Optimal Model Averaging for High-dimensional Predictive Quantile Regression with Persistent Covariates

SONG-YEN CHEN & TZEE-MING HUANG

NCCU

April 29, 2024

Reborn from the old version: optimal model averaging for ultra-high dimensional quantile regression with diverging covariates (2023) (rejected by JMA)

This paper:

- extend the independent case to time series case with possibly high-persistent covariates
- propose the new variable screening and model averaging approaches

PROLOGUE

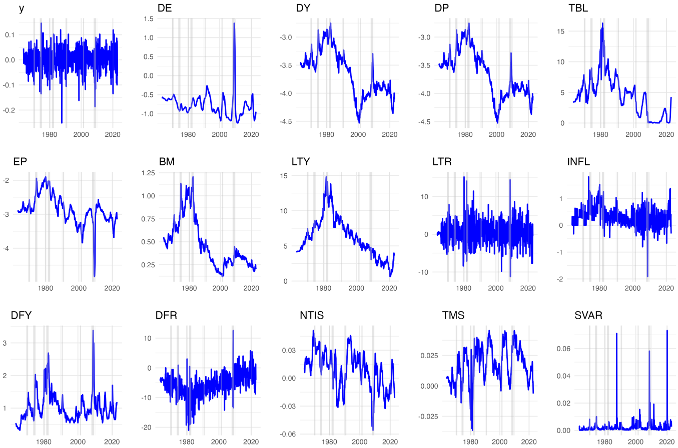
Why "persistence" matters ? (in time series)

- terminology: *highly persistent*, *near unit-root*, *local to unity*, *near stationary*,...,etc.
- recall: recovering the $I(0)$ stationarity by differencing is a common technique, however, something is ambiguous...
- over-differencing: lost of information and a distorted pattern, inaccurate estimates and reducing the predictability of power.
- not all predictors can be differenced : lost of the original meaning, e.g., long-short portfolios, technical indicators, momentum indicators,...,etc.

Var.	Mean	S.D.	P25	P50	P75	Sk.	Kur.	ADF.pv	$AC_{(1)}$	Corr(y, var.)
DP	-3.64	0.41	-3.97	-3.57	-3.36	0.02	2.08	0.36	0.99	-0.10
DY	-3.64	0.41	-3.96	-3.56	-3.35	0.01	2.10	0.38	0.99	0.01
DE	-0.77	0.31	-0.94	-0.81	-0.62	2.93	19.48	0.01	0.99	0.01
EP	-2.87	0.43	-3.11	-2.90	-2.67	-0.51	5.43	0.02	0.99	-0.10
BM	0.47	0.26	0.27	0.39	0.64	0.90	2.83	0.48	0.99	-0.09
TBL(%)	4.46	3.26	1.69	4.64	6.13	0.66	3.63	0.20	0.99	-0.10
LTY(%)	6.22	2.89	4.23	5.99	8.02	0.51	2.98	0.44	0.99	-0.08
LTR(%)	0.56	3.02	-1.22	0.38	2.31	0.37	5.24	0.01	0.06	0.10
INFL(%)	0.32	0.36	0.10	0.30	0.52	0.01	5.85	0.01	0.58	-0.11
DFY(%)	1.02	0.44	0.73	0.91	1.20	1.74	7.45	0.01	0.96	0.02
DFR(%)	-5.62	3.90	-7.77	-5.54	-3.36	-0.28	4.86	0.01	0.53	0.24
NTIS	0.01	0.02	-0.01	0.01	0.02	-0.48	2.92	0.01	0.98	-0.04
TMS	0.02	0.01	0.01	0.02	0.03	-0.24	2.70	0.04	0.96	0.07
RVOL	0.14	0.06	0.09	0.13	0.18	0.46	3.47	0.01	0.96	0.05

Note: DP: dividend price ratio, DY: dividend yield ratio, EP: earnings price ratio, BM: book-to-market ratio, DFY: default yield spread, NTIS: net equity expansion, LTY: long term yield, TBL: treasury bill rates, SVAR: stock sample variation, DFR: default return spread, LTR: long term rate of returns, INFL: inflation.

