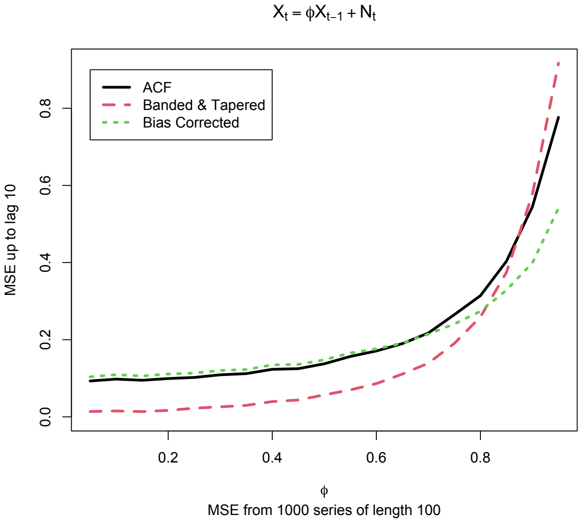
School of Mathematical Sciences | Lancaster University

# Why are we all using biased estimates of autocorrelation? and Why are most time series simulation studies poor?

Rebecca Killick
Joint work with Colin Gallagher, Xiyan Tan, and Xinyi Li (Clemson)
Clemson Nov 2025

# Aim



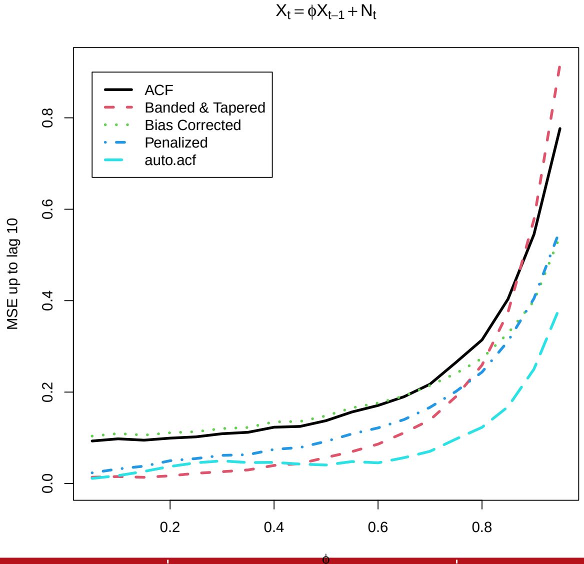

$X_t = \phi X_{t-1} + N_t$
ACF
Banded & Tapered
Bias Corrected
MSE up to lag 10
0.0
0.2
0.4
0.6
0.8
0.2
0.4
0.6
0.8
$\phi$
MSE from 1000 series of length 100

# Aim

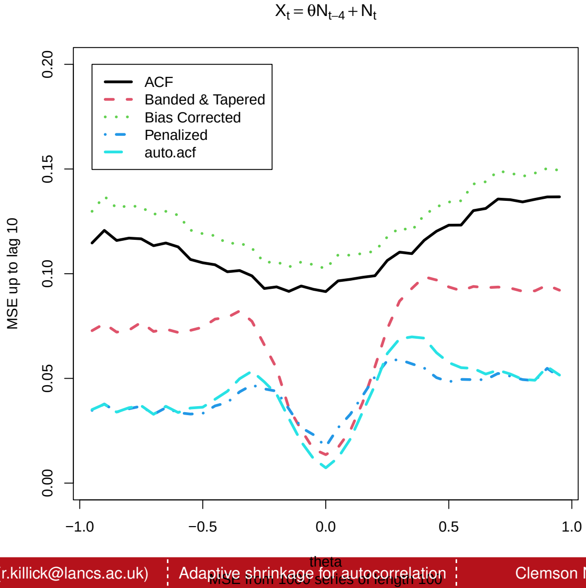
- Provide better estimates of the (p)acf by adaptive shrinkage
- Give more intuition on causality
- Motivate better time series simulation studies.

