Predicting GDP growth across different Quantiles

Honor Thesis

Qinwen (Katherine) Wang

Advisor: Dr. Seunghwa Rho

Committee: Dr. Ruoxuan Xiong   Dr. Bree Ettinger

Presentation Overview

- Why GDP growth estimate?
- Why quantile analysis?

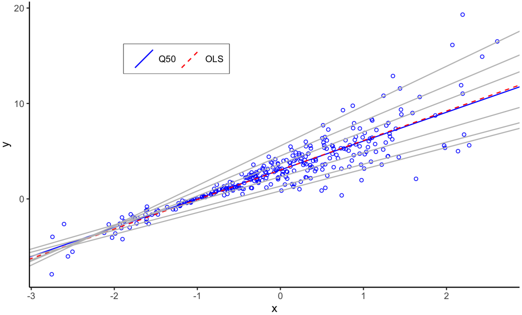


Figure: Comparison of OLS and quantile regression

We choose to analyze the relationship between real GDP growth and National Financial Conditions Index(NFCI).

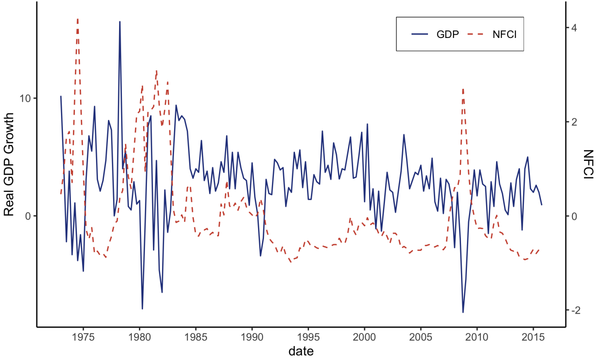


Figure: Quarterly NFCI & real GDP growth

- **Univariate Analysis**
  Estimation of τ^{th} quantile of $ggdp_{t+h}$ is defined as:
  $$\hat{Q}_{ggdp_{t+h}|ggdp_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} ggdp_t$$
  and
  $$\hat{Q}_{ggdp_{t+h}|NFCI_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} NFCI_t$$
- **Bivariate Analysis**
  Estimation of τ^{th} quantile of $ggdp_{t+h}$ under bivariate condition can be defined as:
  $$\hat{Q}_{ggdp_{t+h}|ggdp_t+NFCI_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} ggdp_t + \hat{\beta}_{2,\tau} NFCI_t$$

Method

List of the methods we use for GDP growth estimates:

- **Quantile Regression** – classic QR method
- **IVXQR**[1] – address QR invalidity
- **Forest Based Methods**
  - Quantile Regresssion Forest[2]
  - Generalized Quantile Random Forest[3]

---

[1] David M. Kaplan and Yixiao Sun. "SMOOTHED ESTIMATING EQUATIONS FOR INSTRUMENTAL VARIABLES QUANTILE REGRESSION". In: *Econometric Theory* 33.1 (2017), pp. 105–157.

[2] Nicolai Meinshausen. "Quantile Regression Forests". In: *JOURNAL OF MACHINE LEARNING RESEARCH* 7 (2006), pp. 983–999.

[3] Susan Athey, Julie Tibshirani, and Stefan Wager. *Generalized Random Forests*. 2016.

Quantile Regression

Estimate τ^{th} quantile of y: $Q_{y_{t+h}|x_t}(\tau) = x_t'\beta_\tau$

- OLS:

$$\hat{\beta} = argmin \sum_{i=1}^{n} (y_{t+h} - x_t'\beta)^2$$

- Median ($\hat{\beta}_{0.5}$):

$$\hat{\beta}_{0.5} = argmin \sum_{i=1}^{n} |y_{t+h} - x_t'\beta_{0.5}|$$

- Quantile Regression:

$$\hat{\beta}_\tau = argmin \sum_{t=1}^{T-h} |u| \cdot \{\tau \cdot \mathbb{1}[u_t \geq 0] + (1-\tau) \cdot \mathbb{1}[u_t < 0]\}$$

where $u_t = y_{t+h} - x_t'\beta_\tau$ is the residual.

IVXQR

When NFCI process is approximated with an AR(1) process,

$$NFCI_t = \rho NFCI_{t-1} + \varepsilon_t$$

the AR parameter estimate ρ is around 0.88, indicating that NFCI is highly persistent.