Figure: Time series plots of some common financial predictors



PROLOGUE

- e.g., Campbell and Yogo (2006); Phillips and Magdalinos (2007): $X_t = \phi X_{t-1} + v_t$, $X_0 = o_p(n)$, $t = 1, 2, \cdots, n$, where

$$\phi = 1 + \frac{c}{n}, \text{ for some } \alpha \in [0,1],$$

  1. stationary and weakly persistent: ϕ is fixed and bounded away from unity, $|\phi| \ll 1$.
  2. stationary but strongly persistent: $|\phi| < 1$, $c \in (-\infty, 0)$.

- Consider a classical low-dimensional $(p < n)$ one-step ahead linear predictive regression model:

$$Y_t = X_{t-1}^{\top}\beta + u_t,$$

where $w_t := (u_t, v_t)^{\top} \sim$ martingale differences sequence with $\mathbb{E}[w_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[w_t w_t^{\top}|\mathcal{F}_{t-1}] = \Sigma$.

PROLOGUE

- Cavanagh et al. (1995); Stambaugh (1999):

$$\mathbb{E}[\widehat{\beta}_{LS} - \beta] = -\frac{\Sigma_{uv}}{\Sigma_{vv}}(1 + 3\phi/n) + O(n^{-2}).$$

$$n(\widehat{\beta}_{LS} - \beta) \Rightarrow \mathcal{N}(0, \frac{(\Sigma_{uu} - \Sigma_{uv}\Sigma_{vv}^{-1}\Sigma_{vu}}{(\int_0^1 J_c(r)^2 dr)}) + (\frac{\Sigma_{uv}}{\Sigma_{vv}})\frac{\int_0^1 J_c(r)dW_v(r)}{\sqrt{\int_0^1 J_c(r)^2 dW_v(r)}}.$$

- $$t_{\widehat{\beta}_{LS}} = \frac{\widehat{\beta}_{LS} - \beta}{\widehat{\sigma}_{\widehat{\beta}}} \Rightarrow \sqrt{(1 - \delta^2)}Z + \delta\frac{\int_0^1 J_c(r)dW_v(r)}{\sqrt{\int_0^1 J_c(r)^2 dW_v(r)}},$$

where $\delta = \Sigma_{uv}/(\Sigma_{uu}\Sigma_{vv})^{1/2}$, known as the embedded endogeneity, Z is the standard Gaussian, $J_c(r) = \int_0^r e^{(r-s)c}dW_v(s)$ is an Ornstein-Uhlenbeck process, and $W_v(r)$ is s standard Brownian motion with the process v_t.

- Only if $\Sigma_{uv} = 0$, $\hat{\beta}_{LS}$ is asymptotically unbiased, consistent, and asymptotic normality. In general, $c \neq 0$ or/and $\delta \neq 0$.
- For a predictive linear quantile regression model, for a given quantile $\tau \in (0, 1)$,

$$Q_\tau(Y_t|\mathcal{F}_{t-1}) = X_{t-1}^\top \beta(\tau),\ \mathbb{P}(Y_t \leq Q_\tau(Y_t|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}) = \tau,$$

similarly,

$$t_{\widehat{\beta}(\tau)} = \frac{\widehat{\beta}(\tau) - \beta}{\widehat{\sigma}_{\widehat{\beta}(\tau)}} \Rightarrow \sqrt{(1 - \delta(\tau)^2)} Z + \delta(\tau) OU(c),$$

where $\delta(\tau) = -corr(\boldsymbol{I}(u_t(\tau) < 0) v_t)$ and $OU(c)$ is an Ornstein-Uhlenbeck process, following Lee (2016).
- inaccurate estimates and prediction, wrong statistical inferences.

HD QMA

- (HD MA):
  - When $p \gg n$: overall model space is huge. 2^p possible submodels.
  - Ando and Li (2014; 2017): first proposed a leave-one-out cross validation (LOOCV)-based MA for high dimensional linear regression and generalized linear model after pre-screening, allowing the number of covariates increase as the sample size increases.
  - model screening procedure proposed by Fan and Lv (2008) that pre-screen models using partial linear correlation ranking procedure.
- (QMA & HD QMA):
  - (QMA): difficulty: non-smoothness of the objective function and the lack of the bias-variance decomposition for MSE/MSFE criterion.
  - Most of the existing literature use cross-validation (CV) based criterion, e.g., (Quantile): Lu and Su (2015, LOOCV), (Expectile): Tu and Wang (2020, CV); Bai et al. (2022, J-fold CV); Zhan et al. (2022, LOOCV); Wang et al. (2023).
  - we extend to the time series framework based on high dimensional predictive quantile regression with persistent covariates.

