FIRM RETURNS AND NETWORK CENTRALITY

Zornitsa Todorova *

* Department of Finance, Bocconi University, Italy

OPEN ACCESS

How to cite this paper:
Todorova, Z. (2019). Firm returns and network centrality. *Risk Governance and Control: Financial Markets & Institutions, 9(3),* 74-82.
http://doi.org/10.22495/rgcv9i3p6

ISSN Online: 2077-4303
ISSN Print: 2077-429X

Received: 02.08.2019
Accepted: 23.09.2019

JEL Classification: G01, G12, C58
DOI: 10.22495/rgcv9i3p6

Abstract

Using methods from graph theory and network analysis, this paper identifies, visualizes and analyzes a correlation network of residual stock returns for more than 5,000 US-based publicly traded firms. Building on prior work by Billio et al. (2012), the paper computes a systemic measure of network centrality using principal components analysis. Two main questions are addressed: 1) What is the empirical relationship between expected stock returns and network centrality? and 2) Does network centrality have predictive power to identify firms, which are most at risk during systemic events? First, the paper finds that network centrality has substantial predictive power in out-of-sample tests related to the recent financial crisis. Second, firms that are more central in the network earn higher returns than firms that are located in the periphery. The paper rationalizes this finding by arguing that central firms are characterized by higher market risk because they are more exposed to idiosyncratic shocks passing through the network. Finally, the paper develops a novel factor-mimicking portfolio, weighted by centrality scores. The investment strategy earns an annualized risk premium of 3.38 % controlling for market beta, size and book-to-market.

Keywords: Correlation Networks, Stock Returns, Idiosyncratic Risk, Network Spillovers

Authors' individual contribution: the author is responsible for all the contributions to the paper according to CRediT (Contributor Roles Taxonomy) standards.

1. INTRODUCTION

The standard macroeconomic diversification argument discards the possibility that firm-specific (idiosyncratic) shocks could have an impact on aggregate volatility or asset prices (Lucas, 1977). However, it ignores the fact that firms do not function in isolation but are embedded in intricate business networks. Firms in such a network are economically related: they receive shocks from their business partners through the network, and as a result, they tend to move together. This propensity to co-move with related stocks is called network risk.

Recent work by Gabaix (2012), Carvalho & Gabaix (2013), Acemoglu et al. (2012), Acemoglu, Akcigit and Kerr (2015) and Barrot and Sauvagnat (2016)[1] shows that when the firm-size distribution or network connections are sufficiently fat-tailed, the law of large numbers doesn't apply and sectoral shocks don't cancel out. An immediate implication of this result is that if idiosyncratic shocks are potential drivers of the volatility of the economic system, then companies with greater exposure to idiosyncratic shocks will be characterized by higher levels of market risk, which should be reflected in higher stock returns. Modelling the economy as a network of firms, exposure to network shocks is given by the firm's connectivity i.e. centrality in the network. To capture the network of business relationships, the paper uses information contained in asset prices. The idea is that asset prices are forward-looking and quickly reflect all information available on the market. Looking at asset correlations after having controlled for common market and industry-wide factors it is possible to infer business connections among firms.

Given this framework, the paper explores the asset pricing implications of network connections for an extensive sample of publicly traded firms in the US during the period 2001-2015. It answers two questions: 1) What is the empirical relationship between expected stock returns and network centrality? and 2) Does network centrality have predictive power to identify firms, which are most at risk during systemic events?

The contribution of the paper is mainly empirical. First, it develops an econometric measure of network connectedness building on prior work by Billio et al. (2012). The authors focus on linkages between financial institutions and compute centrality scores based on principal components analysis (PCA) of stock returns correlations. This measure is systemic in nature because it reflects how much a given financial institution

[1] Other examples of recent contributions in this field are: Acemoglu, Ozdzglar, and Tahbez-Salehi (2017), Carvalho (2014), Carvalho and Gabaix (2013), Atalay (2017).

contributes to the risk of the financial system as a whole. Typically, the literature characterizes systemic risk as any event, which threatens the stability of or confidence in the financial system. This paper argues that the concept can also be extended to a more general economic context. An illustrative example why this could be useful is the US government bailout of the automotive industry in 2009. In November 2008, Allan R. Mulally, the CEO of Ford, urged the Senate Banking Committee to support General Motors (GM) and Chrysler, Ford's most prominent rivals. He argued that due to the significant overlap of suppliers and dealers between the three automotive giants, the failure of either GM or Chrysler will have significant ripple effects across the entire US economy (Mulally, 2008). This was also a key argument for the government bailout of several financial institutions during the 2007-2008 crisis.

Moreover, differently, from Billio et al. (2012) this paper computes the connectivity measure after netting out the effect of common factors from asset prices correlations. This is important because co-movement of asset prices could be either due to exposure to common risk factors or due to the existence of linkages. Conceptually, whereas Billio et al. (2012) view connections as a manifestation of market-wide forces, this paper argues that network risk constitutes a fraction of idiosyncratic risk. The first result of the paper is that network connectivity has substantial out-of-sample predictive power to identify companies, which are most likely to suffer losses during systemic events, such as the Global Financial Crisis in 2007/8. This result is important from a practitioner's point of view because it shows that network connections can be used as an early warning or risk management tool.

Second, the paper uses the connectivity scores to construct a novel factor mimicking network risk and to show that it carries a positive risk premium. This result is in line with the findings by Ahern (2013), Kelly, Lustig and Van Nieuwerburgh (2013) and Herskovic (2018), who show that the network structure of the economy is an important factor driving asset prices. Most of the prior work in the field has worked with sectoral linkages obtained from Input-Output Tables or with customer-supplier linkages between large companies, which are unfortunately available on a very low frequency (usually every 5 years). This paper contributes to the literature by studying firm-level connections inferred from market data. The advantage of using market data is that this is an intuitive and easy-to-calculate approach, which allows calculating a time-series of centrality scores on a monthly frequency for small and big firms alike.

