Forecasting and Managing Correlation Risks

Tim Bollerslev Sophia Zhengzi Li Yushan Tang
Duke University Rutgers University Nankai University

2024 AFA Annual Meeting
January, 2024

Motivation

Correlation is central to portfolio construction and risk management

- Comparing with return and volatility forecasting, less is known about correlation forecasting

Novel framework to forecast realized correlation RC via big data + ML

- **25 main features: HAR, factor-driven, EMA features** (150 additional predictors: main feature $\times$ firm-link dummy)
- **LASSO** (Ridge, ENet, PCR, NN)

Motivation

Correlation is central to portfolio construction and risk management

- Comparing with return and volatility forecasting, less is known about correlation forecasting

Novel framework to forecast realized correlation RC via big data + ML

- **25 main features: HAR, factor-driven, EMA features**
  (150 additional predictors: main feature × firm-link dummy)
- **LASSO** (Ridge, ENet, PCR, NN)

Motivation

Correlation is central to portfolio construction and risk management

- Comparing with return and volatility forecasting, less is known about correlation forecasting

Novel framework to forecast realized correlation RC via big data + ML

- **25 main features: HAR, factor-driven, EMA features** (150 additional predictors: main feature × firm-link dummy)
- **LASSO** (Ridge, ENet, PCR, NN)

Motivation

Correlation is central to portfolio construction and risk management

- Comparing with return and volatility forecasting, less is known about correlation forecasting

Novel framework to forecast realized correlation RC via big data + ML

- **25 main features: HAR, factor-driven, EMA features**
  (150 additional predictors: main feature × firm-link dummy)
- **LASSO** (Ridge, ENet, PCR, NN)

Motivation

Correlation is central to portfolio construction and risk management

- Comparing with return and volatility forecasting, less is known about correlation forecasting

Novel framework to forecast realized correlation RC via big data + ML

- **25 main features: HAR, factor-driven, EMA features**
  (150 additional predictors: main feature $\times$ firm-link dummy)
- **LASSO** (Ridge, ENet, PCR, NN)

Motivation

Correlation is central to portfolio construction and risk management

- Comparing with return and volatility forecasting, less is known about correlation forecasting

Novel framework to forecast realized correlation RC via big data + ML

- **25 main features: HAR, factor-driven, EMA features** (150 additional predictors: main feature $\times$ firm-link dummy)
- **LASSO** (Ridge, ENet, PCR, NN)

Preview of Results

Benchmark: HAR model by Corsi (2009)
e.g., lagged daily, weekly, monthly RC to forecast next-month RC

Relative to the HAR benchmark, our LASSO framework is able to:

- Improve R^2_{OOS} of RC forecast by 10%
- Increase pairs trading strategy return from 3.63% to 9.34% per annum based on return convergence approx. by RC forecast
- A one-SD increase in forecasted average RC based on LASSO predicts a rise in market excess return of 18.3% per year
- Produce ex-ante portfolio risk much closer to the realized risk
- Reduce the risk of Global Minimum Variance (GMV) portfolios

Preview of Results

Benchmark: HAR model by Corsi (2009)

e.g., lagged daily, weekly, monthly RC to forecast next-month RC

Relative to the HAR benchmark, our LASSO framework is able to:

- Improve R^2_{OOS} of RC forecast by 10%
- Increase pairs trading strategy return from 3.63% to 9.34% per annum based on return convergence approx. by RC forecast
- A one-SD increase in forecasted average RC based on LASSO predicts a rise in market excess return of 18.3% per year
- Produce ex-ante portfolio risk much closer to the realized risk
- Reduce the risk of Global Minimum Variance (GMV) portfolios

Preview of Results

Benchmark: HAR model by Corsi (2009)

e.g., lagged daily, weekly, monthly RC to forecast next-month RC

Relative to the HAR benchmark, our LASSO framework is able to:

- Improve R^2_{OOS} of RC forecast by 10%
- Increase pairs trading strategy return from 3.63% to 9.34% per annum based on return convergence approx. by RC forecast
- A one-SD increase in forecasted average RC based on LASSO predicts a rise in market excess return of 18.3% per year
- Produce ex-ante portfolio risk much closer to the realized risk
- Reduce the risk of Global Minimum Variance (GMV) portfolios

Preview of Results

Benchmark: HAR model by Corsi (2009)
e.g., lagged daily, weekly, monthly RC to forecast next-month RC

Relative to the HAR benchmark, our LASSO framework is able to:

- Improve R^2_{OOS} of RC forecast by 10%
- Increase pairs trading strategy return from 3.63% to 9.34% per annum based on return convergence approx. by RC forecast
- A one-SD increase in forecasted average RC based on LASSO predicts a rise in market excess return of 18.3% per year
- Produce ex-ante portfolio risk much closer to the realized risk
- Reduce the risk of Global Minimum Variance (GMV) portfolios

Preview of Results

Benchmark: HAR model by Corsi (2009)
e.g., lagged daily, weekly, monthly RC to forecast next-month RC

Relative to the HAR benchmark, our LASSO framework is able to:

- Improve R^2_{OOS} of RC forecast by 10%
- Increase pairs trading strategy return from 3.63% to 9.34% per annum based on return convergence approx. by RC forecast
- A one-SD increase in forecasted average RC based on LASSO predicts a rise in market excess return of 18.3% per year
- Produce ex-ante portfolio risk much closer to the realized risk
- Reduce the risk of Global Minimum Variance (GMV) portfolios

Preview of Results

Benchmark: HAR model by Corsi (2009)
e.g., lagged daily, weekly, monthly RC to forecast next-month RC

Relative to the HAR benchmark, our LASSO framework is able to:

- Improve R^2_{OOS} of RC forecast by 10%
- Increase pairs trading strategy return from 3.63% to 9.34% per annum based on return convergence approx. by RC forecast
- A one-SD increase in forecasted average RC based on LASSO predicts a rise in market excess return of 18.3% per year
- Produce ex-ante portfolio risk much closer to the realized risk
- Reduce the risk of Global Minimum Variance (GMV) portfolios

Preview of Results

Benchmark: HAR model by Corsi (2009)
e.g., lagged daily, weekly, monthly RC to forecast next-month RC

Relative to the HAR benchmark, our LASSO framework is able to:

- Improve R^2_{OOS} of RC forecast by 10%
- Increase pairs trading strategy return from 3.63% to 9.34% per annum based on return convergence approx. by RC forecast
- A one-SD increase in forecasted average RC based on LASSO predicts a rise in market excess return of 18.3% per year
- Produce ex-ante portfolio risk much closer to the realized risk
- Reduce the risk of Global Minimum Variance (GMV) portfolios

Outline

- **Data and Variables**
- Estimation Methodology
- Out-of-sample Forecast Performance
- Applications
- Robustness

Response Variable

Covariance matrix $RCov_t$ can be decomposed into:

$$RCov_t = \sqrt{RV_t} \cdot RC_t \cdot \sqrt{RV_t}$$

- $\sqrt{RV_t}$: diagonal matrix of volatilities
- RC_t: correlation matrix
- RV_t and RC_t different dynamics
- Forecast RV_t and RC_t separately
- Main focus of this paper: forecast RC_t
- RV_t modeled by univariate HAR models; more sophisticated ML-based method to forecast volatility see Li and Tang (2023) Automated Volatility Forecasting

Response Variable

Covariance matrix $RCov_t$ can be decomposed into:

$$RCov_t = \sqrt{RV_t} \cdot RC_t \cdot \sqrt{RV_t}$$

- $\sqrt{RV_t}$: diagonal matrix of volatilities
- RC_t: correlation matrix
- RV_t and RC_t different dynamics
- Forecast RV_t and RC_t separately
- Main focus of this paper: forecast RC_t
- RV_t modeled by univariate HAR models; more sophisticated ML-based method to forecast volatility see Li and Tang (2023) Automated Volatility Forecasting

Response Variable

Covariance matrix $RCov_t$ can be decomposed into:

$$RCov_t = \sqrt{RV_t} \cdot RC_t \cdot \sqrt{RV_t}$$

- $\sqrt{RV_t}$: diagonal matrix of volatilities
- RC_t: correlation matrix
- RV_t and RC_t different dynamics
- Forecast RV_t and RC_t separately
- Main focus of this paper: forecast RC_t
- RV_t modeled by univariate HAR models; more sophisticated ML-based method to forecast volatility see Li and Tang (2023) Automated Volatility Forecasting

Response Variable

Covariance matrix $RCov_t$ can be decomposed into:

$$RCov_t = \sqrt{RV_t} \cdot RC_t \cdot \sqrt{RV_t}$$

- $\sqrt{RV_t}$: diagonal matrix of volatilities
- RC_t: correlation matrix
- RV_t and RC_t different dynamics
- Forecast RV_t and RC_t separately
- Main focus of this paper: forecast RC_t
- RV_t modeled by univariate HAR models; more sophisticated ML-based method to forecast volatility see Li and Tang (2023) Automated Volatility Forecasting

Response Variable

Covariance matrix $RCov_t$ can be decomposed into:

$$RCov_t = \sqrt{RV_t} \cdot RC_t \cdot \sqrt{RV_t}$$

- $\sqrt{RV_t}$: diagonal matrix of volatilities
- RC_t: correlation matrix
- RV_t and RC_t different dynamics
- Forecast RV_t and RC_t separately
- Main focus of this paper: forecast RC_t
- RV_t modeled by univariate HAR models; more sophisticated ML-based method to forecast volatility see Li and Tang (2023) Automated Volatility Forecasting

Response Variable

Covariance matrix $RCov_t$ can be decomposed into:

$$RCov_t = \sqrt{RV_t} \cdot RC_t \cdot \sqrt{RV_t}$$

- $\sqrt{RV_t}$: diagonal matrix of volatilities
- RC_t: correlation matrix
- RV_t and RC_t different dynamics
- Forecast RV_t and RC_t separately
- Main focus of this paper: forecast RC_t
- RV_t modeled by univariate HAR models; more sophisticated ML-based method to forecast volatility see Li and Tang (2023) Automated Volatility Forecasting

Response Variable

Covariance matrix $RCov_t$ can be decomposed into:

$$RCov_t = \sqrt{RV_t} \cdot RC_t \cdot \sqrt{RV_t}$$

- $\sqrt{RV_t}$: diagonal matrix of volatilities
- RC_t: correlation matrix
- RV_t and RC_t different dynamics
- Forecast RV_t and RC_t separately
- Main focus of this paper: forecast RC_t
- RV_t modeled by univariate HAR models; more sophisticated ML-based method to forecast volatility see Li and Tang (2023) Automated Volatility Forecasting

Response Variable - continue

Response y: 1-month ahead realized correlation for all pairs

- 417 S&P 500 stocks with full history over 2003-2020
- 15-minute mid-quote prices from the TAQ database
- Choice of universe, frequency, and mid-quote data effectively mitigate non-synchronous prices and bid-ask bounce effects

Response Variable - continue

Response y: 1-month ahead realized correlation for all pairs

- 417 S&P 500 stocks with full history over 2003-2020
- 15-minute mid-quote prices from the TAQ database
- Choice of universe, frequency, and mid-quote data effectively mitigate non-synchronous prices and bid-ask bounce effects

Response Variable - continue

Response y: 1-month ahead realized correlation for all pairs

- 417 S&P 500 stocks with full history over 2003-2020
- 15-minute mid-quote prices from the TAQ database
- Choice of universe, frequency, and mid-quote data effectively mitigate non-synchronous prices and bid-ask bounce effects

Response Variable - continue

Response y: 1-month ahead realized correlation for all pairs

- 417 S&P 500 stocks with full history over 2003-2020
- 15-minute mid-quote prices from the TAQ database
- Choice of universe, frequency, and mid-quote data effectively mitigate non-synchronous prices and bid-ask bounce effects