Model Framework

- Recall: linear conditional quantile regression:

$$Y_t = \boldsymbol{X}_t'\beta(\tau) + u_t(\tau) \;\; \mathbb{P}(u_t \leq 0|\boldsymbol{X}_t) = \tau \in (0,1).$$

- The QR estimator is given by:

$$\widehat{\beta}(\tau) = \arg\min_{\beta} n^{-1} \sum_{i=1}^{n} \rho_\tau(Y_t - \boldsymbol{X}_t'\beta(\tau)),$$

$\rho_\tau(t) := [\tau - \boldsymbol{I}(t \leq 0)]t$ is a check function.

• Remark

The L^q-QR estimator is given with the objective function,

$$\rho_\tau(t) := |\tau - \boldsymbol{I}(t \leq 0)||t|^q, q \geq 1,$$

$q = 2$ corresponds to the expectile regression estimator.

Model Framework

1. DGP: $Y_t = \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \beta_{j,0} X_{j,t-k-1} + u_t,\ t = 1, 2, \cdots, n.$
2. Consider the one-step ahead linear predictive QR model:

$$Y_t = \boldsymbol{X}_{t-1}^{\top}(\tau)\beta + u_t(\tau),\ \mathbb{P}(u_t(\tau) < 0|\mathcal{F}_{t-1}) = \tau \in (0,1),$$

such that the conditional quantile of Y_t is the linear form of

$$Q_{\tau}(Y_t|\mathcal{F}_{t-1}) = \boldsymbol{X}_{t-1}^{\top}\beta(\tau),$$

and $\boldsymbol{X}_t \in \mathbb{R}^p$, $p > n$ and $\log p = O(n^{1/3})$,

$$\boldsymbol{X}_t = (\boldsymbol{I}_p - C/n^{\alpha})\boldsymbol{X}_{t-1} + v_t,\ \alpha \in [0,1],$$

where $C = \mathrm{diag}(c_1, \cdots, c_p)$, $w_t := (u_t, v_t)^{\top}$ follows an weakly stationary α-mixing process with $\mathbb{E}[w_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[w_t w_t^{\top}|\mathcal{F}_{t-1}] = \Sigma$ with $\Sigma_{uv} = \Sigma_{vu} \neq 0$, $\sigma_{ut}^2 := \Sigma_{uu}$ and $\sigma_{vt}^2 := \Sigma_{vv}$.

Persistence

- 4 possible levels of persistence: let $c_i > 0$,
  1. $|1 - c_i| \ll 1$, $\alpha = 0\ \forall i$;
  2. $c_i \in (0, \infty)$, $\alpha = 1\ \forall i$;
  3. $c_i = 0$, $\alpha = 1\ \forall i$ (unit-root);
  4. $c_i \in (0, \infty)$, $\alpha \in (0, 1)\ \forall i$.
- To cure the persistence and the embedded endogeneity simultaneously, consider a smoothing filter instrumental variable IVX:

$$\boldsymbol{Z}_t = F\boldsymbol{Z}_{t-1} + \Delta\boldsymbol{X}_t = \sum_{k=1}^{t} F^{t-k}\Delta\boldsymbol{X}_k,$$

  where $\boldsymbol{Z}_0 = 0$, Δ is a first-differenced operator, $F = \boldsymbol{I}_p - C_z/n^\gamma$, $C_z = c_z\boldsymbol{I}_p$, and the pre-specified $\gamma \in (0, 1)$. We consider $\gamma = 0.95$.
- $F = 0$, corresponds to $\boldsymbol{Z}_t = \Delta\boldsymbol{X}_t$; $F = \boldsymbol{I}_p$ gives $\boldsymbol{Z}_t = \boldsymbol{X}_t$.

Estimation

- Consider M candidate submodels to used, in which

$$\mathbb{P}(Y_{t(m)} \leq \boldsymbol{X}_{t-1(m)}^{\top}\beta(m)|\boldsymbol{Z}_{t-1(m)}) = \tau \in (0,1),\ m = 1,2,\cdots,M,$$

M can go to infinity as n, and p_m denotes the covariates used in the mth submodel.

