2015 UK Stata Users Group Meeting
Cass Business School, London, UK
September 10–11, 2015

rscore: a Stata module to compute Responsiveness Scores

Giovanni Cerulli

CNR-IRCrES
National Research Council of Italy
Institute for Research on Sustainable Economic Growth
Via dei Taurini 19, 00185 Roma
E-mail: g.cerulli@ceris.cnr.it

Introduction

Responsiveness Scores measure the change of a given outcome y when a given factor x_j changes, conditional on all other factors $\mathbf{x}_{-j}$.

It is the *derivative* of y on x_j, given $\mathbf{x}_{-j}$ (*regression coefficient*), but allowing each observation to get its own **responsiveness score** (*random coefficient regression*).

RSCORES: definition and estimation

Responsiveness Scores (RS) are obtained by an *iterated* ***Random Coefficient Regression*** **(RCR)**. The basic econometrics of this model can be found in Wooldridge (2002, pp. 638-642). The calculation of RS follows this simple protocol:

1. Define y, the outcome (or *response*) variable.
2. Define a set of Q factors thought of as affecting y, and indicate the generic factor with x_j.
3. Define a RCR model linking y to the various x_j, and extract a unit-specific *responsiveness effect* of y to the all set of factors x_j, with j=1, ... , Q.
4. For the generic unit i and factor j, indicate this effect as b_{ij} and collect all of them in a matrix **B**. Finally, aggregate by unit (row) and by factor (column) the b_{ij} getting synthetic unit and factor responsiveness measures.

Analytically, an RS is defined as the "**partial effect**" of an RCR (Wooldridge, 1997; 2002; 2005). Define a RCR model of this kind:

$$\begin{cases} y_i = a_{ij} + b_{ij} x_{ij} + e_i \\ a_{ij} = \gamma_0 + \mathbf{x}_{i,-j}\boldsymbol{\gamma} + u_{ij} \\ b_{ij} = \delta_0 + \mathbf{x}_{i,-j}\boldsymbol{\delta} + v_{ij} \end{cases}$$

where e_i, u_{ij} and v_{ij} are error terms with $E(e_i \mid x_{ij}) = E(u_{ij} \mid x_{ij}) = E(v_{ij} \mid x_{ij}) = 0$.

It is easy to show that the regression parameters, a_{ij} and b_{ij}, are both non constant as depending on all the other inputs x except x_j (this is, in fact, the meaning of the vector $\mathbf{x}_{i,-j}$). Observe that δ_0 and γ_0 are, on the contrary, constant parameters.

According to this model, we can define the **regression line** as:

$$E(y_i \mid x_{ij}, a_{ij}, b_{ij}) = a_{ij} + b_{ij} x_{ij}$$

From this, we define the **RS** of of x_{ij} on y_i as the *derivative* of y_i respect to x_{ij}, that is:

$$\frac{\partial}{\partial x_{ij}} \left[E(y_i \mid x_{ij}, a_{ij}, b_{ij}) \right] = b_{ij}$$

where: b_{ij} is called the ***partial effect*** of x_{ij} on y_i.

We can repeat the same procedure for each x_{ij} (j=1, ..., Q) so that it is possible eventually to define, for each unit i =1 ..., N and factor j=1, ..., Q, the N x Q matrix **B** of "partial effects" as follows:

$$\mathbf{B} = \begin{pmatrix} b_{11} & \cdots & b_{1Q} \\ \vdots & b_{ij} & \vdots \\ b_{N1} & \cdots & b_{NQ} \end{pmatrix}$$

If all variables are standardized, partial effects are **beta coefficients** so that they are independent of the unit of measurement and can be compared and summed.

Once matrix **B** is known, we can define for each unit *i* the Total Unit Responsiveness (TUR) and the Mean Unit Responsiveness (MUR) as:

$$\text{TUR}_i = \sum_{j=1}^{Q} b_{ij} \quad \text{and} \quad \text{MUR}_i = \frac{1}{Q}\sum_{j=1}^{Q} b_{ij}$$

and for each factor *j*, the Total (or Mean) Responsiveness of *y* to factor *j*'s unit change (TFR and MFR) as:

$$\text{TFR}_j = \sum_{i=1}^{N} b_{ij} \quad \text{and} \quad \text{MFR}_j = \frac{1}{N}\sum_{i=1}^{N} b_{ij}$$

In a **cross-section** data setting, the estimation of each b_{ij} can be done by **Ordinary Least Squares** of this regression:

$$y_i = \gamma_0 + \mathbf{x}_{i,-j}\boldsymbol{\gamma} + (\delta_0 + \bar{\mathbf{x}}_{-j}\boldsymbol{\delta})x_{ij} + x_{ij}(\mathbf{x}_{i,-j} - \bar{\mathbf{x}}_{-j})\boldsymbol{\delta} + \eta_i$$
$$\eta_i = u_i + x_{ij}v_i + e_i$$

where: $\bar{\mathbf{x}}_{-j}$ is the vector of the sample means of $\mathbf{x}_{i,-j}$.

