New robust inference for predictive regressions

Anton Skrobotov
Russian Academy of National Economy and Public Administration
and Innopolis University
based on joint work with Rustam Ibragimov and Jihyun Kim

Research supported by a grant from Russian Science Foundaton (Project No. 16-18-10432)
The IAAE grant is gratefully acknowledged

Motivation – Endogeneity

- Consider the following predictive regression

$$y_t = \alpha + \beta x_{t-1} + u_t, \quad (1)$$
$$x_t = \rho x_{t-1} + v_t, \quad (2)$$

  for $t = 1, \ldots, T$, where x_t is some covariate.
- Our purpose to test the null hypothesis of no predictability of y_t (e.g., stock returns).
- In other words, we want to test $H_0 : \beta = 0$.
- If the process x_t be stationary then we estimate β by OLS and construct t-ratio, $t(\hat{\beta}_{OLS})$, which converges to standard Normal distribution.

Motivation – Endogeneity

- Consider the following predictive regression

$$
\begin{aligned}
y_t &= \alpha + \beta x_{t-1} + u_t, && (3)\\
x_t &= \rho x_{t-1} + v_t, && (4)
\end{aligned}
$$

  for $t = 1, \ldots, T$, where x_t is some covariate.
- If, however, x_t is (near) non-stationary,

$$x_t = \left(1 - \frac{c}{T}\right) x_{t-1} + v_t,$$

  we can use standard Normal inference only if the u_t and v_t are uncorrelated.

Motivation – Endogeneity

- Consider the following predictive regression

$$y_t = \alpha + \beta x_{t-1} + u_t, \quad (5)$$
$$x_t = \rho x_{t-1} + v_t, \quad (6)$$

- More presicely, let $\xi_i = (u_i, v_i)'$, and (appropriately normalizing) u_t converges to a bivariate Wiener process $([\sqrt{[rT]}]^{-1} \sum_{i=1}^{[rT]} \xi_i \Rightarrow B(s) = (U(r), V(r))')$ with covariance matrix

$$\Omega = \begin{pmatrix} \omega_y^2 & \omega_{xy} \\ \omega_{xy} & \omega_x^2 \end{pmatrix}$$

- It can be shown that $t(\hat{\beta}_{OLS}) \Rightarrow P + Q$, where Q is normal and P is function of Ornstein-Uhlenbeck process (near-unit root distributions).
- If $\omega_{xy} = 0$, then P vanishes from the limiting distribution, so that $t(\hat{\beta}_{OLS}) \Rightarrow N(0, 1)$. Otherwise, there are serious size distortions.

Motivation – Non-stationary volatility

- Consider the following predictive regression

$$\begin{aligned} y_t &= \alpha + \beta x_{t-1} + u_t, && (7) \\ x_t &= \rho x_{t-1} + v_t, && (8) \end{aligned}$$

- $u_t = \sigma_{t-1}\varepsilon_t$, where ε_t is MDS w.r.t. filtration $\mathcal{F}_{t-1}$, s.t. $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1$. Therefore, $E(u_t^{y,2}|\mathcal{F}_{t-1}) = \sigma_{t-1}^2$.
- The volatility process σ_t may be (nearly) non-stationary: $\sigma_t = \omega(z_t)$, where z_t is (near) unit root process, or $\sigma_t = \omega(t/T)$ (fixed or random), or $\sigma_t = \omega(z_t/\sqrt{T})$ (see Choi et al., 2016).
- Then the limiting distribution of $t(\hat{\beta}_{OLS})$ is non-normal, even when there is no endogeneity (no dependence between u_t and v_t).

