Learning Connections in Financial Time Series

Gartheeban Ganeshapillai GARTHEE@MIT.EDU
John Guttag GUTTAG@MIT.EDU
Andrew W. Lo ALO@MIT.EDU
Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA 02139 USA

Abstract

To reduce risk, investors seek assets that have high expected return and are unlikely to move in tandem. Correlation measures are generally used to quantify the connections between equities. The 2008 financial crisis, and its aftermath, demonstrated the need for a better way to quantify these connections. We present a machine learning-based method to build a connectedness matrix to address the shortcomings of correlation in capturing events such as large losses. Our method uses an unconstrained optimization to learn this matrix, while ensuring that the resulting matrix is positive semi-definite. We show that this matrix can be used to build portfolios that not only "beat the market," but also outperform optimal (i.e., minimum variance) portfolios.

1. Introduction

It is widely accepted that in designing a portfolio of equities there is a tradeoff to be made between risk and return. The root cause of the tradeoff is that more volatile equities typically have higher returns. Much of modern financial theory is based upon the assumption that a portfolio containing a diversified set of equities can be used to control risk while achieving a good rate of return. The basic idea is to choose equities that have high expected returns, but are likely to move down in tandem.

Different investors have different goals. Often, investors begin by choosing a minimum desired expected return as the independent variable. They then formulate portfolio design as an optimization problem with return as a constraint and variance minimization as the objective function. Central to this optimization is the construction of a covariance matrix for the daily returns of the equities in the portfolio.

A problem with this approach is that the covariance matrix uses correlation, which gives equal weight to positive and negative returns and to small and large returns. This is inappropriate in a world in which risk preference plays an increasingly important role. Some investors, for example, hedge fund managers, expect high returns, and in exchange, expect to bear corresponding risks. For such investors, it is critical to control for *tail risk*, the risk of an improbable but potentially catastrophic negative return (Bae, 2003; Forbes & Rigobon, 2002). Hence, selective as opposed to full-cover hedging is gaining popularity (Stulz, 2005). Learning the connectedness between equities in terms of large losses and exploiting this knowledge in portfolio construction is the topic of this paper. We refer to these large losses as *events*.

We formulate the learning problem as given a set of equities A, some of which had events and some of which didn't, which equities in a disjoint set B, are mostly to experience an event on the same day. It may seem that this is useless, because by the time we have the returns for equities in A, we would already know the returns for equities in B. However, the goal of this phase is not to learn to predict events, but to learn historical relationships among equities. This learned relationship will then be used to construct a portfolio containing assets that are less likely to have correlated events in the future.

We apply our method to the daily returns for all 369 companies listed in S&P500 as of Jan 1, 2012 that were publicly traded from Jan 1, 2000 through December 31, 2011.

We use a factor model to describe the daily return of each equity in terms of the equity's active return, market sensitivity, and the daily returns of other eq-

Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28.

uities in the sector (e.g., financial or energy) to which that equity belongs. We then train a regression model on the historical data using regularized least squares and estimate the parameters using gradient descent. In contrast to methods that quantify the connectedness between equities using pairwise relationships, our method accounts for interactions with all other equities in the portfolio. Since extreme events are rare, we use all of the historical data rather than just events. We use a cost function that differentially weights returns of different magnitudes.

In our model, we update the weights daily and predict the returns for the following day. We rank the equities in B using the predicted returns. We compare this list against the true occurrences of events using mean average precision (MAP) scores. Using this metric, our approach consistently outperforms the most popular techniques in the financial literature, e.g., t-copula (Nelsen, 2006).

Our experiments provide strong evidence that by exploiting these learned relationships we can build portfolios that outperform portfolios constructed using techniques drawn from the literature. The comparison is done using minimum daily return, total cumulative return, maximum drawdown, and the Sharpe ratio (Sharpe, 1994).

We make the following contributions in this paper.

- A method of modeling returns using three factors: active return a, market sensitivity b, and the connectedness of equities $w_{j,k}$.
- An alternative to the usual approach of using a correlation matrix to represent relationships among equities. Instead, we use what we refer to as a connectedness matrix. This differs from a correlation matrix in two important ways:
    1. Traditional correlation matrices do not account for interactions among neighbors. Specifically, correlation is calculated between i and j independently of other neighbors. Therefore, these methods may end up incorporating information provided by neighbors multiple times. Our method uses supervised learning to discover the connections between two entities while discounting the influence of others. We use regularization to reduce the impact of over fitting and spurious estimates.
    2. Extreme returns occur rarely and therefore play a minor role in a traditional correlation matrix. The connectedness matrix focuses on extreme returns without ignoring the non-extreme returns.
- Formulating the problem of estimating one return in terms of other returns as a recursive regression problem, and providing a method to solve it using unconstrained least squares optimization. The method ensures that the resulting connectedness matrix G is positive semi-definite.

