How Good Is Likelihood Asymptotic Inference of the First Order?

Kai Wang Ng

Department of Statistics and Actuarial Science
The University of Hong Kong, Hong Kong

Email: kaing@hku.hk

Slides of the talk at DASF III, Fields Institute
Toronto, 1 May 2010, 10:15-11:00 am

A Memory of

an Advice I Heard on an Island:

Scholarship is more than publication

My concerns in this talk

- Large-sample theories so far **do not provide** error bounds between the exact and asymptotic CDF as a function of n, nor any guideline on n.
- With high respect and deep trust in theories, large-sample practitioners act **as if** no requirement on n is needed.
- An obvious example of such **pervasive** practice is in the areas of econometrics and mathematical finance, where the modeling is more and more complex with many parameters while sample size seems never a concern.

First order asymptotic inference ("delta method"):

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
L(x_1,\ldots,x_n;\theta) \arrow[r, "\text{MLE } \hat{\theta}_n"] \arrow[d, leftrightarrow, "\tau = g(\theta)"', "\text{One to one}"] & \hat{\theta}_n \sim N(\theta, \sigma^2(\theta)) \ as\ n \to \infty \arrow[d, leftrightarrow, "\text{Both normal?}"] \\
L(x_1,\ldots,x_n;\tau) \arrow[r, "\text{MLE } g(\hat{\theta}_n)"] & g(\hat{\theta}_n) \sim N(g(\theta), \sigma^2(\theta)(g'(\theta))^2) \ as\ n \to \infty
\end{tikzcd}
$$

Basic: If X is normal, $g(X)$ **cannot** be normal **unless** $g(\cdot)$ is linear.

Logic from basic: For any non-linear $g(\cdot)$, $\hat{\theta}_n$ and $g(\hat{\theta}_n)$ cannot be almost normal simultaneously with the same sample size.

First order asymptotics $\bigcap$ Logic from basic => **Cautions**

Caution (a):

Sample Size is Crucial

Caution (b):

A Statistic's True Nature Is Vital

Caution (a): [First page]
Sample Size is Crucial

The need for this caution is demonstrated with the simplest ARMA(p, q) model where $p = q = 1$:

$$(x_t - \mu) - \phi(x_{t-1} - \mu) = z_t - \theta z_{t-1}$$

where $|\phi| < 1, |\theta| < 1$, and $\phi \neq \theta$ (otherwise not identifiable).

Asymptotic normal distributions of MLE:

$$\hat{\phi} \sim N\left(\phi, \quad \frac{(1-\phi^2)(1-\phi\theta)^2}{n(\phi-\theta)^2}\right)$$

$$\hat{\theta} \sim N\left(\theta, \quad \frac{(1-\theta^2)(1-\phi\theta)^2}{n(\phi-\theta)^2}\right)$$

Typical sample size

Box + Jenkins (1976)

Collection of Time Series Used for Examples in the Text

Series A	Chemical process concentration readings: every two hours. 197
Series B	IBM common stock closing prices: Daily, 17th May 1961–2nd November 1962. 369
Series B′	IBM common stock closing prices: Daily, 29th June 1959–30th June 1960. 255
Series C	Chemical process temperature readings; every minute. 226
Series D	Chemical process viscosity readings: every hour. 310
Series E	Wölfer sunspot numbers: yearly. 100
Series F	Series of 70 consecutive yields from a batch chemical process (tabulated in Table 2.1 in the text). 70
Series G	International airline passengers: Monthly totals (thousands of passengers) January 1949–December 1960. 144
Series J	Gas furnace data. 296
Series K	Simulated dynamic data with two inputs. 64
Series L	Pilot scheme data. 310
Series M	Sales data with leading indicator. 150

Typical sample size

William Wei (2006)

TABLE 7.3 Summary of models fitted to Series W1–W7 (values in the parentheses under estimates refer to the standard errors of those estimates).

Series	No. of observations	Fitted models	$\hat{\sigma}_a^2$
W1	45	$(1 - \underset{(.134)}{.43}B)Z_t = \underset{(.25)}{1.04} + a_t$	.2
W2	302	$(1 - \underset{(.04)}{1.41}B + \underset{(.04)}{.7}B^2)\sqrt{Z_t} = \underset{(.17)}{1.88} + a_t$	1.37
		$(1 - \underset{(.06)}{1.21}B + \underset{(.09)}{.49}B^2 + \underset{(.09)}{.12}B^3 - \underset{(.09)}{.24}B^4 + \underset{(.09)}{.23}B^5 - \underset{(.09)}{.01}B^6 - \underset{(.09)}{.17}B^7 + \underset{(.09)}{.22}B^8 - \underset{(.06)}{.31}B^9)\sqrt{Z_t} = \underset{(.28)}{.80} + a_t$	1.216
		$(1 - \underset{(.04)}{1.23}B + \underset{(.04)}{.52}B^2 - \underset{(.02)}{.2}B^9)\sqrt{Z_t} = \underset{(.23)}{.56} + a_t$	1.222
W3	82	$(1 - \underset{(.07)}{.73}B)Z_t = \underset{(313.36)}{1127.1} + a_t$	773,982.936
W4	500	$(1 - B)Z_t = (1 - \underset{(.04)}{.60}B)a_t$	1,324.727
W5	71	$(1 - B)Z_t = \underset{(.33)}{2.05} + a_t$	7.678
		$(1 - \underset{(.01)}{.98})Z_t = \underset{(1.36)}{4.63} + a_t$	7.283
W6	114	$(1 - B)\ln Z_t = (1 - \underset{(.07)}{.60}B)a_t$	.028
W7	55	$(1 - \underset{(.12)}{.97}B + \underset{(.18)}{.12}B^2 + \underset{(.13)}{.5}B^3)\ln Z_t = \underset{(.81)}{6.41} + a_t$	.124
		$(1 - \underset{(.06)}{1.55}B + \underset{(.06)}{.94}B^2)\ln Z_t = \underset{(.41)}{3.9} + (1 - \underset{(.12)}{.59}B)a_t$	.116

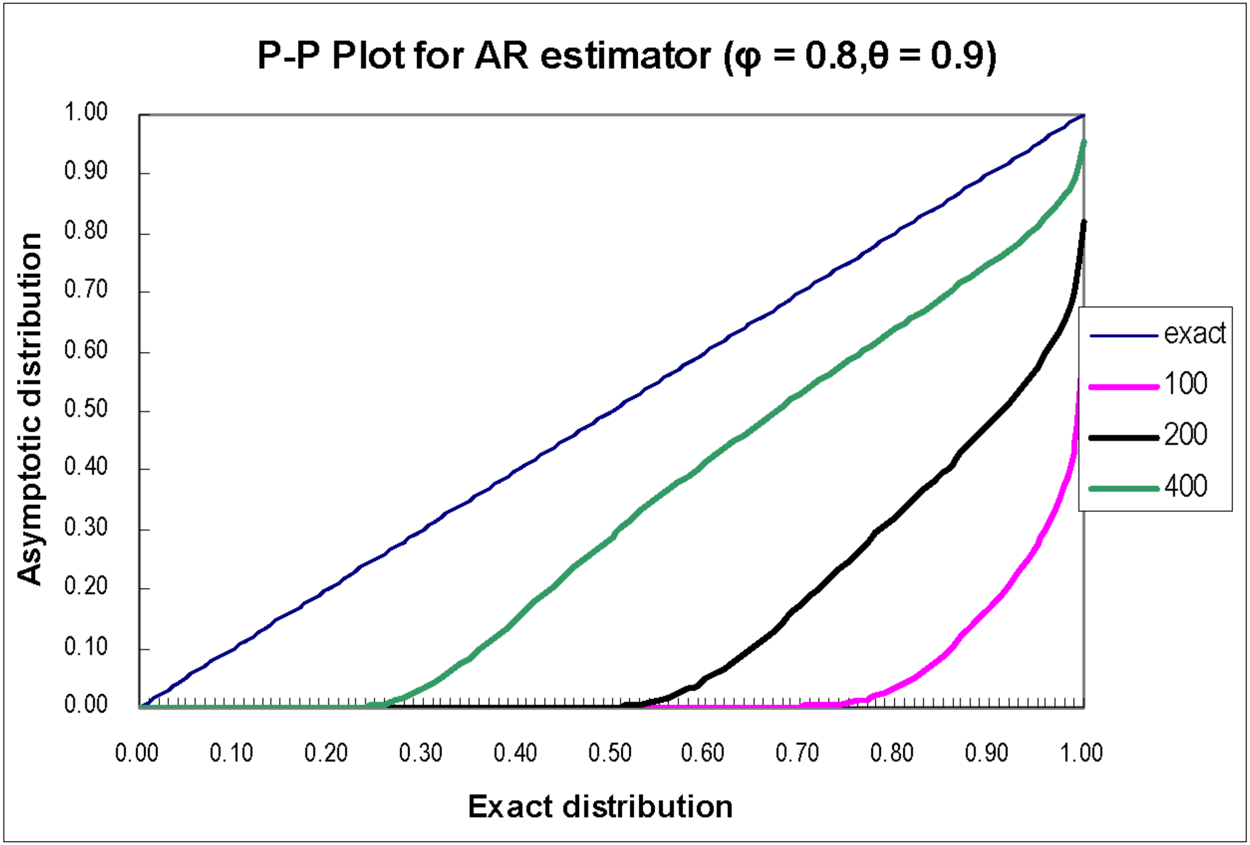
Simulation process for the exact distributions of $\hat{\phi}$ and $\hat{\theta}$:

$$
\left.\begin{array}{ccccc}
(\varphi, \theta) & \xrightarrow{\text{SAS/IML}} & (X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)}\} & \xrightarrow{\text{SAS/ARIMA}} & (\hat{\phi}_{(1)}, \hat{\theta}_{(1)}) \\
 & & (X_1^{(2)}, X_2^{(2)}, \ldots, X_n^{(2)}\} & \longrightarrow & (\hat{\phi}_{(2)}, \hat{\theta}_{(2)}) \\
 & & & \vdots & \vdots \\
 & & (X_1^{(m)}, X_2^{(m)}, \ldots, X_n^{(m)}\} & \longrightarrow & (\hat{\phi}_{(m)}, \hat{\theta}_{(m)})
\end{array}\right] \rightarrow \begin{array}{l}\text{Simulated}\\ \text{Exact}\\ \text{Distribution}\\ \text{for large } m\\ (m = 10,000)\end{array}
$$

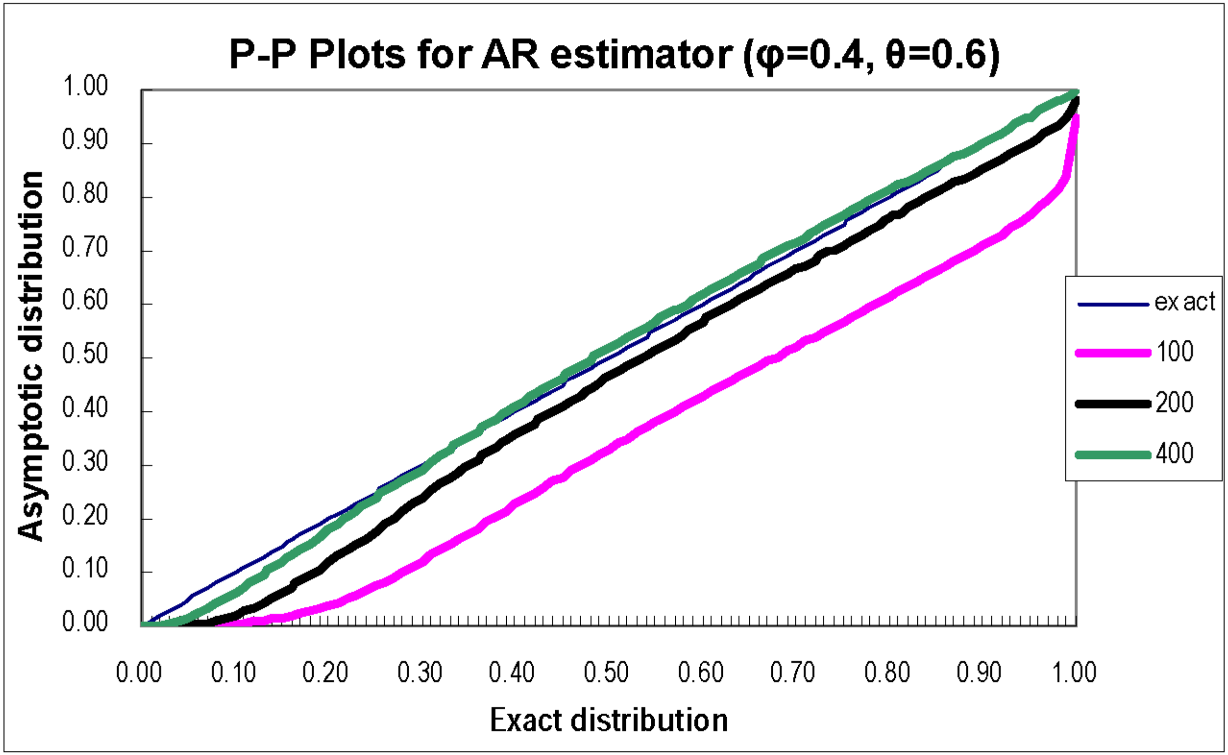
Fix $\mu = 10$ and $\sigma = 1$ throughout to focus on (ϕ, θ).

