An Alternative to Represent Time Series: "the Time Scatter Plot"

Alberto Calderero[1], Hanna Kuittinen[1], and Javier Fernández-Macho[2]

[1] Innovation Systems Unit, LABEIN-TECNALIA
acalderero@labein.es
[2] Department of Applied Economics III, University of the Basque Country

Abstract. This paper presents a substantial improvement for the representation of panel data. The time scatter plot is an alternative type of graph using Cartesian coordinates to display values across the time for two variables of a set of data. The data is displayed as a collection of points, linked by lines and ending by an arrow, each having the value of one variable determining the position on the horizontal axis and the value of another variable determining the position on the vertical axis. The purpose of the lines and the final arrow is to link the consecutive points in time for each case, resulting in a chain of lines in the shape of an irregular arrow which starts in the first period of time and ends in the last. This paper illustrates the usability of time scatter plot with two case studies 1) Research and development as a driver of innovation and growth and 2) The goal of social and economy convergence in Europe.

Key words: time scatter plot, convergence plot, graphing time series, plotting tools.

1 Introduction

Graphical demonstration is a very powerful means to present statistical data. For presenting a long-term development in time or relationship between two variables a graph can compress the data in a very interpretable manner for the use of decision support (e.g. policy makers, managers of companies). As the end-users of statistics are not necessarily very familiar with statistical models, a well design and structured graph can give the requested information in a form that is very easy to understand.

This paper presents a substantial improvement from the deficient present graphical representation of panel data (in Gretl and probably in other statistical packages). Gretl's graphical representation of panel data simply shows each of the "groups"=cases sequentially (with the option of individual graphs if the number of cases is small), which is not satisfactory when the number of cases is relatively high. Besides, this is only useful to present the time evolution of some cases, whilst there is nothing to visualize the evolution of the relationship between two characteristics of the whole set of cases in the panel, which is precisely what the present instrument does.

Our graphical proposal shows an alternative way to represent time series in a scatter plot. When compared to classical scatter plot, the advantage of so called time scatter plot is in its representation capacity. It can show two variables and their evolvement in time for several cases (or individuals) using a simple variation of a traditional scatter plot.

The idea of time scatter plot came up from the practical need to compare the evolution and the relationship of gross domestic product per capita and research and development expenditures of European countries in the same graph. Therefore, our efforts were focused on finding a graphical way to represent simultaneously these two variables and their evolution in time for a subset of European countries.

Traditional scatter plots allow us to represent only two variables (or dimensions) in the same moment, whereas a graph of time series shows the evolution in time for different variables of the same case or only for one variable of different cases. The classical graphing tools are not able to combine in a unique graph the statistical information that we wanted to show. This was the main reason to develop a variant of the traditional scatter plot.

The time scatter plot is an alternative type of graph using Cartesian coordinates to display values across the time for two variables of a set of data. The data is displayed as a collection of points, linked by lines and ending by an arrow, each having the value of one variable determining the position on the horizontal axis and the value of another variable determining the position on the vertical axis. The purpose of the lines and the final arrow is to link the consecutive points in time for each case, resulting in a chain of lines in the shape of an irregular arrow which starts in the first period of time and ends in the last.

The time scatter plot enables a researcher to obtain a visual comparison of two variables in time and helps to determine what kind of relationship there is between the two variables. This approach also allows to integrate in a two axis plot a third additional dimension, the time, providing more information and interpretability to the figure. We present two empirical application of the time scatter plot to demonstrate its usability in practise. The first example demonstrates the already mentioned relationship of R&D expenditures and GDP in Europe for 10 years period of 1995-2005. The second case shows the GDP convergence process of certain European regions.

2 Methodology

The objective of a chart must be to convey the major story being revealed by the data in an unambiguous and illuminating form, transmitting to the researcher or observer the major quantity of information about a set of data in an interpretable, clear and accurate way.

The basis of our alternative display is the simultaneous representation of two variables for each case in the set of data using Cartesian coordinates and including in the same graph the corresponding points to each one of both time series. Until this point we only have a structure of typical scatter plot with the special characteristic that there are represented points in different moments of time.