# Results




$X_t = \phi X_{t-1} + N_t$
ACF
Banded & Tapered
Bias Corrected
Penalized
auto.acf
MSE up to lag 10
0.0
0.2
0.4
0.6
0.8
0.2
0.4
0.6
0.8
ϕ
MSE from 1000 series of length 100

# Results


$X_t = \theta N_{t-4} + N_t$
ACF
Banded & Tapered
Bias Corrected
Penalized
auto.acf
MSE up to lag 10
0.00
0.05
0.10
0.15
0.20
−1.0
−0.5
0.0
0.5
1.0
theta
MSE from 1000 series of length 100

# Method



For $y_t = x_t - \bar{x}$, sample ACF at lag $h$, $\rho(h)$ minimizes:

$$Q_h(\rho_h) = \sum_{t=h+1}^{n+h} (y_t - \rho_h y_{t-h})^2 ,$$

To shrink toward zero (reduce variance) add a penalty:

$$Q_h(\rho_h) + \lambda_h \left( \sum_{t=1}^{n} y_t^2 \right) (\rho_h)^2$$

To move toward a bias corrected estimator, $\hat{\rho}_b$:

$$Q_h(\rho_h) + \lambda_h \left( \sum_{t=1}^{n} y_t^2 \right) (\rho_h - \hat{\rho}_b)^2$$

Let the data tell us which one to use and how much to shrink.

We minimize

$$Q_h(\rho_h) + \lambda_h \left( \sum_{t=1}^{n} y_t^2 \right) (\rho_h - \rho_\tau(h))^2$$

The solution is

$$\tilde{\rho}(h) = w_h \rho_\tau(h) + (1 - w_h)\hat{\rho}(h),$$

where $w_h = \lambda_h/(1+\lambda_h)$, and $\hat{\rho}(h)$ is the ordinary ACF at lag $h$.\

Note: General framework includes, ordinary ACF, shrinkage toward zero, bias correction, Banded & Tapered, ...

# Implementation



We use plug-in estimators of the tuning parameters.

Shrink toward zero if $\hat{\rho}_b(h)$ is small, move toward bias correction if larger.

$$\rho_\tau(h) = \begin{cases} 0 & |\hat{\rho}_b(h)| < \ell_h \\ \hat{\rho}_b(h) & |\hat{\rho}_b(h)| > \ell_h \end{cases}$$

Select $\lambda_h$ ($w$) to move more aggressively near boundaries, and to smoothly transition between making smaller and increasing magnitude:

$$\lambda_h = \begin{cases} \frac{h(\ell_h - r)}{r^2} & r < \ell_h \\ \frac{10\log_{10}(n)h(r-\ell_h)(1-\ell_h)}{(1-r)^2} & r > \ell_h, \end{cases}$$

where $r = |\hat{\rho}_b(h)|$.

# Asymptotics



$\ell_h$ controls the asymptotic distribution of $\tilde{\rho}(h)$ for $\rho(h) = 0$:

1. If $\sqrt{n}\ell_h \to 0$ as $n \to \infty$, $\hat{\rho}$ and $\tilde{\rho}(h)$ have same limiting distribution.
2. If $\sqrt{n}\ell_h \to C$ as $n \to \infty$, $\sqrt{n}\tilde{\rho}(h)$ has a non-normal limit with smaller variance than $\sqrt{n}\hat{\rho}(h)$.
3. If $\ell_h \to 0$ and $\sqrt{n}\ell_h \to \infty$ as $n \to \infty$, $\sqrt{n}\tilde{\rho}(h) \to 0$ in probability.

$\sqrt{n}(\tilde{\rho}(h) - \rho(h))$ has same limit as $\sqrt{n}(\hat{\rho}(h) - \rho(h))$, when$\rho(h) \neq 0$, \
We use

$$\ell_h = c_h\sqrt{\log(n)/n},$$

where $c_h$ depends on the estimated ACF.

# PACF

Similar development for the PACF at lag $h$, $\pi(h)$ gives:

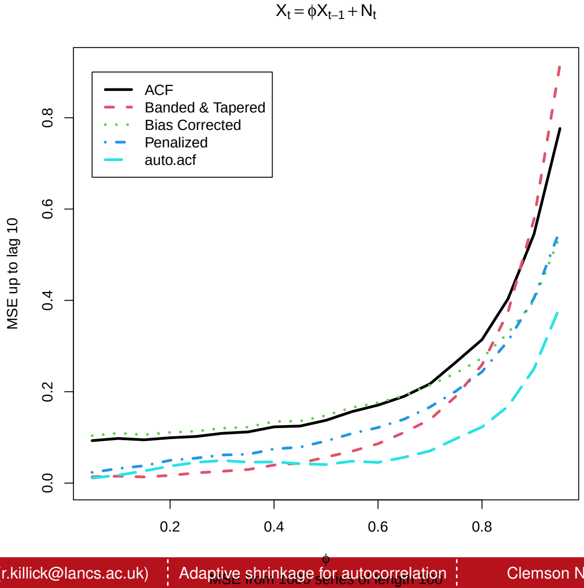
$$\tilde{\pi}(h) = w_h \pi_\tau(h) + (1 - w_h)\hat{\pi}(h),$$

where $w_h = \lambda_h/(1 + \lambda_h)$, and $\hat{\pi}(h)$ is the ordinary PACF at lag $h$.\ Use $\ell_h$, $\hat{\pi}_b(h)$, and $\lambda_h$ as above (but using PACF).

$$\pi_\tau(h) = \left\{ \begin{array}{ll} 0 & |\hat{\pi}_b(h)| < \ell_h \\ \hat{\pi}_b(h) & |\hat{\pi}_b(h)| > \ell_h. \end{array} \right\}$$

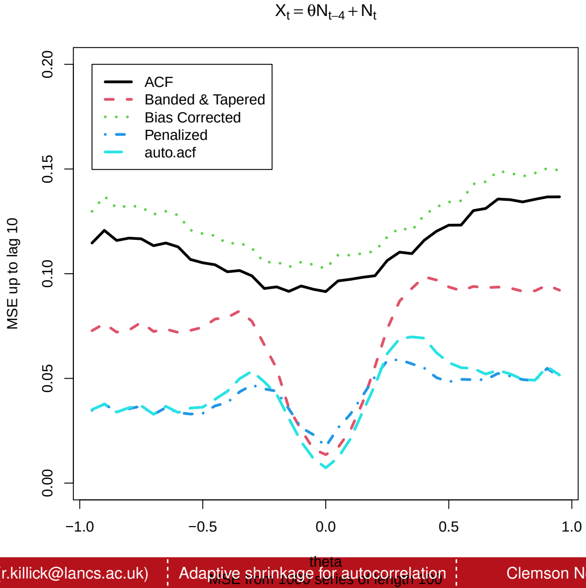
Asymptotic results again controlled by $\ell_h$ (we use our plug-in tuning parameters below).

# Results


$X_t = \phi X_{t-1} + N_t$
ACF
Banded & Tapered
Bias Corrected
Penalized
auto.acf
MSE up to lag 10
0.0
0.2
0.4
0.6
0.8
0.2
0.4
0.6
0.8
$\phi$
MSE from 1000 series of length 100

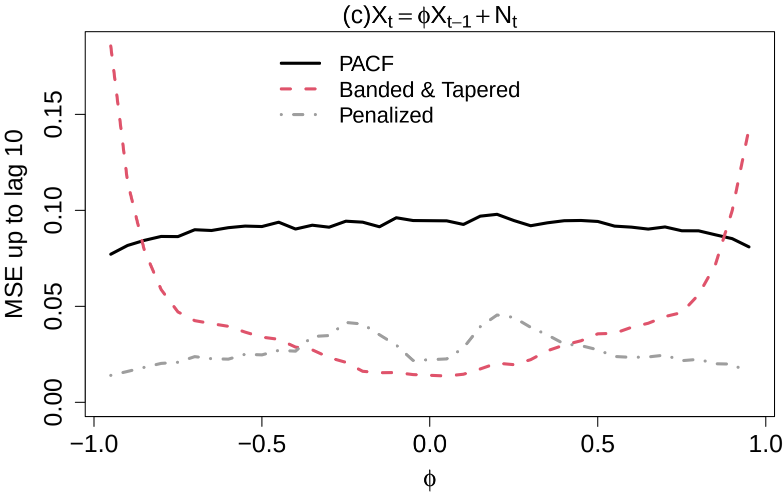
# Results




$X_t = \theta N_{t-4} + N_t$
ACF
Banded & Tapered
Bias Corrected
Penalized
auto.acf
MSE up to lag 10
0.00
0.05
0.10
0.15
0.20
−1.0
−0.5
0.0
0.5
1.0
theta
MSE from 1000 series of length 100