Acknowledgement: My SAS code to get the simulated distributions was implemented by an undergraduate project student, Christne Liang, and a student research assistant, Emily Jiang, the former generating all the pp-plots by Excell while the latter generating the CDF plots and tables by SAS/ODS.

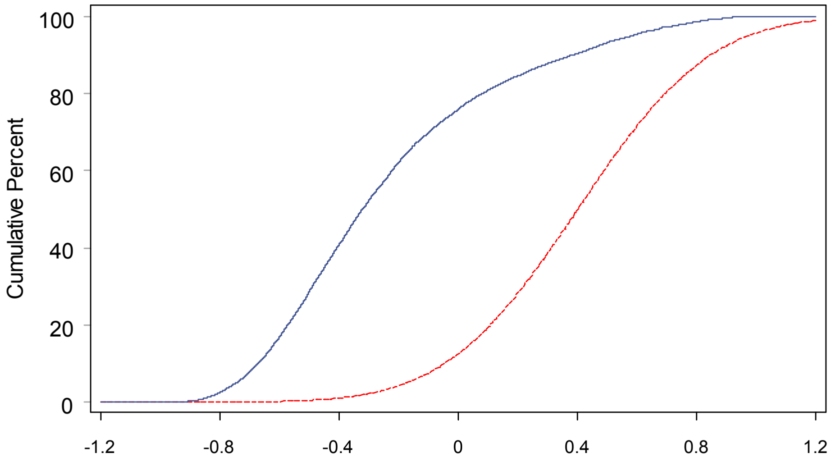
Asymptotic cdf vs. Simulated cdf



Asymptotic cdf vs. Simulated cdf



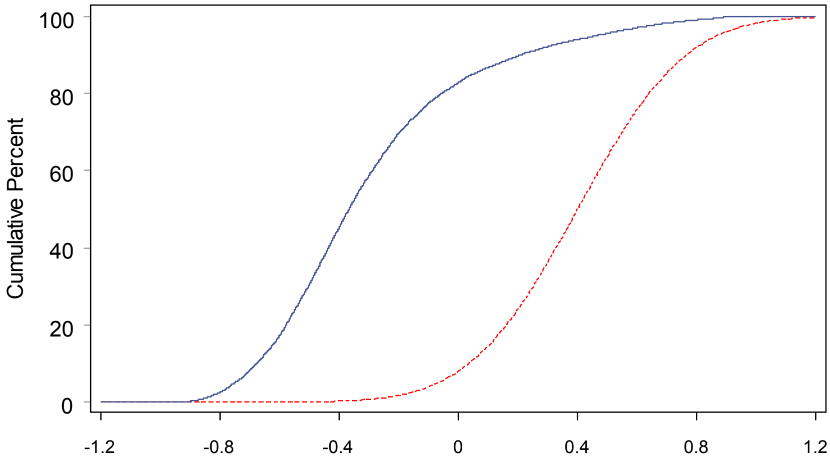
CDF of AR Estimator: Simulated Exact vs. Asymptotic



Simulated CDF (AR=0.4 MA=0.6 n=100 simulation=10000) for AR: ——

Normal Curve: -------- Mu=0.4, Sigma=0.3483

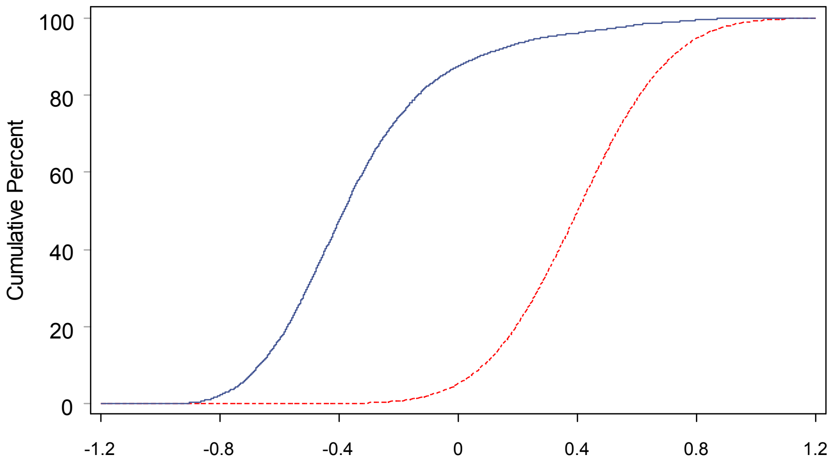
AR ar=0.4 ma=0.6	cum % simulation n=100	cum % normal
-0.9	0.28	0.01
-0.8	2.48	0.03
-0.7	7.88	0.08
-0.6	16.87	0.20
-0.5	28.74	0.49
-0.4	40.80	1.08
-0.3	52.33	2.22
-0.2	62.25	4.25
-0.1	69.85	7.56
-0.0	75.95	12.54
0.1	80.81	19.45
0.2	84.53	28.29
0.3	87.75	38.70
0.4	90.40	50.00
0.5	93.18	61.30
0.6	95.40	71.71
0.7	97.26	80.55
0.8	98.66	87.46
0.9	99.59	92.44
1.0	100.00	95.75



Simulated CDF (AR=0.4 MA=0.6 n=150 simulation=10000) for AR: ——

Normal Curve: -------- Mu=0.4, Sigma=0.2844

AR ar=0.4 ma=0.6	cum % simulation n=150	cum % normal
-0.9	0.19	0.00
-0.8	2.55	0.00
-0.7	8.21	0.01
-0.6	17.15	0.02
-0.5	30.46	0.08
-0.4	45.27	0.25
-0.3	58.61	0.69
-0.2	69.71	1.74
-0.1	77.52	3.93
-0.0	82.75	7.98
0.1	86.78	14.57
0.2	89.78	24.09
0.3	92.12	36.25
0.4	93.99	50.00
0.5	95.61	63.75
0.6	97.10	75.91
0.7	98.27	85.43
0.8	99.07	92.02
0.9	99.81	96.07
1.0	100.00	98.26

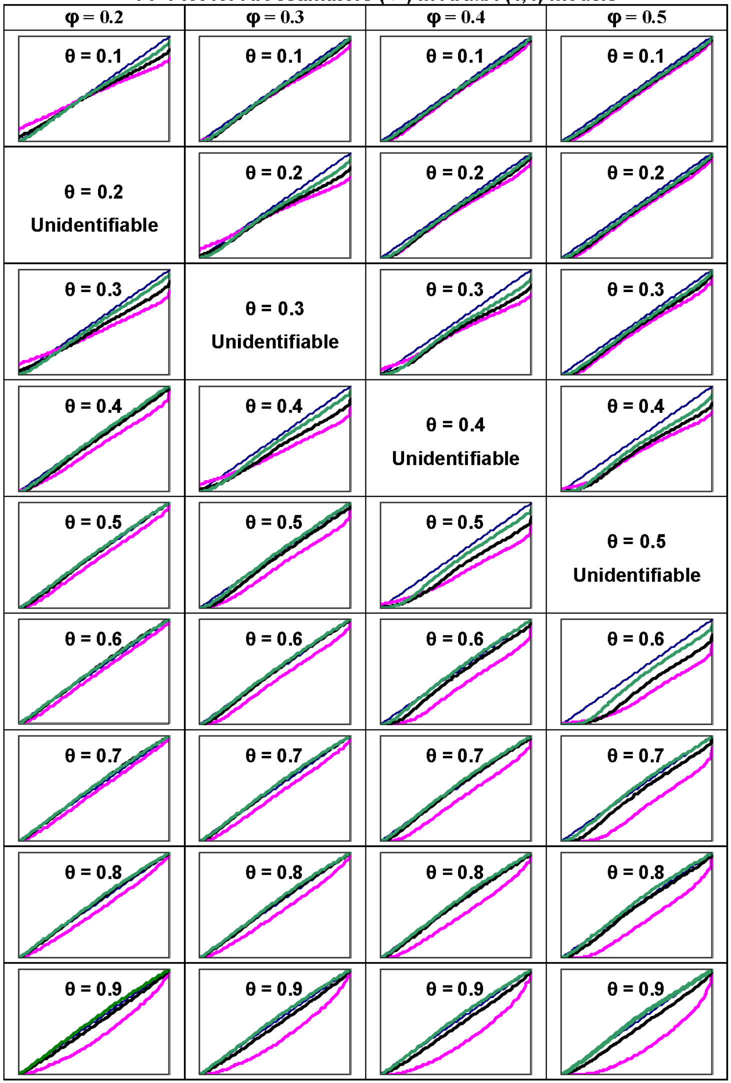


Simulated CDF (AR=0.4 MA=0.6 n=200 simulation=10000) for AR: ——

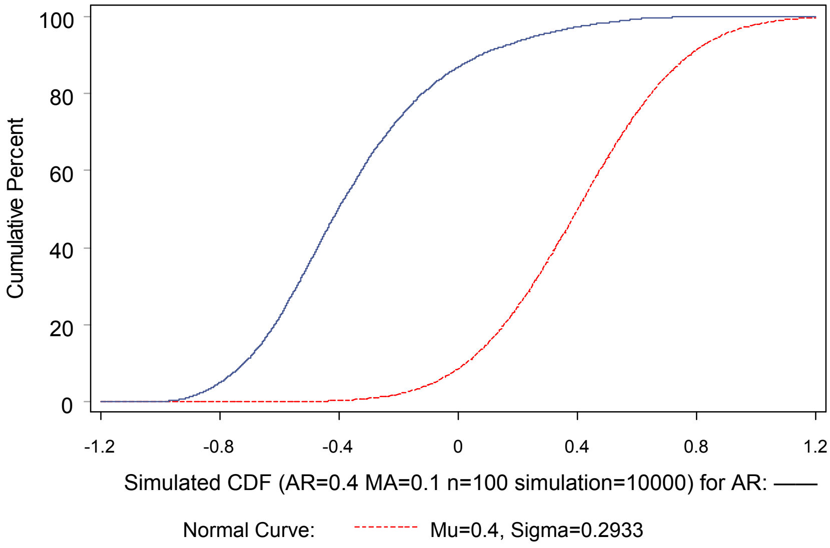
Normal Curve: -------- Mu=0.4, Sigma=0.2463

AR ar=0.4 ma=0.6	cum % simulation n=200	cum % normal
-0.9	0.23	0.00
-0.8	2.18	0.00
-0.7	7.29	0.00
-0.6	16.74	0.00
-0.5	31.09	0.01
-0.4	47.37	0.06
-0.3	62.72	0.22
-0.2	74.22	0.74
-0.1	82.50	2.12
-0.0	87.51	5.22
0.1	90.94	11.16
0.2	93.39	20.84
0.3	95.14	34.23
0.4	96.12	50.00
0.5	97.16	65.77
0.6	98.19	79.16
0.7	98.87	88.84
0.8	99.45	94.78
0.9	99.84	97.88
1.0	100.00	99.26

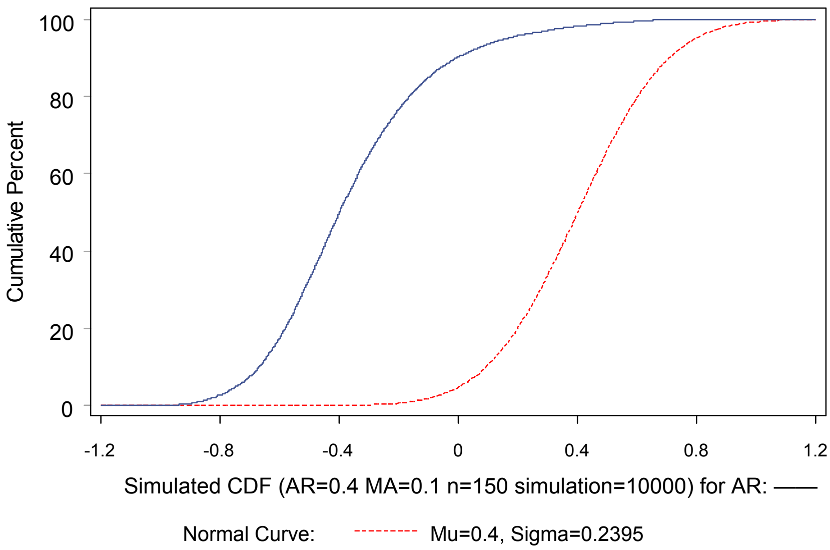
PP-Plot for AR estimators ($\hat{\phi}$) in ARMA (1,1) models