Our value added to the graph is that we link the lines between each point of a case in a moment with the next point in time for the same case marking the points in time with arrows. In this way we obtain a chain of linked lines creating an irregular arrow for each case. These arrow shapes start in the first period of the time series and end in the last available period (missing values in the time series can be supported by extrapolation). Therefore, the time scatter plot represents as many chains of arrows as the number of cases of the data. The direction of these arrows shows the evolution in the time of each case for the two selected variables and the position of each point show the relationship between the variables.

Another distinguishing feature of our alternative display is the possibility of divide our chart in four quadrants from the start and/or the end point of a target case. In our approach it was relevant to compare the evolution of certain countries with a concrete target, the European average. Both divisions in quadrants, one related on the initial period and other related on the last period of the objective case, can be also interpreted in terms of speed of convergence or divergence to the target. This is a powerful characteristic of our chart raising the information and improving the interpretability of the chart.

Depending on each variable incorporated to the graph the quadrants can be interpreted in different manners. The context and the domain of study contribute to determinate how the distinct quadrants must be interpreted.

3 Implementation

Several programs provide excellent statistical and graphical interfaces to edit and compose different kind of pre-formatted graphics for compare and analyse information. The problem appears when the idea that you have is impossible to implement with these standard formats or when you need to insert certain additional information to the plot. In all of these cases, open-source software offers a great opportunity to develop and design new alternatives exploiting the code already existing. A clear example of this is Gretl and other statistical software like R-project.

Gretl [1] has implemented several models of graphs to represent the most important kinds of statistical output, in particular, those related to cross section, time series and panel data analysis. Also, with the command line window, the user can add and modify the standard graphs by including additional parameters to the call of the gnuplot function. This function controls the transference of information between Gretl and a plotting utility, Gnuplot, which is used to generate graphs.

Gnuplot [2] is a portable command-line driven interactive data and function plotting utility for different operative systems. The software is copyrighted but freely distributed and support many uses, including web scripting and integration as a plotting engine for third-party applications like Octave, Gretl, etc. This means that any display or graphical model built in Gnuplot could be supported by Gretl. This is a great opportunity to generate new models and graphical designs to explain statistical information, which could be provided by Gretl.

For the implementation of the time scatter plot we have different options. One of them is to download the source code of Gretl and implement the new model of graph by the modification of those functions involved in the plotting process (mainly, in graphing file). An easier option is however to control the flow of information by scripting or launching the command line console of Gretl. Gnuplot function in Gretl accepts literal commands to control the appearance of the plot. In this way it is possible to build the proposed time scatter plot by editing the script with the appropriate commands. These scripts need a lot of numerical information and commands which make them preferable to write by code. Finally, for control in a more detailed way the output, as in this case, it is possible to operate directly, and with the same commands, in Gnuplot application. An easy R-script can generate, as in this case, the literal code for Gnuplot (also the graphical tool of R can support this kind of representation, as in the original version).

4 Practical application

Next two different practical applications of the alternative display are shown, the first of them relative to the R&D effort and its influence in the GDP per capita of the European states and the second relative to one of the traditional objectives of the European Union, the social and economic convergence, analyzing the current situation and the speed of convergence of certain regions of the EU.

4.1 Research and development as driver of innovation and growth

In the context of a global economy and from the Lisbon Strategy point of view, research is a component of a knowledge triangle (the other two being education and innovation) meant to boost growth and employment in the European Union [3], [4].

Relevance of the case Knowledge has been increasingly acknowledged to be a main driver of economical growth [5], [6]. Consequently, the member states of European Union agreed in 2002 to increase the research and development investments following the Lisbon Strategy. Specifically they set the goal of R&D expenditure to reach 3% of GDP by 2010 [7].

The European Commission intends to play a central role in driving and coordinating European research across the Framework Programmes to have a leverage effect on national research spending, in order to achieve the objective of 3% R&D expenditure of GDP on research in Europe [8].

Analyse the current position and speed of convergence with regard to this target is a relevant purpose for the Commission, comparing the effort of the European nations in R&D expenditures (percentage of GDP) and its impact on the income expressed by GDP per capita.

