Testing for Marginal Linear Effects in Quantile Regression

Huixia Judy Wang

Department of Statistics
George Washington University

August 12, 2017

Outline

Detection of Overall Covariate Effect

- Assume the linear quantile regression model:

$$Q_\tau(Y|\mathbf{X}) = \alpha_0(\tau) + \mathbf{X}^T\beta_0(\tau), \tag{1}$$

  where $\mathbf{X} = (X_1, \ldots, X_p)^T$.

- Interest: test if any component of $\mathbf{X}$ has an effect on the τth quantile of Y.
- Wald-type, likelihood-ratio-type test: requires fitting the full model of p predictors. Problem: can quickly become **prohibitive and less powerful** for large p.
- Post-selection inference (e.g. penalized estimator)
  - First select a few important predictors using some variable selection method, and then carry out conventional statistical inference assuming that the chosen model is the correct one.
  - Problem: **ignores the uncertainty** involved in the model selection step, leading to inflated family-wise error rate.

Idea of QMET

- A new test that provides valid post-selection inference based on marginal quantile regression, where quantile regression of Y is performed on each predictor separately.
- Test statistic:

$$T_n(\tau) \doteq \frac{n^{1/2}\hat{\theta}_{\hat{k}_n}(\tau)}{\hat{\sigma}_{\hat{k}_n}(\tau)},$$

  where $\hat{\theta}_{\hat{k}_n}(\tau)$ is the marginal quantile slope estimator of the most "predictive" covariate at the τth quantile.
- An adaptive bootstrap test is developed to control the FWER by calibrating the critical values to adapt to the non-regular asymptotic behavior caused by variable selection.

Model Setup

- Working marginal quantile regression (MQR) models:

$$Q_\tau(Y|X_k) = \alpha_k(\tau) + \theta_k(\tau)X_k, k = 1, \ldots, p, \tag{2}$$

  - $\theta_k(\tau)$: approximates the linear effect of the kth predictor X_k on the τth quantile of Y, **Quantile Marginal linear Effect (QME)** of X_k.
  - Due to model misspecification, in general

$$\theta_k(\tau) \neq \frac{Q_\tau(Y|\mathbf{X})}{\partial X_k}.$$

- Define $k_{0,\tau} = \arg\min_{k=1,\ldots,p} E[\rho_\tau(Y - \alpha_k(\tau) - \theta_k(\tau)X_k)]$: index of the "most predictive covariate" at the τth quantile.
- Let $\theta_0(\tau) = \theta_{k_{0,\tau}}(\tau)$.
- Consider testing $H_0 : \theta_0(\tau) = 0$ versus $H_a : \theta_0(\tau) \neq 0$.

Proposed Test Statistic

- Sample QME estimators

$$(\hat{\alpha}_k(\tau), \hat{\theta}_k(\tau)) = \text{argmin}_{\alpha,\theta} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - \theta x_{i,k}).$$

- Define

$$\hat{k}_{n,\tau} = \text{argmin}_{k=1,\ldots,p} \sum_{i=1}^{n} \rho_\tau\{y_i - \hat{\alpha}_k(\tau) - \hat{\theta}_k(\tau) x_{i,k}\},$$

  and $\hat{\theta}_n(\tau) \equiv \hat{\theta}_{\hat{k}_{n,\tau}}(\tau)$.

- For any given $k = 1, \ldots, p$, let $\sigma_k^2(\tau)$ be the asymptotic variance of $\hat{\theta}_k(\tau)$, and $\hat{\sigma}_k^2(\tau)$ a consistent estimator of $\sigma_k^2(\tau)$.
- Proposed test statistic:

$$T_n(\tau) = n^{1/2} \hat{\theta}_n(\tau) / \hat{\sigma}_n(\tau),$$

  where $\hat{\sigma}_n(\tau) = \hat{\sigma}_{\hat{k}_{n,\tau}}(\tau)$.

Challenges

The standardized estimator T_n is a natural statistic for testing H_0, but establishing its limiting distribution is challenging.

