Empirical estimation of non linear input-output models: an Entropy Econometrics approach

Esteban Fernández-Vázquez
University of Oviedo (Spain)

-Literature on non linear input output modeling is mainly theoretical (Lahiri, 1973; Lahiri and Pyatt, 1980; Chander, 1983; Fujimoto, 1986, or Dietzenbacher, 1994)

-Empirical estimation prevented by data availability: number of parameters to estimate higher than available data points

-CGE models: calibration techniques

- Proposal here: to estimate the parameters of non linear IO models

- Exploit new IO databases

- Econometric estimation based on Entropy Econometrics (EE), econometric technique suitable for ill-conditioned datasets

Outline of the presentation

1. A simple non linear IO model
2. Overview of entropy econometrics
3. Estimating non linear IO models with EE
4. An illustration: a non linear IO model for Spain
5. Final remarks

1. A simple non linear IO model

- The traditional linear IO model (1):

$$z_{ij} = (a_{ij})x_j$$

$$0 \leq a_{ij} \leq 1$$

$$\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{f} = \boldsymbol{L}\boldsymbol{f}$$

- A non-linear (scale-dependent) IO model (2):

$$z_{ij}(x_j) = \alpha_{ij} x_j^{\beta_{ij}}$$

$$\alpha_{ij}, \beta_{ij} \geq 0$$

$$\boldsymbol{x} = \boldsymbol{A}^*(\boldsymbol{x})\boldsymbol{x} + \boldsymbol{f} = \boldsymbol{L}^*(\boldsymbol{x})\boldsymbol{f}$$

based on Sancho and Guerra (2014)

- (2) is equivalent to (1) when β_{ij}=1

Multipliers with IO models

- A linear model:

$$\boldsymbol{r} = \hat{\boldsymbol{c}}\boldsymbol{x} = \hat{\boldsymbol{c}}(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{f} = \boldsymbol{M}\boldsymbol{f}$$

$$c_j = r_j / x_j$$

Scale-independent coefficients

- A model with scale-dependent multipliers:

$$\boldsymbol{r} = \hat{\boldsymbol{c}}^*(\boldsymbol{x})\boldsymbol{x} = \hat{\boldsymbol{c}}^*(\boldsymbol{x})\big(\boldsymbol{I} - \boldsymbol{A}^*(\boldsymbol{x})\big)^{-1}\boldsymbol{f} = \boldsymbol{M}^*(\boldsymbol{x})\boldsymbol{f}$$

$$c_j^*(x_j) = r_j / x_j = \alpha_j^r x_j^{\beta_j^r} / x_j = \alpha_j^r x_j^{(\beta_j^r - 1)}$$

Use of variable *r* depends on the output

Parameters (α, β and α^r, β^r) in the equations condition the effect of final demand shocks

2. Overview of entropy econometrics

- a discrete random variable that can take $M \geq 2$ values $X = \{x_1, \ldots, x_M\}$ with probabilities $\boldsymbol{p} = \{p_1, \ldots, p_M\}$. Shannon's entropy function measures the uncertainty in X:

$$\underset{\boldsymbol{p}}{\text{Max}}\, H(\boldsymbol{p}) = -\sum_{m=1}^{M} p_m \, ln(p_m)$$

- $H(\mathbf{p})$ achieves its unconstrained maximum for the uniform distribution. $H(\mathbf{p}) = 0$ means no uncertainty

- if $\boldsymbol{p}$ was unknown what would be our best guess? The distribution that maximizes H(**p**). If we have some information about X, we now maximize H(**p**) subject to (but only to) the information we have

-) The underlying ideas of the ME methodology can be applied for estimating the parameters of general linear models (GME)

-) The general form (T observations, H parameters) is:

$$y = X\beta + \epsilon$$

-) The beta parameters (and the error) do not behave as probabilities

-) each parameter β_h is modeled as a discrete random variable that can take M≥2 values

-) define a support vector ***b*** that contains the M≥2 possible outcomes of the random variable:

$$\boldsymbol{b}' = (b_1, \ldots, b_M)$$

-) each element of ***b***' has an unknown probability:

$$\boldsymbol{p'}_h = (p_{h1}, \ldots, p_{hM})$$

-) each parameter β_h is given by the following expression (same procedure for the errors):

$$\beta_h = \boldsymbol{b}'\boldsymbol{p_h} = \sum_{m=1}^{M} b_m \, p_{hm}; \; \forall h = 1, \ldots, H$$

