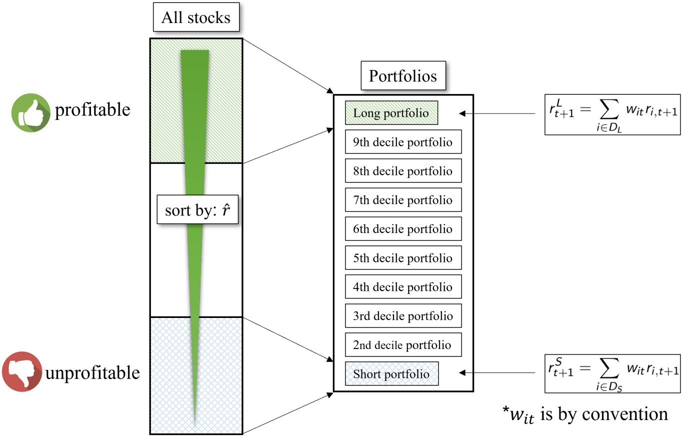
Cross-Sectional Analysis of Conditional Stock Returns: Quantile Regression with Machine Learning

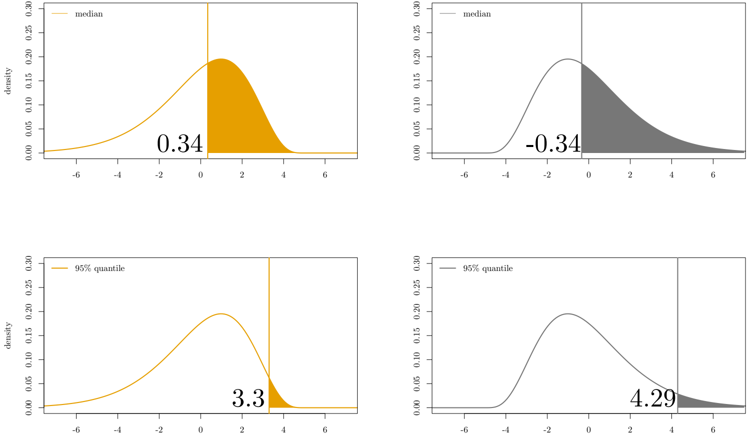
Haitao Li   Guoliang Ma   Cindy Yu

June 3, 2022

Asset pricing and long-short investing.
Key to profitable investment: **profitability indicator (sorting var.)**



Expectation is not enough



1. Setup and Goals

Setup

Long-short investment that requires profitability indicators (sorting variables).

Goals

- To recover conditional distributional information.
- To identify indicators (sorting variables) using distributional information.

2. Methodology: Distributional Prediction

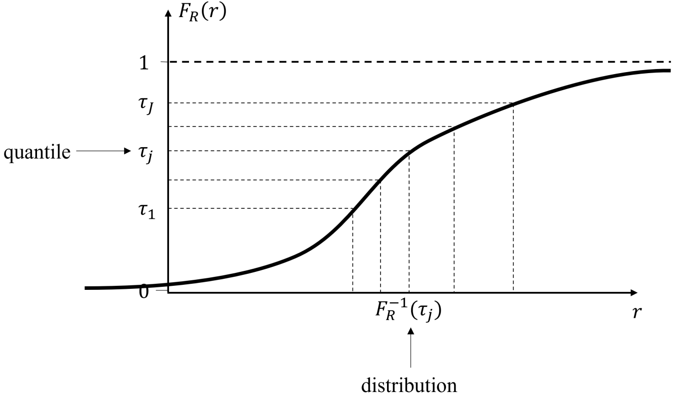
Distributions prediction in the literature

- specifying distributions (GARCH):
  - Kon (1984), Nelson (1991), Peiro (1994), Baixauli and Alvarez (2004), Kelly and Jiang (2014), and Hohberg, Pütz, and Kneib (2020)
- specifying local information then aggregate:
  - Quantiles: Machado and Mata (2005) and Zhao (2013)
  - Local odds ratio: Foresi and Peracchi (1995) and Anatolyev and Baruník (2019)