Once previous regression parameters have been estimated, we can get for the generic unit i an estimation of the partial effect of factor x_j on y as:

$$\hat{b}_{ij} = \hat{\delta}_0 + \mathbf{x}_{i,-j}\hat{\boldsymbol{\delta}}$$

By repeating this procedure for each unit i and factor j, we can finally obtain $\hat{\mathbf{B}}$, the estimation of matrix **B**.

When a **longitudinal** dataset is available, the estimation of **B** can be obtained either by using random-effect or fixed-effects estimation of this **panel regression**:

$$y_{it} = \gamma_0 + \mathbf{x}_{it,-jt}\boldsymbol{\gamma} + (\delta_0 + \bar{\mathbf{x}}_{-jt}\boldsymbol{\delta})x_{ijt} + x_{ijt}(\mathbf{x}_{it,-jt} - \bar{\mathbf{x}}_{-jt})\boldsymbol{\delta} + \alpha_i + \eta_{it}$$

where the added parameter α_i represents a **unit-specific effect** accounting for *unobserved heterogeneity*.

Fixed-effect estimation, by assuming free correlation between α_i and η_{it}, can mitigate a **potential endogeneity** bias due to misspecification of previous equation and measurement errors in the variables considered in the model (Wooldridge, 2010, pp. 281-315).

As such, a panel dataset allows for more reliable estimates of the true responsiveness scores than usual OLS.

The Stata command rscore

```
help rscore
-------------------------------------------------------------------------------------------------------------------------------

Title

    rscore-      Estimation of responsiveness scores


Syntax

        rscore outcome [varlist] [if] [in] [weight], model(modeltype) [factor(varlist_f) xlist(varlist_c)]


    fweights, iweights, and pweights are allowed; see weight.



Description

    rscore computes unit-specific responsiveness scores using an iterated Random-Coefficient-Regression (RCR).  The basic econometrics of this model can
    be found in Wooldridge (2002, pp. 638-642).  The model estimated by rscore considers a regression of a response variable y, i.e. (outcome), on a
    series of factors (or regressors) x, i.e. varlist, by assuming a different reaction (or "responsiveness") of each unit to each factor contained in x.
    rscore allows for: (i) ranking units according to the level of the responsiveness score obtained; (ii) detecting factors that are more influential in
    driving unit performance; (iii) studying, more in general, the distribution (heterogeneity) of the factors' responsiveness scores across units.



Options

    model(modeltype) specifies the model to be estimated, where modeltype must be one of the following models: "ols", "fe", "re". It is always required to
        specify one model.

    factor(varlist_f) specifies that factor variables have to be included among the regressors. It is optional for both models.

    xlist(varlist_c) specifies that control variables (which are not factors) have to be included among the regressors. It is optional for both models.

  modeltype_options            Description
  -----------------------------------------------------------------------------------------------------------------------------
  Model
  ols                          regression estimated by ordinary least squares (OLS)
  fe                           panel data fixed-effect regression (FE)
  re                           panel data random-effect regression (RE)
  -----------------------------------------------------------------------------------------------------------------------------
```

rscore creates a number of variables:

_bvarname is the responsiveness scores variable related to varname. They are as many as the variables and/or factors considered in model specification.

Remarks

Please remember to use the update query command before running this program to make sure you have an up-to-date version of Stata installed.

Examples

```
. rscore y x1 x2 x3 , model(ols) factor(f1 f2)
. rscore y x1 x2 x3 , model(fe)  factor(f1 f2) xlist(x4 x5)
```

Reference

Wooldridge, J. M. 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge.