The paper rationalizes the positive relationship between stock returns and network centrality by arguing that more connected firms are more exposed to shocks that pass through the network. Crucial for the argument is the idea that some firms are disproportionately well connected, which means that a shock originating in their business is distributed to many firms and dies out at a lower rate. In these circumstances, firm-specific shocks do not cancel out. Network risk becomes unverifiable and investors demand a premium to hold central firms in their portfolios. The paper extends our understanding of how idiosyncratic shocks transmitted through the economy and provides a micro foundation for market risk by emphasizing the role of firm-level network connections.

This paper is organized as follows. Section 2 reviews the relevant literature. Section 3 introduces the methodology used to conduct the empirical analysis. Section 4 provides the main results of the paper. Section 5 discusses the implications of the results and presents additional checks, which corroborate the main results of the paper. Finally, Section 6 concludes and outlines areas for future research.

2. LITERATURE REVIEW

This paper speaks to two main strands of literature. First, the paper relates to the literature studying the asset pricing implications of networks. Using Bureau of Economic Analysis (BEA) industry input-output tables, Ahern (2013)[2] discovers that a factor mimicking portfolio of returns long in the highest quintile of centrality and short in the lowest quintile of centrality is positively priced in the cross-section of returns. Barrot & Sauvagnt (2016) document empirically that idiosyncratic shocks (natural disasters) transmit from suppliers to customers, especially when they produce specific inputs. On a related vein, Atalay (2017) develops a multi-sector general equilibrium model to quantify the effect of sectoral shocks to business cycle fluctuations in output. Herskovic (2018) finds that two properties of the network, concentration and sparsity, are priced in the cross-section of returns. Richmond (2019) studies network centrality in the context of currency risk premia and shows that countries, which are more central in a trade network have lower interest rates and currency risk premia. This paper contributes to the literature by addressing one of its challenges: the availability of data to identify network linkages. Data is usually available only for a subset of large firms/sectors at a yearly or 5-year frequency. This paper constructs a monthly firm-level time-series of centrality scores for 5,203 US-based publicly traded companies. From an academic standpoint, it enriches our understanding of how idiosyncratic shocks transmit at a more disaggregated level. On the other hand, from an investor's point of view, such a database could be useful for risk management and forecasting, which usually require inputs at a high frequency.

Second, by analyzing asset return commonalities, the strength of linkages between individual stocks and the sensitivities of these connections to changing economic conditions, the paper broadly relates to the literature on systemic risk and identification of systemic events. Three measures have been developed recently to capture linkages between financial institutions: conditional value-at-risk (Adrian & Brunnermeier, 2011), systemic expected shortfall (Acharya et al., 2010) and a distressed insurance premium (Huag et al., 2011). On the asset pricing side, Buraschi & Tebaldi (2017) study the implications of systemic risk in network economies. The authors show theoretically that idiosyncratic shocks could generate aggregate fluctuations and that this largely depends on the topology of the network. The paper improves our understanding of the mechanisms of systemic risk by studying it from a novel perspective. The toolkit of networks provides us with a useful framework to extend the concept of systemic risk and contagion to the broader economy.

3. DATA AND METHODOLOGY

3.1. Data

Monthly stock prices are obtained from the Center for Research in Security Prices (CRSP) database for January 2000 to December 2015. Data on common stock traded at the New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and NASDAQ Stock Exchange[3]. A firm-month observation is included if the stock has no missing monthly observations in the following 36

[2] Using a similar dataset, Ahern and Harford (2014) study the propagation of merger waves through production networks.
[3] This corresponds to CRSP share code of 10, 11 or 12.

months. The requirement for non-missing observations is quite stringent and a relevant concern could be that it introduces a selection bias. However, the monthly set of stocks in the sample changes relatively slowly, with the average replacement rate (the ration between the number of new companies, from a month to the next, and total number of companies) being less than 5%.

Economic fundamentals are obtained from the COMPUSTAT Database[4]. To avoid look-ahead bias, when calculating size and book-to-market the paper follows the matching procedure by Fama and French (1992), which imposes a minimum gap of 6 months between fiscal year-ends and firm returns. Monthly returns from July of calendar year *T* are matched to June of calendar year *T+1*. Market equity (ME) is calculated as share price times number of shares outstanding. The variable is measured in June of year *T*, for the returns between July of year *T* and June of year *T+1*. Book-to-market (BM) is computed as the ratio between book equity and market equity in December of year *T-1*. Book equity is calculated by summing shareholders' equity, balance sheet deferred taxes and investments and subtracting the book value of preferred stock. The book value of preferred stock is depending on availability of data using the following (in this order): redemption, liquidation or par value. Variables are log-transformed and winsorized at the 1st and 99th percentiles level to avoid the effect of outliers.

The dataset includes 481,609 firm-month observations from all sectors of the economy. Table 1 provides a split by a number of observations by standard industry classification (SIC) code at the division level (11 groups).

Table 1. Observations by industry division

Industry division	*Observations*
Agriculture, Forestry, Fishing	779
Mining	18,528
Construction	5,602
Manufacturing	196,915
Transportation, Utilities	40,218
Wholesale Trade	19,369
Retail Trade	29,175
Finance, Insurance and Real Estate	90,274
Services	80,702
Public Administration	44
Non-classifiable	0

The dataset provides a holistic representation of the economy. The sector with the largest number of observations is Manufacturing and the one with the smallest number in Public Administration. This is not surprising as it reflects a large degree industry size. One could be concerned that financial companies are inherently different from all other companies. They are heavily regulated, their stock has higher turnover and they usually lead the market when processing new information. If financial companies are over-represented in the sample, this could induce spurious results. Table 1 shows that this is not the case: although sizeable, the proportion of financial companies in the data is 18.74 %.