QR method will lead to size distortion if one or more predictors are persistent. Therefore, we apply IVXQR to address the invalidity.

IVXQR

IVXQR adopts IVX filtering[4] method to define the instrument variable $\tilde{z}_t$ in the following way:

$$\tilde{z}_t = R\tilde{z}_{t-1} + \Delta x_t$$

- $R \to 0$, $\tilde{z}_t$ boils down to the first difference transformation.
- $R \to 1$, $\tilde{z}_t$ becomes the level of the variable without transformation.

[4] Tassos Magdalinos and Peter C. B. Phillips. "Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors". In: *Econometric Theory* 25.2 (2009), pp. 482–526.

For the model we consider, the following moment conditions need to be satisfies:

$$\sum_{t=1}^{n} (\tau - \mathbb{1}[u_t < 0]) = o_p(1)$$

$$\sum_{t=1}^{n} ggdp_t \, (\tau - \mathbb{1}[u_t < 0]) = o_p(1)$$

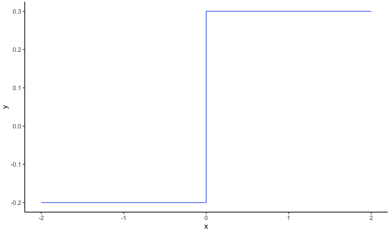
$$\sum_{t=1}^{n} \tilde{z}_t \, (\tau - \mathbb{1}[u_t < 0]) = o_p(1)$$

where

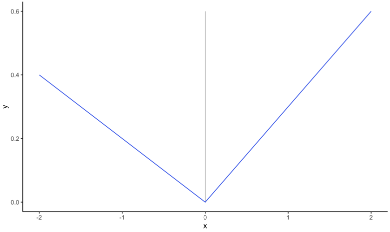
$$u_t = y_{t+h} - \hat{\beta}_{0,\tau} - \hat{\beta}_{1,\tau} ggdp_t - \hat{\beta}_{2,\tau} NFCI_t,$$

$$\tilde{z}_t = (1 - 5/n^{\delta}) \tilde{z}_{t-1} + \Delta NFCI_t.$$

IVXQR-ivqr



(a) indicator



(b) IVXQR-ivqr

IVXQR-see

IVXQR-see smoothes out the moment condition by replacing the indicator function, $\mathbb{1}[u_t < 0]$, with

$$G\left(\frac{-u_t}{h}\right).$$

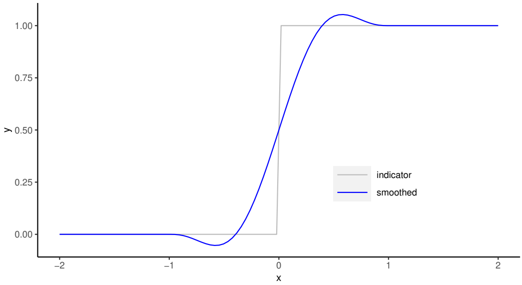


Figure: IVXQR-see smooth visualization

Tree Based Methods

We used two branches for Tree Based Methods Analysis:

- Quantile Regression Forest
- Generalized Quantile Random Forest

Different from random forest, quantile regression forest obtained the distribution $F(y|X = x) = Pr[Y \leq y|X = x]$ instead of the mean only.

$$\hat{F}[y|X = x] = \sum_{i=1}^{n} w_i(x)\mathbb{1}[Y_i \leq y]$$

where the weight can be defined as:

$$w_i(x) = k^{-1}\sum_{t=1}^{k} \frac{\mathbb{1}[x_i \in l(x, \theta_t)]}{\#l(x, \theta_t)}$$

Consider we have k trees in total, $l(x, \theta_t)$ denotes the leaf that x falls into for tree $T(\theta)$,
and $\#l(x, \theta)$ is the total number of observations in the node.

The generalized random forest (**grf**) and quantile regression forest have different definitions for weight. **grf** estimates the weight based on closeness to $(NFCI_t = c_1, ggdp_t = c_2)$.