- When there is no marginal effect of any of the components of $\mathbf{X}$ on the τth conditional quantile of Y, $k_{0,\tau}$ is **not identifiable** — it can be any of the p indices.
- The distribution of $n^{1/2}\{\hat{\theta}_n(\tau) - \theta_0(\tau)\}$ does not converge uniformly with respect to $\theta_0(\tau)$ in the neighborhood of $\theta_0(\tau) = 0$, so the normal limiting distribution that holds away from $\theta_0(\tau) = 0$ cannot be used to construct rejection regions.
- Solution: find the local asymptotic distribution of $\hat{\theta}_n(\tau)$.

Local Model

- Local linear quantile regression model:

$$Q_\tau(Y|\mathbf{X}) = \mathbf{X}^T \boldsymbol{\beta}_n(\tau), \quad \boldsymbol{\beta}_n(\tau) = \boldsymbol{\beta}_0(\tau) + n^{-1/2}\mathbf{b}_0(\tau), \tag{3}$$

  where $\mathbf{b}_0(\tau) \in \mathbb{R}^p$ is the local parameter.

- Define $\bar{k}_{n,\tau}(\mathbf{b}_0) = \text{argmin}_{k=1,\ldots,p} \min_{\alpha,\theta} E[\rho_\tau(Y - \alpha - X_k\theta)]$.
- Rewrite $\theta_0(\tau)$ as $\theta_n(\tau) \doteq \theta_{\bar{k}_{n,\tau}(\mathbf{b}_0)}(\tau)$ and $k_{0,\tau}$ as $k_{n,\tau}$.
- If $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$ and $k_{0,\tau}$ is unique: $\bar{k}_{n,\tau}(\mathbf{b}_0) \to k_{0,\tau}$ and $\theta_n(\tau)$ is asymptotically bounded away from zero **(non-local alternative)**.
- If $\boldsymbol{\beta}_0(\tau) = \mathbf{b}_0(\tau) = \mathbf{0}$: $\bar{k}_n(\mathbf{b}_0)$ is not well defined and $\theta_n(\tau) = 0$ **(null)**.

- Under the local model, $\bar{k}_{n,\tau}(\mathbf{b}_0)$ is still **"weakly identifiable"** when $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$ but $\mathbf{b}_0(\tau) \neq \mathbf{0}$.

Asymptotic Properties

Theorem

Under some regularity conditions, assume that $k_0 \equiv \bar{k}_n(\mathbf{0})$ *is unique when* $\beta_0 \neq \mathbf{0}$, *and* $\bar{k}_n(\mathbf{b}_0) \to \kappa(\mathbf{b}_0) \in \{1, \ldots, p\}$ *when* $\beta_0 = \mathbf{0}$ *with* $\mathbf{b}_0 \neq 0$,

$$\frac{n^{1/2}(\hat{\theta}_n - \theta_n)}{\hat{\sigma}_n} \xrightarrow{d} \begin{cases} \dfrac{M_{2k_0}(\beta_0)\pi_{k_0}(\beta_0) - M_{1k_0}(\beta_0)\mu_{k_0}(\beta_0)}{V_{k_0}(\beta_0)\sigma_{k_0}} & \text{if } \beta_0 \neq \mathbf{0}, \\ \dfrac{M_{2K}(\mathbf{0})\pi(\mathbf{0}) - M_1(\mathbf{0})\mu_K(\mathbf{0})}{V_K(\mathbf{0})\sigma_K} + \left(\dfrac{\mathbf{C}_K}{V_K(\mathbf{0})\sigma_K} - \dfrac{\mathbf{C}_{\kappa(\mathbf{b}_0)}}{V_{\kappa(\mathbf{b}_0)}(\mathbf{0})\sigma_K}\right)^T \mathbf{b}_0 & \text{if } \beta_0 = \mathbf{0}, \end{cases}$$