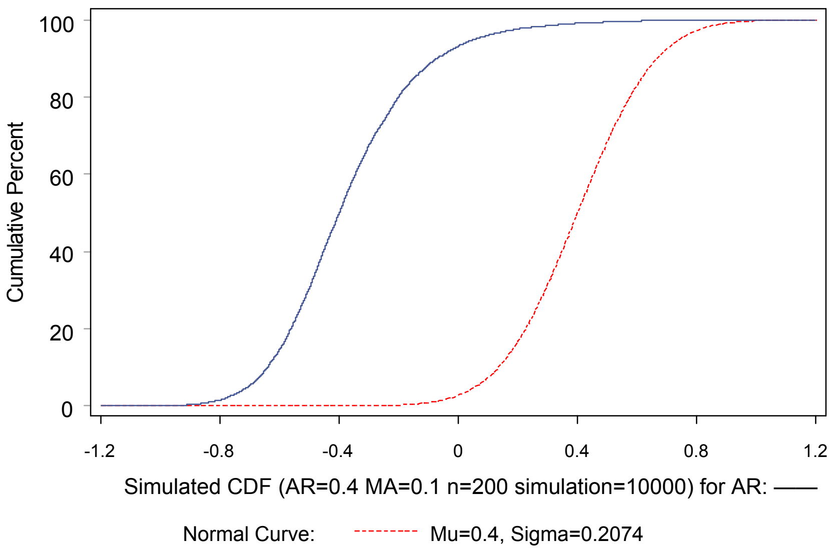
CDF of AR Estimator: Simulated Exact vs. Asymptotic



AR ar=0.4 ma=0.1	cum % simulation n=100	cum % normal
-0.9	1.26	0.00
-0.8	5.05	0.00
-0.7	11.36	0.01
-0.6	21.79	0.03
-0.5	35.90	0.11
-0.4	50.55	0.32
-0.3	63.28	0.85
-0.2	73.34	2.04
-0.1	81.22	4.41
-0.0	86.84	8.63
0.1	90.90	15.32
0.2	93.49	24.76
0.3	95.64	36.66
0.4	97.22	50.00
0.5	98.42	63.34
0.6	99.35	75.24
0.7	99.73	84.68
0.8	99.97	91.37
0.9	99.99	95.59
1.0	100.00	97.96

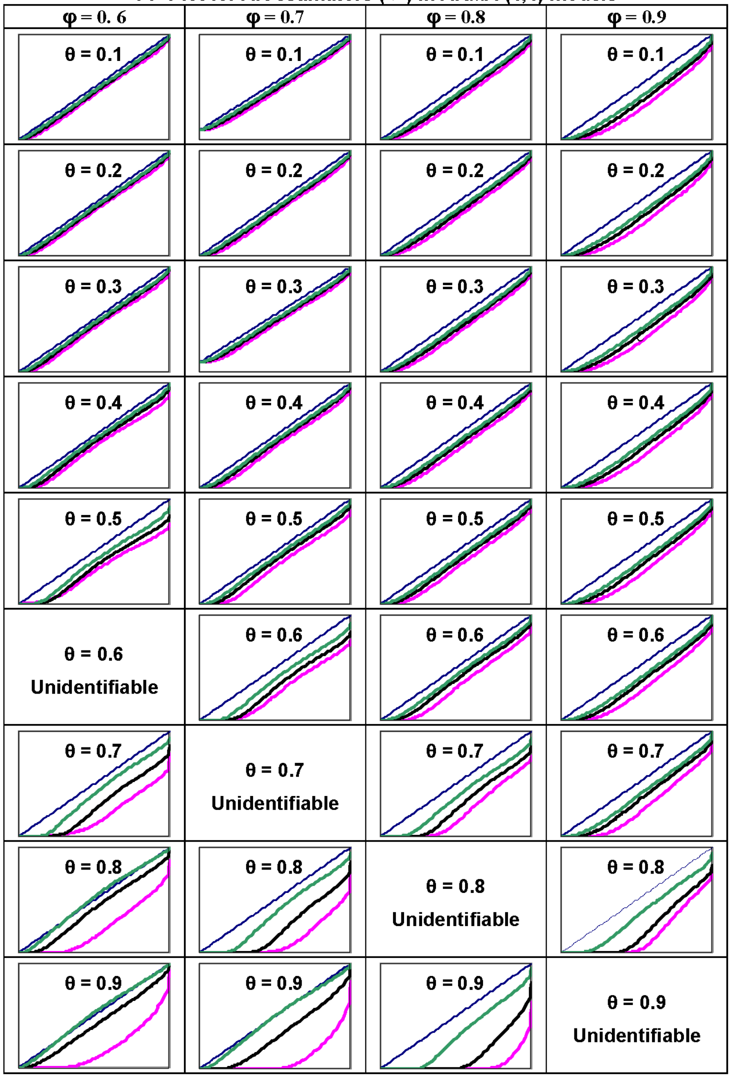


AR ar=0.4 ma=0.1	cum % simulation n=150	cum % normal
-0.9	0.45	0.00
-0.8	2.71	0.00
-0.7	7.47	0.00
-0.6	17.38	0.00
-0.5	32.63	0.01
-0.4	49.81	0.04
-0.3	64.88	0.17
-0.2	76.72	0.61
-0.1	85.01	1.84
-0.0	90.35	4.74
0.1	93.62	10.51
0.2	95.81	20.18
0.3	97.08	33.81
0.4	98.25	50.00
0.5	98.92	66.19
0.6	99.52	79.82
0.7	99.88	89.49
0.8	100.00	95.26

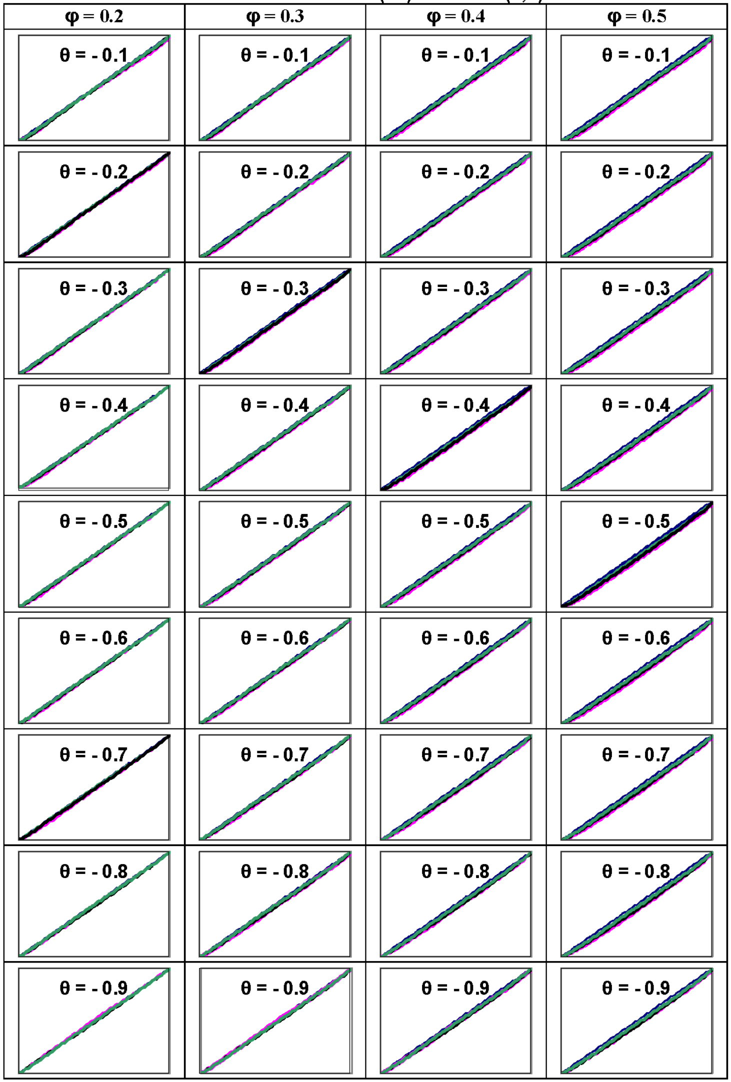


AR ar=0.4 ma=0.1	cum % simulation n=200	cum % normal
-0.9	0.27	0.00
-0.8	1.49	0.00
-0.7	5.26	0.00
-0.6	14.47	0.00
-0.5	30.42	0.00
-0.4	49.67	0.01
-0.3	67.44	0.04
-0.2	80.10	0.19
-0.1	88.35	0.80
-0.0	93.19	2.69
0.1	96.08	7.40
0.2	97.64	16.74
0.3	98.56	31.48
0.4	99.16	50.00
0.5	99.51	68.52
0.6	99.74	83.26
0.7	99.90	92.60
0.8	99.99	97.31
0.9	100.00	99.20

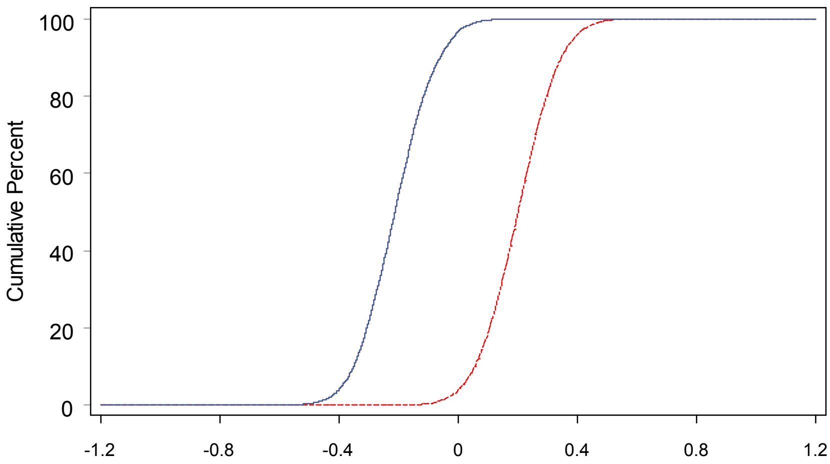
PP-Plot for AR estimators ($\hat{\phi}$) in ARMA (1,1) models



PP-Plot for AR estimators ($\hat{\phi}$) in ARMA (1,1) models



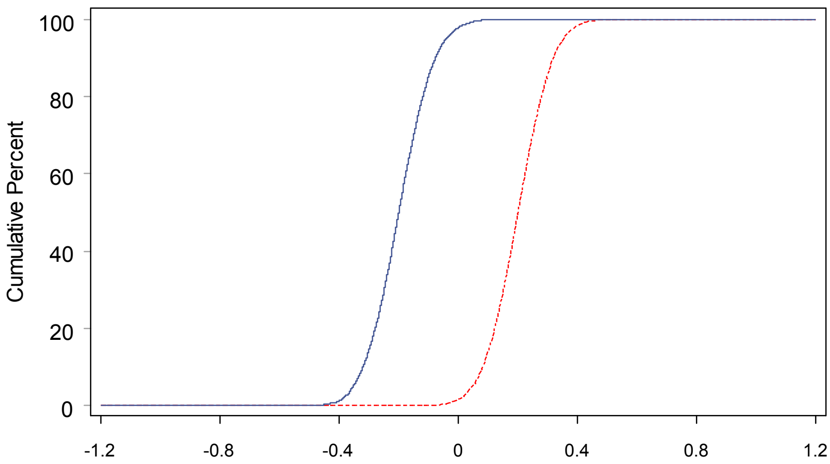
CDF of AR Estimator: Simulated Exact vs. Asymptotic