Therefore, the problem becomes:

$$\frac{1}{n}\sum_{t=1}^{T} w(NFCI_t, ggdp_t; c_1, c_2)\begin{pmatrix} 1 \\ ggdp_t \\ \tilde{z}_t \end{pmatrix}(\tau - \mathbb{1}[u_t > 0]) = 0$$

where $u_t = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} ggdp_t + \hat{\beta}_{2,\tau} NFCI_t - y_{t+h}$,
and $w(NFCI_t, ggdp_t; c_1, c_2)$ denotes the weight.

Univariate Analysis

Recall our model:

$$\hat{Q}_{ggdp_{t+h}|ggdp_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} ggdp_t$$

and

$$\hat{Q}_{ggdp_{t+h}|NFCI_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} NFCI_t$$

Objective

- Compare $\beta_{1,\tau}$ calculated using QR and IVXQR methods with OLS.

Univariate Analysis

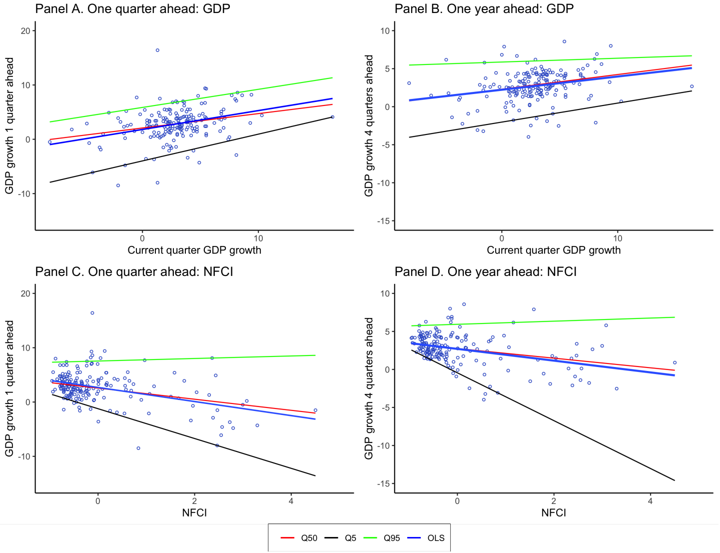
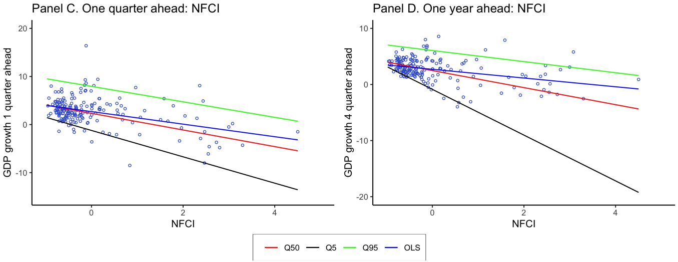
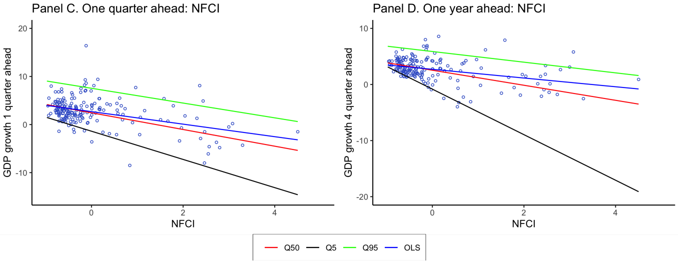


Figure: Quantile Regression

Univariate Analysis



(a) IVXQR-ivqr



(b) IVXQR-see

Bivariate Analysis

Recall our model:

$$\hat{Q}_{ggdp_{t+h}|ggdp_t+NFCI_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} ggdp_t + \hat{\beta}_{2,\tau} NFCI_t$$

Objective:

- Determine if estimates across quantiles are meaningful.
- Compare the estimates using different models.
- Present predicted distribution for one- and four-quarter GDP growth.

Bivariate Analysis

How to obtain the confidence interval?