2. Related Work

We begin by discussing general methods that have been used to study correlations among returns. We then move on to discuss work specific to understanding extreme returns.

2.1. Correlation and Partial Correlation

If one knows the correlation of equity e with all other equities, one can estimate the expected return of e as a weighted average over known returns of other equities.

Correlation measures give equal weight to small and large returns, and therefore the differential impact of large returns may be hidden. Since the absolute values of returns increase during volatile periods, unconditional correlation values also rise even when the connectedness between two equities may remain the same (Longin & Solnik, 1999). To address this, researchers have proposed conditional correlations to focus on certain segments (Stărică, 1999).

However, it has been shown that conditional correlation of multivariate normal returns will always be less than the true correlation. This effect also exists when a GARCH model generates the returns (Longin & Solnik, 1999).

Longin & Solnik (1999) provides a formal statistical method, based on extreme value theory, to model the correlation of large returns. First, they model the tails of marginal distributions using generalized Pareto distribution (GPD)(Castillo & Hadi, 1997). Then, they learn the dependence structure between two univariate distributions of extreme values. Semi-parametric models have since been proposed to address the inflexibilities of such parametric models (Boldi & Davison, 2007). A downside of these methods is that the linkage is learned between two time series independently of the rest.

Partial correlation measures the degree of association between two time series while discounting the influence of others. It is calculated by fitting a regression model for each of these two time series on the rest. The correlation between the residuals of these regression models gives the partial correlation (Kendall & Stuart, 1973). But, partial correlation doesn't distinguish extreme values.

2.2. Understanding Extreme Returns

Correlation between stocks has traditionally been used when measuring co-movements of prices, and discovering contagion in financial markets (Richards, 1995; Bae, 2003). Researchers have also used partial correlations to build correlation-based networks. These networks are then used to identify the dominant stocks that drive the correlations present among stocks (Kenett et al., 2010).

Bae (2003) distinguishes extreme returns in establishing the linkages between financial time series. They capture the transmission of financial shocks to answer questions such as how likely is it that two Latin American countries will have extreme returns on a day given that two countries in Asia have extreme returns on that or the preceding day. There has been extensive research on multivariate extreme values (Coles & Tawn, 1991; Pekasiewicz, 2007). Chen & Chihying (2007) provides a method to model the temporal sequence associations for rare events. Arnold et al. (2007) examines a host of algorithms that, loosely speaking, fall under the category of graphical Granger methods to quantify the connectedness in time series.

3. Method

If the closing prices of the equity j on day T and $T-1$ are $p_{T,j}$ and $p_{T-1,j}$, the return for equity j on day T is given by $r_{T,j} = (p_{T,j} - p_{T-1,j})/p_{T-1,j}$. On day $T+1$, we are given historical daily returns for m equities in a $T \times m$ matrix $\mathbf{R} = \{r_{t,j}\}; 1 \le t \le T, 1 \le j \le m$. We use indexing t for days, and j, k for equities. When $r_{t,j} < -0.1$ (10% loss), we say that equity j had an event on day t.

We assume that daily returns (rows of $\mathbf{R}$) are independent. While daily returns are generally believed to be heteroskedastic (White, 1980), since we focus on large returns that are rare, we can safely assume that the modeling errors are uncorrelated. We use regularization to tackle over fitting. The regularization parameter λ is determined by cross validation.

Factor model representation of returns is common in finance and econometrics (Longin & Solnik, 1999; Khandani & Lo, 2007).

We model the return of equity k on day t by

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

In this model, we explicitly learn the factors for equity k: equity's active return a_k, equity's sensitivity to the market b_k, and equity's connectedness with other equities $w_{j,k}$, where $1 \le j \le m; j \neq k$. The S&P500 index return $(r_{t,\Lambda})$ averages the returns of all the equities on a given day.

Our focus is on negative returns, which are not normally distributed. Therefore, we apply the Box-Cox transformation to make the daily return samples more normal distribution-like (Box & Cox, 1964).

We use least squares minimization to estimate the weights. We find that better performance is achieved, when we capture the differential impact of certain values by weighting with a cost function $f(x)$.