The GME program

$$\underset{\boldsymbol{P},\boldsymbol{U}}{\text{Max}}\, H(\boldsymbol{P},\boldsymbol{U}) = -\sum_{h=1}^{H}\sum_{m=1}^{M} p_{hm}\, ln(p_{hm}) - \sum_{n=1}^{N}\sum_{j=1}^{J} u_{tj}\, ln(u_{tj})$$

subject to:

$$\boxed{y_t = \sum_{k=1}^{K}\sum_{m=1}^{M} b_m p_{hm} x_{kt} + \sum_{j=1}^{J} v_j\, u_{tj}\,;\ \ \forall t = 1, \dots, T}$$

$$\sum_{m=1}^{M} p_{hm} = 1\,;\ \forall h = 1, \dots, H$$

$$\sum_{j=1}^{J} u_{tj} = 1\,;\ \forall t = 1, \dots, T$$

The probabilities are estimated by maximizing entropy **conditional on the observations**

- once the ***p*** probabilities are recovered, we have point-estimates of the parameters:

$$\tilde{\beta}_h = \sum_{m=1}^{M} b_m \, \tilde{p}_{hm}; \ \forall h = 1, \dots, H$$

- under some mild assumptions, GME estimators are asymptotically consistent and normally distributed:

$$\widehat{\boldsymbol{\beta}} \rightarrow N[\boldsymbol{\beta}, \widehat{\boldsymbol{\sigma}}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}];$$

$$\widehat{\boldsymbol{\sigma}}^2{}_{\boldsymbol{h}} = \sigma_e^2 \left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_v^2} \right), \forall h = 1, \dots, H;$$

t-ratio statistics can be calculated

3. Estimating non linear IO models with EE

-EE can be applied to recover the parameters of nonlinear IO equations. We need:

- a dataset (a time series or a cross-section) of IO tables

- to assume that the production technology is constant along all the observations

- T observations of a $(n \times n)$ IO matrix, n^2 equations like:

$$\Delta ln(z_{ijt}) = \Delta\beta_{ij}\, ln(x_{jt}) + \varepsilon_{ijt}$$

- Support vectors:
  - "natural" centers and bounds
  - 3-sigma rule for the errors, with J=3, **v**=(-3s,0,3s)
  - with M=3, **b**=$(b_1,b_2,b_3)=(b_2-d,b_2,b_2+d)$=(0,1,2)

A global GME program for ($n \times n$) equations

$$Max\, H(\boldsymbol{P_b}, \boldsymbol{W})$$

subject to:

$$\boxed{\Delta ln\,(\boldsymbol{Z}) = \boldsymbol{BP}\Delta\, ln(\boldsymbol{X}) + \boldsymbol{VW}}$$

$$\boldsymbol{e}'\boldsymbol{p_{ij}} = 1; \forall i, j$$

$$\boldsymbol{e}'\boldsymbol{w_{ijt}} = 1; \forall i, j, t$$

Without any information the GME solution is the linear model (all β_{ij}=1)

$$\boldsymbol{e}'\widehat{\boldsymbol{A}}^{*}(\boldsymbol{x_t}) < 1$$

$$[\boldsymbol{I} - \widehat{\boldsymbol{A}}^{*}(\boldsymbol{x_t})]^{-1} = \left[\boldsymbol{I} + \widehat{\boldsymbol{A}}^{*}(\boldsymbol{x_t}) + [\widehat{\boldsymbol{A}}^{*}(\boldsymbol{x_t})]^{2} + [\widehat{\boldsymbol{A}}^{*}(\boldsymbol{x_t})]^{3} + \cdots\right]$$

4. An illustration: a non linear IO model for Spain

-A non-linear IO model will be estimated for Spain in order to estimate output and labor multipliers

- 1995-2009 series of annual industry-by-industry IO tables (from WIOD)

- 1995-2009 series of annual labor figures by industry (from WIOD)

