Replication

Gleb Pashchenko

New Economic School

(Dated: December 11, 2019)

I. INTRODUCTION

Throught article I will have two pictures side by side: result from my replication on the left and result from Billio [1] on the right.

II. DATA

At first, I did not have access to data that Billio et al. used. I used Bloomberg to download data for all types using the same SIC code as in paper. I did not limit my search to currently active funds. I faced two problems. First of all, there was no data on hedge funds and second of all observations for early dates were very scarce, for example there were only 14 broker-dealers for 1994 while I needed 25 securities for the analysis similar to Billio et al. I did not have data on hedge funds, so I asked my classmate who used data from CISDM Morningstar database for hedge funds returns. This is not the same database as in Billio et al. (they use Lipper Hedge Fund Database), but I had to use what I had.

Later I found out that Mila Sherman actually responded to my message and shared data as well as partial MatLab code. I found this message in my spam folder and was really surprised. Data on hedge funds was missing but I recreated it using data that was shared by my classmate.

I also added new data to the files shared with me, so now I had observations until Jan 2018. The shared files where in xls format with 144 sheets, where each had top 100 financial institutions and observations for 36 month rolling period. To extend data I followed the same procedure as in Billio et al., sorting by market cap at each 36 month rolling period and selecting top 25 in the following categories: brokers, insurers,banks and hedge funds.

Below is summary of data that replicates figure 1. Full sample results are different because I also added data from 2008 onwards. Billio perhaps confusingly uses annualized values for mean and median but simple monthly min and max. I calculated all statistics for annualized variables.

Table 1
Summary statistics. Summary statistics for monthly returns of individual hedge funds, broker/dealers, banks, and insurers for the full sample: January 1994 to December 2008, and five time periods: 1994–1996, 1996–1998, 1999–2001, 2002–2004, and 2006–2008. The annualized mean, annualized standard deviation, minimum, maximum, median, skewness, kurtosis, and first-order autocorrelation are reported. We choose the 25 largest financial institutions (as determined by average AUM for hedge funds and average market capitalization for broker/dealers, insurers, and banks during the time period considered) in each of the four financial institution sectors.

Full sample								
	Mean (%)	SD (%)	Min (%)	Max (%)	Median (%)	Skew.	Kurt.	Autocorr.
Hedge funds	12	11	−7	8	12	−0.24	4.40	0.14
Brokers	23	39	−21	32	14	0.23	3.85	−0.02
Banks	16	26	−17	19	17	−0.05	3.71	−0.09
Insurers	15	28	−17	21	15	0.04	3.84	−0.06
January 1994–December 1996								
	Mean (%)	SD (%)	Min (%)	Max (%)	Median (%)	Skew.	Kurt.	Autocorr.
Hedge funds	14	15	−8	12	12	0.25	3.63	0.08
Brokers	23	29	−15	22	21	0.26	3.63	−0.09
Banks	29	23	−12	16	29	−0.05	2.88	0.00
Insurers	20	22	−11	17	16	0.20	3.18	−0.06
January 1996–December 1998								
	Mean (%)	SD (%)	Min (%)	Max (%)	Median (%)	Skew.	Kurt.	Autocorr.
Hedge funds	13	18	−15	11	18	−1.12	6.13	0.15
Brokers	31	43	−29	37	26	0.06	5.33	−0.03
Banks	34	30	−23	22	35	−0.53	5.17	−0.10
Insurers	24	29	−19	21	24	−0.13	3.60	−0.03
January 1999–December 2001								
	Mean (%)	SD (%)	Min (%)	Max (%)	Median (%)	Skew.	Kurt.	Autocorr.
Hedge funds	14	11	−6	9	11	0.08	3.99	0.15
Brokers	28	61	−26	55	−2	0.76	4.19	−0.03
Banks	13	33	−19	24	8	0.21	3.26	−0.10
Insurers	10	41	−22	34	2	0.62	4.21	−0.16
January 2002–December 2004								
	Mean (%)	SD (%)	Min (%)	Max (%)	Median (%)	Skew.	Kurt.	Autocorr.
Hedge funds	9	7	−4	5	9	−0.03	4.05	0.21
Brokers	10	32	−20	21	10	−0.11	3.13	−0.01
Banks	14	22	−14	15	15	−0.12	3.18	−0.12
Insurers	12	24	−17	16	14	−0.19	3.81	0.02
January 2006–December 2008								
	Mean (%)	SD (%)	Min (%)	Max (%)	Median (%)	Skew.	Kurt.	Autocorr.
Hedge funds	1	13	−12	5	10	−1.00	5.09	0.26
Brokers	−5	40	−33	27	6	−0.52	4.69	0.16
Banks	−24	37	−34	22	−8	−0.57	5.18	0.05
Insurers	−15	39	−40	28	1	−0.84	8.11	0.07

