

# **Evidence from micro-data as a complement to aggregate analyses on regional disparities.**

Enrique López-Bazo  
AQR-IREA, Univ. Barcelona

Vicente Royuela  
AQR-IREA, Univ. Barcelona



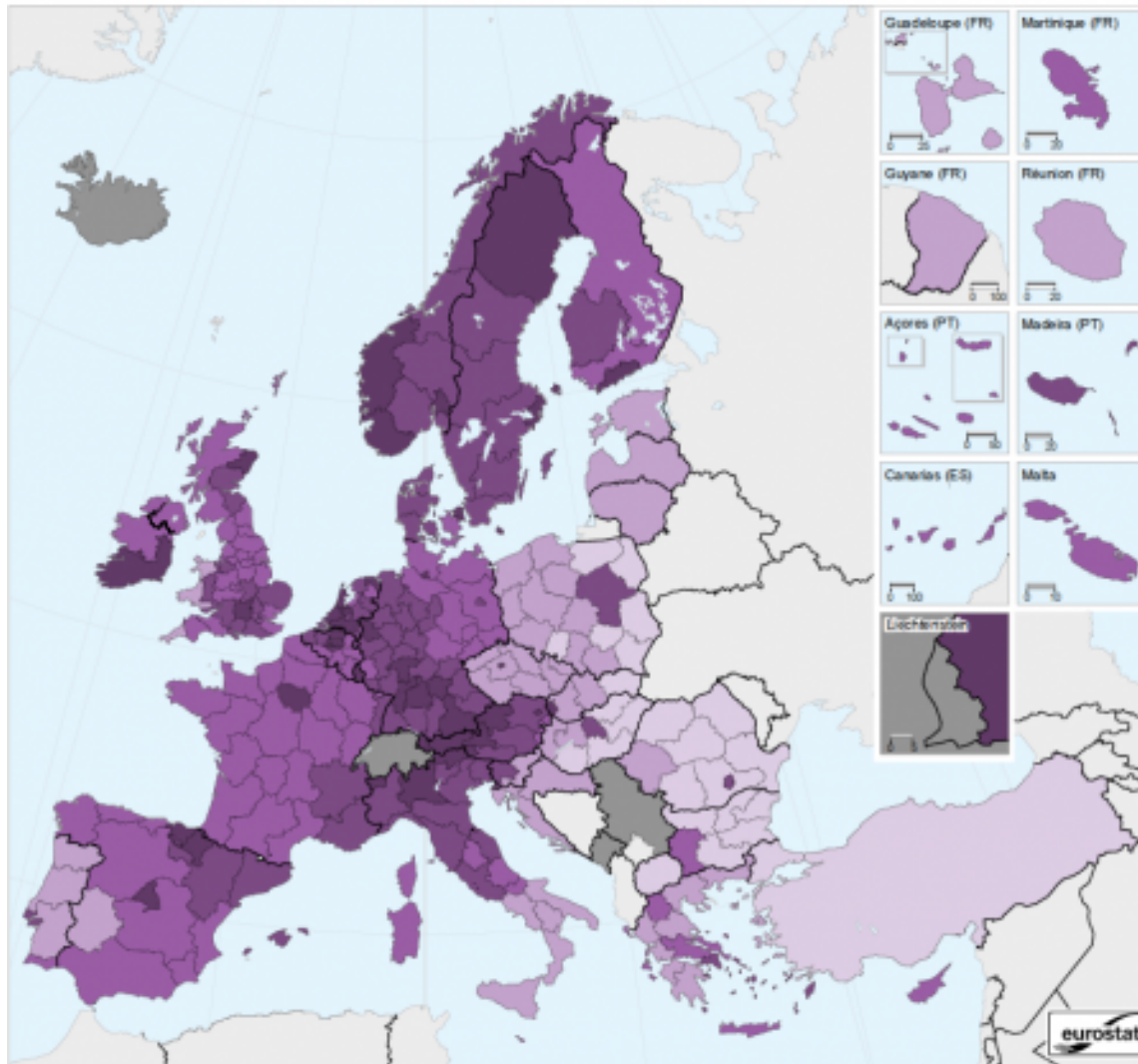
Regional Quantitative  
Analysis Research Group



Most analyses of regional disparities –the so-called *regional problem*– use aggregate data at some territorial level. (for instance OECD, 2011 and EC, 2011)