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Features

1. HAR features RC_t^d, RC_t^w, RC_t^m, RC_t^{d-}, RC_t^{w-}, RC_t^{m-}
   Refer to model based on HAR features as "SHAR" Model
2. Factor-driven features FRC^d, FRC^w, FRC^m
   Refer to 6 HAR features + 3 FRC features as "SHAR-F" model
3. EMA features EMA of lagged RC & semi RC + within-sector RC
   Denote 6 HAR + 3 FRC + 16 EMA features as "SHAR-F-Exp"

One major contribution:

- A large and novel feature set for correlation prediction

To the best of our knowledge, we are the first to:

- Use EMA terms with sector risk to predict correlation
- Use observable firm char to back out factor-driven realized features instead of constructing high-frequency factors
- Combine features from econometrics, statistics, and finance literature

Outline

- Data and Variables
- **Estimation Methodology**
- Out-of-sample Forecast Performance
- Applications
- Robustness

Training and Validation

In parallel to other machine learning algorithms, LASSO requires a validation set for tuning its hyperparameter

Training-validation-testing scheme:

- "Pooled models" based on panel data for all stock pairs
- A training set consisting of data from year $t-4$ to year $t-1$, a validation set consisting of year t data, and a testing set consisting of year $t+1$ data
- Refit the models every year by rolling the training, validation, and testing sets one year forward

Allows the features selected by LASSO to dynamically enter and exit the prediction models based on changing market conditions

Training and Validation

In parallel to other machine learning algorithms, LASSO requires a validation set for tuning its hyperparameter

Training-validation-testing scheme:

- "Pooled models" based on panel data for all stock pairs
- A training set consisting of data from year $t-4$ to year $t-1$, a validation set consisting of year t data, and a testing set consisting of year $t+1$ data
- Refit the models every year by rolling the training, validation, and testing sets one year forward

Allows the features selected by LASSO to dynamically enter and exit the prediction models based on changing market conditions

Training and Validation

In parallel to other machine learning algorithms, LASSO requires a validation set for tuning its hyperparameter

Training-validation-testing scheme:

- "Pooled models" based on panel data for all stock pairs
- A training set consisting of data from year $t-4$ to year $t-1$, a validation set consisting of year t data, and a testing set consisting of year $t+1$ data
- Refit the models every year by rolling the training, validation, and testing sets one year forward

Allows the features selected by LASSO to dynamically enter and exit the prediction models based on changing market conditions

Training and Validation

In parallel to other machine learning algorithms, LASSO requires a validation set for tuning its hyperparameter

Training-validation-testing scheme:

- "Pooled models" based on panel data for all stock pairs
- A training set consisting of data from year $t-4$ to year $t-1$, a validation set consisting of year t data, and a testing set consisting of year $t+1$ data
- Refit the models every year by rolling the training, validation, and testing sets one year forward

Allows the features selected by LASSO to dynamically enter and exit the prediction models based on changing market conditions

Training and Validation

In parallel to other machine learning algorithms, LASSO requires a validation set for tuning its hyperparameter

Training-validation-testing scheme:

- "Pooled models" based on panel data for all stock pairs
- A training set consisting of data from year $t-4$ to year $t-1$, a validation set consisting of year t data, and a testing set consisting of year $t+1$ data
- Refit the models every year by rolling the training, validation, and testing sets one year forward

Allows the features selected by LASSO to dynamically enter and exit the prediction models based on changing market conditions

Training and Validation

In parallel to other machine learning algorithms, LASSO requires a validation set for tuning its hyperparameter

Training-validation-testing scheme:

- "Pooled models" based on panel data for all stock pairs
- A training set consisting of data from year $t-4$ to year $t-1$, a validation set consisting of year t data, and a testing set consisting of year $t+1$ data
- Refit the models every year by rolling the training, validation, and testing sets one year forward

Allows the features selected by LASSO to dynamically enter and exit the prediction models based on changing market conditions

Model Fitting and LASSO

Simple linear combinations of the different features $f(x_{ij,t};\theta) \equiv x'_{ij,t}\theta$

Unlike OLS, LASSO estimates θ through a penalized L_1 loss function

$$\mathscr{L}^{LASSO}(\theta;\lambda) = \frac{1}{N}\sum_{(ij,t)\in\mathscr{T}}(RC^m_{ij,t+1} - x'_{ij,t}\theta)^2 + \lambda\sum_{p=1}^{P}|\theta_p|$$

- λ: the shrinkage parameter that controls the degrees of penalty
- $\lambda = 0$ collapses to standard OLS; $\lambda > 0$ performs feature selection

Empirically, features are normalized by training-set mean and standard deviation to have comparable magnitudes for meaningful feature selection

Model Fitting and LASSO

Simple linear combinations of the different features $f(x_{ij,t};\theta) \equiv x'_{ij,t}\theta$

Unlike OLS, LASSO estimates θ through a penalized L_1 loss function

$$\mathscr{L}^{LASSO}(\theta;\lambda) = \frac{1}{N}\sum_{(ij,t)\in\mathscr{T}}(RC^m_{ij,t+1} - x'_{ij,t}\theta)^2 + \lambda\sum_{p=1}^{P}|\theta_p|$$

- λ: the shrinkage parameter that controls the degrees of penalty
- $\lambda = 0$ collapses to standard OLS; $\lambda > 0$ performs feature selection

Empirically, features are normalized by training-set mean and standard deviation to have comparable magnitudes for meaningful feature selection

Model Fitting and LASSO

Simple linear combinations of the different features $f(x_{ij,t};\theta) \equiv x'_{ij,t}\theta$

Unlike OLS, LASSO estimates θ through a penalized L_1 loss function

$$\mathscr{L}^{LASSO}(\theta;\lambda) = \frac{1}{N}\sum_{(ij,t)\in\mathscr{T}}(RC^{m}_{ij,t+1} - x'_{ij,t}\theta)^2 + \lambda\sum_{p=1}^{P}|\theta_p|$$

- λ: the shrinkage parameter that controls the degrees of penalty
- $\lambda = 0$ collapses to standard OLS; $\lambda > 0$ performs feature selection

Empirically, features are normalized by training-set mean and standard deviation to have comparable magnitudes for meaningful feature selection

Model Fitting and LASSO

Simple linear combinations of the different features $f(x_{ij,t};\theta) \equiv x'_{ij,t}\theta$

Unlike OLS, LASSO estimates θ through a penalized L_1 loss function

$$\mathscr{L}^{LASSO}(\theta;\lambda) = \frac{1}{N}\sum_{(ij,t)\in\mathscr{T}}(RC^{m}_{ij,t+1} - x'_{ij,t}\theta)^2 + \lambda\sum_{p=1}^{P}|\theta_p|$$

- λ: the shrinkage parameter that controls the degrees of penalty
- $\lambda = 0$ collapses to standard OLS; $\lambda > 0$ performs feature selection

Empirically, features are normalized by training-set mean and standard deviation to have comparable magnitudes for meaningful feature selection

Model Fitting and LASSO

Simple linear combinations of the different features $f(x_{ij,t};\theta) \equiv x'_{ij,t}\theta$

Unlike OLS, LASSO estimates θ through a penalized L_1 loss function

$$\mathscr{L}^{LASSO}(\theta;\lambda) = \frac{1}{N}\sum_{(ij,t)\in\mathscr{T}}(RC^m_{ij,t+1} - x'_{ij,t}\theta)^2 + \lambda\sum_{p=1}^{P}|\theta_p|$$

- λ: the shrinkage parameter that controls the degrees of penalty
- $\lambda = 0$ collapses to standard OLS; $\lambda > 0$ performs feature selection

Empirically, features are normalized by training-set mean and standard deviation to have comparable magnitudes for meaningful feature selection

Outline

- Data and Variables
- Estimation Methodology
- **Out-of-sample Forecast Performance**
- Applications
- Robustness

Performance Evaluation Measures

- Out-of-sample R^2's relative to the HAR model

$$R^2_{OOS}(\theta) = 1 - \frac{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}(RC^m_{ij,t} - \widehat{RC}^{m,\theta}_{ij,t})^2}{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}(RC^m_{ij,t} - \widehat{RC}^{m,HAR}_{ij,t})^2}$$

- $\omega_{ij,t} = 1 \Longrightarrow R^{2,EW}_{OOS}$; $\omega_{ij,t} =$ product of market caps $\Longrightarrow R^{2,VW}_{OOS}$

- a positive $R^2_{OOS}(\theta)$ indicates that model θ achieves smaller out-of-sample prediction mean squared errors than HAR

- Modified Diebold and Mariano test for pairwise comparison of two models

- based on the difference in the out-of-sample squared error losses

- equal-weighted and value-weighted versions

Performance Evaluation Measures

- Out-of-sample R^2's relative to the HAR model

$$R^2_{OOS}(\theta) = 1 - \frac{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}\left(RC^m_{ij,t} - \widehat{RC}^{m,\theta}_{ij,t}\right)^2}{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}\left(RC^m_{ij,t} - \widehat{RC}^{m,HAR}_{ij,t}\right)^2}$$

- $\omega_{ij,t} = 1 \Longrightarrow R^{2,EW}_{OOS}$; $\omega_{ij,t} =$ product of market caps $\Longrightarrow R^{2,VW}_{OOS}$

- a positive $R^2_{OOS}(\theta)$ indicates that model θ achieves smaller out-of-sample prediction mean squared errors than HAR

- Modified Diebold and Mariano test for pairwise comparison of two models

  - based on the difference in the out-of-sample squared error losses

  - equal-weighted and value-weighted versions

Performance Evaluation Measures

- Out-of-sample R^2's relative to the HAR model

$$R^2_{OOS}(\theta) = 1 - \frac{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}(RC^m_{ij,t} - \widehat{RC}^{m,\theta}_{ij,t})^2}{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}(RC^m_{ij,t} - \widehat{RC}^{m,HAR}_{ij,t})^2}$$

- $\omega_{ij,t} = 1 \Longrightarrow R^{2,EW}_{OOS}$; $\omega_{ij,t} =$ product of market caps $\Longrightarrow R^{2,VW}_{OOS}$

- a positive $R^2_{OOS}(\theta)$ indicates that model θ achieves smaller out-of-sample prediction mean squared errors than HAR

- Modified Diebold and Mariano test for pairwise comparison of two models

- based on the difference in the out-of-sample squared error losses

- equal-weighted and value-weighted versions

Performance Evaluation Measures

- Out-of-sample R^2's relative to the HAR model

$$R^2_{OOS}(\theta) = 1 - \frac{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}\left(RC^m_{ij,t} - \widehat{RC}^{m,\theta}_{ij,t}\right)^2}{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}\left(RC^m_{ij,t} - \widehat{RC}^{m,HAR}_{ij,t}\right)^2}$$

- $\omega_{ij,t} = 1 \Longrightarrow R^{2,EW}_{OOS}$; $\omega_{ij,t} =$ product of market caps $\Longrightarrow R^{2,VW}_{OOS}$

- a positive $R^2_{OOS}(\theta)$ indicates that model θ achieves smaller out-of-sample prediction mean squared errors than HAR

- Modified Diebold and Mariano test for pairwise comparison of two models

- based on the difference in the out-of-sample squared error losses

- equal-weighted and value-weighted versions

Performance Evaluation Measures

- Out-of-sample R^2's relative to the HAR model

$$R^2_{OOS}(\theta) = 1 - \frac{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}(RC^m_{ij,t} - \widehat{RC}^{m,\theta}_{ij,t})^2}{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t}(RC^m_{ij,t} - \widehat{RC}^{m,HAR}_{ij,t})^2}$$

- $\omega_{ij,t} = 1 \Longrightarrow R^{2,EW}_{OOS}$; $\omega_{ij,t} =$ product of market caps $\Longrightarrow R^{2,VW}_{OOS}$