- Apply 4-order Vector Autoregression (VAR)

$$\begin{pmatrix} NFCI_t \\ ggdp_t \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \boldsymbol{X}' + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

- Restore with randomly generated residuals
- Record quantile estimates for each iterations
- Obtain confidence interval from the esimates

Bivariate Analysis

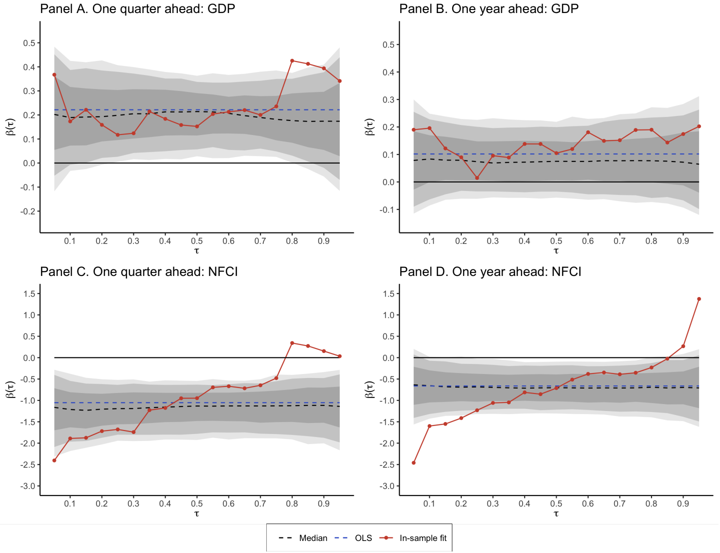


Figure: Estimated Quantile Regression Coefficients

Bivariate Analysis

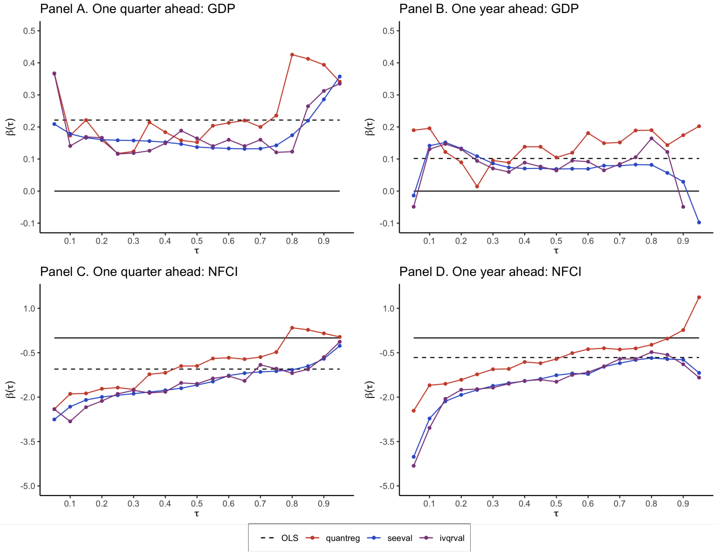


Figure: Compare across models

Bivariate Analysis

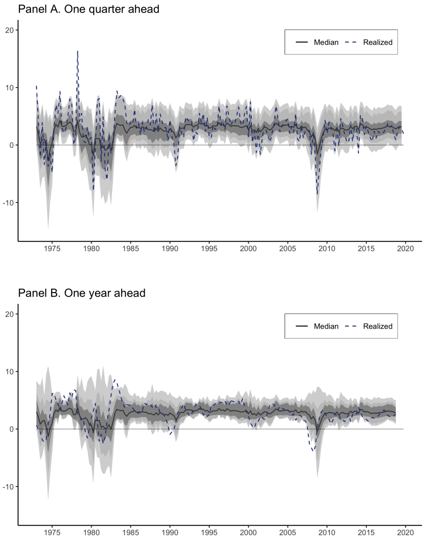
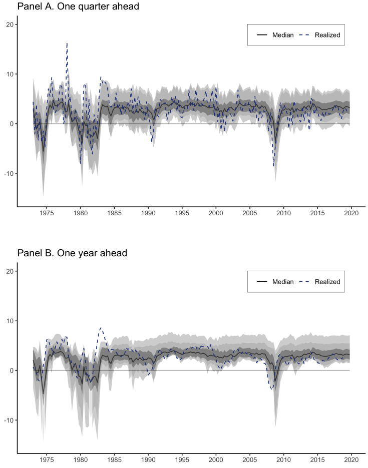
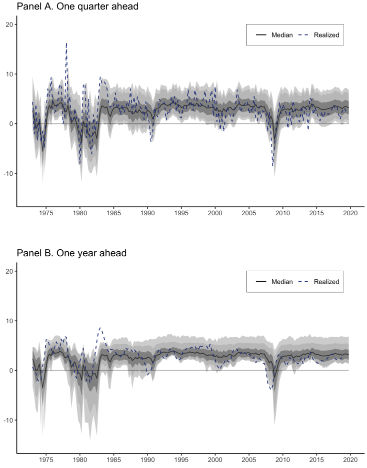


