

# Self-declared *Professional Status*

---

## Additional Questions to Better Measure and How to Link the Mismatches Produced

Javier Orche Galindo

Miguel Ángel García Martínez

**Spanish Labour Force Survey (LFS)**



Gdańsk, May 2013

# The self-declared “Professional Status” in the Spanish LFS

## Index

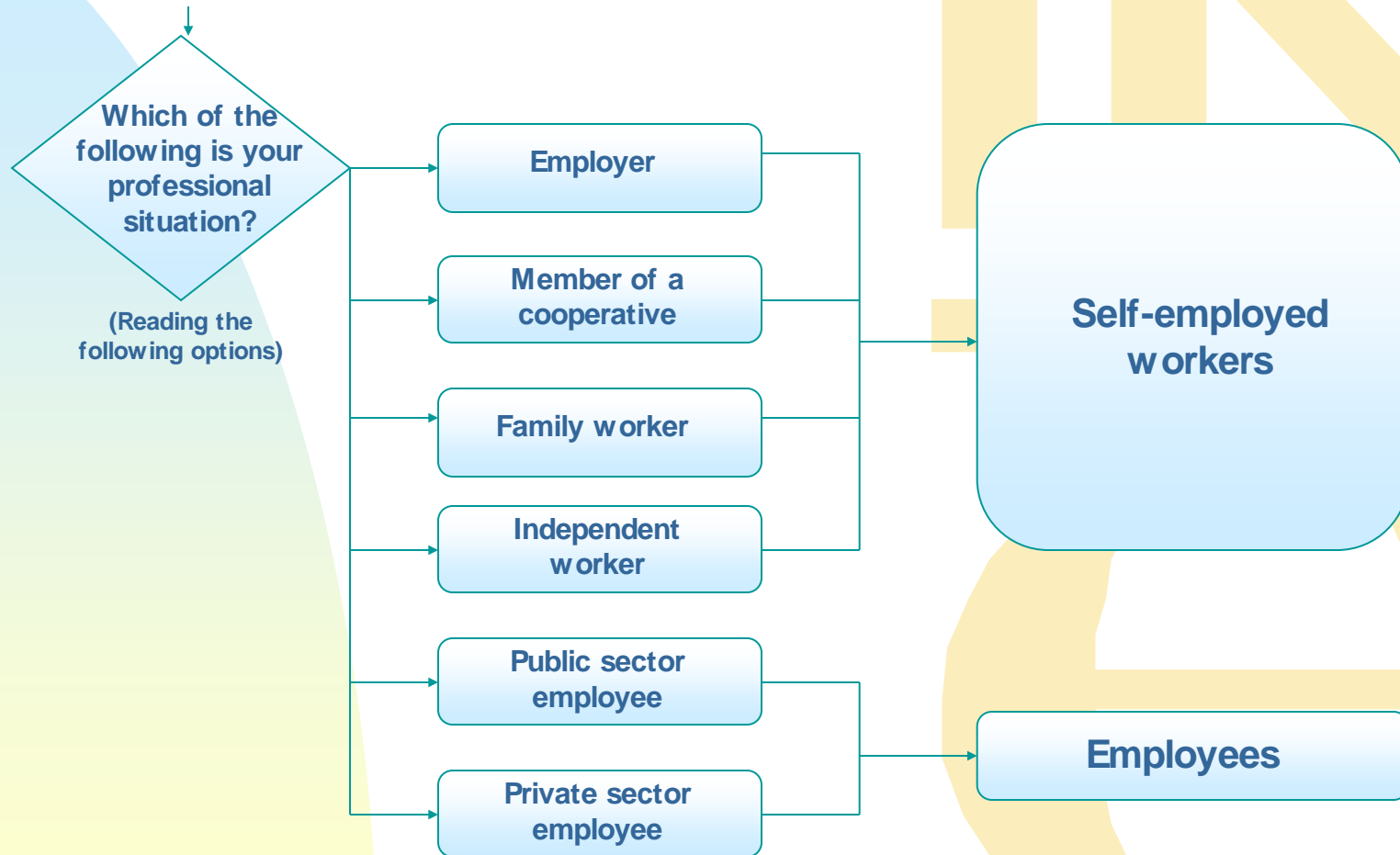
- I. Additional questions to better measure the self-declared *professional status*.
- II. How to link the mismatches produced in series through an econometric model.
- III. Final conclusions.

