Learning Connections in Financial Time Series

Gartheeban G*, John Guttag, Andrew Lo

CSAIL
MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

We propose a method for learning the **connections between equities** focusing on **large losses** and exploiting this knowledge in portfolio construction.

Problem Statement

- We refer to the large losses as events.
- Given a set of equities A, some of which had events and some of which didn't, learn which equities in a disjoint set B, are mostly to experience an event on the same day?
- Use this learned relationship to construct a portfolio containing assets that are less likely to have correlated events in the future.

Background

- Correlation measures give equal weight to small and large returns, and therefore the differential impact of large returns may be hidden.
  - **Conditional Correlation:** *Starica, C. Multivariate extremes for models with constant conditional correlations. Journal of Empirical Finance. 1999.*

- But conditional correlation of multivariate normal returns will always be less than the true correlation.
  - **Extreme value theory**: *Longin, F. and Solnik, B. Correlation structure of international equity markets during extremely volatile periods. 1999.*
  - **Semi-parametric methods**: *Boldi, M O and Davison, A C. A mixture model for multivariate extremes. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2007.*

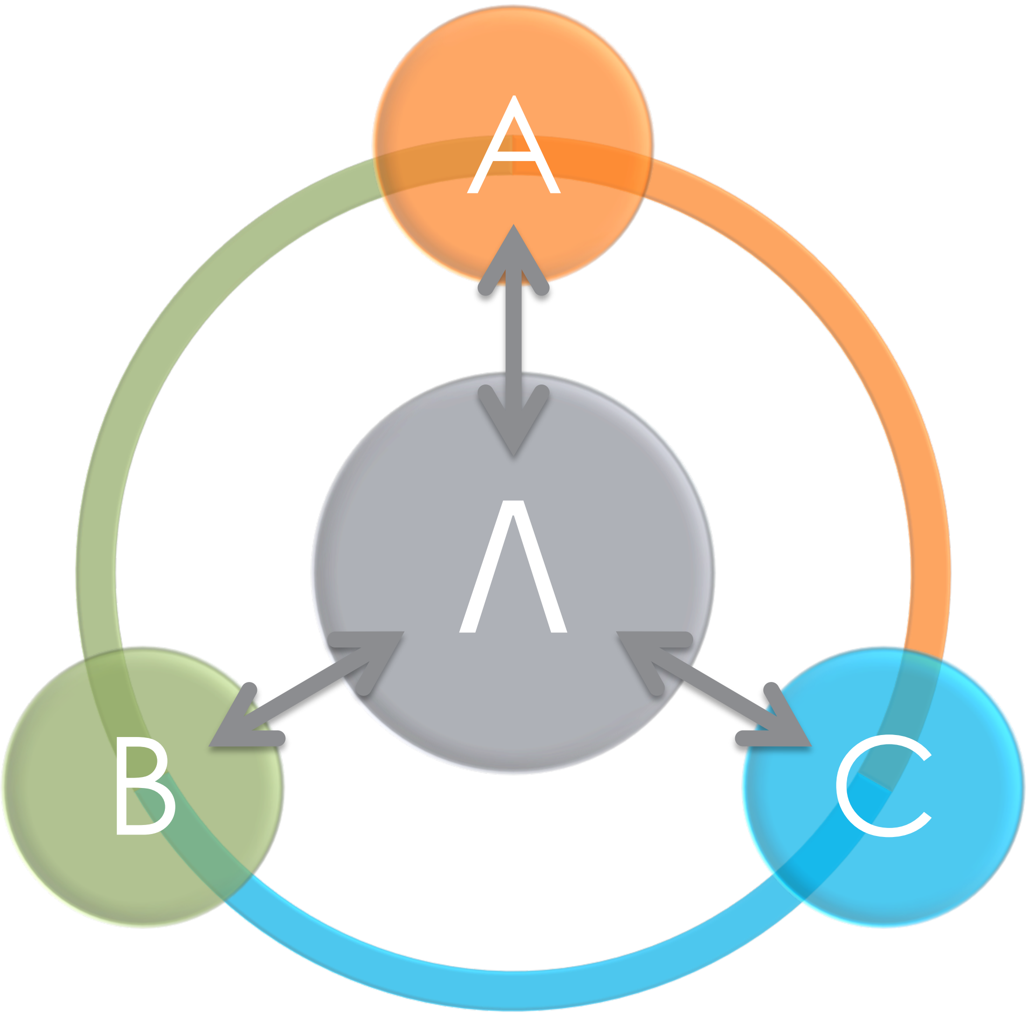
Background

- A downside of these methods is that the linkage is learned independently for pairs of time series. Other time series might be lurking variables.
  - **Partial correlation:** *Kendall, M.G. and Stuart, A. The Advanced Theory of Statistics. Vol. 2: Inference and: Relationship. 1973.*