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

The flexibility in choosing the cost functions allows us to optimize different aspects of the problem. For top-k ranking evaluation, we define events at or below -10% daily return. We use $f(x) = e^{-(x-r_\theta)^2}; r_\theta = -0.1$ to achieve higher accuracy because the maximum ambiguity is at the boundary. For portfolio construction problems, we consider all negative returns, and use a different cost function

We can efficiently compute model parameters ($\theta = \{(a_k, b_k, w_{j,k} | 1 \le k \le m, 1 \le j \le m, j \neq k\}$) by estimating the inner products. However, estimating the weights directly on the observed data is prone to overfitting (Bell & Koren, 2007). Therefore, we learn the parameters by solving the following regularized least squares problem:

$$\min_{a*,b*,w*} \sum_{\substack{t=1:T \\ k=1:m}} f(r_{t,k})(r_{t,k} - \hat{r}_{t,k})^2 + \lambda(a_k^2 + b_k^2 + |w|^2) \quad (3)$$

We use gradient descent to minimize the regularized square errors. For each $r_{t,k} \in \mathbf{R}$, we update the parameters by:

$$\begin{aligned} a_k &\leftarrow a_k + \eta(e_{t,k} - \lambda . a_k) \\ b_k &\leftarrow b_k + \eta(e_{t,k} \cdot r_{t,\Lambda} - \lambda \cdot b_k) \\ w_{j,k} &\leftarrow w_{j,k} + \eta(e_{t,k}(r_{t,j} - d_{t,j}) - \lambda \cdot w_{j,k}) \ \forall j \neq k \end{aligned}$$

Here, η is the learning rate that is dynamically adjusted using line search, and $e_{t,k} \stackrel{def}{=} f(r_{t,k})(r_{t,k} - \hat{r}_{t,k})$. We use the last 500 days in the historical data to train our model. We iterate 100 times for the initial estimate of the parameters. The model is updated daily to make predictions for the next day. Since this new training set differs from the previous for only two days, convergence is achieved within a few iterations.

3.1. Connectedness Matrix

In portfolio construction, connectedness between equities is used to find less correlated assets. Generally, a covariance matrix C is used to find the optimal diversification, i.e., minimum variance portfolio. In contrast, our model uses the connectedness matrix G, which is learned using the factor model. We assume that the portfolio is built with m equities that belong to a sector and the S&P500 index (SPX).

Table 1. Connectedness matrix construction

	1	...	j	...	m	Λ
1	σ_1^2	...	$w_{1,j}$	...	$w_{1,m}$	b_1
...	...	...	...	...	...	...
j	$w_{j,1}$	...	σ_j^2	...	$w_{j,m}$	b_j
...	...	...	...	...	...	...
m	$w_{m,1}$	...	$w_{m,j}$	...	σ_m^2	b_m
Λ	b_1	...	b_j	...	b_m	σ_Λ^2

We build the connectedness matrix $G \in \Re^{(m+1)\times(m+1)}$ from the interpolation weights of the factor model, and demonstrate that it provides better diversification. A direct construction is given in Table 1. Here, σ_j^2 is the variance of the daily returns of the equity j. Λ denotes SPX.

Such a matrix should be positive semi-definite to be used in portfolio optimization involving quadratic programming. Since any positive semi-definite matrix G can be decomposed into PP^T, where $P \in \Re^{(m+1)\times(m+1)}$, we reformulate Equations 1 and 3 as:

$$\hat{r}_{t,k} = \underbrace{a_k + \sum_v P_{k,v} P_{\Lambda,v} r_{t,\Lambda} + \sum_{\substack{j\neq k\\ v}} P_{k,v} P_{j,v}(r_{t,j} - d_{t,j})}_{d_{t,k}} \tag{4a}$$

$$\min_{a*,P*} \sum_{\substack{t=1:T\\ k=1:m}} f(r_{t,k})(r_{t,k} - \hat{r}_{t,k})^2 + \lambda(a_k^2 + |P|^2) \tag{4b}$$

We begin the training by initializing P to $\tilde{P}$, where $\tilde{P}$ is the Cholesky factorization of the covariance matrix C, i.e., $C = \tilde{P}\tilde{P}^T$. We compute the covariance matrix C on historical data. For each $r_{t,k} \in \mathbf{R}$, we update P and a_k by moving against the gradient. For $1 \leq j \leq m; j \neq k$, and $1 \leq v \leq m$, the updates are:

$$\begin{aligned}
\tau_{t,j} &= r_{t,j} - d_{t,j} \\
P_{k,v} &\leftarrow P_{k,v} + \eta(e_{t,k}(P_{\Lambda,v} \cdot r_{t,\Lambda} + \sum_{j\neq k} P_{j,v}\tau_{t,j}) - \lambda \cdot P_{k,v}) \\
P_{j,v} &\leftarrow P_{j,v} + \eta(e_{t,k} \cdot P_{k,v}\tau_{t,j} - \lambda \cdot P_{j,v}) \\
P_{\Lambda,v} &\leftarrow P_{\Lambda,v} + \eta(e_{t,k} \cdot P_{k,v} \cdot r_{t,\Lambda} - \lambda \cdot P_{\Lambda,v}) \\
a_k &\leftarrow \alpha_k + \eta(e_{t,k} - \lambda \cdot \alpha_k)
\end{aligned}$$