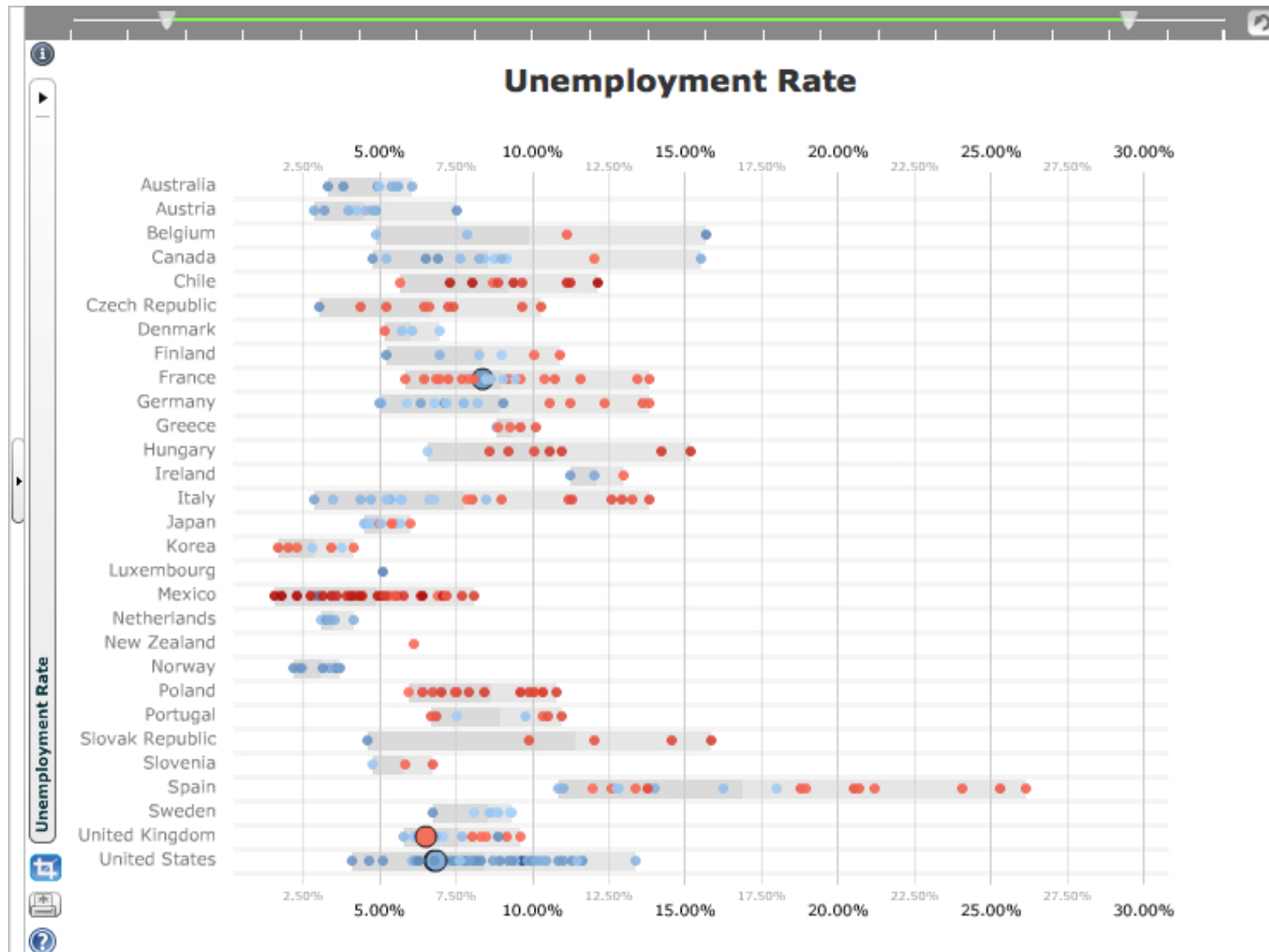
# MOTIVATION

Gross domestic product (GDP) per inhabitant, in purchasing power standard (PPS),  
by NUTS 2 regions, 2010 (\*)  
(% of the EU-27 average, EU-27 = 100)



Source: EUROSTAT.