The robust test for no predictability

- First assume $\alpha = 0$ in predictive regression. Also assume MDS assumption of u_t (no non-stationary volatility).
- Choi et al., 2016 proposed so called Cauchy estimator of β,

$$\hat{\beta}_T = \left(\sum_{t=1}^{T} |x_{t-1}| \right)^{-1} \sum_{t=1}^{T} sgn(x_{t-1}) y_t, \tag{9}$$

  where $sgn(\cdot)$ is a sign function such that $sign(x) = 1$ for $x \geq 0$ and $sign(x) = -1$ for $x < 0$.
- Extention 1: $\alpha \neq 0$ – recursive de-meaning (the limiting distribution is the same).
- Extention 2: u_t follows non-stationary volatility assumption – Time Change in continuous time framework (the limiting distribution is the same).

The robust test for no predictability

- It can be shown that under the null hypothesis of $\beta = 0$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \Rightarrow \int_0^1 \sigma(s) dW(s) =_d MN(0, Q),$$

where Q is some (random) variance depending on the volatility process ($Q = \int_0^1 \sigma(s)^2 ds$), and therefore

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} sgn(x_{t-1}) y_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} sgn(x_{t-1}) u_t \Rightarrow \int_0^1 \sigma(s) dW(s) =_d MN(0, Q).$$

- Therefore, the numerator of the Cauchy estimator,

$$\hat{\delta}_T := \sum_{t=1}^{T} sgn(x_{t-1}) y_t \tag{10}$$

has approximately mixed normal distribution with mean zero and variance $Q \times T$.

Brief review of t-statistic based robust inference

- Bakirov and Szekely (2005), Ibragimov and Mueller (2010, 2016): Usual small sample t-test of level $\alpha \leq 5\%$: conservative for independent heterogeneous Gaussian observations (not $\alpha = 10\%$)
- $X_j \sim N(\mu, \sigma_j^2)$, $j = 1, \dots, q$: $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$

  t-statistic $t = \sqrt{q}\frac{\bar{X}}{s_X}$

  $\bar{X} = q^{-1} \sum_{j=1}^{q} X_j$, $s_X^2 = (q-1)^{-1} \sum_{j=1}^{q} (X_j - \bar{X})^2$

  $cv_q(\alpha)$ = critical value of $T_{q-1} : P(|T_{q-1}| > cv_q(\alpha)) = \alpha$
- $P(|t| > cv(\alpha)|H_0) \leq P(|t| > cv(\alpha)|H_0, \sigma_1^2 = \cdots = \sigma_q^2) = \alpha$
- Holds under heavy tails, mixtures of normals (stable, Student-t)

Brief review of t-statistic based robust inference

- Regression: Assume data can be classified in a finite number q of groups that allow asymptotically independent normal inference about the (scalar) parameter of interest β, so that $\hat{\beta}_j \sim idN(\beta, v_j^2)$ for $j = 1, \ldots, q$. Time series example: Divide data into $q = 4$ consecutive blocks, and estimate the model 4 times.
- Treat $\hat{\beta}_j$ as observations for the usual t-statistic, and reject a 5% level test if t-statistic is larger than usual critical value for $q - 1$ degrees of freedom. Results in valid inference by small sample result.
- Exploits information $\hat{\beta}_j \sim idN(\beta, v_j^2)$ in an efficient way.
- Does not rely on single asymptotic model of sampling variability for estimated standard deviation

Monte Carlo Results

Same design as in Andrews (1991): Linear Regression, 5 regressors, 4 nonconstant regressors are independent draws from stationary Gaussian AR(1), as are the disturbances, $+$ heteroskedasticity. $T = 128$, 5% level test about coefficient of one nonconstant regressor.

	t-statistic (q)			$\hat{\omega}^2_{QA}$	$\hat{\omega}^2_{PW}$	$\hat{\omega}^2_{BT}(b)$			
	2	4	8			0.05	0.1	0.3	1
ρ					Size				
0	4.9	4.7	4.6	7.1	8.1	6.7	6.6	6.0	6.2
0.5	4.8	4.6	4.6	10.4	9.9	9.4	8.4	7.5	7.0
0.8	4.8	4.9	5.4	19.1	17.3	18.6	15.6	12.8	11.9
0.9	4.9	5.1	6.1	28.9	25.4	29.9	24.9	20.5	18.8
ρ					Size Adjusted Power				
0	15.1	38.4	53.7	62.7	60.6	60.7	58.6	51.9	47.2
0.5	14.5	38.2	55.9	57.0	56.2	56.0	53.5	48.4	44.2
0.8	15.4	45.1	66.0	52.9	51.7	54.0	52.6	46.9	42.4
0.9	17.2	56.7	77.6	57.5	54.6	58.7	57.5	51.4	46.6