From a theoretical point of view, it is expected that high level of investments in research and development are resulting in higher level of GDP in future. On the other hand, also the countries with higher initial level of GDP are prone to have relatively higher level of investments in R&D because of more capital available for investing and vice versa the countries with low initial level of GDP are not expected to have very high level of R&D investments.

Showing the results The following time scatter plot shows the R&D expenditures (% of GDP) and GDP per capita (euros) of European Union member countries in years 1995-2005. Each country is having a line with arrows marking each year. The starting point for each arrow is the level of R&D expenditures and GDP per capita in year 1995 (expect for certain countries with missing values in the first periods) and accordingly, the ending point of the arrow remarks the levels of year 2005 (for all the countries).

For a better representation of the points and arrow, we use in this case a square scale in both axes. The black lines crossing the axes are representing the EU-27 averages and the red line crossing the y-axis shows the R&D expenditure target for 2010.

Generally the graph demonstrates rather linear relationship between R&D expenditures and GDP of EU-27 countries -the higher the level of R&D expenditures, the higher the GDP per capita.

Quadrant A shows the observations with higher than EU average spending on R&D but lower than average GDP per capita. As expected, there are very few points in this quadrant. Similarly, there are only view countries with high

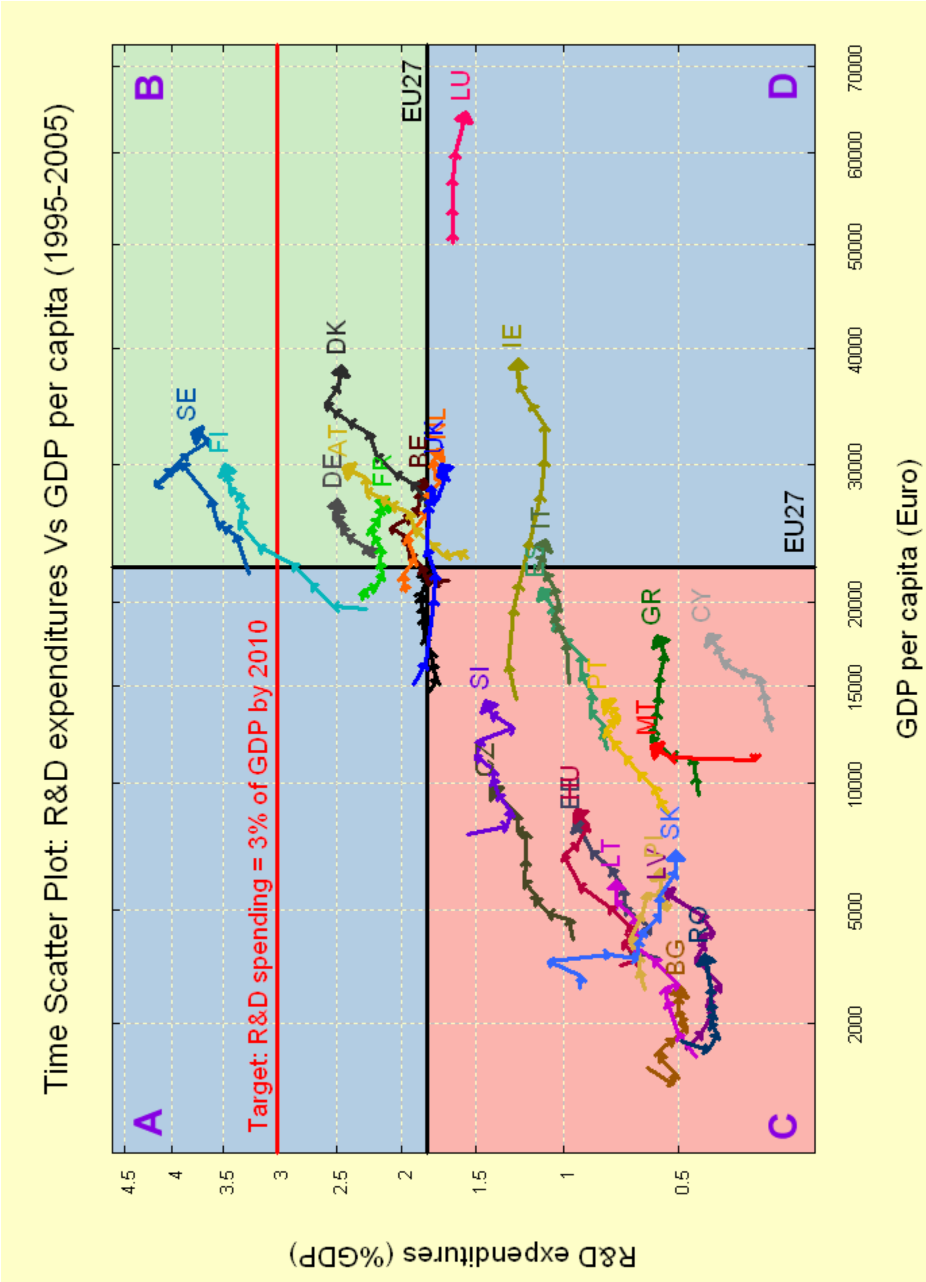

Fig. 1. Time scatter plot: R&D expenditures vs. GDP per capita (1995-2005)

GDP per capita but lower than average spending on R&D as it can be seen from quadrant D.

Most of the EU-15 countries are in quadrant B implying that they have high level of investment on R&D and rather high level of GDP. Only two countries, Sweden and Finland, have reached the EU R&D spending target of 3% but also others like Denmark, Austria and Germany are showing a rather fast growth in R&D expenditures.

The quadrant C represents the countries with below the average R&D spending and GDP per capita lower than EU-27. Relative increases in both R&D expenditure and GDP per capita are the highest in this quadrant.

4.2 The goal of social and economy convergence in Europe

Gross Domestic Product per capita is an excellent indicator for measuring the level and speed of economic and social convergence in Europe. One of the objectives of EU is to increase the welfare of its citizens, reducing the income disparities between different countries and regions.