- $(M_{11}(\beta_0), \ldots, M_{1p}(\beta_0), M_{21}(\beta_0), \ldots, M_{2p}(\beta_0))^T$ *is a zero-mean normal random vector*
- $K = \text{argmax}_k \left\{\mathbf{M}_k(\mathbf{0}) + \mathbf{B}_k^T \mathbf{b}_0\right\}^T \mathbf{J}_k(\mathbf{0})^{-1} \left\{\mathbf{M}_k(\mathbf{0}) + \mathbf{B}_k^T \mathbf{b}_0\right\}$, $\mathbf{M}_k(\beta_0) = (M_{1k}(\beta_0), M_{2k}(\beta_0))^T$
- $\mathbf{J}_k(\beta_0) = \lim_{n\to\infty} E\left\{f_\epsilon(e_k|\mathbf{X})\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^T\right\}$ *with* $e_k = \alpha_k + X_k\theta_k - \alpha_0 - \mathbf{X}^T\beta_n$, $V_k(\beta_0) = |\mathbf{J}_k(\beta_0)|$, $\mathbf{B}_k = E\left\{f_\epsilon(0|\mathbf{X})\mathbf{X}\tilde{\mathbf{X}}_k^T\right\}$, *and* $\mathbf{C}_k = E\left\{f_\epsilon(0|\mathbf{X})\right\} E\left\{f_\epsilon(0|\mathbf{X})X_k\mathbf{X}\right\}$ $-E\left\{f_\epsilon(0|\mathbf{X})X_k\right\} E\left\{f_\epsilon(0|\mathbf{X})\mathbf{X}\right\}$ *with* $\epsilon = Y - \alpha_0 - \mathbf{X}^T\beta_n$.

Adaptive Bootstrap 1

- Based on Theorem 1, we may obtain the asymptotic critical values for $T_n(\tau)$ by simulating its asymptotic representation under the null hypothesis (with $\mathbf{b}_0 = \mathbf{0}$ and $\boldsymbol{\beta}_0 = \mathbf{0}$).
- However, this requires estimating the weighted covariance matrix of $\mathbf{X}$ with weights to accommodate heteroscedasticity, and thus does not perform well in finite samples with large p.
- For practical purposes, we propose to adopt the idea in McKeague and Qian (2015) and develop an adaptive bootstrap procedure.

Adaptive Bootstrap 2

- Obtain the bootstrap sample $\{(y_i^*, \mathbf{x}_i^*), i = 1, \ldots, n\}$ by sampling the observed data with replacement.
- Naive bootstrap version of $R_n(\tau) \doteq n^{1/2}\{\hat{\theta}_n(\tau) - \theta_n(\tau)\}/\hat{\sigma}_n(\tau)$ is

$$R_n^\star(\tau) \doteq \frac{n^{1/2}\{\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau)\}}{\hat{\sigma}_n^*(\tau)}. \qquad \text{Invalid}$$

- Idea of the adaptive bootstrap to account for the discontinuity in the limiting distribution of $R_n(\tau)$: carry out a pre-test by comparing $|T_n(\tau)|$ with some threshold $\lambda_n > 0$ to capture the behaviors of $R_n(\tau)$ in two scenarios:
  - $|T_n(\tau)| > \lambda_n \Rightarrow \beta_0(\tau) \neq 0$ **(away from the null)**. Use standard bootstrap.
  - $|T_n(\tau)| < \lambda_n \Rightarrow \beta_0(\tau) = \mathbf{0}$ **(in the $n^{-1/2}$ neighborhood of the null)**. Bootstrap the asymptotic representation under $\beta_0(\tau) = \mathbf{0}$.

Adaptive Bootstrap 3

- Modified bootstrap version of R_n:

$$R_n^*(\tau) = \begin{cases} \frac{n^{1/2}\{\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau)\}}{\hat{\sigma}_n^*(\tau)} & \text{if } |T_n(\tau)| > \lambda_n \text{ or } |T_n^*(\tau)| > \lambda_n, \\ \mathbb{V}_{n,\tau}(\mathbf{b}) & \text{otherwise,} \end{cases} \tag{4}$$

  where $T_n^*(\tau) = n^{1/2}\hat{\theta}_n^*(\tau)/\hat{\sigma}_n^*(\tau)$, and $\mathbb{V}_{n,\tau}(\mathbf{b})$ is a process in $\mathbf{b} \in \mathbb{R}^p$, denoting the asymptotic representation of $R_n(\tau)$ in the local model with $\beta_n(\tau) = n^{-1/2}\mathbf{b}$.

- For calibration of the test statistic, only need compute $\mathbb{V}_{n,\tau}(\mathbf{b})$ with $\mathbf{b} = \mathbf{0}$.

- The $\alpha/2$ and $(1-\alpha/2)$th percentiles of $R_n^*(\tau)$ can be used as critical values for a level α test.