Related Work

- Transmission of financial shocks by measuring by measuring the linkages on extreme returns
  - *Bae, K H. A New Approach to Measuring Financial Contagion. Review of Financial Studies. 2003*

- Linkages on multivariate extreme values
  - *Coles, S.G. and Tawn, J.A. Modelling extreme multivariate events. Journal of the Royal Statistical Society. Series B. 1991.*


A
Λ
B
C

Method

Active return

Market Dependency

Connectedness with other equities

$$\hat{r}_{t,k} = \underbrace{a_k + b_k r_{t,\Lambda}}_{d_{t,k}} + \sum_{j=1:m;j\neq k} w_{j,k}(r_{t,j} - d_{t,j})$$

Daily return of stock *k* on day *t*

Method

Active return

Market Dependency

Connectedness with other equities

$$\hat{r}_{t,k} = \underbrace{a_k + b_k r_{t,\Lambda}}_{d_{t,k}} + \sum_{j=1:m;j\neq k} w_{j,k}(r_{t,j} - d_{t,j})$$

Daily return of stock *k* on day *t*

Cost function

Regularization factor

$$\min_{a*,b*,w*} \sum_{\substack{t=1:T\\k=1:m}} f(r_{t,k})(r_{t,k}-\hat{r}_{t,k})^2+\lambda(a_k^2+b_k^2+|w|^2)$$

- Cost function that weights the tails more
- E.g., $f(x) = e^{-x/0.05}$

Connectedness Matrix

- From the interpolation weights, we built connectedness matrix **G**.

- Such a matrix should be positive semi-definite to be used in portfolio optimization involving quadratic programming.
  - Any positive semi-definite matrix **G** can be decomposed into $\mathbf{P}\mathbf{P}^{T}$, where $\mathbf{P} \in \mathbf{R}^{m \times m}$

Connectedness Matrix

Active return

Market Dependency

Connectedness with other equities

$$\hat{r}_{t,k} = \underbrace{a_k + b_k r_{t,\Lambda}}_{d_{t,k}} + \sum_{j=1:m;j\neq k} w_{j,k}(r_{t,j} - d_{t,j})$$

Daily return of stock k on day t

$$\hat{r}_{t,k} = \underbrace{a_k + \sum_{v} P_{k,v} P_{\Lambda,v} r_{t,\Lambda}}_{d_{t,k}} + \sum_{\substack{j\neq k \\ v}} P_{k,v} P_{j,v}(r_{t,j} - d_{t,j})$$

$$\hat{r}_{t,k} = \underbrace{a_k + \sum_{v} P_{k,v} P_{\Lambda,v} r_{t,\Lambda}}_{d_{t,k}} + \sum_{\substack{j \neq k \\ v}} P_{k,v} P_{j,v} (r_{t,j} - d_{t,j})$$

$$\tau_{t,j} = r_{t,j} - d_{t,j}$$

$$P_{k,v} \leftarrow P_{k,v} + \eta(e_{t,k}(P_{\Lambda,v} \cdot r_{t,\Lambda} + \sum_{j \neq k} P_{j,v} \tau_{t,j}) - \lambda \cdot P_{k,v})$$

$$P_{j,v} \leftarrow P_{j,v} + \eta(e_{t,k} \cdot P_{k,v} \tau_{t,j} - \lambda \cdot P_{j,v})$$

$$P_{\Lambda,v} \leftarrow P_{\Lambda,v} + \eta(e_{t,k} \cdot P_{k,v} \cdot r_{t,\Lambda} - \lambda \cdot P_{\Lambda,v})$$

$$a_k \leftarrow \alpha_k + \eta(e_{t,k} - \lambda \cdot \alpha_k)$$

Experiments

- We use daily return data from CRSP.
- Consider 369 companies that were in the S&P500 from 2000 to 2011.
- Build Markowitz portfolios for each sector.
- Compare the connectedness matrix built using our method (FAC) against covariance (COV) estimated on historical returns.
- Distribute the capital equally across sectors.

Top-K Ranking

- Given all returns for days *1* to *T*, and returns on day *T+1* for equities in *A*, we predict which equities from *B* will have events (losses greater than 10%) on that day.

- Rank equities from *B* according to their likelihoods of having events on day ***T+1***.
  - Evaluate using Mean Average Precision (MAP).
  - In our experiments, we randomly pick ***80%*** of the equities for set *A*, and the rest for set *B*, and repeat the process *100* times.

