# 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)

# CONTENTS

- Motivation
- 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)
- Conclusions

# I. 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*).

# 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*.

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

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

$M_{oo}$=‘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*)

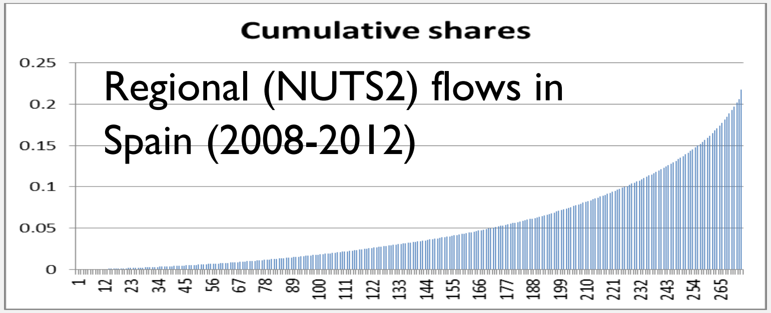
# II. GPROBIT: AN ALTERNATIVE

## II.1. Specification #ii

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

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

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



Regional (NUTS2) flows in Spain (2008-2012)

Non-linear GProbit model of flows

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

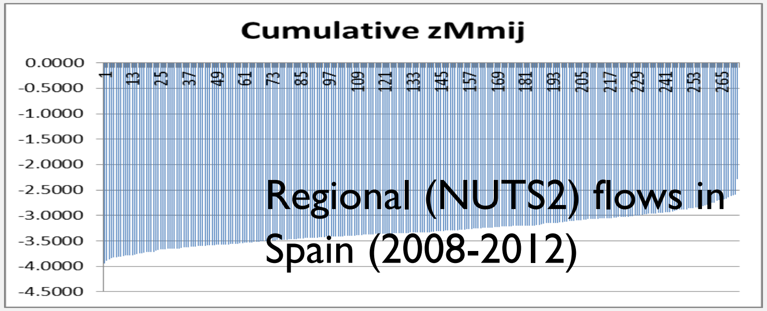
Can be linearized:

(*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 $P_{od}$



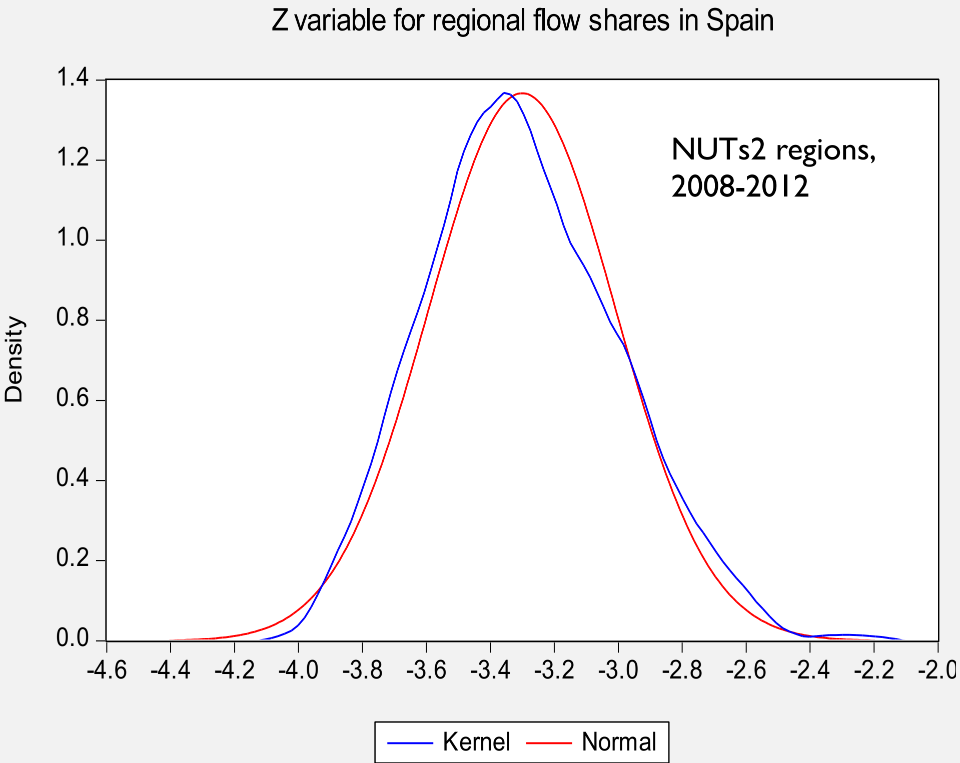
Regional (NUTS2) flows in Spain (2008-2012)