- The model can be written as a conditional expectation,

$$\mathbb{E}[\tau - \boldsymbol{I}(Y_{t(m)} - \boldsymbol{X}_{t-1,(m)}^{\top}\beta(m) \leq 0)|\boldsymbol{Z}_{t-1(m)}] = 0,$$

which implies the unconditional estimating equations

$$\mathbb{E}[g_t(\beta(m))] := \mathbb{E}[\boldsymbol{Z}_{t-1(m)}(\tau - \boldsymbol{I}(Y_{t(m)} - \boldsymbol{X}_{t-1(m)}^{\top}\beta(m) \leq 0))] = 0.$$

- The pseudo-true vector of parameters (β^*) is identified by

$$\mathbb{E}[g_t(\beta^*(m))] = 0,\ a.s.$$

- We assume $\beta^* = \beta_0 + o_p(n^{1/2})$.

- The τth predictive quantile regression estimator in mth submodel can be solved by:

$$\widehat{\boldsymbol{\beta}}(m) = \arg\min_{\boldsymbol{\beta}(m)} Q_n(\boldsymbol{\beta}) := \frac{1}{2}\Big[n^{-1}\sum_{t=1}^{n} g_t(\beta(m)\Big]^{\top}\widehat{\Omega}(\bar{\beta}(m))\Big[n^{-1}\sum_{t=1}^{n} g_t(\beta(m))\Big],$$

  whre $\widehat{\Omega}(\bar{\beta}(m)) = \widehat{S}^{-1}(\bar{\beta}(m))$ is an weighting matrix with some consistently initial estimator of long-run variance estimator $\widehat{S}(\bar{\beta}(m))$, such that $\widehat{S}^{-1} \xrightarrow{\text{p}} S^{-1}$. If dim$(\boldsymbol{X})$ =dim$(\boldsymbol{Z})$, $\overline{\Omega}(\beta(m)) = \boldsymbol{I}$ is convenient.
- The predictive quantile regression model averaging estimator of $Q_\tau(Y_t|\mathcal{F}_{t-1})$ is given by

$$\widehat{Q}_\tau(Y_t;\boldsymbol{w}) = \sum_{m=1}^{M} w_m \boldsymbol{X}_{t-1(m)}^{\top}\widehat{\beta}(m),$$

  where $\boldsymbol{w} \in W := \{\boldsymbol{w} \in [0,1]^M : \sum_{m=1}^{M} w_m = 1\}$, which is a simplex.

IMPROVEMENT

- To improve the efficiency with the serial dependence, we employ the blocking technique to preserve the dependence among the underlying data.
- Let J and L be two integers denoting the block length and separation between adjacent blocks, respectively. Then, the total number of blocks is, $B = [(n - J)/L + 1]$, where $[.]$ is the integer truncation operator. For each $s = 1, \ldots, B$, the sth data block $b_s = (W_{(s-1)L+1} : W_{(s-1)L+J})$, $W_t = (Y_t, X_{t-1}^\top, Z_{t-1}^\top)^\top$.
- The refined estimating functions over the sth block is

$$m(b_s(\beta^*(m))) = J^{-1/2} \sum_{j=1}^{J} g_{(s-1)L+j}(\beta(m)),$$

- The pseudo-true vector of parameters (β_*) is identified by

$$\mathbb{E}[m(b_s(\beta^*(m))] = 0, \ a.s..$$

- $$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} QC(\boldsymbol{w}),$$
- $$QC(\boldsymbol{w}) := Q_n(\boldsymbol{w}) + \psi(\boldsymbol{w}),$$

  where $Q_n(\boldsymbol{w}) := Q_n(\widehat{\beta}(\boldsymbol{w}))$, $\psi(\boldsymbol{w}) := \gamma_n \tau(1-\tau)s(\tau)\widehat{S}(M)\sum_{m=1}^{M} w_m p_m$.
- $\gamma_n > 0$ is a penalty term regularizing the increasing dimension of the fitted models, and $s(\tau) := 1/f(F^{-1}(\tau))$ is called a sparsity function. A consistent estimator of $s(\tau)$ is adopted by the analog of Powell's (1991) kernel estimator.
- $$\hat{s}(\tau) := \Big[\frac{1}{nh}\sum_{i=1}^{n} K(\frac{\|Y_t - \boldsymbol{X}'_{t-1}\widehat{\beta}(M)\|}{h})\Big]^{-1},$$

  where $K(t) = 2^{-1}\boldsymbol{I}(|t| \leq 1)$ is the uniform kernel, and h is a bandwith such that $h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$. In practice, $h = O(n^{-1/3})$.