Relevance of the case Economical convergence has been in interest of economic research for long -first by neo-classical growth theory and later by new growth theory-. Neo-classical growth theory [9] rationalized convergence by diminishing returns on investments of physical capital, thus implying faster growth rates for countries with lower initial level of capital.

The theory expected the economical growth in long-run to be dependent only on exogenous production factors (availability of labour, technological development) and economical convergence between countries occurred when the levels of these factors were equal between economies. However, the majority of empirical evidence was not consistent with this theory as it seemed that richer countries were growing faster than poorer ones.

So called new growth theory on the other hand believed that the growth of economy is endogenous and depends on new knowledge created by the accumulated capital [10], [11]. Instead of diminishing returns, new growth theory suggests that the returns remains steady thus benefiting the countries with higher initial level of capital. The theory also points out how public policies can effect on growth by governmental investments on R&D [12].

European Union's 27 member states form an internal market for 493 million citizens. The economical and social inequalities are still however very large amongst the 268 regions of Europe. Monitoring the convergence or divergence of EU regions is very important as union's ultimate goal is to promote economical, social and territorial cohesion [13].

The recent enlargements of European Union have significantly increased the economical disparities between member states and its regions. The richest member state (measured by per-capita income) Luxembourg, has seven times larger income per capita than Romania. The differences are even larger at regional level where Inner London reaches 290% of EU-27 income per capita average and at the same time the poorest region Nord-Est Romania has only the level of 23% of EU-27 average [14].

To respond this economical disparity of the EU regions, the European Council agreed on the budget of $347 billion on Structural and Cohesion Funds for the period of 2007-2013 [15]. In total, the structural and cohesion funds yield about 30% share of the budget of EU. Majority of these funds (81.5%) are allocated to convergence regions and targeted on investments on infrastructure and human capital. These investments are expected to result on faster growth rate of economy and better employment performance. Consequently, there is a need for monitoring the regional development policies and especially the economical catch up speed of the convergence regions.

Showing the results The Table 1 presents the list of selected regions, all from NUTS 2 level, which we included in the second exercise for showing the applicability of the time scatter plot. In all case, the researcher is the one who must select the number and adequacy of the cases to be represented in the plot.