- a positive $R^2_{OOS}(\theta)$ indicates that model θ achieves smaller out-of-sample prediction mean squared errors than HAR

- Modified Diebold and Mariano test for pairwise comparison of two models

- based on the difference in the out-of-sample squared error losses

- equal-weighted and value-weighted versions

Performance Evaluation Measures

- Out-of-sample R^2's relative to the HAR model

$$R^2_{OOS}(\theta) = 1 - \frac{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t} (RC^m_{ij,t} - \widehat{RC}^{m,\theta}_{ij,t})^2}{\sum_{(ij,t)\in\mathscr{T}'} \omega_{ij,t} (RC^m_{ij,t} - \widehat{RC}^{m,HAR}_{ij,t})^2}$$

- $\omega_{ij,t} = 1 \Longrightarrow R^{2,EW}_{OOS}$; $\omega_{ij,t} =$ product of market caps $\Longrightarrow R^{2,VW}_{OOS}$

- a positive $R^2_{OOS}(\theta)$ indicates that model θ achieves smaller out-of-sample prediction mean squared errors than HAR

- Modified Diebold and Mariano test for pairwise comparison of two models

- based on the difference in the out-of-sample squared error losses

- equal-weighted and value-weighted versions

Forecast Performance - R^2_{OOS} relative to HAR

Model	Feature Set	Equal-weighted	Value-weighted
(1) SHAR	3 RC^h + 3 RC^{h-} (# of features = 6)	0.22%	0.11%

Forecast Performance - R^2_{OOS} relative to HAR

Model	Feature Set	Equal-weighted	Value-weighted
(1) SHAR	$3\ RC^h + 3\ RC^{h-}$ (# of features = 6)	0.22%	0.11%
(2) SHAR-F	$3\ RC^h + 3\ RC^{h-}$ $+\ 3\ FRC^h$ (# of features = 9)	1.71%	1.30%

Forecast Performance - R^2_{OOS} relative to HAR

Model	Feature Set	Equal-weighted	Value-weighted
(1) SHAR	$3\ RC^h + 3\ RC^{h-}$ (# of features = 6)	0.22%	0.11%
(2) SHAR-F	$3\ RC^h + 3\ RC^{h-}$ $+ 3\ FRC^h$ (# of features = 9)	1.71%	1.30%
(3) SHAR-F-Exp	$3\ RC^h + 3\ RC^{h-}$ $+ 3\ FRC^h$ $+ 4\ ExpRC^h + 4\ ExpRC^{h-}$ $+ 4\ ExpScRC^h + 4\ ExpScRC^{h-}$ (# of features = 25)	9.82%	7.31%

Forecast Performance - R^2_{OOS} relative to HAR

Model	Feature Set	Equal-weighted	Value-weighted
(1) SHAR	$3\ RC^h + 3\ RC^{h-}$ (# of features = 6)	0.22%	0.11%
(2) SHAR-F	$3\ RC^h + 3\ RC^{h-}$ $+ 3\ FRC^h$ (# of features = 9)	1.71%	1.30%
(3) SHAR-F-Exp	$3\ RC^h + 3\ RC^{h-}$ $+ 3\ FRC^h$ $+ 4\ ExpRC^h + 4\ ExpRC^{h-}$ $+ 4\ ExpScRC^h + 4\ ExpScRC^{h-}$ (# of features = 25)	9.82%	7.31%
(4) LASSO	All 25 main features	10.16%	8.05%

Forecast Performance - Modified DM Tests

Panel B: DM *t*-statistics (equal-weighted)

	Model	HAR	(1)	(2)	(3)
(1)	SHAR	11.55			
(2)	SHAR-F	29.32	27.58		
(3)	SHAR-F-Exp	39.08	39.84	35.24	
(4)	LASSO	47.70	48.93	43.43	6.31

Forecast Performance - Modified DM Tests

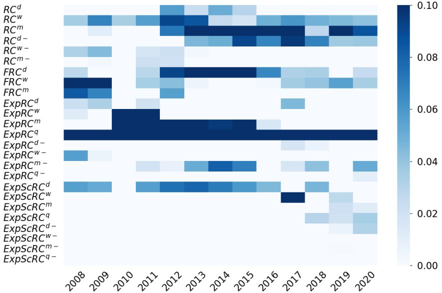
Panel B: DM *t*-statistics (equal-weighted)

	Model	HAR	(1)	(2)	(3)
(1)	SHAR	11.55			
(2)	SHAR-F	29.32	27.58		
(3)	SHAR-F-Exp	39.08	39.84	35.24	
(4)	LASSO	47.70	48.93	43.43	6.31

Panel C: DM *t*-statistics (value-weighted)

	Model	HAR	(1)	(2)	(3)
(1)	SHAR	4.99			
(2)	SHAR-F	13.56	13.41		
(3)	SHAR-F-Exp	16.21	16.29	15.51	
(4)	LASSO	17.85	17.91	17.41	8.99

Feature Selection

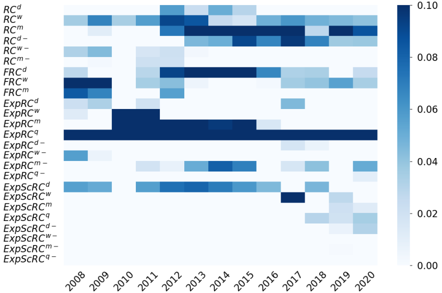


10 features on average

- $ExpRC^q$: 13/13, 50%
- RC^m: 10/13, 11%
- FRC^d, FRC^w, $ExpScRC^d$
- $ExpRC^m$: 7/13, 15%

- Several long-term predictors are consistently selected over time
- Different short-term signals enter and exit the models
- Most sparse set for 2010 to adapt to changing market conditions

Feature Selection

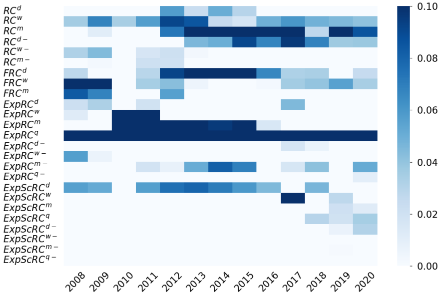


10 features on average

- $ExpRC^q$: 13/13, 50%
- RC^m: 10/13, 11%
- FRC^d, FRC^w, $ExpScRC^d$
- $ExpRC^m$: 7/13, 15%

- Several long-term predictors are consistently selected over time
- Different short-term signals enter and exit the models
- Most sparse set for 2010 to adapt to changing market conditions

Feature Selection

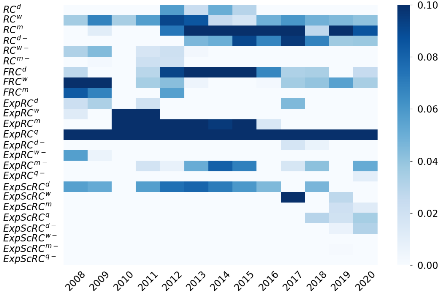


10 features on average

- $ExpRC^q$: 13/13, 50%
- RC^m: 10/13, 11%
- $FRC^d, FRC^w, ExpScRC^d$
- $ExpRC^m$: 7/13, 15%

- Several long-term predictors are consistently selected over time
- Different short-term signals enter and exit the models
- Most sparse set for 2010 to adapt to changing market conditions

Feature Selection



10 features on average

- $ExpRC^q$: 13/13, 50%
- RC^m: 10/13, 11%
- $FRC^d, FRC^w, ExpScRC^d$
- $ExpRC^m$: 7/13, 15%

- Several long-term predictors are consistently selected over time
- Different short-term signals enter and exit the models
- Most sparse set for 2010 to adapt to changing market conditions

Outline

- Data and Variables
- Estimation Methodology
- Out-of-sample Forecast Performance
- **Applications**
- Robustness

Application

- The improvements in out-of-sample R^2 based on LASSO framework are well demonstrated, **open question:**

  Can statistical improvements translate into economic gains?

- Evaluate the economic significance by considering four practical applications:
  1. Augmented pairs trading strategy
  2. Equity premium prediction
  3. Risk-targeting
  4. Global Minimum Variance (GMV) portfolio construction

Application

- The improvements in out-of-sample R^2 based on LASSO framework are well demonstrated, **open question:**

  Can statistical improvements translate into economic gains?

- Evaluate the economic significance by considering four practical applications:

  1. Augmented pairs trading strategy
  2. Equity premium prediction
  3. Risk-targeting
  4. Global Minimum Variance (GMV) portfolio construction

Application

- The improvements in out-of-sample R^2 based on LASSO framework are well demonstrated, **open question:**

  Can statistical improvements translate into economic gains?

- Evaluate the economic significance by considering four practical applications:

  1. Augmented pairs trading strategy
  2. Equity premium prediction
  3. Risk-targeting
  4. Global Minimum Variance (GMV) portfolio construction

Application

- The improvements in out-of-sample R^2 based on LASSO framework are well demonstrated, **open question:**

  Can statistical improvements translate into economic gains?

- Evaluate the economic significance by considering four practical applications:

  1. Augmented pairs trading strategy
  2. Equity premium prediction
  3. Risk-targeting
  4. Global Minimum Variance (GMV) portfolio construction

Application

- The improvements in out-of-sample R^2 based on LASSO framework are well demonstrated, **open question:**

  Can statistical improvements translate into economic gains?

- Evaluate the economic significance by considering four practical applications:

  1. Augmented pairs trading strategy
  2. Equity premium prediction
  3. Risk-targeting
  4. Global Minimum Variance (GMV) portfolio construction

Application

- The improvements in out-of-sample R^2 based on LASSO framework are well demonstrated, **open question:**

  Can statistical improvements translate into economic gains?

- Evaluate the economic significance by considering four practical applications:

  1. Augmented pairs trading strategy
  2. Equity premium prediction
  3. Risk-targeting
  4. Global Minimum Variance (GMV) portfolio construction

Application

- The improvements in out-of-sample R^2 based on LASSO framework are well demonstrated, **open question:**

  Can statistical improvements translate into economic gains?

- Evaluate the economic significance by considering four practical applications:

  1. Augmented pairs trading strategy
  2. Equity premium prediction
  3. Risk-targeting
  4. Global Minimum Variance (GMV) portfolio construction

Application 1 - Pairs Trading

Bets on price convergence: stocks with return above/below its pair portfolio are likely overvalued/undervalued (Chen et al., 2016)

Return divergence (*RetDiff*): return difference between stock i and its pair portfolio

$$RetDiff_{i,t} = \beta_{i,t}(PRet_{i,t} - r_{f,t}) - (Ret_{i,t} - r_{f,t})$$

- β_i: regression coefficient from regressing stock i's returns on its pair portfolio returns using daily data between month $t-12$ and $t-1$
- Define the top 20 stocks with the highest one-year historical correlation with stock i as its pairs

Application 1 - Pairs Trading

Bets on price convergence: stocks with return above/below its pair portfolio are likely overvalued/undervalued (Chen et al., 2016)

Return divergence ($RetDiff$): return difference between stock i and its pair portfolio

$$RetDiff_{i,t} = \beta_{i,t}(PRet_{i,t} - r_{f,t}) - (Ret_{i,t} - r_{f,t})$$

- β_i: regression coefficient from regressing stock i's returns on its pair portfolio returns using daily data between month $t-12$ and $t-1$
- Define the top 20 stocks with the highest one-year historical correlation with stock i as its pairs

Application 1 - Pairs Trading

Bets on price convergence: stocks with return above/below its pair portfolio are likely overvalued/undervalued (Chen et al., 2016)

Return divergence (*RetDiff*): return difference between stock i and its pair portfolio

$$RetDiff_{i,t} = \beta_{i,t}(PRet_{i,t} - r_{f,t}) - (Ret_{i,t} - r_{f,t})$$

- β_i: regression coefficient from regressing stock i's returns on its pair portfolio returns using daily data between month $t-12$ and $t-1$
- Define the top 20 stocks with the highest one-year historical correlation with stock i as its pairs