- (Out-of-Sample Quantile Prediction Errors (QPE))

$$QPE(\boldsymbol{w}) := \mathbb{E}[\rho_\tau(Y_{t|t-1} - \sum_{m=1}^{M} \boldsymbol{X}_{t-1(m)}^\top \widehat{\boldsymbol{\beta}}(\boldsymbol{w}))|D_{t-1}],$$

where $D_{t-1} := \{(Y_{t-1}, \boldsymbol{X}_{t-1}) : 1 \le t \le n\}$ is the in-sample information set associated with $\mathcal{F}_{t-1}$.

ASYMPTOTIC UNBIASEDNESS

Theorem (4.2.)

$$\mathbb{E}[QC(\boldsymbol{w})] = QPE(\boldsymbol{w}) + o_p(1).$$

• Remark

1. This theorem extends Hansen (2007, 2008) that the Mallows Cp criterion has been shown to be unbiased for the mean squared errors (MSE) up to a constant, and is an approximately unbiased estimate of the mean squared forecast errors (MSFE).
2. We show that a quantile-based Mallows type criterion is asymptotically unbiased for the mean quantile prediction errors. In particular, we allow for increasing numbers of the covariates on high dimension. In other words, the minimizer $\widehat{\boldsymbol{w}}$ that minimizing $QC(\boldsymbol{w})$ can be asymptotically equivalent to minimize $QPE(\boldsymbol{w})$, in probability.

ASYMPTOTIC OPTIMALITY

Under Assumption [B1]-[B3],

Theorem (4.3.)

$$\frac{QPE_n(\widehat{\boldsymbol{w}})}{\inf_{\boldsymbol{w}\in W} QPE_n(\boldsymbol{w})} = 1 + o_p(1)$$

as $n \to \infty$, where $\widehat{\boldsymbol{w}} := \arg\inf_{\boldsymbol{w}\in W} QC(\boldsymbol{w})$.

• Remark

It shows that the Cp-type quantile regression model averaging weight can be asymptotically efficiency in terms of minimizing the out-of-sample quantile prediction errors.

Quantile Screening

- p is too huge, we need to pre-screen predictors, first.
- For a given quantile level $\tau \in (0,1)$, we say that the conditional quantile of Y given X_j does not depend on X_j at the τth quantile level if

$$Q_\tau(Y|X_j) = Q_\tau(Y), a.s. \ \forall \tau \in (0,1),$$

  where $Q_\tau(Y|X_j)$ is the τth CQF of Y given X_j, and $Q_Y(\tau)$ is the τth unconditional quantile of Y.
- Let $F_Y^{-1}(\tau)$ be the unique solution of the τth unconditional quantile,

$$F_Y^{-1}(\tau) = \inf\{t : \mathbb{P}(Y \leq t) \geq \tau\} =: Q_Y(\tau),$$

  and $\hat{Q}_Y(\tau) := F_{Y,n}^{-1}(\tau)$, is the empirical quantile function.

Averaging Quantile Correlation Screening (AQCS)

- $$\hat{u}_j = (nK)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} |X'_{t-1,j} \hat{\beta}_j(\tau_k) - F^{-1}_{Y,n}(\tau_k)|.$$

  $u_j := \mathbb{E}\,|X'_{ij}\beta_{*j} - F_Y^{-1}(\tau)| = \mathbb{E}\,|Q_{Y|X_j}(\tau|X_j) - Q_Y(\tau)|$, in which $\beta_{*j} = \arg\min_{\beta_j} \mathbb{E}[\rho_\tau(Y_i - X'_{ij}\beta_j)]$ and $\hat{\beta}_j = \arg\min_{\beta_j} n^{-1} \sum_{i=1}^{n} \rho_\tau(Y_i - X'_{ij}\beta_j)$.
- We consider $\tau_k \in \{0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99\}$.

2-step Screening Procedure

- step 1: $A := \{1, \ldots, p\}$ be the set of the first d_A mostly correlated covariates.

$$\widehat{A} := \{j \in \{1, \ldots, p\} : \widehat{ACF}_j(1) \leq v_A\},$$

  $\widehat{ACF}_j(1)$ is the 1-order sample autocorrelation coefficient function of X_j, we set $v_A = 0.95$; i.e., the non-stationary and explosive-root predictors are excluded.