3.2. Network representation of the stock market

This study considers a correlation-based network obtained by analyzing the return dynamics of a set of stocks simultaneously traded in the US stock market. Such an approach generates a network starting from a set of time-series. Following the pioneering work of Mantegna and Stanley (1999), which introduced concepts of statistical physics in the description of financial systems, a number of recent studies have used correlations between asset returns to infer network connections (Tse et al., 2010; Curme et al., 2015; Bariggozi & Brownlees, 2019). The main theoretical assumption of these studies is that network connections are reflected in asset prices. In an efficient market, stock prices incorporate all information available to market participants, and, in equilibrium correspond to the discounted value of future dividends. The presence of a high-degree of cross-correlation is a well-known empirical fact in financial markets (Bonanno et al., 2001; Kelly, Lustig, & Van Nieuwerburgh, 2013). However, correlation in equity markets could be due to two very different mechanism. One way is due to a common macro shock, such as a monetary policy shock, which shifts the entire market. A second way is due to a firm-specific shock, which transmits to other related firms though the network and induces co-movement. This paper considers the latter mechanism. The first contribution of the paper is to provide a cleaner measure of network connectivity by first netting out the effect of common risk factors and working directly with idiosyncratic returns.

Idiosyncratic returns are constructed on a rolling 36-month basis using the following model[5]:

$$r_{it} - r_{ft} = \alpha + \beta_{MKTRF,it}MKTRF_t + \beta_{SMB,it}SMB_t + \beta_{HML,it}HML_t + \varepsilon_{it} \tag{1}$$

where r_{it} is the return of firm[6] i at time t, r_{ft} is the risk-free rate and t denotes a monthly observation from $t = 1,..,36$. *MKTRF, SMB* and *HML* are the three Fama-French factors (1992): the return on the market portfolio, the size factor and book-to-market factor respectively. The market portfolio is the return on a well-diversified portfolio of returns in excess of the risk-free rate[7]. β_{MKTRF}, β_{SMB} and β_{HML} are asset-specific exposures to the common risk-factors. ε_{it} denote idiosyncratic returns *i.e* returns orthogonal to the common risk factors. In each month, the correlation matrix of idiosyncratic returns Ψ is used to *infer* network connections[8].

A network, or a graph object in mathematics $G(V,E)$, is a set of vertices $V = \{v_1, v_2, \dots, v_N\}$ linked by edges E (Mantegna, 1999). In this paper vertices are represented by assets traded on the stock market and edges give undirected connections between them. Networks can be graphically represented by a square *adjacency matrix A* with entries a_{ij} such that:

$$= a_{ij} \begin{cases} \psi_{ij} & if\ i \neq j \\ 0 & if\ i = j \end{cases} \tag{2}$$

Entries on the main diagonal of A are set equal to zero i.e. self-loops are excluded. The strength of the link between two assets i and j is given by their correlation. The result of this procedure is an undirected, time-varying and fully connected network. Note that the paper is agnostic with respect to the nature of connections: it could be a direct relationship as in the case of a supplier or a customer, or an indirect link such as a marketing consultancy, a financial audit firm or a logistics service provider.

3.3. Network centrality

Due to the finiteness of the financial time-series data, the correlation network described in Section 3.2 contains some degree of noise. In order to remove the

[4] Data (subject to subscription) is available to download from https://wrds-web.wharton.upenn.edu/wrds/

[5] A detailed version of this paper is available for download from the author's website at https://zornitsa-todorova.com/research

[6] The terms firms, assets and stocks are used inter-changeably in this paper.

[7] Data on *MKTRF, SMB* and *HML* is obtained from Kenneth French's online data library available at: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[8] Note that adding industry fixed effects to the econometric model in (1) would be possible to purge within-industry connections. However, the paper takes the stand that both between within-industry and across-industry connections contain valuable information for asset prices, and, so differentiating between the two is outside of the scope of the paper.

less relevant information, a suitable filtering procedure has to be performed. The most common approach to address this problem is principal components analysis (PCA) of the correlation matrix of the data (Bonanno et al., 2004).

The paper follows closely the PCA procedure suggested by Billio et al. (2012). The idea behind PCA is to describe the covariance structure of a given set of variables by identifying the primary sources of variation. By identifying the sources, PCA reduces the dimensionality of the data to a few common orthogonal factors of decreasing explanatory power. In order to do so, PCA computes the eigenvalues and the associated eigenvectors of the covariance matrix. More formally, the univariate measure of the connectivity of each stock to the system is:

$$PCAS_{i,k} = \sum_{k=1}^{K} \frac{\sigma_i^2}{\sigma_S^2} L_{ik}^2 \lambda_k \,|K < N \quad (3)$$

where σ_i^2 is the idiosyncratic variance of a stock, σ_S^2 is total system idiosyncratic variance and L is a matrix of eigenvectors, associated with the eigenvalues λ_k. The *PCAS* measure is systemic in nature: it gives the contribution of each stock towards the riskiness of the system as a whole. Intuitively, *PCAS* reflects endogenous risk transmitted through the system due to the fact that stocks are connected in a network.

Central to the theoretical and the empirical validity of the paper is the correct specification of the number of principal components, the loadings on which are used to derive *PCAS*. Following the Kaiser criterion, only those eigenvectors, where the associated eigenvalue is greater than 1, are retained. Then, three centrality measures are computed: *PCAS1* (using the 1st principal component), *PCAS5* (using principal components 1 to 5) and *PCAS10* (using principal components 1 to 10). Since *PCASs* are highly skewed, the variables are log-transformed and summary statistics are presented in Table 2.

Table 2. Summary statistics of centrality

	logPCAS1	*logPCAS5*	*logPCAS10*
Mean	0.710	1.372	1.530
Standard Dev.	1.134	0.606	0.547
Variance	1.280	0.367	0.299
Skewness	-0.911	0.154	0.325
Kurtosis	4.061	2.839	2.857

Note: LogPCAS1 is negatively skewed with extra kurtosis of 1.061 compared to the normal distribution. LogPCAS5 and LogPCAS10 look similar to each other: they are positively skewed and have kurtosis less than 3. In terms of volatility, PCAS1 scores highest with standard deviation equal to 1.134, followed by PCAS5 with 0.606 and PCAS10 with 0.547.

3.4. Alternative measures

A number of measures have been developed in the literature to characterize centrality in networks. Some examples include in-degree and out-degree, closeness, betweenness, diameter and eigenvector centrality and Katz centrality. The reader is referred to Borgatti (2003) for a detailed discussion of the use of these measures and the assumptions underlying them.

