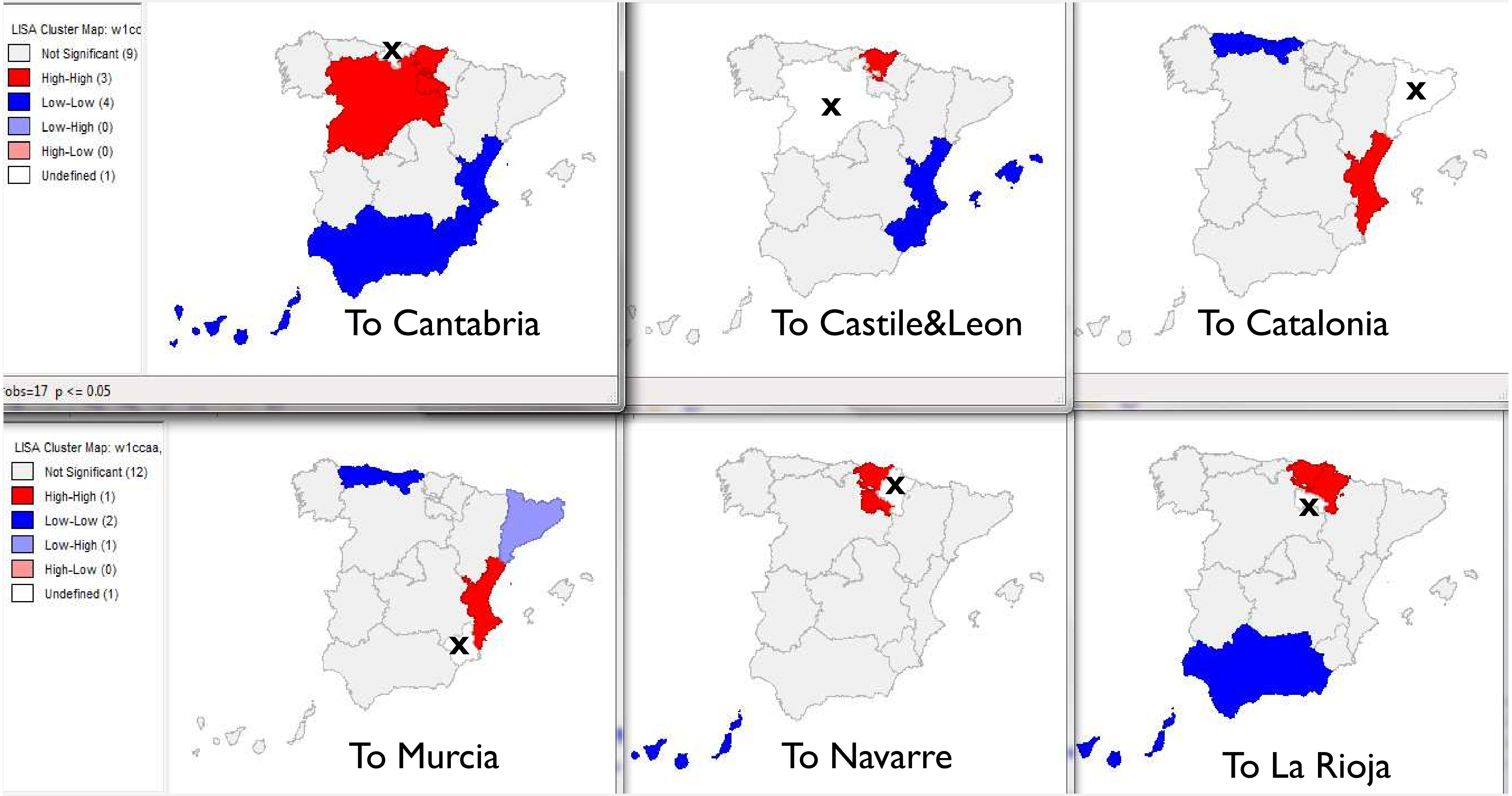
- Adjusted R^2 takes a very low value, particularly for the gravity model estimation, which is in line with other previous analysis in the literature.
- Spanish interregional migration has long been resistant to traditional economic explanations., even to core variables of income and employment (Mulhern & Watson, 2009).
- The strong rigidity of the Spanish labor market, centrally controlled by the trade unions, and a very high national unemployment discourages internal migration (Bover & Velilla, 1999) and instead promotes migration to other countries.