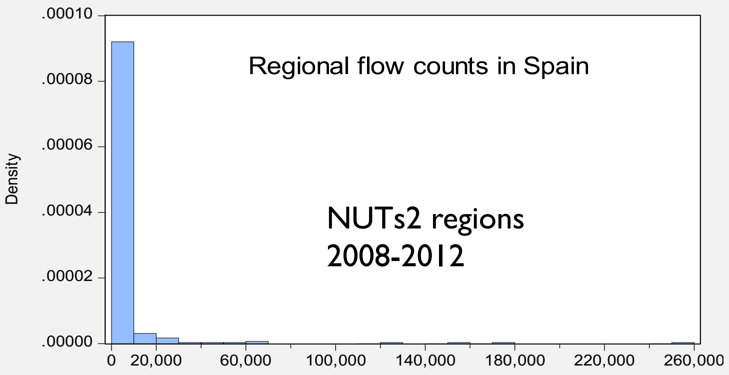
# II. GPROBIT: AN ALTERNATIVE

## II.2. Solutions for model problems #i

| 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 i |

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

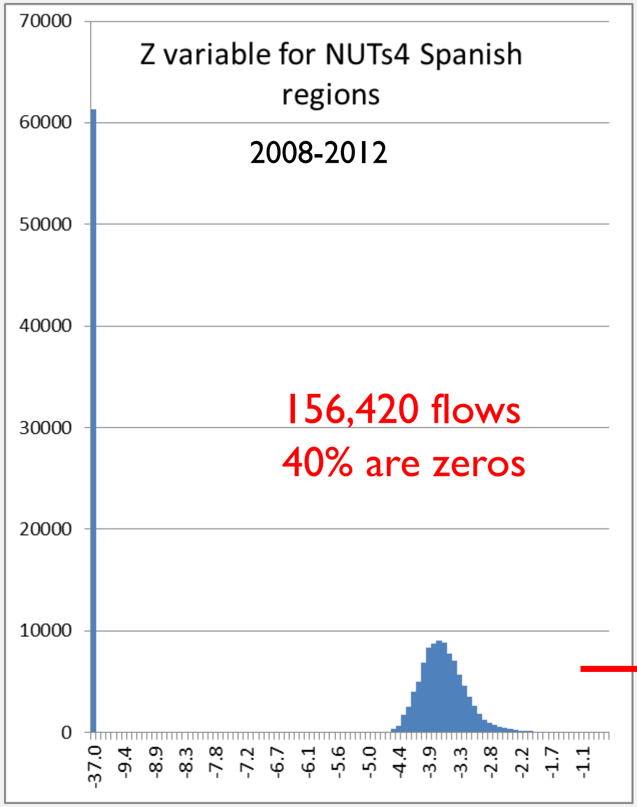




# II. GPROBIT: AN ALTERNATIVE

## II.2. Solutions for model problems #ii

| 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 |


Z variable for NUTs4 Spanish regions

2008-2012

156,420 flows

40% are zeros