Testing Across Multiple Quantiles

- Assess the relationship between Y and $\mathbf{X}$ at multiple quantiles.
- A grid of quantile levels: $\tau_1 < \cdots < \tau_L$.
- Hypotheses
  - $\tilde{H}_0 : \theta_0(\tau_1) = \cdots = \theta_0(\tau_L) = 0$ v.s.
  - $\tilde{H}_a$: at least one of the $\theta_0(\tau_l)$'s is nonzero,
  - $\theta_0(\tau_l)$: the slope of the most informative predictor at the τ_lth quantile.
- Test statistic:

$$S_n = \sum_{l=1}^{L} T_n^2(\tau_l)$$

- Test calibration via adaptive bootstrap: the modified bootstrap statistic $S_n^* = \sum_{l=1}^{L} R_n^{*2}(\tau_l)$.

Tuning Parameter λ_n

- The tuning parameter λ_n is involved in the pre-test.
- $\lambda_n = 0 \Leftrightarrow$ standard bootstrap (inflated Type I error).
- Large $\lambda_n \Rightarrow$ a conservative test (lack of power in finite samples).
- Rule of thumb: $\lambda_n = c\sqrt{\tau(1-\tau)\log(n)}$ with $c \in [4, 10]$ works well for large samples.
- For finite samples: use a double bootstrap to choose c.

Consistency of the Adaptive Bootstrap

The distribution of $R_n^*(\tau)$ provides a consistent estimation of the distribution of $R_n(\tau)$.

Theorem

Suppose assumptions A1-A4 hold, $\lambda_n = o(n^{1/2})$ and $\lambda_n \to \infty$, then under model (3), $R_n^*(\tau)$ converges to the limiting distribution of $R_n(\tau)$ conditionally on the observed data.

Forward Regression via QMET

- The proposed bootstrap test can be extended in a forward stepwise manner to select multiple significant predictors.
- If the test detects a significant predictor in the first stage, we can then use the residuals $Y - \hat{\theta}_n(\tau) X_{\hat{k}_{n,\tau}}(\tau)$ as a new outcome variable and carry out the marginal quantile regression over the remaining predictors.
- The procedure can be repeated until no more significant predictors are detected.

Simulation Design

Model: $Y = \mathbf{X}^T\boldsymbol{\beta} + (1 + \gamma X_1)\epsilon$, $\mathbf{X} = (X_1, \ldots, X_p)$

- **Case 1:** $\boldsymbol{\beta} = \mathbf{0}$ **(null)**, $\gamma = 0$ (homoscedastic), $\epsilon \sim N(0, 1)$.
- **Case 2**: $\boldsymbol{\beta} = (1/3, 0, \ldots, 0)^T$ (one unique active predictor).
- **Case 3**: $\beta_1 = \cdots = \beta_5 = 0.25$, $\beta_6 = \cdots = \beta_{10} = -0.15$ and $\beta_j = 0$ for $j = 10, \ldots, p$ ($k_{0,\tau}$ is not unique), $\epsilon \sim t_2$.
- Case 4 (heteroscedastic): $\gamma = 0.45$, $\epsilon \sim N(0, 1)$ (null at $\tau = 0.5$; alternative at $\tau \neq 0.5$).
- $n = 200$, $p = 10, 100, 200, 400, 1000$, $\tau = 0.5, 0.75$.

Methods For Comparison

- QMET: proposed quantile marginal effect test at single quantiles $\tau = 0.5, 0.75$ and across three quartiles (Mtau).
- CPB: standard centered percentile bootstrap approach.
- AOV: analysis-of-variance-type test. Feasible only in cases where $p < n$.
- BONF: multiple testing using Bonferroni adjustment (across p tests).

Rejection Percentages in Case 1 (Null Model)

	$\tau = 0.5$				$\tau = 0.75$				Mtau
p	CPB	AOV	BONF	QMET	CPB	AOV	BONF	QMET	QMET
10	**27.8**	6.6	5.2	3.8	**24.8**	6.2	3.2	5.6	2.4
100	**35.4**	0.0	2.4	3.0	**37.2**	0.0	2.4	4.8	2.8
200	**39.4**		2.2	4.6	**34.6**		1.8	5.6	2.6
400	**42.2**		2.0	4.4	**42.4**		2.0	8.0	2.4
1000	**48.8**		1.6	4.6	**42.0**		1.2	4.8	1.6

Rejection Percentages in Case 2 (Alternative Model)