Top-K Ranking Results

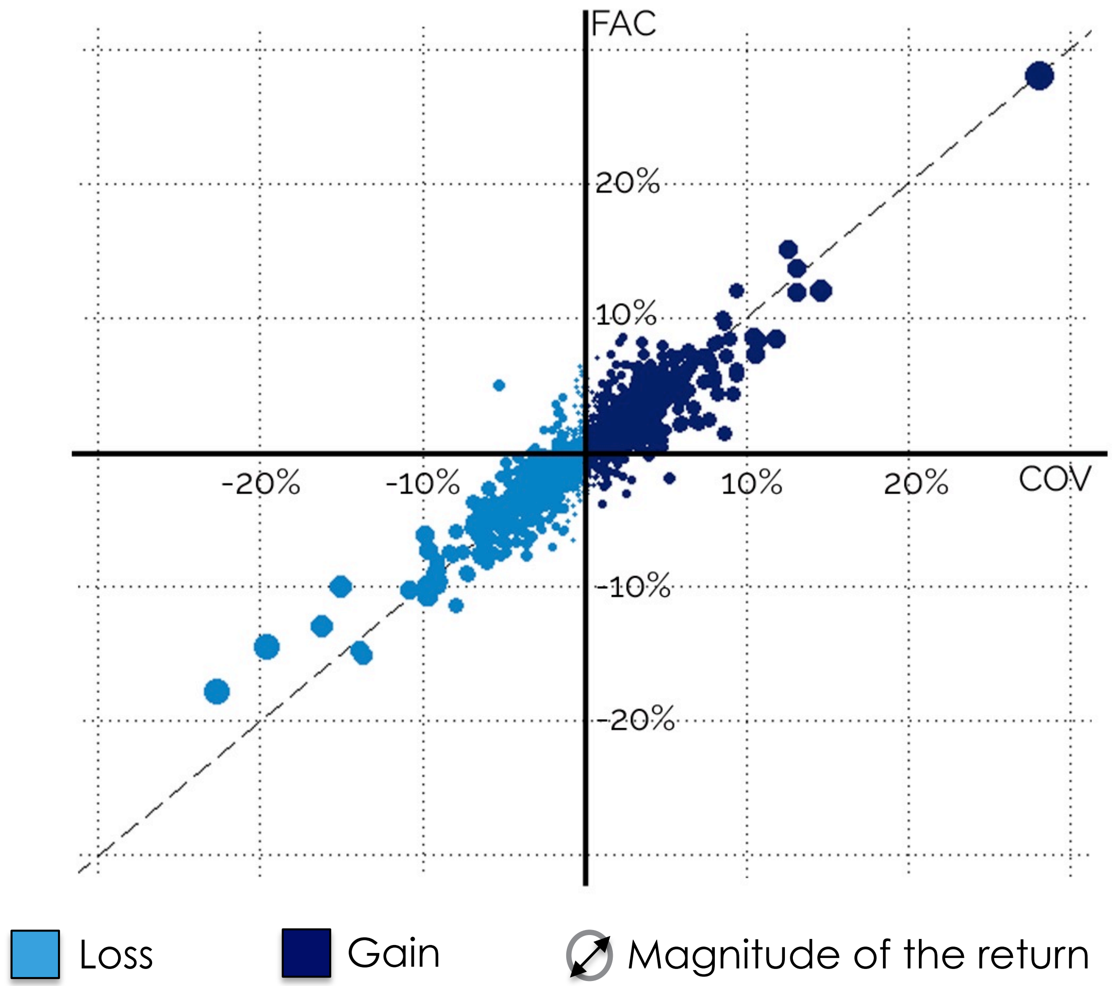
Sectors	Ours (FAC)	Correlation (CR)	Partial Correlation (PCR)	Extreme Value Correlation(EVCR)
Consumer Discrt.	**0.72**±0.082	0.30±0.075	0.45±0.114	0.34±0.101
Energy	**0.81**±0.044	0.62±0.073	0.71±0.073	0.86±0.081
Financials	**0.74**±0.051	0.44±0.055	0.62±0.062	0.65±0.114
Health Care	**0.78**±0.161	0.33±0.144	0.58±0.212	0.27±0.073
Industrials	**0.81**±0.087	0.33±0.095	0.56±0.112	0.26±0.067
Information Tech.	**0.61**±0.054	0.41±0.057	0.52±0.049	0.42±0.071
Materials	**0.91**±0.089	0.70±0.105	0.84±0.215	0.73±0.195