The criteria for selecting one or other case should answer to the interest or relevance of the study. In our case it was very important to analyze the trend of Basque Country during the last years because of the great effort executed by the Basque governance institutions for reaching the EU15 average. The other regions are shown as reference to our case of interest, facilitating the development of benchmarking activities to identify best practices in other leader regions.

The second time scatter plot (Fig. 2) represents the convergence of GDP per capita for selected EU-15 regions. In this graph, the axes are GDP per capita (euro) and its annual variation rate (%). This figure helps to analyse the convergence (or divergence) of different regions in comparison with the European average.

The European averages of GDP and its growth rates for both the beginning (1996) and ending periods (2005) are marked with black lines crossing the axes. Each line with arrows marks the development of a selected European region, start point of the line showing the level and growth rate of GDP at beginning period (1996) and the arrow ending the line pointing the situation at ending period (2005).

Table 1. Set of regions in the analysis by code (NUTS)

Code	Region Name	Country
DE11	Stuttgart	Germany
DE21	Oberbayern	Germany
DE24	Oberfranken	Germany
DED2	Dresden	Germany
ES21	Basque Country	Spain
FI18	Etelä-Suomi	Finland
GR43	Kriti	Greece
NL11	Groningen	Netherlands
NL31	Utrecht	Netherlands
PT11	Norte	Portugal
PT15	Algarve	Portugal

To explain the second plot we must characterize the different quadrants in the graph, especially, those determinate in the last period of time by European average.