- step 2: $T := \{j \in \{1, \ldots, p\} : u_j \geq v\}$: the set of active covariates. $|T| = s < p$, non-sparsity size, where v is user-specified thresholding value be the set of active covariates.

- $$\widehat{T} := \{j \in \{1, \ldots, p\} : \hat{u}_j \geq v_T\},$$

- $|\widehat{I}| := |\widehat{T} \cap \widehat{A}| = d$, be the largest size of all screened submodels.
- p is reduced to d. Construct M submodels by d covariates.

QUANTILE SURE SCREENING PROPERTY

Under Assumption [A1], [A2], and $\log p = o(n^{(1-2\xi)/3})$, $0 < \xi < 1/2$,

Theorem (4.1.)

(i) (Sure Screening Property) For any constant $c_1 > 0$,
(a)

$$\mathbb{P}\Big(\sup_{1\le j\le p} |\hat{u}_j - u_j| \ge c_1 n^{-\xi}\Big) \to 0, \text{ as } n \to \infty.$$

Under $v_T = c_1 n^{-\xi}$,
(b)

$$\mathbb{P}(T \subseteq \widehat{T}) \to 1, \text{as } n \to \infty.$$

(c)

$$\mathbb{P}\left(|\widehat{T}| \le O(n^{\xi} \sum_{j=1}^{p} u_j)\right) \to 1, \text{ as } n \to \infty.$$

QUANTILE SURE SCREENING PROPERTY

Under Assumption [A1]-[A3],

Theorem (4.1.)

(ii) (Ranking Selection Property)

$$\mathbb{P}\left(\liminf_{n\to\infty} \{ \inf_{j\in T} \hat{u}_j - \sup_{j\in T^c} \hat{u}_j \} > 0 \right) = 1.$$

QUANTILE SURE SCREENING PROPERTY

- Remark

1. (a) tells us that the threshold value of selecting important covariates asymptotically, (b) implies that the important covariates whose marginal regression functions make significant contribution will be selected with probability approaching one, and (c) provides a bound for the size of screened models with high probability, as well as controlling the false discovery of selecting covariates.
2. It provides the information about that the important covariates should be ranked higher than the unimportant covariates, asymptotically. We can separate them asymptotically by using a suitable thresholding value. That is, an important covariate cannot be dropped while an unimportant covariate is chosen simultaneously.

- Remark

Recall that model selection Oracle consistent property implies that,

$$\mathbb{P}(T = \widehat{T}) \to 1, \text{ as } n \to \infty.$$

which is stronger than the sure screening property.

COMPARISON

- Denote our screening procedure as AQCS.
- Some different model screening approaches have been proposed in the literature, for example, the sure independent screening (SIS) of Fan and Lv (2008), Hall and Miller (2009)'s generalized correlation screening (GCS), the robust rank correlation based screening (RRCS) of Li et al. (2012), the model-free sure independent ranking screening (SIRS) of Zhu et al. (2011), the quantile-based adaptive sure independent screening (QaSIS) of He et al. (2013), the nonparametric independence screening (NIS) of Fan et al. (2011).

COMPARISON

- The regularized IVX quantile regression with LASSO, EN (elastic net), SCAD (smoothly clipped absolute deviation) and MCP (minimax concave penalty) penalties are described as follows:

$$\widehat{\beta}(\tau) = \arg\min_{\beta} Q_n(\beta) + \sum_{i=1}^{p} P_\lambda(|\beta_j(\tau)|),$$

  where $\lambda > 0$ is the tuning parameter selected by h_v-blocked time series cross validation proposed by Racine (2000),

- and $P_\lambda(|\beta_j|)$ are given by the l_1-norm, SCAD and MCP penalty functions as following, respectively:

1. $P_\lambda(|\beta_j|) := \lambda|\beta_j|$,
2. $P_\lambda(|\beta_j|) = \lambda\Big(\alpha|\beta_j| + (1-\alpha)|\beta_j|^2\Big)$, $\alpha \in [0,1]$,
3. $$P_\lambda(|\beta_j|) := \begin{cases} \lambda|\beta_j|, & \text{if } |\beta_j| \le \lambda \\ \frac{2a\lambda|\beta_j| - \beta_j^2 - \lambda^2}{2(a-1)}, & \text{if } \lambda < |\beta_j| < a\lambda \\ (a+1)\lambda^2/2, & \text{if } |\beta_j| \ge a\lambda, \end{cases}$$

   with the gradient

   $$P'_\lambda(|\beta_j|) := \lambda\Big\{\boldsymbol{I}(|\beta_j| \le \lambda) + \frac{a\lambda - |\beta_j|}{(a-1)\lambda}\boldsymbol{I}(|\beta_j| > \lambda)\Big\},$$

   for some $a > 2$, and $a = 3.7$ is suggested by Fan and Li (2001).
4. $$P_\lambda(|\beta_j|) := \begin{cases} \lambda|\beta_j| - \beta_j^2/2a, & \text{if } |\beta_j| \le a\lambda \\ a\lambda^2/2, & \text{if } |\beta_j| > a\lambda, \end{cases}$$

   with the gradient

   $$P'_\lambda(|\beta_j|) = (\lambda - |\beta_j|/a)_+,$$

   where $a = 3$ is suggested by Zhang (2010).

SIMULATION STUDY: screening

- DGP:

$$Y_t = 2(\boldsymbol{X}_{t-1,1} + 0.95\boldsymbol{X}_{t-1,2} + 0.7\boldsymbol{X}_{t-1,3} + 0.5\boldsymbol{X}_{t-1,4} + 0.4\boldsymbol{X}_{t-1,5}) + \epsilon_t,$$

  where $\epsilon_t := \exp(\boldsymbol{X}_{t-1,11} + \boldsymbol{X}_{t-1,13} + \boldsymbol{X}_{t-1,15})u_t$,
  $R^2 = [\mathrm{Var}(Y) - \mathrm{Var}(\epsilon)]/\mathrm{Var}(Y)$,
- $\boldsymbol{X}$ follows the individual AR (1): $X_t = AX_{t-1} + v_t$, $A = 0.9$ and has the multivariate Gaussian error term $(u_t, v_t)^\top \sim \mathcal{N}(0, \Sigma)$ with the zero mean, the unit variance and the covariance matrix $\Sigma_{uv} = \rho = \{0, -0.95\}$.
- $MC := 200$ replications. The sample size $n = 100, 200$, with the number of the covariates $p = \{200, 400\}$.
- the number of the important covariates is $s_0 := 8$.

SIMULATION STUDY

ρ	s_0	p	n	AQCS	QaSIS5	QaSIS75	NIS	SIRS
0	8	200	100	176	176	174	194	157
0	8	200	200	64	61	103	114	25
0	8	400	100	232	225	278	296	298
0	8	400	200	129	138	146	292	182
-0.95	8	200	100	70	57	107	199	18
-0.95	8	200	200	16	14	53	198	6
-0.95	8	400	100	94	93	257	398	53
-0.95	8	400	200	16	21	64	298	6

SIMULATION STUDY: model averaging

Evaluation:

- the 100 out of sample observations $\{x_{n+k}, y_{n+k}\}_{k=1}^{100}$ are generated and use to calculate the prediction error at τ-quantile,

$$PE(\tau)_r := \frac{1}{100}\sum_{k=1}^{100}\rho_\tau\Big(Y_k - \sum_{m=1}^{M}\hat{w}_m \boldsymbol{X}'_{k(m)}\widehat{\beta}(m)\Big),$$

- $$\widehat{QPE}(\tau) := \frac{1}{MC}\sum_{r=1}^{MC} PE(\tau)_r.$$

- $p = 400,\ M = 10.$

SIMULATION STUDY

/Users/jerryc/Documents/jerrycDocuments/r_files/100_.png

Figure: The performances of empirical quantile prediction errors under $\tau = 0.5$ and $n = 100$.

APPLICATION

- We expand the 544 anomaly portfolios returns constructed from Chen and Zimmermann (2022), together with 14 financial predictors, 14 technical indicators, which includes (i) moving average rules, (ii) momentum rules and, (iii) on-balance volume rules. After excluding the missing values and ineffective variables, our study covers the total effective sample period from $1963/7$ to $2022/12$ at monthly frequency.
- The estimation strategy is based on the recursive window estimation, to obtain the one-step ahead forecast; and the first window length is 240.