CONCLUSIONS

- Only a few push & pull factors explain internal migration flows among Spanish regions.
- Physical distance in straight line from OD regional capital citiers works better as a deterrence variable than travel time.
- Only socioeconomic agglomeration (population, house price and R&D investment), joint to climate variables explain internal flows among the Spanish regions.
- Pending: coefficient interpretation and estimation of spatial dependence Gprobit models (which imply overcoming some methodological problems, which are present in the spatial depedence gravity models).

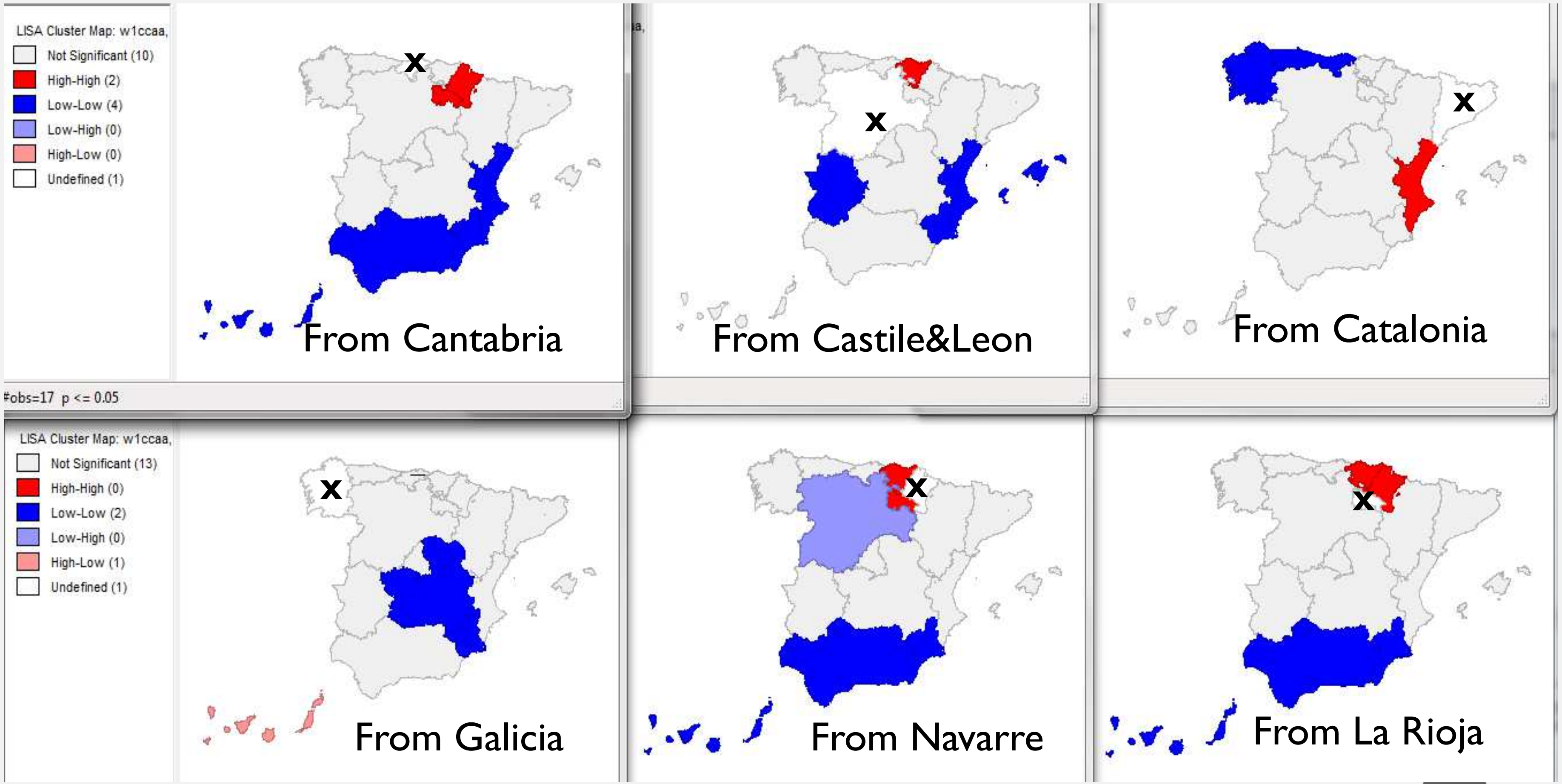
CONCLUSIONS

Spatial dependence of migration flows:



CONCLUSIONS

Spatial dependence of migration flows:



Regions with significant origin/destination local autocorrelation are the same.

THANK YOU!

Coro Chasco

(Universidad Autónoma de Madrid – UAM, Spain)

Luc Anselin

(University of Chicago, USA)

Patricio Aroca

(Universidad Adolfo Ibáñez, Chile)