

The Structure of Economic Connections between Industries: Non-Scaling Behaviour

Camila C. S. Caiado¹ and Paul Ormerod²

- 1 Department of Mathematical Sciences, University of Durham, UK
c.c.d.s.caiado@durham.ac.uk
- 2 Volterra Partners LLP, London and Department of Anthropology, University of Durham pormerod@volterra.co.uk corresponding author

Abstract

There is considerable evidence of the existence of scaling behaviour (power law relationships) in a number of aspects of economic activity. Here, we examine the evidence on the connections between different industrial sectors, in terms of the value of output which each industry sells to each of the others, and the value of output which the others sell to it.

Information on these connections between industries is available in the input-output tables in the national economic accounts.

We analyse a database in which activity in the United Kingdom economy is disaggregated into 123 separate industries. We find that although the statistical distributions of the connections are highly non-Gaussian, there are marked departures from scaling behaviour, whether in the distribution of the connections from each individual industry to all the others, or in the distribution of connections from all industries into each individual one.

1. Introduction

Evidence of scale-free networks which connect agents in social and economic systems was established at least a decade ago [for example Pastor-Sattoras et.al. 2001, Albert et.al. 1999, Lahererre and Sornette 1998, Liljeros et.al. 2001]. Further, power law relationships were shown to give a good description of several important aspects of economic activity relating to firms. For example, the size distribution of firms [Axtell 2001], the variability of output according to the size of the firm [Stanley et.al. 1997], and the frequency and size of firm extinctions [Cook and Ormerod 2003, deGuilmi et.al. 2004].

Firms are the basic building block of economic activity, and produce most of the output of the economy, at least in advanced economies of the type prevalent in the West. There is some analysis of inter-firm relationships using a Japanese database, an inter-firm network with some 1 million nodes (firms) and 4 million directed links (customer-supplier relationships). Ohnishi et.al. (2010) examine motifs in the network, with the clique in which all component firms are reciprocally interconnected is the strongest motif. An undirected network between these firms, in other words not distinguishing whether firm A sells to B or vice versa, is analysed in Konno (2008) and a scaling relationship is claimed, although no formal analysis is presented. More realistically, Fujiwara and Aoyama (2008) extract the directed networks from the data, examining both the in-degree (the number of suppliers to each firm) and the out degree (the number of other firms each firm has as a customer). Both the relationships approximate a power law over a section of the data, but in each case the number of links is under-predicted quite substantially at the upper end of the distribution.

However, data limitations mean that in general at present we are only able to analyse the connections between industries, which represent aggregations of firm. So the aim of this paper is to examine some initial evidence on the distribution of connections between industries in the economy. Section 2 discusses both the data and the empirical results. Section 3 gives a brief conclusion.

2. The data and the results

Detailed information is available on the connections between *industries* in so-called input-output tables in the national economic accounts. These show the value of output which each industry sells to each of the others, and the value of output which the others sell to it.

The 2005 input-output tables have been derived from supply use tables compiled using 108 industry input-output groups (IOGs) consistent with the UK's Standard Industrial Classification 2003 (SIC (03)) for industries and Eurostat's Classification of Products by Activity (CPA (02)) for products, extended to 123 IOGs by separating components of the non-market output produced by general government and NPISHs (Non-Profit Institutions Serving Households) from the output produced by the market sectors, in order to allow for their different roles in the economy.

The size of individual industries identified in the input-output tables varies enormously. The total value of output in the UK economy is around £1500 billion¹. Large industries such as construction and retail distribution have output levels of £180 billion and £100 billion respectively. The smallest, 'jewellery and related products' has an output of only £0.6 billion. There are in fact 36 industries, almost all of which are in the manufacturing sector, whose output is less than £5 billion.

However, much less significance can be attached to these differences in size than may appear to be the case at first sight. In particular, information on manufacturing industries is provided in much greater detail than for other sectors of the economy. No fewer than 77 out of the total of 123 categories of industry are within manufacturing, even though this sector as a whole makes up no more than 12 per cent of total output.

Information is not provided at a similar level of disaggregation for other sectors of the economy. In part, this is due to historical accident. When the national economic accounts were first constructed in the 1930s and 1940s, manufacturing was a much

¹ as gross value added at basic prices.

bigger percentage of the economy than it is today. In part, it is because it is easier to estimate the value of things – manufactured articles – than it is of services. More direct information is available with which to compile estimates, and hence it is done in more detail. In principle, the contribution of, say, computer services - a non-trivial industry with an output of £54 billion – could also be broken down into more detailed, smaller component parts, exactly as is done with manufacturing.

We examine the data in two dimensions:

- how many industries are connected into a given industry
- how many industries a given industry is connected into

For purposes of description, we refer subsequently to the first category as ‘column connections’ and the second as ‘row connections’.

In the first instance, we investigate the degree of sparseness of the connections in each of these dimensions. In other words, to examine the total number of non-zero connections for each industry².

Figures 1 and 2 below show the histograms of the distribution of non-zero row and column connections

² There are certain obscure principles of national accounting which lead one of the individual sets of row connections and one of the column connections to consist entirely of zeros, and this is omitted from the analysis.