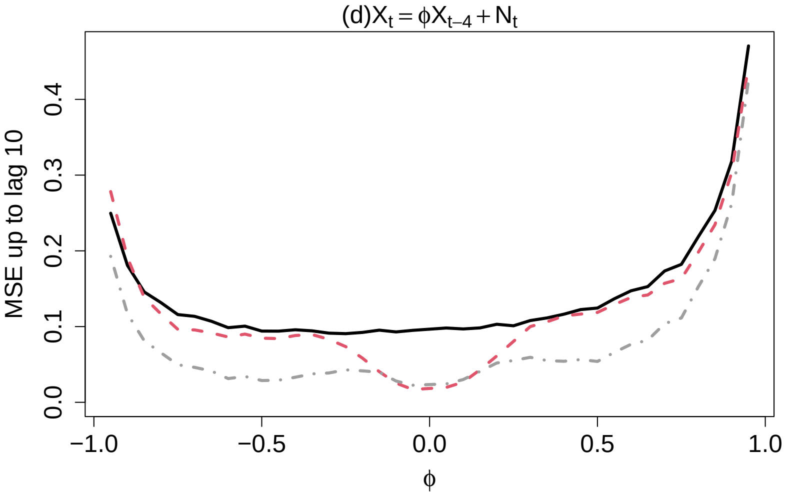
# Results




(c)$X_t = \phi X_{t-1} + N_t$
PACF
Banded & Tapered
Penalized
MSE up to lag 10
0.00
0.05
0.10
0.15
−1.0
−0.5
0.0
0.5
1.0
$\phi$

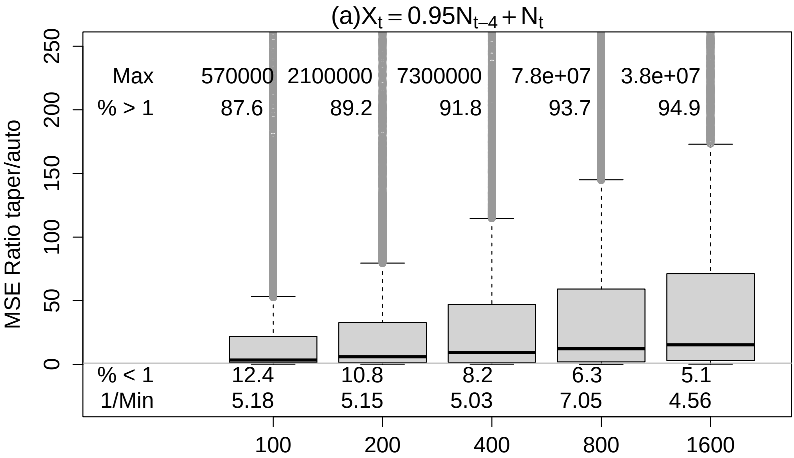
# Results



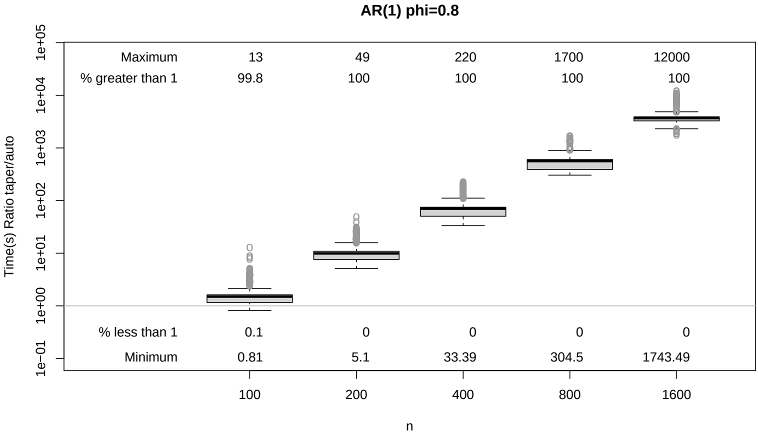


# Grow N




$(a) X_t = 0.95N_{t-4} + N_t$
MSE Ratio taper/auto
0
50
100
150
200
250
Max
570000
2100000
7300000
7.8e+07
3.8e+07
% > 1
87.6
89.2
91.8
93.7
94.9
% < 1
12.4
10.8
8.2
6.3
5.1
1/Min
5.18
5.15
5.03
7.05
4.56
100
200
400
800
1600