APPLICATION

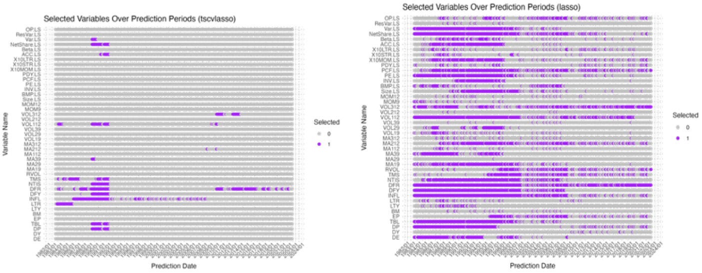


Figure: The selection of predictors under $\tau = 0.5$, recursive window estimation.

APPLICATION

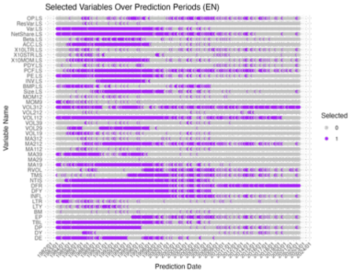


Figure: The selection of predictors under $\tau = 0.5$, recursive window estimation.

APPLICATION

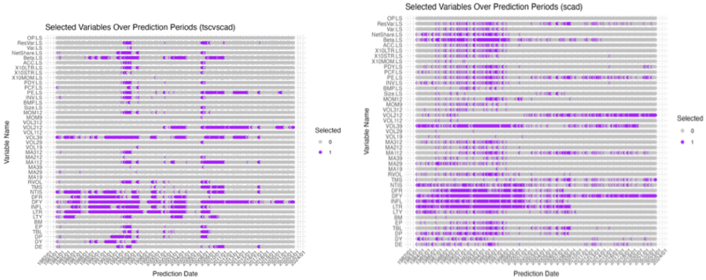
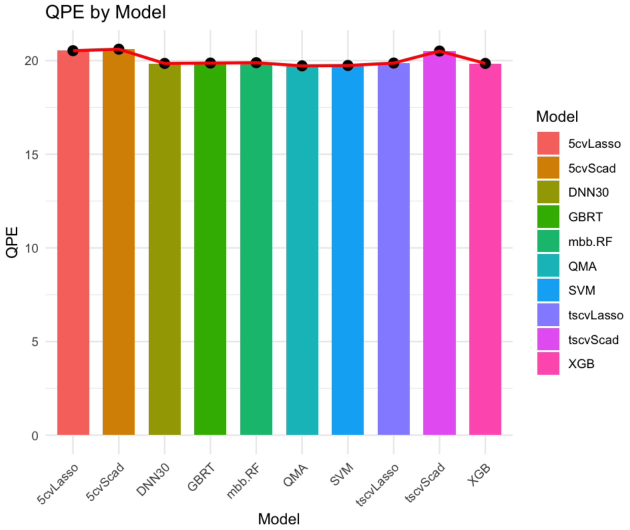


Figure: The selection of predictors under $\tau = 0.5$, recursive window estimation.

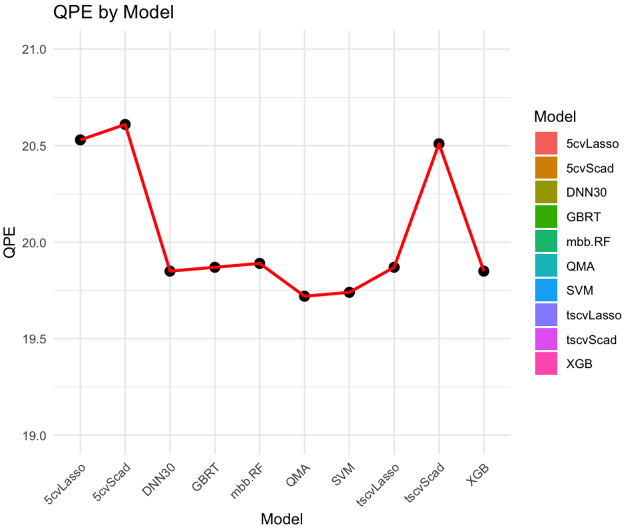
APPLICATION

Method	MSFE
5cv.lasso	20.52
tscv.lasso	19.87
5cv.EN	20.74
tscv.EN	19.88
5cv.SCAD	20.60
tscv.SCAD	20.51
5cv.MCP	20.11
tscv.MCP	20.68

APPLICATION



APPLICATION



THANK YOU.