3.5. Hypothesis

The main hypothesis of this paper is that a stock's true market riskiness is, in part, influenced by its relative position in the network of stock returns. If idiosyncratic shocks can be accumulated to form aggregate shocks, then such local shocks will affect asset prices. If shocks are transmitted through the business network, then those companies that are more connected to the economy are more exposed to such shocks. Moreover, stocks in business networks are economically related, their asset prices tend to move together and such commonality will be reflected in the correlation matrix. Finally, companies cannot fully protect themselves against these shocks through diversification. Consequently, network risk becomes undiversified and the risk should be priced in the cross-section of returns.

4. RESULTS

4.1. Network centrality and traditional risk measures

If the hypothesis that network centrality captures (in part) a firm's riskiness is true, then centrality should be positively correlated with other risk measures traditionally used in the finance literature. Table 3 presents the relationship between *PCAS* and risk measure standardly used in the asset pricing literature: market beta. Market betas, $\beta_{MKTRF,it}$, are the exposures to the market factors, calculated from the model specified in Equation (1). $\beta_{MKTRF,it}$ is interpreted as a firm-specific measure of riskiness relative to the market. The market itself has a beta of 1; beta > 1 means that a stock is riskier than the market; beta < 1 means the stock is less risky than the market.

Table 3. Risk measures

	logPCAS1	*logPCAS5*	*logPCAS10*
$\beta_{MKTRF,it}$	0.12***	0.28***	0.33***
	(<0.00)	(<0.00)	(<0.00)

*Note: the table reports correlation coefficients between centrality measures PCAS1, PCAS5, PCAS10 and market beta. Values in the parentheses are p-values. Statistical significance at the 10%, 5%, and 1% level is indicated by *, ** and ***.*

The results indicate that all centrality measures are positively correlated with market beta, which offers first preliminary evidence that central, connected firms are characterized by higher levels of risk.

Applying the *PCAS* measure outlined in Section 3.3 it is possible to calculate a centrality score for each stock *i*, which is updated every month t. The advantage of using market data is that they are easily available, have higher frequency (allows to update the links on a continuous basis) and are forward-looking in contrast to connections extracted from accounting data, which provide "snapshots" and could be considered as backward looking. The practical value-added of this approach can be described in the following way. Every month the distribution of centralities is divided into 5 quintiles such that quintile 1 contains the least central stocks and quintile 5 contains the most central stocks. If stocks do not migrate at all between quintiles over time, then having a high-frequency time-series of centrality scores is not very useful. It is possible to calculate (average) transition probabilities i.e the probability that a stock is going to change quintiles in the following month. Table 4 shows that in fact stocks centralities change over time. For example, the probability that a firm assigned to quintile 5 is to going to be assigned to quintile 1 in one month is 0.13. This information is particularly valuable for active portfolio managers, who rebalance continuously to achieve higher performance targets. Instead, if the manager relies on accounting type of linkages available at a yearly frequency, she would be missing out an important aspect of the market dynamics.

Table 4. Transition probabilities

State	*1*	*2*	*3*	*4*	*5*
1	0.3034	0.2327	0.1824	0.1503	0.1312
2	0.2337	0.2603	0.2078	0.1663	0.1318
3	0.1814	0.2090	0.2558	0.2075	0.1462
4	0.1517	0.1647	0.2059	0.2895	0.1882
5	0.1300	0.1335	0.1484	0.1867	0.4014

Note: the table gives the probability that a stock is going to migrate from one state to another in the following month. States 1 to 5 correspond to quintiles 1 to 5 of the network centrality distribution. Results are shown for PCAS1.

4.2. Factor-mimicking portfolio

To test empirically whether centrality is associated with higher risk premia, centrality scores for each stock in the sample are used to create a factor mimicking portfolio of network risk. A mimicking portfolio is a portfolio of assets designed in a way to stand for a background factor. Such a design is usually preferred to directly using the factor in asset pricing tests especially when the realizations of the factor are not returns. Using a mimicking portfolio instead, the researcher retains only the information in the factor relevant for asset returns and reduces the amount of noise present in the model.

Let X be an $N \times 1$ vector of centralities in month t, normalized such that $X'X = 1$. Then, it follows that $CNTR_t = X'\varepsilon_t$ is a common factor, calculated as a weighted average of idiosyncratic returns. The weights are given by centrality scores and ε_t is a vector of idiosyncratic returns obtained from Equation (1). To obtain exogeneity of the weights, network centralities are lagged by 12 months[9].

4.3. Two-pass cross-sectional regression

To test whether $CNTR_t$ is priced in the cross-section of returns the paper follows the two-pass procedure by Fama and Macbeth (1973): a time-series regression, followed by a cross-sectional regression. In the first step, each asset's idiosyncratic return $\varepsilon_{i,t}$ is regressed against the $CNTR_t$ time series to determine how exposed it is to the factor:

$$\varepsilon_{i,t} = \alpha_i + \beta_{CNTR,i} CNTR_{t-12} + \eta_{i,t} \quad i = 1, \dots, N \tag{4}$$

N regressions are calculated to obtain factor exposures, $\beta_{CNTR,i}$, for each $i = 1,..,N$. In the second pass, the cross-section of excess returns is regressed on $\beta_{CNTR,i}$, at each time step:

$$r_{it} - r_{ft} = \lambda_{0t} + \lambda_{CNTR,t}\beta_{CNTR,i} + v_{it} \quad t = 1, \dots T \tag{5}$$

The main object of interest in the T cross-sectional regressions is the λ's, which measure the price of risk or risk premia. The result of the second step is to give a time-series of risk premia. The insight of the Fama-Macbeth procedure is to average these coefficients, which gives the expected premium for a unit of exposure to the centrality risk factor over time. Hence, $\lambda_{CNTR} = \frac{\sum_{t=1}^{T}\lambda_{CNTR,t}}{T}$. If λ_{CNTR} is positive, this means that the market compensates investors for accumulating network risk.