Simulated CDF (AR=0.2 MA=-0.8 n=100 simulation=10000) for AR: ——

Normal Curve: -------- Mu=0.2, Sigma=0.1137

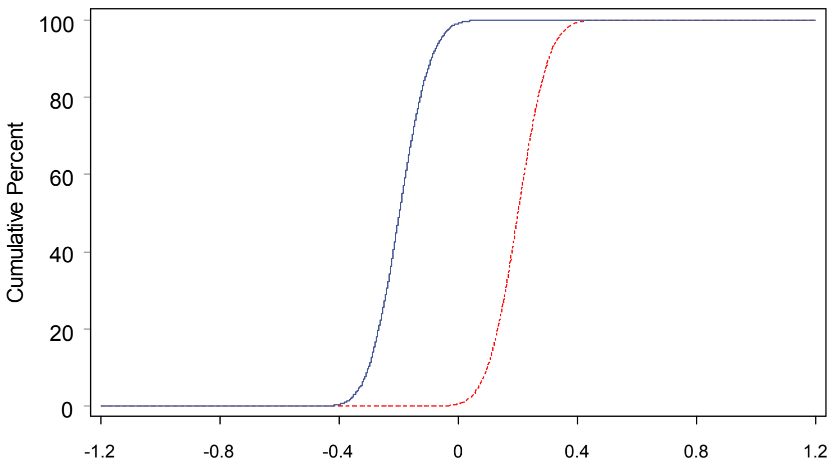
AR ar=0.2 ma=-0.8	cum % simulation n=100	cum % normal
-0.8	0.01	0.00
-0.6	0.03	0.00
-0.5	0.40	0.00
-0.4	4.10	0.00
-0.3	21.84	0.00
-0.2	54.78	0.02
-0.1	83.56	0.42
-0.0	96.77	3.92
0.1	99.69	18.95
0.2	99.97	50.00
0.3	100.00	81.05



Simulated CDF (AR=0.2 MA=-0.8 n=150 simulation=10000) for AR: ——

Normal Curve: -------- Mu=0.2, Sigma=0.0928

AR ar=0.2 ma=-0.8	cum % simulation n=150	cum % normal
-0.5	0.03	0.00
-0.4	1.27	0.00
-0.3	14.37	0.00
-0.2	49.96	0.00
-0.1	85.14	0.06
-0.0	97.86	1.56
0.1	99.89	14.06
0.2	100.00	50.00

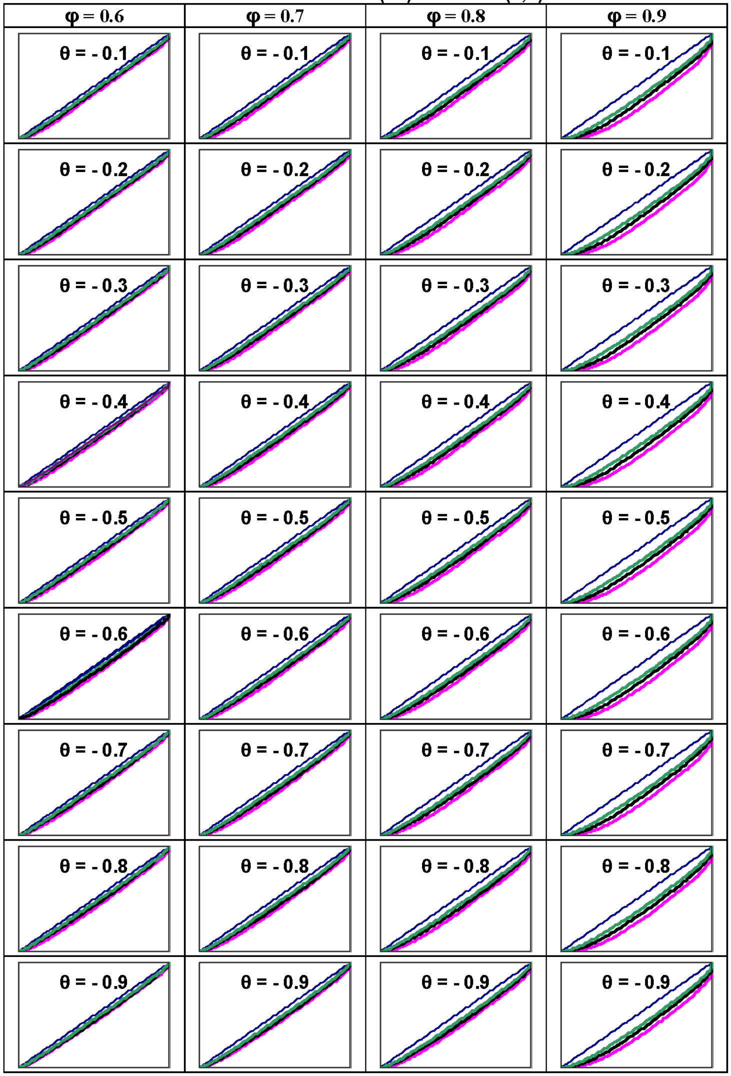


Simulated CDF (AR=0.2 MA=-0.8 n=200 simulation=10000) for AR: ——

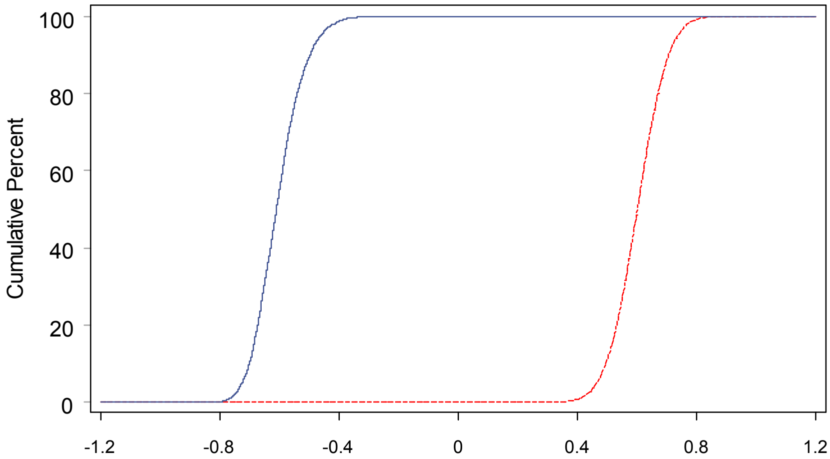
Normal Curve: -------- Mu=0.2, Sigma=0.0804

AR ar=0.2 ma=-0.8	cum % simulation n=200	cum % normal
-0.4	0.49	0.00
-0.3	10.24	0.00
-0.2	48.79	0.00
-0.1	87.83	0.01
-0.0	99.17	0.64
0.1	99.98	10.67
0.2	100.00	50.00

PP-Plot for AR estimators ($\hat{\phi}$) in ARMA (1,1) models



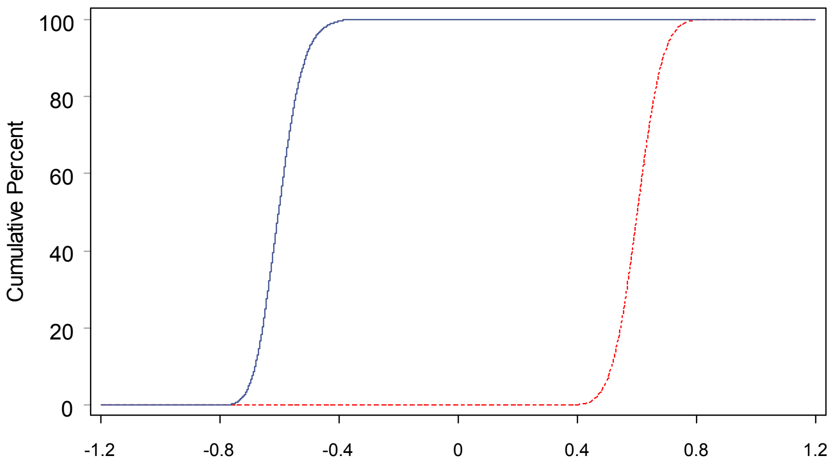
CDF of AR Estimator: Simulated Exact vs. Asymptotic



Simulated CDF (AR=0.6 MA=-0.9 n=100 simulation=10000) for AR: ——

Normal Curve: - - - - - - Mu=0.6, Sigma=0.0821

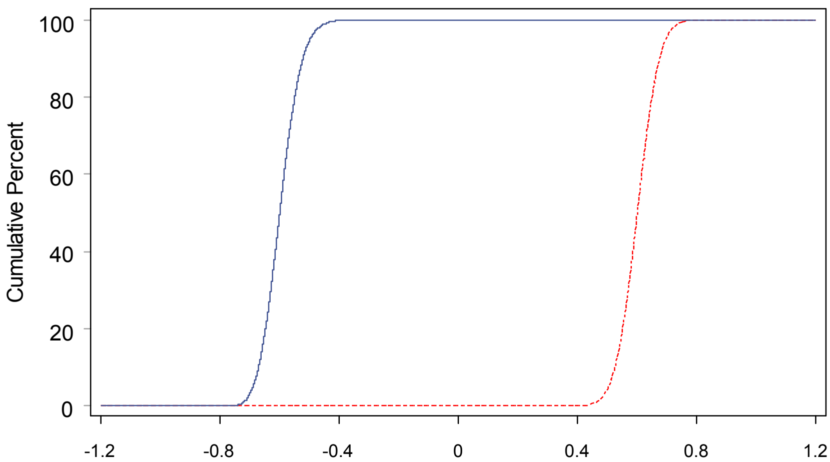
AR ar=0.6 ma=-0.9	cum % simulation n=100	cum % normal
-0.8	0.10	0.00
-0.7	10.38	0.00
-0.6	54.98	0.00
-0.5	89.29	0.00
-0.4	98.75	0.00
-0.3	99.97	0.00
-0.2	100.00	0.00



Simulated CDF (AR=0.6 MA=-0.9 n=150 simulation=10000) for AR: ——

Normal Curve: - - - - - - Mu=0.6, Sigma=0.0671

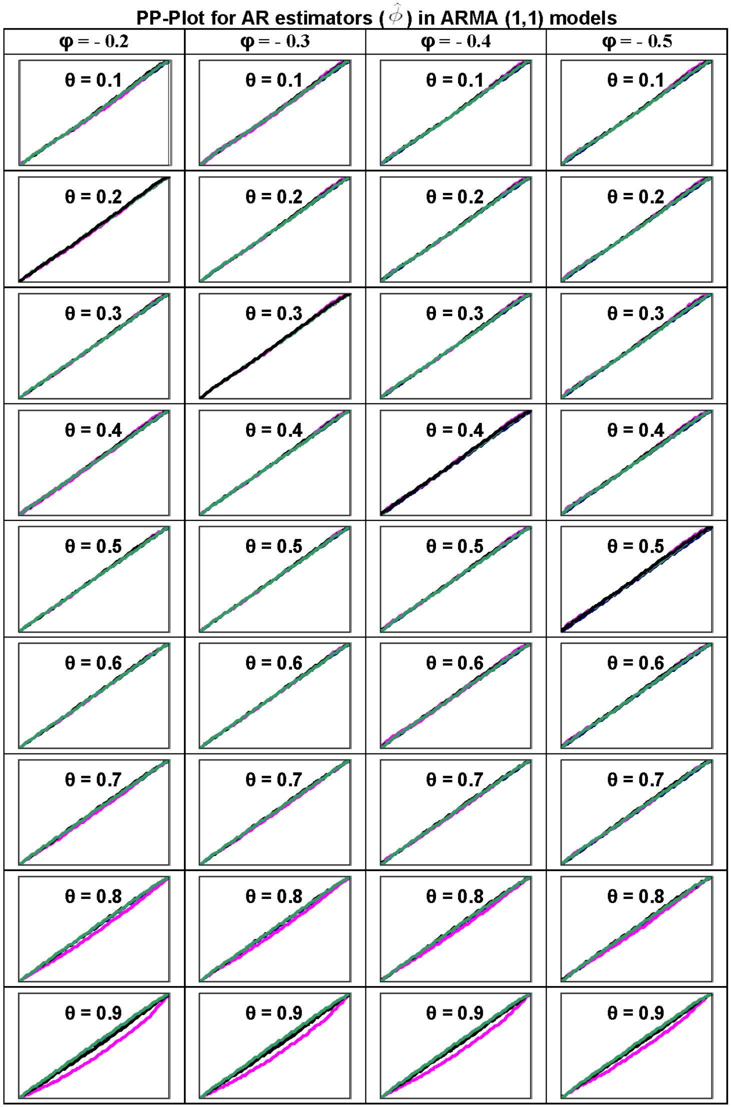
AR ar=0.6 ma=-0.9	cum % simulation n=150	cum % normal
-0.8	0.02	0.00
-0.7	5.30	0.00
-0.6	51.70	0.00
-0.5	92.52	0.00
-0.4	99.53	0.00
-0.3	99.98	0.00
-0.2	100.00	0.00



