Advances in Complex Systems, Vol. 10, No. 1 (2007) 29–51

THE NETWORK OF INTER-REGIONAL DIRECT INVESTMENT STOCKS ACROSS EUROPE

STEFANO BATTISTON* and JOAO F. RODRIGUES†

Chair for Systems Design, ETH Zurich,
Kreuzplatz 5, 8032 Zurich, Switzerland
**sbattiston@ethz.ch*
†jrodrigues@ethz.ch

HAMZA ZEYTINOGLU

Clarifix Ltd., Suite 302,
95 Wilton Road, London SW1V 1BZ, UK
hamza@clarifix.net

Received 22 August 2005
Revised 13 June 2006

We present an analysis of inter-regional investment stocks within Europe from a complex networks perspective. We consider two different levels: first, we compute the inward–outward investment stocks at the level of firms, based on ownership shares and number of employees; then we estimate the inward–outward investment stock at the level of regions in Europe, by aggregating the ownership network of firms, based on their headquarter location. To our knowledge, there is no similar approach in the literature so far, and we believe that it may lead to important applications for policy making. In the present paper, we focus on the statistical distributions and the scaling laws, while in further studies we will analyze the structure of the network and its relation to geographical space. We find that while outward investment and activity of firms are power law distributed with a similar exponent, for regions these quantities are better described by a log-normal distribution. At both levels we also find scaling laws relating investment to activity and connectivity. In particular, we find that investment stock scales as a power law of the connectivity, as previously found for stock market data.

Keywords: Geographical networks; complex networks; scaling laws; investment.

1. Introduction

In this work, we study the network of investment stocks in Europe at two different levels of graining, a finer and a coarser level. We start by studying the network of investment stocks between individual firms and we then proceed to study the network of investment stocks between European regions. At each level we focus on two specific aspects of such networks: on one hand the statistical distributions of activity, investment stock and connectivity degree, and on the other hand the scaling relations between such quantities. In this respect, this work is related to a number of previous works in the fields of complex networks, economic networks, and industrial economics.

Several authors within the complex networks scientific community have recently studied the statistical properties of some economic networks. Among these, the most relevant to the present work are the so-called World Trade Web (WTW) (i.e. the network of import–export trade among countries in the world) [12, 25], and firm ownership networks. The firm ownership network of the largest firms in Germany has been found to exhibit small world properties [19]. For the firms quoted on US stock markets, power law distributions of connectivity degree and scaling relations between degree and invested volume have been found [13]. However, it has become clear that the connectivity degree alone does not provide a satisfactory description: it has in fact been observed that network structures may differ from each other in terms of control concentration and still look similar from the point of view of the degree distribution [6].

All of these works have proven the value of studying economic networks from a complex network perspective. However, while economic networks are naturally embedded in a geographical space, the geographical aspect has seldom been addressed [30]. Interestingly, several authors within the complex networks scientific community have started to focus recently on the geographical aspects of networks in other domains such as the worldwide air traffic network [4,17]. The results found in such systems can only motivate the study of economic networks in combination with their geographical aspects.

In the economic literature, some works consider the geographical embedding of economic networks. It has been argued that domestic rivalry and geographic industry concentration are especially important in creating dynamic clusters [20]. From a more general perspective, some authors assume the existence of a global world economy in which, since its inception in the sixteenth century, the periphery is assigned the function of supplying the core with cheap labor and raw materials, while the core has the role of producing manufactured goods requiring capital intensive technology [29].

Another topic related to the present work is the statistical distribution of wealth among individuals and firm sizes, which has been the object of a long and ongoing debate. At the end of the nineteenth century Pareto, after examination of the upper tails of the income distributions in a number of countries and epochs, found a remarkably close fit to a power law probability distribution [24]. Since the exponent was in all cases within a narrow range, this result was interpreted as a kind of "law" holding for all human societies [8]. Pareto's results also initiated a debate about the relationship between economic growth and income inequality.

Later empirical studies showed that the power law distribution did not describe accurately the left part of the income distribution. In a very influential work of 1931, Gibrat proposed to fit empirical data with the two-parameter log-normal distribution, providing also a dynamical model to explain the emergence of such a distribution [15]. His so-called "law of proportional effect," or "Gibrat's law," was applied to individual income but also to firm size. In this context, Gibrat's model assumed that growth is a random process with growth rates independent of firm size. The log-normal distribution was later found to be particularly appropriate for income distributions within a same category of the workforce and for firm size distribution within industrial sectors [2].

However, the aggregate of several distinct log-normal distributions may not itself be log-normal and indeed, more recently, the investigation of very large cross-sectional data sets have shown that firm size is better described by the power law distribution rather than by the log-normal [3]. Moreover, such a result was found to hold for data from multiple years and for various definitions of firm size, at least in the US. With the purpose of improving the fit of empirical data, several other distributions have been proposed during the last decades, in particular the generalized gamma and beta functions (three and four-parameter distributions which include power law and log-normal distributions as special or limiting cases) [22].

The debate about the appropriate distribution is still open, and so is the debate about the relation between inequality and economic growth [1, 8]. Moreover, it should be remarked that for power law distributions with exponent between 2 and 3, the second moment is infinite, and therefore one cannot speak of a "mean size plus or minus a standard deviation." There is not such thing as a "typical size" in a power law distribution. This may raise some difficulties for economic theories based on the representative firm.

Interestingly, the log-normal and the power law distribution can be the result of quite similar dynamical processes. In fact, the log-normal distribution can be explained with a simple multiplicative stochastic process, as suggested by Gibrat, while the power law can be explained with a simple additional ingredient to that very same process. This can be a lower reflecting barrier (representing, for example, a bankruptcy threshold below which a firm disappears and a new one is created), or a reset event [26]. More generally, it is known from the works of Kesten in the 1970s that under some general conditions, a combination of random multiplicative and additive processes can give rise to power law distributions [18]. Therefore, a number of models in the economic literature reproduce firm size distributions by assuming variants of these simple multiplicative stochastic processes.

The weak point of this modeling approach is, however, that interaction between firms is not considered at all. We know that firms cannot exist in isolation and that they interact with other firms through supply–customer relations, but also ownership, partnership and other relations. Such relations should then play a role in the statistical properties observed at a macroscopic scale. Or, if they do not, this is a surprising result that needs explanation. Therefore, there is a fundamental lacunae about firm–firm interactions, both on the side of the empirical analysis and the side of the modeling.

Besides firm size, the distribution of firm growth rates [11] and firm indebtedness [9] have been studied in several countries and contexts. Some scaling laws have been found to relate growth and its standard deviation [27]. However, very few works have investigated the statistical properties of quantities related to the pairwise interaction among firms [6, 13]. It must be said that a reason for this may be the availability of firm–firm interaction data, which are mainly limited to ownership relations. On the other hand, in the models including interaction, firms usually do not interact directly with other specific firms but rather with a "mean" firm, via some global coupling. This is the case for instance for the model of Ref. 10 in which a firm bankruptcy affects indirectly other firms through the interest rate of the central bank.

In this paper, we start filling the aforementioned gap from the empirical side, by focusing on the distribution and the scaling properties of investments of firms in other firms. Our aim is to contribute to the understanding of how interdependency between firms gives rise to well-defined statistical distributions, both at the level of firms and at the level of regions.