The fact that factor exposures ($\beta_{CNTR,i}$) and the factor premium (λ_{CNTR}) are not calculated on the same data sample is a crucial point here. Lagging centralities by 12 months, the estimation sample does not suffer from look-ahead bias. This allows making statements about risk out-of-sample, which is particularly useful for investment purposes.

In some specifications of the Fama-Macbeth regressions control for market beta, size (ME), book-to-market (BM), turnover and idiosyncratic volatility (IVOL) are included. Market betas are estimated from the time-series regressions specified in Equation (1) and IVOL is the standard deviation of the residuals (ε_{it}) from this model. Turnover measures liquidity and is calculated as trading volume divided by a number of shares outstanding. All controls are log-transformed and lagged by 1 month. Table 5 shows summary statistics of the variables used in the Fama-Macbeth regressions. Excess returns have a mean on 1.2 %, a standard deviation of 15% and a positive skew. The average market beta is approximately 1 and the average network beta $\beta_{CNTR} = 0.71$. Two things are worth mentioning here. First, the magnitude of the network beta is very close to that of the market beta and, second, the network beta is considerably more positively skewed than the market beta. This result suggests that there is a small number of firms that load very heavily on the network risk factor.

The logarithm of size has a mean of 3.82 and a standard deviation of 2.161. LogBM, LogTrnv and LogIVOL have negative means and very low values of skewness, which is a result of the logarithmic transformation.

Table 5. Summary statistics of variables

	Mean	***St. Deviation***	***Q1***	***Q3***	***Skew***
Panel A: Stock characteristics					
Exret	0.012	0.15	-0.055	0.064	4.00
Beta	1.05	0.96	0.47	1.49	1.08
β_{CNTR}	0.71	0.45	0.43	0.85	2.61
LogSize	3.82	2.161	2.22	5.3	0.18
LogBM	-0.58	0.86	-1.05	-0.09	-0.40
LogTrnv	-0.16	1.28	-0.88	0.71	-0.73
LogIVOL	-2.13	0.52	-2.50	-1.70	0.31

Table 6 contains the main result of the paper. The estimates in Column (1) indicate that the network risk factor is positively priced in the cross-section of returns. Economically, the result means that an investor who buys network risk is expected to earn a monthly risk-premium of 0.352 % in excess of the risk-free rate. On an annualized basis, this amount to 4.22 %.

The results in Table 3 indicate that the position of stock in the network is strongly related to market risk. Controlling for market beta in Column (2) reduces the magnitude of the risk premium to 0.23 %, but it does not affect its statistical significance. Moreover, the effect is robust to including size and book-to-market in Column (3) and volatility and turnover in Column (4). This result suggests that the centrality factor accounts for a different dimension of risk, which is not captured by asset pricing factors previously studied in the literature.

One of the arguments of the paper is that the connectivity measure, which Billio et al. (2012) apply to the financial services sector only, can be extended to the economy as a whole. Therefore, a major concern could be that the results in Table 6 (see Appendix) are driven by financial companies. To address this concern, Column (5) repeats the baseline regression from Column (1) by excluding all financial companies[10]. The magnitude and significance of the risk premium estimate remain largely unchanged. This evidence is reassuring because it shows that the concept of network risk is not confined to the financial industry.

4.3 Out-of-sample results during the financial crisis

Given its systemic nature, Billio et al. (2012) suggest that one important application of connectivity measures is to provide early warning signals to financial regulators and risk managers. It would be interesting to test whether the *PCAS* measures have any power to predict losses during systemic events outside of the finance sector.

To test the out-of-sample performance of the measure, stocks are first ranked based on *PCAS1*, *PCAS5* and *PCAS10*. Then, the variable *Maximum % Loss* is computed and stocks are also ranked according to it.

[9] Results in the subsequent section are based on $X = logPCAS10$. The results and conclusions do not change if instead *logPCAS1* or *logPCAS5* are used. These results are omitted for brevity and are available from the author upon request.

[10] Financial companies are denoted by SIC codes 6000 – 6799.

This is the maximum percentage loss suffered by each stock in the sample during the recent financial crisis period from July 2007 to December. The maximum percentage loss for a firm is defined to be the difference between market capitalization in the end of June 2007 and the minimum market capitalization during the period from July 2007 to December 2008, divided by the market capitalization in the end of June 2007. Following Acharya, Pedersen, Philippon and Richardson (2017), two estimation samples for the *PCAS* measures are used: October 2002 – September 2005 and July 2004 – June 2007. Table 7 gives the estimates of regressing *Maximum % Loss* on the stock's PCAS rankings.

Table 7. Predictive power of network centrality

	Maximum % Loss			
	Coeff	*t-stat*	*p-value*	*Kendall τ*
	October 2002-September 2005			
PCAS 1	0.01	1.14	0.256	0.01
PCAS 1-5	0.12***	7.00	0.00	0.08
PCAS 1-10	0.17***	9.49	0.00	0.11
	July 2004-June 2007			
PCAS 1	0.05***	3.22	0.001	0.03
PCAS 1-5	0.19***	10.92	0.00	0.12
PCAS 1-10	0.24***	14.19	0.00	0.16

*Note: the table shows coefficients, t-stats, p-values and Kendall (1938) τ rank-correlation coefficients of regressions of Maximum % Loss on PCAS 1, PCAS 1-5 and PCAS 1-10. Estimates are shown for two samples: October 2002- September 2005 and July 2004 to June 2007. Statistical significance at the 10%, 5%, and 1% level is indicated by *, ** and ***.*

Stocks that were more exposed to network risk, *i.e.* with larger *PCAS* loadings, are more likely to suffer considerable losses during the recent financial crisis. The beta-coefficients are significant at the 5 % level, which suggests that the centrality measure correctly identifies stocks that will be more affected during crisis periods.