FIG. 1: Summary Table from [1]

My Summary

TABLE I: Full sample

type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev	Skew	kurtosis	autocorrelation
banks	-9.70	-0.42	0.16	0.13	0.70	25	1.36	2.99	56.72	-0.02
brokers	-10	-0.54	0.17	0.18	0.82	35	1.47	1.92	22.51	0
hedge funds	-7.70	-0.05	0.09	0.09	0.22	5.70	0.40	0.46	25.58	0.14
insurers	-10	-0.38	0.15	0.15	0.65	20	1.28	2.61	45.88	-0.05

TABLE II: January 1994-December 1996

type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev	Skew	kurtosis	autocorrelation
banks	-2.30	-0.23	0.27	0.29	0.80	5.30	0.82	0.31	1.87	0.01
brokers	-4.20	-0.40	0.18	0.23	0.82	5.60	1.01	0.25	1.46	-0.07
hedge funds	-2.10	-0.04	0.11	0.12	0.25	3.20	0.42	0.99	10.12	0.21
insurers	-3.10	-0.27	0.16	0.20	0.67	5.80	0.82	0.66	3.91	-0.09

TABLE III: January 1996-December 1998

type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev	Skew	kurtosis	autocorrelation
banks	-4.40	-0.20	0.35	0.34	0.91	5.60	1.04	-0.18	2.83	-0.09
brokers	-10	-0.42	0.24	0.31	1.10	20	1.67	2.87	40.02	0.01
hedge funds	-3.20	0.004	0.12	0.13	0.28	3.30	0.50	0.43	9.78	0.10
insurers	-7.90	-0.35	0.24	0.24	0.86	8.30	1.08	-0.23	8.86	-0.01

TABLE IV: January 1999-December 2001

type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev	Skew	kurtosis	autocorrelation
banks	-6.10	-0.61	0.06	0.13	0.83	5.90	1.21	0.27	2.24	-0.06
brokers	-5.50	-0.93	0.08	0.28	1.20	19	2.42	2.18	11.49	-0.01
hedge funds	-2.80	0.002	0.10	0.12	0.26	3.50	0.53	0.80	8.37	0.16
insurers	-7.10	-0.77	0.01	0.10	0.75	10	1.46	0.82	5.05	-0.14

TABLE V: January 2002-December 2004

type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev	Skew	kurtosis	autocorrelation
banks	-3.20	-0.30	0.17	0.14	0.60	3.60	0.78	-0.15	1.26	-0.09
brokers	-3.30	-0.62	0.14	0.10	0.77	4.90	1.14	0.11	1.04	0.04
hedge funds	-1.60	0.004	0.07	0.09	0.16	2	0.26	0.87	12.06	0.21
insurers	-5.90	-0.35	0.12	0.12	0.61	3.30	0.87	-0.52	4.16	0.03

TABLE VI: January 2006-December 2008

type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev	Skew	kurtosis	autocorrelation
banks	-9.40	-0.68	-0.07	-0.24	0.41	10	1.40	-1.05	10.19	0.07
brokers	-9	-0.71	0.09	-0.05	0.72	14	1.48	0.25	12.96	0.17
hedge funds	-2.50	-0.13	0.07	0.03	0.22	3.20	0.42	-0.44	7.62	0.25
insurers	-10	-0.51	0.02	-0.15	0.46	11	1.46	-0.58	17.09	0.12

III. IN-SAMPLE RESULTS

A. GARCH

I start with reproducing Garch (1,1) results. For this I calculate mean of all institutions for each rolling 36 month window and apply Garch (1,1) model to predict variance of the system.

First of all, residuals do not look normal. The p-value of the null that residuals are normally distributed is very small. However, closer examination gives insight that it is mainly because of 2008 where returns were abnormal which skewed the results. Billio did not use log transformation. I tried doing log transformation but this did not affect the results. The Box-Ljung test's p-value is well above zero which means that we cannot reject the null that squared residuals do not exhibit serial correlations after applying Garch.