Author

Giovanni Cerulli
IRCrES-CNR
Research Institute for Sustainable Economic Growth, National Research Council of Italy
E-mail: giovanni.cerulli@ircres.cnr.it

Also see

Online: ivregress

Application

ELSEVIER
www.elsevier.com/locate/worlddev

World Development Vol. 59, pp. 147–165, 2014

0305-750X/$ - see front matter

http://dx.doi.org/10.1016/j.worlddev.2014.01.019

The Impact of Technological Capabilities on Invention: An Investigation Based on Country Responsiveness Scores

GIOVANNI CERULLI *
Ceris-CNR, Roma, Italy

Summary. — This study explores the impact of "technological capabilities" (TCs) on invention (measured by "patenting intensity") in a dataset of 42 emerging and advanced countries observed over 13 years (1995–2007). By computing country responsiveness scores we are able to: (i) rank countries according to their inventive responsiveness; (ii) detect more influential TCs factors; (iii) test the presence of increasing/decreasing patenting returns to TCs. Results show an inverted-U relation between invention responsiveness and TCs intensity. We conclude that self-reinforcing mechanisms characterize the early stage of TCs accumulation (increasing returns), and weakening mechanisms higher levels of TCs intensity (decreasing returns). Findings are widely discussed.

Key words — technological capabilities, responsiveness scores, patenting returns, countries' technological performance

Illustrative example: drivers of GDP growth

Research questions

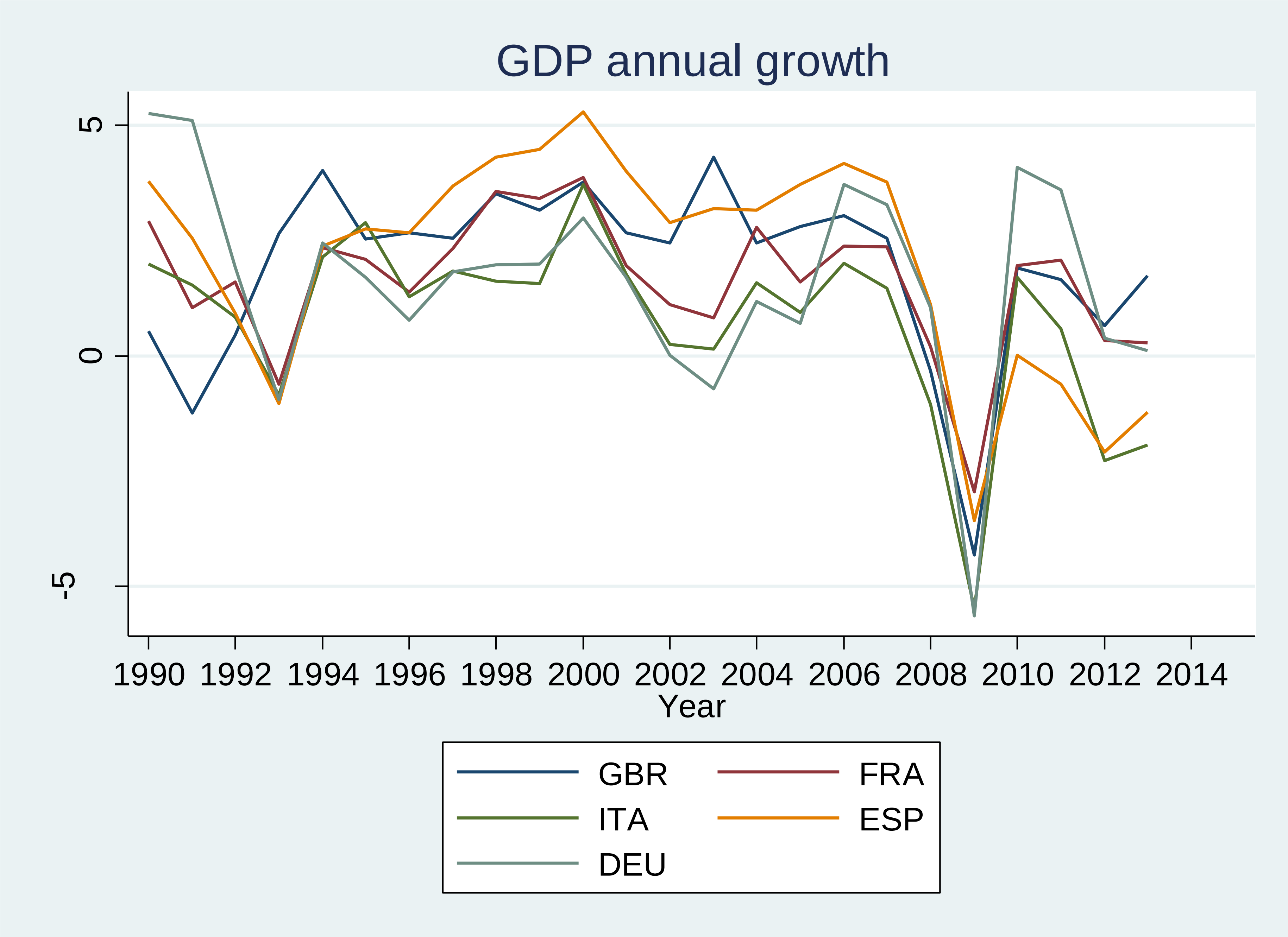
Factor importance rank: Among countries' GDP components, what are those whose growth change produces larger/smaller response in terms of GDP growth?