Application 1 - Pairs Trading

Bets on price convergence: stocks with return above/below its pair portfolio are likely overvalued/undervalued (Chen et al., 2016)

Return divergence (*RetDiff*): return difference between stock i and its pair portfolio

$$RetDiff_{i,t} = \beta_{i,t}(PRet_{i,t} - r_{f,t}) - (Ret_{i,t} - r_{f,t})$$

- β_i: regression coefficient from regressing stock i's returns on its pair portfolio returns using daily data between month $t-12$ and $t-1$
- Define the top 20 stocks with the highest one-year historical correlation with stock i as its pairs

Application 1 - Pairs Trading

A key implicit assumption behind the above pairs trading strategy is the persistence of correlations

To improve the strategy performance, we explicitly incorporate correlation predictions into the portfolio construction

- Use $\Delta RC^{\theta}_{i,t} = \widehat{RC}^{\theta}_{i,t} - RC^{h}_{i,t}$ to capture the persistence of correlations
- Keep the subset of stocks in the highest ΔRC^{θ} quintile
- Form pairs trading strategy with this subset of stocks

First demonstrate the failure of traditional pairs trading

Then show how better corr forecasts could enhance pairs trading strategy

Application 1 - Pairs Trading

A key implicit assumption behind the above pairs trading strategy is the persistence of correlations

To improve the strategy performance, we explicitly incorporate correlation predictions into the portfolio construction

- Use $\Delta RC_{i,t}^{\theta} = \widehat{RC}_{i,t}^{\theta} - RC_{i,t}^{h}$ to capture the persistence of correlations
- Keep the subset of stocks in the highest ΔRC^{θ} quintile
- Form pairs trading strategy with this subset of stocks

First demonstrate the failure of traditional pairs trading

Then show how better corr forecasts could enhance pairs trading strategy

Application 1 - Pairs Trading

A key implicit assumption behind the above pairs trading strategy is the persistence of correlations

To improve the strategy performance, we explicitly incorporate correlation predictions into the portfolio construction

- Use $\Delta RC_{i,t}^{\theta} = \widehat{RC}_{i,t}^{\theta} - RC_{i,t}^{h}$ to capture the persistence of correlations
- Keep the subset of stocks in the highest ΔRC^{θ} quintile
- Form pairs trading strategy with this subset of stocks

First demonstrate the failure of traditional pairs trading

Then show how better corr forecasts could enhance pairs trading strategy

Application 1 - Pairs Trading

A key implicit assumption behind the above pairs trading strategy is the persistence of correlations

To improve the strategy performance, we explicitly incorporate correlation predictions into the portfolio construction

- Use $\Delta RC_{i,t}^{\theta} = \widehat{RC}_{i,t}^{\theta} - RC_{i,t}^{h}$ to capture the persistence of correlations
- Keep the subset of stocks in the highest ΔRC^{θ} quintile
- Form pairs trading strategy with this subset of stocks

First demonstrate the failure of traditional pairs trading

Then show how better corr forecasts could enhance pairs trading strategy

Application 1 - Pairs Trading

A key implicit assumption behind the above pairs trading strategy is the persistence of correlations

To improve the strategy performance, we explicitly incorporate correlation predictions into the portfolio construction

- Use $\Delta RC^{\theta}_{i,t} = \widehat{RC}^{\theta}_{i,t} - RC^{h}_{i,t}$ to capture the persistence of correlations
- Keep the subset of stocks in the highest ΔRC^{θ} quintile
- Form pairs trading strategy with this subset of stocks

First demonstrate the failure of traditional pairs trading

Then show how better corr forecasts could enhance pairs trading strategy

Application 1 - Pairs Trading

A key implicit assumption behind the above pairs trading strategy is the persistence of correlations

To improve the strategy performance, we explicitly incorporate correlation predictions into the portfolio construction

- Use $\Delta RC^{\theta}_{i,t} = \widehat{RC}^{\theta}_{i,t} - RC^{h}_{i,t}$ to capture the persistence of correlations
- Keep the subset of stocks in the highest ΔRC^{θ} quintile
- Form pairs trading strategy with this subset of stocks

First demonstrate the failure of traditional pairs trading

Then show how better corr forecasts could enhance pairs trading strategy

Application 1 - Pairs Trading

A key implicit assumption behind the above pairs trading strategy is the persistence of correlations

To improve the strategy performance, we explicitly incorporate correlation predictions into the portfolio construction

- Use $\Delta RC_{i,t}^{\theta} = \widehat{RC}_{i,t}^{\theta} - RC_{i,t}^{h}$ to capture the persistence of correlations
- Keep the subset of stocks in the highest ΔRC^{θ} quintile
- Form pairs trading strategy with this subset of stocks

First demonstrate the failure of traditional pairs trading

Then show how better corr forecasts could enhance pairs trading strategy

Application 1 - Pairs Trading

Panel A: Equal-weighted portfolio sorted by return divergence

	1 (Low)	2	3	4	5 (High)	HML
Unconditional	6.92%	5.75%	7.09%	6.45%	8.07%	1.15% (0.47)
HAR	9.53%	6.44%	9.25%	7.90%	13.16%	3.63% (0.88)
LASSO	3.50%	5.12%	6.08%	5.67%	12.84%	9.34% (2.30)

Application 1 - Pairs Trading

Panel A: Equal-weighted portfolio sorted by return divergence

	1 (Low)	2	3	4	5 (High)	HML
Unconditional	6.92%	5.75%	7.09%	6.45%	8.07%	1.15% (0.47)
HAR	9.53%	6.44%	9.25%	7.90%	13.16%	3.63% (0.88)
LASSO	3.50%	5.12%	6.08%	5.67%	12.84%	9.34% (2.30)

Panel B: Value-weighted portfolio sorted by return divergence

	1 (Low)	2	3	4	5 (High)	HML
Unconditional	6.05%	4.58%	6.05%	4.76%	4.86%	-1.20% (-0.45)
HAR	6.42%	6.63%	9.02%	7.90%	12.56%	6.14% (1.60)
LASSO	1.90%	5.68%	6.63%	6.28%	10.75%	8.85% (2.20)

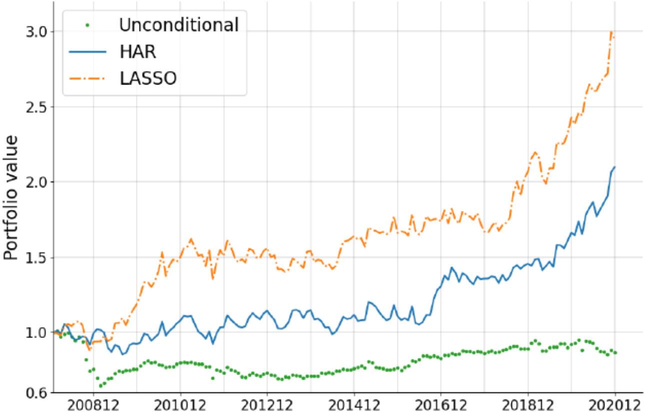
Application 1 - Pairs Trading

Panel C: Fama-MacBeth regressions

	Unconditional		HAR		LASSO	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.50	4.42	0.55	5.79	0.13	6.52
	(1.28)	(3.78)	(1.18)	(2.91)	(0.31)	(3.60)
RetDiff	0.03	0.04	0.08	0.12	0.14	0.16
	(0.54)	(0.97)	(1.02)	(1.80)	(1.87)	(2.33)
Controls	No	Yes	No	Yes	No	Yes
Adj-R^2	0.59%	12.01%	0.92%	11.73%	1.06%	12.82%
N	64,635	64,635	13,020	13,020	13,020	13,020

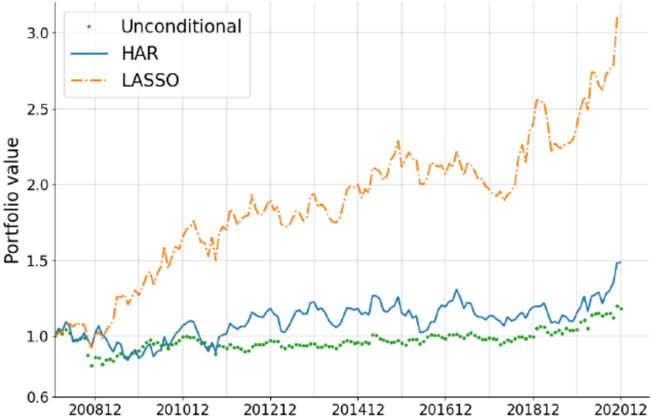
Application 1 - Pairs Trading

Cumulative profits of the equal-weighted strategy



Application 1 - Pairs Trading

Cumulative profits of the value-weighted strategy



Application 2 - Equity Premium Prediction

Average correlation among stocks manifests aggregate systematic risks and therefore predicts future market returns (Pollet and Wilson, 2010)

$$AvgCorr_t^{\theta} = \sum_{i=1}^{N} \sum_{j=1,\, j\neq i}^{N} \omega_{ij,t} \widehat{RC}_{ij,t+1}^{m,\theta}$$

- Originally, the average lagged pairwise correlation, $AvgCorr^{RC}$, is used to approx. the expected future average correlation
- By the same logic, the use of superior correlation predictions, $AvgCorr^{\theta}$, should result in stronger return predictive power
- Also include the eight commonly used macroeconomic predictors from Welch and Goyal (2008) as controls

Application 2 - Equity Premium Prediction

Average correlation among stocks manifests aggregate systematic risks and therefore predicts future market returns (Pollet and Wilson, 2010)

$$AvgCorr_t^{\theta} = \sum_{i=1}^{N} \sum_{j=1,\, j \neq i}^{N} \omega_{ij,t} \widehat{RC}_{ij,t+1}^{m,\theta}$$

- Originally, the average lagged pairwise correlation, $AvgCorr^{RC}$, is used to approx. the expected future average correlation
- By the same logic, the use of superior correlation predictions, $AvgCorr^{\theta}$, should result in stronger return predictive power
- Also include the eight commonly used macroeconomic predictors from Welch and Goyal (2008) as controls

Application 2 - Equity Premium Prediction

Average correlation among stocks manifests aggregate systematic risks and therefore predicts future market returns (Pollet and Wilson, 2010)

$$AvgCorr_t^{\theta} = \sum_{i=1}^{N} \sum_{j=1,\, j\neq i}^{N} \omega_{ij,t} \widehat{RC}_{ij,t+1}^{m,\theta}$$

- Originally, the average lagged pairwise correlation, $AvgCorr^{RC}$, is used to approx. the expected future average correlation
- By the same logic, the use of superior correlation predictions, $AvgCorr^{\theta}$, should result in stronger return predictive power
- Also include the eight commonly used macroeconomic predictors from Welch and Goyal (2008) as controls

Application 2 - Equity Premium Prediction

Average correlation among stocks manifests aggregate systematic risks and therefore predicts future market returns (Pollet and Wilson, 2010)

$$AvgCorr_t^{\theta} = \sum_{i=1}^{N} \sum_{j=1,\, j \neq i}^{N} \omega_{ij,t} \widehat{RC}_{ij,t+1}^{m,\theta}$$

- Originally, the average lagged pairwise correlation, $AvgCorr^{RC}$, is used to approx. the expected future average correlation
- By the same logic, the use of superior correlation predictions, $AvgCorr^{\theta}$, should result in stronger return predictive power
- Also include the eight commonly used macroeconomic predictors from Welch and Goyal (2008) as controls