**The self-declared “Professional Status” in the Spanish LFS**

**I. Additional questions  
to better measure the self-declared  
*professional status***

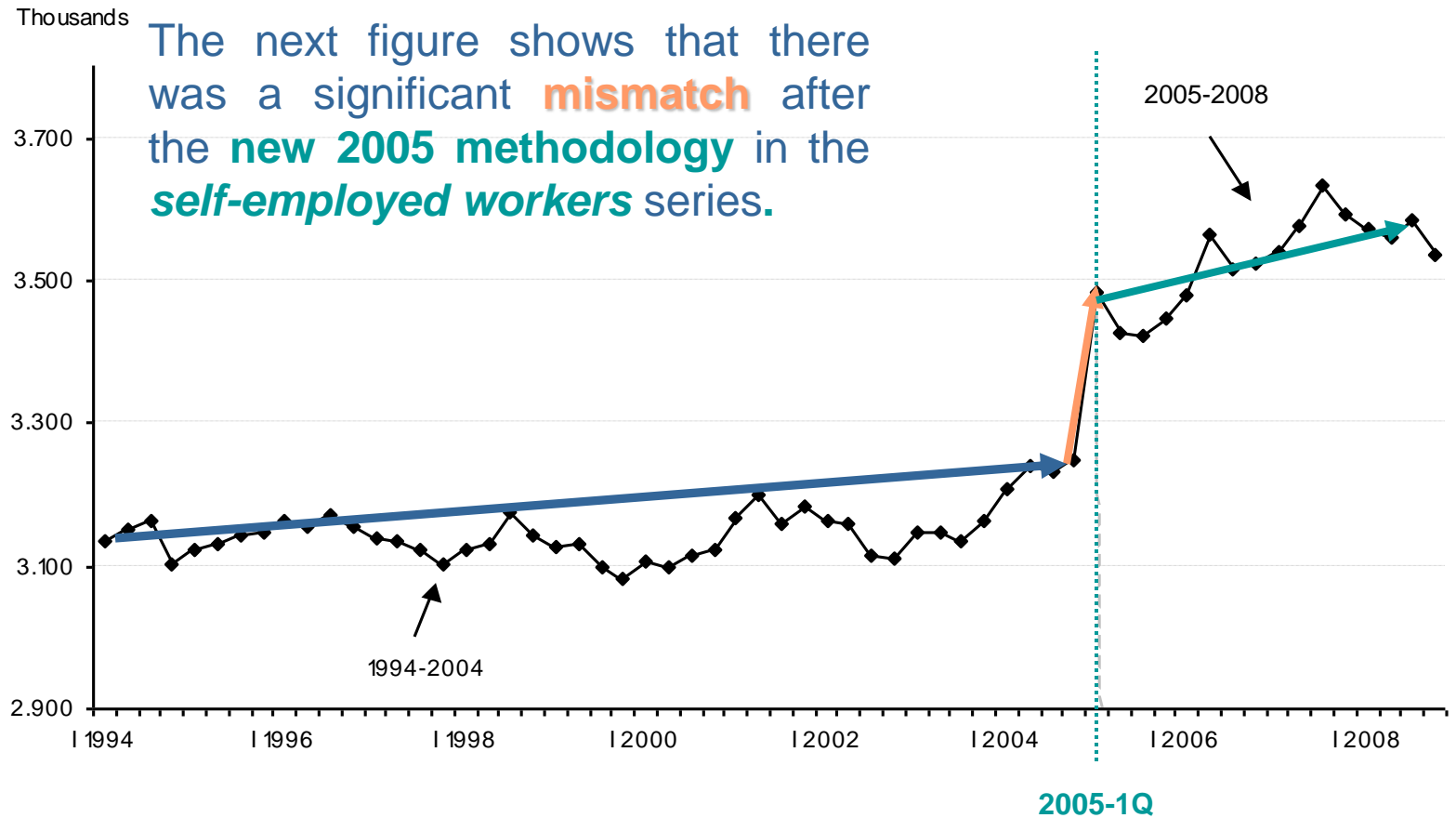
## I.I. The self-declared professional status (2005-2008 series)