1.1. *Foreign direct investments and inter-regional direct investments*

Concerning investment there is a tendency to study investment stocks or flows between countries, referred to as Foreign Direct Investment (FDI), without considering a higher resolution. Contrary to this tendency, in this paper we focus on investment stocks between regions of Europe, which can be referred to as Inter-Regional Direct Investment (IRDI). In this sense, as discussed later, some concepts will be borrowed from the FDI literature and applied in the IRDI context. The statistical characterization of the network of IRDI stocks in Europe is a first step towards relating investment flow patterns at a global level to local and regional dynamics.

FDI is defined as "investment that adds to, deducts from or acquires a lasting interest in an enterprise operating in an economy other than that of the investor," the purpose of which is to have an "effective voice in the management of the enterprise," equivalent to holding 10% or more in the foreign enterprise [31]. Foreign affiliates are made up of subsidiaries, associates, and branches. Subsidiaries are majority- or wholly-owned by the parent companies. Associates are companies in which the investing firm participates in the management but does not exercise control. Branches are permanent establishments set up by the parent company in which there is no equity share capital apart from that of the parent. For associates and subsidiaries, FDI flows consist of the net sales of shares and loans (including non-cash acquisitions made against equipment, manufacturing rights, etc.) to the parent company plus the parent firm's share of the affiliate's reinvested earnings plus total net intra-company loans (short- and long-term) provided by the parent company. For branches, FDI flows consist of the increase in reinvested earnings plus the net increase in funds received from the foreign direct investor. FDI flows with a negative sign (reverse flows) indicate that at least one of the components in the above definition is negative and not offset by positive amounts of the remaining components.

However, the magnitude of FDI flow can also be measured as number of jobs created or increased. Contribution to more favorable employment status in the host country is critical when evaluating FDI, especially for countries that are battling high unemployment rates or want to increase the quality of their workforce. When quantifying FDI, "flow" (function of time) and "stock" (cross-sectional/cumulative) are measured and headquarter location of the investor plays a critical role [20].

Due to standard legal, institutional and policy attributes that are common in a given country, it indeed makes sense to focus on FDI. However, nowadays firms may perceive some regions in another country as more similar than regions within national borders. This might indicate that the process of European integration has reduced the national specificities perceived by multinationals and that regions now are competing to attract FDIs more across than within countries. Therefore, it is important to gain insight into investment flows at a higher spatial resolution. In this paper, we define Inter-Regional Direct Investment (IRDI) flow, in analogy to FDI, as the flow between two administratively separated regions irrespective of their countries. In contrast to the FDI definition given above, we consider as IRDI all investments, not just those larger than 10% of equity.

1.2. *Relevance of FDI/IRDI for the global economy*

Just like FDI, the study of IRDI has prominent policy making implications, in particular for governing bodies trying to tackle economic growth and employment creation [32]. There are certain general factors that consistently determine which countries/regions attract the most investment [7]. In particular, investors cite the following: market size and growth prospects of the host, wage-adjusted productivity of labor, the availability of infrastructures, reasonable levels of taxation and the overall stability of the tax regime. Recent crises have magnified perceptions of regulatory risks and greater attention is now being focused on the legal framework and the rule of law. Thus, the decision process in investment is multi-factorial, whereas the success or a higher productivity of the investment holds only when the host country/region has a minimum threshold stock of human capital. Thus, investment contributes to economic growth only when a sufficient absorptive capability of the advanced technologies is available in the host economy [21]. Promotional efforts to attract investment have become the focal point of competition among developed and developing countries. This competition is maintained even when countries are pursuing economic integration at another level. And it also extends to the subnational level, with different regional authorities pursuing their own strategies and assembling their own basket of incentives to attract new investments. While some see countries lowering standards to attract FDI as a "race to the bottom," others praise FDI for raising standards and welfare in recipient countries. The targets for these promotional efforts are predominantly the major players of FDI, namely the Transnational Corporations (TNCs), who for their part push for newer markets (in the last few decades the global trend of privatization has been a very important means towards this end). Public interest driven policies meanwhile continue to serve as the balancing force, and sustainable development concerns are dealt with accordingly. The EU has attracted over 40% of total world flows of FDI in the 1990s, becoming the largest recipient of multinational activity: multinationals account for a growing share of gross fixed capital formation in Europe (from 6% in 1990 to over 50% in 2000). However, this increasing inflow of FDI in Europe has not been equally distributed across countries and regions [5].

While statistics are still mostly examined at the country level, investments are actually made in specific regions with geographic features, local administrative constraints, and cultural profiles. There seem to be a substantial gap in our understanding of the role of region-to-region investment flows in the global economy. For this reason, we hereby propose to define the IRDI stock network and we investigate some of its statistical properties as a first stage of a more comprehensive study to be continued in the future. In Sec. 2, we describe the data set analyzed and discuss some methodological issues. In Secs. 3 and 4, we present and discuss the results of our analysis. In Sec. 5, we draw the conclusions and list some possible extensions of the present work.

2. Data Sets and Methods

In this section, we first describe the content of the firm database we used for our analysis. We then introduce the quantities we have measured on the data set and include some methodological remarks.

For the firm information, we used data collected in December 2004 from the Amadeus database of Bureau Van Dijk (BvD).

Access to the database was kindly granted by Prof. Delli Gatti of Università Cattolica di Milano. The database provides information on about 8 million firms in 38 European countries. In particular, it provides headquarter address and geographical region, financial profile, number of employees, industrial classification, names of shareholders and board of directors. The regions of the headquarters of the firms correspond in most cases to regions in the level 3 of the EU NUTS classification. As usual in most firm databases, only the shareholders with the most significant shares in the firms are listed.

Because data are available only in files of limited size we were forced to restrict ourselves to a subset of all the firms in the database. We have chosen to select the firms with number of employees larger or equal to 100. The resulting data set consists of 181,945 firms uniquely identified by their Bureau Van Dijk identification number (BVDID). For a given firm, shareholders can be individuals/families, governments or other institutions that are not listed as firms. Moreover, even shareholders which are firms may have less than 100 employees, and therefore location and profile for them is not available in our data set. We have chosen to restrict our analysis to the network of firms for which we have a profile in the data set. As such, we do not consider in our analysis shareholders that are individuals/families, governments or other institutions not listed as firms as well as firms with fewer than 100 employees. The final set of data we use for the network analysis includes 29,314 firms and 22,174 links located over 1,288 different regions.

The selection of a subset of the ownership links induces of course an underestimation of the total hosted investment stock. Still, investigating how investment size among this set of firms is distributed in the network and among geographical regions is a very interesting point to address.

2.1. *Defining the quantities of interest*

For each firm i, we consider the following quantities: the activity a_i measured as number of employees in the firm, the shares w_{ij} of firm i owned by any other firm j, and the headquarter region R_i of firm i. The number of employees is one of the standard quantities used to measure firm size [3], and in the following we will measure also investment stocks in terms of number of employees. There are of course other possible measures of investment stocks, based on capital rather than on human resources, but from the point of view of labor market and economic impact at a local scale, it is relevant to have an estimate of how many employees of a firm or a region depend on the investment coming from outside.