3.2. Discussion

In our model (Equation 1), we represent the relationship between the returns of equities after discounting their interactions with the market. Thus, interpolation weights $w_{j,k}$ resemble partial correlation estimates. In the factor model, we simultaneously fit the regression model and learn the correlation weights. Further, regularization is employed in our model to reduce the likelihood of spurious estimates.

For a matrix X, a column vector Y, and a regression problem expressed as $Xw = Y$, an explicit solution, denoted by $\hat{w}$ is given by: $\hat{w} = (X^T X)^{-1} X^T Y$. If the variables are mean adjusted, Σ is the covariance of X, and C is the covariance between Y and each column of X, it can be rewritten as, $\hat{w} = \Sigma^{-1} C$. In Equation 1, we can observe the similarity between these variables (X and Y) and adjusted returns, i.e., $y_t \approx (r_{t,k} - d_{t,k})$, and $x_{t,j} \approx (r_{t,j} - d_{t,j})$. The connectedness matrix G, built from the interpolation weights, models the pairwise connection between the adjusted returns of two equities while discounting the connectedness among all other equities.

4. Results

Though univariate time series of daily equity returns lack both significant autocorrelation and stationarity, multivariate time series of returns exhibit consistent correlation that persist over time. This persistent correlation is what makes portfolio diversification possible (Borodin et al., 2004; Kalai & Vempala, 2000; Kenett et al., 2010).

4.1. Data

We use daily return data from CRSP[1]. We examine all 369 companies that were in the S&P500 from 2000 to 2011. This time period contains two major financial crises (2001 and 2008). The set of companies are

[1]CRSP, Center for Research in Security Prices. Graduate School of Business, The University of Chicago (2004). Used with permission. All rights reserved. www.crsp.uchicago.edu

Table 2. MAP scores for different methods.

Sectors	FAC	CR	PCR	EVCR
Consumer Discretionary	**0.72** $\pm$0.082	0.30 $\pm$0.075	0.45 $\pm$0.114	0.34 $\pm$0.101
Energy	**0.81** $\pm$0.044	0.62 $\pm$0.073	0.71 $\pm$0.073	0.68 $\pm$0.081
Financials	**0.74** $\pm$0.051	0.44 $\pm$0.055	0.62 $\pm$0.062	0.65 $\pm$0.114
Health Care	**0.78** $\pm$0.161	0.33 $\pm$0.144	0.58 $\pm$0.212	0.27 $\pm$0.073
Industrials	**0.81** $\pm$0.087	0.33 $\pm$0.095	0.56 $\pm$0.112	0.26 $\pm$0.067
Information Technology	**0.61** $\pm$0.054	0.41 $\pm$0.057	0.52 $\pm$0.049	0.42 $\pm$0.071
Materials	**0.91** $\pm$0.089	0.70 $\pm$0.105	0.84 $\pm$0.215	0.73 $\pm$0.195

from ten sectors: consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, telecommunications services, and utilities.

4.2. 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. We produce an ordered list of equities from B, ranked by their likelihoods of having events on day $T+1$ based on their predicted returns $\hat{r}_{T+1}$.

4.2.1. Evaluation

We use mean average precision (MAP) to evaluate the correctness of the ordered list. Average precision (AP) is a popular measure that is used to evaluate an ordered list while taking into account both recall and precision (Zhu, 2004). MAP is the mean of average precision across the test set. For an ordered list of top-k items, MAP and AP are given by:

$$AP(k) = \sum_{j=1}^{k} p(j)\Delta r(j) \quad (5)$$

$$MAP(k) = \sum_{i=1}^{n} AP_i(k)/n \quad (6)$$

Here, n is the size of the test set, $p(j)$ is the precision at cut-off j, and $\Delta r(j)$ is the change in the recall from $j-1$ to j. We produce a top-10 list, and evaluate with $MAP(10)$.