In the **Spanish LFS** (after implement a new methodology in **2005**), there was the following **only one question** to obtain the **professional status** variable:



# I.I. The self-declared professional status (2005-2008 series)

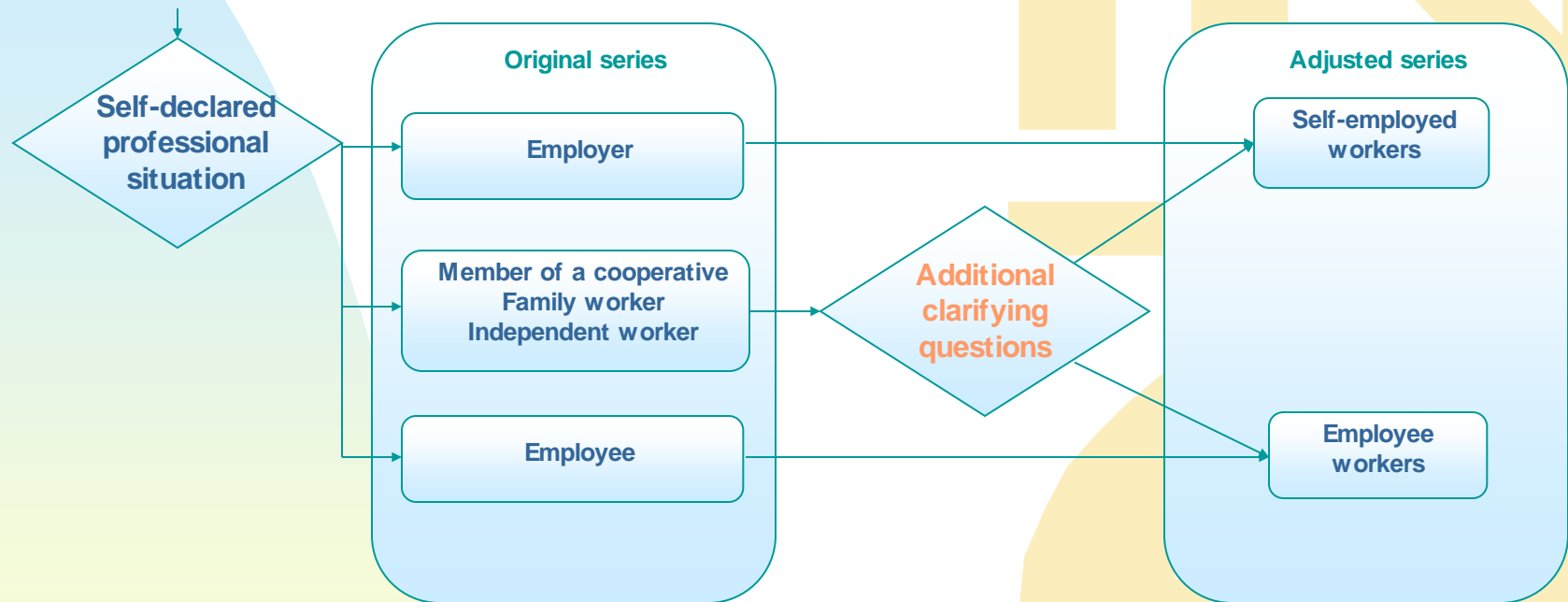
Figure 1. Self-employed workers: total