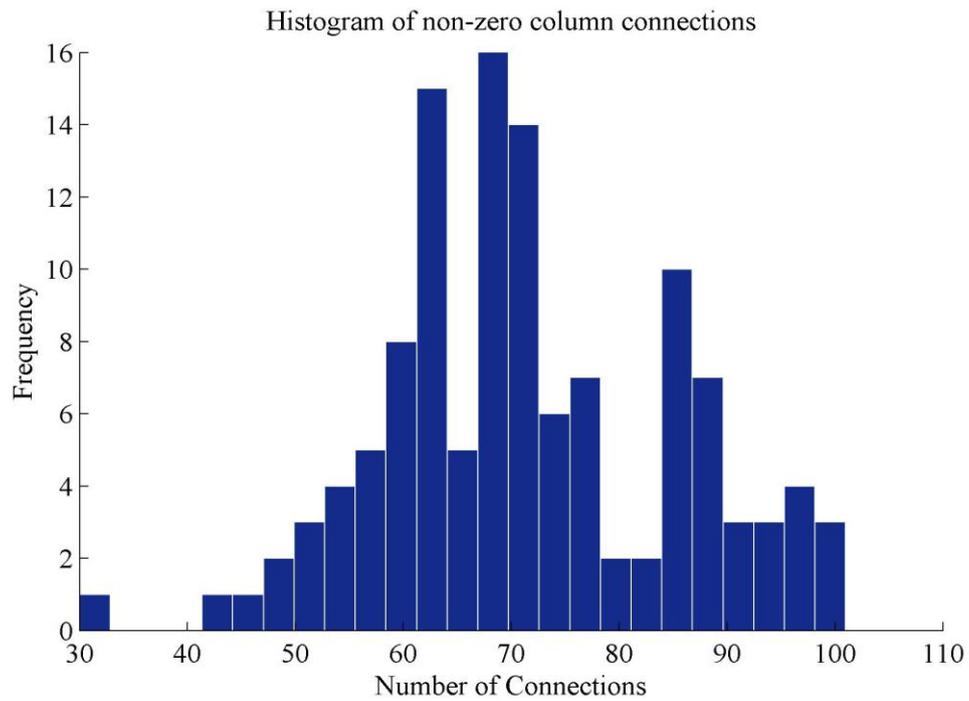


Figure 1 *UK national accounts input-output tables: 123 industries. Histogram of how many industries a given industry provides with products*

The row connections for an industry show the number of industries into which it sells. As Figure 1 shows, a substantial proportion of industries are connected in this way to most of the other industries. The mean number of connections is 82 and the median 91, with a standard deviation of 36.4.

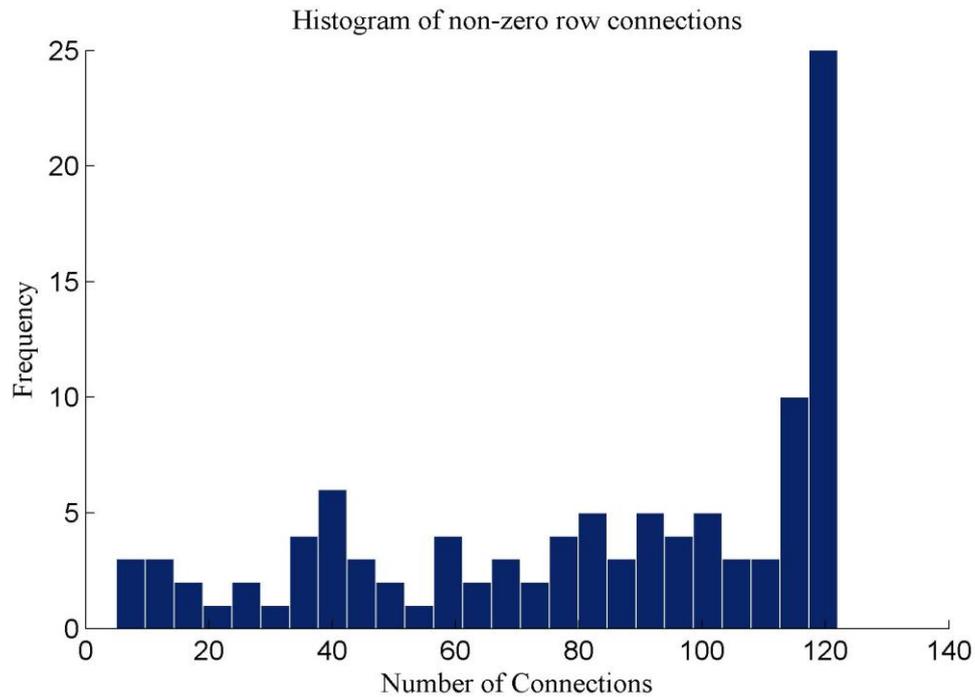


Figure 2 *UK national accounts input-output tables: 108 industries. Histogram showing the number of industries a given industry buys products from*

The column connections of an industry show the number of other industries from which it buys. The mean value is high, at 71, the median 70 and the standard deviation 13.6.