4.4. Discussion

The empirical evidence outlined in the previous sections supports the hypothesis that firms that are central in the network of business relationships earn higher returns as compensation for higher exposure to network risk. This result is consistent with previous studies in the field, such as Ahern (2013), which finds that centrality bears a positive monthly premium. The paper acknowledges that a positive loading on *CNTR* could be reconciled with multiple hypotheses. The most natural explanation could be that it captures systematic risk that is not accounted by the other standard factors. Alternatively, as Ahern (2013) argues, it could be the case that the ex-post beta coefficient is measured poorly, and thus, *CNTR* gives a better estimate of ex-ante exposure to market risk.

5. ROBUSTNESS

5.1. Sub-samples analysis

The financial crisis of 2007-2008 induced a key structural break in the US economy: liquidity was scarce, financial regulation tightened and interest rates plummeted close to the zero-lower bound as a result of expansionary monetary policy. Performing the analysis over two different sub-samples, crisis (2007-2008) and non-crisis (2001-2006; 2009-2015), shows that the network risk-premium was nearly 2 times higher during the crisis period. One explanation for this result could be that periods of the economic crisis are characterized by high stock market volatility and high instability, which translates into higher risk-aversion. Therefore, risk-averse investors should require a higher risk premium during troughs than during expansions.

Table 8. Sub-sample analysis

	(1) No crisis	*(2) Crisis*
betaCNTR	0.322*** (<0.00)	0.502*** (0.014)
Constant	0.97*** (<0.00)	0.83*** (<0.00)

*Note: this table presents results of two-stage monthly cross-sectional regressions of average excess firm-level returns on centrality beta. The crisis period refers to the time frame 2007-2009. The CNTR factor is calculated using residuals weighted by LogPCAS10 centrality. Coefficients estimates are in percentages. Statistics are computed using the Newey-West procedure with lag length 1. Values in the parentheses are p-values. Statistical significance at the 10%, 5%, and 1% level is indicated by *, ** and ***.*

5.2. Test assets

This section tests whether the *CNTR* factor is able to price other test assets besides firm-level returns. Three sets of assets are used: 25 portfolios sorted on size and book-to-market, 25 portfolios sorted on centrality and size and 25 portfolios sorted on centrality and book-to-market and firm-level returns. Portfolio betas are estimated in time-series regressions controlling for the three Fama-French factors (1992): *MKTRF* (the return spread of the value-weighted CRSP portfolio minus the risk-free rate), *SMB* (the return spread of small minus big stocks, i.e. size effect) and *HML* (the return spread of high/cheap minus low/expensive Book-to-Market stocks, i.e. value effect).

Table 9. Robustness: Fama-Macbeth cross-sectional regression

	(1) 25 Size and BM Portfolios	*(2) 25 Size and Centrality Portfolios*	*(3) 25 BM and Centrality Portfolios*
betaMktrf	-0.62 (0.613)	-0.037 (0.650)	-0.19 (0.861)
betaSMB	0.19 (0.415)	0.155 (0.640)	0.41 (0.717)
betaHML	0.019 (0.715)	0.37 (0.650)	0.35 (<0.434)
betaCNTR	0.44 (0.675)	0.93** (0.03)	1.14** (0.022)
Constant	1.50 (0.283)	0.73 (0.353)	1.89*** (<0.00)
R^2	17.85	19.93	20.12

*Note: this table presents results of two-stage monthly cross-sectional regressions of average excess firm-level returns on centrality beta and a set of controls over the period December 2002: December 2015. The CNTR factor is calculated using residuals weighted by LogPCAS10 centrality. Betas are estimated from time-series regressions of Fama-French 3 Factor Model augmented with CNTR. Coefficients estimates are in percentages. T -statistics are computed using the Newey-West procedure with lag length 1. Values in the parentheses are p-values. Statistical significance at the 10%, 5%, and 1% level is indicated by *, ** and ***.*

Table 9, Column (1) presents results that are consistent with prior research: there is a negative market beta and a positive constant. The factor premium on *CNTR* is positive, but not significant and hence, *CNTR* does not help explain variation in these portfolios. In contrast, when portfolios sorted on size and centrality (Column (2) and BM and centrality Column (3)) are used, *CNTR* is statistically significant and carries a positive risk premium of 0.93 % and 1.14 % respectively.

5.3. Relationship to macroeconomic variables

This section studies the relationship between centrality and network risk and macroeconomic variables. Three macroeconomic variables are considered: unemployment, recessions and consumption growth. The recession variable is based on the US National Bureau of Economic Research (NBER) definitions[11]. NBER maintains the most comprehensive chronology of the US business cycle. The chronology contains alternating dates of peaks and troughs of the economy. A recession is defined as a period between a trough and a peak. During the time period that the paper considers, there are 4 announcement dates: March 2001 (peak), November 2001 (trough), December 2007 (peak) and June 2009 (trough). The analysis introduces a business cycle variable, which equals 1 during troughs and 0 during peaks. In the months between these announcement dates, the recession variable is filled with linearly interpolated values. Consumption growth is defined as the growth in nondurable consumption per capita in the future 6 months.

Table 10. Cross-sectional regression with macro variables

	Mktrf	*SMB*	*HML*	*CNTR*
Unemployment	-0.235 (-0.57)	-0.057 (-0.23)	-0.000 (-0.22)	-0.503 (-0.62)
Recession	-3.192 (-1.31)	-2.010 (-1.54)	-0.010 (-0.82)	-7.41 (-1.62)
Consumption	0.405** (2.34)	-0.043 (-0.59)	-0.001 (-1.42)	0.311** (2.15)
Time Trend	Yes	Yes	Yes	Yes
Observations	157	157	157	157
R^2	0.093	0.047	0.044	0.119

*Note: the table presents the results of regressions of 3FF Factors and Centrality on macroeconomic outcomes over the period December 2002 to December 2015. Recession is based on NBER business cycle announcement dates and equals 1 for troughs and 0 for peaks, with values in the intervening months being linearly interpolated. Consumption growth is defined as the growth rate in nondurable consumption per capita in the future 6 months. T-statistics are given in brackets. Statistical significance at the 10%, 5%, and 1% level is indicated by *, ** and ***.*

In Table 10 cross-sectional regressions of *MKTRF, SMB, HML* and *CNTR* are ran on unemployment, recession, consumption growth and a time trend. Consumption growth is positively related to *Mktrf* and *CNTR* with coefficients of nearly equal magnitudes. Unfortunately, the other two macroeconomic outcomes are not related to any of the factors. In the case of recession, one potential explanation could be the limited amount of announcement dates, which makes the interpolation procedure less precise and reduces the power of these tests.