The robust test for no predictability

- It can be shown that under the null hypothesis of $\beta = 0$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \Rightarrow \int_0^1 \sigma(s)dW(s) =_d MN(0,Q),$$

where Q is some (random) variance depending on the volatility process ($Q = \int_0^1 \sigma(s)^2 ds$), and therefore

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} sgn(x_{t-1})y_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} sgn(x_{t-1})u_t \Rightarrow \int_0^1 \sigma(s)dW(s) =_d MN(0,Q).$$

- Therefore, the numerator of the Cauchy estimator,

$$\hat{\delta}_T := \sum_{t=1}^{T} sgn(x_{t-1})y_t \qquad (11)$$

has approximately mixed normal distribution with mean zero and variance $Q \times T$.

The robust test for no predictability

- Consider partition of the data into $q \geq 2$ approximately equal groups $\mathcal{G}_j = \{s : (j-1)T/q < s \leq jT/q\}$, $j = 1, \ldots, q$. Then we have q subsamples for which we calculate q numerators of Cauchy estimator of (11), $\hat{\delta}_{T,j}$, $j = 1, \ldots, q$. Therefore, we have

$$(\hat{\delta}_{T,1}, \hat{\delta}_{T,2}, \ldots, \hat{\delta}_{T,q}) \Rightarrow N\left(0, diag(TQ_1, \ldots, TQ_q)\right) \tag{12}$$

under assumption about asymptotically independence of $\delta_{T,j}$.

- Following Ibragimov and Muller (2010), asymptotic Gaussianity then allow us to construct asymptotically valid (conservative) test of level $\alpha \leq 0.083$ of $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ by rejecting H_0 when $|t_{IV,\beta}|$ exceed $(1-\alpha/2)$ percentile of Student t-distribution with $q-1$ degrees of freedom, where $t_{IV,\beta}$ is constructed as

$$t_{IV,\beta} = \sqrt{q} \times \bar{\hat{\delta}}/s_{\hat{\delta}} \tag{13}$$

with $\bar{\hat{\delta}} = q^{-1}\sum_{j=1}^{q} \hat{\delta}_{T,j}$ and $s_{\hat{\delta}} = (q-1)^{-1}\sum_{j=1}^{q}\left(\hat{\delta}_{T,j} - \bar{\hat{\delta}}\right)^2$.

The robust test for no predictability

- Consider partition of the data into $q \geq 2$ approximately equal groups $\mathcal{G}_j = \{s : (j-1)T/q < s \leq jT/q\}$, $j = 1, \ldots, q$. Then we have q subsamples for which we calculate q numerators of Cauchy estimator of (11), $\hat{\delta}_{T,j}$, $j = 1, \ldots, q$. Therefore, we have

$$(\hat{\delta}_{T,1}, \hat{\delta}_{T,2}, \ldots, \hat{\delta}_{T,q}) \Rightarrow N\left(0, diag(TQ_1, \ldots, TQ_q)\right) \qquad (12)$$

  under assumption about asymptotically independence of $\delta_{T,j}$.
- Following Ibragimov and Muller (2010), asymptotic Gaussianity then allow us to construct asymptotically valid (conservative) test of level $\alpha \leq 0.083$ of $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ by rejecting H_0 when $|t_{IV,\beta}|$ exceed $(1-\alpha/2)$ percentile of Student t-distribution with $q-1$ degrees of freedom, where $t_{IV,\beta}$ is constructed as

$$t_{IV,\beta} = \sqrt{q} \times \bar{\hat{\delta}} / s_{\hat{\delta}} \qquad (13)$$

  with $\bar{\hat{\delta}} = q^{-1} \sum_{j=1}^{q} \hat{\delta}_{T,j}$ and $s_{\hat{\delta}} = (q-1)^{-1} \sum_{j=1}^{q} \left(\hat{\delta}_{T,j} - \bar{\hat{\delta}}\right)^2$.