Figure: Predicted Distribution using QR

Bivariate Analysis

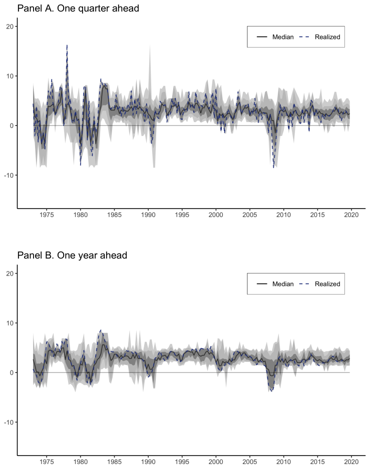


(c) IVXQR-ivqr

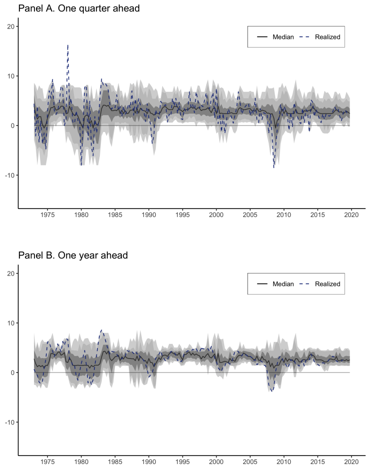


(d) IVXQR-see

Bivariate Analysis



(g) Quantile Regression Forest



(h) Generalized Quantile Random Forest

Out-of-sample Prediction

We conduct out-of-sample predictions using Expanding Window Forecast.

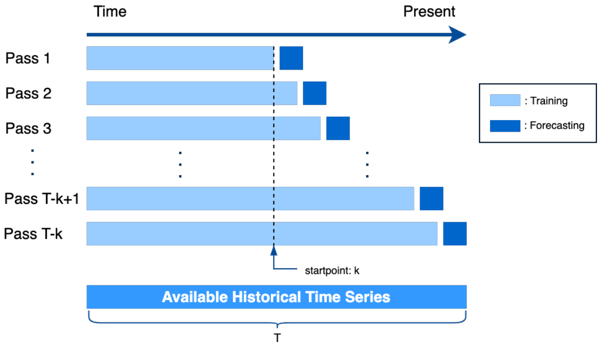


Figure: Expanding Window Forecast

Final Prediction Error:

$$\frac{1}{T-k}\sum_{s=k}^{T}\rho_\tau(y_s-\hat{y}_s)$$

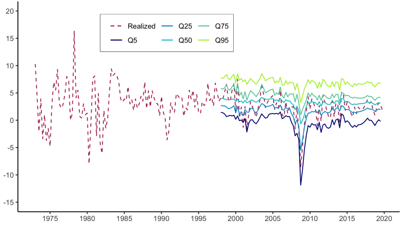
where $\hat{y}_s$ is prediction for τ. Based on FPE, we choose the optimal configurations for each models:

`rangerts`	1-quarter ahead	maxnode=11, minnode=10, blocksize=20
	4-quarter ahead	maxnode=12, minnode=10, blocksize=12
`grf`	1-quarter ahead	minnode=10
	4-quarter ahead	minnode=3