The results are consistent with the hypothesis that central firms face more economic risks than peripheral ones. When *CNTR* is high i.e. the returns of more central firms are higher, then future consumption growth increases as well. Furthermore, the fact that *CNTR* bears similarities to the market premium means that centrality could enrich our understanding of the driving forces of market risk.

5.4. Economic mechanism

To discuss in more detail the economic intuition behind the main result of the paper, Figure 1 (see Appendix) plots the empirical densities of the three centrality measures. Estimates are based on a normal kernel function, which is evaluated at a 100 equally-spaced points. It is immediate to observe that the distribution is heavily skewed to the right, which suggests that there is a small number of firms that are highly connected to the rest of the economy. To further characterize such heavy-tailed distributions, Panel B plots the empirical counter-cumulative distribution function on a log-log scale. The nearly linear relationship indicates that the right tail of the distribution can be approximated by a power-law distribution. The presence of a fat right tail is crucial because it shows how the diversification argument could break down. A shock hitting a particularly central firm is transmitted through the network and is translated into aggregate volatility. Hence, investors demand a premium to hold central firms due to network risk.

6. CONCLUSION

This paper investigates the empirical relationship between network centrality and firm returns. By identifying links between firms on the stock market, their strength and sensitivity to changing market conditions and quantifying the individual contribution of stock to the idiosyncratic volatility of the system, the paper offers insights into how micro shocks translate into the aggregate economy.

The contribution of this paper is mainly empirical. First, it creates a monthly database of centrality scores for more than 5,000 US-based publicly traded firms. It documents substantial heterogeneity in the degree of connectivity of firms and shows that firm-level centrality changes over time. These features of data make a high-frequency time-series of centrality a useful tool to follow market dynamics and monitor the buildup of risk, from which regulators, portfolio managers and risk managers could benefit. Additionally, the paper shows that an essential feature of linkages is its fat-tailed distribution. The fact that there is a small number of firms, who play a disproportionately important role in the economy, becomes extremely important when the regulator has to decide whether to bail out a distressed company. Taking a network perspective would help evaluate the scope and magnitude of potential spillovers and help the regulator make an informed decision.

Second, using the cross-sectional variation in the data the paper constructs a novel factor mimicking network risk using centrality scores as weights. The paper finds that higher centrality predicts higher expected returns on the stock market. From a single firm's perspective, a better understanding of interlinkages between firms is crucial because it has implications for corporate policy decisions (e.g., hiring, compensation, investment) as well as for firm value and asset prices.

One caveat of this paper is that estimating centrality scores using PCA precludes the possibility of assigning directionality to the inferred connections. In fact, PCA allows to establish whether there is a connection between firm A and firm B, but it does not allow to establish the direction of the link. One potential extension for future research could be to use Granger-causality to establish the direction of links and then to use the information to study the dynamic propagation of shocks from close to distant firms in the network.

A second avenue for future return could be to study how firm-specific shocks transmit through such a directed correlation network. One way to cleanly identify idiosyncratic shocks is to use the occurrence of natural disasters. A third avenue to follow relates to portfolio construction. It would be interesting to see whether incorporating information about the linkages between stocks improves asset allocation and portfolio performance. Fourth, it is a well-known fact that the correlation between a portfolio of stocks diminishes as the time horizon used to compute stock returns is decreased (Bonanno, 2001). The existence of this phenomenon, known as the "Epps Effect", motivates the investigation of the properties of the correlation network as a function of the horizon of the return time series used to reconstruct it.

[11] A detailed list of NBER announcement dates is available at https://www.nber.org/cycles.html