Simulated CDF (AR=0.6 MA=-0.9 n=200 simulation=10000) for AR: ——

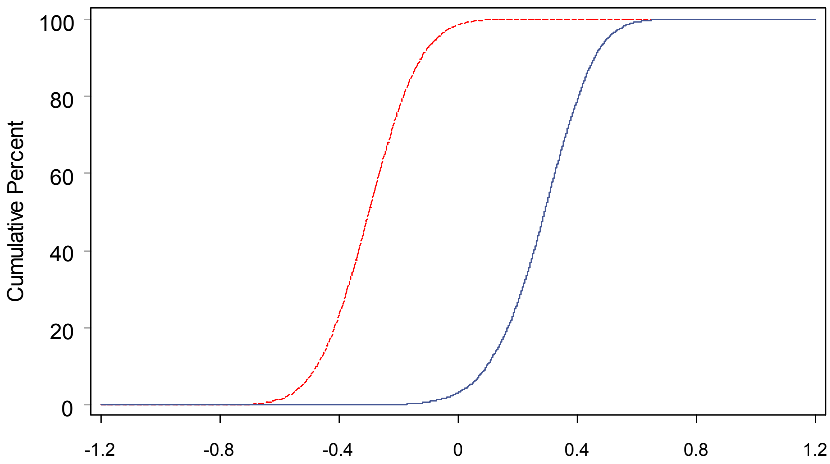
Normal Curve: - - - - - - Mu=0.6, Sigma=0.0581

AR ar=0.6 ma=-0.9	cum % simulation n=200	cum % normal
-0.7	3.01	0.00
-0.6	49.25	0.00
-0.5	94.48	0.00
-0.4	99.88	0.00
-0.3	100.00	0.00



PP-Plot for AR estimators ($\hat{\phi}$) in ARMA (1,1) models
φ = - 0.2
φ = - 0.3
φ = - 0.4
φ = - 0.5
θ = 0.1
θ = 0.2
θ = 0.3
θ = 0.4
θ = 0.5
θ = 0.6
θ = 0.7
θ = 0.8
θ = 0.9

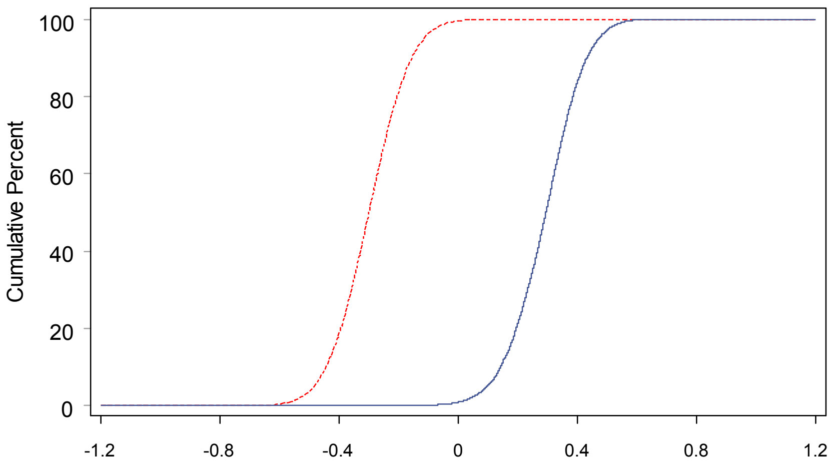
CDF of AR Estimator: Simulated Exact vs. Asymptotic



Simulated CDF (AR=-0.3 MA=0.5 n=100 simulation=10000) for AR: ——

Normal Curve: -------- Mu=-0.3, Sigma=0.1371

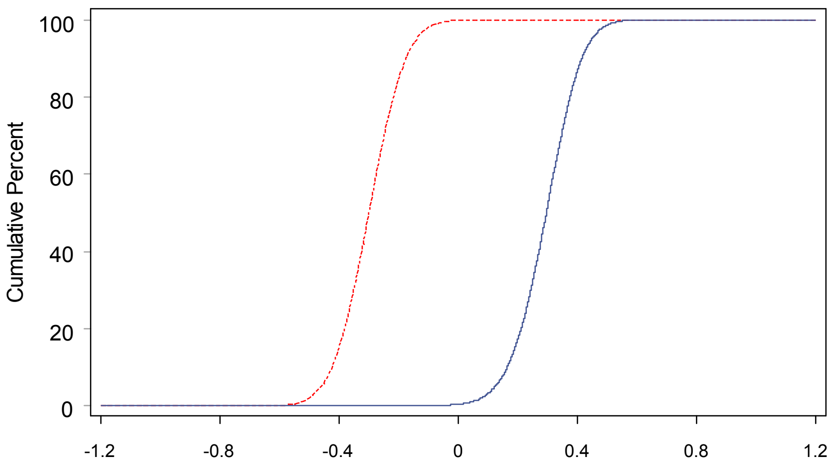
AR ar=-0.3 ma=0.5	cum % simulation n=100	cum % normal
-0.3	0.02	50.00
-0.2	0.13	76.71
-0.1	0.82	92.76
-0.0	3.18	98.57
0.1	10.62	99.82
0.2	26.88	99.99
0.3	53.00	100.00
0.4	79.09	100.00
0.5	94.87	100.00
0.6	99.35	100.00
0.7	99.96	100.00
0.8	100.00	100.00



Simulated CDF (AR=-0.3 MA=0.5 n=150 simulation=10000) for AR: ——

Normal Curve: -------- Mu=-0.3, Sigma=0.112

AR ar=-0.3 ma=0.5	cum % simulation n=150	cum % normal
-0.3	0.01	50.00
-0.1	0.12	96.30
-0.0	0.88	99.63
0.1	5.17	99.98
0.2	21.48	100.00
0.3	52.18	100.00
0.4	84.01	100.00
0.5	97.31	100.00
0.6	99.84	100.00
0.7	100.00	100.00

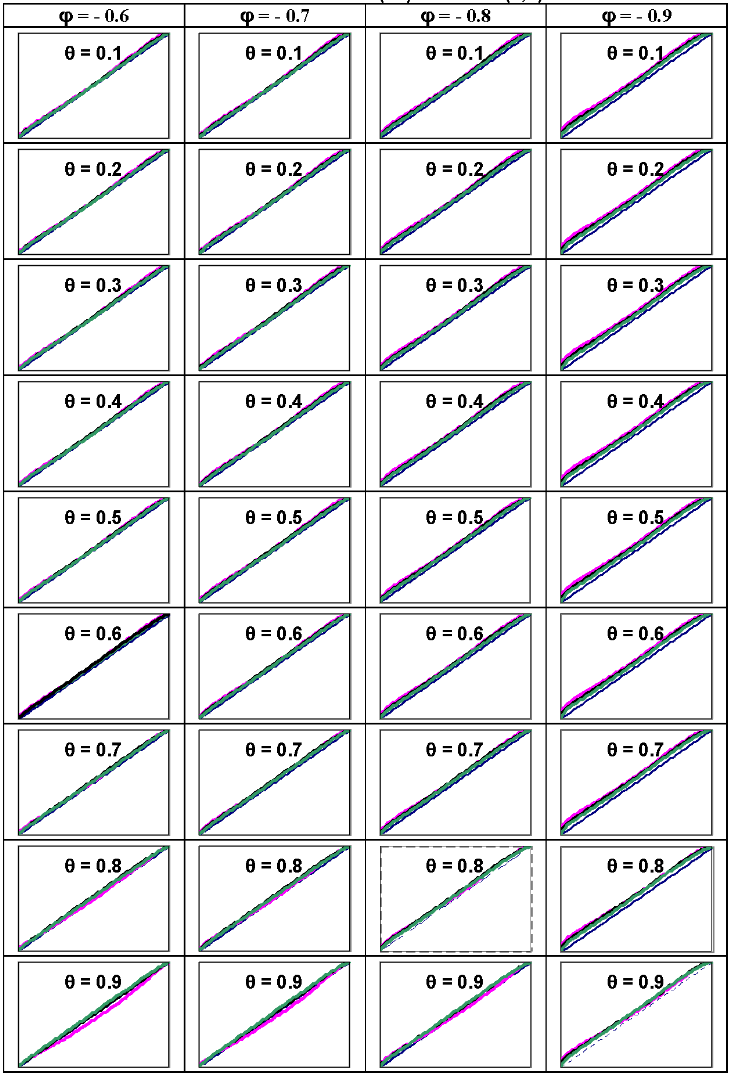


Simulated CDF (AR=-0.3 MA=0.5 n=200 simulation=10000) for AR: ——

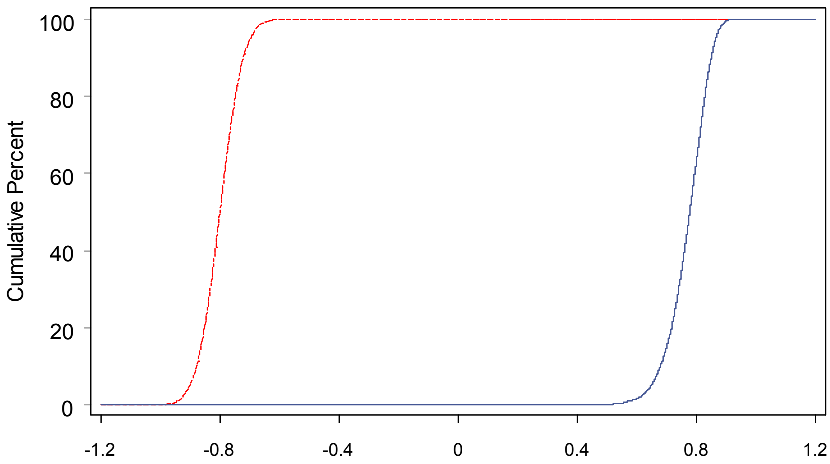
Normal Curve: -------- Mu=-0.3, Sigma=0.097

AR ar=-0.3 ma=0.5	cum % simulation n=200	cum % normal
-0.2	0.01	84.88
-0.1	0.03	98.04
-0.0	0.36	99.90
0.1	3.27	100.00
0.2	17.33	100.00
0.3	52.00	100.00
0.4	87.10	100.00
0.5	98.70	100.00
0.6	99.98	100.00
0.7	100.00	100.00

PP-Plot for AR estimators ($\hat{\phi}$) in ARMA (1,1) models



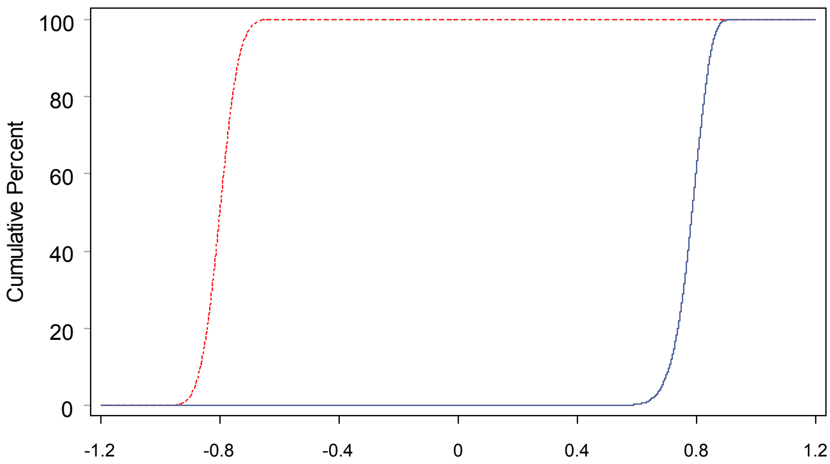
CDF of AR Estimator: Simulated Exact vs. Asymptotic



Simulated CDF (AR=-0.8 MA=0.8 n=100 simulation=10000) for AR: ——

Normal Curve: - - - - Mu=-0.8, Sigma=0.0615

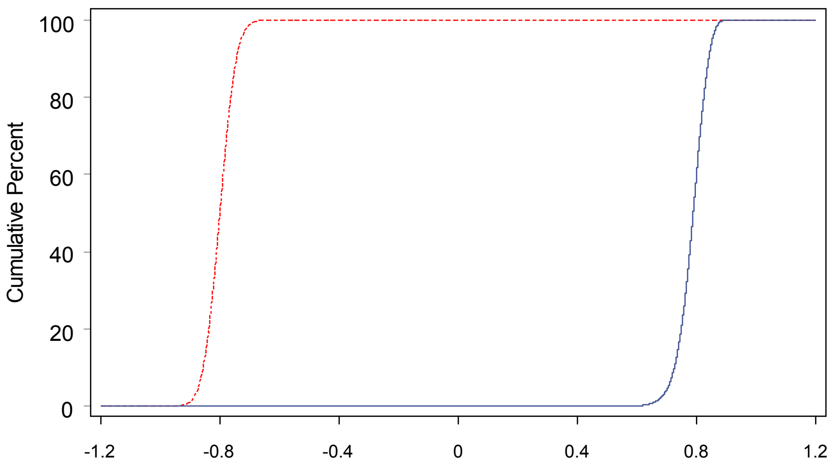
AR ar=-0.8 ma=0.8	cum % simulation n=100	cum % normal
0.5	0.13	100.00
0.6	1.74	100.00
0.7	15.45	100.00
0.8	63.79	100.00
0.9	99.52	100.00
1.0	100.00	100.00