# Grow N


AR(1) phi=0.8
Time(s) Ratio taper/auto
1e+05
1e+04
1e+03
1e+02
1e+01
1e+00
1e−01
Maximum 13 49 220 1700 12000
% greater than 1 99.8 100 100 100 100
% less than 1 0.1 0 0 0 0
Minimum 0.81 5.1 33.39 304.5 1743.49
100
200
400
800
1600
n

# Sims

- Want to simulate across the causal space
- For a wide set of AR, MA and ARMA processes

HOW?

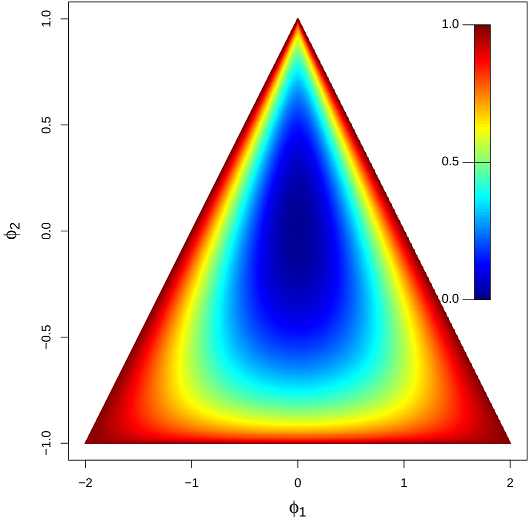
# Sims

- Want to simulate across the causal space
- For a wide set of AR, MA and ARMA processes

HOW?

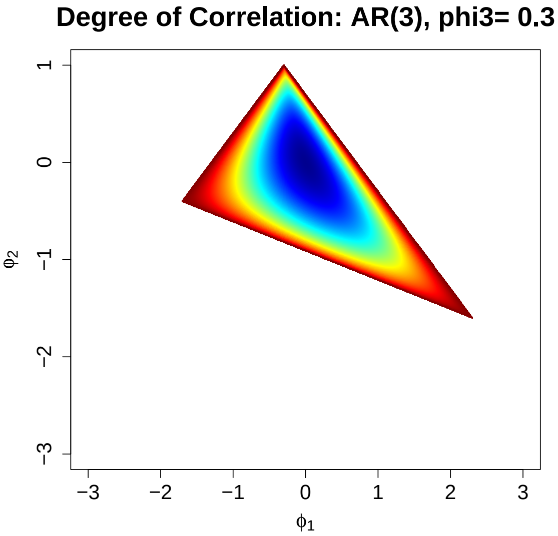
- JTSA published 44 papers in 2024, 39 had an AR simulation
- Only 2 papers selected coefficients randomly (in AR(1))
- Of the remaining 37, 25 used a single AR parameter setting!
- Only 12/39 considered orders larger than 1!

# Solution AR2



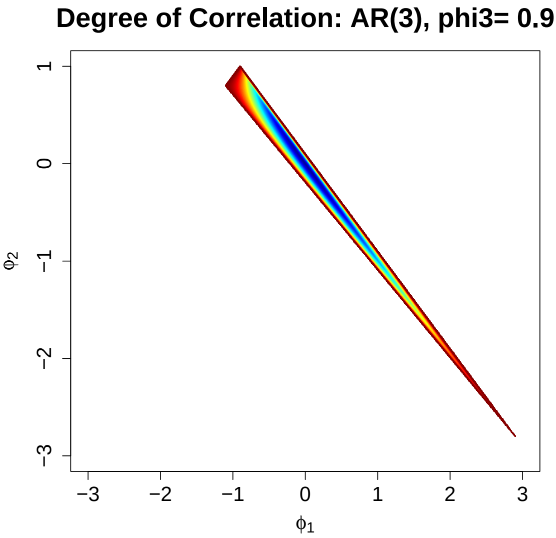
# Solution AR3





# Solution AR3





# Problem - Causality



AR(p):

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \epsilon_t \quad t = 0, \pm 1, \pm 2, \dots,$$

Causal:

$$Y_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j} \quad \sum_{j=0}^{\infty} |c_j| < \infty;$$

How to check causality?