In both cases, the degree of sparseness is low. In other words, the technological structure of production requires that:

- most industries require as inputs, products produced by most other industries
- most industries sell their own products as inputs into most other industries

It is evident from inspection of Figures 1 and 2 that the overall distribution is different from that of the row connections. This is confirmed by a formal test. The null hypothesis that the two distributions are the same is rejected very decisively on a Anderson-Darling test (1952) test (at $p = 0.0000$).

Many years ago Goodwin (1949) conjectured that such an asymmetry could be a determinant of the business cycle, the persistent short-term fluctuations in national output (GDP) observed in the advanced market economies. In a dynamic context, the input-output matrix describes how output at time t is transformed into output at time $t+1$. In general, Goodwin argued that given that non-symmetric, non-negative square matrices have some eigenvalues with non-zero imaginary parts, the structure of production itself is a factor determining the business cycle. This has long since disappeared from mainstream economic theory, which following its inability to understand the recent economic crisis certainly needs to draw on wider sources for inspiration [see, for example, Farmer et.al. 2012].

The distributions of the two types of connections are clearly highly non-Gaussian. However, a power law does not provide a good account of the distributions of either the column or the row connections. We initially fitted the data using the technique described in Clauset et.al. (2009). The approach calculates the p-value for a given power-law fit, using the Kolmogorov-Smirnov test, using the discrete method described in the paper. The method of maximum likelihood is used to carry out a grid search for the power law exponent, α and its normalizing constant, x_{min} . The Clauset method correctly rejects the hypothesis of a power law distribution as we expected from the inspection of the histograms in Figures 1 and 2. The distributions of these two variables are likely to be mixtures of two or more distributions. However, we can still look for a power law behaviour by analysing the rank of the number of row and column connections. Zipf's law or the power law is widely observed in nature, physics and economics [see Zipf 1949] but somehow overused.

Here we apply a power law to both cases by using the non-linear robust fit algorithm provided in Matlab (Seber and Wild, 2003) of a power law. The robust fit procedure is not based on the squared distance between actual and fitted as in standard least squares. The method minimises a weighted sum of squares, where the weight given to each data point depends on how far the point is from the fitted line. Points near the line get full weight, points farther from the line get reduced weight. Points that are farther from the line than would be expected by random chance get zero weight. The results we describe are obtained using this technique.

Figure 3 plots the actual and power law fitted data for columns with the 95 per cent confidence lines around the fit. The adjusted R squared is 0.866 and the estimate of α 0.222 (95 per cent confidence interval 0.204-0.240). The standard error of the equation is 2962.

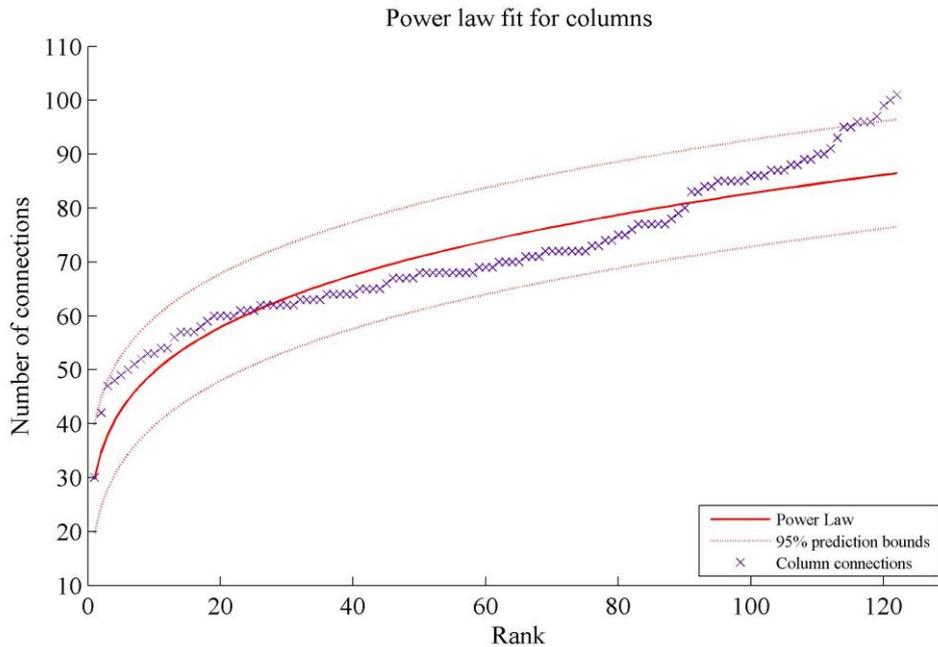


Figure 3 *Power law fit for ranked column counts. The power law fails to capture the upper range of the data and the concavity in the middle.*

The fit of the power law can be improved markedly by adding a constant term which accounts for the vertical shift in the data from zero, but even so the standard error of the power law equation (with linear constant term) is 314.6 We examined a range of polynomial fits to the column data, and the best fit is obtained by a cubic relationship, with the actual and fitted plotted in Figure 4. The adjusted R-squared is 0.993 and the standard error of the equation is 155.6, considerably less than the power law.