Shares are usually defined as a percentage, but it is more convenient to define w_{ij} as a fraction of ownership and, therefore, as a real number in $[0, 1]$. Not all shares are necessarily held by entities external to the firm. Moreover, some entities are not firms and we only look at shares held by firms, as discussed above. Therefore, it holds that

$$\sum_{j \neq i} w_{ij} \leq 1. \tag{2.1}$$

If we take the number of employees as a measure of activity of a firm, it is natural to compute the quantity s_{ij}:

$$s_{ij} = w_{ij} a_i, \tag{2.2}$$

representing the investment stock of firm i held by firm j.

We can define the *firm network* as a graph $G_F = (V_F, E_F)$, Fig. 1 (left), where V_F is a set of nodes representing firms and E_F is a set of directed edges between nodes. An edge (i, j) represents the fact that firm j owns shares of firm i. The order of the pair in this notation is natural for the layout of most databases of firms and we adopt it. However, it should be kept in mind that ownership and investment have opposite directions. In fact, it is more natural to define *in-degree* and *out-degree* of connectivity with respect to investments rather than to ownership. We define as *in-degree* k_{in} of a firm the number of firms investing in i (holding shares of i). Similarly, we define as *out-degree* k_{out} of a firm the number of outside firms in which firm i invests. The *connectivity degree* or simply degree k of a node is the number of edges entering or departing from that node.

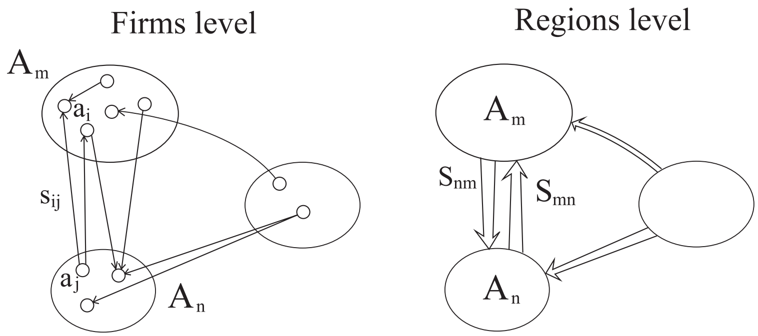

Fig. 1. Diagram illustrating how the (right) region network GR is built up from the (left) firm network G_F.

We can associate to each edge (i, j) the normalized weight w_{ij}, representing the number of shares that firm j owns in i. But we can also associate the absolute weight s_{ij}. We then define the *inward investment stock* a_i^{in} of firm i as the total stock invested in firm i by other firms and *outward investment stock* a_i^{out} as the total stock invested by firm i in other firms:

$$a_i^{in} = \sum_j s_{ij}, \quad a_i^{out} = \sum_k s_{ki}. \tag{2.3}$$

We can now define analogous quantities aggregated by region. The *activity of region m* is defined as the sum of the activity of the firms with headquarters in that region:

$$A_m = \sum_{i \in m} a_i. \tag{2.4}$$

In other words, A_m is the total number of workers employed by firms of that region. Since we analyze only a subset of all firms, the activity of a region can be much smaller than the number of individuals employed in that region. The sum of the investments made by firms of region n in firms in region m is defined as

$$S_{mn} = \sum_{i \in m, j \in n} s_{ij}. \tag{2.5}$$

It can be seen both as the outward investment stock of region n in region m or as the hosted investment stock in region m coming from region n.

It is very natural at this point to define the *region network* as a graph $G_R = (V_R, E_R)$, where nodes represent regions and a directed edge (m, n) from region m to region n represents the fact that some firms of region n own shares in some firms of region m. The diagram in Fig. 1 illustrates the procedure of building the network of regions. Small circles represent firms with their associated values of activity. Edges represent ownership relations. Larger circles represent regions in which firms have their headquarters. The edges in firm network among all firms in regions m and n sum up to form an edge between m and n in the region network.

We associate to the edge (m, n) the absolute weight S_{mn}. The degree is defined as for the firm region and in particular in-degree and out-degree are defined with respect to investments.

The sum of the investments made in firms of a region by firms of any other region will be called the *inward investment stock of region* m [Eq. (2.6)]. In the following, we will refer to this quantity also as *hosted investment stock*. Similarly, the sum of the investments made by the firms of a region in firms of other regions will be called the *outward investment stock of region* m [Eq. (2.6)]:

$$A_m^{in} = \sum_n S_{mn}, \quad A_m^{out} = \sum_n S_{nm}. \tag{2.6}$$

Both definitions are chosen in analogy with the terms used in the literature about Foreign Direct Investments. But instead of looking at investments between different countries, we increase the spatial resolution to the level of regions.

As already mentioned, only the shareholders with the most significant shares in the firms are listed in firm databases; therefore, the in-degree of firms is a biased quantity with little meaning for our purposes. However, the out-degree is not affected by any bias. Moreover, when aggregating by region, the in-degree of a region represents the number of other regions in which the top shareholders of firms in the focal region have their headquarters. This quantity is not limited *a priori* and as such it makes sense to study its distribution.

2.2. *Measuring the quantities of interest*

Our perspective in this work is to try and relate the microscopic and macroscopic aspects of the network of investments between firms and between regions. We want to look at statistical distributions and not at single or average values, the aim being to understand the detailed factors that underlie the macroscopic properties. Therefore, we will focus on the probability distributions of the quantities defined above.

In order to study the probability density function (PDF) of a variable x, it is useful to plot its *complementary cumulative distribution function* (CDF), defined in standard statistics textbooks as

$$P_C(\hat{x}) = \int_{x \geq \hat{x}} p(x) dx, \tag{2.7}$$

where $x \to p(x)$ is the probability density function. In words, the CDF gives the fraction of a randomly chosen sample of the variable x that lies above the value $\hat{x}$. A simple way of constructing $P(x)$ is the following. Consider the vector x of N real numbers. We rank x in ascending order. Clearly, now all values are larger or equal to the first data point. So the probability distribution starts from 1 and decreases. The kth component of the vector x has ascending rank k and there are $(N - k)$ values larger or equal to $x(k)$. The fraction of data larger or equal to $x(k)$ is $(N - k)/N$. We therefore simply plot the pair $(x(k), (N - k)/N)$ for all k. If some values of x are repeated and in particular if x is a discrete variable, then the plot will display "stairs." In this case, it is preferable to count the fraction P of data that is larger or equal to each value x and then to plot P versus x.

If the distribution of the variable x is a power law with exponent $-\gamma$, then its complementary cumulative distribution is still a power law with exponent $-\gamma+1$. On a log–log scale, they appear as straight lines with different slopes:

$$p(x) = c_1 x^{-\gamma}, \tag{2.8}$$

$$P(x) = c_2 x^{-\gamma+1}, \tag{2.9}$$

where c_1, c_2 are normalization factors. If instead the distribution of the variable x is log-normal with coefficients μ and σ [Eq. (2.10)], then it appears as a quadratic curve on a log–log scale:

$$p(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \tag{2.10}$$

$$\log(p(x)) \sim -\log(x) - \frac{(\log x - \mu)^2}{2\sigma^2}. \tag{2.11}$$

However, its complementary cumulative distribution does not have an analytical expression.