Simulated CDF (AR=-0.8 MA=0.8 n=150 simulation=10000) for AR: ——

Normal Curve: - - - - Mu=-0.8, Sigma=0.0502

AR ar=-0.8 ma=0.8	cum % simulation n=150	cum % normal
0.5	0.02	100.00
0.6	0.32	100.00
0.7	8.33	100.00
0.8	62.60	100.00
0.9	99.87	100.00
1.0	100.00	100.00

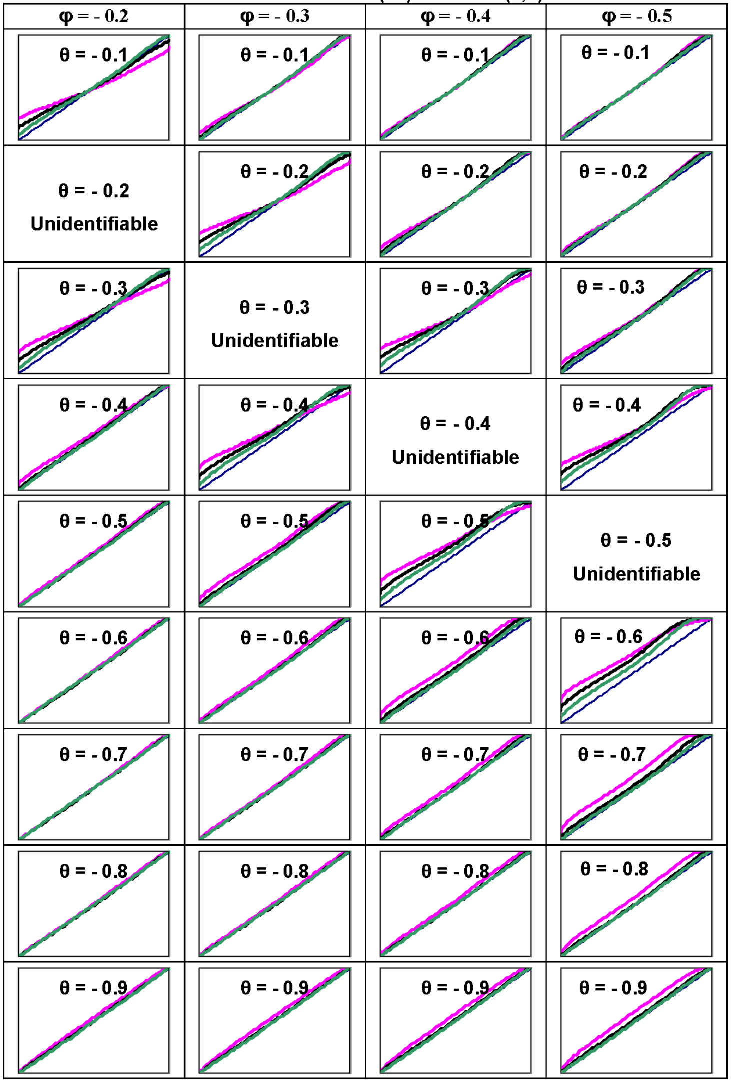


Simulated CDF (AR=-0.8 MA=0.8 n=200 simulation=10000) for AR: ——

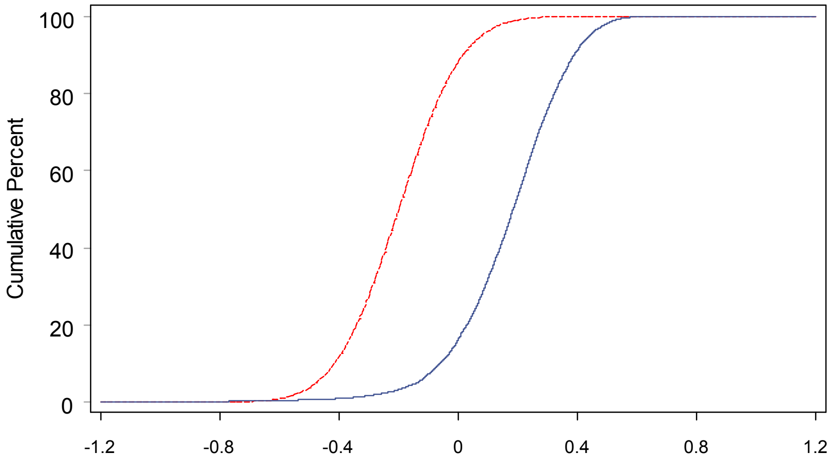
Normal Curve: - - - - Mu=-0.8, Sigma=0.0435

AR ar=-0.8 ma=0.8	cum % simulation n=200	cum % normal
0.5	0.01	100.00
0.6	0.09	100.00
0.7	4.44	100.00
0.8	61.18	100.00
0.9	99.98	100.00
1.0	100.00	100.00

PP-Plot for AR estimators ($\hat{\phi}$) in ARMA (1,1) models



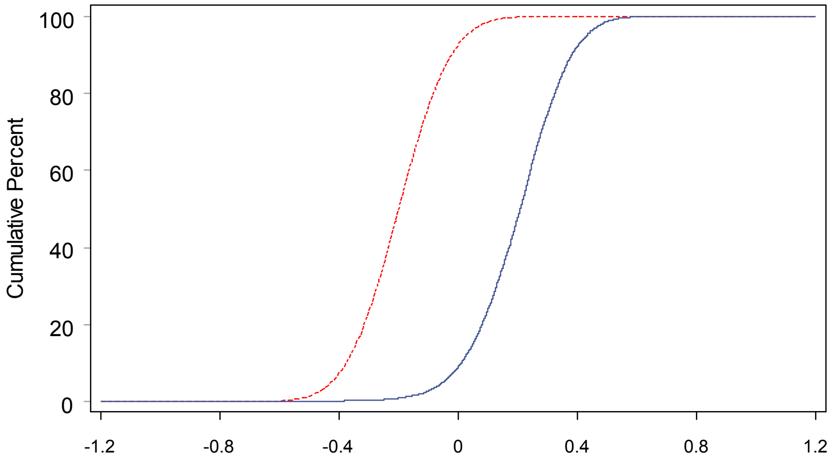
CDF of AR Estimator: Simulated Exact vs. Asymptotic



Simulated CDF (AR=-0.2 MA=-0.7 n=100 simulation=10000) for AR: ——

Normal Curve: -------- Mu=-0.2, Sigma=0.1685

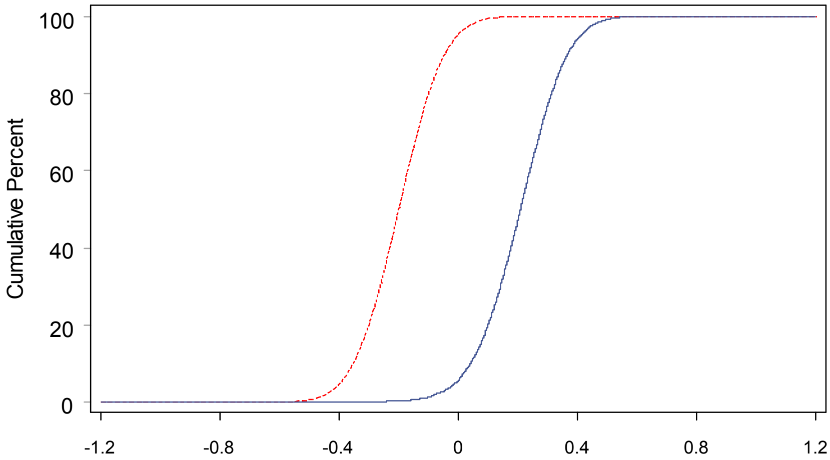
AR ar=-0.2 ma=-0.7	cum % simulation n=100	cum % normal
-0.9	0.12	0.00
-0.8	0.17	0.02
-0.7	0.32	0.15
-0.6	0.42	0.88
-0.5	0.64	3.75
-0.4	0.98	11.77
-0.3	1.68	27.65
-0.2	3.24	50.00
-0.1	7.31	72.35
-0.0	16.20	88.23
0.1	32.04	96.25
0.2	53.70	99.12
0.3	75.95	99.85
0.4	91.31	99.98
0.5	98.26	100.00
0.6	99.90	100.00
0.7	100.00	100.00



Simulated CDF (AR=-0.2 MA=-0.7 n=150 simulation=10000) for AR: ——

Normal Curve: -------- Mu=-0.2, Sigma=0.1376

AR ar=-0.2 ma=-0.7	cum % simulation n=150	cum % normal
-0.9	0.02	0.00
-0.8	0.04	0.00
-0.7	0.05	0.01
-0.6	0.09	0.18
-0.5	0.12	1.46
-0.4	0.16	7.30
-0.3	0.36	23.37
-0.2	0.92	50.00
-0.1	2.79	76.63
-0.0	9.02	92.70
0.1	23.98	98.54
0.2	47.88	99.82
0.3	74.76	99.99
0.4	92.32	100.00
0.5	98.69	100.00
0.6	99.87	100.00
0.7	100.00	100.00

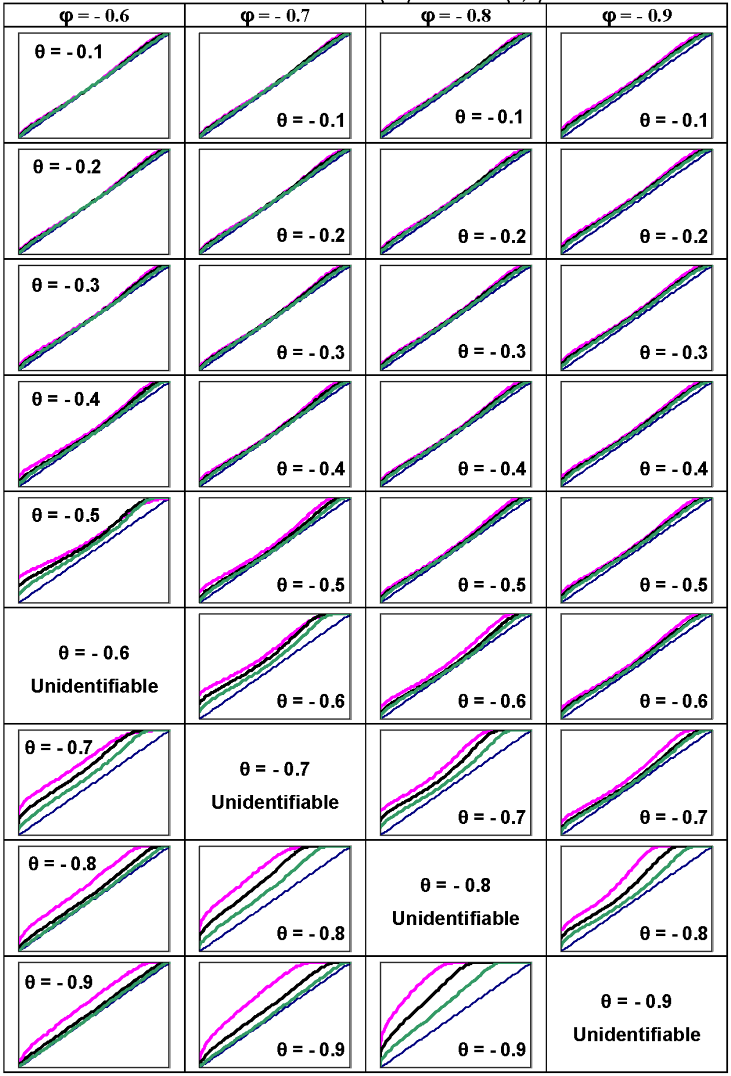


Simulated CDF (AR=-0.2 MA=-0.7 n=200 simulation=10000) for AR: ——

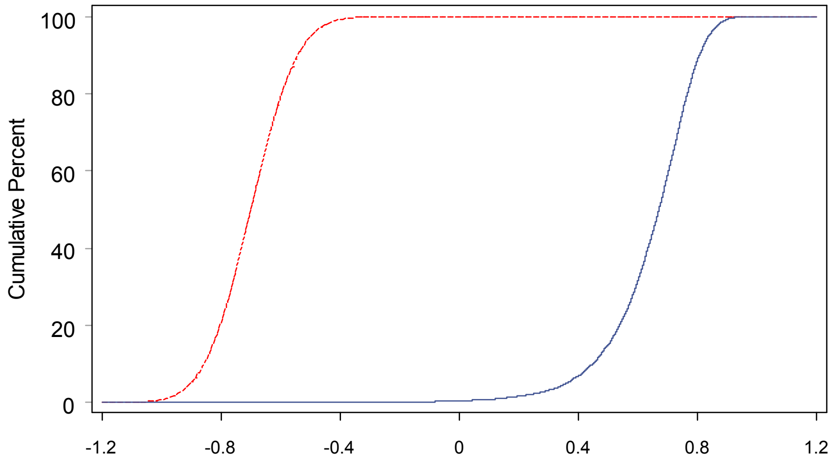
Normal Curve: -------- Mu=-0.2, Sigma=0.1192