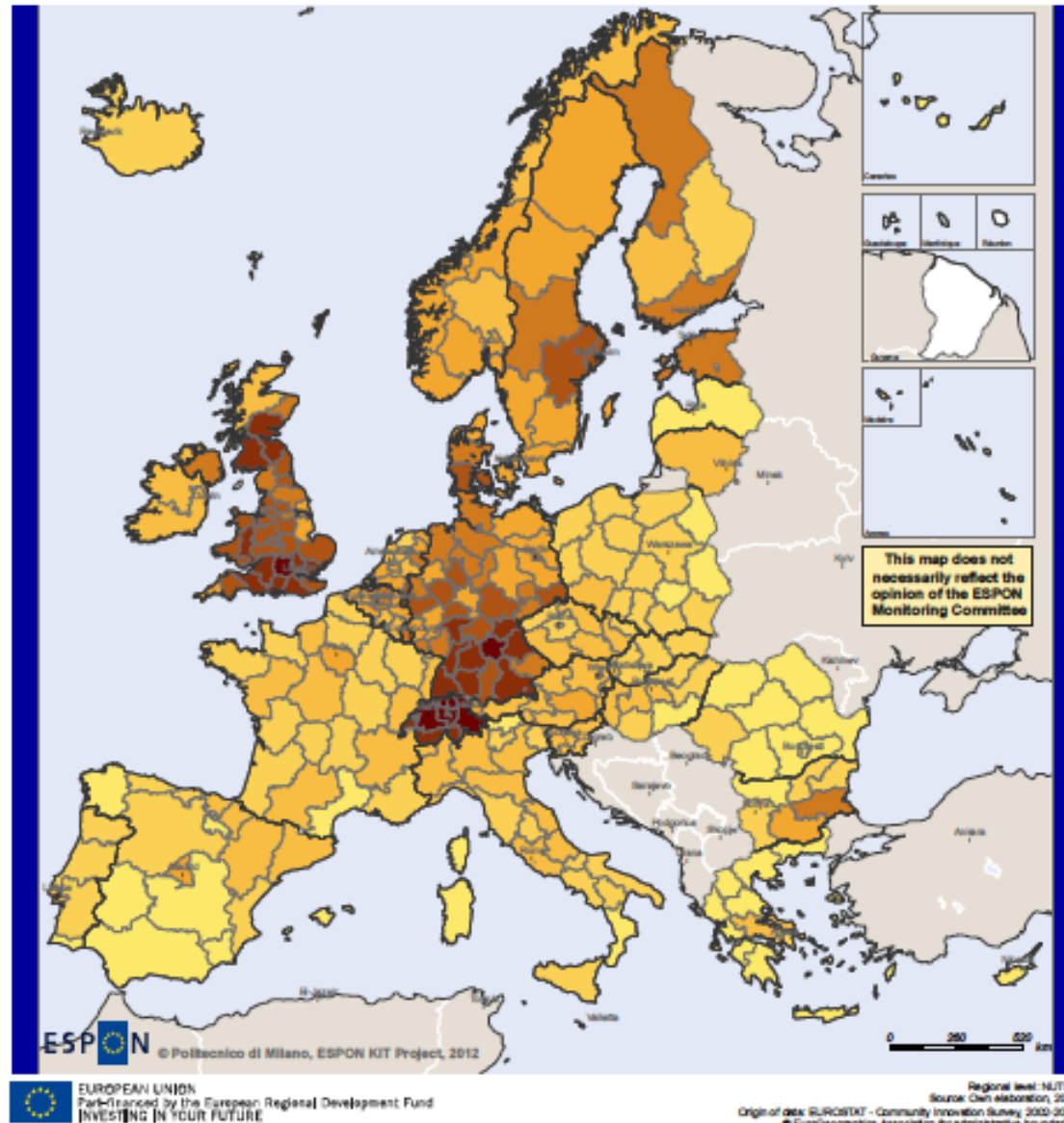
# MOTIVATION



Source: OECD Regions at a Glance, Interactive.

# MOTIVATION

Share of firms introducing product innovation only



Source: KIT, ESPON

Most analyses of regional disparities –the so-called *regional problem*– use aggregate data at some territorial level. (for instance OECD, 2011 and EC, 2011)

**Additional (complementary) evidence can be obtained by using micro-data (firm-level and individual-level) in studies of territorial development and cohesion.**

Introduction

Regional disparities: analyses using aggregate data

Regional disparities: analyses exploiting micro-data

Regional effect of educational human capital

Innovation and firm's export status

Challenges for future research

Traditional approach in studies of territorial disparities, cohesion, regional growth:

- i) Collection of variables at the territorial level, i.e. aggregate or average values.
- ii) Description of the amount of disparities by means of measures of dispersion.
- iii) Explanation of sources of disparities by regression analysis, in which the coefficients linking the variable of interest with its determinants are estimated.



## Aim of this contribution:

To stress that the aggregate approach, being of interest, should be complemented with evidence from micro-level data (for individuals and firms).

## Analyses based on aggregate data:

- i) only focus on the average or representative individual in each region,
- ii) only partially account for the regional distribution of individual characteristics,
- iii) impose the same effect of the determinants in all territories.

Next, some examples to illustrate how using micro-level data can improve our knowledge on the amount and origin of regional disparities:

*On the regional effect of educational human capital.*

*Innovation and firm's export status.*

*Within region inequality and growth*

Implications include support for smart specialization and place-based policies.

Evidence on territorial disparities in, for instance, income per capita, productivity, wage level, unemployment rate, R&D intensity, human capital endowment,...