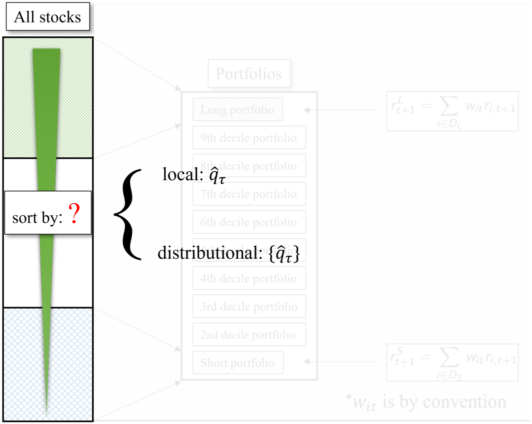
Benefits of quantile regression with machine learning

- No restrictions on distributions
- Ability to handle non-linear relationships

Recover distributions from quantiles with inverse probability transform



2. Methodology: Forming profitability indicators



Local quantile information

$$\hat{q}_{t,\tau_j}(\boldsymbol{x}_{i,t}),\ j = 1, \dots, J = 100. \tag{1}$$

2. Methodology: Forming profitability indicators

Distributional information

$$\Pr\left[r_{i,t+1} \geq r_c | \boldsymbol{x}_{i,t}\right] = \frac{1}{J} \sum_{j=1}^{J} \mathbb{1}(\hat{q}_{t,\tau_j}(\boldsymbol{x}_{i,t}) \geq r_c). \quad (2)$$

$$\mathbb{E}_t^{QM}\left[r_{i,t+1} | \boldsymbol{x}_{i,t}\right] = \frac{1}{J} \sum_{j=1}^{J} \hat{q}_{t,\tau_j}(\boldsymbol{x}_{i,t}). \quad (3)$$

2. Methodology: Quantile Regression

r_{it+1} is the stock return for firm i at time $t+1$ and $\boldsymbol{x}_{i,t}$ is the firm level characteristic at time t.

Existing work with Squared Error Loss

1. Linear regression

$$\hat{\beta}_t = \min_{\beta \in \mathbb{R}^p} \sum_{s \in \mathcal{T}_t} \sum_{i=1}^{N_s} (r_{i,s+1} - \beta^\top \boldsymbol{x}_{i,s})^2 \qquad (4)$$

$$\mathbb{E}[r_{it+1} | \boldsymbol{x}_{it}] = \hat{\beta}_t^\top \boldsymbol{x}_{it}.$$

2. Nonlinear machine learning

$$\hat{f}_t = \min_{f \in \mathcal{F}} \sum_{s \in \mathcal{T}_t} \sum_{i=1}^{N_s} (r_{i,s+1} - f(\boldsymbol{x}_{i,s}))^2 \qquad (5)$$

$$\mathbb{E}[r_{it+1} | \boldsymbol{x}_{it}] = \hat{f}_t(\boldsymbol{x}_{it}).$$

2. Methodology: Quantile Regression

Conditional quantile

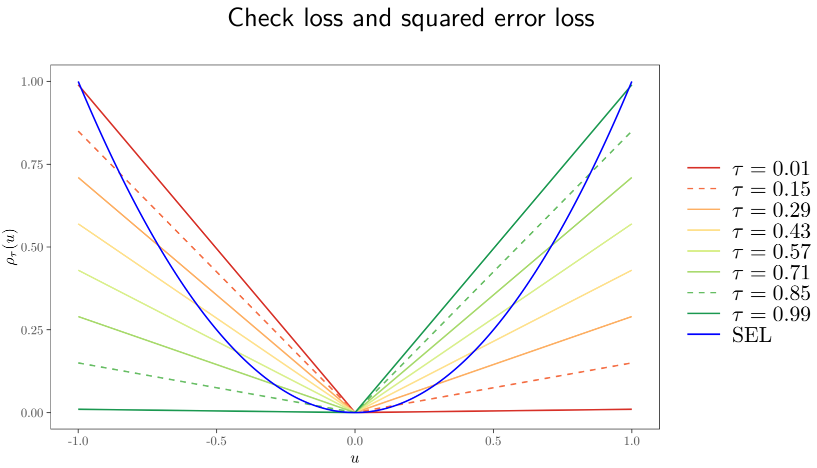
$q_{t,\tau}(\boldsymbol{x}_{i,t})$ is defined as a function satisfying

$$\tau = \Pr\left[r_{i,t+1} < q_{t,\tau}(\boldsymbol{x}_{i,t}) | \boldsymbol{x}_{i,t}\right]. \tag{6}$$

Proposed work with Check Loss

$$\hat{q}_{t,\tau}(\cdot) = \min_{q \in \mathcal{F}} \sum_{s \in \mathcal{T}_t} \left\{ \sum_{i=1}^{N_s} \rho_\tau \left(r_{i,s+1} - q(\boldsymbol{x}_{i,s}) \right) \right\}, \tag{7}$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}_{u<0})$ and $\hat{q}_{t,\tau}(\boldsymbol{x}_{i,t})$ is the predicted τ-th quantile of conditional distribution.