- results are averaged over 100 runs
- the larger the better

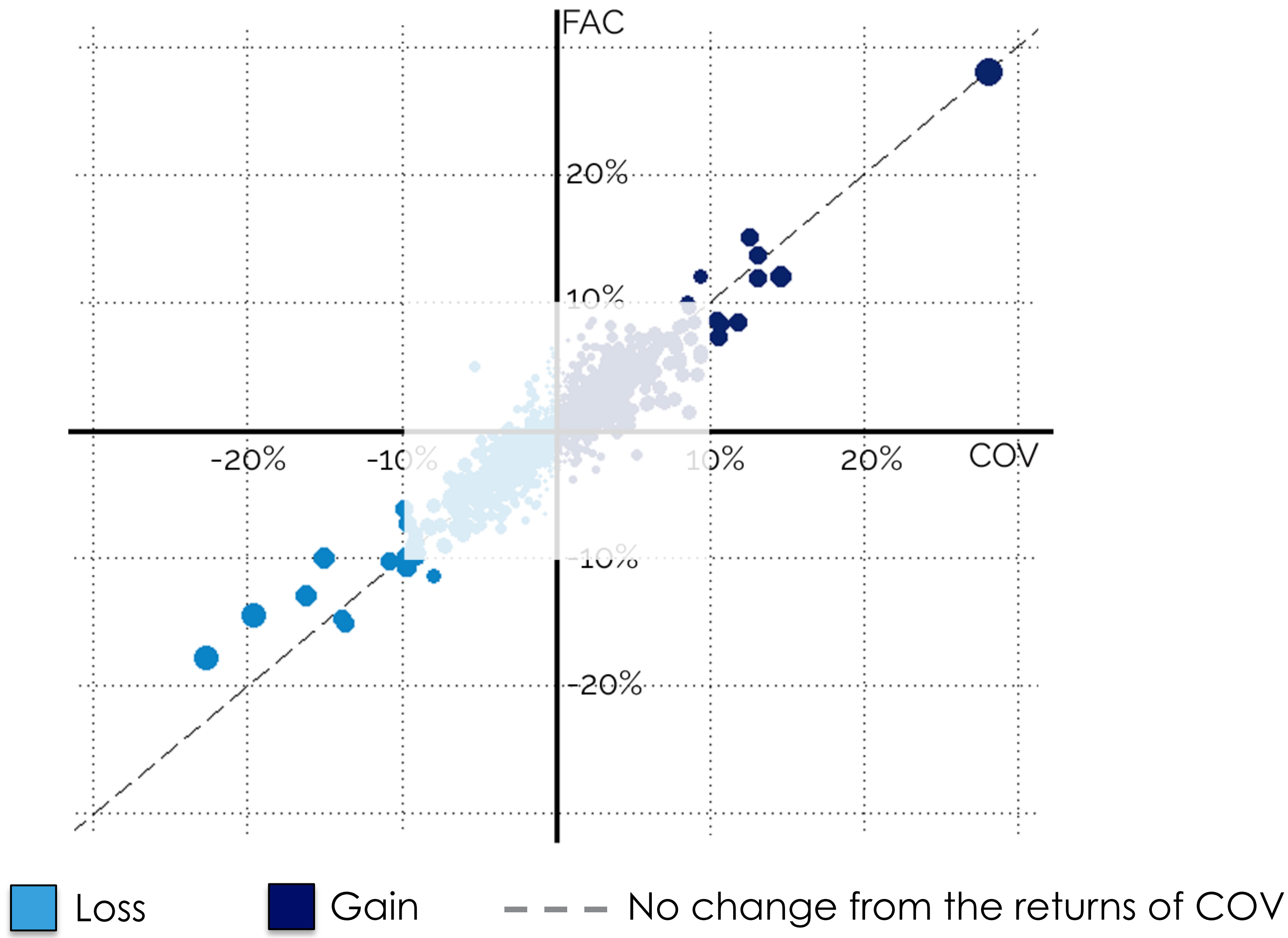
Portfolio Construction

- Markowitz portfolio for each sector, from 2001 to 2011.
- Baseline is an MVP portfolio (COV) built using the estimated covariance matrix.
- FAC is the portfolio built using connectedness matrix G that was learnt in our model.