	$\tau = 0.5$				$\tau = 0.75$				Mtau
p	CPB	AOV	BONF	QMET	CPB	AOV	BONF	QMET	QMET
10	86.5	61.0	79.6	75.0	78.8	50.6	73.0	64.8	83.8
100	85.7	0.0	67.2	67.0	75.4	0.2	55.8	54.4	70.2
200	83.2		**59.2**	61.5	77.2		**47.6**	59.2	65.6
400	84.8		**58.8**	62.1	74.8		**43.0**	60.0	70.1
1000	85.7		**50.6**	64.6	75.6		**38.2**	61.2	65.4

Rejection Percentages in Case 3 (Alternative Model)

	$\tau = 0.5$				$\tau = 0.75$				Mtau
p	CPB	AOV	BONF	QMET	CPB	AOV	BONF	QMET	QMET
10	94.0	94.2	91.6	85.4	75.0	70.2	70.6	62.0	84.2
100	88.2	0.0	75.6	71.8	71.0	0.4	48.8	47.6	72.5
200	89.6		74.2	72.0	71.2		52.2	50.2	71.3
400	84.4		**67.4**	71.6	69.4		**44.8**	49.4	68.8
1000	87.4		**64.0**	74.4	70.2		**39.0**	52.6	70.8

Rejection Percentages in Case 4

- $\tau = 0.5$: null is true
- $\tau = 0.75$: alternative is true

	$\tau = 0.5$				$\tau = 0.75$				Mtau
p	CPB	AOV	BONF	QMET	CPB	AOV	BONF	QMET	QMET
10	**24.0**	7.2	6.6	6.0	90.8	47.8	**70.8**	80.6	77.4
100	**26.2**	0.0	4.4	3.2	86.0	0.0	**53.4**	61.4	46.8
200	**25.0**		3.0	4.0	86.4		**44.0**	58.4	43.4
400	**30.6**		2.2	5.8	85.8		**41.4**	57.0	30.7
1000	**28.6**		2.2	6.6	87.4		**38.2**	58.2	26.8

- For homogeneous alternatives, Mtau is more powerful than single-quantile test.
- For heterogeneous alternatives, better to apply the test at a single quantile with stronger signal than the omnibus test across quantiles.

Under Misspecified Nonlinear Models

- Case 5: $Y = X_1^2/3 + \epsilon$ (cannot detect nonlinear effect)
- Case 6: $Y = X_1^2/3 + X_1/3 + \epsilon$

		p				
Case	Method	10	100	200	400	1000
5	QMET	3.6	1.6	2.1	2.3	4.1
	BONF	4.0	3.2	6.0	2.6	3.2
6	QMET	65.8	50.0	49.0	53.1	60.5
	BONF	65.0	46.0	38.6	40.2	34.2

Possible extension to detect nonlinear covariate effects: marginally regressing Y on some polynomial or basis functions of each covariate separately.

Study of HIV Drug Resistance

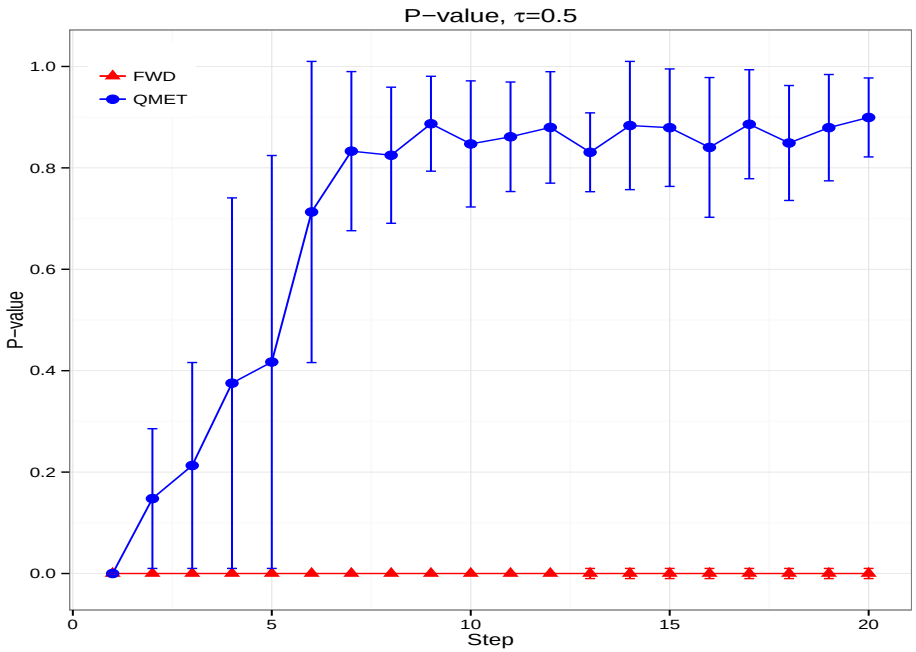
- After a patient starts the therapy, HIV virus can form new mutations and become drug resistance.
- Interest: understand the impact of mutations on drug resistance.
- Susceptibility data for the drug efavirenz (EFV)
  - 1472 HIV isolates
  - Y: $\log_{10}$-susceptibility (fold decrease in susceptibility of a virus isolate compared with a control isolate)
  - $\mathbf{X}$: indicating the presence of a mutation of interest at 197 locations, $p = 197$
- The response distribution is highly non-normal even after log transformation.
- Upper quantiles: associated with stronger drug resistance.