Factor response heterogeneity: Is country GDP growth response more or less heterogeneously/homogenously distributed among its driving factors?

Unit responsiveness rank: which units do have larger/smaller responsiveness scores for each given factor?

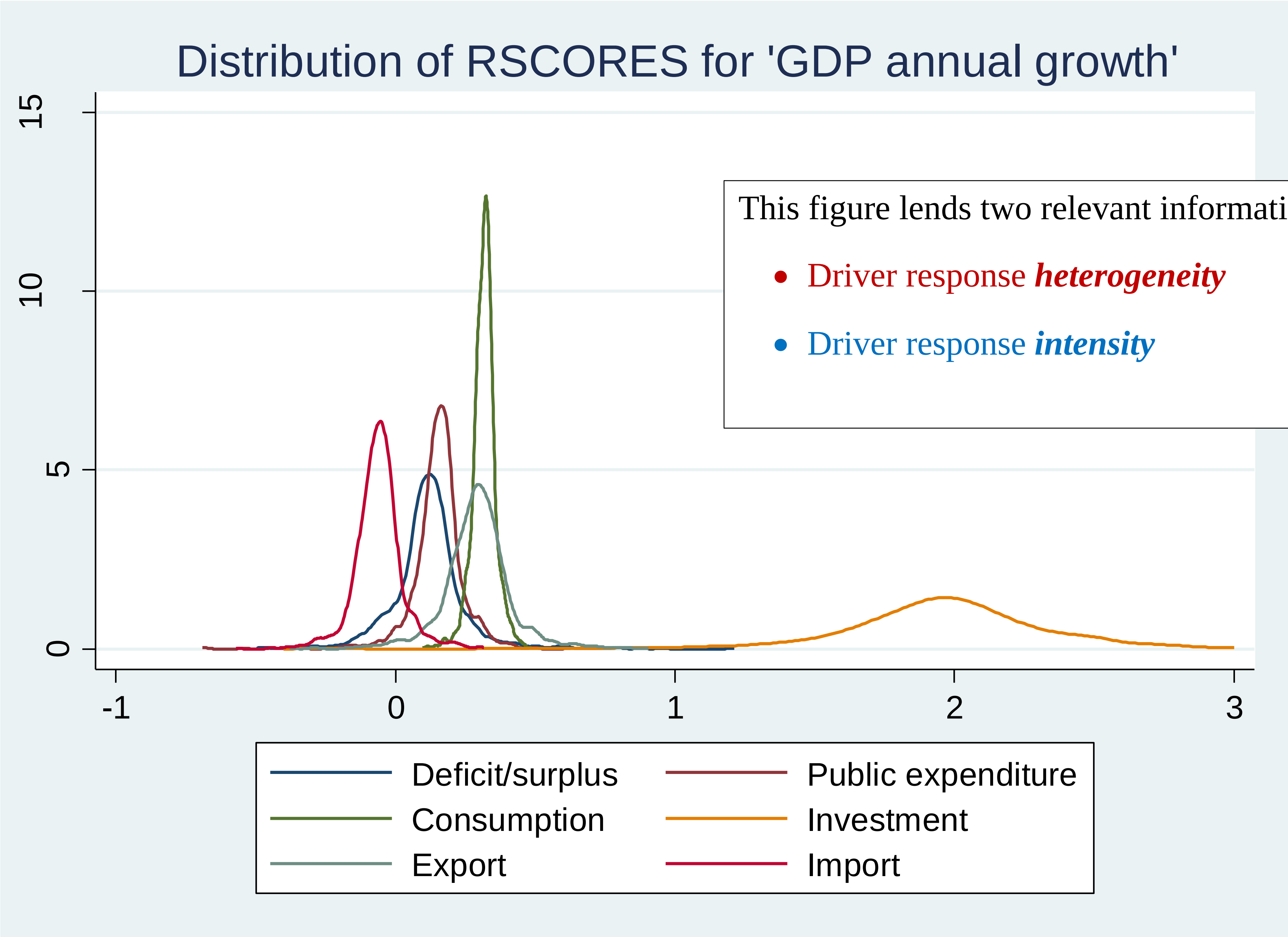
Dataset: World Bank's "Economy & Growth" indicators (283 macroeconomic indicators, 250 countries, 1960-2014, 13,695 observation, longitudinal).

```
************************
* GDP annual growth
************************
global time year>=1990
tw ///
line ny_gdp_mktp_kd_zg year  if countrycode == "GBR" & $time , sort ||  ///
line ny_gdp_mktp_kd_zg year  if countrycode == "FRA" & $time , sort ||  ///
line ny_gdp_mktp_kd_zg year  if countrycode == "ITA" & $time , sort ||  ///
line ny_gdp_mktp_kd_zg year  if countrycode == "ESP" & $time , sort ||  ///
line ny gdp mktp kd zg year  if countrycode == "DEU" & $time , sort     ///
xlabel(1990(2)2015, gmax angle(horizontal)) ///
legend(label(1 "GBR") label(2 "FRA") label(3 "ITA") label(4 "ESP")label(5 "DEU")) ///
title("GDP annual growth")
```