- aggregation into 16 sectors:

$$z_{ijt}(x_{jt}) = \alpha_{ij} x_{it}^{\beta_{ij}} \quad \text{256 equations}$$

$$r_{jt}(x_{jt}) = \alpha_j^r x_{jt}^{\beta_j^r} \quad \text{16 equations}$$

Appendix: industry classification

Industry number	Industry description
i1	Agriculture, Hunting, Forestry and Fishing
i2	Mining and Quarrying
i3	Food, Beverages and Tobacco
i4	Textiles, Leather and Footwear products
i5	Wood, pulp and paper
i6	Coke, Refined Petroleum and Nuclear Fuel
i7	Chemicals, rubber, plastics and non-metallic mineral
i8	Basic Metals and Fabricated Metal
i9	Machinery, equipment and n.e.c. manufacturing
i10	Electricity, Gas and Water Supply
i11	Construction
i12	Sale, maintenance and trade
i13	Hotels and Restaurants
i14	Transport, post and telecommunications
i15	FIRE services and other business activities
i16	Other services

- Equations are estimated in FD

- Point estimates and t-ratio statistics for the (α, β) parameters are computed

- t-ratios are used to test for linearity

    - H_0: $\beta_{ij}=1$

    - H_1: $\beta_{ij}\neq 1$

Table 1: GME estimates of the β_{ij} parameters

$\widehat{\beta}_{ij}$	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16
i1	1.067	0.914	0.980	0.916	0.990	0.470**	1.089	0.838*	0.949	0.706**	0.590**	0.874	0.865	0.943	0.889	0.817
i2	1.038	1.117*	1.076	1.120	1.082	1.169	1.139	1.039	1.086	0.641**	0.932	0.972	1.004	1.001	0.983	0.998
i3	0.944	0.894	0.957	0.803*	0.914	0.561**	0.911	0.838	0.704**	0.911	0.780**	0.817	1.012	0.844	0.862	0.888
i4	1.006	0.987	1.026	1.057	1.060	0.454**	0.990	1.026	0.960	0.919	1.022	0.975	0.953	1.015	0.981	0.989
i5	1.059	0.892	0.897**	1.074	0.971	0.295**	0.981	0.959	0.952	0.948	0.874*	0.977	0.815	0.797*	0.827	0.908
i6	0.771	0.922	1.048	1.098	1.023	1.850**	0.910	0.976	0.944	0.942	0.747**	0.949	0.865	0.904	0.872	0.881
i7	0.906	0.829*	1.027	0.920	0.938	0.513**	0.974	0.886*	0.918	0.865	0.891*	0.993	0.510**	0.900	0.889	0.943
i8	1.027	1.003	0.967	1.093	1.033	0.381**	1.088	1.075**	1.001	0.971	0.951	1.073	0.850	0.978	0.974	0.945
i9	1.004	1.072	1.103	1.102	1.011	0.556**	1.059	1.055	1.079	0.949	0.979	1.075	0.980	1.052	0.988	0.965
i10	0.989	0.991	1.103	1.041	1.032	0.874	1.073	1.026	0.868	1.313**	0.908	1.058	1.071	0.963	1.023	1.061
i11	1.005	0.995	0.969	1.022	1.023	0.489**	1.055	0.941	0.912	1.046	1.103	1.054	0.959	1.115	1.021	1.041
i12	0.999	0.914	1.013	0.907*	1.050	1.319**	0.994	0.943	0.890	0.910	0.889**	0.982	1.044	1.051	0.951	1.031
i13	1.068	0.898	1.069	1.041	0.983	0.338**	0.985	0.845	0.907	0.978	0.974	0.943	1.005	0.983	0.927	1.003
i14	1.086	1.005	1.071	0.824	0.974	1.254	0.977	0.932	0.803**	1.115	0.935	1.045	1.100	1.055	0.933	1.032
i15	1.027	1.028	0.941	0.897	0.957	0.784	1.001	0.896	0.859	1.150	0.974	1.073*	1.084	1.053	1.024	1.069
i16	0.990	0.894*	1.066	0.936	1.051	0.519**	1.000	0.888	0.800**	1.052	0.987	0.988	1.082	1.098	1.008	1.106**

** and * stands for estimates significantly different from 1 at 5% and 10% respectively

- Confidence intervals for multipliers can be computed to complement our point estimates