# Problem - Causality



AR(p):

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \epsilon_t \quad t = 0, \pm 1, \pm 2, \dots,$$

Causal:

$$Y_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j} \quad \sum_{j=0}^{\infty} |c_j| < \infty;$$

How to check causality?

Roots of the polynomial:

$$\Phi(x) = 1 - \phi_{p,1} x - \phi_{p,2} x^2 - \cdots \phi_{p,p} x^p$$

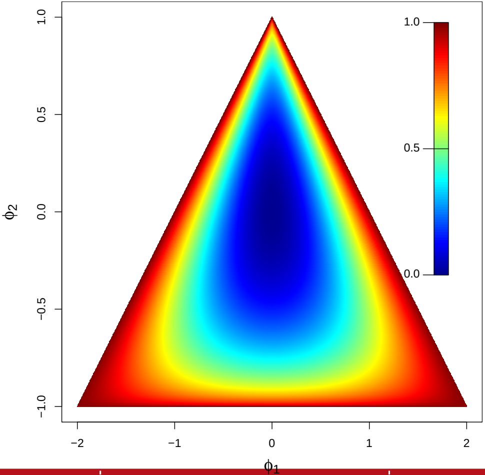
Are the roots outside the unit circle?

# Known solution spaces



AR1: $\phi_{1,1} \in (-1, 1)$

AR2: What about higher orders?



# A different way



PACF at lag $h$: $\alpha(h)$.

$$|\alpha(h)| < 1 \quad \text{for} \quad 1 \leq h \leq p \quad \text{and} \quad \alpha(h) = 0 \quad \text{for} \quad h > p.$$

Ramsey (1974) noted a 1:1 mapping. Any point in $(-1, 1)^p$ corresponds to a unique point in the causal parameter space:

$$(\phi_{p,1}, \ldots, \phi_{p,p}) = g\,(\alpha(1), \ldots, \alpha(p))$$

$$(\alpha(1), \ldots, \alpha(p)) = g^{-1}\left(\phi_{p,1}, \ldots, \phi_{p,p}\right).$$

# Durbin-Levinson

The map $g$ can be described via the Durbin-Levinson algorithm, which provides

$$\phi_{k,k} = \alpha(k)$$
$$\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j} \qquad k > 1, 1 < j < k-1.$$

# Equivalent approaches



Let $R_p$ denote the causal parameter space for an order $p$ AR. Any of the following conditions can be used to assess if $\phi \in R_p$.

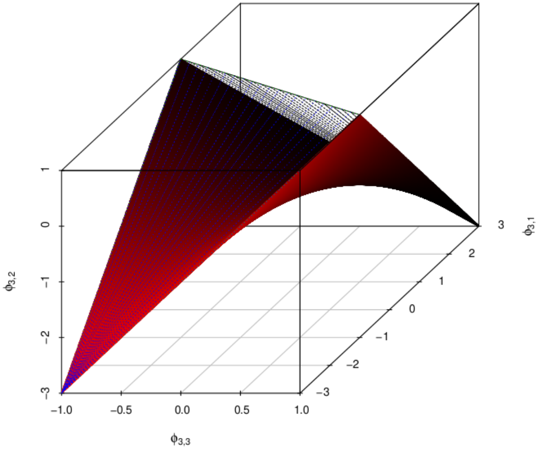
1. The roots of $\Phi(x) = 1 - \phi_{p,1}x - \phi_{p,2}x^2 - \cdots \phi_{p,p}x^p$ are outside the unit circle.
2. $|\phi_{h,h}| < 1$ for $1 \leq h \leq p$, and for each $k = 2, \ldots, p$ and $j = 1, \ldots, k$, $\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}$.
3. $|\phi_{p,p}| < 1$, $\phi_{2i,2i} > -1$ for $i < p/2$, and for each $k = 2, \ldots, p$ and $j = 1, \ldots, k$,

   $$\sum_{j=1}^{k} \phi_{k,j} < 1 \quad \text{and} \quad \sum_{j=1}^{k} (-1)^{j+1}\phi_{k,j} > -1,$$

   with $\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}$.
4. For $p > 1$, $(\phi_{p-1,1}, \ldots, \phi_{p-1,p-1}) \in R_{p-1}$, $|\phi_{p,p}| < 1$ and $\phi_{p,j} = \phi_{p-1,j} - \phi_{p,p}\phi_{p-1,p-j}$ for $j = 1, \ldots, p-1$.