It is usually more accurate and safe to estimate the exponent from the CDF rather than from the PDF, because of the fluctuations in the frequency of high values of x. This is usually done by fitting the CDF on a log–log scale with a line and computing the slope. However, it has been recently remarked that this method introduces a systematic bias [16]. An alternative method that does not make use of a graphic fit is described in Ref. 23. The formula for the exponent is $\gamma = 1 + N[\sum_i \log \frac{x_i}{x_{\min}}]^{-1}$, where $x_{\min}$ is the lower limit of the range of data following the power law. Confidence intervals for γ can be computed with the standard bootstrap technique. When a relation holds between two variables, then their respective probability distributions are also related through the equation: $p(y) = p(f(x)) = p(x)\frac{df}{dx}$. An important consequence is that if, x is power law distributed and y scales as a power law of x, then y is also power law distributed, and a relation holds for all the exponents involved (power law distributions are closed with respect to the operation of power law rescaling). If x is log-normal distributed, and y scales as a power law of x, then the distribution of y converges to a log-normal function for large y. As a consequence, when two variables which are expected to be related, are both power law distributed, the relation between them could be a power law scaling.

To study of the correlation between quantities related to firms and regions that span several order of magnitude and have to be studied on a log–log scale, we proceed as follows. Consider the variables (X, Y). We compute the $\log_{10}$ of both variables and produce a scatter plot of $(\tilde{X}, \tilde{Y}) = (\log_{10} X, \log_{10} Y)$. We then divide the x-axis into k bins of equal size. For each bin centered on the value x_i, we compute the mean y_i and the standard deviation σ_{y_i} of the values of Y for which the corresponding abscissa falls in the bin k. We obtain a new set of data points

(x, y) that allows us to better display the trend of the original data (X, Y). A linear fit is then performed on the (x, y) data points, and values of slope, intercept and correlation coefficient are computed. We recall that a linear relationship between x and y implies that, after taking the exponential, a power law relationship holds between the original variables:

$$\tilde{Y} = m \cdot \tilde{X} + q, \tag{2.12}$$

$$\log_{10}(Y) = m \cdot \log_{10}(X) + q, \tag{2.13}$$

$$Y = C \cdot X^m, \tag{2.14}$$

where $C = 10^q$. Of course, while the operation of binning and averaging over bins is a standard procedure one should be careful in this case since this operation does not commute with the operation of taking the exponential. However, we are not aware of any documented bias introduced by this procedure and, if the fit is reasonably good, we conclude that Y scales as a power law of X and take m as the exponent of the scaling law.

3. Analysis at the Level of Firms

We first report the complementary cumulative frequency distribution of activity and investment stocks of firms. We also investigate how investment stock scales with firm activity and with firm connectivity degree. We then report the analogous results for regions where activity, investment stock and connectivity degree of regions are defined as in Sec. 2.1. Similarly, we investigate how investment stock scales with region activity and region connectivity degree.

3.1. *Firms: Distributions of activity, investment and connectivity degree*

In Fig. 2, we report the complementary cumulative distribution (CDF) of activity and investment stock of firms computed from our data set. The onset at the value 100 for the activity is simply due to the restriction of the data set to firms with more than 100 employees. Investment stock can of course take smaller values as it is measured as a fraction of the number of employees per firm. The complementary cumulative distributions display a linear decay over three decades or more. However, some "bumps" deviating from linearity are visible. The curves can either be fitted with log-normal distributions with very large standard deviation or reasonably fitted with a power law. In both cases, the meaning is the same: the probability of finding large firms is decreasing approximately as the exponent of the power law fitting the curve. Computing the exponent from the linear fit is known to introduce bias [16]; hence, following a known procedure [23] the values of the exponents and confidence intervals were computed as described in Sec. 2. The values of the exponents and their confidence intervals are reported in Table 1. As a check, we also computed the exponents with the more usual method of fitting the CDF in log–log

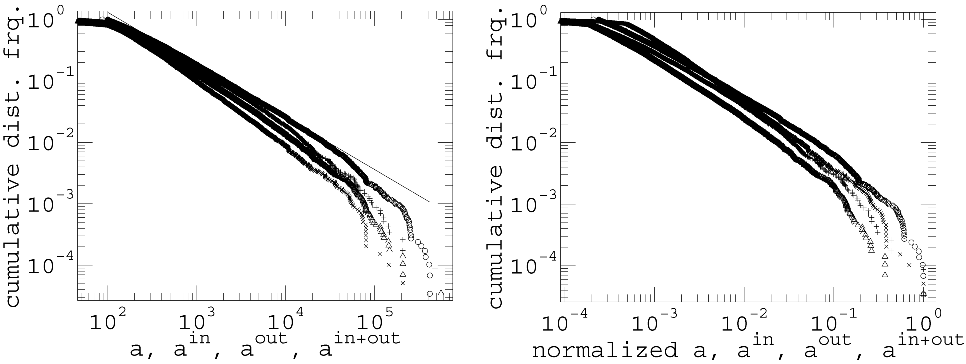

Fig. 2. Firms. Cumulative distribution of: (left) activity a (o); hosted investment stock a^{in} ($\times$); outward investment stock a^{out} (+); total investment stock a^{in+out} (Δ) and (right) normalized values with respect to the maximum value.

Table 1. The values of the exponents and their confidence intervals.

Data	γ	σ_γ
a	1.7829	0.0038
a^{in}	1.9307	0.0064
a^{out}	1.7684	0.0061
a^{in+out}	1.8480	0.0047
k^{out}	3.849	0.044

scale with a line and as expected [16], we found values to be systematically higher by around 5–10%.

Our finding concerning the activity is in line with what is generally known in the literature for firm size distributions of many countries and historical epochs. It implies that firm activity is very heterogeneous and that, roughly speaking, very large values and very small values of activity are much more frequent than in normal distributions. We remind the reader that the data set analyzed includes only firms involved in an ownership relationship in Europe and not all firms indiscriminately. However, the value of the exponent γ is not far from the results obtained in previous studies. For instance, Axtell reports 2.056 for the US firm activity distribution [3]. Fujiwara *et al.* report 1.995 for the UK based on data from the Amadeus Database [11].

On the other hand, the fact that the investment stock distribution is also a power law is to our knowledge a novel result. In particular, the values of the exponent γ for activity and outward investment stock are very close.

It may seem reasonable to make the hypothesis that outward investment is proportional to size. In this case investment would trivially follow a power law distribution with the same exponent as for size, in line with our observations. However,

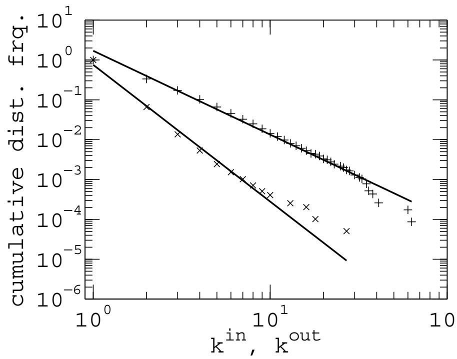

Fig. 3. Firms. Cumulative distribution of connectivity degree: in-degree (×), out-degree (+).

to our knowledge there is no empirical evidence of such simple proportionality, at least in the context of firm–firm interaction. A general way of studying the relation between investment and size from a statistical point of view on large data sets is to investigate the existence of a scaling relation between the two variables (as discussed in Se. 2.2). This will be done in the next section.