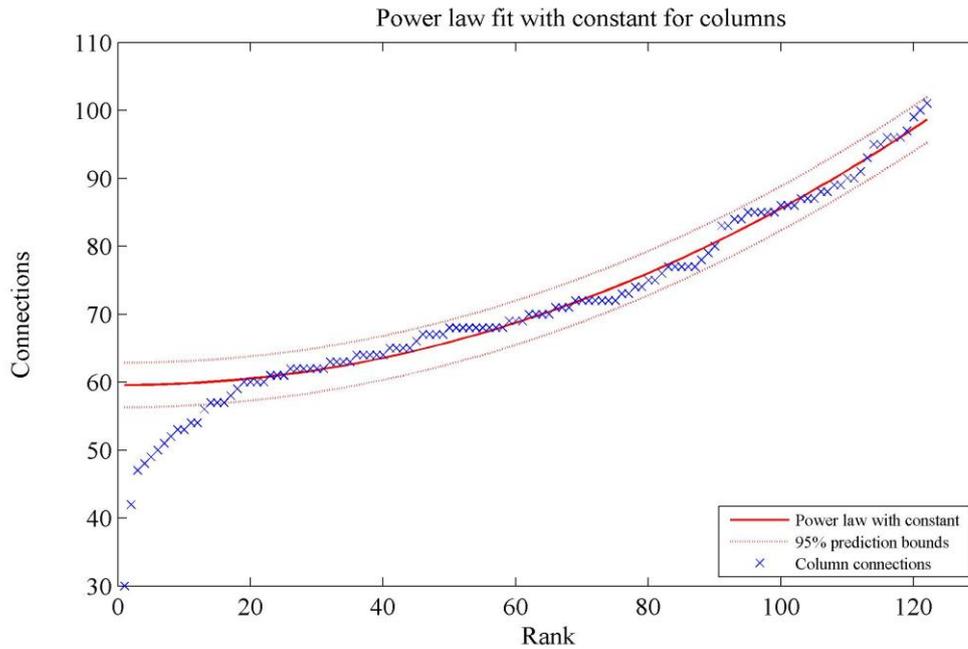


Figure 4 *Power law fit with constant for ranked column counts. The power law fails to capture the lower range of the data but there is an improvement on concavity fit.*

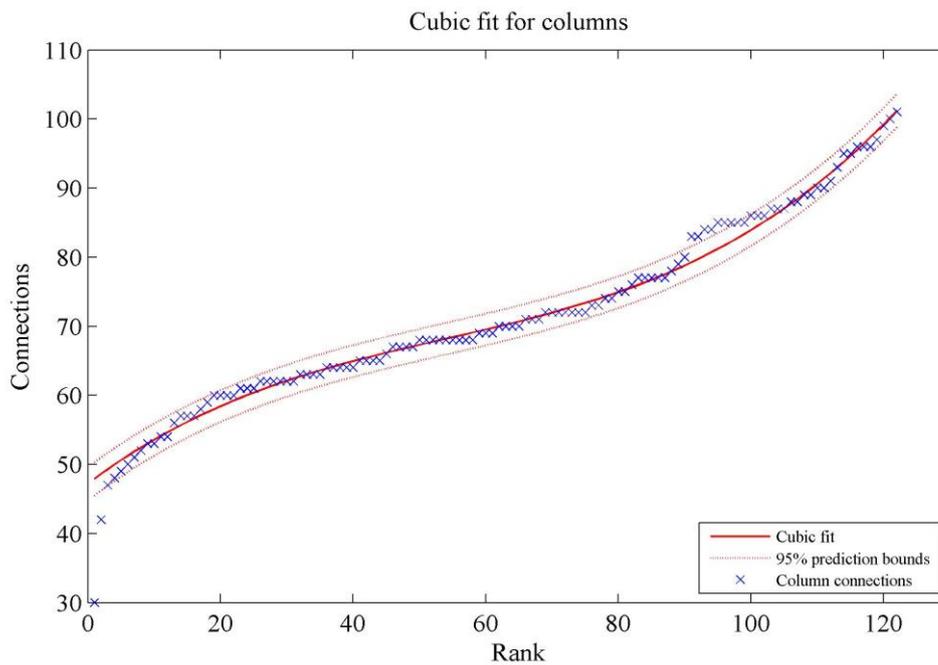


Figure 5 *Cubic fit for ranked column counts. The cubic fit better captures the overall trend of the data*

We now consider the row data. Again using the robust non-linear algorithm in Matlab to estimate a power law, we obtain a value of α of 0.652 (range 0.625 – 0681). The adjusted R squared is 0.976 and the standard error is 3338. The actual and fitted with the 95 per cent confidence intervals are plotted in Figure 6. In this case, it is the upper tail of the data which is not fitted by the power law.