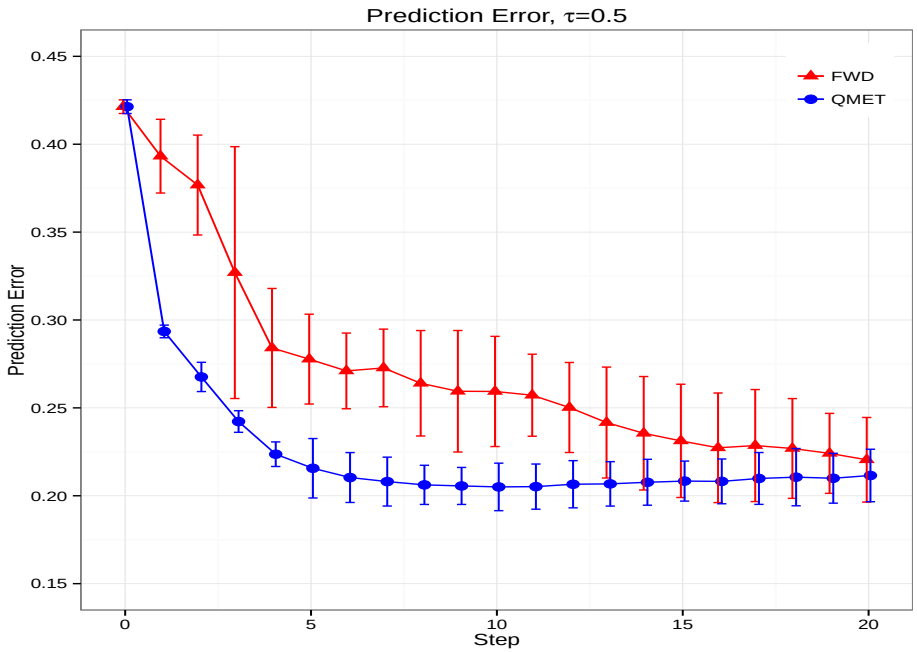
Cross Validation For Quantile Prediction

- Randomly split data to training set $n = 190$, testing set: size 1282.
- For each split, carry out 20 steps of forward quantile regression using the training data, and then use the selected model to predict τth quantile of log susceptibility of the testing data.
- Two methods:
  - **QMET**: the QMET procedure is applied in each step by treating residuals from the previous stage as the new outcomes, and identifies the predictor that gives the smallest quantile loss in marginal regression.
  - **FWD**: conventional forward selection that choses the predictor that gives the smallest Wald-type p-value conditional on predictors that have entered prior to the current step.