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

$$\boldsymbol{Z = \Phi^{-1}(P_{od})}$$

$$\textbf{Domain = [0, 1]}$$

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

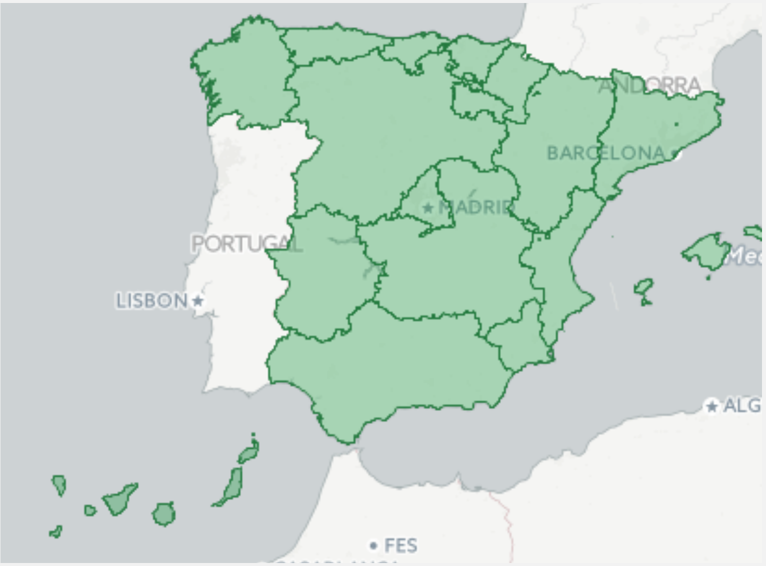
$Z = \Phi^{-1}(P_{od})$ → Domain = $[10^{-323}, (1-2^{-53})]$.

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

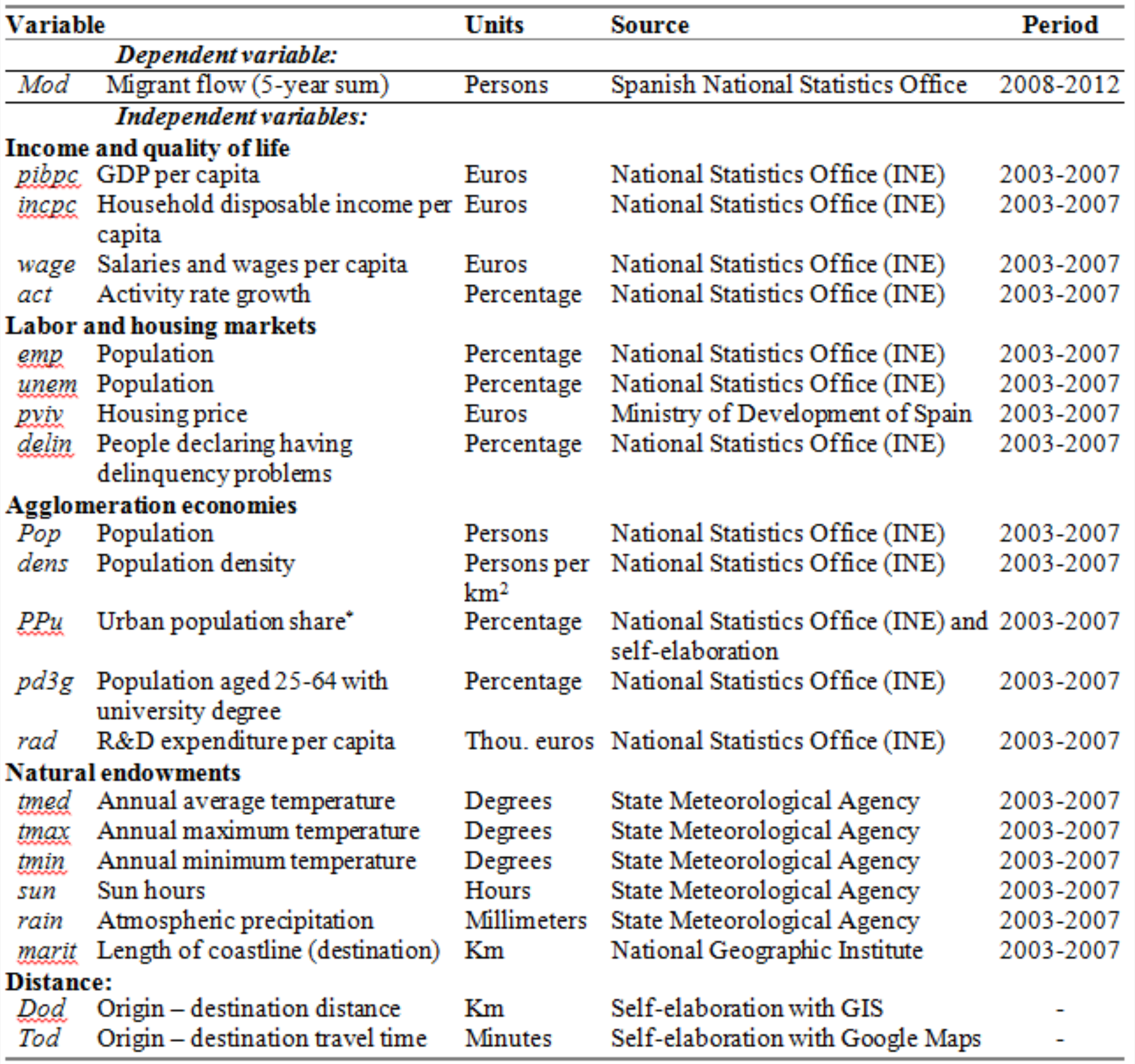
Zeros are possible values for *Z*, but always **problematic** when presented largely in a variable.