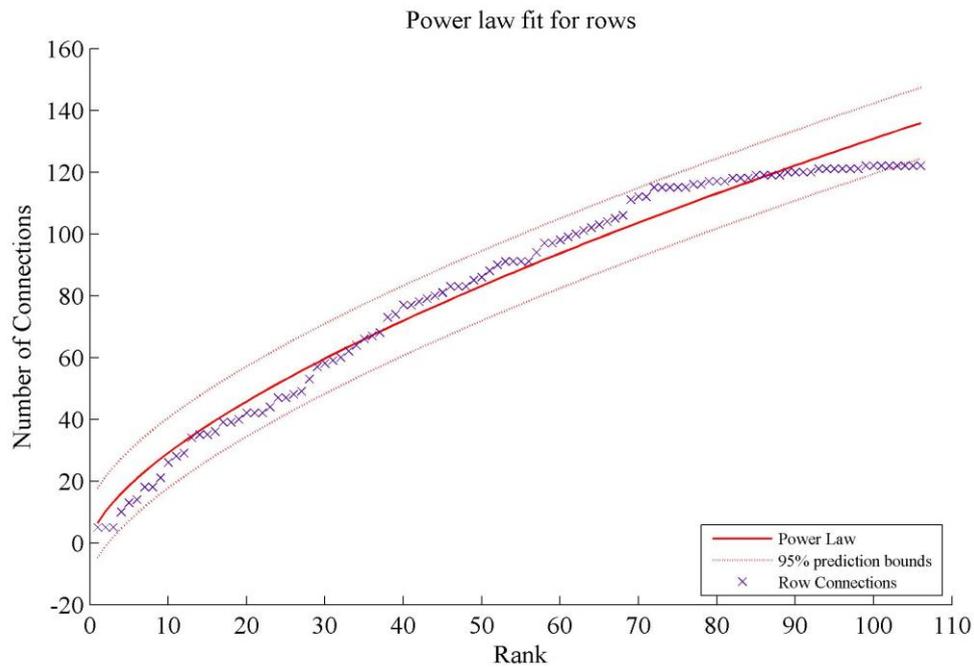


Figure 6 *Power law fit for ranked row connections. The power law fails to capture the upper range of the data.*

A power law captures the row data considerably better than the column data, but again it fails to capture the upper tail.

Adjusting the power law relationship with a linear constant again improves the fit as shown in Figure 7, but the standard error is still high at 2243. We examined polynomial fits, and found that the best is obtained by a quadratic. The adjusted R squared is 0.996 and the standard error of the equation is only 484, much lower than for the best power law fit. Most importantly, the upper range of the data is better explained as we can see in Figure 8.

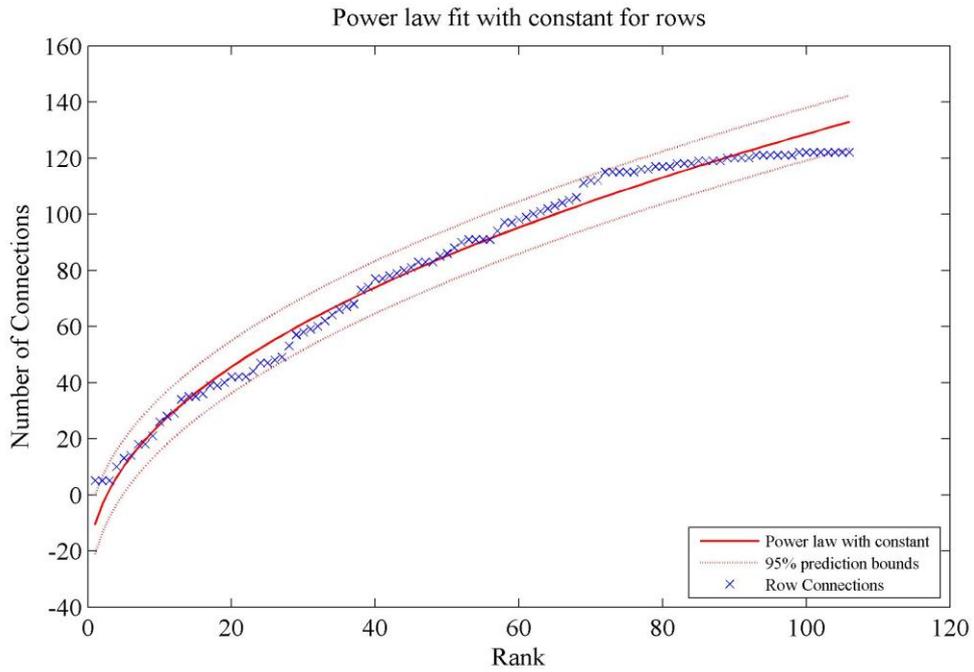


Figure 7 *Power law fit with constant for ranked row counts. The power law fails to capture the upper range of the data but and there is an improvement on concavity fit.*