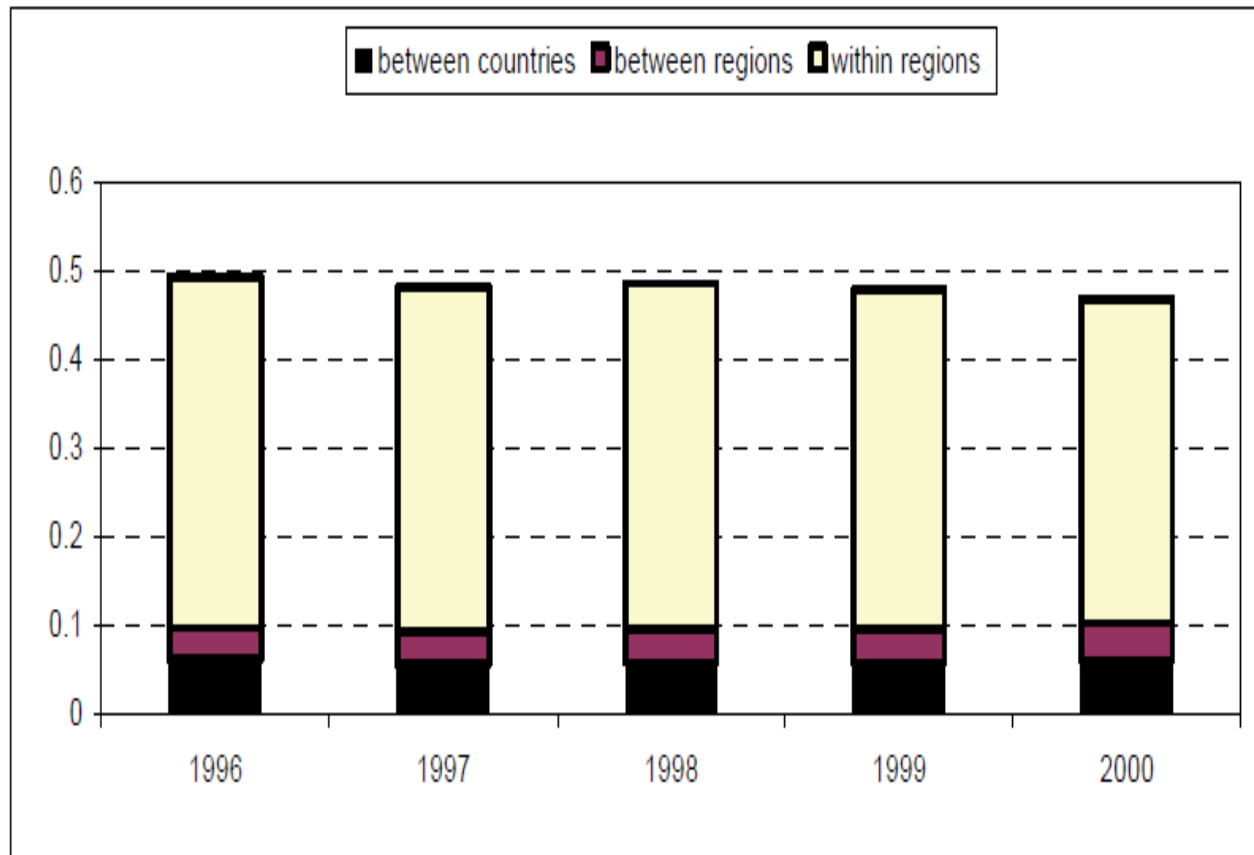
Aggregate or average (per capita, per worker) values of the variable(s) of interest in a set of regions used to:

- Compute **descriptive statistics** (traditional and spatial) with the aim of summarising the amount of dispersion or inequality and its evolution

e.g. GDPpc: choropleth map, coeff. of variation, Gini index, density function

Assumption: a single figure for each region, e.g. the level of GDP per capita, represents well all the economic agents in the region

**Figure 7 Three-level Income Decomposition by Theil Index for the EU from 1996 to 2000**



Source: Rodríguez-Pose & Tselios (2009)

Aggregate or average (per capita, per worker) values of the variable(s) of interest in a set of regions used to:

- Compute **descriptive statistics** (traditional and spatial)
- Analyse the impact of the determinants of the variable of interest, usually by a **regression model**:

$$Y_r = f(X_r, \beta) + \varepsilon_r$$

e.g.  $Y = \text{GDPpc}$

$X$  includes average years of schooling of pop

- Analyse the impact of the determinants of the variable of interest, usually by a **regression model**:

$$Y_r = f(X_r, \beta) + \varepsilon_r$$

Assumption: structural stability in the vector of coefficients  $\beta$ , which means that the response of  $Y$  to changes in  $X$  is similar in all regions.

Imposing the same coefficient to all regions means that one assumes that the response of economic agents is the same in all regions regardless of:

- the particular characteristics of the agents in each region
- the intrinsic features of each territory (such as the institutional, social and economic framework, the endowment of certain factors, and the geographic location).

- Analyse the impact of the determinants of the variable of interest, usually by a **regression model**:

$$Y_r = f(X_r, \beta) + \varepsilon_r$$

Implication of imposing similar coefficients: differences across regions in the variable under analysis are assumed to be caused by differences in the endowment of the determinants only.