4.2.2. Experimental Results

Since diversification inevitably involves using equities from multiple sectors, we focused on the question of which equities to own within sectors. We randomly select 20% of the companies in each sector for set A, and the rest for set B. We ran our experiments from 2001 to 2011, so that at the start of the experiment we will have at least a year of historical data to train on. We evaluated our methods only on days that had at least two events. Three sectors had less than 5 such days in the last decade, and therefore were excluded in the experiments. Across all the sectors, there were 539 days that had two events out of 3019 days in the full dataset.

Table 2 compares the MAP scores (higher is better) for our factor model (FAC) with the scores for benchmark methods: correlation (CR), correlation of extreme values (EVCR), and partial correlation (PCR). The results are averaged over 100 runs. The best result for each sector is in bold face. Results for FAC are statistically different (p-value < 0.001) from the results of every other method under paired t-test.

For CR and PCR we use standard implementations. For EVCR, we apply a GARCH model to remove serial dependencies, if there are any. Then, we fit a univariate generalized Pareto distribution with tail fraction 0.1 on the innovations of the GARCH model. Finally, we model the multivariate distribution with a t-Copula, and learn the linear correlations on extreme values (Cherubini et al., 2004).

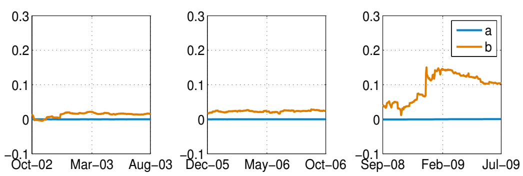
Our factor model consistently outperforms the other methods. EVCR often underperforms PCR, and at times, CR. We conjecture that the inflexibility of the parametric modeling of the tails and not considering the relationship between non-extreme values contribute to this failure. The poor performance of EVCR is striking because t-copula is widely used in financial risk assessment, especially in the pricing of collateralized debt obligations (CDO) (Meneguzzo & Vecchiato, 2004; MacKenzie, 2008).

Table 3 compares the MAP scores for different sizes of known and unknown sets. We change the size of B as a fraction of the total number of companies available, using 10%, 20%, and 40%. Set A contains the rest. Even when we train on only 60% of the companies and test on 40%, our method remains effective. Notice that FAC trained with only 60% of the data outperforms other methods trained with 80% of the data.

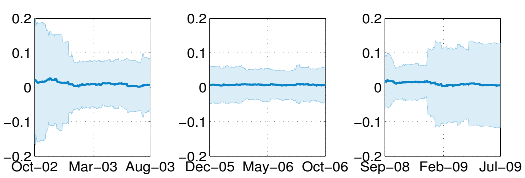
Table 3. MAP scores for different test sizes.

SECTORS	10%	20%	40%
CONSUMER DISC.	0.81	0.72	0.6
ENERGY	0.83	0.81	0.71
FINANCIALS	0.78	0.74	0.61
HEALTH CARE	0.95	0.78	0.72
INDUSTRIALS	0.86	0.81	0.52
INFORMATION TECH.	0.71	0.61	0.51
MATERIALS	0.94	0.91	0.81

As an example, we look at an S&P500 constituent, Bank of America (BAC). Between 2001 and 2011, BAC had 29 events, i.e., daily losses of at least 10%. Figure 1 shows the parameters learned using the factor model for BAC. Notice that the market dependence drastically increases during the 2008 crisis. Further, the "herding effect," as given by the spread in the weights, widens during the crises of 2001 and 2008/9. BAC becomes heavily connected to smaller number of other equities.



(a) Market dependence (b_k)



(b) Co-movements with neighbors ($w_{j,k}$)

Figure 1. Connectedness estimated by the factor model for BAC: (a) Active return a, and market sensitivity b for BAC.(b) Median and $10^{\text{th}} - 90^{\text{th}}$ percentile range of the correlation weights $w_{j,k}$ between BAC and neighbors.

4.3. Portfolio Construction

The major application of our method is the reduction of large losses in equity portfolios. Since there is a tradeoff between risk and expected return, portfolio design usually starts by the investor choosing a desired level of expected return (or a risk tolerance). For a given desired expected return r_e, in the absence of any side information, the minimum variance portfolio (MVP) is the optimal portfolio (Markowitz, 1959). For the MVP, portfolio weights ω are derived by solving the optimization problem:

$$\begin{aligned} &\min_{\omega} \frac{1}{2}\omega^T C\omega \qquad (7)\\ &\text{subject to} \sum_{j=1}^{m+1} \bar{r}_j \omega_j \geq r_e \\ &\qquad \sum_{j=1}^{m+1} \omega_j = 1; 0 \leq \omega_j \leq 1, j = 1, ..., m+1 \end{aligned}$$

Here, C is the covariance matrix of returns, and $\bar{r}_j$ is the expected return of equity j. Typically, the covariance and the expected return are calculated from historical data. Here, we assume fixed capital (no leverage), no short positions, daily readjusted portfolio weights, and we ignore the costs of transactions.