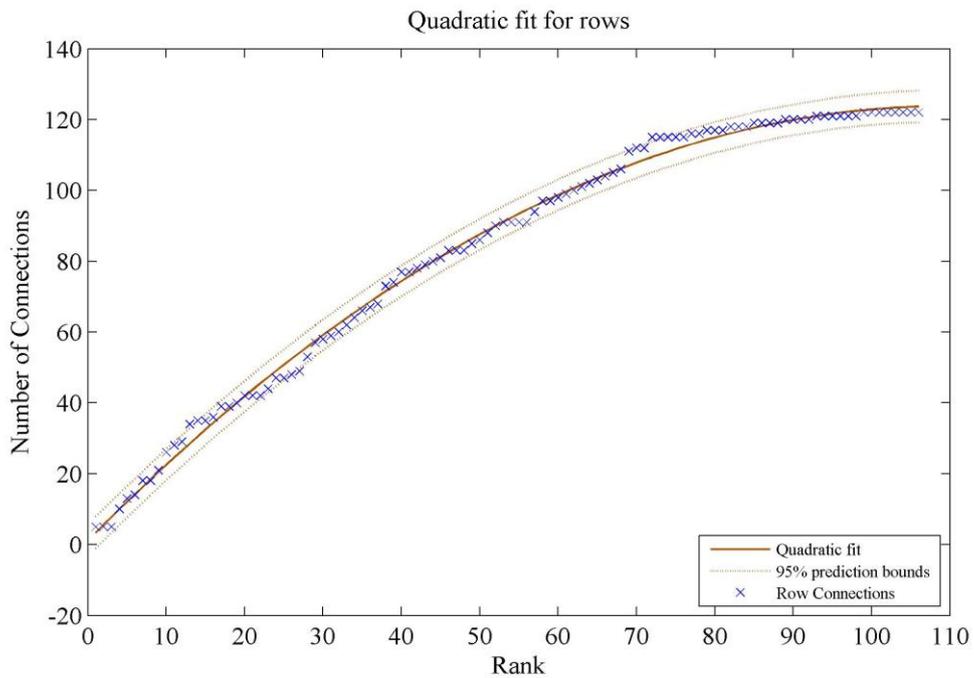


Figure 8 *Quadratic fit for ranked row counts. The quadratic fit better captures the overall trend of the data including the shape of the upper range resulting in tighter prediction bounds.*

The problem with the power law fits for both the column and row data is the fact that it does not capture the distribution in the upper tail. We therefore examined whether a power law could describe the upper part of the data, using as the sample those values which are greater than the mean. There are 54 such observations for the column data and 61 for the row. In this case, it is essential to fit the power law with a constant ($y = bx^{-\alpha} + c$), the constant represents the shift of the minimum value from zero to the mean.

The fit for the column data is considerably improved. The value of α is 1.304 (1.158-1.451), the adjusted R squared is 0.979 and the standard error of the equation 77.78. In this case, a polynomial does not improve the fit very much, a quadratic having adjusted R squared of 0.982 and standard error of 69.74. This finding is consistent with the conjecture of Perline (2005) that the discovery of power law relationships is often associated with truncated data sets which only try to fit the upper part of a distribution.