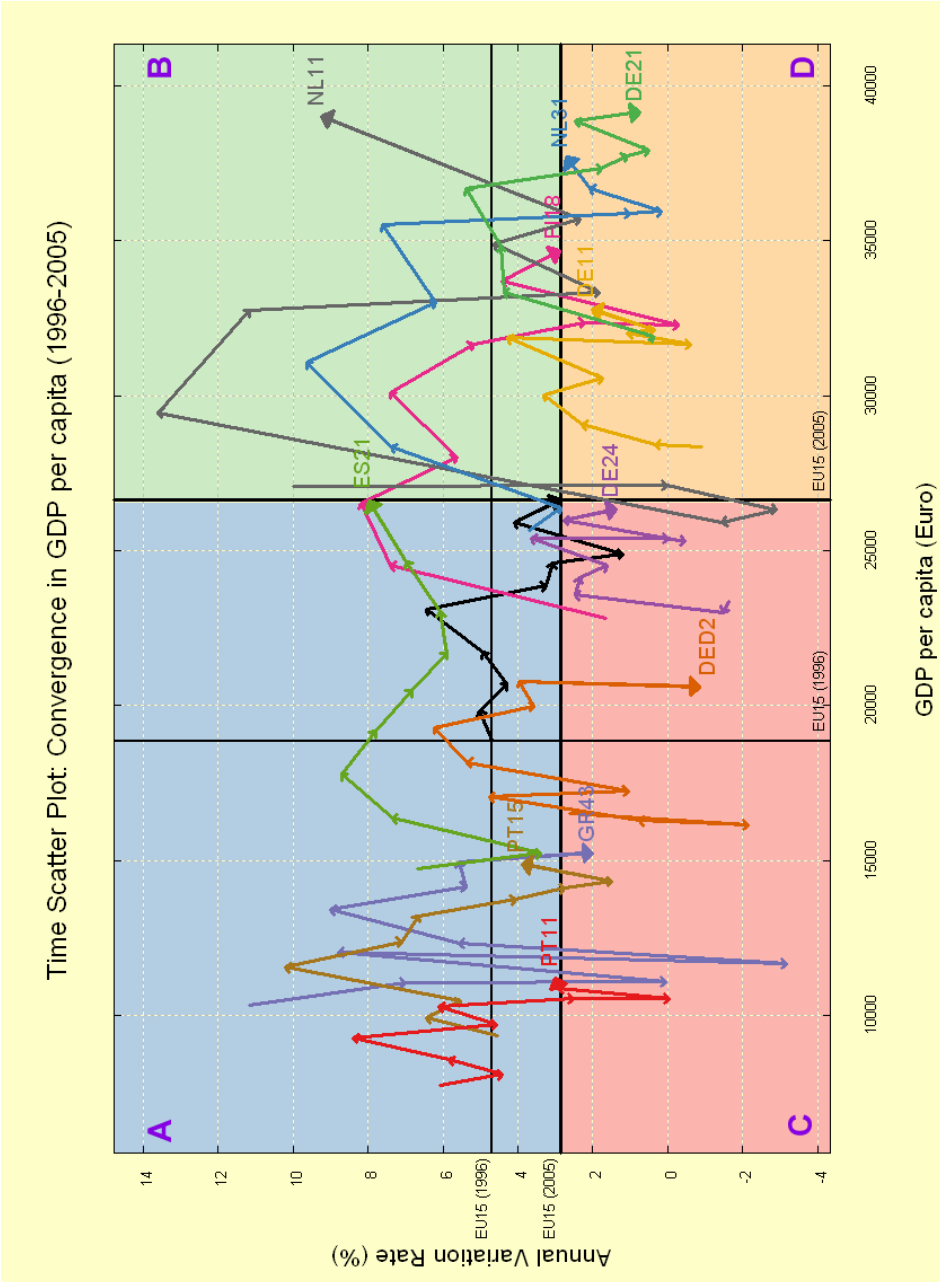

Fig. 2. Time scatter plot: convergence in GDP per capita (1996-2005)

- **Quadrant A (blue colour): Regions with disadvantage but converging**
  This area is relevant to identify regions in a disadvantaged position in terms of GDP per capita, which means these regions have low level of GDP per capita in comparison with the absolute convergence value (the European average). Nevertheless, the regions in this quadrant reached higher annual variation rates. Therefore the relative speed of growth in terms of GDP per capita is higher for all of these regions, generating a very positive process of convergence.

- **Quadrant B (green colour): Regions with advantage and diverging (high relative growth)**
  This quadrant corresponds to regions with a high GDP per capita, higher than the European average. Also these regions have an annual variation rate greater than the mean. Regions in this quadrant are in advanced position, since they have a relatively higher level of GDP and a better variation rate. However, from European point of view, this situation increases the divergence between European regions: rich regions are having a higher growth speed.

- **Quadrant C (red colour): Regions with disadvantage and diverging (low relative growth)**
  In the red area there are placed those regions with disadvantaged position from both point of views. They have low speed of growth and lower level of GDP. In other words, they are regions in disadvantaged economical level and they are diverging instead of converging due lower speed of growth. Consequently, these regions are worsening the European convergence.

- **Quadrant D (orange colour): Regions with advantage but converging**
  Finally, quadrant D covers another converging area. It includes those regions with a good relative level in terms of GDP per capita but with low growth. From convergence view this is a positive situation (Europe progress towards the global convergence), nevertheless these regions must remain alert because of their lower growth rate.

From the point of view of strict convergence in European frame, the optimum position for the regions is in the quadrants A and D. Placing all European regions in both quadrants Europe would reach the perfect convergence in terms of GDP per capita. On the contrary, regions in quadrants B and C contribute to increase the economic divergence in Europe, promoting income inequalities.

Generally, Figure 2 helps us to evaluate, in an analytical way, the condition and the convergence (or divergence) level of different regions in comparison with the European average (EU15) that is considered as reference level.

Analysing the regions by quadrants, Algarve (PT15) and Norte (PT11) regions in Portugal (quadrant A) were initially in unfavourable positions and were growing above the European average during last years. This has reduced the existing gap of these regions with the European reference. Though their position is not favourable yet, the current trend is encouraging.

Similarly, the Basque Country (ES21) starting in the same quadrant with an unfavourable position, has grown well above the European average during last years, reaching in the last analysed year (2005) the European convergence point, maintaining a high speed of growth.

During the observation period the Basque economy presented high rates of GDP growth, which can be explained through some competitive advantages gained from the past -such as appropriate public policies, human capital with high educative level and greater efforts in R&D. In order to maintain the differential growth achieved, the Basque Country may base its future on a clear increment of the productivity level, accelerating the transition towards the economy based on knowledge and technology.