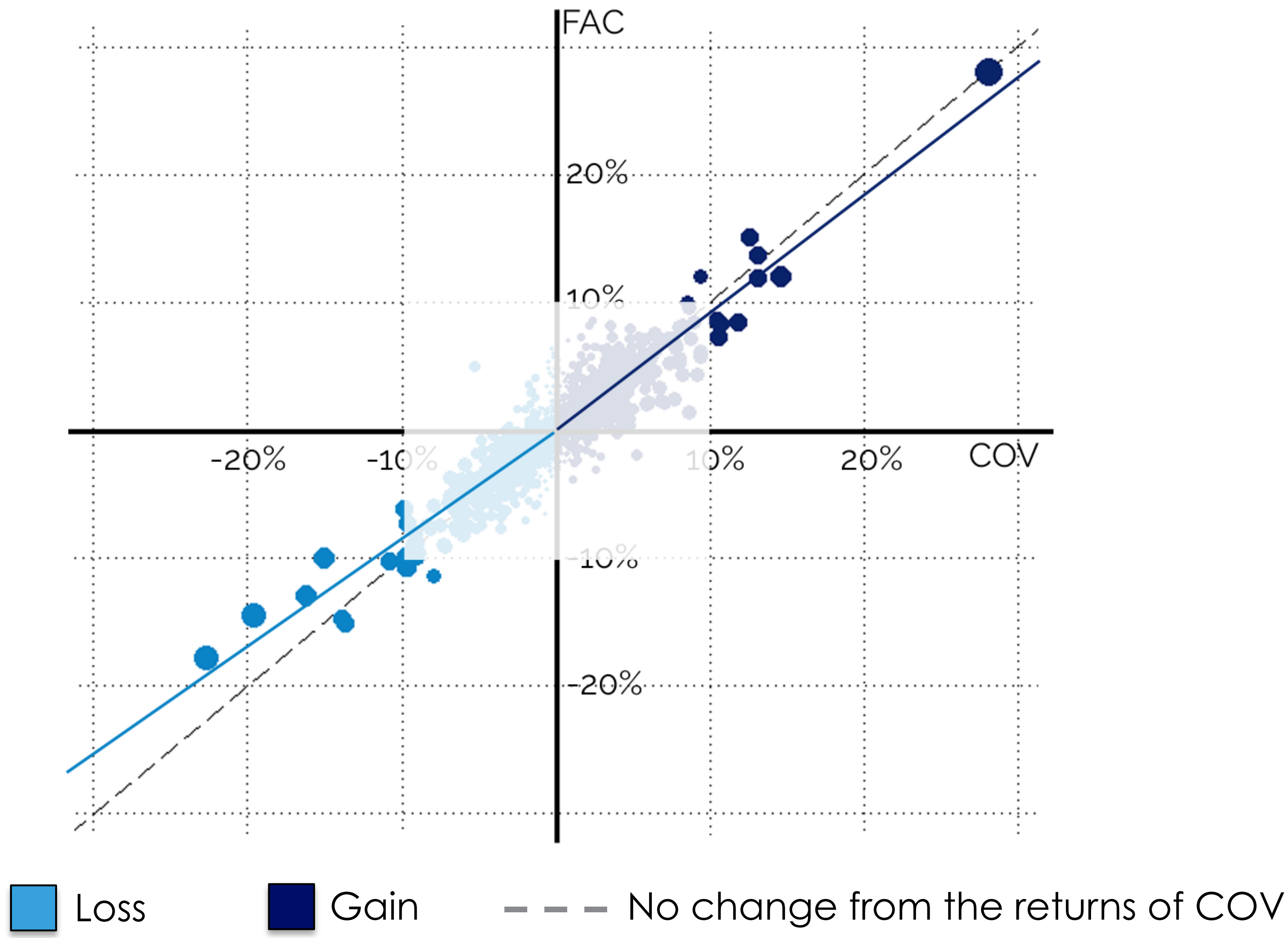
Daily Returns – Energy Sector



Daily Returns – Energy Sector



Daily Returns – Energy Sector



Cumulative Returns – Energy Sector

Market Wide Results

Measures	FAC	COV	PCR	EVCR	EW	CVaR	SPX
Worst Day	**-0.04**	-0.11	-0.12	-0.09	-0.1	-0.11	-0.09
Expected Shortfall (5%)	**-0.02**	-0.03	-0.04	-0.03	-0.02	-0.04	-0.2
Max Drawdown	**-0.13**	-0.6	-0.59	-0.6	-0.6	-0.58	-1.02
Cumulative Return	**6.52**	3.68	5.29	1.89	3.2	6.36	-0.09
Sharpe Ratio	0.14	**0.09**	0.1	0.2	0.17	0.13	-

- COV,PCR, EVCR are portfolios using correlation, partial-correlation, and extreme-value correlation respectively.
- EW: Equi-weighted portfolio
- CVaR: portfolio where the optimization minimizes conditional variance at 5% level.
- SPX: S&P500 index

Learning Connections in Financial Time Series.
Proceedings of the 30th International Conference on Machine Learning (ICML-13). 2013.
Also appearing in JMLR Volume 28

This work was supported by Quanta Computers Inc