As is usual in the study of networks, we report the complementary cumulative distribution functions (CDF) of the degree of connectivity of firms (Fig. 3). We distinguish between (total) degree, in-degree and out-degree as defined in Sec. 2.1. The distributions span a short range of less than two decades. The out-degree displays a linear decay on a log–log scale that can be fitted by a power law with exponent 3.85, which is a little larger than the typical values observed in many empirical complex networks, which are typically in the range 1.8–3. The result implies that the number of connections between firms is moderately heterogeneous and decreases slower than exponentially. The curve for the in-degree instead is not meaningful for the reasons mentioned in Sec. 2.2 and simply shows how many records of shareholders are available per firm. The linear fit was performed just for the sake of completeness.

However, it is important to remark that the connectivity out-degree represents the number of ownership relations in which the firm is involved, regardless of the size of shares involved in each relation. Because the degree does not take into account the size of the shares, a large out-degree does not mean that a firms really controls a lot of other firms. Alternative quantities are needed to characterize the ownership concentration such as those introduced in Ref. 6, and they will be applied to the present data set in a future study.

3.2. *Firms: Correlations between activity, investment and connectivity degree*

In order to understand why the investments of firms are power law distributed and why distributions of outward investment and activity have exponents very close to each other, we investigate the correlations between activity, investment and connectivity out-degree. The plots in Figs. 4 and 5 are produced by taking the $\log_{10}$ of the quantities and binning the data on the x-axis as described in Sec. 2.2.

Table 2 reports the value of slope and correlation coefficients for the linear fit of the binned data. The correlation coefficients are all quite close to 1 so, with the caveat mentioned in Sec. 2.2, we conclude that the data indicate the existence of scaling laws between investment and activity and between out-degree and activity.

We notice that the exponent 0.925 for the scaling of hosted investment versus activity is smaller than but still close to 1. This means that firms tend to host investment in amount almost just proportional to their activity. This is not surprising if we consider that all firms in the data set analyzed, are owned to some extent by

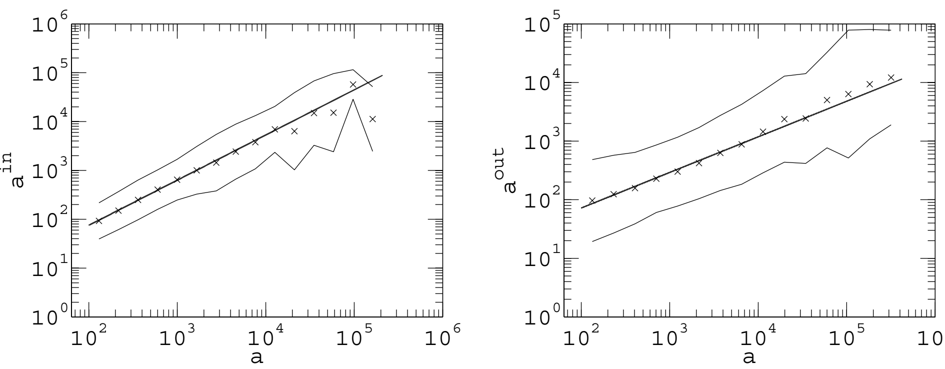

Fig. 4. Firms. Plot of (left) hosted invested stock a^{in} and (right) outward invested stock a^{out} of firms versus their activity a in log–log scale. Data are binned. Mean ± standard deviation of the values in each bin are plotted as the continuous lines.

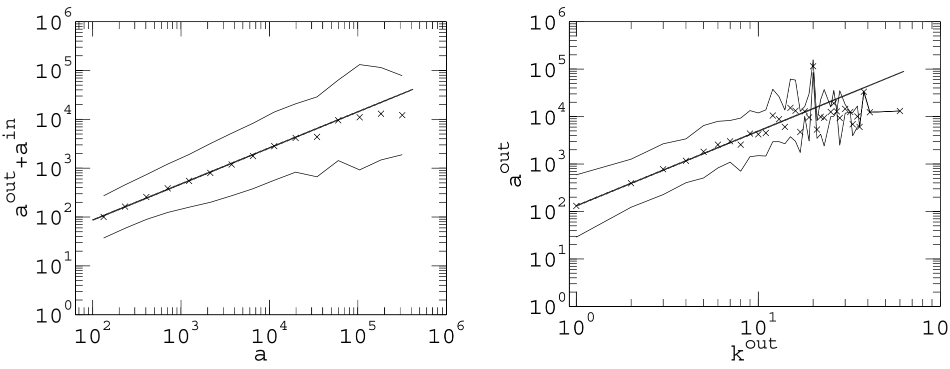

Fig. 5. Firms. Left: plot of total invested stock a^{in+out} of firms versus their activity a. Right: plot of connectivity out-degree k^{out} of firms versus their hosted investment stock a^{out} on a log–log scale. Plot data are binned. The mean ± standard deviation of the values in each bin are plotted as continuous lines.

Table 2. Values of slope and correlation coefficient relative to Fig. 4 and Fig. 5 (left).

y-axis	m	q	Corr. coef.
a^{in}	0.925	0.026	0.974
a^{out}	0.607	0.644	0.997
a^{in+out}	0.739	0.460	0.992

Table 3. Values of slope and correlation coefficient relative to Fig. 5 (right).

y-axis	m	q	Corr. coef.
a^{out}	1.574	2.120	0.855

some other firm in the data set, and in many cases the share is large and close to 1. A deeper understanding would require an investigation of the statistics of the ownership concentration and will be carried out in a future work. On the other hand, the exponent 0.61 for the scaling of outward investment versus activity implies that although more active firms tend to invest more, the investment increases less than linearly as a function of the activity (sub-linear increase). This means that very large and active firms invest proportionately less than smaller ones. Such a finding becomes quite important if an institution in charge of attracting investments from firms is trying to estimate the expected investment of firms based on their activity.

Finally, the exponent 1.57 for the scaling of activity versus out-degree is quite interesting. It implies that the larger the firm the larger the number of investments, but the increase is sub-linear. Interestingly, the value of the exponent is not far from the values found for the scaling law between invested volume and degree in some stock markets: 1.1 for Nasdaq, 1.43 for NYSE, 1.59 for MIB [13]. In that case, the invested volume is exactly the analogous quantity to the outward investment. Moreover, the scaling law resembles the one observed in the context of air traffic networks for a quantity analogous to the outward investment that has been recently introduced as node strength [4]. In that case, the exponent is found to be 1.7. These findings might support the idea that a universal scaling law for node strength holds in complex networks where the weight plays a crucial role. In a future work, we will investigate possible network formation models leading to the emergence of such scaling law in economic networks.