Let  $f$  be a linear function and assume  $\varepsilon_r=0$ :

$$Y_r = X_r \beta$$

The difference between region A and region B in the level of  $Y$ :

$$Y_A - Y_B = (X_A - X_B) \beta$$

*If the effects differ across regions:*  $Y_r = X_r \beta_r$

*→ the difference in the level of  $Y$  between regions  $A$  and  $B$  is:*

$$Y_A - Y_B = (X_A - X_B) \beta_A + (\beta_A - \beta_B) X_B$$

(Oaxaca-Blinder gap decomposition)

Contribution to the gap between regions  $A$  and  $B$ :

- difference in the endowment of the variables in  $X$
- differences in the effect of these variables

**A complete analysis of the origin of the territorial disparities requires an estimate of  $\beta_r$ , which is impossible to obtain by exploiting aggregate regional data.**



- Analyse the impact of the determinants of the variable of interest, usually by a **regression model**:

$$Y_r = f(X_r, \beta) + \varepsilon_r$$

Additional implication:

If the focus is on the growth effect of:

- agglomeration economies
- intra-regional inequality

$X$  must include an appropriate measure of these variables.

**However, it is not feasible to compute appropriate measures of intra-regional agglomeration and inequality from aggregate regional data.**

- Increasing availability of micro-level data (individuals, firms)
  - empirical microeconomic evidence complementing that from a macro perspective, using aggregate data (from samples of countries and industries).
- Some studies in regional and urban economics used micro-level data to analyse spatial phenomena.
- However, most analyses of territorial disparities have exploited only aggregate data and, correspondingly, methods from the empirical macro toolkit.

---

Why?

When studying regional disparities from a micro perspective we need **samples of individuals representing the population of each region.**

- Surveys sometimes include the geographic location of the economic agents.
- But it is less frequent that they are designed to guarantee the territorial representativeness.

3 examples to illustrate the potential benefits of complementing the evidence on territorial disparities obtained with aggregate data, with that from micro-level data.

---

## **The regional effect of educational human capital**

**Objective:** To provide evidence on regional differences in the effect of workers schooling on wages and on unemployment

### **Datasets:**

- European Community Household Panel (2000) & Earnings Structure Survey (2002, 2006, 2010) → wages, NUTS2
- Labour Force Survey (quarterly) → unemployment, NUTS2 & NUTS3

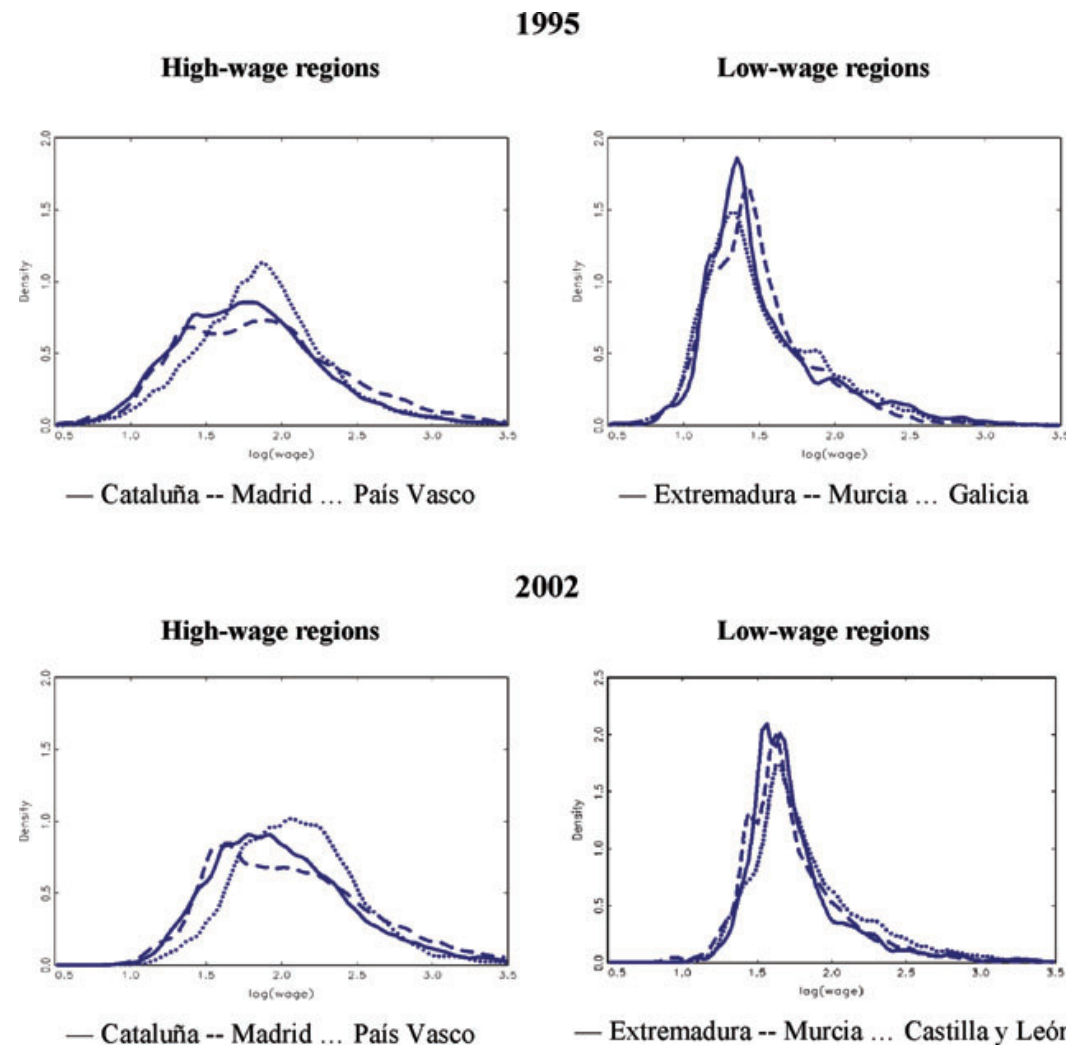
## The regional effect of educational human capital on wages