Application 2 - Equity Premium Prediction

	Panel A: $AvgCorr^{EW}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.00	-0.02	-0.02	0.54	0.49	0.53
	(0.05)	(-0.94)	(-1.41)	(1.53)	(1.33)	(1.52)
AvgCorr						
RC	0.03			0.03		
	(0.88)			(0.52)		
HAR		0.11			0.08	
		(1.47)			(0.73)	
LASSO			0.13			0.24
			(2.00)			(2.40)
dp				0.12	0.11	0.13
				(1.65)	(1.49)	(1.77)
ep				-0.00	-0.00	-0.01
				(-0.17)	(-0.27)	(-0.31)
bm				-0.12	-0.13	-0.15
				(-0.83)	(-0.88)	(-0.99)
ntis				0.22	0.20	0.11
				(0.65)	(0.62)	(0.33)
tbl				-0.99	-0.91	-0.63
				(-1.36)	(-1.25)	(-0.87)
tms				0.11	0.10	0.09
				(0.87)	(0.83)	(0.73)
dfy				-2.99	-2.91	-4.84
				(-1.74)	(-1.74)	(-2.62)
svar				-0.18	-0.15	-0.23
				(-0.31)	(-0.27)	(-0.41)
Adj-R^2	-0.15%	0.74%	1.91%	1.76%	1.94%	5.33%
N	155	155	155	155	155	155

Application 2 - Equity Premium Prediction

	Panel A: $AvgCorr^{EW}$						Panel B: $AvgCorr^{VW}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.00	-0.02	-0.02	0.54	0.49	0.53	0.00	-0.02	-0.03	0.55	0.50	0.56
	(0.05)	(-0.94)	(-1.41)	(1.53)	(1.33)	(1.52)	(0.18)	(-0.86)	(-1.49)	(1.54)	(1.37)	(1.63)
AvgCorr												
RC	0.03			0.03			0.03			0.02		
	(0.88)			(0.52)			(0.75)			(0.37)		
HAR		0.11			0.08			0.10			0.06	
		(1.47)			(0.73)			(1.39)			(0.61)	
LASSO			0.13			0.24			0.13			0.25
			(2.00)			(2.40)			(2.08)			(2.66)
dp				0.12	0.11	0.13				0.12	0.11	0.14
				(1.65)	(1.49)	(1.77)				(1.66)	(1.52)	(1.90)
ep				-0.00	-0.00	-0.01				-0.00	-0.00	-0.00
				(-0.17)	(-0.27)	(-0.31)				(-0.14)	(-0.23)	(-0.21)
bm				-0.12	-0.13	-0.15				-0.12	-0.13	-0.13
				(-0.83)	(-0.88)	(-0.99)				(-0.82)	(-0.85)	(-0.93)
ntis				0.22	0.20	0.11				0.23	0.22	0.10
				(0.65)	(0.62)	(0.33)				(0.69)	(0.66)	(0.31)
tbl				-0.99	-0.91	-0.63				-0.99	-0.93	-0.67
				(-1.36)	(-1.25)	(-0.87)				(-1.36)	(-1.27)	(-0.92)
tms				0.11	0.10	0.09				0.12	0.11	0.09
				(0.87)	(0.83)	(0.73)				(0.92)	(0.87)	(0.73)
dfy				-2.99	-2.91	-4.84				-2.89	-2.83	-5.03
				(-1.74)	(-1.74)	(-2.62)				(-1.70)	(-1.70)	(-2.74)
svar				-0.18	-0.15	-0.23				-0.17	-0.17	-0.41
				(-0.31)	(-0.27)	(-0.41)				(-0.29)	(-0.30)	(-0.73)
Adj-R^2	-0.15%	0.74%	1.91%	1.76%	1.94%	5.33%	-0.29%	0.60%	2.12%	1.67%	1.82%	6.16%
N	155	155	155	155	155	155	155	155	155	155	155	155

Application 3 - Risk Targeting

Consider a portfolio manager who allocates her funds into N risky assets based on a long-short trading strategy

- Set portfolio weight for stock i to $\omega_{i,t} = 1(-1)$ if the stock is in the long-leg (short-leg) of the strategy
  - Use simple HAR model for $\widehat{RV}_t$
- Average risk-targeting ratios across testing samples
  - $AvgRatio^{\theta} = \frac{1}{T}\sum_{t=1}^{T} \frac{\omega_t' \widehat{RCov}_t^{\theta} \omega_t}{\omega_t' RCov_t \omega_t}$
- Consider 15 different long-short strategies

More accurate corr forecasts $\Longrightarrow$ average risk targeting ratio close to 1

Application 3 - Risk Targeting

Consider a portfolio manager who allocates her funds into N risky assets based on a long-short trading strategy

- Set portfolio weight for stock i to $\omega_{i,t} = 1(-1)$ if the stock is in the long-leg (short-leg) of the strategy
  - Use simple HAR model for $\widehat{RV}_t$
- Average risk-targeting ratios across testing samples
  - $AvgRatio^{\theta} = \frac{1}{T}\sum_{t=1}^{T} \frac{\omega_t' \widehat{RCov}_t^{\theta} \omega_t}{\omega_t' RCov_t \omega_t}$
- Consider 15 different long-short strategies

More accurate corr forecasts $\Longrightarrow$ average risk targeting ratio close to 1

Application 3 - Risk Targeting

Consider a portfolio manager who allocates her funds into N risky assets based on a long-short trading strategy

- Set portfolio weight for stock i to $\omega_{i,t} = 1(-1)$ if the stock is in the long-leg (short-leg) of the strategy
  - Use simple HAR model for $\widehat{RV}_t$
- Average risk-targeting ratios across testing samples
  - $AvgRatio^{\theta} = \frac{1}{T}\sum_{t=1}^{T} \frac{\omega_t' \widehat{RCov}_t^{\theta} \omega_t}{\omega_t' RCov_t \omega_t}$
- Consider 15 different long-short strategies

More accurate corr forecasts $\Longrightarrow$ average risk targeting ratio close to 1

Application 3 - Risk Targeting

Consider a portfolio manager who allocates her funds into N risky assets based on a long-short trading strategy

- Set portfolio weight for stock i to $\omega_{i,t} = 1(-1)$ if the stock is in the long-leg (short-leg) of the strategy
  - Use simple HAR model for $\widehat{RV}_t$
- Average risk-targeting ratios across testing samples
  - $AvgRatio^{\theta} = \frac{1}{T}\sum_{t=1}^{T} \frac{\omega_t' \widehat{RCov}_t^{\theta} \omega_t}{\omega_t' RCov_t \omega_t}$
- Consider 15 different long-short strategies

More accurate corr forecasts $\Longrightarrow$ average risk targeting ratio close to 1

Application 3 - Risk Targeting

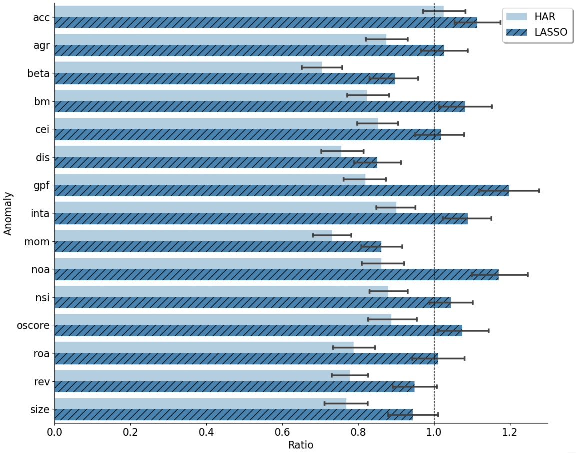
Consider a portfolio manager who allocates her funds into N risky assets based on a long-short trading strategy

- Set portfolio weight for stock i to $\omega_{i,t} = 1(-1)$ if the stock is in the long-leg (short-leg) of the strategy
  - Use simple HAR model for $\widehat{RV}_t$
- Average risk-targeting ratios across testing samples
  - $AvgRatio^{\theta} = \frac{1}{T}\sum_{t=1}^{T} \frac{\omega_t' \widehat{RCov}_t^{\theta} \omega_t}{\omega_t' RCov_t \omega_t}$
- Consider 15 different long-short strategies

More accurate corr forecasts $\Longrightarrow$ average risk targeting ratio close to 1

Application 3 - Risk Targeting

Risk-targeting ratios of long-short strategies



Application 4 - Global Minimum Variance Portfolio

Global Minimum Variance (GMV) portfolio is often used for evaluating covariance matrix forecasts

- Portfolio weights only depend on the covariance matrix, "clean" comparison
- Empirically achieve higher out-of-sample Sharpe ratios than MV optimized tangent portfolios (Jagannathan and Ma, 2003)

Calculate optimal portfolio weight vector $\omega_t^\theta = \frac{(\widehat{RCov}_t^\theta)^{-1}}{\mathbf{1}'(\widehat{RCov}_t^\theta)^{-1}\mathbf{1}}$, compare

- Portfolio returns $\omega_t^{\theta\prime} r_t$
- Realized portfolio risks $\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Portfolio Sharpe ratios $(\omega_t^{\theta\prime} r_t - r_{f,t})/\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Realized utility gains from switching forecasting models

Application 4 - Global Minimum Variance Portfolio

Global Minimum Variance (GMV) portfolio is often used for evaluating covariance matrix forecasts

- Portfolio weights only depend on the covariance matrix, “clean” comparison
- Empirically achieve higher out-of-sample Sharpe ratios than MV optimized tangent portfolios (Jagannathan and Ma, 2003)

Calculate optimal portfolio weight vector $\omega_t^\theta = \frac{(\widehat{RCov}_t^\theta)^{-1}}{\mathbf{1}'(\widehat{RCov}_t^\theta)^{-1}\mathbf{1}}$, compare

- Portfolio returns $\omega_t^{\theta\prime} r_t$
- Realized portfolio risks $\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Portfolio Sharpe ratios $(\omega_t^{\theta\prime} r_t - r_{f,t}) / \sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Realized utility gains from switching forecasting models

Application 4 - Global Minimum Variance Portfolio

Global Minimum Variance (GMV) portfolio is often used for evaluating covariance matrix forecasts

- Portfolio weights only depend on the covariance matrix, "clean" comparison
- Empirically achieve higher out-of-sample Sharpe ratios than MV optimized tangent portfolios (Jagannathan and Ma, 2003)

Calculate optimal portfolio weight vector $\omega_t^\theta = \frac{(\widehat{RCov}_t^\theta)^{-1}}{\mathbf{1}'(\widehat{RCov}_t^\theta)^{-1}\mathbf{1}}$, compare

- Portfolio returns $\omega_t^{\theta\prime} r_t$
- Realized portfolio risks $\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Portfolio Sharpe ratios $(\omega_t^{\theta\prime} r_t - r_{f,t}) / \sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Realized utility gains from switching forecasting models

Application 4 - Global Minimum Variance Portfolio

Global Minimum Variance (GMV) portfolio is often used for evaluating covariance matrix forecasts

- Portfolio weights only depend on the covariance matrix, "clean" comparison
- Empirically achieve higher out-of-sample Sharpe ratios than MV optimized tangent portfolios (Jagannathan and Ma, 2003)

Calculate optimal portfolio weight vector $\omega_t^\theta = \frac{(\widehat{RCov}_t^\theta)^{-1}}{\mathbf{1}'(\widehat{RCov}_t^\theta)^{-1}\mathbf{1}}$, compare

- Portfolio returns $\omega_t^{\theta\prime} r_t$
- Realized portfolio risks $\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Portfolio Sharpe ratios $(\omega_t^{\theta\prime} r_t - r_{f,t}) / \sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Realized utility gains from switching forecasting models

Application 4 - Global Minimum Variance Portfolio

Global Minimum Variance (GMV) portfolio is often used for evaluating covariance matrix forecasts

- Portfolio weights only depend on the covariance matrix, "clean" comparison
- Empirically achieve higher out-of-sample Sharpe ratios than MV optimized tangent portfolios (Jagannathan and Ma, 2003)