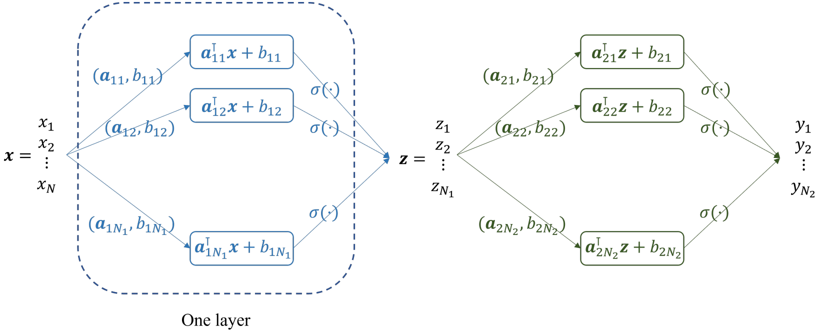

Check loss and squared error loss
1.00
0.75
0.50
0.25
0.00
$\rho_\tau(u)$
-1.0
-0.5
0.0
0.5
1.0
u
$\tau = 0.01$
$\tau = 0.15$
$\tau = 0.29$
$\tau = 0.43$
$\tau = 0.57$
$\tau = 0.71$
$\tau = 0.85$
$\tau = 0.99$
SEL

2. Methodology: Machine Learning

- Principal component regression (PCR)
- Lasso
- Neural network
- Boosting trees (LightGBM of Ke et al. (2017))

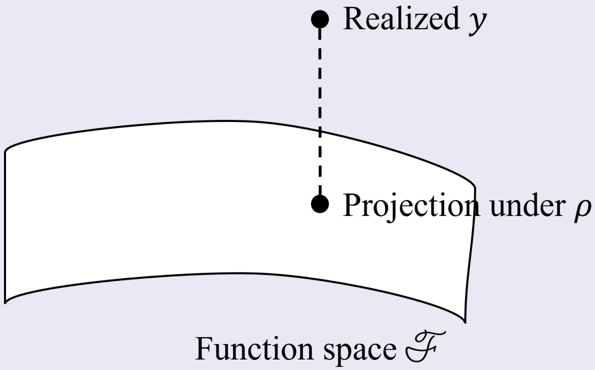
2. Methodology: Machine Learning

Neural network

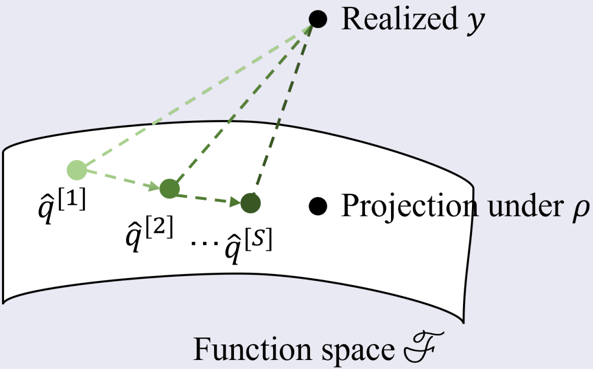


Our neural network has three layers of size 64, 16, and 8.