	Average
<b>1995</b>	
Cataluña	7.09
Madrid	7.85
País Vasco	7.47
<b>High-Wage regions</b>	<b>7.44</b>
Extremadura	4.50
Galicia	5.05
Murcia	4.64
<b>Low-Wage regions</b>	<b>4.83</b>
<b>Relative Gap (HW vs. LW)</b>	<b>0.54</b>
<b>2002</b>	
Cataluña	9.07
Madrid	9.53
País Vasco	9.48
<b>High-Wage regions</b>	<b>9.31</b>
Castilla y León	6.34
Extremadura	5.37
Murcia	5.98
<b>Low-Wage regions</b>	<b>6.04</b>
<b>Relative Gap (HW vs. LW)</b>	<b>0.54</b>

Source: Motellón, López-Bazo, El-Attar, 2009, JRegSci

TABLE 2: Descriptive of Wages in the Groups of Low- and High-Wage Regions

## The regional effect of educational human capital on wages



Source: Motellón, López-Bazo, El-Attar, 2009, JRegSci

FIGURE 2: Density Functions for Real Hourly Wages in the Regions under Analysis.

## The regional effect of educational human capital on wages

	IV Heckit	
	Conditional	Unconditional
Andalusia	0.0819	0.2161
Aragon	0.0739	0.2050
Asturias	0.0124	0.1110
Balearic Islands	0.0709	0.1804
Basque Country	0.0752	0.1813
Canary Islands	0.0588	0.1500
Cantabria	0.0996	0.1711
Castile-La Mancha	0.0790	0.2112
Castile-Leon	0.0788	0.1864
Catalonia	0.0957	0.1765
Extremadura	0.1057	0.2916
Galicia	0.1032	0.2065
La Rioja	0.0780	0.1572
Madrid	0.0676	0.1846
Murcia	0.1057	0.2047
Navarre	0.0890	0.1517
Valencia	0.0873	0.2106
Spain	0.0785	0.1881