We demonstrate our method's utility by building portfolios with our connectedness matrix G (Section 3.1), and compare their performance to portfolios built using methods drawn from the financial literature.

- Our baseline is an MVP portfolio built using the estimated covariance matrix C. This is the conventional approach; we refer it as COV.

- For the factor model, we replace C with connectedness matrix G and $\bar{r}_j$ with active return a_j. We learn both G and a_j using our factor model. We use the same optimization as MVP, i.e., Equation 7.

  1. FAC_1 is the factor model with cost function $f(x) = e^{-x/0.05}$ applied. This model focuses on minimizing the co-occurrences of large losses. This risk avoidance results in smaller overall return compared to FAC_2.
  2. FAC_2 is the factor model without any cost functions. It captures the connections on large returns. It produces significantly larger overall returns, at the cost of larger worst case daily losses.

When a portfolio is heavily diversified, the expected return is smaller. Therefore, in our formulation the desired expected return r_e governs the amount of diversification. The range of achievable values for r_e is the minimum and maximum of the expected returns of the equities. Maximum expected return is achieved by owning the equity with the highest historical return. Minimum risk relative to the market is achieved by owning everything in that market.

Table 4. Portfolio Returns

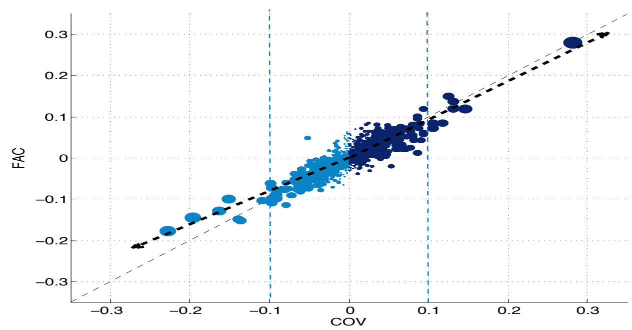
Measures		Energy			Financials			Materials		
		COV	FAC_1	FAC_2	COV	FAC_1	FAC_2	COV	FAC_1	FAC_2
Largest loss in a day	$\min(r_t)$	-0.23	**-0.16**	-0.18	-0.11	**-0.11**	-0.16	-0.19	**-0.08**	-0.12
Expected Shortfall (5%)		-0.06	**-0.05**	-0.06	-0.04	**-0.03**	-0.05	-0.06	**-0.04**	-0.05
Max Drawdown		-0.72	-0.74	**-0.66**	-0.77	**-0.69**	-0.74	-0.79	**-0.31**	-0.69
Total Cumulative Return	R_T	3.58	11.72	**13.53**	1.82	2.71	**8.13**	3.28	6.1	**7.87**
Sharpe Ratio	$\frac{E(r-r_\Lambda)}{\sigma_{(r-r_\Lambda)}}$	0.04	0.03	**0.06**	0.04	0.01	**0.06**	0.04	0.03	**0.06**

It has been shown that 90% of the maximum benefit of diversification is achieved with portfolios containing roughly 5% of the market constituents (Reilly & Brown, 2011). This led us to set r_e to the 95$^{\text{th}}$ percentile of the expected returns of the equities. This setting causes the optimization to choose about $3-5$ equities per sector (5% to 10%).

Table 4 summarizes the return characteristics for the three sectors with the most events. We re-weight the portfolio daily, and estimate the returns daily. Cumulative return R_T from day 1 to day T is given by $R_T = \prod_{t=1}^{T}(r_t+1) - 1$. Total cumulative return is the overall return from year 2001 to 2011, i.e., R_T on December 31, 2011. Maximum drawdown is the largest drop from the maximum cumulative return. The Sharpe ratio measures the excess return for additional risks taken (Sharpe, 1994). It is given by $S = E(r-r_\Lambda)/\sqrt{var(r-r_\Lambda)}$, where r is the daily return of the portfolio and r_Λ is the reference return (return on the S&P500 index). A positive Sharpe ratio implies that excess return is greater than the additional risk taken. The expected shortfall (also known as CVaR) at 5% level gives the expected return on the portfolio in the worst 5% of the cases. Table 4 shows that by learning the connectedness between equities, our portfolios cannot only beat the market (positive Sharpe Ratio), but also beat the optimal (minimum-variance) portfolios.