AR ar=-0.2 ma=-0.7	cum % simulation n=200	cum % normal
-0.9	0.02	0.00
-0.8	0.03	0.00
-0.7	0.05	0.00
-0.6	0.06	0.04
-0.4	0.07	4.66
-0.3	0.14	20.07
-0.2	0.35	50.00
-0.1	1.31	79.93
-0.0	5.65	95.34
0.1	20.03	99.41
0.2	47.36	99.96
0.3	77.07	100.00
0.4	94.17	100.00
0.5	99.27	100.00
0.6	99.97	100.00
0.7	100.00	100.00

PP-Plot for AR estimators ($\hat{\phi}$) in ARMA (1,1) models



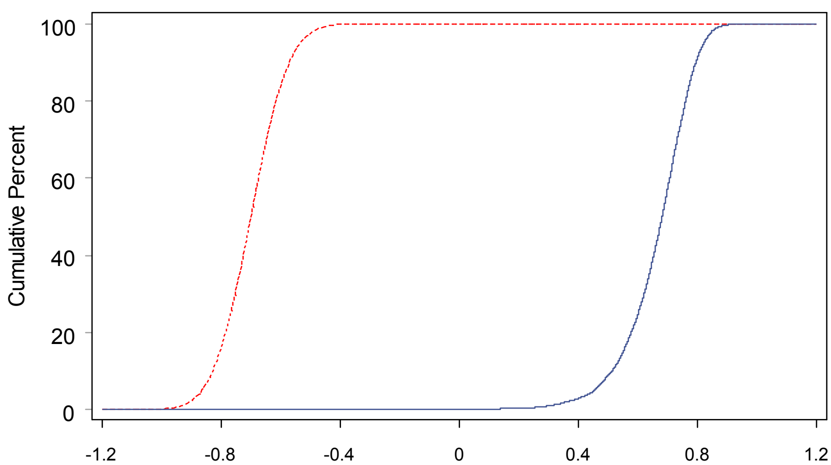
CDF of AR Estimator: Simulated Exact vs. Asymptotic



Simulated CDF (AR=-0.7 MA=-0.2 n=100 simulation=10000) for AR: ——

Normal Curve: - - - - - Mu=-0.7, Sigma=0.1228

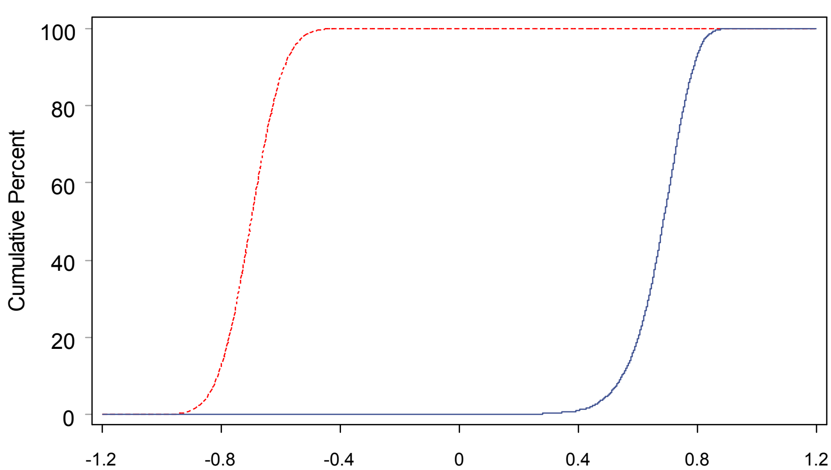
AR ar=-0.7 ma=-0.2	cum % simulation n=100	cum % normal
-0.4	0.02	99.27
-0.3	0.07	99.94
-0.2	0.10	100.00
-0.1	0.21	100.00
-0.0	0.39	100.00
0.1	0.79	100.00
0.2	1.57	100.00
0.3	3.16	100.00
0.4	6.89	100.00
0.5	15.02	100.00
0.6	31.84	100.00
0.7	59.40	100.00
0.8	88.97	100.00
0.9	99.33	100.00
1.0	100.00	100.00



Simulated CDF (AR=-0.7 MA=-0.2 n=150 simulation=10000) for AR: ——

Normal Curve: - - - - - Mu=-0.7, Sigma=0.1003

AR ar=-0.7 ma=-0.2	cum % simulation n=150	cum % normal
-0.1	0.03	100.00
-0.0	0.06	100.00
0.1	0.18	100.00
0.2	0.28	100.00
0.3	0.98	100.00
0.4	2.87	100.00
0.5	9.02	100.00
0.6	25.00	100.00
0.7	57.75	100.00
0.8	91.60	100.00
0.9	99.75	100.00
1.0	100.00	100.00



Simulated CDF (AR=-0.7 MA=-0.2 n=200 simulation=10000) for AR: ——

Normal Curve: - - - - - Mu=-0.7, Sigma=0.0869

AR ar=-0.7 ma=-0.2	cum % simulation n=200	cum % normal
0.1	0.02	100.00
0.2	0.07	100.00
0.3	0.32	100.00
0.4	1.04	100.00
0.5	5.08	100.00
0.6	19.82	100.00
0.7	56.64	100.00
0.8	93.25	100.00
0.9	99.97	100.00
1.0	100.00	100.00

You may wonder what about a general ARMA(p, q) model:

$$(x_t - \mu) - \phi_1(x_{t-1} - \mu) - \cdots - \phi_p(x_{t-p} - \mu) = z_t - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q}$$

with a total of $r = 2 + p + q$ parameters, $(\mu, \sigma^2, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$.

And for those more complex models that have additional variance structure like GARCH(r, s) for z_t:

$$z_t = h^{1/2} e_t, \quad e_t \sim iidN0, 1), \quad h_t = \omega + \sum_{i=1}^{s} \alpha_i z_{t-1}^2 + \sum_{j=1}^{r} \gamma_j h_{t-j}$$

and further variance structures like EGARCH, IGARCH, HIGARCH ....

Caution (a): [Last page]
General Suggestions

Neighborhood Validation by Simulation: With the estimates at hand after standard analysis, do simulations for parameter values in a neighborhood around the estimate.

Compare asymptotically equivalent approaches: Do the 3 tests, LR, Wald and Score concur? Do the Wald C.I. concur with the profile likelihood intervals?

Try transformation: Fisher, the founder of likelihood asymptotics of the first order, did not apply asymptotic normality directly to r_n, but transformed it to $log(1+r)-log(1-r)$ before using asymptotic normality.

Caveat (b):
<u>A Statistic's True Nature Is Vital</u>

- The stereo-type and immediate application of large-sample theory directly lands on normal inference *without nuanced thinking*.
- Such practice *might miss the true nature* of the statistic in use.
- Consider the case of testing the Random-walk Model that underpins the Efficient Market Hypothesis.

Random-walk Model

- P_t: Price at time t, $\quad t = 0, \ldots, n$.
- $R_t = \log(P_t) - \log(P_{t-1})$: 1-period return
- $R_t^{(q)} = \log(P_t) - \log(P_{t-q})$: q-period return
- Null hypothesis of iid Gaussian random-walk

$$H_0 : R_t = \mu + \varepsilon_t,\ \varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2),\ t = 1, \ldots, n$$

- Under H_0, $\mathrm{Var}(R_t^{(q)}) = q\mathrm{Var}(R_t)$
- So test $\mathrm{VR}(q) = \mathrm{Var}(R_t^{(q)})/(q\mathrm{Var}(R_t)) = 1$

Large-Sample Tests

Lo & MacKinlay (1988) showed that **under** H_0**:**

- Non-overlapping VR: **When** $n \to \infty$**,**

$$\sqrt{n}\,(\widehat{\mathbf{VR}}(q) - 1) \overset{\cdot}{\sim} N(0, 2(q-1))$$

- Overlapping VR: **When** $n \to \infty$**,**

$$\sqrt{n}(\widehat{VR}(q) - 1) \overset{\cdot}{\sim} N\left(0, \frac{2(q-1)(q-1)}{3q}\right)$$

VR as an optimal test:

Econometrica, Vol. 60, No. 5 (September, 1992), 1215–1226

WHEN ARE VARIANCE RATIO TESTS FOR SERIAL DEPENDENCE OPTIMAL?

BY JON FAUST[1]

This paper considers a class of statistics that can be written as the ratio of the sample variance of a filtered time series to the sample variance of the original series. Any such statistic is shown to be optimal under normality for testing a null of white noise against some class of serially dependent alternatives. A simple characterization of the class of alternative models is provided in terms of the filter upon which the statistic is based. These results are applied to demonstrate that a variance ratio test for mean reversion is an optimal test for mean reversion and to illustrate the forms of mean reversion it is best at detecting.

KEYWORDS: Variance ratio statistic, serial dependence, optimal test.

Doubts on Asymptotics by Simulations

- Cecchetti and Lam (1994, J. Bus. & Eco. Stat.) found that **"VR tests based on asymptotic approximations are often misleading"** and thus unreliable in finite samples.
- Their Monte Carlo experiments showed that **"there is substantial size distortion in searching over many horizons to decide whether to reject a model."**

As ratio of quadratic forms (Faust, 1992)

It is useful to write the statistic as a ratio of quadratic forms; to do so, write the vector R^ϕ as ΦR where Φ $((T-m)\times T)$ is given by

$$\Phi = \begin{bmatrix} \phi_m & \cdots & \phi_0 & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & \phi_m & \cdots & \phi_0 \end{bmatrix}.$$

Next, in order to simplify the proofs, pick a slight modification of the standard sample variance for the numerator of the statistic. Defining $\bar{R}$ as the sample mean of R, estimate the mean of R^ϕ by $\phi(1)\bar{R}$.[2] This allows the sample variance of R^ϕ to be written $(R-\bar{R}i)'\Phi'\Phi(R-\bar{R}i)/T$ (where i is a conformable vector of ones), suggesting the variance ratio statistic

$$(1) \qquad \nu_\phi = \frac{(R-\bar{R}i)'\Phi'\Phi(R-\bar{R}i)}{(R-\bar{R}i)'(R-\bar{R}i)}.$$

The Durbin-Watson (DW) statistic provides the simplest example of an FVR statistic. This statistic is defined as $\Sigma_{t=2}^T(\hat{\varepsilon}_t-\hat{\varepsilon}_{t-1})^2/\Sigma_{t=1}^T\hat{\varepsilon}_t^2$, where the $\hat{\varepsilon}_t$ are residuals from some regression. If the regression has a constant, then the sample mean of $\hat{\varepsilon}$ is zero, and the DW statistic is an FVR statistic based on the filter $\phi(L)=(1-1L)$.

Non-overlapping VR & One-way ANONVA

$$\Phi_{k\times kq} = \begin{bmatrix} 1\cdots 1 & 0\cdots 0 & 0\cdots & 0\cdots 0 \\ 0\cdots 0 & 1\cdots 1 & 0\cdots & 0\cdots 0 \\ \vdots & & & \\ 0\cdots\cdot & \cdots\cdots\cdot & \cdots 0 & 1\cdots 1 \end{bmatrix}$$

Under H_0 in one-way ANOVA of k groups, each of q observations, F=BMS/WMS $\sim$ $F(k-1, k(q-1))$, hence v_ϕ=BSS/TSS $\sim$ Beta$((k-1)/2, k(q-1)/2)$.

Peking U technical report in 1998 by G. Tian and Y. Zhang, "A Note on the Exact Distributions of Variance Ratio Statistics" based on an elaborated proof.

Overlapping VR

Large

$$\Phi_{(n-q+1)\times n} = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 1 \end{bmatrix}$$

- 2004: Y.K. Tse, K.W. Ng and X.B. Zhang, "A Small Overlapping Variance-Ratio Test," JTSA, 25(1), 127-135. Beta approximations.
- 2004, 2008: Raymond Kan "Exact Variance Ratio Tests with Overlapping Data," manuscript.

Fang, Kotz and Ng (1990) "Symmetric Multivariate and Related Distributions"

2.7 Robust statistics and regression model

2.7.1 *Robust statistics*

Let $\mathbf{x} = (X_1, \ldots, X_n)'$ be an exchangeable random vector which can be viewed as a sample from a population with the distribution of X_1. In the case when $\mathbf{x}$ is normally distributed several statistics useful for inferential purposes, such as t- and F-statistics, are available. The following theorem shows that these statistics, being invariant under scalar multiplication, are robust in the class of spherical distributions.