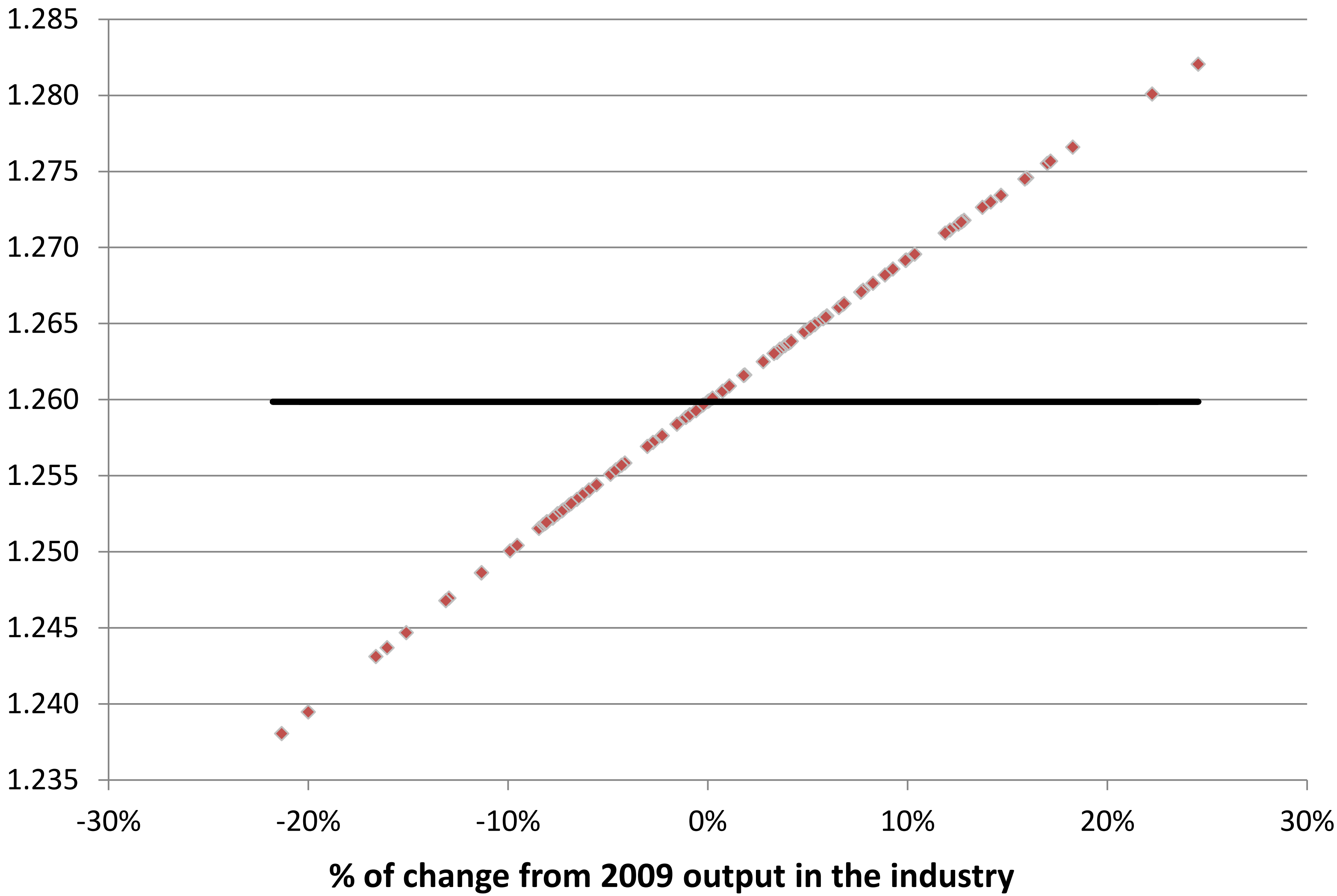
-Note that the multipliers now are scale-dependent: the effects of final demand shocks will be different depending on the output level

-Numerical simulation to explore the reaction of multipliers to changes in the output

- $x_{js} = x_{j2009} + N(0{,}0.1x_{j2009})$

- 100 simulated outputs for each industry j

Δ Output of Electricity, Gas and Water Supply/Δ final demand



Δ Output of Mining and Quarrying/Δ final demand Elec., Gas and Water Sup.