Source: Spanish LFS

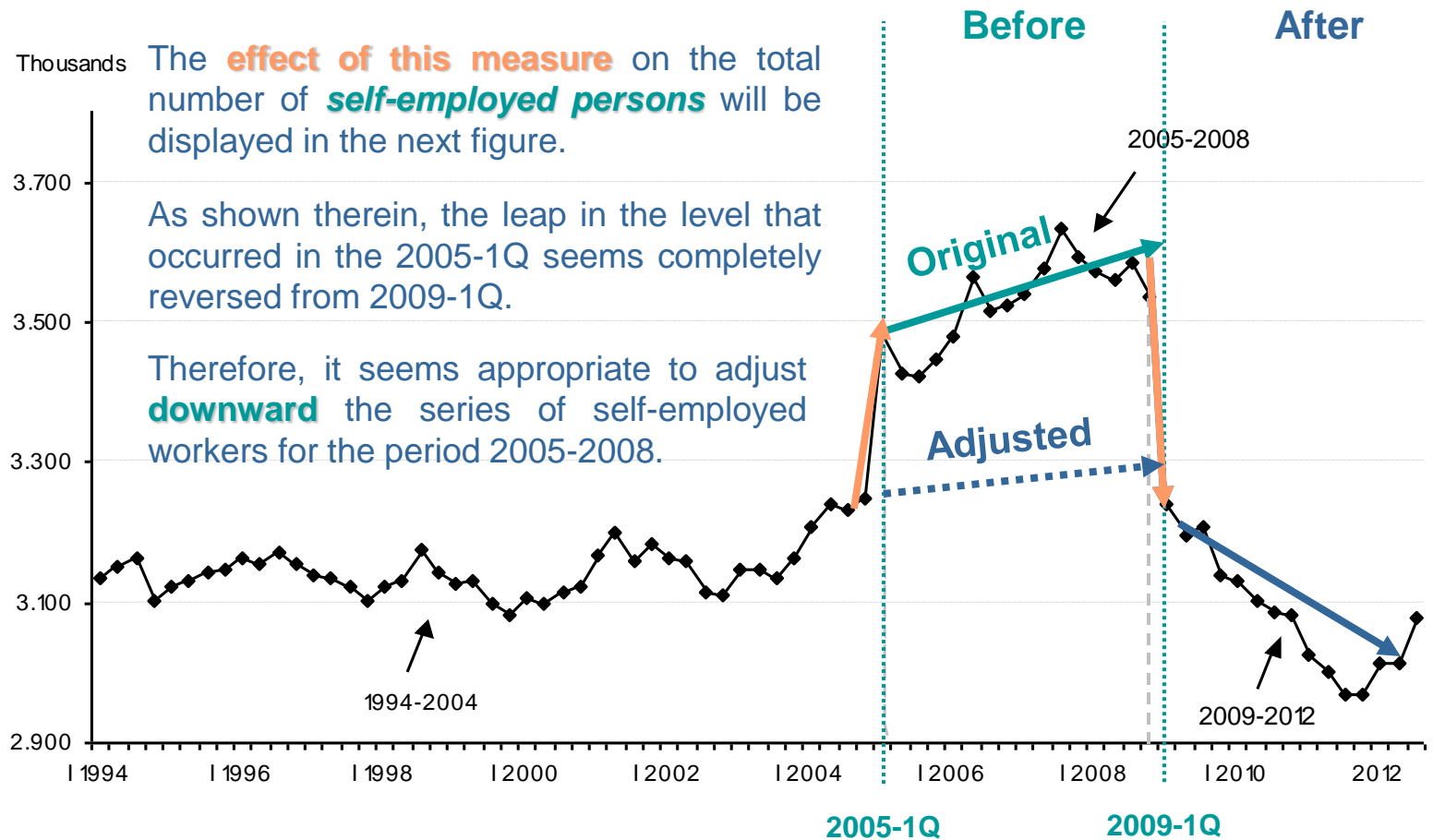
## I.II Additional questions to better measure (since 2009)

Analysing the problem, it was decided (since 2009) to include **additional questions** in **some self-classifications** for clearing the **nature of employment**:



## I.II Additional questions to better measure (since 2009)

Figure 2. Self-employed workers: total



Source: Spanish LFS

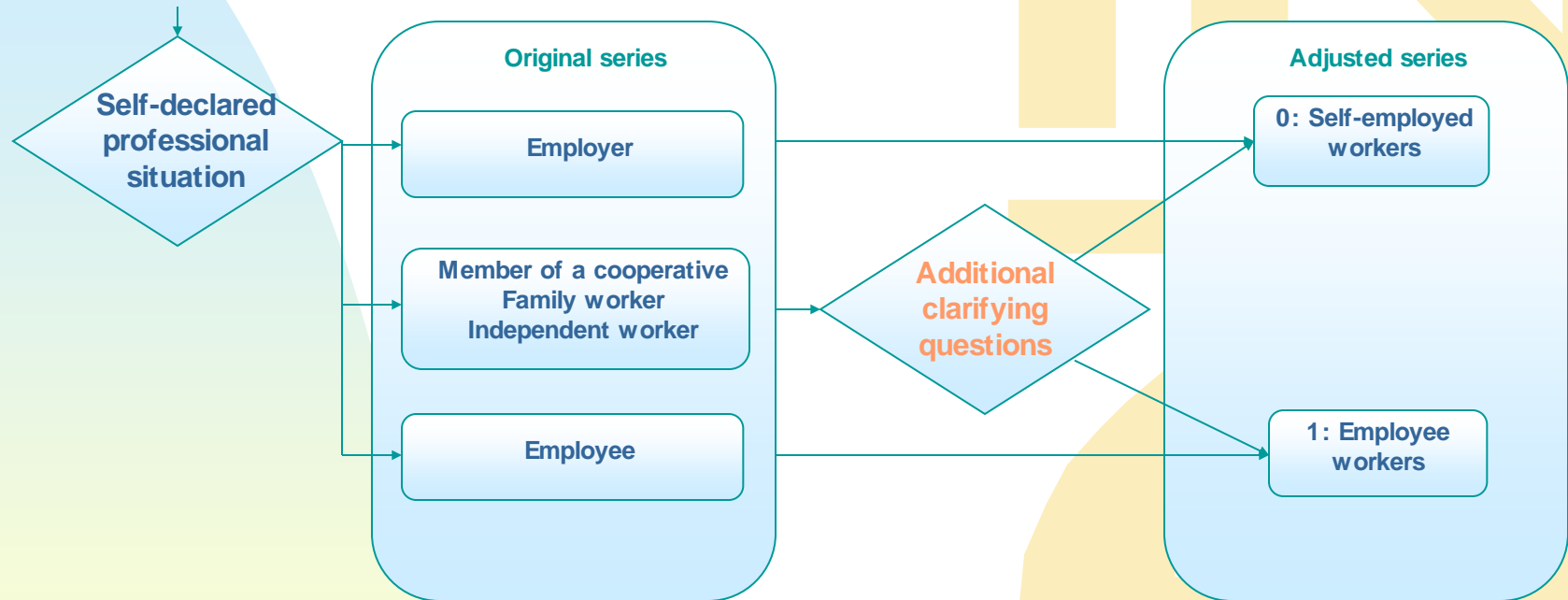
*The self-declared “Professional Status” in the Spanish LFS*