Figure 2 shows the impact of our method on returns in the energy sector. Until the 2008 crisis, because the energy sector remained calm, our FAC model performed comparably to COV. Note that 2001 crisis, unlike 2008 crisis, was limited to few sectors not including energy. After the collapse in May 2008, our model began learning new connectivities related to large negative returns and was able to reduce large losses (Figure 2(a)) late that year and going forward. It took about 200 days to learn the new model, but it persisted long enough to be useful. Figure 2(b) demonstrates the effectiveness of our method in making the large negative returns smaller without significantly affecting positive and small negative returns. The largest daily loss

(a) Cumulative Returns



(b) Daily Returns

Figure 2. Cumulative returns (a), and daily returns (b) of COV and FAC_1 models for the energy sector. In figure (b), positive returns with COV are highlighted with the darker color. Size of the points correspond to the absolute values of the returns with COV. Returns above the line correspond to an improvement with FAC_1. A clockwise shift that is more prominent in the negative side (lighter region) is noticeable.

dropped 30%, i.e., from 23% to 16%.

Figure 3 shows the equity weights learned using the connectedness matrix for the financials sector. Until August 2008, our factor model based portfolio consistently focused on two equities in the financial sector: BlackRock (ID 19), Inc. and Ventas, Inc. (ID 55). In the aftermath of the 2008 crisis, our model increases the diversification.

Finally, in Table 5 we demonstrate the effectiveness of our model in constructing a market wide portfolio. We build a market wide portfolio by combining the portfolios built for each sector weighted equally.

Table 5. Market wide portfolios.

MEASURES	FAC_1	FAC_2	COV	PCR	EVCR	EW	MIN-CVAR	SPX
WORST DAY	**-0.04**	-0.10	-0.11	-0.12	-0.09	-0.1	-0.11	-0.09
EXPECTED SHORTFALL(5%)	**-0.02**	-0.04	-0.03	-0.04	-0.03	-0.02	-0.04	-0.02
MAX DRAWDOWN	**-0.13**	-0.5	-0.60	-0.59	-0.6	-0.6	-0.58	-1.02
TOTAL CUMULATIVE RETURN	6.52	**10.25**	3.68	5.29	1.89	3.2	6.36	-0.09
SHARPE RATIO	0.14	**0.21**	0.09	0.1	0.02	0.17	0.13	N/A

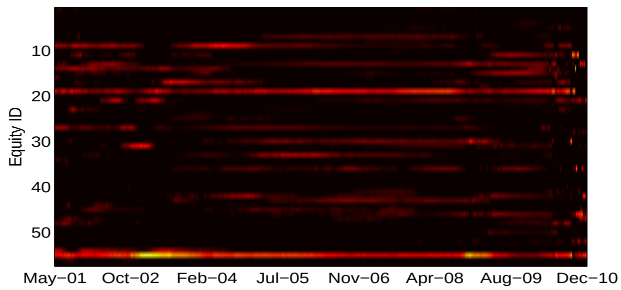


Figure 3. Heat map of the weights of the equities for the portfolios built for the financials sector.

We compare our method with COV, PCR and EVCR. We also compare these "MVP-like" portfolios with other benchmark portfolios: a equi-weighted portfolio (EW), where the portfolio is rebalanced to equal weights daily, a Min-CVaR portfolio where the optimization minimizes conditional variance at 5% level (Rockafellar & Uryasev, 2000), and the S&P500 index (SPX). FAC_2 achieves annualized sharpe ratio of $0.21 \times \sqrt{260} = 3.39$. It is significant that, in every measure, the portfolio built with our method outperforms the "hard-to-beat" equal-weighted portfolio (Plyakha et al., 2012).

4.4. Limitations

Our portfolios achieve returns several times greater than minimum-variance portfolios. Like minimum-variance portfolios, our method rebalances the portfolio daily, and therefore incurs transaction costs. These costs are ignored here, therefore absolute returns are overstated for both methods. COV leads to slightly more transactions than our method, and would therefore incur higher transaction costs.

Relationships among equities can be viewed as occurring on three time scales: permanent (e.g., sector grouping), long-term (e.g., based on fundamentals and technicals), and short-term (e.g., based on real-time news and announcements). In this work, we capture only the long-term relationships. This may result in a slower response when market conditions change rapidly as in 2008 (Figure 2). We intend to address this in future work by attempting to incorporate news and sentimental analysis into our model.

We show that by limiting the large losses, risk management can be a source of excess returns. During the 2008 crisis, all equities became heavily correlated as the market crashed, and market risk governed returns (Figure 1). Without short positions (or derivatives that simulate short positions), this kind of risk cannot be diversified away. Our current portfolio construction method (Equation 7) does not permit negative weights. We are currently working on extending our method to cover more kinds of portfolios.