REFERENCES

1. Acemoglu, D., Akcigit, U., & Kerr, W. (2015). *Networks and the macroeconomy: An empirical exploration* (MIT Department of Economics Working Paper No. 15-05). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2632793
2. Acemoglu, D., Carvalho, M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica, 80(5),* 1977-2016. https://doi.org/10.3982/ECTA9623
3. Acharya, V., Pedersen, L., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies, 30(1),* 2-47. https://doi.org/10.1093/rfs/hhw088
4. Adrian, T., & Brunnermeier, M. (2011). *CoVaR* (National Bureau of Economic Research Working Paper 17454). https://doi.org/10.3386/w17454
5. Ahern, K. R. (2013). *Network centrality and the cross-section of stock returns* (National Bureau of Economic Research Working Paper). https://doi.org/10.2139/ssrn.2197370
6. Ahern, K. R., & Harford, J. (2014). The importance of industry links in merger waves. *The Journal of Finance*, *69(2),* 527-576. https://doi.org/10.1111/jofi.12122
7. Atalay, E. (2017). How important are sectoral shocks? *American Economic Journal*, *9(4),* 254-280. https://doi.org/10.1257/mac.20160353
8. Barigozzi, M., & Brownlees, C. (2019). Nets: Network estimation for time series. *Journal of Applied Econometrics, 34(3),* 347-364. https://doi.org/10.1002/jae.2676
9. Barrot, J. N., & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics, 131(3),* 1543-1592. https://doi.org/10.1093/qje/qjw018
10. Billio, M., Getmansky, M., Co, W., & Pellizon, L. (2012). Econometric measures of connectedness and systemic risk in finance and insurance sectors. *Journal of Financial Economics, 104(3)*, 535-559. https://doi.org/10.1016/j.jfineco.2011.12.010
11. Bonanno, G., Lillo, F., & Mantegna, R. N. (2001). High-frequency cross-correlation in a set of stocks. *Quantitative Finance, 1(1),* 96-104. https://doi.org/10.1088/1469-7688/1/1/306
12. Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B*, *38(2),* 363-371. https://doi.org/10.1140/epjb/e2004-00129-6
13. Borgatti, S. (2003). The key player problem. In R. Breiger, K. Carley, & P. Pattison, *Dynamic social media network modeling and analysis* (pp. 241-252). New York: National Academy of Sciences Press. https://doi.org/10.2139/ssrn.1149843
14. Buraschi, A., & Tebaldi, C. (2017). *Asset pricing in network economies with systemic risk.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3074012
15. Carvalho, V., & Gabaix, X. (2013). The great diversification and its undoing. *The American Economic Review, 103(5),* 1697-1727. https://doi.org/10.1257/aer.103.5.1697
16. Curme, C., Tumminello, M., Mantegna, R. N., Stanley, H. E., & Kenett, D. Y. (2015). Emergence of statistically validated financial intraday lead-lag relationships. *Quantitative Finance*, *15(8),* 1375-1386. https://doi.org/10.1080/14697688.2015.1032545
17. Fama, E., & French, K. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47(2)*, 427-465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x
18. Fama, E., & Macbeth, J. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy, 81(3),* 607-636. https://doi.org/10.1086/260061
19. Gabaix, X. (2012). The granular origins of aggregate fluctuations. *Econometrica, 79(3),* 733-772. https://doi.org/10.3982/ECTA8769
20. Herskovic, B. (2018). Networks in production: Asset pricing implications. *Journal of Finance*, *73(4)* , 1785-1818. https://doi.org/10.1111/jofi.12684
21. Huang, X., Zhou, H., & Zhu, H. (2012). Systemic risk contributions. *Journal of Financial Services Research, 42(1),* 55-83. https://doi.org/10.1007/s10693-011-0117-8
22. Kelly, B., Lustig, H., & Van Nieuwerburgh, S. (2013). *Firm volatility in granular networks.* (National Bureau of Economic Research Working Paper 19466). https://doi.org/10.3386/w19466
23. Lucas, R. (1977). Understanding business cycles. *Carnegie-Rochester Conference Series on Public Policy, 5.* https://doi.org/10.1016/0167-2231(77)90002-1
24. Mantegna, R., & Stanley, H. (1999). *Introduction to econophysics: Correlations and complexity in finance.* Cambridge, United Kingdom: Cambridge University Press. https://doi.org/10.1017/CBO9780511755767
25. Richmond, R. J. (2019). Trade network centrality and currency risk premia. *The Journal of Finance*, *74(3),* 1315-1361. https://doi.org/10.1111/jofi.12755
26. Tobias, A., & Brunnermeier, M. K. (2016). CoVaR. *The American Economic Review, 106(7),* 1705-1741.
27. Tse, C. K., Liu, J., & Lau, F. C. (2010). A network perspective of the stock market. *Journal of Empirical Finance, 17(10),* 535-559. https://doi.org/10.1016/j.jempfin.2010.04.008

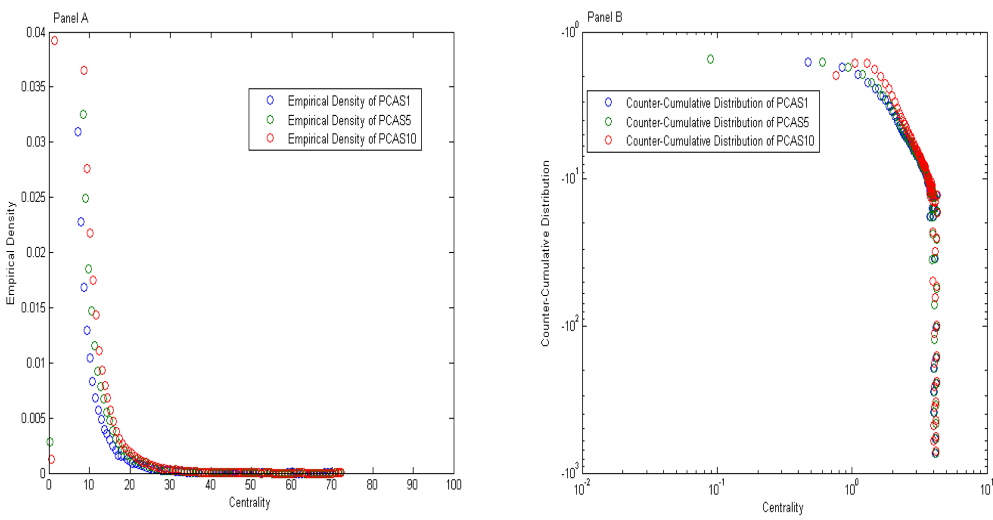
APPENDIX

Table 6. Fama-Macbeth cross-sectional regression

	(1)	*(2)*	*(3)*	*(4)*	*(5)*
betaCNTR	0.352*** (<0.00)	0.234*** (0.008)	0.165* (<0.08)	0.282** (<0.07)	0.344*** (<0.00)
betaMktrf		0.12*** (<0.00)	0.125** (0.03)	0.05 (0.127)	
LogSize			-0.172*** (<0.00)	-0.261*** (<0.00)	
LogBM			0.455*** (<0.00)	0.481*** (<0.00)	
LogTrnv				0.317*** (<0.00)	
LogIVOL				-0.02 (0.75)	
Constant	0.864*** (<0.00)	0.88*** (<0.00)	0.89*** (<0.00)	0.237*** (<0.00)	1.05*** (<0.00)
Includes FIN	YES	YES	YES	YES	NO
R^2	0.009	0.015	0.070	0.077	0.010

*Note: this table presents results of two-stage monthly cross-sectional regressions of average excess firm-level returns on centrality beta and a set of controls over the period December 2002: December 2015. The CNTR factor is calculated using residuals weighted by LogPCAS10 centrality. Coefficients estimates are in percentages. Statistics are computed using the Newey-West procedure with lag length 1. Values in the parentheses are p-values. Statistical significance at the 10%, 5%, and 1% level is indicated by *, ** and ***.*

Figure1. Empirical density of centrality



Note: Panel A shows the empirical density function of PCAS, PCAS5 and PCAS10. The estimate is based on a normal kernel function and is evaluated at 100 equally-spaced points. Panel B gives the counter-cumulative distribution on a log-log scale.