## II. How to **link** the mismatches produced in series through an **econometric model**



## II.1 The econometric model proposed

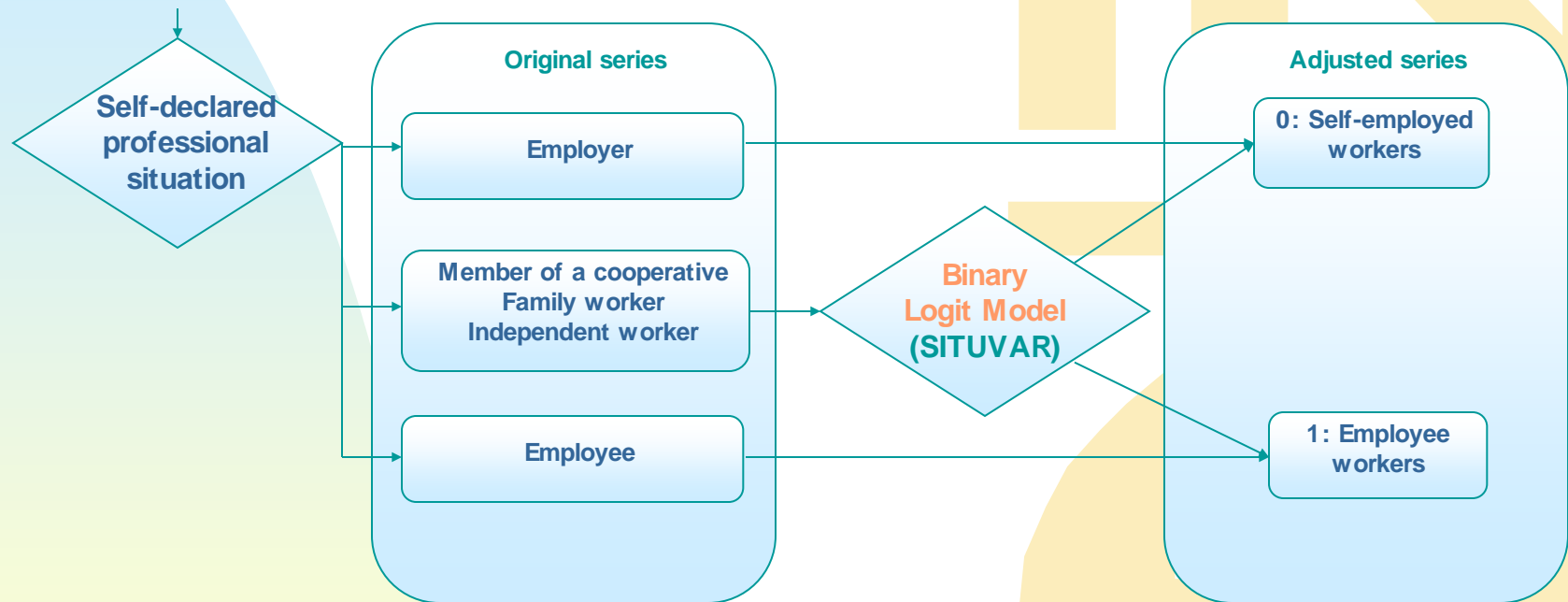
The Spanish LFS quarterly series (from **2009** to **2012**) collected the professional status classification in respect to the **original and fitted series** with **new questions** incorporated in that year.



Therefore, we can forecast using an **econometric model**, the lower self-employed level and the resulting increase in the same amount of employees in the quarterly series from **2005** to **2008**.

## II.1 The econometric model proposed

The **model** which is used to fit the model it is a **binary logit model**, where the response variable (**SITUVAR**) takes on only two possible values **0** (holding the same classification of self-employed with new questions incorporated) and **1** (if varies becoming employee).



Therefore, this model would be shown as the **probabilities of change** in each subgroup according to **explanatory variables** collected in the rest of the **LFS questionnaire**.