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

- 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.



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



| 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.

# GPROBIT: OLS

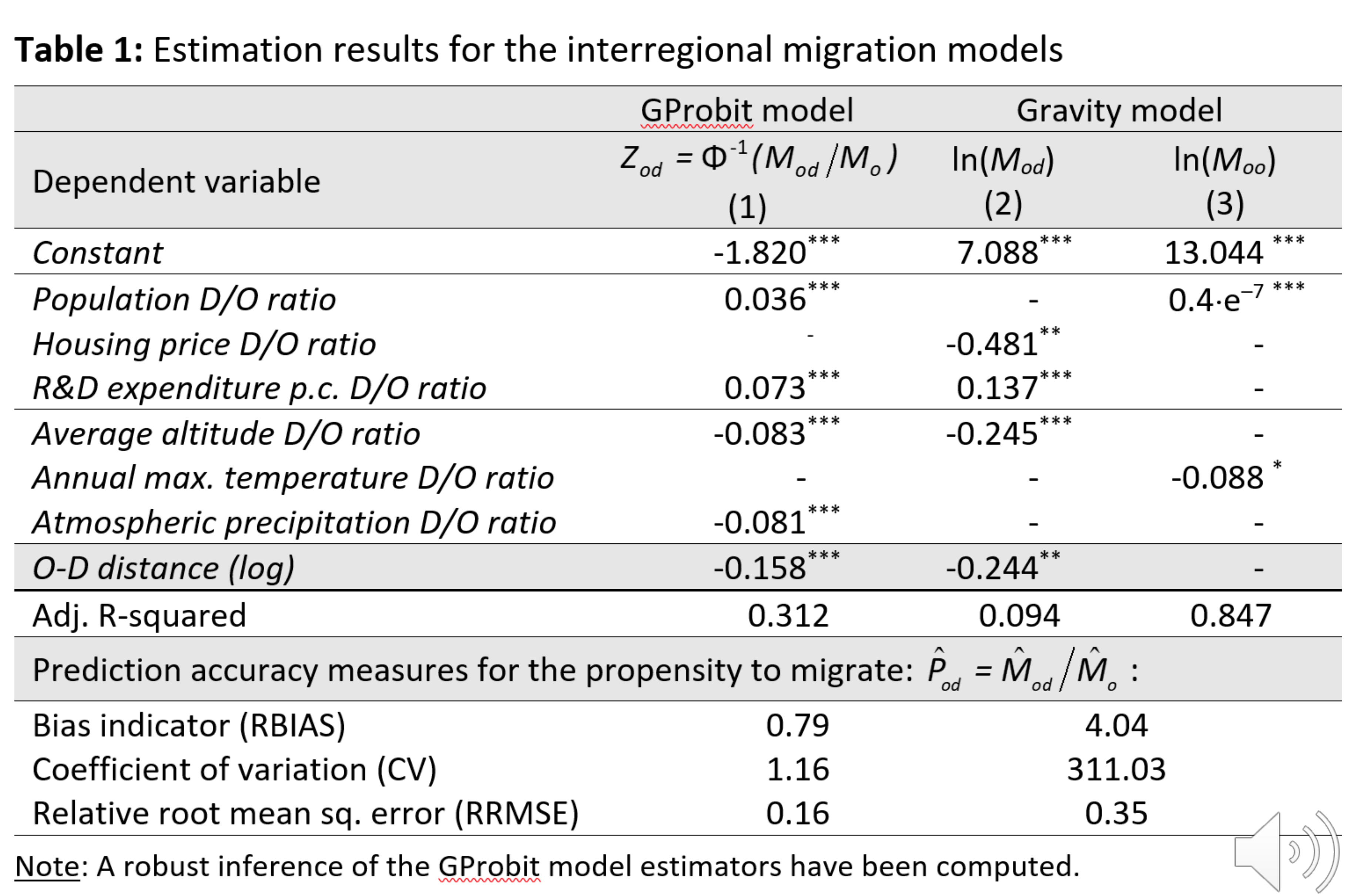
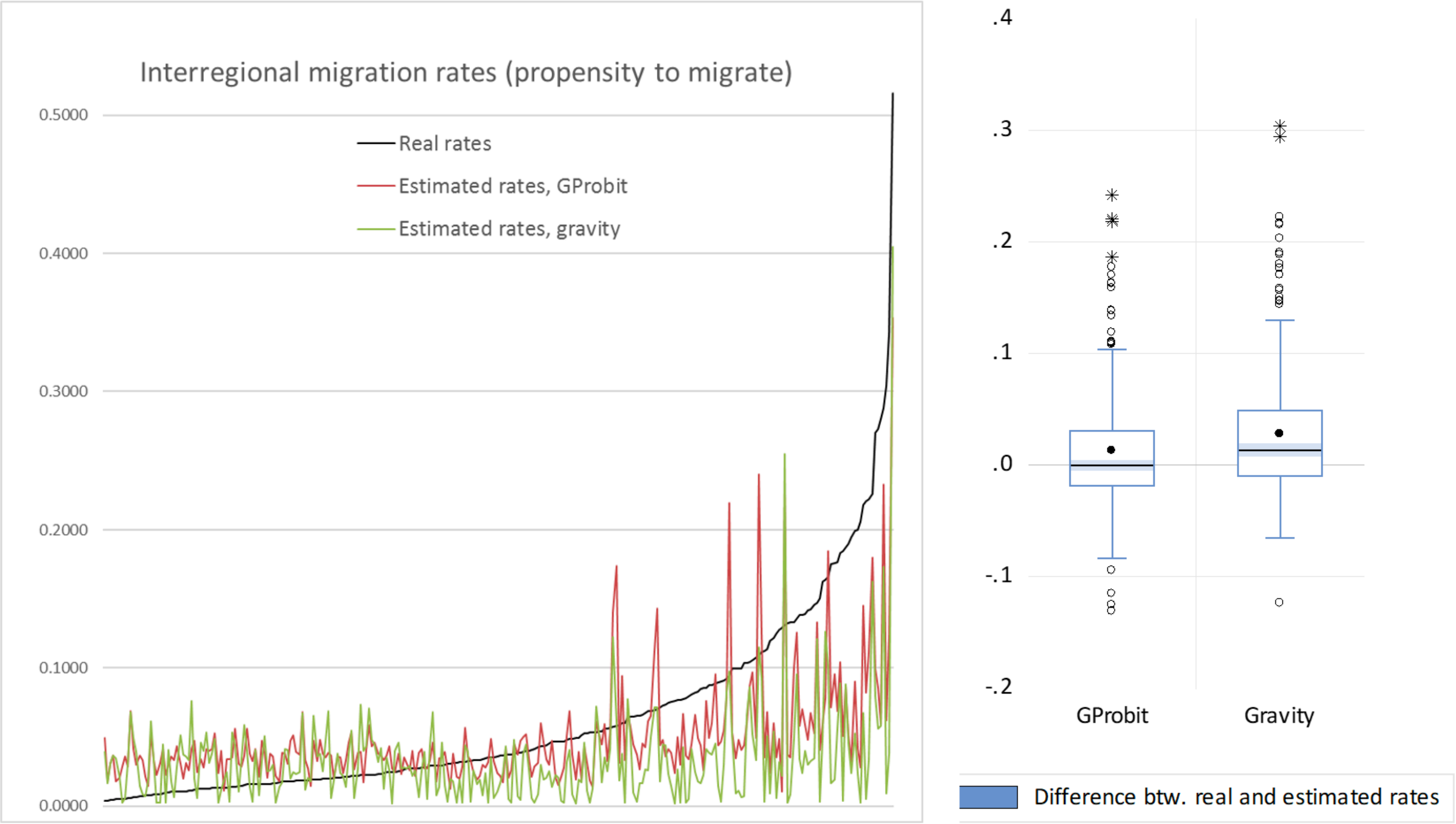