Calculate optimal portfolio weight vector $\omega_t^\theta = \frac{(\widehat{RCov}_t^\theta)^{-1}}{\mathbf{1}'(\widehat{RCov}_t^\theta)^{-1}\mathbf{1}}$, compare

- Portfolio returns $\omega_t^{\theta\prime} r_t$
- Realized portfolio risks $\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Portfolio Sharpe ratios $(\omega_t^{\theta\prime} r_t - r_{f,t}) / \sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Realized utility gains from switching forecasting models

Application 4 - Global Minimum Variance Portfolio

Global Minimum Variance (GMV) portfolio is often used for evaluating covariance matrix forecasts

- Portfolio weights only depend on the covariance matrix, "clean" comparison
- Empirically achieve higher out-of-sample Sharpe ratios than MV optimized tangent portfolios (Jagannathan and Ma, 2003)

Calculate optimal portfolio weight vector $\omega_t^{\theta} = \frac{(\widehat{RCov}_t^{\theta})^{-1}}{\mathbf{1}'(\widehat{RCov}_t^{\theta})^{-1}\mathbf{1}}$, compare

- Portfolio returns $\omega_t^{\theta\prime} r_t$
- Realized portfolio risks $\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^{\theta}}$
- Portfolio Sharpe ratios $(\omega_t^{\theta\prime} r_t - r_{f,t}) / \sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^{\theta}}$
- Realized utility gains from switching forecasting models

Application 4 - Global Minimum Variance Portfolio

Global Minimum Variance (GMV) portfolio is often used for evaluating covariance matrix forecasts

- Portfolio weights only depend on the covariance matrix, "clean" comparison
- Empirically achieve higher out-of-sample Sharpe ratios than MV optimized tangent portfolios (Jagannathan and Ma, 2003)

Calculate optimal portfolio weight vector $\omega_t^\theta = \frac{(\widehat{RCov}_t^\theta)^{-1}}{\mathbf{1}'(\widehat{RCov}_t^\theta)^{-1}\mathbf{1}}$, compare

- Portfolio returns $\omega_t^{\theta\prime} r_t$
- Realized portfolio risks $\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Portfolio Sharpe ratios $(\omega_t^{\theta\prime} r_t - r_{f,t}) / \sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Realized utility gains from switching forecasting models

Application 4 - Global Minimum Variance Portfolio

Global Minimum Variance (GMV) portfolio is often used for evaluating covariance matrix forecasts

- Portfolio weights only depend on the covariance matrix, "clean" comparison
- Empirically achieve higher out-of-sample Sharpe ratios than MV optimized tangent portfolios (Jagannathan and Ma, 2003)

Calculate optimal portfolio weight vector $\omega_t^\theta = \frac{(\widehat{RCov}_t^\theta)^{-1}}{\mathbf{1}'(\widehat{RCov}_t^\theta)^{-1}\mathbf{1}}$, compare

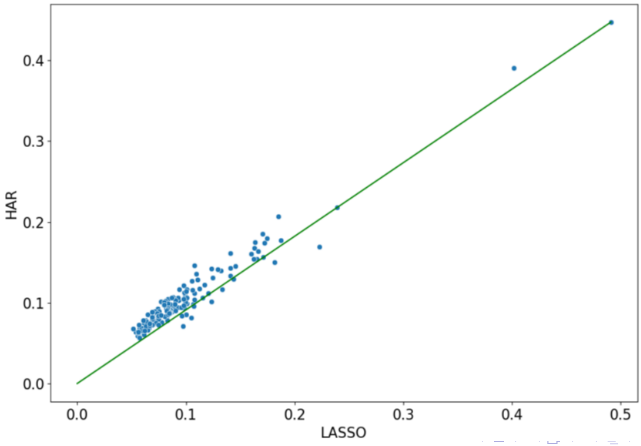
- Portfolio returns $\omega_t^{\theta\prime} r_t$
- Realized portfolio risks $\sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Portfolio Sharpe ratios $(\omega_t^{\theta\prime} r_t - r_{f,t}) / \sqrt{\omega_t^{\theta\prime} RCov_t \omega_t^\theta}$
- Realized utility gains from switching forecasting models

Application 4 - Global Minimum Variance Portfolio

	Mean Ret	St.Dev.	Sharpe Ratio	γ=2	γ=5	γ=10
HAR	10.27%	36.42%	0.36			
LASSO	10.90%	34.49%	0.48	0.77%	0.98%	1.37%

Application 4 - Global Minimum Variance Portfolio

	Mean Ret	St.Dev.	Sharpe Ratio	γ=2	γ=5	γ=10
HAR	10.27%	36.42%	0.36			
LASSO	10.90%	34.49%	0.48	0.77%	0.98%	1.37%



Application 4 - Beta-neutral GMV Portfolio

Consider a beta-neutral GMV portfolio following Cosemans et al. (2016)

Augment the traditional GMV optimization problem with the additional constraint that the portfolio's beta equals zero

$$\frac{\omega_t' \widehat{RCov}_t^{\theta} m_t}{m_t' \widehat{RCov}_t^{\theta} m_t} = 0$$

where m_t denotes the $N \times 1$ vector of firm market capitalization normalized to sum to unity

Compare returns, risks, Sharpe Ratios, and realized betas of the resulting beta-neutral GMV portfolios

Application 4 - Beta-neutral GMV Portfolio

Consider a beta-neutral GMV portfolio following Cosemans et al. (2016)

Augment the traditional GMV optimization problem with the additional constraint that the portfolio's beta equals zero

$$\frac{\omega_t' \widehat{RCov}_t^{\theta} m_t}{m_t' \widehat{RCov}_t^{\theta} m_t} = 0$$

where m_t denotes the $N \times 1$ vector of firm market capitalization normalized to sum to unity

Compare returns, risks, Sharpe Ratios, and realized betas of the resulting beta-neutral GMV portfolios

Application 4 - Beta-neutral GMV Portfolio

Consider a beta-neutral GMV portfolio following Cosemans et al. (2016)

Augment the traditional GMV optimization problem with the additional constraint that the portfolio's beta equals zero

$$\frac{\omega_t' \widehat{RCov}_t^{\theta} m_t}{m_t' \widehat{RCov}_t^{\theta} m_t} = 0$$

where m_t denotes the $N \times 1$ vector of firm market capitalization normalized to sum to unity

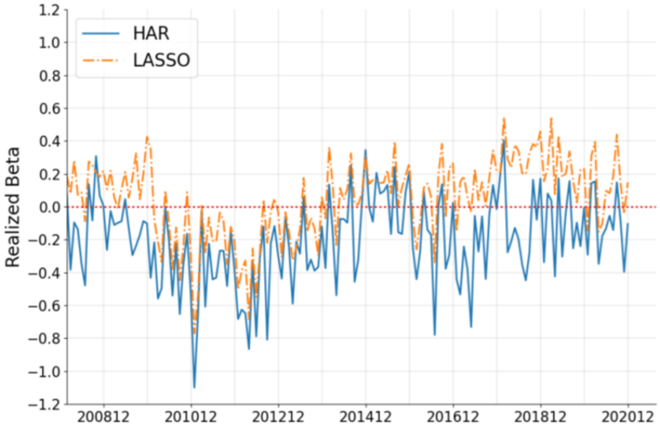
Compare returns, risks, Sharpe Ratios, and realized betas of the resulting beta-neutral GMV portfolios

Application 4 - Beta-neutral GMV Portfolio

	Mean Ret	St.Dev.	Sharpe Ratio	Realized Beta	γ=2	γ=5	γ=10
HAR	12.13%	51.17%	0.26	-0.20 (-7.19)			
LASSO	13.14%	43.12%	0.39	0.05 (1.58)	1.81%	3.36%	6.07%

Application 4 - Beta-neutral GMV Portfolio

	Mean Ret	St.Dev.	Sharpe Ratio	Realized Beta	γ=2	γ=5	γ=10
HAR	12.13%	51.17%	0.26	-0.20 (-7.19)			
LASSO	13.14%	43.12%	0.39	0.05 (1.58)	1.81%	3.36%	6.07%



Outline

- Data and Variables
- Estimation Methodology
- Out-of-sample Forecast Performance
- Applications
- **Robustness**

Subsample Analysis: Equal-Weighted

Panel A: Equal-weighted

	Model	Feature set	R^2_{OOS} relative to HAR		
			2008-2011	2012-2015	2016-2020
(1)	SHAR	$3\ RC^h + 3\ RC^{h-}$ (# of features = 6)	0.12%	0.33%	0.23%
(2)	SHAR-F	$3\ RC^h + 3\ RC^{h-}$ $+ 3\ FRC^h$ (# of features = 9)	2.34%	0.64%	1.97%
(3)	SHAR-F-Exp	$3\ RC^h + 3\ RC^{h-}$ $+ 3\ FRC^h$ $+ 4\ ExpRC^h + 4\ ExpRC^{h-}$ $+ 4\ ExpScRC^h + 4\ ExpScRC^{h-}$ (# of features = 25)	6.95%	9.95%	11.89%
(4)	LASSO	All 25 main features	7.87%	10.70%	11.51%

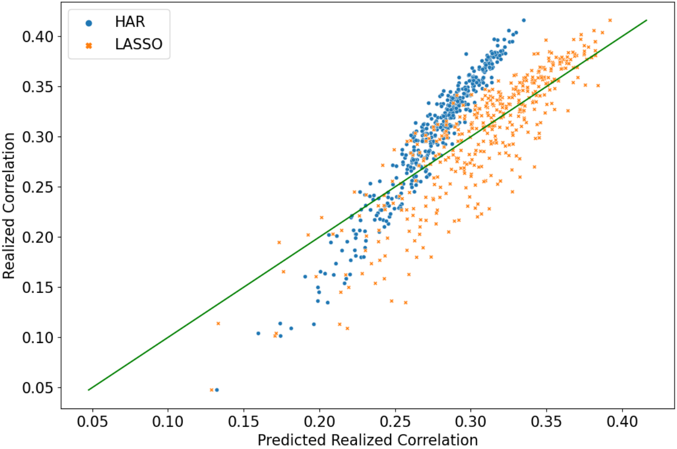
Subsample Analysis: Value-Weighted

Panel B: Value-weighted

	Model	Feature set	R^2_{OOS} relative to HAR 2008-2011	2012-2015	2016-2020
(1)	SHAR	$3\ RC^h + 3\ RC^{h-}$ (# of features = 6)	0.08%	0.25%	0.05%
(2)	SHAR-F	$3\ RC^h + 3\ RC^{h-} + 3\ FRC^h$ (# of features = 9)	2.24%	0.04%	1.44%
(3)	SHAR-F-Exp	$3\ RC^h + 3\ RC^{h-} + 3\ FRC^h + 4\ ExpRC^h + 4\ ExpRC^{h-} + 4\ ExpScRC^h + 4\ ExpScRC^{h-}$ (# of features = 25)	3.66%	9.40%	8.47%
(4)	LASSO	All 25 main features	5.76%	10.10%	8.31%

Subsample Analysis - Covid

Out-of-sample predictions during the peak of Covid



Alternative Features and Machine Learning Techniques

Consider six additional economically-motivated firm-linkage variables:

- Distance between two firms' headquarters (Parsons et al., 2020)
- Text-based network industry classifications (Hoberg and Phillips, 2010, 2016)
- Industry supply chain dependence (Menzly and Ozbas, 2010)
- Common analyst coverage (Israelsen, 2016)
- Common active mutual fund ownership (Antón and Polk, 2014)
- Common passive mutual fund ownership (Appel et al., 2016)

Alternative Features and Machine Learning Techniques

After turning firm-linkage variables into simple dummies using medians as cutoffs, construct two alternative feature sets:

- Original 25 features plus the 6 dummies
- Original 25 features plus the 150 additional features obtained by interacting each of the original features with the 6 dummy variables