## II.II The significant explanatory variables observed (2009-2012)

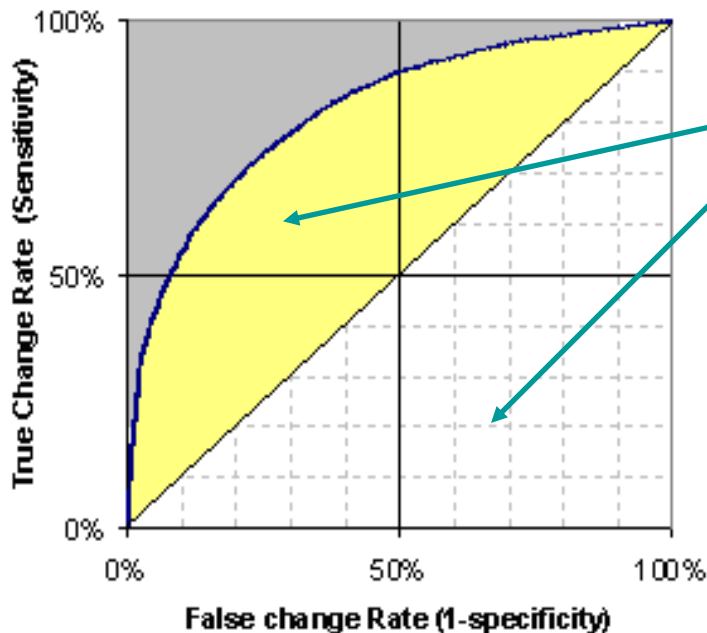
Model: <b>binary logit</b> - Response variable: <b>SITUVAR</b>
Filtered Units : Self-declared independent worker, member of a cooperative or family worker
Number of observations used (from 2009 to 2012): <b>481,271</b> units
No change the self-classification: <b>92.3%</b> Change the self-classification to employee: <b>7.7%</b>

Explanatory <b>LFS variables</b> selection (at 0.05 level to enter)	- 2 LOG L	p-value
Step 0. <b>Intercept</b> (initial intercept entered)	<b>20,370.7</b>	
Step 1. <b>Self-declared professional status</b>	<b>18,782.0</b>	<b>&lt; 0.0001</b>
Step 2. <b>Occupation</b> (ISCO1D)	<b>17,482.4</b>	<b>&lt; 0.0001</b>
Step 3. <b>Interaction between self-declared and occupation</b>	<b>17,108.1</b>	<b>&lt; 0.0001</b>
Step 4. <b>Supervisory responsibilities</b>	<b>16,905.0</b>	<b>&lt; 0.0001</b>
Step 5. <b>Seniority</b> (months working)	<b>16,694.1</b>	<b>&lt; 0.0001</b>
Step 6. <b>Activity</b> of the local unit (NACE1D)	<b>16,513.3</b>	<b>&lt; 0.0001</b>
Step 7. <b>Interaction between self-declared and supervisory</b>	<b>16,427.9</b>	<b>&lt; 0.0001</b>
Step 8. <b>Age</b> (years old)	<b>16,393.3</b>	<b>&lt; 0.0001</b>
Step 9. <b>Interaction between self-declared and seniority</b>	<b>16,324.6</b>	<b>&lt; 0.0001</b>
Step 10. <b>Region</b> of household (NUTS2D)	<b>16,291.0</b>	<b>&lt; 0.0001</b>
* <b>No more LFS variables</b> were entered or removed at 0.05 level	<b>16,291.0</b>	

## II.III Evaluation of the Goodness of Fitted Model (2009-2012)

Association of Predicted Probabilities and Observed Responses (from 2009 to 2012)	
Percent Concordant	82.2%.
Percent Discordant	17.2%.
Percent Tied	0.6%.