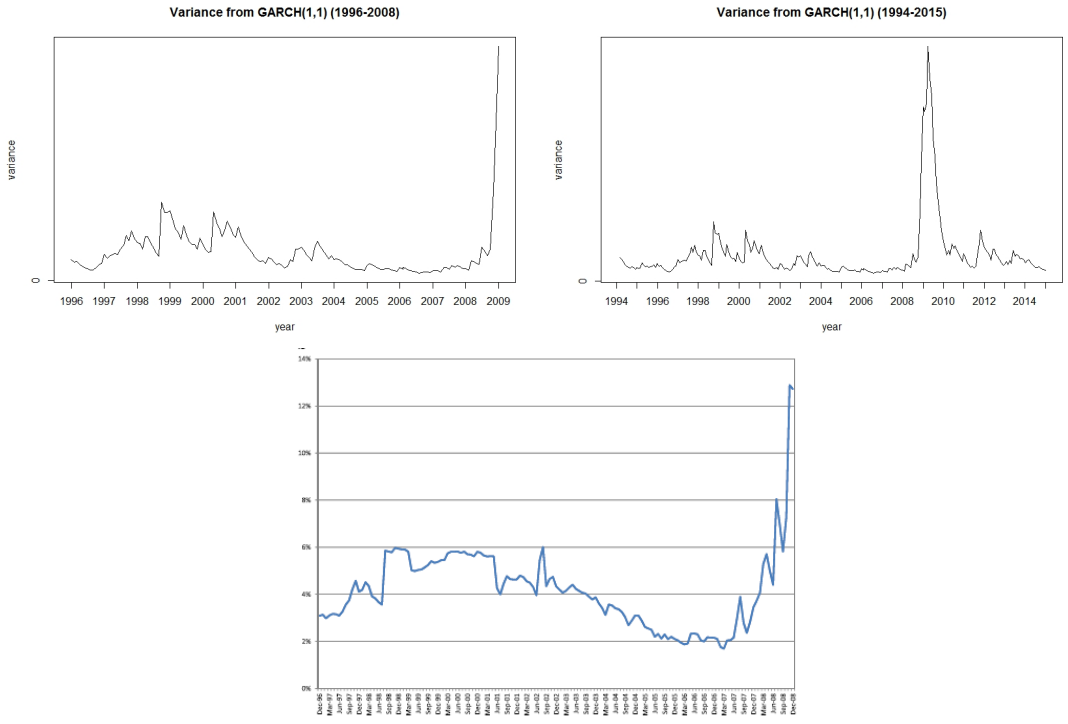

FIG. 2: Results from GARCH estimation for sample similar to [1] (left), for full sample(right) and from [1] (bottom)

B. PCA

I reproduce PCA results calculating variance explained by first principal components. For each 36 month rolling window I transform the data into lower dimensional space. Eigenvalues of the matrix represent variance of the data after transformation (they are matched with vectors representing principal components). Using all 36 principal components would mean that there is no loss of information in the transformed data and therefore we can calculate proportion of variance left relative to the case of 36 components.

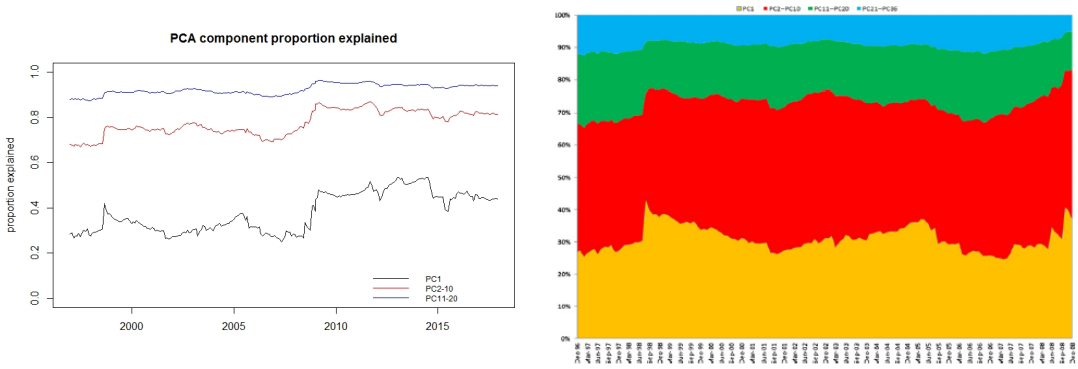

FIG. 3: Proportion explained by principal components from PCA

C. Granger Causality Graph

Time-series i Granger causes time-series j if adding lags of i to lags of j improves prediction of j. Billio et. al use BIC criterion to specify number of lags included in the model. For simplicity, I just used the lag of size 1. The density of resulting graph serves as a proxy of interconnectedness of the system. The resulting graphs for two time periods are plotted on figures 4 and 5