Finite Sample Properties

- We use continuous time model as in Choi et al. (2016)
-

$$dY_t = \frac{\bar{\beta}}{T} X_t dt + dU_t, \tag{14}$$

$$dX_t = -\frac{\bar{\kappa}}{T} X_t dt + \sigma_t dV_t, \tag{15}$$

$$dU_t = \sigma_t \left(dW_t + \int_{\mathbb{R}} x \Lambda(dt, dx) \right), \tag{16}$$

  where V_t and W_t are Brownian motion with correlation coefficient -0.98 and $\alpha = 0$. The continuous data are generated to b eobserved at δ-intervals over T years, so that there are δT daily observations. See Choi et al. (2016).
- Compare with Bonferroni Q-test of Campbell and Yogo, 2006 (BQ) and restricted likelihood ratio test of Chen and Deo, 2009 (RLRT)

Finite Sample Properties

Volatility process:

- Model CNST. *Constant volatility*: $\sigma_t^2 = \sigma_0^2$, $\sigma_0 = 1$.
- Model SB. *Structural break in volatility*: $\sigma_0 + (\sigma_1 - \sigma_0)\mathbb{I}(t \geq 4T/5)$ with $\sigma_0 = 1$ and $\sigma_1 = 4$.
- Model GBM. *Geometric Brownian motion*: $d\sigma_t^2 = \frac{1}{2}\frac{\bar{\omega}^2}{T}\sigma_t^2 dt + \frac{\bar{\omega}^2}{\sqrt{T}}\sigma_t^2 dZ_t$, Z_t is Brownian motion which correlated with W_t with correlation coefficient -0.4 and $\bar{\omega}$ is set to be 9.
- Model RS. *Regime switching*: $\sigma_t = \sigma_0(1 - s_t) + \sigma_1 s_t$, where s_t be a homogeneous Markov process indicating the current state of the world which is independent on both Y_t and X_t with transition matrix with the state space $\{0, 1\}$

$$P_t = \begin{pmatrix} 0.8 & 0.2 \\ 0.8 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.2 & -0.2 \\ -0.8 & 0.8 \end{pmatrix} \exp\left(-\frac{\bar{\lambda}}{T}t\right)$$

  with $\bar{\lambda} = 60$, $\sigma_0 = 1$ and $\sigma_1 = 4$. s_t is initialized by its invariant distribution.

Finite Sample Property

Constant volatility: $\sigma_t = \sigma$

Table: Sizes of tests

		$\bar{\kappa} = 0$			$\bar{\kappa} = 5$			$\bar{\kappa} = 10$		
		5	20	50	5	20	50	5	20	50
CNST	OLS	42.2	42.0	43.0	19.5	19.5	19.7	11.1	11.2	10.9
	BQ	8.6	4.9	4.3	7.5	4.5	4.2	8.6	4.1	3.2
	RLRT	8.5	7.7	8.1	5.4	5.9	5.6	4.8	5.2	5.3
	Cauchy RT	5.3	4.9	5.3	5.2	5.4	4.7	5.5	5.1	5.1
	q=4	5.3	5.2	5.1	5.3	4.8	5.3	5.2	5.3	4.8
	q=8	5.1	4.9	5.0	5.1	5.1	4.9	5.0	5.6	4.8
	q=12	5.4	5.0	5.2	5.2	5.0	5.2	4.9	5.3	5.0
	q=16	5.1	4.9	5.1	4.9	4.9	5.1	4.8	5.6	4.9