Also consider the use of alternative machine learning algorithms:

- Ridge Regression (Ridge)
- Elastic Net (ENet)
- Principal Component Regression (PCR)
- Feed-forward Neural Networks (FNN)

Alternative Features and Machine Learning Techniques

After turning firm-linkage variables into simple dummies using medians as cutoffs, construct two alternative feature sets:

- Original 25 features plus the 6 dummies
- Original 25 features plus the 150 additional features obtained by interacting each of the original features with the 6 dummy variables

Also consider the use of alternative machine learning algorithms:

- Ridge Regression (Ridge)
- Elastic Net (ENet)
- Principal Component Regression (PCR)
- Feed-forward Neural Networks (FNN)

Alternative Features and Machine Learning Techniques

R^2_{OOS} relative to HAR

Panel A: Equal-weighted

Feature set	LASSO	Ridge	ENet	PCR	FNN
All 25 main features	10.16%	9.83%	10.14%	10.44%	10.12%
All 25 main features + 6 dummies (# of features = 31)	10.24%	9.96%	10.19%	9.61%	9.97%
All 25 main features + 150 feature × dummy combinations (# of features = 175)	10.35%	9.95%	10.31%	8.76%	9.88%

Panel B: Value-weighted

Feature set	LASSO	Ridge	ENet	PCR	FNN
All 25 main features	8.05%	7.31%	8.07%	8.31%	7.56%
All 25 main features + 6 dummies (# of features = 31)	8.05%	7.38%	8.09%	7.66%	6.98%
All 25 main features + 150 feature × dummy combinations (# of features = 175)	8.20%	7.54%	8.24%	7.68%	7.02%

Firm-link features do not have much incremental value; LASSO performs well relative to other algorithms

Conclusion

Use big data and machine learning to forecast realized correlation

- Feature engineering: build a large and novel feature set based on insights from various literature
- Scale of experiment: large in terms of stock universe and feature set
- OOS performance: improve R^2_{OOS}, triple pairs trading profit, enhance market equity premium prediction, produce ex-ante portfolio risk much closer to the realized risk, reduce risk of GMV portfolios

The same ideas and techniques could also be used in the construction of forecasting models for other commonly used risk measures, including measures of precision and factor risk exposures

Conclusion

Use big data and machine learning to forecast realized correlation

- Feature engineering: build a large and novel feature set based on insights from various literature

- Scale of experiment: large in terms of stock universe and feature set

- OOS performance: improve R^2_{OOS}, triple pairs trading profit, enhance market equity premium prediction, produce ex-ante portfolio risk much closer to the realized risk, reduce risk of GMV portfolios

The same ideas and techniques could also be used in the construction of forecasting models for other commonly used risk measures, including measures of precision and factor risk exposures

Conclusion

Use big data and machine learning to forecast realized correlation

- Feature engineering: build a large and novel feature set based on insights from various literature
- Scale of experiment: large in terms of stock universe and feature set
- OOS performance: improve R^2_{OOS}, triple pairs trading profit, enhance market equity premium prediction, produce ex-ante portfolio risk much closer to the realized risk, reduce risk of GMV portfolios

The same ideas and techniques could also be used in the construction of forecasting models for other commonly used risk measures, including measures of precision and factor risk exposures

Conclusion

Use big data and machine learning to forecast realized correlation

- Feature engineering: build a large and novel feature set based on insights from various literature
- Scale of experiment: large in terms of stock universe and feature set
- OOS performance: improve R^2_{OOS}, triple pairs trading profit, enhance market equity premium prediction, produce ex-ante portfolio risk much closer to the realized risk, reduce risk of GMV portfolios

The same ideas and techniques could also be used in the construction of forecasting models for other commonly used risk measures, including measures of precision and factor risk exposures

Conclusion

Use big data and machine learning to forecast realized correlation

- Feature engineering: build a large and novel feature set based on insights from various literature
- Scale of experiment: large in terms of stock universe and feature set
- OOS performance: improve R^2_{OOS}, triple pairs trading profit, enhance market equity premium prediction, produce ex-ante portfolio risk much closer to the realized risk, reduce risk of GMV portfolios

The same ideas and techniques could also be used in the construction of forecasting models for other commonly used risk measures, including measures of precision and factor risk exposures

Appendix - Correlation Signature Plot

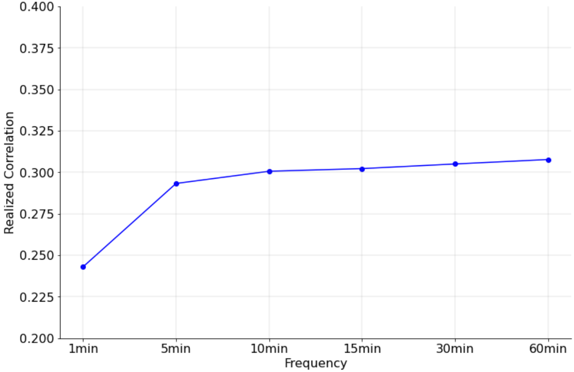
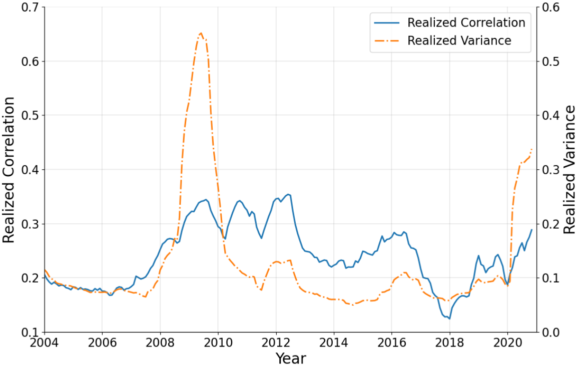


Figure A.1: Signature plots for monthly realized correlation

Appendix - Anomaly Characteristics

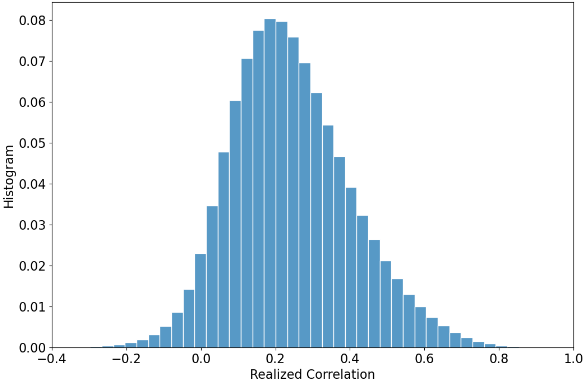
Variable	Acronym	Mean	Std	P1	P25	Median	P75	P99
Accruals	acc	0.00	0.04	-0.12	-0.01	0.00	0.02	0.11
Asset growth	agr	0.10	0.25	-0.30	-0.01	0.05	0.13	1.29
Beta	beta	1.04	0.51	0.12	0.67	0.97	1.31	2.63
Book-to-market	bm	0.47	0.42	-0.09	0.22	0.37	0.62	1.82
Composite equity issues	cei	-0.08	0.23	-0.75	-0.10	-0.06	-0.03	0.36
Distress	dis	-6.50	5.41	-8.57	-7.42	-6.86	-6.01	0.50
Gross profitability	gpf	0.30	0.23	-0.01	0.12	0.26	0.42	1.02
Investment-to-assets	inta	0.06	3.70	-0.17	0.01	0.03	0.06	0.39
Momentum	mom	0.13	0.37	-0.61	-0.06	0.11	0.28	1.31
Net operating assets	noa	0.53	0.35	-0.20	0.36	0.54	0.67	1.53
Net stock issues	nsi	0.13	0.93	-0.15	-0.03	0.00	0.01	3.09
O-score	oscore	-3.91	1.60	-7.64	-4.78	-3.95	-3.16	0.77
Return on assets	roa	0.01	0.02	-0.07	0.00	0.01	0.03	0.08
Reversal	rev	0.01	0.10	-0.25	-0.03	0.01	0.06	0.28
Size	size	16.20	1.24	13.23	15.36	16.21	17.04	19.09

Appendix - Response Variable



RV and *RC* exhibit different dynamic dependencies (RC relatively stable); justify modeling *RV* and *RC* separately

Appendix - Response Variable



Though the time series of realized correlations appear relatively stable, the unconditional distribution still reveals non-trivial variation

Appendix - (1) HAR Features

Extend HAR model by Corsi (2009) and Semi-HAR by Patton and Sheppard (2015) for volatility modelling to correlation forecasting

- RC_t^d, RC_t^w, RC_t^m: Lagged daily, weekly, monthly realized correlations constructed using 15-min mid-quote returns

  "HAR Model"

- RC_t^{d-}, RC_t^{w-}, RC_t^{m-}: Lagged daily, weekly, monthly realized negative semicorrelations constructed using negative returns only
  - contain different info; help improve portfolio risk forecast (Bollerslev et al. 2020, Econometrica)

  "SHAR Model"

Appendix - (1) HAR Features

Extend HAR model by Corsi (2009) and Semi-HAR by Patton and Sheppard (2015) for volatility modelling to correlation forecasting

- RC_t^d, RC_t^w, RC_t^m: Lagged daily, weekly, monthly realized correlations constructed using 15-min mid-quote returns

  "HAR Model"

- RC_t^{d-}, RC_t^{w-}, RC_t^{m-}: Lagged daily, weekly, monthly realized negative semicorrelations constructed using negative returns only
  - contain different info; help improve portfolio risk forecast (Bollerslev et al. 2020, Econometrica)

  "SHAR Model"

Appendix - (1) HAR Features

Extend HAR model by Corsi (2009) and Semi-HAR by Patton and Sheppard (2015) for volatility modelling to correlation forecasting

- RC_t^d, RC_t^w, RC_t^m: Lagged daily, weekly, monthly realized correlations constructed using 15-min mid-quote returns

  "HAR Model"

- RC_t^{d-}, RC_t^{w-}, RC_t^{m-}: Lagged daily, weekly, monthly realized negative semicorrelations constructed using negative returns only
  - contain different info; help improve portfolio risk forecast (Bollerslev et al. 2020, Econometrica)

  "SHAR Model"

Appendix - (2) Factor-driven Features

- Assume returns on N assets are driven by K common factors:

$$r = Lf + \varepsilon$$

return r is $N \times 1$, factor f is $K \times 1$, factor exposure L is $N \times K$

$$Cov(r) = LCov(f)L' + \Sigma_\varepsilon$$

- $LCov(f)L' + Diag(\Sigma_\varepsilon)$ factor-driven covariance matrix component

  factor-driven realized features are the off-diagonal elements from the correlation matrix of $LCov(f)L' + Diag(\Sigma_\varepsilon)$ (i.e., denoised lagged realized corr)

- Q: How do we obtain $LCov(f)L'$?
  - Existing method: construct HF factors (Fan, Furger, and Xiu, 2016)
  - New approach: uses low-freq firm char and $Cov(r)$; computationally more efficient

Appendix - (2) Factor-driven Features

- Assume returns on N assets are driven by K common factors:

$$r = Lf + \varepsilon$$

return r is $N \times 1$, factor f is $K \times 1$, factor exposure L is $N \times K$

$$Cov(r) = LCov(f)L' + \Sigma_{\varepsilon}$$

- $LCov(f)L' + Diag(\Sigma_{\varepsilon})$ factor-driven covariance matrix component

  factor-driven realized features are the off-diagonal elements from the correlation matrix of $LCov(f)L' + Diag(\Sigma_{\varepsilon})$ (i.e., denoised lagged realized corr)

- Q: How do we obtain $LCov(f)L'$?
  - Existing method: construct HF factors (Fan, Furger, and Xiu, 2016)
  - New approach: uses low-freq firm char and $Cov(r)$; computationally more efficient