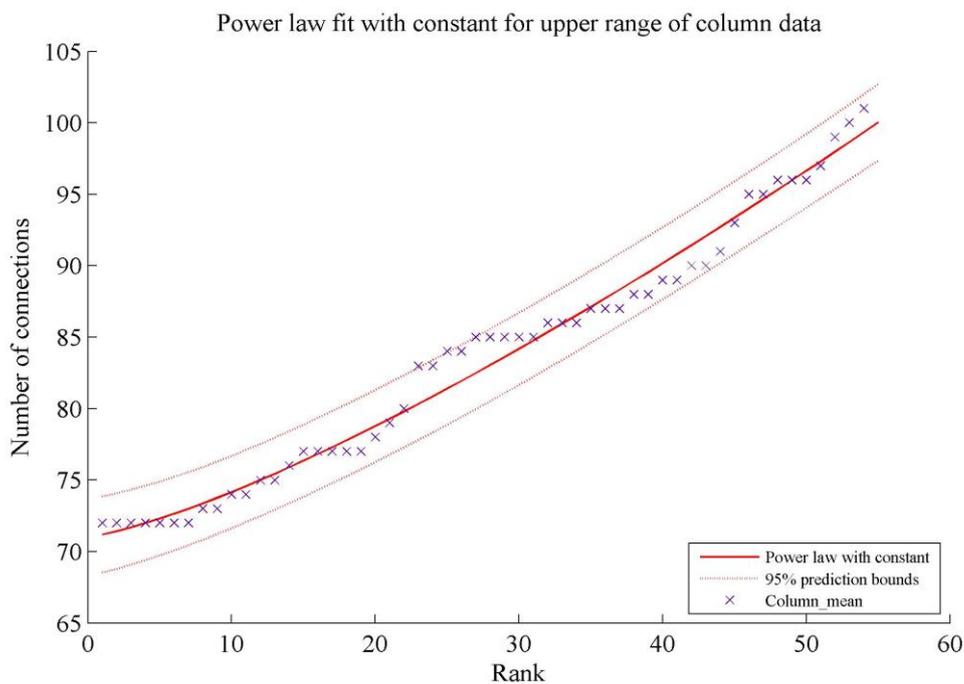


Figure 9 *Power law fit with constant for column counts greater than the mean.*

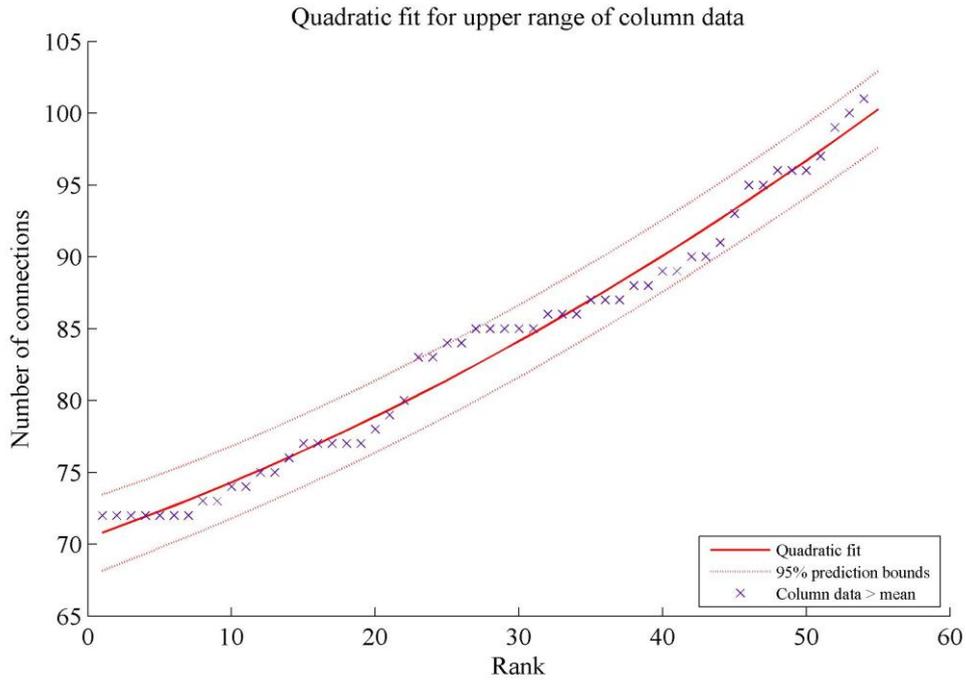


Figure 10 *Quadratic fit for column counts greater than the mean.*

For the row data, again restricting the sample to values above the mean and using the power law with a constant, the fit is not as good. The adjusted R squared is 0.957 and the standard error 414.5. The power law fit again under-predicts the upper end of the data. The quadratic is in this case clearly superior, with adjusted R squared of 0.988 and standard error of 118.9