Source: López-Bazo, Motellón, 2012, RegStud

Table 4. Estimated returns to education in the Spanish regions

## The regional effect of educational human capital on unemployment

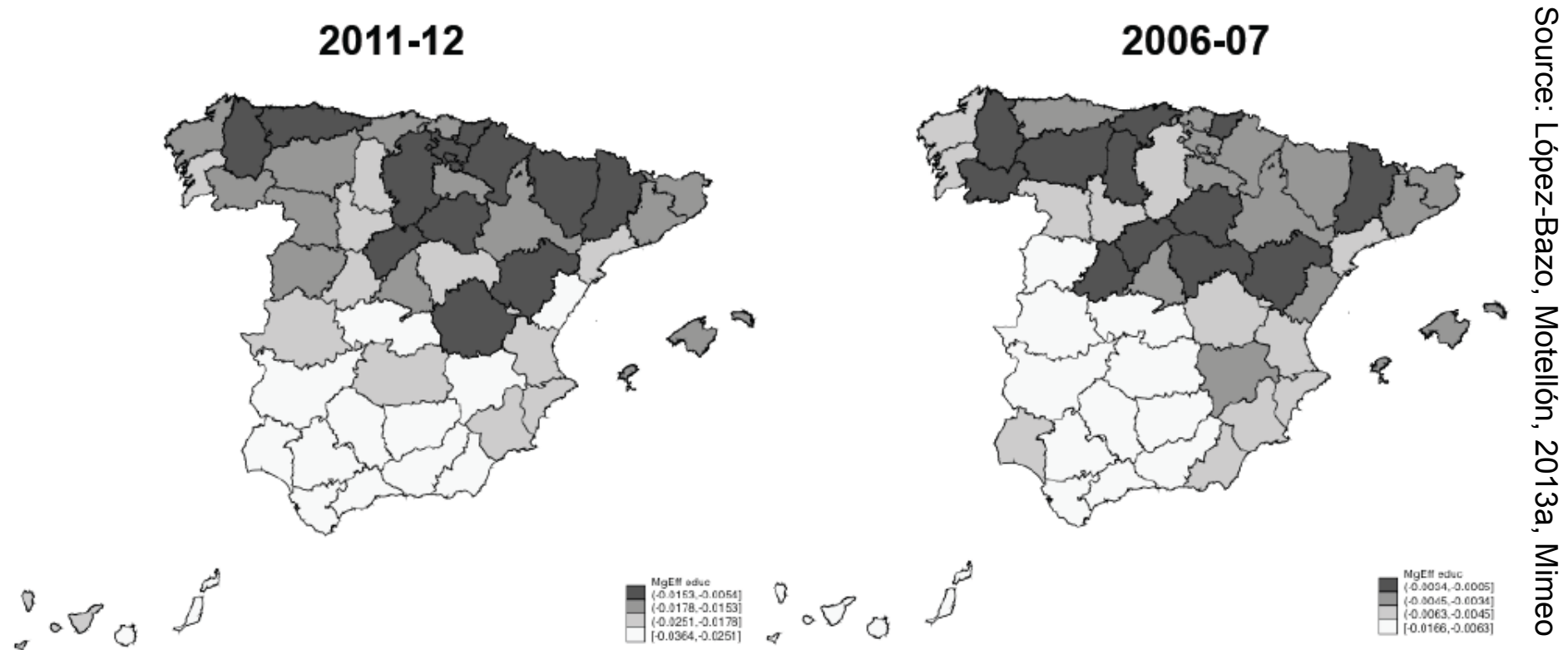


Figure 6. Marginal effect of schooling in Spanish provinces



---

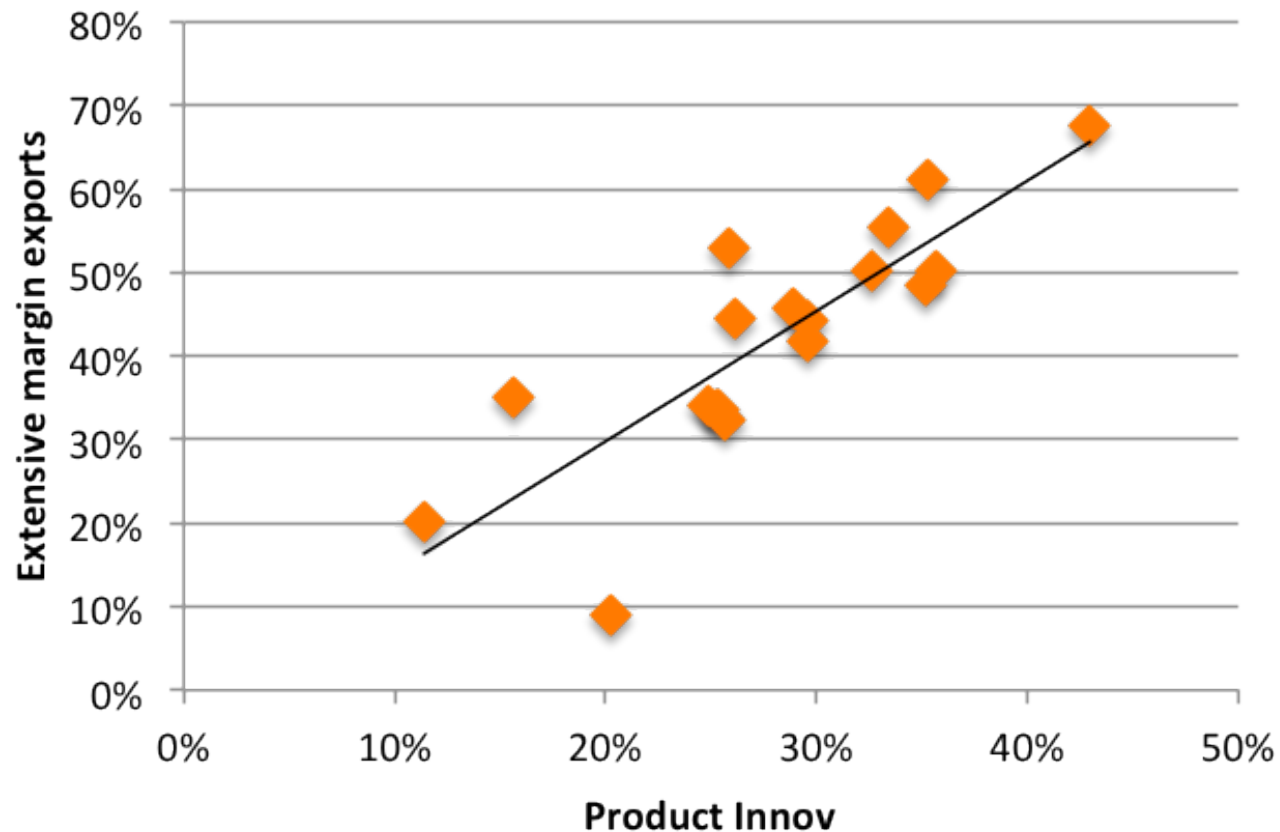
## **The regional effect of innovation on firm exports**

**Objective:** To provide evidence on regional differences in the effect of innovation on firm's export activity

**Dataset:**

- *Encuesta de Innovación en las Empresas* – Enterprises Innovation Survey, produced by INE [ $\sim$ CIS] (yearly 2002–2011)

## The regional effect of innovation on firm exports



Source: López-Bazo, Motellón, 2013b, Mimeo

Figure 1. Share of innovative firms and share of exporting firms in the Spanish regions.