4. Analysis at the Level of Regions

4.1. *Regions: Distributions of activity, investment and connectivity degree*

In Fig. 6 (left), we report the complementary cumulative distribution of activity A and inward/outward/total investment stocks A^{in}, A^{out}, A^{in+out} for regions. These quantities were computed from the data set as described in Sec. 2.1. As done for

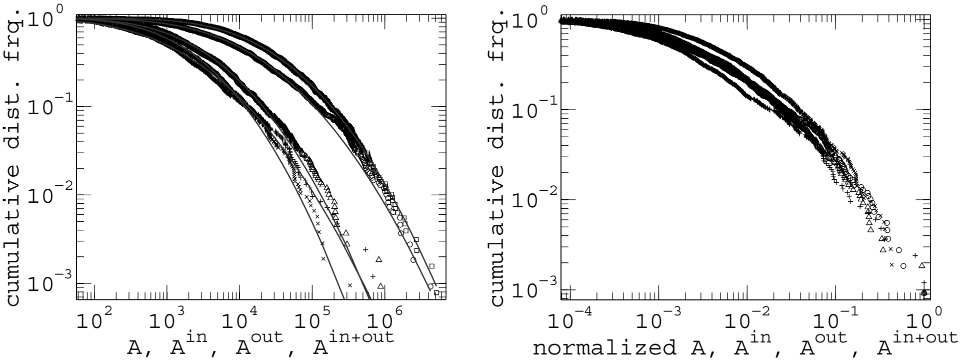

Fig. 6. Regions. Cumulative distribution of: (left) activity A (for all firm profiles) ($\square$); activity A (o); hosted investment stock A^{in} ($\times$); outward investment stock A^{out} ($+$); total investment stock A^{in+out} (Δ) and (right) the same quantities normalized with respect to the maximum value.

the case of firms, in order to check to what extent they overlap we also normalized such quantities and we report their complementary cumulative distribution in Fig. 6 (right). As can be seen, the overlap is only partial. This may suggest that a nonlinear scaling holds between the variables, and this hypothesis will be investigated in the next section.

It must be remarked that the way regions are defined within a country depends on the country's administrative system. To give an example, the regions provided in the data set are at the level of "provincia" for Italy and "department" for France, which have comparable surface area on average.

In contrast to the case of firms, the distributions do not display a linear decay on a log–log scale, but rather a quadratic one, which we tried to fit with a log-normal distribution.

The fit with the log-normal shows some discrepancy in the right tail. In particular, the distributions of A^{out} and A show bumps between 2×10^4 and 1×10^5, after which the observations are systematically above the log-normal estimate. This evidence could be an artefact in the data and will be investigated in the future by comparing a more recent data set.

A standard test for goodness-of-fit to a normal distribution is the Jarque–Bera (J–B) test. This test checks the null hypothesis that the data come from a normal distribution and was applied in our case to the logarithm of the data. We choose the usual value of 5% significance level to test the hypothesis. This means that the test fails when, assuming a process with a normal distribution, the probability that this process would generate a distribution identical to the tested data is less than 5%.

The J–B test is known to be very sensitive to artefacts and not surprisingly, the distributions of A^{in} and A^{in+out} pass the J–B test while the distributions of A^{out} and A do not.

Table 4. Values of coefficients of the log-normal fit relative to Fig. 6.

Data	μ	$\Delta\mu$	σ	$\Delta\sigma$
A	9.06	0.12	1.97	0.08
$A_{all\ data\ set}$	9.79	0.10	1.82	0.07
A^{in}	7.28	0.10	1.64	0.07
A^{out}	6.89	0.14	2.02	0.10
A^{in+out}	7.69	0.10	1.76	0.07
K	1.98	0.07	1.18	0.05
K^{in}	1.50	0.06	1.00	0.04
K^{out}	1.52	0.08	1.16	0.06

Table 5. Values of coefficients of the power-law fit relative to Fig. 6.

Data	γ	σ_γ
A	2.29	0.14
$A_{all\ data\ set}$	2.23	0.09
A^{in}	2.58	0.19
A^{out}	2.40	0.20
A^{in+out}	2.21	0.11
K	2.53	0.09
K^{in}	3.00	0.15
K^{out}	2.40	0.10

On the other hand, only the very last portion of the distribution could be fitted with a power law. Values for the coefficients of the log-normal fit are given in Table 4. For comparison we also report the exponents of the power law fit of the rightmost tail (Table 5). It can be seen that they are systematically higher than those of the firms, implying that the distributions decay faster than in the case of firms.

In conclusion, in the case of regions, the distributions of all variables cannot be fitted with a power law over the whole range, while the fit to a log-normal seems by far more reasonable. The fit could possibly be improved with generalized four-parameter functions that have been proposed in the literature to describe income and wealth distributions [22].

At a first sight the finding that the above distributions are close to log-normal is puzzling, as one may expect that regional activity and investment stocks also scale with a power law. However, the following remark is relevant at this point. Cities range from small villages of a few inhabitants to metropolises of 10–20 millions. Firms also range from one-person enterprises to multinationals with a few hundred thousand employees. On the other hand, while the position of the boundaries of a region surely is the result of a historical process, the area and the population within a region is probably limited by administrative constraints, in the sense that if the administrative load becomes too heavy, the region is split into two. For instance, within no country are there regions that are orders of magnitude larger in area than other regions. On the other hand, it is known that the economic activity

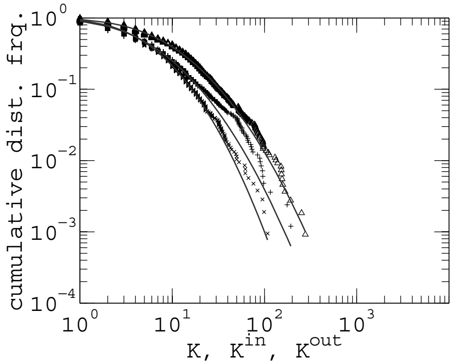

Fig. 7. Network of Regions. Cumulative distribution of connectivity degree: in-degree ($\times$), out-degree ($+$), total degree (Δ).

of countries measured by Gross Domestic Product is power law distributed with exponent 1 [12]. So it appears that regions are clearly less heterogeneous than firms and countries. However, before trying to draw some implications for economic development policies, it would be interesting to normalize the activity of the regions by area and/or active population. Unfortunately, these data were not available to us.

In Fig. 7, we report the CDF of the connectivity degree for regions. The curves are clearly not power laws, so we tried to fit them with log-normal functions. Values of coefficients are reported in Table 4. The curve for the in-degree is systematically above the one for the out-degree. Given the fact that the PDF is the derivative of the CDF, the plot implies that in the range $[1, 50]$ the in-degree is typically smaller than the out-degree, while above 50 the opposite holds. We do not have an explanation for this finding. On the other hand, we recall that as in the case of firms, the number of connections is not necessarily meaningful. In fact, this is exactly why we have introduced the inward and outward investment stock.

4.2. *Regions: Activity, investment and connectivity degree correlations*

In order to understand why investment between regions is log-normally distributed and why distributions of inward/outward investment and activity display a partial overlap after normalization (see Sec. 2.2), we investigate the correlations between activity, investment and connectivity in/out-degree of regions. With the same procedure used for the case of firms (see Sec. 3.2), we produced the plots in Figs. 8 and 9 (see Sec. 2.2).