**Table 1:** Estimation results for the interregional migration models

| | GProbit model | Gravity model | |
|---|---|---|---|
| Dependent variable | $Z_{od} = \Phi^{-1}(M_{od}/M_o)$<br>(1) | $\ln(M_{od})$<br>(2) | $\ln(M_{oo})$<br>(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



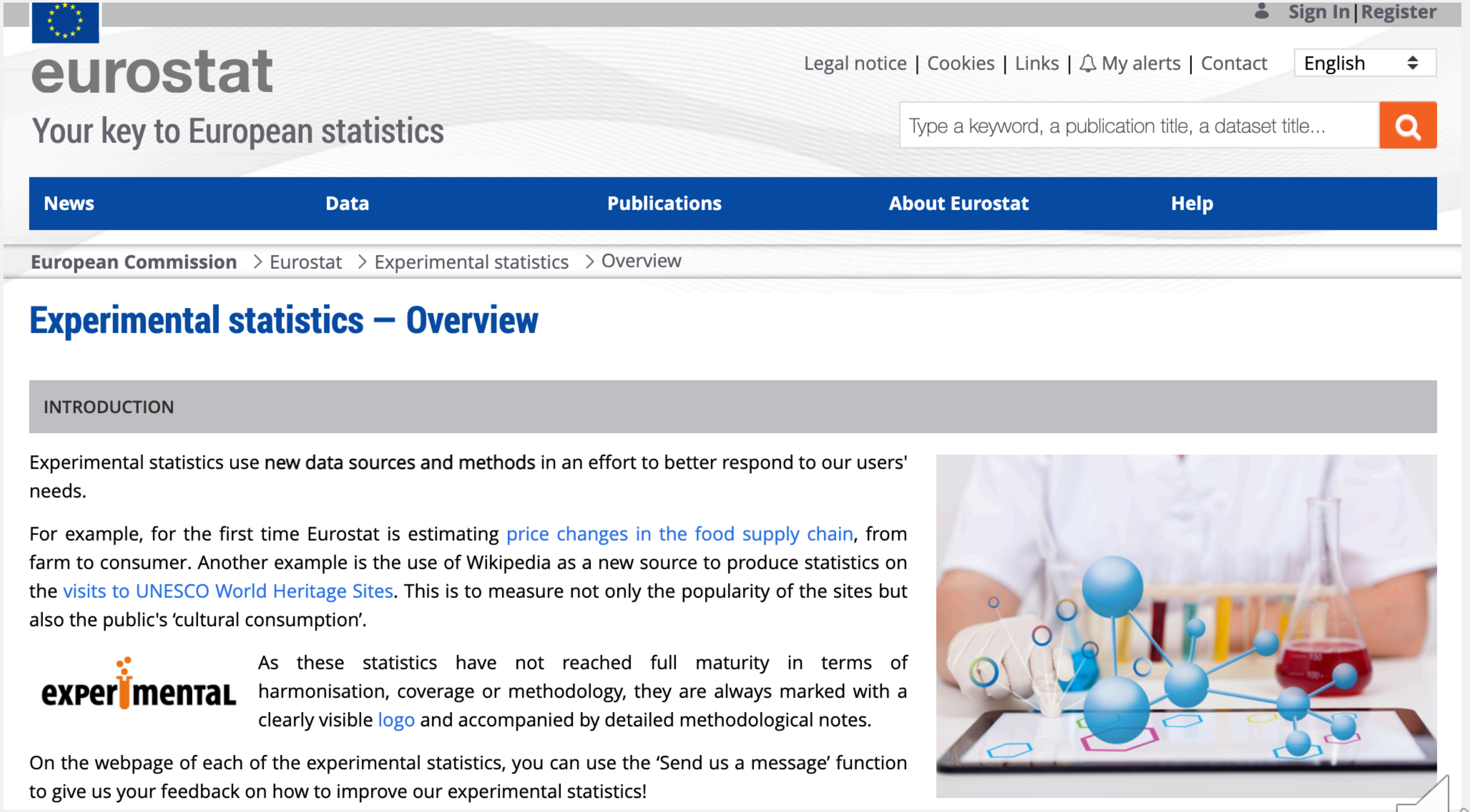
# IV. CONCLUSIONS

- 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.