Finite Sample Properties

Single break in vlatility: $\sigma_t = 1$ for $t \in [0, 4T/5]$ and $\sigma_t = 4$ for $t \in [4T/5, T]$

Table: Sizes of tests

		$\bar{\kappa} = 0$			$\bar{\kappa} = 5$			$\bar{\kappa} = 10$		
		5	20	50	5	20	50	5	20	50
SB	OLS	38.3	38.8	39.9	29.6	30.8	31.2	24.3	26.4	26.0
	BQ	18.1	12.9	11.9	17.0	15.1	14.1	17.4	14.8	14.3
	RLRT	23.8	22.8	23.6	21.0	21.9	21.8	22.4	24.5	23.6
	Cauchy RT	5.6	5.0	5.1	5.2	5.3	5.0	5.4	5.0	4.9
	q=4	3.0	2.9	3.2	3.2	2.8	3.4	3.2	3.3	2.8
	q=8	4.3	3.5	4.1	4.1	3.9	3.7	3.6	4.0	3.8
	q=12	5.1	4.2	4.5	4.4	4.4	4.4	4.2	4.2	4.6
	q=16	4.9	4.6	4.8	4.7	4.7	4.4	4.4	4.7	4.6

Finite Sample Properties

Regime switching model: $\sigma_t = \sigma_0(1 - s_t) + \sigma_1 s_t$, s_t is homogeneous Markov process independent of both Y and X

Table: Sizes of tests

		$\bar{\kappa} = 0$			$\bar{\kappa} = 5$			$\bar{\kappa} = 10$		
		5	20	50	5	20	50	5	20	50
RS	OLS	42.9	43.6	44.6	22.0	23.4	24.5	14.9	18.9	19.5
	BQ	8.8	6.3	6.0	9.8	7.2	6.8	12.6	8.9	8.4
	RLRT	9.3	10.0	10.7	7.5	9.4	9.6	9.6	13.0	14.2
	Cauchy RT	5.0	4.8	5.2	4.9	4.9	4.9	5.4	5.1	4.8
	q=4	4.4	4.2	4.9	4.8	5.1	4.5	4.8	5.1	4.7
	q=8	4.6	4.4	4.9	5.0	4.8	4.5	4.8	4.9	4.6
	q=12	4.8	4.4	5.1	4.9	5.1	4.3	4.9	4.7	4.8
	q=16	5.0	4.6	5.3	4.8	4.8	4.7	5.0	4.9	4.6

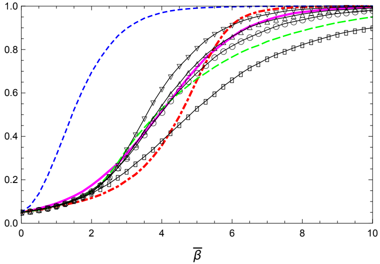
Finite Sample Properties

Geometric Brownian motion: $d\sigma_t^2 = \frac{\omega^2}{2}\sigma_t^2 dt + \omega\sigma_t^2 dZ_t$, Z_t – BM, correlated with W (with -0.4), $\omega = 9/\sqrt{T}$.