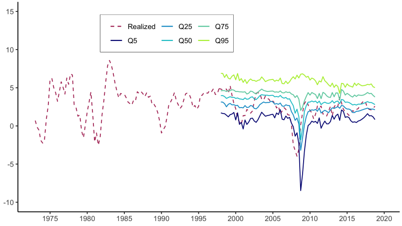
Out-of-sample Prediction

Result – Quantile Regression

Panel A. 1 quarter prediction

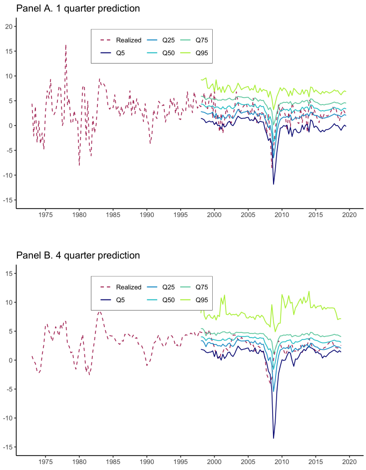


Panel B. 4 quarter prediction

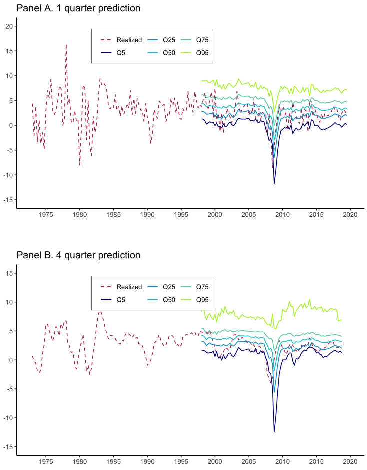


Out-of-sample Prediction

Result – IVXQR



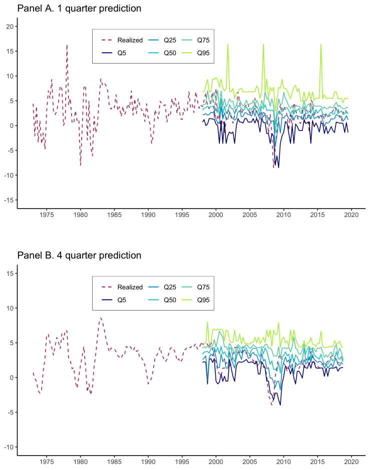
(e) ivqr



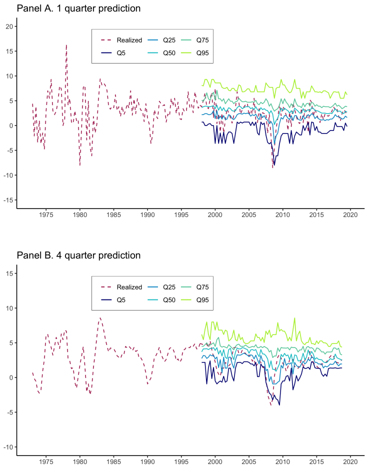
(f) see

Out-of-sample Prediction

Result – Quantile Random Forest



(i) quantile regression forest



(j) generalized quantile random forest

Final Prediction Error across models

A summary of FPE across different models:

	quantreg		rangerts		grf		ivqr		see	
	h1	h4	h1	h4	h1	h4	h1	h4	h1	h4
$\tau = 0.05$	0.2706	0.1977	0.3195	0.1615	0.2430	0.1379	0.2692	0.1625	0.2565	0.1599
$\tau = 0.25$	0.6752	0.5128	0.7432	0.5610	0.6848	0.4752	0.6772	0.5269	0.6779	0.5200
$\tau = 0.50$	0.8236	0.6649	0.8471	0.7000	0.8029	0.6345	0.8428	0.7159	0.8497	0.7178
$\tau = 0.75$	0.6939	0.5184	0.6966	0.5468	0.6736	0.5202	0.7143	0.5496	0.7479	0.5755
$\tau = 0.95$	0.2392	0.1823	0.2392	0.1648	0.2518	0.1796	0.2493	0.3087	0.2619	0.2882

- Lower FPE at h4 than h1.
- Lower FPE at the tails.
- Generalized Quantile Random Forest has the best performance.

References

Athey, Susan, Julie Tibshirani, and Stefan Wager. *Generalized Random Forests*. 2016.

Kaplan, David M. and Yixiao Sun. “SMOOTHED ESTIMATING EQUATIONS FOR INSTRUMENTAL VARIABLES QUANTILE REGRESSION”. In: *Econometric Theory* 33.1 (2017), pp. 105–157.

Magdalinos, Tassos and Peter C. B. Phillips. “Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors”. In: *Econometric Theory* 25.2 (2009), pp. 482–526.

Meinshausen, Nicolai. “Quantile Regression Forests”. In: *JOURNAL OF MACHINE LEARNING RESEARCH* 7 (2006), pp. 983–999.