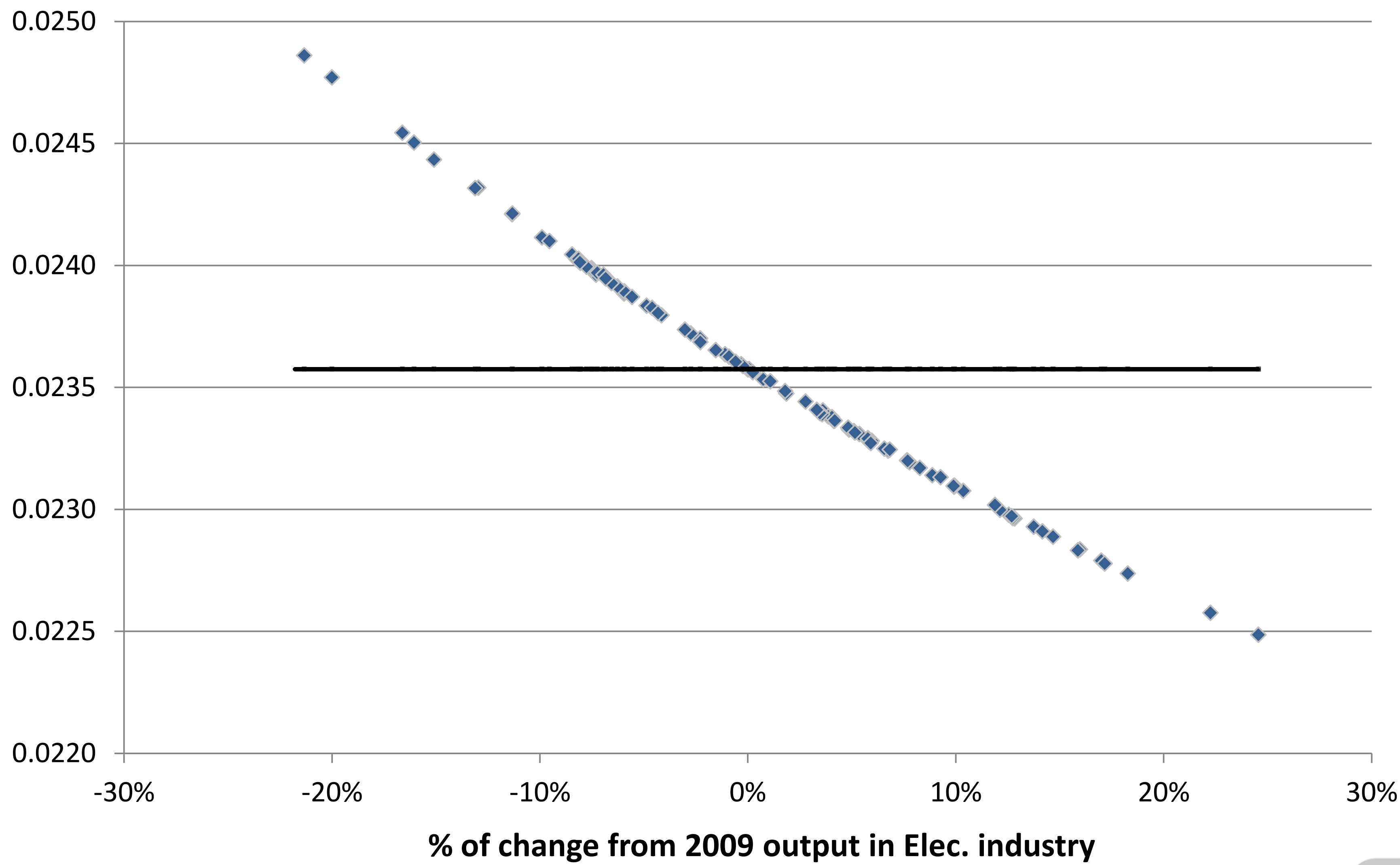
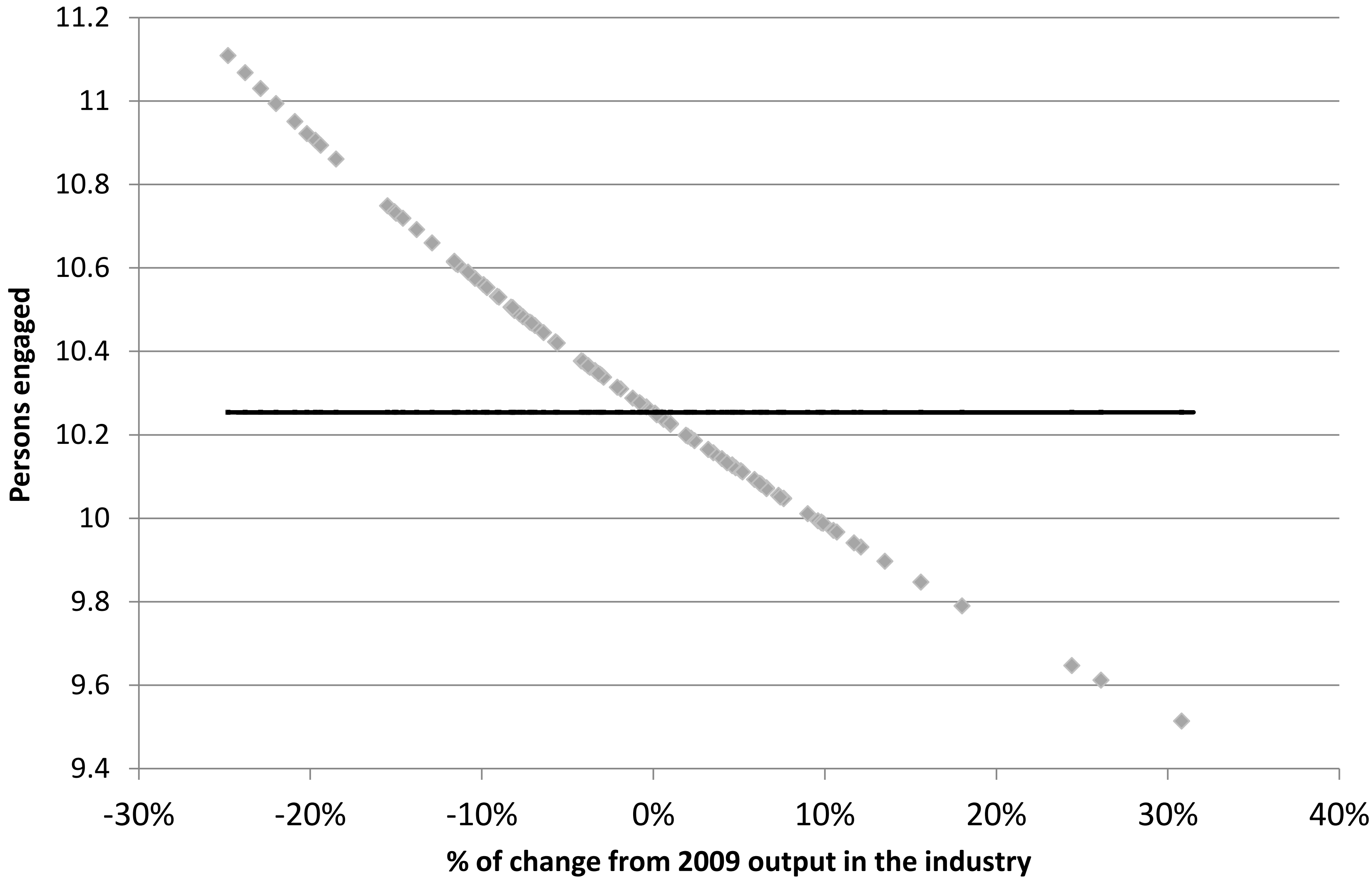


Table 2: GME estimates of the β_j^r parameters

Industry number	$\widehat{\beta}_j^r$
i1	0.651**
i2	0.754**
i3	0.654**
i4	0.773**
i5	0.694**
i6	0.649**
i7	0.691**
i8	0.748**
i9	0.696**
i10	0.633**
i11	0.688**
i12	0.667**
i13	0.656**
i14	0.639**
i15	0.654**
i16	0.625**

** and * stands for estimates significantly different from 1 at 5% and 10% respectively

Δ labor in Construction/Δ final demand



5. Final remarks (and future research lines)

- We have the data and the estimation techniques to apply non linear analysis in our IO framework
- Computationally more demanding but it would allow for more flexible analysis
- Some points in the research agenda:
  - More factors in our equations
  - Non linearities in the context of SUT's

THANK YOU

Esteban Fernández-Vázquez
University of Oviedo (Spain)
evazquez@uniovi.es