GDP annual growth
5
0
-5
1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014
Year
GBR
FRA
ITA
ESP
DEU

```
**************************************************************
* Estimate RSCORES for the "GDP annual growth" (Y)
**************************************************************
global xvars B_std  G_std  C_std  I_std  E_std  M_std
* OLS
rscore  Y_std $xvars , model(ols)
order _b*
```

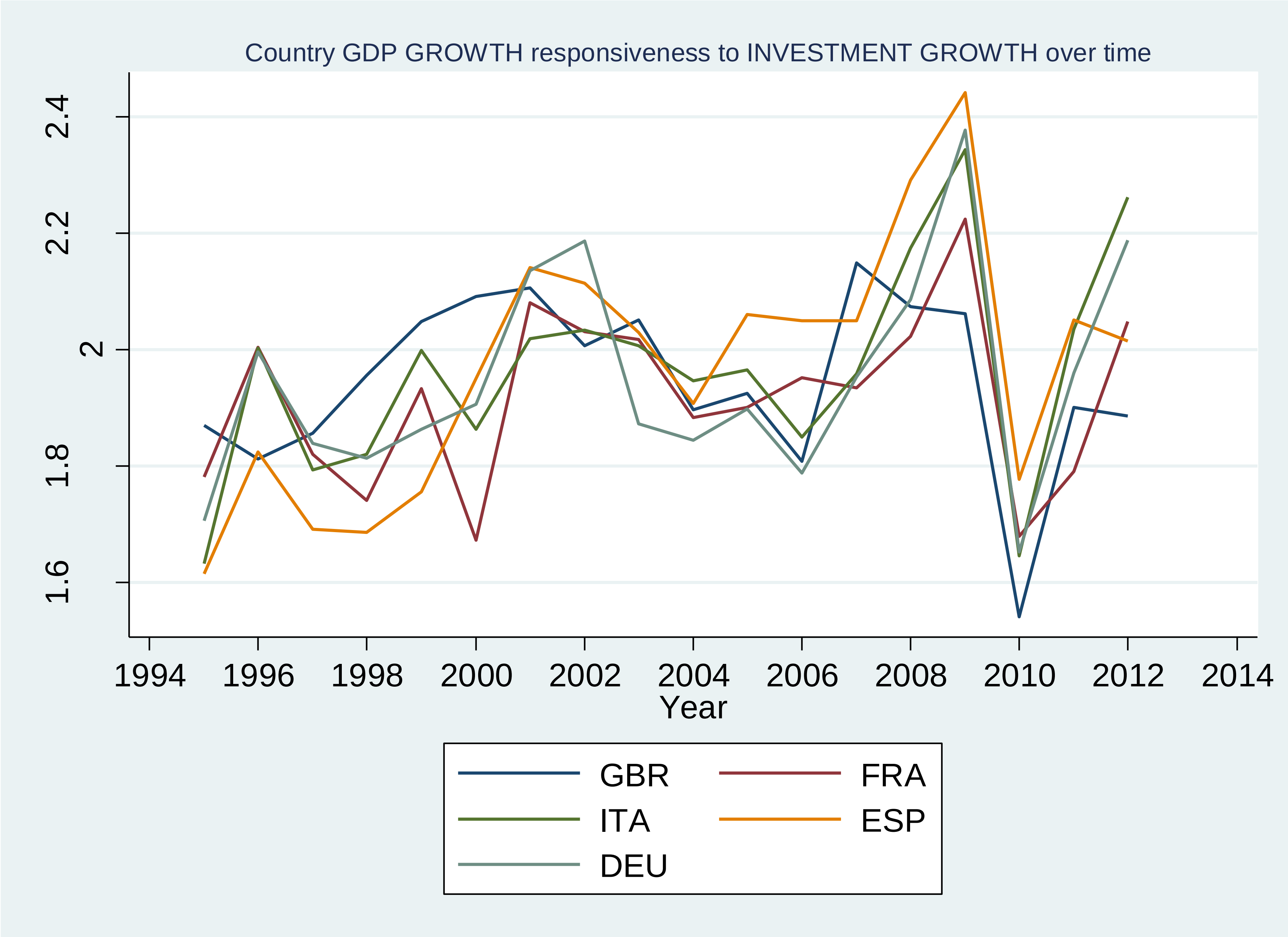
```
**********************************************
* Distribution of RSCORES for 'GDP annual growth'
**********************************************
global cond  if _bx5>-0.5 & _bx1>-0.5  & _bx4>-0.5 & _bx4<3
tw ///
kdensity _bx1  $cond   ||  ///
kdensity _bx2  $cond   ||  ///
kdensity _bx3  $cond   ||  ///
kdensity _bx4  $cond   ||  ///
kdensity _bx5  $cond   ||  ///
kdensity _bx6  $cond    ,  ///
xtitle("") legend(label(1 "Deficit/surplus") label(2 "Public expenditure") ///
label(3 "Consumption") label(4 "Investment") label(5 "Export") label(6 "Import")) ///
title("Distribution of RSCORES for 'GDP annual growth'")
```