## The regional effect of innovation on firm exports

	Only Product	Only Process	Innovation (Prod/Proc)
Spain	0.342***	0.327***	0.299***
Andalusia	0.279***	0.244***	0.220***
Aragon	0.441***	0.352***	0.322***
Asturias	0.393***	0.356***	0.322***
Balearic Isl.	0.289*	0.118	0.135
Canary Isl.	0.144	0.020	0.050
Cantabria	0.248***	0.350***	0.252***
Castile Leon	0.215***	0.165**	0.148**
Castile La Mancha	0.260**	0.213***	0.207***
Catalonia	0.354***	0.352***	0.319***
Valencia	0.378***	0.380***	0.320***
Extremadura	0.307*	0.308**	0.320***
Galicia	0.335***	0.318***	0.316***
Madrid	0.363***	0.373***	0.356***
Murcia	0.294***	0.358***	0.336***
Navarra	0.310***	0.342***	0.292***
Basque Country	0.437***	0.370***	0.369***
La Rioja	0.240**	0.332***	0.273***

Source: López-Bazo, Motellón, 2013b, Mimeo

Table 4. Marginal effects from the bivariate probit model for the extensive margin of exports in Spanish regions.

---

## **The effect of within region inequality on growth**

**Objective:** To provide evidence on the relationship between income inequality (within region inequality) and growth in the European regions.

### **Datasets:**

- ECHP & EU-SILC for computing inequality indexes for sample of EU regions.
- Cambridge Econometrics Dataset & other sources for GDP growth and aggregate controls

## The effect of within region inequality on growth

Gini index in the EU regions (individual income).

	1994	2001	2004	2010
Min.	0.21	0.22	0.22	0.22
Max.	0.41	0.36	0.38	0.35

Source: Ramos, Royuela, 2013, Mimeo

## The effect of within region inequality on growth

Unbalanced panel GVA conditional convergence

GVA growth 1996-2010	(1) Coeff.	(2) Coeff.	(3) Coeff.	(4) Coeff.	(5) Coeff.
Initial GVA	-0.051***	-0.148***	-0.160***	-0.136***	-0.196***
Initial Gini		-0.040	-0.085*	-0.082*	-0.295**
Gini Growth			-0.039		
Regional fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	No	No	No	Yes	Yes
Observations	1360	668	549	668	556
R2	0.043	0.164	0.177	0.598	

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Balanced panel GVA conditional convergence

GVA growth 1996-2010	(1) Coeff.	(2) Coeff.	(3) Coeff.	(4) Coeff.	(5) Coeff.
Initial GVA	-0.048***	-0.131***	-0.132***	-0.094***	-0.125***
Initial Gini		-0.089**	-0.120***	-0.090***	-0.207**
Gini Growth			-0.139*		
Regional fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	No	No	No	Yes	Yes
Observations	680	441	372	441	372
R2	0.061	0.241	0.254	0.678	

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Source: Ramos, Royuela, 2013, Mimeo

To extend the studies on regional disparities based on microeconomic evidence, more surveys are required that:

- include the **geographic location of individuals**
- guarantee the **regional representativeness**.

However, we are aware that this will:

- **raise the cost** of the statistical survey operations (increase in sample size)
- **affect confidentiality**, since adding the territorial dimension facilitates the identification of individuals with particular characteristics.

In the European Union context:

- it would be important to have **access to homogeneous official micro-level data** for as much Member States as possible (extending the regional dimension in the LFS, the EES, the CIS and the EU-SILC, and providing additional survey data on firm performance),
- ***universal and democratic access*** to micro-data for research,
- appropriate ways for **easing the access** of researchers to these datasets should be implemented (ensuring data confidentiality).



---

Micro-data and knowledge on territorial disparities:

- **Empirical procedures ready** to be applied for obtaining evidence on regional disparities exploiting micro-data.

They have been developed and applied in other fields of economics and social sciences.

- Increasing evidence on regional disparities based on micro-level data will stimulate **further theoretical developments** and efforts to develop a framework for integrating the macro and the micro approaches.
- (Hopefully) We have proved that analyses exploiting regional micro-data provide particular information for each region that remain hidden in the aggregate studies → useful evidence on the ex-ante and ex-post evaluation of the effectiveness of **smart specialization** and **place-based policies**.

# **Evidence from micro-data as a complement to aggregate analyses on regional disparities.**

Enrique López-Bazo  
AQR-IREA, Univ. Barcelona

Vicente Royuela  
AQR-IREA, Univ. Barcelona



Regional Quantitative  
Analysis Research Group