Boosting trees



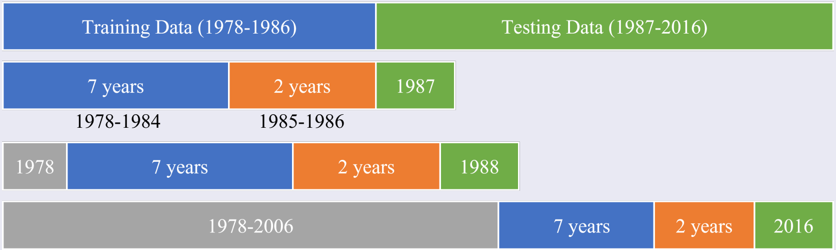
Boosting trees



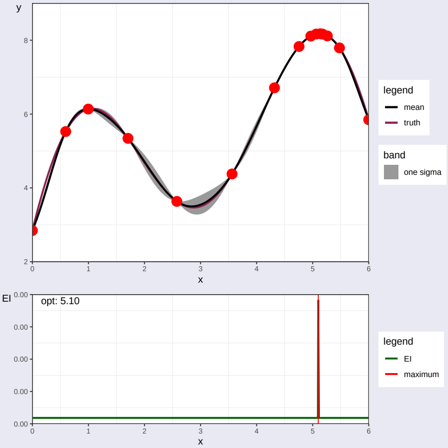
2. Methodology: Machine Learning

Training for cross sections

7+2+1 Training Scheme



Bayesian optimization with Gaussian process



3. Simulation

Setup

$$\begin{aligned} r_{i,t+1} &= g^*(\boldsymbol{x}_{i,t}) + \beta_{t+1}^{\intercal}\boldsymbol{x}_{i,t} + \varepsilon_{i,t+1}, \\ \beta_{t+1} &\overset{iid}{\sim} \mathcal{N}\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_3\right), \\ \varepsilon_{i,t+1} &\text{ is the random error to be specified} \end{aligned} \tag{8}$$

Data

- 200 firms
- 180 months
- 36,000 total observations
- 50 covariates
- 14,400 observations for prediction

Main driver

$$g^*(\boldsymbol{x}_{i,t}) = 0.3 + 0.3x_{i,t,1} + 0.2x_{i,t,2} + w_t x_{i,t,3}, \tag{9a}$$

$$\boxed{g^*(\boldsymbol{x}_{i,t}) = 1 + 0.4x_{i,t,1} + 2x_{i,t,1}x_{i,t,2}^2.} \tag{9b}$$

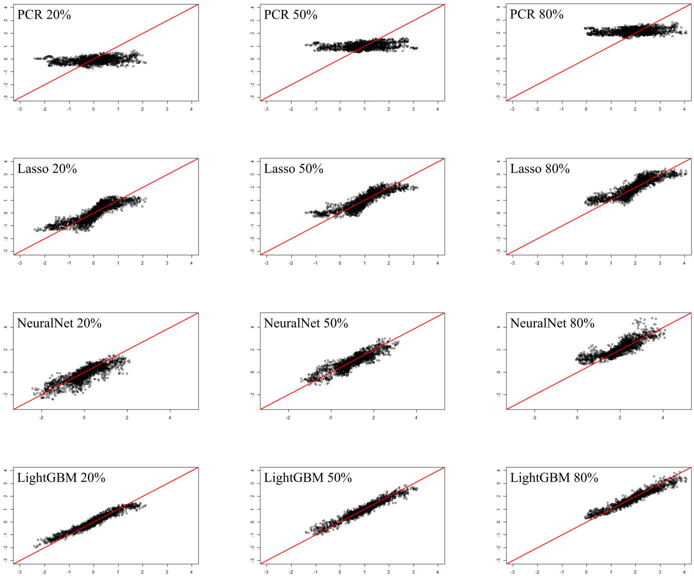
$$w_t = \rho w_{t-1} + u_t, u_t \overset{iid}{\sim} \mathcal{N}\left(0, 1-\rho^2\right), \text{ for } \rho = 0.9. \tag{10}$$

Firm-time-specific error

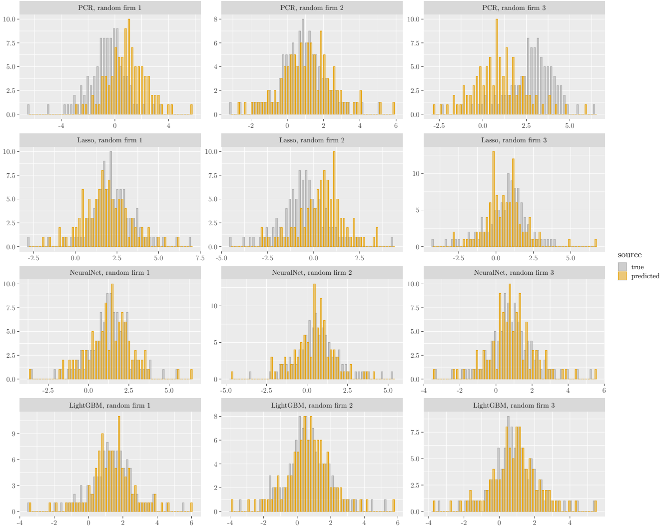
$$\varepsilon_{i,t+1} \overset{iid}{\sim} t_5(\sigma_\varepsilon^2), \sigma_\varepsilon = 0.05, \tag{11a}$$