# IV. 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: : analyze different types of migration flows by gender, age and nationality. Additionally, we also would like to apply this model approach to **other kind of flow data**.

# INE EXPERIMENTAL STATISTICS


Sign In | Register
eurostat
Your key to European statistics
Legal notice | Cookies | Links | My alerts | Contact
English
Type a keyword, a publication title, a dataset title...
News
Data
Publications
About Eurostat
Help
European Commission > Eurostat > Experimental statistics > Overview

## Experimental statistics – Overview

INTRODUCTION

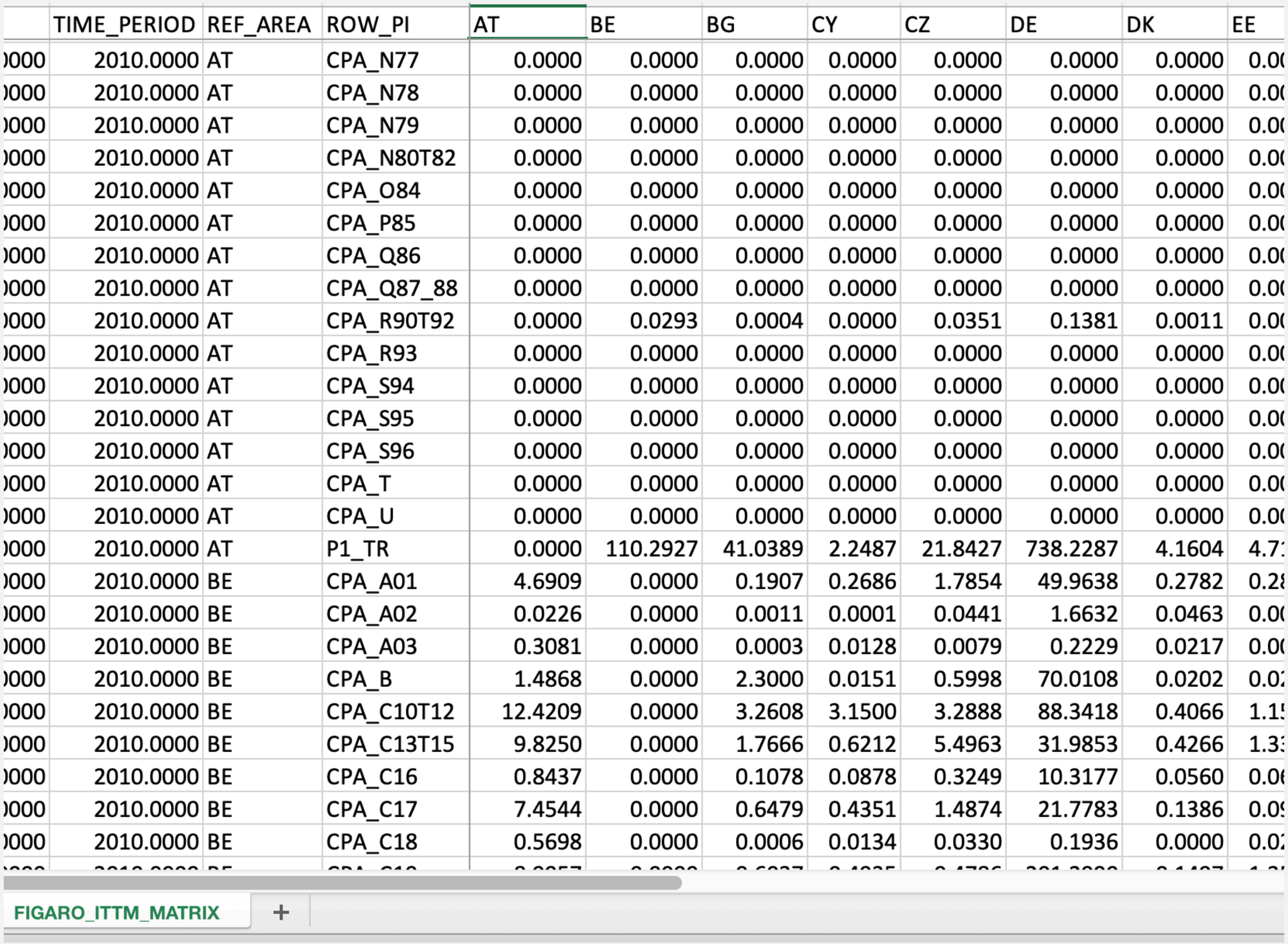
Experimental statistics use **new data sources and methods** in an effort to better respond to our users' needs.