Appendix - (2) Factor-driven Features

- Assume returns on N assets are driven by K common factors:

$$r = Lf + \varepsilon$$

return r is $N \times 1$, factor f is $K \times 1$, factor exposure L is $N \times K$

$$Cov(r) = LCov(f)L' + \Sigma_\varepsilon$$

- $LCov(f)L' + Diag(\Sigma_\varepsilon)$ factor-driven covariance matrix component

  factor-driven realized features are the off-diagonal elements from the correlation matrix of $LCov(f)L' + Diag(\Sigma_\varepsilon)$ (i.e., denoised lagged realized corr)

- Q: How do we obtain $LCov(f)L'$?
  - Existing method: construct HF factors (Fan, Furger, and Xiu, 2016)
  - New approach: uses low-freq firm char and $Cov(r)$; computationally more efficient

Appendix - (2) Factor-driven Features - Continue

- Use observable characteristics as factor loadings L to back out f (Fama and French, 2020)

$$r = Lf + \varepsilon$$

$$f = (L'L)^{-1}L'r$$

$$Cov(f) = (L'L)^{-1}L'Cov(r)L(L'L)^{-1}$$

$$LCov(f)L' = L(L'L)^{-1}L'Cov(r)L(L'L)^{-1}L'$$

- Use lagged daily, weekly, monthly realized $Cov(r)$ to back out $LCov(f)L'$ at three different speeds $\longrightarrow$ three factor-driven realized features, denoted by FRC^d, FRC^w, FRC^m
- Empirically, use 15 characteristics to construct L, including 11 mispricing anomalies from Stambaugh et al. (2012) + CAPM Beta, Size, Book-to-Market, and Reversal.
- Refer to model based on previous 6 realized features + 3 FRC features as SHAR-F model

Appendix - (2) Factor-driven Features - Continue

- Use observable characteristics as factor loadings L to back out f (Fama and French, 2020)

$$r = Lf + \varepsilon$$

$$f = (L'L)^{-1}L'r$$

$$Cov(f) = (L'L)^{-1}L'Cov(r)L(L'L)^{-1}$$

$$LCov(f)L' = L(L'L)^{-1}L'Cov(r)L(L'L)^{-1}L'$$

- Use lagged daily, weekly, monthly realized $Cov(r)$ to back out $LCov(f)L'$ at three different speeds $\longrightarrow$ three factor-driven realized features, denoted by FRC^d, FRC^w, FRC^m

- Empirically, use 15 characteristics to construct L, including 11 mispricing anomalies from Stambaugh et al. (2012) + CAPM Beta, Size, Book-to-Market, and Reversal.

- Refer to model based on previous 6 realized features + 3 FRC features as SHAR-F model

Appendix - (2) Factor-driven Features - Continue

- Use observable characteristics as factor loadings L to back out f (Fama and French, 2020)

$$r = Lf + \varepsilon$$

$$f = (L'L)^{-1}L'r$$

$$Cov(f) = (L'L)^{-1}L'Cov(r)L(L'L)^{-1}$$

$$LCov(f)L' = L(L'L)^{-1}L'Cov(r)L(L'L)^{-1}L'$$

- Use lagged daily, weekly, monthly realized $Cov(r)$ to back out $LCov(f)L'$ at three different speeds $\longrightarrow$ three factor-driven realized features, denoted by FRC^d, FRC^w, FRC^m
- Empirically, use 15 characteristics to construct L, including 11 mispricing anomalies from Stambaugh et al. (2012) + CAPM Beta, Size, Book-to-Market, and Reversal.
- Refer to model based on previous 6 realized features + 3 FRC features as SHAR-F model

Appendix - (2) Factor-driven Features - Continue

- Use observable characteristics as factor loadings L to back out f (Fama and French, 2020)

$$r = Lf + \varepsilon$$

$$f = (L'L)^{-1}L'r$$

$$Cov(f) = (L'L)^{-1}L'Cov(r)L(L'L)^{-1}$$

$$LCov(f)L' = L(L'L)^{-1}L'Cov(r)L(L'L)^{-1}L'$$

- Use lagged daily, weekly, monthly realized $Cov(r)$ to back out $LCov(f)L'$ at three different speeds $\longrightarrow$ three factor-driven realized features, denoted by FRC^d, FRC^w, FRC^m
- Empirically, use 15 characteristics to construct L, including 11 mispricing anomalies from Stambaugh et al. (2012) + CAPM Beta, Size, Book-to-Market, and Reversal.
- Refer to model based on previous 6 realized features + 3 FRC features as SHAR-F model

Appendix - (2) Factor-driven Features - Continue

- Use observable characteristics as factor loadings L to back out f (Fama and French, 2020)

$$r = Lf + \varepsilon$$

$$f = (L'L)^{-1}L'r$$

$$Cov(f) = (L'L)^{-1}L'Cov(r)L(L'L)^{-1}$$

$$LCov(f)L' = L(L'L)^{-1}L'Cov(r)L(L'L)^{-1}L'$$

- Use lagged daily, weekly, monthly realized $Cov(r)$ to back out $LCov(f)L'$ at three different speeds $\longrightarrow$ three factor-driven realized features, denoted by FRC^d, FRC^w, FRC^m
- Empirically, use 15 characteristics to construct L, including 11 mispricing anomalies from Stambaugh et al. (2012) + CAPM Beta, Size, Book-to-Market, and Reversal.
- Refer to model based on previous 6 realized features + 3 FRC features as SHAR-F model

Appendix - (2) Factor-driven Features - Continue

- Use observable characteristics as factor loadings L to back out f (Fama and French, 2020)

$$r = Lf + \varepsilon$$

$$f = (L'L)^{-1}L'r$$

$$Cov(f) = (L'L)^{-1}L'Cov(r)L(L'L)^{-1}$$

$$LCov(f)L' = L(L'L)^{-1}L'Cov(r)L(L'L)^{-1}L'$$

- Use lagged daily, weekly, monthly realized $Cov(r)$ to back out $LCov(f)L'$ at three different speeds $\longrightarrow$ three factor-driven realized features, denoted by FRC^d, FRC^w, FRC^m
- Empirically, use 15 characteristics to construct L, including 11 mispricing anomalies from Stambaugh et al. (2012) + CAPM Beta, Size, Book-to-Market, and Reversal.
- Refer to model based on previous 6 realized features + 3 FRC features as SHAR-F model

Appendix - (2) Factor-driven Features - Continue

- Use observable characteristics as factor loadings L to back out f (Fama and French, 2020)

$$r = Lf + \varepsilon$$
$$f = (L'L)^{-1}L'r$$
$$Cov(f) = (L'L)^{-1}L'Cov(r)L(L'L)^{-1}$$
$$LCov(f)L' = L(L'L)^{-1}L'Cov(r)L(L'L)^{-1}L'$$

- Use lagged daily, weekly, monthly realized $Cov(r)$ to back out $LCov(f)L'$ at three different speeds $\longrightarrow$ three factor-driven realized features, denoted by FRC^d, FRC^w, FRC^m
- Empirically, use 15 characteristics to construct L, including 11 mispricing anomalies from Stambaugh et al. (2012) + CAPM Beta, Size, Book-to-Market, and Reversal.
- Refer to model based on previous 6 realized features + 3 FRC features as SHAR-F model

Appendix - (3) EMA Features

- $ExpRC^{d}$, $ExpRC^{w}$, $ExpRC^{m}$, $ExpRC^{q}$, $ExpRC^{d-}$, $ExpRC^{w-}$, $ExpRC^{m-}$, $ExpRC^{q-}$:
  Exponential moving average of lagged daily realized correlations and negative semicorrelations with half-life between one day and one quarter

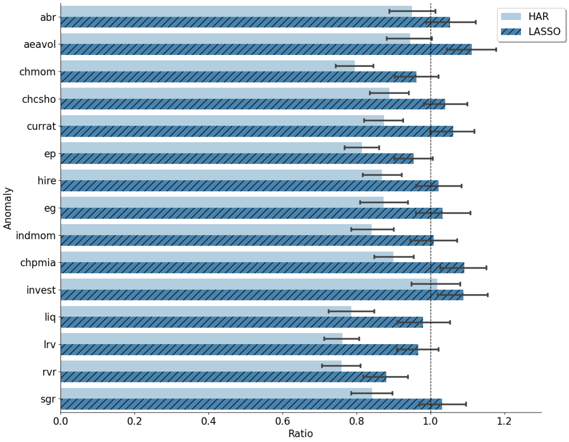
- $ExpScRC^{d}$, $ExpScRC^{w}$, $ExpScRC^{m}$, $ExpScRC^{q}$, $ExpScRC^{d-}$, $ExpScRC^{w-}$, $ExpScRC^{m-}$, $ExpScRC^{q-}$:
  Exponential moving average of lagged within-sector average realized correlations to exploit stronger within-sector correlation

- Denote SHAR-F model with all EMA features as "SHAR-F-Exp" model

Appendix - Additional Anomaly Characteristics

Variable	Acronym	Mean	Std	P1	P25	Median	P75	P99
Abnormal earnings announcement return	abr	0.00	0.02	-0.05	-0.01	0.00	0.01	0.05
Abnormal earnings announcement volume	aeavol	0.87	0.96	-0.35	0.26	0.65	1.20	4.50
Change in 6-month momentum	chmom	0.01	0.37	-0.86	-0.17	-0.01	0.17	1.08
Change in shares outstanding	chcsho	0.04	0.22	-0.14	-0.02	0.00	0.01	1.05
Current ratio	currat	2.57	4.65	0.50	1.09	1.53	2.34	24.58
Earnings to price	ep	0.03	0.22	-0.56	0.03	0.05	0.07	0.17
Employee growth rate	hire	0.04	0.17	-0.38	-0.02	0.02	0.08	0.72
Expected growth	eg	0.00	0.02	-0.05	0.00	0.00	0.01	0.05
Industry momentum	indmom	0.12	0.29	-0.48	-0.04	0.11	0.24	1.11
Industry-adjusted change in profit margin	chpmia	0.52	7.43	-15.81	-0.17	0.00	0.12	37.83
Investment	invest	1.00	0.45	0.30	0.85	0.98	1.13	1.99
Liquidity	liq	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Long-term reversals	lrv	0.33	0.72	-0.76	-0.03	0.24	0.54	2.71
Residual variance	rvr	0.04	0.02	0.02	0.02	0.03	0.04	0.11
Sales growth	sgr	0.08	0.22	-0.44	0.00	0.06	0.13	0.83

Appendix - Additional Risk-targeting Ratios



Appendix - Correcting Non-positive Definite Matrices

Challenge: 10% of the LASSO-based correlation matrix forecasts in our sample are not positive definite

Solution: apply a simple convexity correction on any non-positive-definite correlation matrix prediction

- $\widehat{R}_t^{LASSO*} = \alpha \widehat{R}_t^{HAR} + (1-\alpha)\widehat{R}_t^{LASSO}$
- choose the minimum value of $\alpha > 0$ s.t. $\widehat{R}_t^{LASSO*}$ is P.D.

Importantly, however, our GMV-related model comparison results remain robust to the exclusion of Non-P.D. months

Appendix - Traditional Firm-linkage Measures

Variable	Definition
ZipDist	Zip code distance between two firms' headquarters
TNIC3	Text-based Network Industry Classifications based on firm pairwise similarity scores from text analysis of firm 10-K product descriptions
IndSuppDep	Industry supply chain dependence measured by fraction of industry-by-industry purchases from input-output tables
CmnAnalys	Common analyst coverage as # of common analysts following the stock pair over # of total unique analysts
CmnActOwn	Common active mutual fund ownership defined as the total dollar value of a stock pair held by common active mutual funds over the total dollar value of shares outstanding for the stock pair
CmnPssOwn	Common passive mutual fund ownership defined as the total dollar value of a stock pair held by common passive mutual funds over the total dollar value of shares outstanding for the stock pair