$$\boxed{\varepsilon_{i,t+1} \sim ALD(\kappa_i), \kappa_i \sim \mathcal{N}(1, 0.15).} \tag{11b}$$

3. Simulation Results (Local)



3. Simulation Results (Distributional)



4. Empirical application

U.S. stock market

- From 1978 to 2016
- 1,809,418 monthly observations
- 106 covariates

$F^{-1}(\tau|\boldsymbol{x}_{i,t}) = q_{t,\tau}(\boldsymbol{x}_{i,t})$. Draw τ_j $(j = 1, 2, \ldots, J) \sim \mathcal{U}_{[0,1]}$, $q_{t,\tau_j}(\boldsymbol{x}_{i,t})$ are a random sample from the conditional density $f(r_{i,t+1}|\boldsymbol{x}_{i,t})$.

Proposed profitability indicators

- (A) Median: $\hat{q}_{t,0.5}(\boldsymbol{x}_{i,t})$
- (B) Quantile mean: $\mathbb{E}_t^{QM}[r_{i,t+1}|\boldsymbol{x}_{i,t}] = \frac{1}{J}\sum_{j=1}^{J} \hat{q}_{t,\tau_j}(\boldsymbol{x}_{i,t})$

Evaluating density forecast

Table 1: **Summary statistics of *p*-values**

ML	test	min	Q1	median	Q3	max	mean	$P(>0.05)$
PCR	KS	0.0000	0.0652	0.2341	0.5167	0.9997	0.3138	0.7864
	Shapiro	0.0000	0.0733	0.2464	0.5198	1.0000	0.3209	0.8018
	LR*	0.0000	0.0301	0.2018	0.5328	1.0000	0.3044	0.7020
Lasso	KS	0.0000	0.0314	0.1706	0.4626	0.9998	0.2736	0.6992
	Shapiro	0.0000	0.0677	0.2477	0.5319	0.9998	0.3253	0.7871
	LR	0.0000	0.0107	0.1272	0.4314	1.0000	0.2525	0.6239
NeuralNet	KS	0.0000	0.0554	0.2072	0.4762	0.9997	0.2920	0.7633
	Shapiro	0.0000	0.0812	0.2459	0.5046	0.9999	0.3189	0.7956
	LR	0.0000	0.0399	0.2290	0.5470	1.0000	0.3173	0.7049
LightGBM	KS	0.0000	0.0783	0.2571	0.5312	0.9995	0.3262	0.8096
	Shapiro	0.0000	0.1131	0.3030	0.5709	0.9999	0.3598	0.8614
	LR	0.0000	0.0821	0.3003	0.6083	1.0000	0.3623	0.8033

*Likelihood ratio test proposed in Berkowitz (2001).