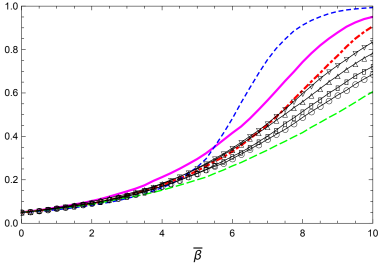
Table: Sizes of tests

		$\bar{\kappa}=0$			$\bar{\kappa}=5$			$\bar{\kappa}=10$		
		5	20	50	5	20	50	5	20	50
GBM	OLS	51.9	54.0	53.7	28.6	31.5	31.3	23.2	25.6	27.4
	BQ	16.9	12.5	11.6	13.9	12.9	12.9	15.8	11.7	13.0
	RLRT	21.3	22.6	22.5	16.3	18.3	18.7	21.8	23.9	24.4
	Cauchy RT	3.7	4.4	4.6	4.1	4.6	4.3	4.2	4.6	4.8
	q=4	2.4	2.3	2.1	2.6	2.4	2.4	2.4	2.4	2.6
	q=8	3.1	3.3	3.0	3.1	3.3	3.2	3.1	3.2	3.5
	q=12	3.8	4.0	3.8	3.7	4.0	3.7	3.6	3.6	4.1
	q=16	4.0	4.0	3.7	3.9	4.4	4.0	4.0	3.9	4.3

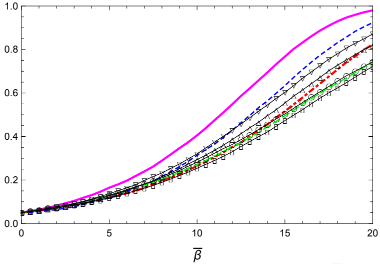
OLS: ━━ ,Bonf.Q: ━ · ━ ,RLRT: ╺ ╺ ,CauchyRT: ╺ ╺ ,q=4: □ ,q=8: △ ,q=12: ▽ ,q=16: ○



(a) CNST, $\bar{\kappa} = 0$, $T = 5$

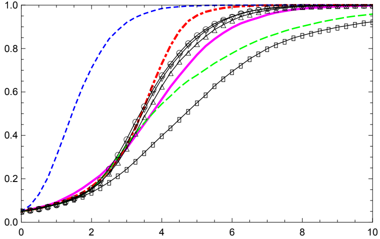


(b) CNST, $\bar{\kappa} = 5$, $T = 5$

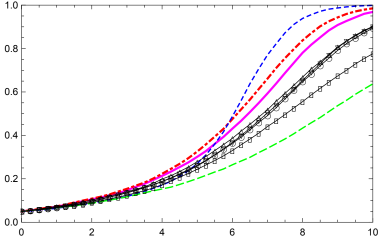


(c) CNST, $\bar{\kappa} = 20$, $T = 5$

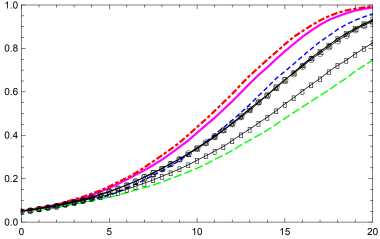
OLS: ,Bonf.Q: ,RLRT: ,CauchyRT: ,q=4: ,q=8: ,q=12: ,q=16:



(a) CNST, $\bar{\kappa} = 0$, $T = 50$

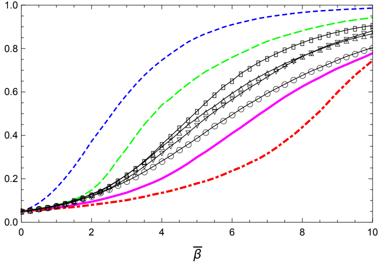


(b) CNST, $\bar{\kappa} = 5$, $T = 50$

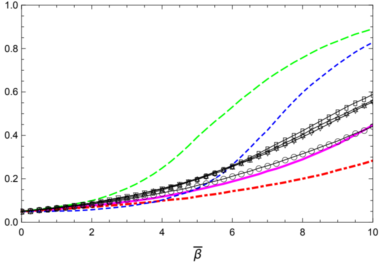


(c) CNST, $\bar{\kappa} = 20$, $T = 50$

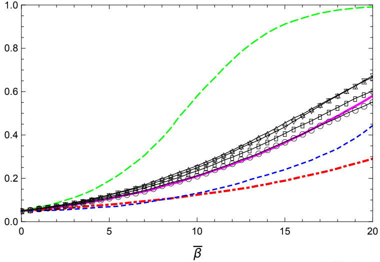
OLS: ,Bonf.Q: ,RLRT: ,CauchyRT: ,q=4: ,q=8: ,q=12: ,q=16:



(a) SB, $\bar{\kappa} = 0$, $T = 5$

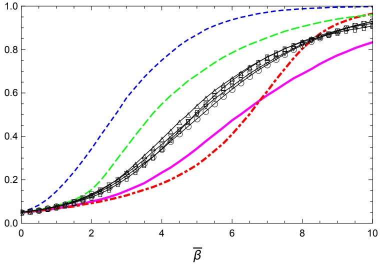


(b) SB, $\bar{\kappa} = 5$, $T = 5$

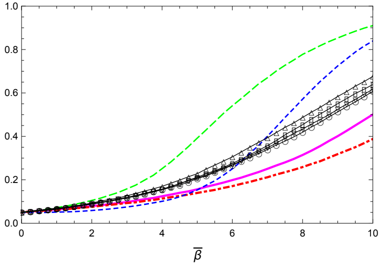


(c) SB, $\bar{\kappa} = 20$, $T = 5$

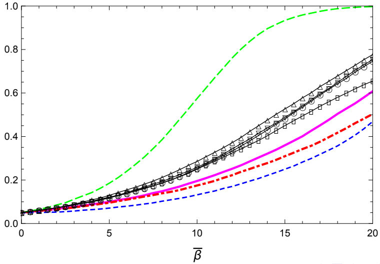
OLS: ,Bonf.Q: ,RLRT: ,CauchyRT: ,q=4: ,q=8: ,q=12: ,q=16:



(a) SB, $\bar{\kappa} = 0$, $T = 50$

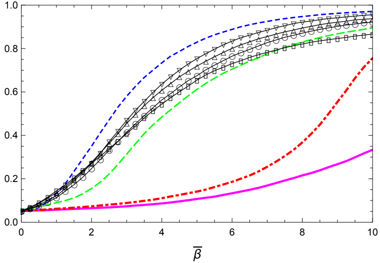


(b) SB, $\bar{\kappa} = 5$, $T = 50$

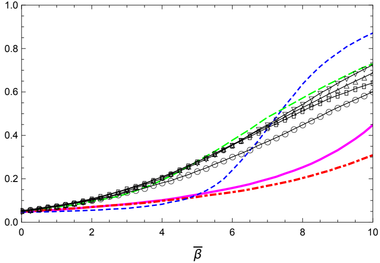


(c) SB, $\bar{\kappa} = 20$, $T = 50$

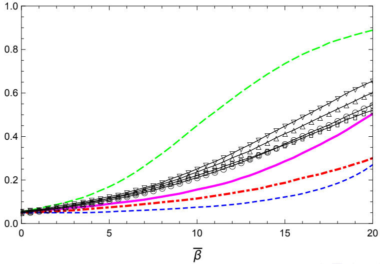
OLS: ,Bonf.Q: ,RLRT: ,CauchyRT: ,q=4: ,q=8: ,q=12: ,q=16:



(a) GBM, $\bar{\kappa} = 0$, $T = 5$

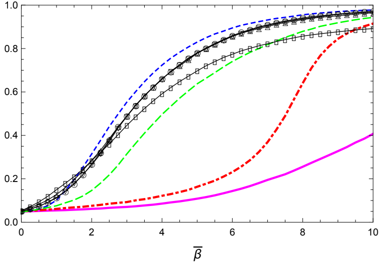


(b) GBM, $\bar{\kappa} = 5$, $T = 5$

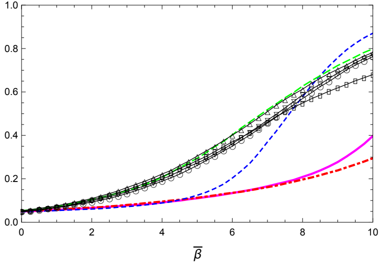


(c) GBM, $\bar{\kappa} = 20$, $T = 5$

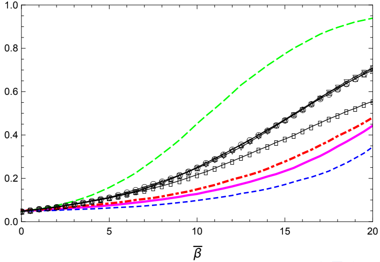
OLS: ,Bonf.Q: ,RLRT: ,CauchyRT: ,q=4: ,q=8: ,q=12: ,q=16:



(a) GBM, $\bar{\kappa} = 0$, $T = 50$

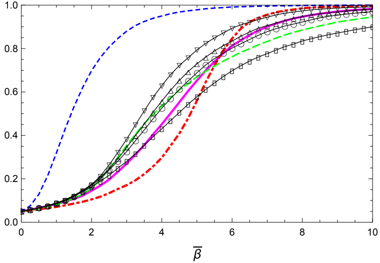


(b) GBM, $\bar{\kappa} = 5$, $T = 50$

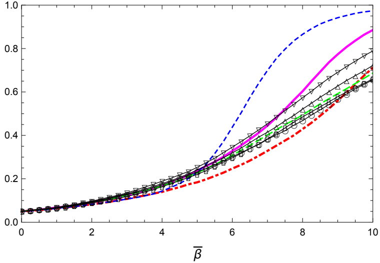


(c) GBM, $\bar{\kappa} = 20$, $T = 50$

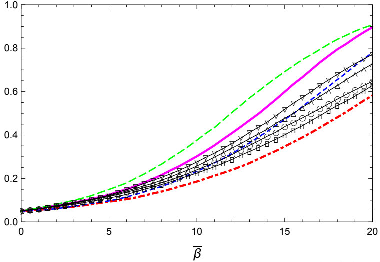
OLS: ,Bonf.Q: ,RLRT: ,CauchyRT: ,q=4: ,q=8: ,q=12: ,q=16:



(a) RS, $\bar{\kappa} = 0$, $T = 5$

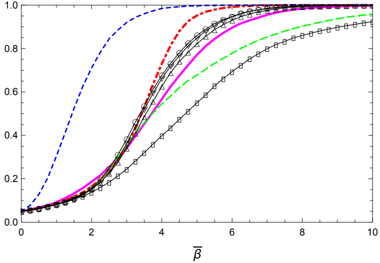


(b) RS, $\bar{\kappa} = 5$, $T = 5$

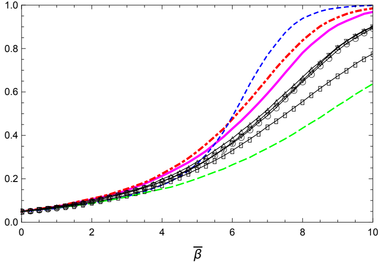


(c) RS, $\bar{\kappa} = 20$, $T = 5$

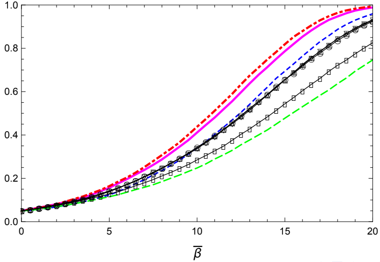
OLS: ,Bonf.Q: ,RLRT: ,CauchyRT: ,q=4: ,q=8: ,q=12: ,q=16:



(a) RS, $\bar{\kappa} = 0$, $T = 50$



(b) RS, $\bar{\kappa} = 5$, $T = 50$



(c) RS, $\bar{\kappa} = 20$, $T = 50$

Concluding remarks

- New approach to robust inference in predictive regression
- Based on instrumental variable estimator (Cauchy estimator) which allows the endogeneity between variables
- The inference then based on splitting the sample and obtaining robust Student t-test
- The obtaining approach is robust to a wide class of errors: dependence, heterockedasticity, nonstationary volatility and heavy tails

Thank you for attention!