Tables 6 and 7 report the values of the slope and correlation coefficients for the linear fit of the binned data. Again, the correlation coefficients are all quite close to 1 so, with the caveat mentioned in Sec. 2.2, we conclude that the data indicate the existence of scaling laws in the network of regions between investment and activity and between connectivity degree and activity.

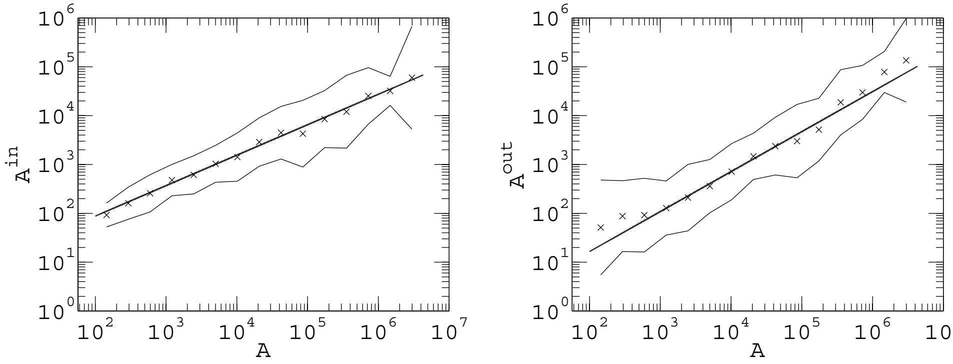

Fig. 8. Regions. Plot of (left) hosted invested stock A^{in} and (right) outward invested stock A^{out} of regions versus their activity A on a log–log scale. Data are binned. Mean ± standard deviation of the values in each bin are plotted as the continuous lines.

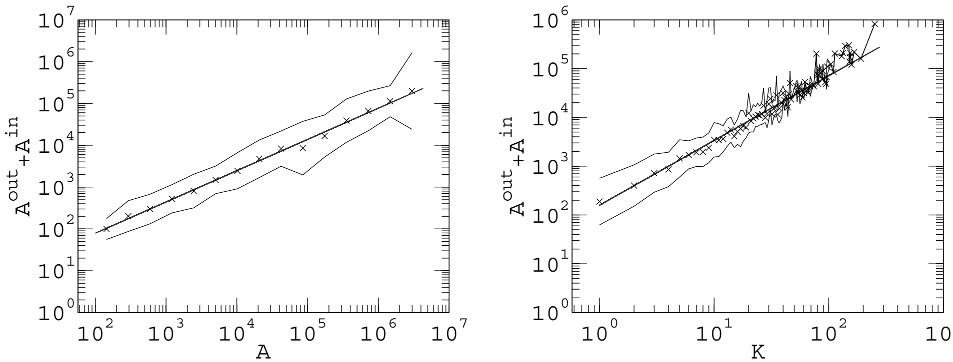

Fig. 9. Regions. Plot of (left) total invested stock A^{in+out} of regions versus their activity A and (right) total connectivity degree k^{tot} of regions versus their hosted investment stock A^{in+out} on a log–log scale. Data are binned. Mean ± standard deviation of the values in each bin are plotted as the continuous lines.

Table 6. Values of slope and correlation coefficient relative to Fig. 8 and Fig. 9 (left).

y-axis	m	q	Corr. coef.
A^{in}	0.623	0.697	0.997
A^{out}	0.819	−0.420	0.990
A^{in+out}	0.748	0.401	0.998

We notice that the exponents 0.62 and 0.82 for the scaling of hosted investment versus activity and outward investment versus activity, respectively, are smaller than 1.

As seen at the firm level, it follows that although more active regions tend to make and host more investment, the investment increases less than linearly as a

Table 7. Values of slope and correlation coefficient relative to Fig. 9 (right) and Fig. 10.

y-axis	m	q	Corr. coef.
A^{in}	1.467	2.022	0.965
A^{out}	1.370	2.268	0.965
A^{in+out}	1.326	2.201	0.979

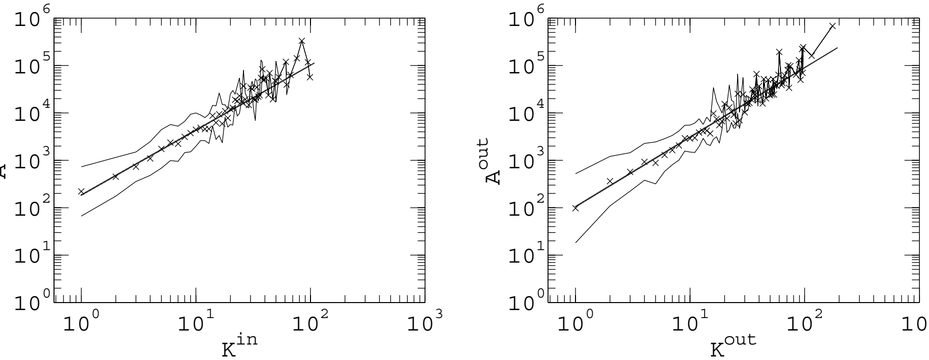

Fig. 10. Regions. Plot of connectivity (left) in-degree k^{in} and (right) out-degree of regions versus their hosted investment stock A^{in} and outward investment stock A^{out} respectively on a log–log scale. Data are binned. Mean $\pm$ standard deviation of the values in each bin are plotted as the continuous lines.

function of the activity (sub-linear increase). This means that very active regions invest proportionately less than smaller regions. However, the increase of outward investment with activity is stronger for regions ($m = 0.82$) than for firms ($m = 0.61$; see Table 2). This is an interesting result that we will address in the future.

The amount of investment made or received by regions in relation to their activity is relevant to institutions in charge of fostering regional development. Although these findings cannot provide detailed predictions, they could help develop multi-agent based models trying to reproduce the observed features with the aim of designing possible incentive strategies.

Finally, the exponents for the scaling of activity versus in/out-degree are again not far from the values found for the scaling laws in other works for such variables as invested volume versus degree in stock markets and node strength versus degree in air traffic networks.

5. Conclusions

In this paper, we propose a simple but novel procedure to analyze the network of inter-regional investment based on the number of employees of firms in each region and their network of ownership. In this network representation, the connectivity in-degree of a region is the number of other regions from which firms invest in the focal region, while the out-degree is the number of regions in which firms of the focal region invest. The sum of the weights over the incoming links represents the hosted investment stock of a region in terms of employees. The sum over the outgoing links represents the outward investment stock of a region in terms of employees.

We study the statistical properties of investment stock networks at the level of firms and at the level of regions. Our first result is that investment stock of firms is power law distributed and that, in particular, the exponent of outward investment is very close to the one of firm activity. As is well known, this fact may result from a power law scaling relation between activity and outward investment. This is neither obvious nor documented in the literature, so it has to be checked empirically. At first sight, activity and investment are quite scattered and span a few orders of magnitude. However, by taking the logarithm of the values of activity and investments, and binning the data, we indeed find that investment scales as a power of the activity. Moreover, power law scaling relations also hold between investment stock and connectivity degree.

On the other hand, in the case of regions, we find log-normal distributions for activity, investments and degree. It can be argued that this result might simply be related to the distribution of population size across regions (unfortunately, we do not have data to test this hypothesis at the moment). Even so, it is a remarkable fact that such probability distributions for regions clearly differ both from those of firms as well as from those of countries in the world. The impact of this fact on the design of global policies to foster investment and economic development should be investigated.