IV. OUT OF SAMPLE RESULTS

V. EXTENSION

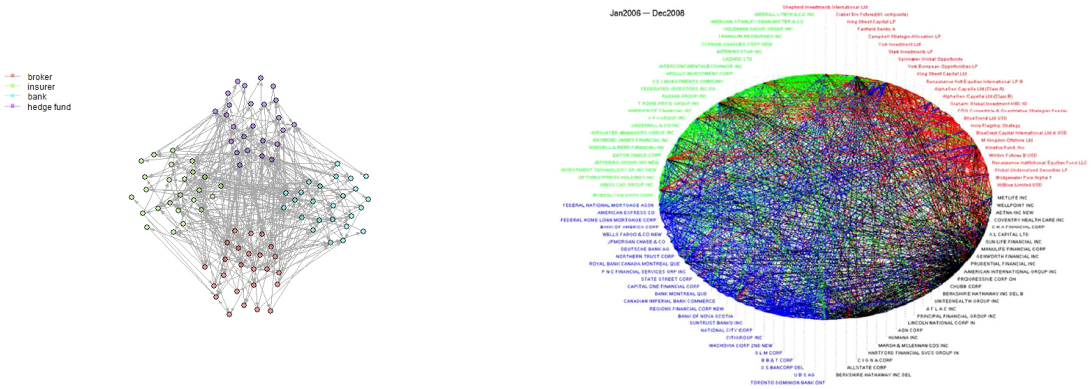

FIG. 4: Graph for 1994-1996. Mine on the left, theirs on the right

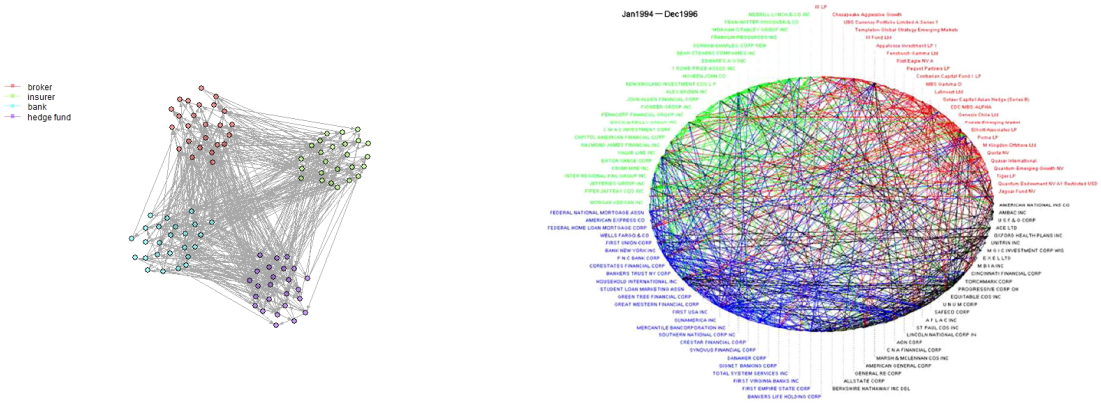

FIG. 5: Graph for 2006-2008. Mine on the left, theirs on the right

APPENDIX: WHAT I DID

1. Data

I used EQS function in Bloomberg to screen for securities. File "returns.xlsx" has all the tickers from Bloomberg with market capitalizations at 7 different points at time. For speed, I used Python to filter top 25 for each period in time (file "clean.ipynb"). At the end I produced file "picks.xlsx" which summarizes whether given ticker was included in the top 25 for a given period. I also saved into "backup" folder csv files with top 25 in each category across the time (by using number of non-null observations as sorting mechanism and average market cap if number of tickers with non-null observations is higher than 25).

All data cleaning related to hedge funds is in folder "hedge fund" . I was given files "active_info.csv", "active_returns.csv", "dead funds_info.csv", "dead funds_returns.csv". I selected hedge funds conditioning on having at least 140 returns in a 180 month period and maximum average NAV (file "selection.csv") and "hedge funds.csv".

Files that were shared by Sherman are in the folder "FilesFromSherman". To recreate hedge fund data I used Python for speed, (Python file is hedge fund/newattempt/hedge fund data.ipynb, output file is in hedge fund/newattempt/hf_result.xlsx). The output files of data extension are in "shared/new". Code to produce this was is in "hedge fund/new attempt/get_new_data.ipynb".

---

[1] Billio, Getmansky, Lo and Pelizzon, Journal of Financial Economics, 104 (2012)