For example, for the first time Eurostat is estimating price changes in the food supply chain, from farm to consumer. Another example is the use of Wikipedia as a new source to produce statistics on the visits to UNESCO World Heritage Sites. This is to measure not only the popularity of the sites but also the public's 'cultural consumption'.

As these statistics have not reached full maturity in terms of harmonisation, coverage or methodology, they are always marked with a clearly visible logo and accompanied by detailed methodological notes.

On the webpage of each of the experimental statistics, you can use the 'Send us a message' function to give us your feedback on how to improve our experimental statistics!

# EUROSTAT EXPERIMENTAL STATISTICS



|  | TIME_PERIOD | REF_AREA | ROW_PI | AT | BE | BG | CY | CZ | DE | DK | EE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 000 | 2010.0000 | AT | CPA_N77 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_N78 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_N79 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_N80T82 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_O84 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_P85 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_Q86 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_Q87_88 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_R90T92 | 0.0000 | 0.0293 | 0.0004 | 0.0000 | 0.0351 | 0.1381 | 0.0011 | 0.00 |
| 000 | 2010.0000 | AT | CPA_R93 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_S94 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_S95 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_S96 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_T | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | CPA_U | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00 |
| 000 | 2010.0000 | AT | P1_TR | 0.0000 | 110.2927 | 41.0389 | 2.2487 | 21.8427 | 738.2287 | 4.1604 | 4.7 |
| 000 | 2010.0000 | BE | CPA_A01 | 4.6909 | 0.0000 | 0.1907 | 0.2686 | 1.7854 | 49.9638 | 0.2782 | 0.2 |
| 000 | 2010.0000 | BE | CPA_A02 | 0.0226 | 0.0000 | 0.0011 | 0.0001 | 0.0441 | 1.6632 | 0.0463 | 0.00 |
| 000 | 2010.0000 | BE | CPA_A03 | 0.3081 | 0.0000 | 0.0003 | 0.0128 | 0.0079 | 0.2229 | 0.0217 | 0.00 |
| 000 | 2010.0000 | BE | CPA_B | 1.4868 | 0.0000 | 2.3000 | 0.0151 | 0.5998 | 70.0108 | 0.0202 | 0.0 |
| 000 | 2010.0000 | BE | CPA_C10T12 | 12.4209 | 0.0000 | 3.2608 | 3.1500 | 3.2888 | 88.3418 | 0.4066 | 1.1 |
| 000 | 2010.0000 | BE | CPA_C13T15 | 9.8250 | 0.0000 | 1.7666 | 0.6212 | 5.4963 | 31.9853 | 0.4266 | 1.3 |
| 000 | 2010.0000 | BE | CPA_C16 | 0.8437 | 0.0000 | 0.1078 | 0.0878 | 0.3249 | 10.3177 | 0.0560 | 0.0 |
| 000 | 2010.0000 | BE | CPA_C17 | 7.4544 | 0.0000 | 0.6479 | 0.4351 | 1.4874 | 21.7783 | 0.1386 | 0.0 |
| 000 | 2010.0000 | BE | CPA_C18 | 0.5698 | 0.0000 | 0.0006 | 0.0134 | 0.0330 | 0.1936 | 0.0000 | 0.0 |

Figaro Tables of international & sectoral trade flows (in current prices).

Regional trade flows?

# THANK YOU!

**Coro Chasco**

**(Universidad Autónoma de Madrid – UAM, Spain)**

Luc Anselin

(University of Chicago, USA)

Patricio Aroca

(Universidad Adolfo Ibáñez, Chile)