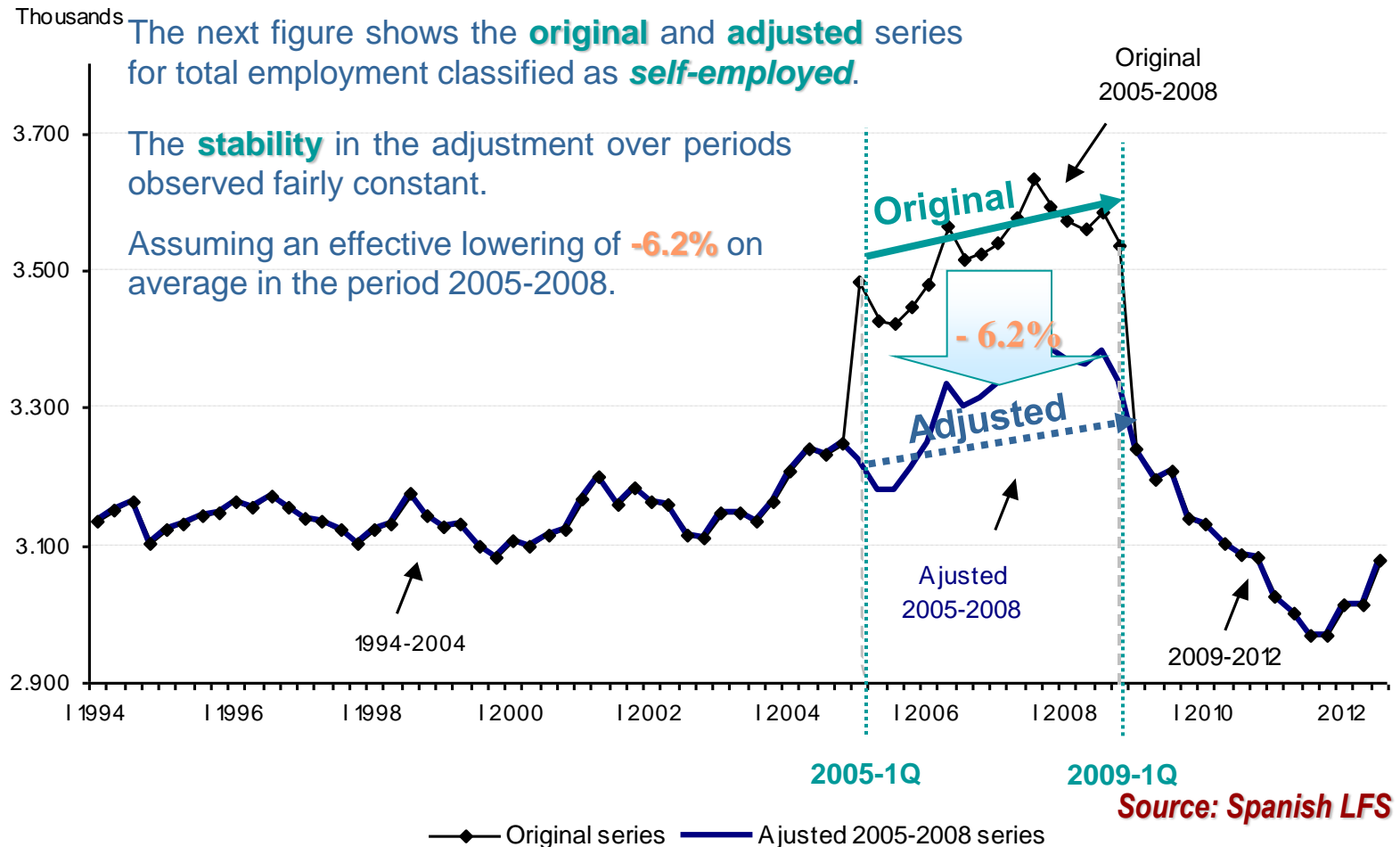
Receiver Operating Characteristic (ROC) curve



The area under the ROC curve is given by the “**c statistic**” and in this case is **0.825**.

## II.IV Final adjustment in 2005-2008 professional status LFS series

Figure 3. Self-employed workers: total



*The self-declared “Professional Status” in the Spanish LFS*

### **III. Final conclusions.**

LINE

### III. Final conclusions

The method of “**logistic regression**” allows to predict the behaviour of a (qualitative or discrete) **response variable** based on (qualitative or quantitative) **explanatory variables**.

This technique is especially useful in **social surveys** (like **LFS**) where most variables are **qualitative variables** with few quantitative variables and it can be used by statistical offices for:

1. **Imputation** of missing values in the questionnaire.
2. **Backcasting**. Micro conversion of the historical series.

In this case, this technique allows the **micro conversion** of **historical series** in some groups of **self-employed workers** through a **probabilities of reallocation** and according to the other variables in the **LFS questionnaire**.

**Thank you very much!**



**Javier Orche Galindo**  
**Miguel Ángel García Martínez**  
**Spanish Labour Force Survey (LFS)**