Cross Validation Results —$\tau = 0.5$

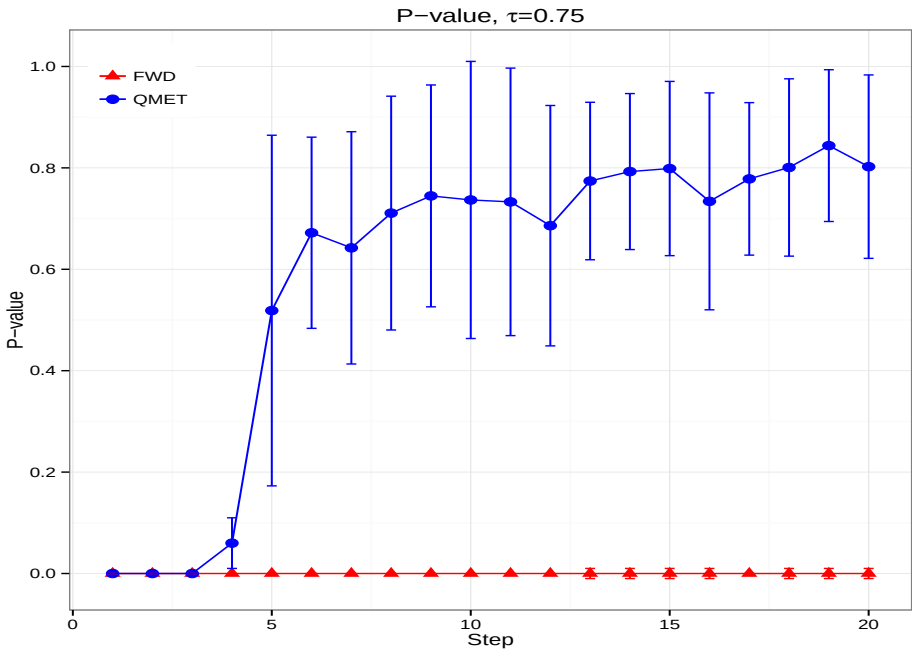
- Left: training-set p-values (median $\pm$ MAD) for the newly entered predictor at each step.
- Right: prediction errors (median $\pm$ MAD) in the testing sets.

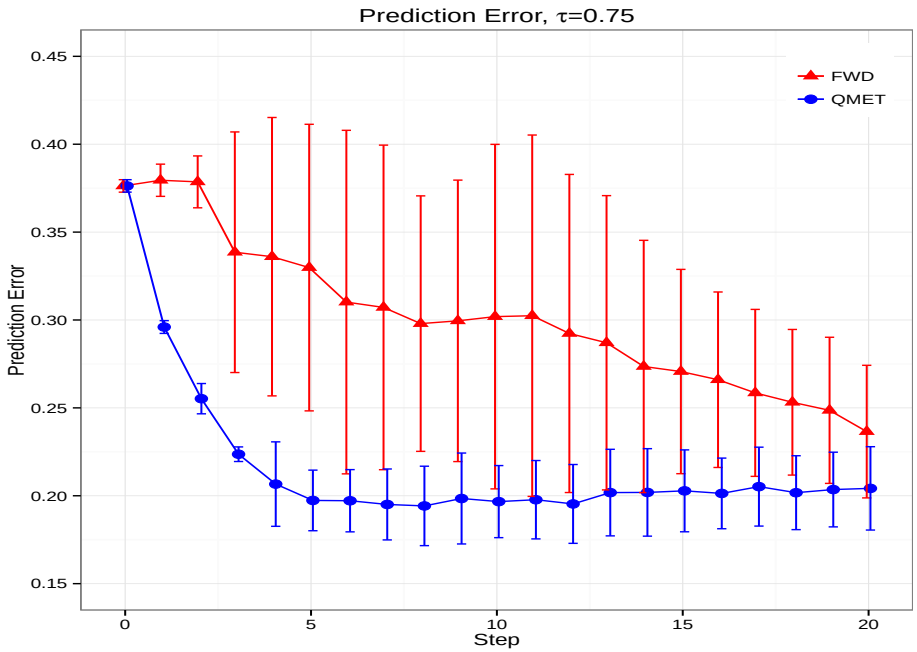




Cross Validation Results ($\tau = 0.75$)