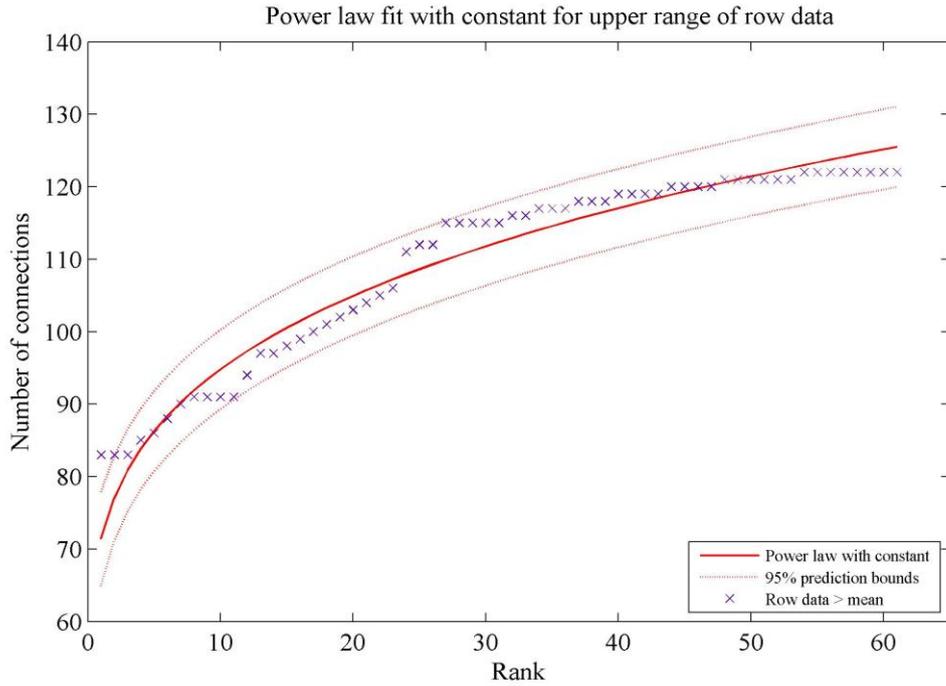


Figure 11 *Power law fit with constant for row counts greater than the mean.*

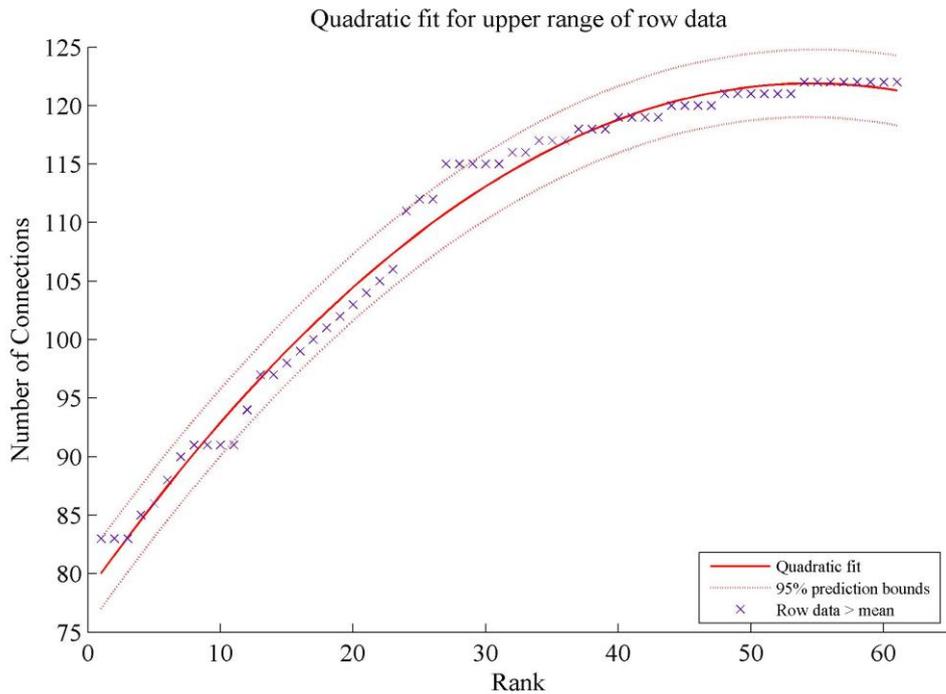


Figure 12 *Quadratic fit for row counts greater than the mean. It is clear that the upper tail of the data is better fitted with a quadratic than with a power law.*

3. Conclusion

There is increasing interest in the existence or otherwise of scaling behaviour in the economy. We examine the evidence on the connections between different industrial sectors, in terms of the value of output which each industry sells to each of the others, and the value of output which the others sell to it.

Using data from the input-output tables in the UK national accounts, we find that at the detailed level of 123 industries, there is a high degree of connectivity between industries. Most industries require as inputs, products produced by most other industries. And most industries sell their own products as inputs into most other industries.

The statistical distributions of these two types of connection are significantly different from each other and clearly non-Gaussian. However, they also do not follow a power law distribution and are likely to be from a mixed nature. Even so, one would still expect that the relationship between the rank and the number of connections would present a scaling behaviour but that does not seem to be the case here. In other words, the economic connections between industries follow non-scaling behaviour.

Acknowledgment We acknowledge financial support from the Tipping Points programme at the IHRR – Durham University funded by the Leverhulme Trust

References

- R. Pastor-Satorras, A.Vasquez and A.Vespignani, Dynamical and correlation properties of the Internet, *Physical Review Letters*, 87, 258702, 2001
- R.Albert, H.Jeong and A-L.Barabasi, Diameter of the world wide web *Nature*, 401, 130, 1999
- J.Lahererre and D.Sornette, Stretched exponential distributions in Nature and Economy: ‘Fat tails’ with characteristic scales *European Physics Journal B*, 2, 525, 1998
- F.Liljeros, C.R.Edling, L.A.N.Amaral, H.E.Stanley and Y.Berg, The web of human sexual contacts *Nature*, 411, 907, 2001

- R.L.Axtell, Zipf distribution of US firm sizes *Science*, 293, 1818, 2001
- M.H.R.Stanley, L.A.N Amaral, S.V.Buldyrev, S.Havlin, H.Leschorn, P.Maass,M.A.Salinger, H.E.Stanley, Scaling behaviour in the growth of companies *Nature*, 379, 1997
- W.Cook and P.Ormerod, Power law distribution of the frequency of demises of US firms *Physica A* , 324, 207-212, 2003
- C Di Guilmi, M Gallegati and P Ormerod, Scaling invariant of firms' exits in OECD countries *Physica A*, 334, 267-273, 2004
- T Ohnishi, H Takayasu, and M Takayasu, Network motifs in an inter-firm network *Journal of Economic Interaction and Control*, 171-180, 2010
- T Konno, Network structure of Japanese firms – hierarchy and degree correlation: analysis from 800,000 firms Economics Discussion Papers Institut für Weltwirtschaft, No. 2008-39, <http://hdl.handle.net/10419/27475>, 2008
- Y Fujiwara and H Aoyama, Large scale structure of a nationwide production network *European Physics Journal B*, 77, 565-580, 2008
- R M Goodwin, The multiplier as a matrix, *Economic Journal*, LIX, 537-551, 1949
- JD Farmer, M Gallegati, C Hommes, A Kirman, P Ormerod, S Cincotti, A Sanchez and D Helbing, A complex systems approach to constructing better models for managing financial markets and the economy, *European Physics Journal*, forthcoming, 2012
- TW Anderson and DA Darling, Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes *Annals of Mathematical Statistics* 23, 193–212, 1952
- A Clauset, CR Shalizi, and MEJ Newman, Power law distributions in empirical data, *SIAM Review* 51, 661-703, 2009
- GAF Seber CJ Wild, *Nonlinear Regression*, John Wiley and Sons, 2003
- R Perline, Strong, weak and false inverse power laws, *Statistical Science*, 20, 68-88, 2005
- GK Zipf, *Human Behaviour and the Principle of Least Effort*, Cambridge, Massachussets, 1949.