Again, the fact that similar distributions emerge for activity, investment and connectivity degree of regions suggests that some relation should hold among them. In particular, we find that investment of firms scales as a power of the degree with exponent 1.57 (out-degree), while for regions it scales with exponents 1.37 and 1.47 (in-degree and out-degree respectively). Interestingly, previous studies on different data sets have investigated the scaling law of two quantities analogous to the investment stock (invested volume and node strength in Refs. 4 and 13). There, the exponent was found to be range between 1.1 and 1.7. The existence of scaling laws relating investment, activity and connectivity both in firms and regions is an interesting and novel result relevant to the fields of complex networks, industrial economics and geography.

On the other hand, such scaling laws should be of interest for policy making. For instance, we find that very active regions invest proportionately less than smaller ones. The same holds for firms, although the coefficients governing the relation between investment and activity are different. This kind of result allows us to make some statistical predictions about the investments that regions will receive or make based on their activity and connectivity. The present work is a first step towards understanding the relation between the local dynamics of investment flows and the macroeconomic facts emerging at a global level. One can ask, for example, whether such a distribution of investments is desirable with respect to some societal goals that might be at stake at the country or at the EU level. If it is not, one can investigate if introducing some incentive policies can improve the distribution with respect to these goals. In this sense, the findings reported here should stimulate the investigation of models for managing the development of regions and optimally allocating resources. Overall, we believe that these results open the way for further studies with potential long-run implications for policy making at the level of EU investment promotion, support for underdeveloped EU regions and optimization of investment flows.

Acknowledgments

This manuscript greatly benefited from Frank Schweitzer's comments. We also thank Domenico Delli Gatti of Università Cattolica di Milano for granting us access to the Amadeus database. S. Battiston acknowledges the support of CNRS during the phase of data collection and the initial stage of this work. S. Battiston and J. Rodrigues acknowledge the support of the European project MMCOMNET (IST Contract No. 12999) and the EXYSTENCE Thematic Institute CHIEF (held at University Politecnica delle Marche, Ancona, IT, May 2–21, 2005), where this collaboration started.

References

[1] Aghion, P. and Williamson, J. G., *Growth Inequality and Globalization: Theory, History, and Policy* (Cambridge University Press, 1999).

[2] Aitchinson, J. and Brown, J. A. C., *The Log-Normal Distribution, with Special Reference to Its Use in Economics* (Cambridge University Press, New York, 1969).

[3] Axtell, R. L., Zipf distribution of U.S. firm sizes, *Science* **293** (2001) 1818.

[4] Barrat, A., Barthélemy, M. and Vespignani, A., The effects of spatial constraints on the evolution of weighted complex networks, physics/0504029v1 (2005).

[5] Basile, R., Castellani, D. and Zanfei, A., Location choices of multinational firms in Europe: The role of national boundaries and EU policy, *Quad. Econ. Mat. Stat.* **76**, June (2003) [http://www.fscpo.unict.it/PDF/Castellani].

[6] Battiston, S., Inner structure of capital control networks, *Physica A* **338**(1 & 2) (2004) 107–112.

[7] Borensztein, E., De Gregorio, J. and Lee, J. W., How does foreign direct investment affect economic growth?, *J. Int. Econ.* **45** (1998) 115–135.

[8] Cowell, F., *Measuring Inequality: Techniques for the Social Sciences* (Oxford University Press, 1999).

[9] Delli Gatti, D., Di Guilmi, C. and Gallegati, M., On the Empirics of "Bad Debt", mimeo, Università Politecnica delle Marche (2003).

[10] Delli Gatti, D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M. and Palestrini, A., A new approach to business fluctuations: Heterogeneous interacting agents, scaling laws and financial fragility, to appear in *J. Econ. Behav. Organ.* cond-mat/0312096.

[11] Fujiwara, Y., Di Guilmi, C., Aoyama, H., Gallegati, M. and Souma, W., Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms, cond-mat/0310061v2 (2003).

[12] Garlaschelli, D. and Loffredo, M. I., Fitness-dependent topological properties of the World Trade Web, *Phys. Rev. Lett.* **93** (2004) 188701 [cond-mat/0403051].
[13] Garlaschelli, D., Battiston, S., Castri, M., Servedio, V. D. P. and Caldarelli, G., The scale free topology of market investments, *Physica A* **350**(2 & 4) (2005) 491–499.
[14] Gaffeo, E., Gallegati, M. and Palestrini, A., *Physica A* **324** (2003) 117.
[15] Gibrat, R., *Les Inégalités Économiques* (Recueil Sirey, Paris, 1931).
[16] Goldstein, M. L., Morris, S. A. and Yen, G. G., Problems with fitting to the power-law distribution, to appear in *Eur. Phys. J.* cond-mat/0402322v3 (2004).
[17] Guimerà, R., Mossa, S., Turtschi, A. and Amaral, L. A. N., The world-wide air transportation network: Anomalous centrality, community structure, and cities' global roles, *Proc. Nat. Acad. Sci. USA* **102** (2005) 7794–7799.
[18] Kesten, H., Random difference equations and renewal theory for products of random matrixes, *Acta Math.* **131** (1973) 207–248.
[19] Kogut, B. and Walker, G., The small world of Germany and the durability of national ownership networks, *Am. Soc. Rev.* **66**(3) (2001) 317–335.
[20] Jakobsen, S. E. and Onsager, K., Head office location — Agglomeration, clusters or flow nodes? *Urban Studies* **42**(9) (2005) 5.
[21] Loungani, P. and Razin, A., How beneficial is foreign direct investment for developing countries? *Financ. Dev.* **38**(2) (2001) 1.
[22] McDonald, J. B., Some generalized functions for the size distribution of income, *Econometrica* **52** (1984) 647–663.
[23] Newman, M. E. J., Power laws, Pareto distributions and Zipf's law, cond-mat/0412004v2 (2005).
[24] Pareto, V., *Cours d'Economie Politique* (F. Rouge, Lausanne, 1896).
[25] Serrano, M. A. and Boguñá, M., Topology of the World Trade Web, *Phys. Rev. E* **68** (2003) 015101 [cond-mat/0301015].
[26] Sornette, D., Multiplicative processes and power laws, *Phys. Rev. E* **57**(4) (1998) 4811–4813.
[27] Stanley, M. H. R., Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A. and Stanley, H. E., Scaling behaviour in the growth of companies, *Nature* **379** (1996) 804–806.
[28] Sutton, J., *J. Econ. Lit.* **35** (1997) 40.
[29] Wallerstein, I., *The Modern World System I: Capitalist Agriculture and the Origins of the European World — Economy in the Sixteenth Century* (Academic Press, New York, 1974).
[30] Weisbuch, G. and Battiston, S., From production networks to geographical economics, to appear in *J. Econ. Behav. Control* (2007).
[31] Office for National Statistics (ONS), UK, FDI Profile UNCTAD.
[32] Policies to Attract Foreign Direct Investment, World Bank/IFC (Web site: Papers and Links — Privatization, Infrastructure and Business Environment).