The regions in quadrant B, Etelä-Suomi (FI18) and Groningen (NL11), are in a very favoured position. Their GDP per head was higher than the European average (advantage) and the growth rate was over European Union's average during the whole observation period (divergence trend). Consequently though, they did not contribute towards the aim of converging Europe.

On the other hand, regions like Oberfranken (DE24) and Dresden (DED2) from Germany are placed in the quadrant C. Also Kriti (GR43) in Greece ends up in this quadrant, although the growth rate was more erratic than the average (high variance). The situation for these regions is not so good because they did not reach an acceptable speed of growth and their levels of GDP per capita were not able to reduce the gap with the European average.

Finally, in the last quadrant, the regions of Stuttgart (DE11), Oberbayern (DE21) and, also in the last periods, Utrecht (NL31), began from a very favoured position, but presented a GDP growth below the European Union during the observation period and therefore were converging with the European average. Notwithstanding, the excellent starting positions of these regions have permitted them to keep rather favourable positions also at the end of the observation period.

5 Conclusions

This paper presented a new graphical application to represent time series data called time scatter plot and two empirical cases to illustrate its usability. The advantage of the time scatter plot is that it is able to show the possible correlation of two variables and in addition their evolvement in time. This is a very illustrative manner to present statistical data, especially with relatively small number of cases.

Flexibility offered by open-source programs like Gretl is essential when applying new modified manners to represent statistical data in graphical forms.

Gretl provides us with an excellent environment to develop and implement new graphical designs and statistical algorithms, due to its flexibility and the accessibility of the source code.

The two empirical cases that we presented as illustrative examples of usability of time scatter plot showed its advantages in presenting time series data of two variables for multiple cases. Although combining the relationship of two variables and dimension of time in same plot may sound complex, the illustration of time scatter plot is very clear and easy to interpret. This is a clear advantage of time scatter plot in the face of end-users of statistical data who may not be experts in statistical models.

In future other type of graphical models applied in different knowledge areas could be adopted to measure and analyze this kind of economic aspects. For example, a violin plot (a combination of a box plot and a kernel density plot) could be applied to show simultaneously the rank and the dispersion of a synthetic index.

Bibliography

[1] Cottrell, A., Lucchetti, R.: Gretl User's Guide - Gnu Regression, Econometrics and Time-series (September, 2008)

[2] Williams, T., Kelley, C.: Gnuplot - An Interactive Plotting Program (March, 2007)

[3] European Council: Presidency Conclusions - Lisbon European Council 23 and 24 (March 2000), http://tinyurl.com/bks2bq

[4] Commission of the European Communities: Facing the challenge - The Lisbon strategy for growth and employment (2004), http://europa.eu.int/comm/lisbon_strategy/index_en.html

[5] Schumpeter, J.: The Theory of Economic Development. Harvard University Press, Cambridge, MA (1934)

[6] Grossman, G. M., Helpman, E.: Innovation and Growth in the Global Economy - Cambridge, MA: MIT Press (1991)

[7] European Council: Presidency Conclusions: Barcelona European Council 15 and 16 (March 2002), http://tinyurl.com/bks2bq

[8] Commission of the European Communities: Implementing the Community Lisbon Programme: A policy framework to strengthen EU manufacturing - towards a more integrated approach for industrial policy. Brussels 5.10.2005. Communication from the Commission, COM (2005)474final

[9] Solow, R.: "A Contribution to the Theory of Economic Growth". Quarterly Journal of Economics 70 (1): 65-94. (1956)

[10] Romer, P. M.: "Increasing Returns and Long Run Growth" Journal of Political Economy, 94, 1002-38 (1986)

[11] Romer, P. M.: "Endogenous Technological Change", Journal of Political Economy 98, 71-102 (1990)

[12] Aghion, P., Howitt, P.: "A Model of Growth Through Creative Destruction", Econometrica 60: 322-351 (1992)

[13] EUR-Lex: Treaty on European Union, Official Journal C 191, (29 July 1992)

[14] European Union: Working for the Regions - EU Regional Policy 2007-2013. (2008) http://tinyurl.com/bqn6q6

[15] Commission of the European Communities: Investing in our future - The European Union's Financial Framework 2007-2013. Second revised edition (March 2008)