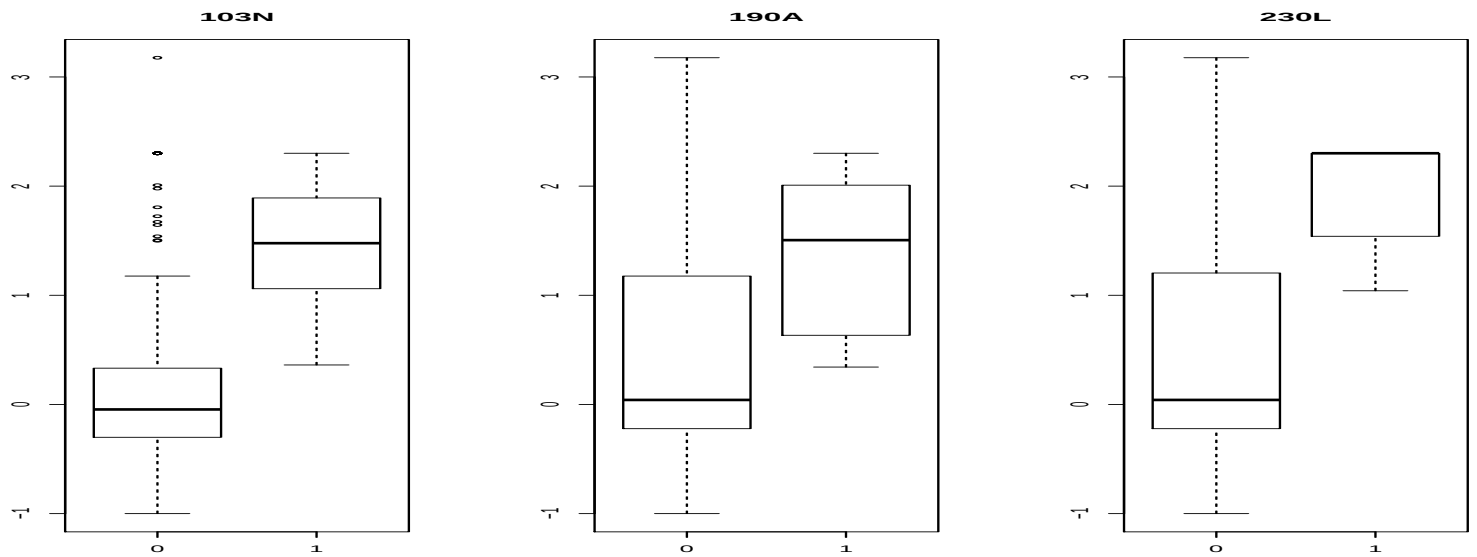
- Left: training-set p-values (median $\pm$ MAD) for the newly entered predictor at each step.
- Right: prediction errors (median $\pm$ MAD) in the testing sets.





One Example

- QMET sequentially selects 103N, 190A and 230L at $\tau = 0.75$, but only the first at $\tau = 0.5$.
- Isolates with these three mutations are associated with higher drug resistance at both quantiles.
- After accounting for the effect of the first mutation, the effects of 190A and 230L become insignificant at median but remains significant at $\tau = 0.75$, that is, for the isolates that are more drug resistant.



Main Conclusions

- Developed a new procedure for detecting the presence of marginal effects in quantile regression.
- The proposed test is shown to be effective and has stable performance, providing adequate control of family-wise error rate along with competitive power, even for high-dimensional cases with $p \gg n$ (although the asymptotic theory we used to calibrate the test assumes fixed p).
- The proposed method can also be extended to detect the overall significance of covariates across multiple quantiles.
- The new approach allows the most active predictors to vary at different quantile levels of the response distribution
  - more flexible than mean regression;
  - added advantage of being robust against outliers.

Limitations and Extensions

- The method is designed to detect linear covariate effect, though it also works reasonably well in some misspecified nonlinear models that exhibit some linear trend.
  - Extension: regression on polynomial or basis functions of each predictor separately.
- The asymptotic theory assumes fixed p. Theory for diverging p will be different (work in progress).
- Marginal regression/screening is more susceptible to the problem of “unfaithfulness" when correlations between covariates increase, leading to high false discovery rate.
  - Conditional MQR-based test to identify the presence of significant predictors conditional on some pre-chosen covariates. Such a conditional test in forward regression can help recover hidden significant covariates (paper in revision jointly with Yanlin Tang and Emre Barut on least squares regression).

Acknowledgement

- NSF CAREER Award DMS-1149355.
- KAUST OSR-2015-CRG4-2582.
- Collaborators:
  - Ian McKeague and Min Qian at Columbia University (QMET)
  - Yanlin Tang and Emre Barut (Conditional ART)
  - Yanlin Tang, Yinfeng Wang, Qing Pan (MQR for diverging p)

Thank you all for your attention!