This figure lends two relevant information:

- Driver response ***heterogeneity***
- Driver response ***intensity***

```
* Country "GDP growth" responsiveness to "GROWTH IN FIXED INVESTMENTS" over time
global time _bx4!=.
tw ///
line _bx4 year  if countrycode == "GBR" & $time , sort ||  ///
line _bx4 year  if countrycode == "FRA" & $time , sort ||  ///
line _bx4 year if countrycode == "ITA" & $time  , sort ||  ///
line _bx4 year if countrycode == "ESP" & $time  , sort ||  ///
line _bx4 year  if countrycode == "DEU" & $time , sort ///
xlabel(1994(2)2014, gmax angle(horizontal)) ///
legend(label(1 "GBR") label(2 "FRA") label(3 "ITA") label(4 "ESP")label(5 "DEU")) ///
title("Country GDP GROWTH responsiveness to INVESTMENT GROWTH over time", size(small))
```


Country GDP GROWTH responsiveness to INVESTMENT GROWTH over time
1.6
1.8
2
2.2
2.4
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
Year
GBR
FRA
ITA
ESP
DEU

Unit responsiveness rank

```
set more off
sort _bx4
list countryname year _bx4 if _bx4>=3 & _bx4!=.
```

```
      +---------------------------------------+
      |        countryname   year        _bx4 |
      |---------------------------------------|
1864. |            Belarus   1992    3.230418 |
1865. |         Azerbaijan   1999    3.232215 |
1866. |               Mali   2006    3.271389 |
1867. |        Congo, Rep.   2008    3.311152 |
1868. |         Seychelles   2008    3.314989 |
      |---------------------------------------|
1869. |            Nigeria   2012    3.334117 |
1870. |   Macao SAR, China   2009    3.413818 |
1871. | Trinidad and Tobago   2007     3.46852 |
1872. |          Indonesia   1999    3.514539 |
1873. |          Argentina   2002    3.515946 |
      |---------------------------------------|
1874. |           Bulgaria   1991    3.667125 |
1875. |  Iran, Islamic Rep.   1994    3.690769 |
1876. |           Bulgaria   1990     4.40006 |
1877. |            Nigeria   2004    5.845391 |
      +---------------------------------------+
```

Conclusions

`rscore` can be useful to detect both *factor importance* and *factor heterogeneous response*

`rscore` allows to *fixed-effect* estimation to mitigate potential factor *endogeneity*

`rscore` allows to rank both factors and observations, thus providing more detailed idiosyncratic information