5. Summary

We presented a method for learning connections between financial time series. We modeled daily returns using three factors: active return, market sensitivity, and connectedness of returns. We learned these factors using a recursive regression. We solved the regression problem using an unconstrained least squares optimization that ensures that the resulting matrix is positive semi-definite so that it can be used in portfolio construction.

We evaluated our method in two ways. First, we evaluated its accuracy in producing a list of equities ordered by their likelihoods of having large losses, given information about the behavior of other equities. We then presented and demonstrated the potential real world utility of a method that constructs portfolios using the learned relationships. The performance of portfolios constructed using our methods were compared to the performance of portfolios constructed using conventional approaches, including traditional correlation-matrix based methods. Portfolios constructed using our method not only "beat the market," but also beat the so-called "optimal portfolios."

Acknowledgments

We thank the anonymous reviewers for their helpful comments. This work was supported by Quanta Computers Inc.

References

Arnold, Andrew, Liu, Yan, and Abe, Naoki. Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD*, KDD '07, pp. 66–75. ACM, 2007.

Bae, K H. A New Approach to Measuring Financial Contagion. *Review of Financial Studies*, 16(3):717–763, July 2003.

Bell, R.M. and Koren, Y. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. *ICDM 2007. Seventh IEEE International Conference on*, pp. 43–52, 2007.

Boldi, M O and Davison, A C. A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):217–229, 2007.

Borodin, A., El-Yaniv, R., and Gogan, V. Can we learn to beat the best stock. *J. Artif. Intell. Res. (JAIR)*, 21: 579–594, 2004.

Box, G.E.P. and Cox, D.R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 211–252, 1964.

Castillo, E. and Hadi, A.S. Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association*, 92(440):1609–1620, 1997.

Chen, P. and Chihying, H. Learning causal relations in multivariate time series data. *Economics: The Open-Access, Open-Assessment E-Journal*, 1:11, 2007.

Cherubini, U., Luciano, E., and Vecchiato, W. *Copula methods in finance*. Wiley, 2004.

Coles, S.G. and Tawn, J.A. Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 377–392, 1991.

Forbes, K J and Rigobon, R. No contagion, only interdependence: measuring stock market comovements. *The Journal of Finance*, 57(5):2223–2261, 2002.

Kalai, A and Vempala, S. Efficient algorithms for universal portfolios. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pp. 486–491, 2000.

Kendall, M.G. and Stuart, A. *The Advanced Theory of Statistics. Vol. 2: Inference and: Relationsship*. Griffin, 1973.

Kenett, Dror Y, Tumminello, Michele, Madi, Asaf, Gur-Gershgoren, Gitit, Mantegna, Rosario N, and Ben-Jacob, Eshel. Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market. *PLoS ONE*, 5(12):e15032, December 2010.

Khandani, A. and Lo, A. What happened to the quants in august 2007? *Available at SSRN 1015987*, 2007.

Longin, F. and Solnik, B. Correlation structure of international equity markets during extremely volatile periods. *Available at SSRN 147848*, 1999.

MacKenzie, D. End-of-the-world trade. *London Review of Books*, 30(9):24–26, 2008.

Markowitz, H. Portfolio selection: efficient diversification of investments. *New Haven, CT: Cowles Foundation*, 94, 1959.

Meneguzzo, D. and Vecchiato, W. Copula sensitivity in collateralized debt obligations and basket default swaps. *Journal of Futures Markets*, 24(1):37–70, 2004.

Nelsen, R.B. *An introduction to copulas*. Springer, 2006.

Pekasiewicz, D. *Methods of multivariate statistical analysis and their applications*. Wydawnictwo Uniwersytetu Lodzkiego, 2007.

Plyakha, Y., Uppal, R., and Vilkov, G. Why does an equal-weighted portfolio outperform value-and price-weighted portfolios? *Available at SSRN 1787045*, 2012.

Reilly, F.K. and Brown, K.C. *Investment Analysis and Portfolio Management*. South-Western Pub, 2011.

Richards, A.J. Comovements in national stock market returns: Evidence of predictability, but not cointegration. *Journal of Monetary Economics*, 36(3):631–654, 1995.

Rockafellar, R Tyrrell and Uryasev, Stanislav. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

Sharpe, W.F. The Sharpe Ratio, 1994.

Stărică, C. Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance*, 6(5):515–553, 1999.

Stulz, R M. Rethinking risk management. *Journal of applied corporate finance*, 9(3):8–25, 2005.

White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pp. 817–838, 1980.

Zhu, M. Recall, precision and average precision. *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo*, 2004.