Theorem 2.22

The distribution of a statistic $t(\mathbf{x})$ remains unchanged as long as $\mathbf{x} \sim S_n^+(\phi)$ (cf. Section 2.1) provided that

$$t(a\mathbf{x}) \stackrel{d}{=} t(\mathbf{x}) \qquad \text{for each } a > 0. \tag{2.57}$$

In this case $t(\mathbf{x}) \stackrel{d}{=} t(\mathbf{y})$ where $\mathbf{y} \sim N_n(\mathbf{0}, \mathbf{I}_n)$.

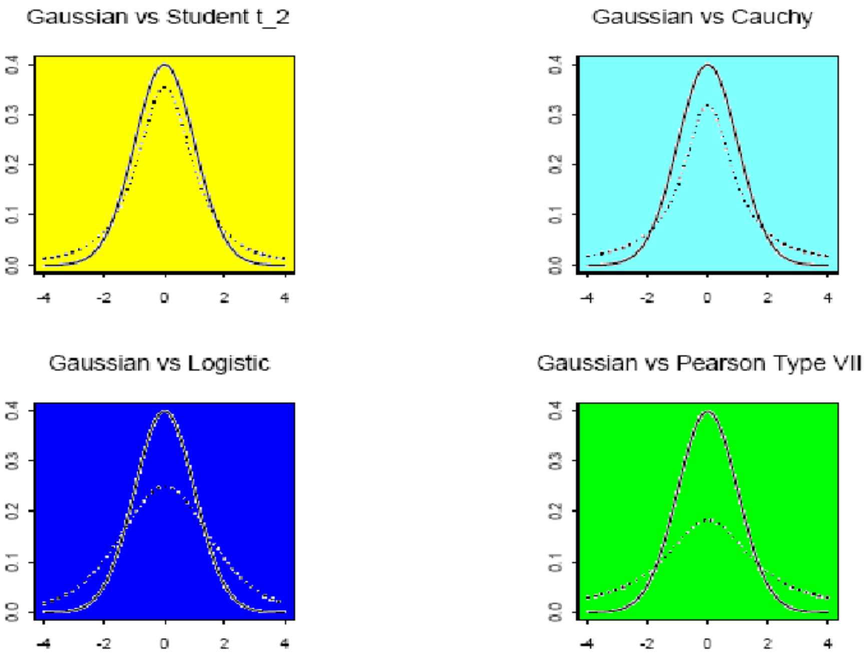
The exact distribution of $\widehat{\mathrm{VR}}(q)$, overlapping or not, under H_0 that $(\varepsilon_1, \ldots, \varepsilon_n)$ are iid standard Gaussian, **remains the same under a bigger null** $H_0^* : R_t = \mu + \varepsilon_t, \ t = 1, \ldots, n$, where $(\varepsilon_1, \ldots, \varepsilon_n)$ has a **multivariate spherical distribution** $S_n(\mu\mathbf{1}, \sigma^2\mathbf{I})$ with a common location parameter μ and a common scale parameter σ.

Spherical: Uncorrelated, Dependent, and Identically Distributed (u.d.i.d) Except the Gaussian which is i.i.d.

Table 3.1 *Some subclasses of n-dimensional spherical distributions*

Type	Density function $f(\mathbf{x})$ or c.f. $\psi(\mathbf{t})$
Kotz type	$f(\mathbf{x}) = c(\mathbf{x}'\mathbf{x})^{N-1}\exp(-r(\mathbf{x}'\mathbf{x})^s)$, $r, s > 0$, $2N + n > 2$
Multinormal	$f(\mathbf{x}) = c\exp(-\frac{1}{2}\mathbf{x}'\mathbf{x})$
Pearson Type VII	$f(\mathbf{x}) = c(1 + \mathbf{x}'\mathbf{x}/s)^{-N}$, $N > n/2$, $s > 0$
Multivariate t	$f(\mathbf{x}) = c(1 + \mathbf{x}'\mathbf{x}/s)^{-(n+m)/2}$, $m > 0$ an integer
Multivariate Cauchy	$f(\mathbf{x}) = c(1 + \mathbf{x}'\mathbf{x}/s)^{-(n+1)/2}$, $s > 0$
Pearson Type II	$f(\mathbf{x}) = c(1 - \mathbf{x}'\mathbf{x})^m$, $m > 0$
Logistic	$f(\mathbf{x}) = c\exp(-\mathbf{x}'\mathbf{x})/[1 + \exp(-\mathbf{x}'\mathbf{x})]^2$
Multivariate Bessel	$f(\mathbf{x}) = c(\lVert\mathbf{x}\rVert/\beta)^a K_a(\lVert\mathbf{x}\rVert/\beta)$, $a > -n/2$, $\beta > 0$, where $K_a(\cdot)$ denotes the modified Bessel function of the third kind
Scale mixture	$f(\mathbf{x}) = c\int_0^\infty t^{-n/2}\exp(-\mathbf{x}'\mathbf{x}/2t)\mathrm{d}G(t)$, $G(t)$ a c.d.f.
Stable laws	$\psi(\mathbf{t}) = \exp\{r(\mathbf{t}'\mathbf{t})^{\alpha/2}\}$, $0 < \alpha \leqslant 2$, $r < 0$
Multiuniform	$\psi(\mathbf{t}) = {}_0F_1(n/2; -\frac{1}{4}\lVert\mathbf{t}\rVert^2)$, ${}_0F_1(\cdot;\cdot)$ is a generalized hypergeometric function

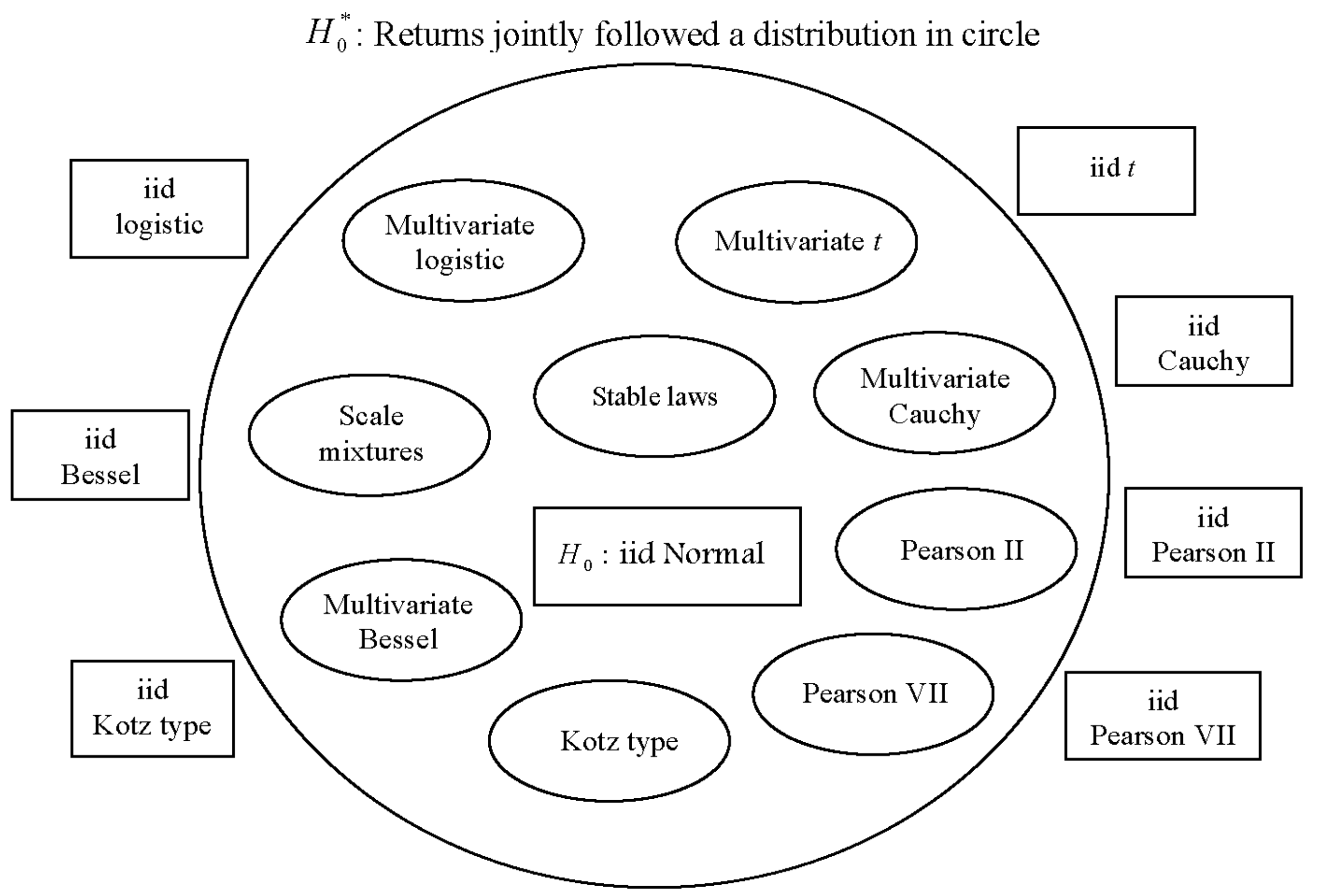
Spherical Distributions Include Heavy-Tailed Marginal pdf



Statisticians normally LONG for robustness, right?

In this case then, is it really GOOD news?

Is VR Testing What You Intended?



What If VR Test Does Not Reject?

Relax ..., some good news.

The u.d.i.d Case with Spherical Dist.

- n past returns: $R_1, \ldots, R_n$
- m future returns: $Y_1, \ldots, Y_m$
- $(R_1, \ldots, R_n, Y_1, \ldots, Y_m)$ has a spherical dist. $S_{n+m}(\mu\mathbf{1}, \sigma^2\mathbf{I})$, be it $N_{n+m}(\mu\mathbf{1}, \sigma^2\mathbf{I})$, multivariate t, multivariate Cauchy, multivariate logistic, or multivariate stable distributions, ...
- **Inference** on any function of $(Y_1, \ldots, Y_m)$ is based on the conditional dist. on $(R_1, \ldots, R_n)$: available joint work with Guo-liang Tian

Weighted Sum

- A $100\gamma\%$ prediction interval for a weighted sum, $\sum_{j=1}^{m} w_j Y_j$, $w_j > 0$ and $\sum_{j=1}^{m} w_j = 1$, is:

$$\bar{R} \pm t\left(n-1; \frac{1+\gamma}{2}\right) S_R \sqrt{\sum_{j=1}^{m} w_j^2 + 1/n}$$

where

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i, \qquad S_R^2 = \frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2.$$

Only Next Day

- The $100\gamma\%$ forecast interval for tomorrow's return, Y_1 , is:

$$\bar{R} \pm t(n-1;(1+\gamma)/2) S_R \sqrt{1+1/n}$$

- A $100\gamma\%$ forecast interval for tomorrow's price, P_{n+1}, is:

$$P_{n+1} \times$$
$$\exp\left\{\bar{R} \pm t\left(n-1;\frac{1+\gamma}{2}\right) S_R\sqrt{1+1/n}\right\}$$

Average

- **Let** $w_j = 1/m$**, then** $\sum_{j=1}^{m} w_j^2 = 1/m$**.** Hence, a two-sided prediction interval for $\bar{Y}$ with γ confidence level is

$$\bar{R} \pm t(n-1;(1+\gamma)/2) S_R \sqrt{1/m + 1/n}$$

Lower Limit of rth Smallest

- Let $T(.|\nu, \delta)$ denote the cdf of noncentral t with DF ν, and

$$f_r(z) = r \binom{m}{r} [\Phi(z)]^{r-1}[1 - \Phi(z)]^{m-r}\phi(z)$$

A $100\gamma\%$ lower limit for $Y_{r,m}$ $(1 \le r \le m)$ is

$$\bar{R} - t(n, m, r; \gamma) S_R/\sqrt{n},$$

where $t(n, m, r; \gamma)$ is a solution of t in

$$\gamma = \int_{-\infty}^{+\infty} T(t|n - 1, -\sqrt{n}\, z) \cdot f_r(z)\, dz,$$

A threshold for at least s of the m

- The probability that at least s of the m future observations $\{Y_1, \ldots, Y_m\}$ exceed the pre-specified return value L_0 is

$$\gamma = \int_{-\infty}^{+\infty} T\Big(\sqrt{n}(\bar{R} - L_0)/S_R \Big| n-1, -\sqrt{n}\, z\Big)$$

$$\times f_{m-s+1}(z)\, dz$$

where $T(.|\nu, \delta)$ denotes the cdf of t distribution with DF ν and noncentrality δ

Thank you for listening!