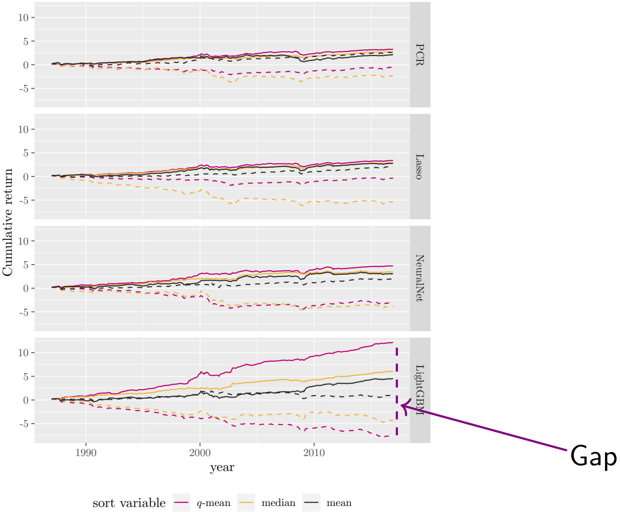
4. Empirical application

Larger gap is better

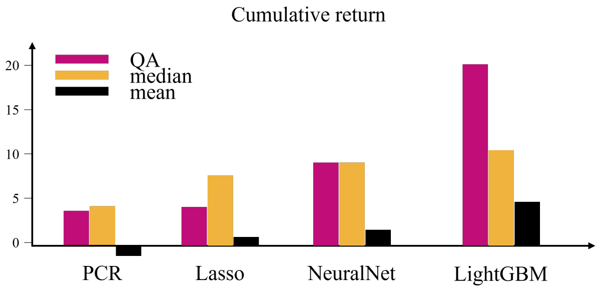
proposal A

proposal B

traditional

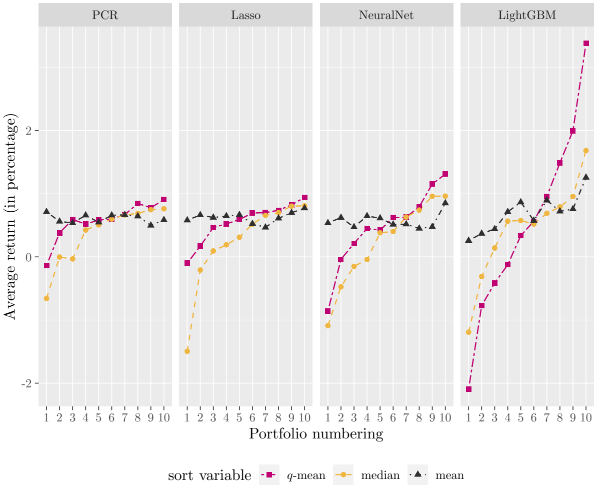


4. Empirical application



4. Empirical application

Clearer monotonic trend is better



5. Conclusion

- We extend the empirical asset pricing literature by providing a cross-sectional analysis of conditional stock returns through the lens of conditional quantiles and distributions.
- We develop machine learning methods to forecast the conditional quantiles of stock returns in the cross section through quantile regression.
- Empirical results based on the U.S. data show that measures constructed from conditional distributions can identify stocks with extreme positive or negative returns and achieve superior performance in long-short investing.

Thank you!

Reference I

Anatolyev, Stanislav and Jozef Baruník. 2019. "Forecasting dynamic return distributions based on ordered binary choice." *International Journal of Forecasting* 35 (3):823–835.

Baixauli, J Samuel and Susana Alvarez. 2004. "Analysis of the conditional stock-return distribution under incomplete specification." *European Journal of Operational Research* 155 (2):276–283.

Berkowitz, Jeremy. 2001. "Testing density forecasts, with applications to risk management." *Journal of Business & Economic Statistics* 19 (4):465–474.

Foresi, Silverio and Franco Peracchi. 1995. "The conditional distribution of excess returns: An empirical analysis." *Journal of the American Statistical Association* 90 (430):451–466.

Hohberg, Maike, Peter Pütz, and Thomas Kneib. 2020. "Treatment effects beyond the mean using distributional regression: Methods and guidance." *PloS one* 15 (2):e0226514.

Reference II

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A highly efficient gradient boosting decision tree." *Advances in Neural Information Processing Systems* 30:3146–3154.

Kelly, Bryan and Hao Jiang. 2014. "Tail risk and asset prices." *The Review of Financial Studies* 27 (10):2841–2871.

Kon, Stanley J. 1984. "Models of stock returns—a comparison." *The Journal of Finance* 39 (1):147–165.

Machado, José AF and José Mata. 2005. "Counterfactual decomposition of changes in wage distributions using quantile regression." *Journal of Applied Econometrics* 20 (4):445–465.

Nelson, Daniel B. 1991. "Conditional heteroskedasticity in asset returns: A new approach." *Econometrica* :347–370.

Peiro, Amado. 1994. "The distribution of stock returns: International evidence." *Applied Financial Economics* 4 (6):431–439.

Reference III

Zhao, Yizhen. 2013. "Forecasting the stock return distribution using macro-finance variables." *Job Market Paper* .

Advantage of quantile mean sorting

Advantageous q-mean