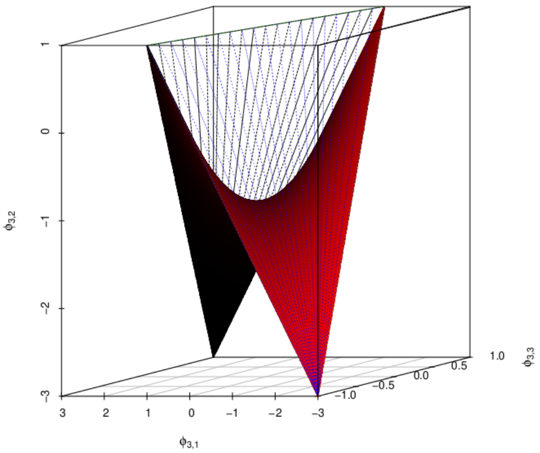
# Causal Region AR3





# Causal Region AR3





# Causal Region AR4

School of Mathematical Sciences | Lancaster University

Dropbox Link

# So what?



Need to summarise: $R_n(i,j) = corr(Y_i, Y_j)$, for $1 \leq i,j \leq n$

Measure of total correlation:

$$\rho(\{Y_t\}) = 1 - \det(R_n) = 1 - \prod_{k=1}^{p}(1-\alpha(k)^2)^{p+1-k}.$$

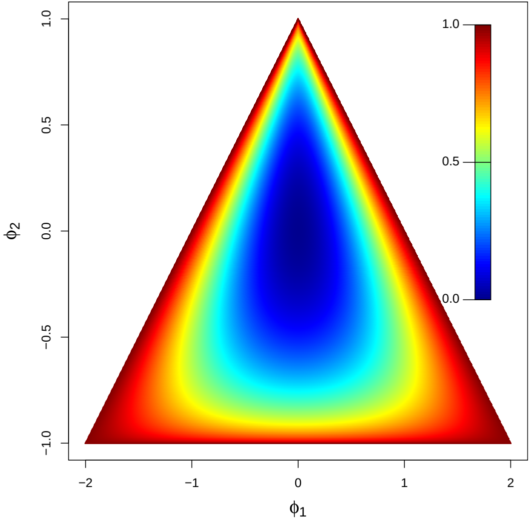
where $\det$ denotes the determinant.

$\rho(\{Y_t\}) \to 1$, high correlation

$\rho(\{Y_t\}) \to 0$, low correlation

# AR2 Total Correlation





# Simulation Studies



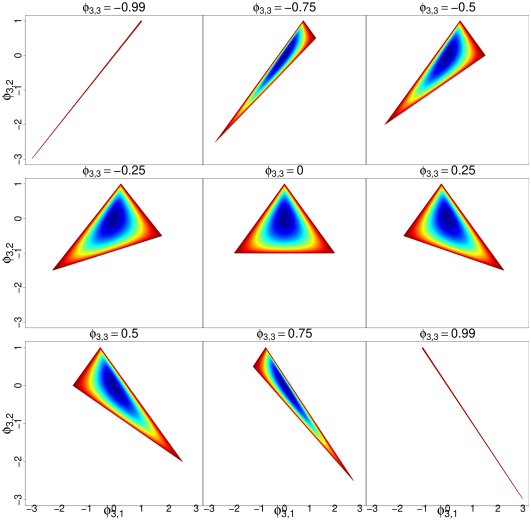
1. Generate $\{X_i\}_{i\in(1,\dots,p)}$ independently having, $0 \le \ell, u \le 1$ Uniform density
$$f(x) = \frac{1}{(u-\ell)}\mathbb{I}_{\ell<x<u}.$$
2. Generate $\{R_i\}$ independent of $\{X_i\}$, $P(R=1) = P(R=-1) = 0.5$.
3. Let each $\alpha(i) = R_i X_i$ and solve for $\{\phi_{p,j}\}$.

Larger values of $(\ell, u)$ for one or more $i$ correspond to more correlation.

# AR(3)

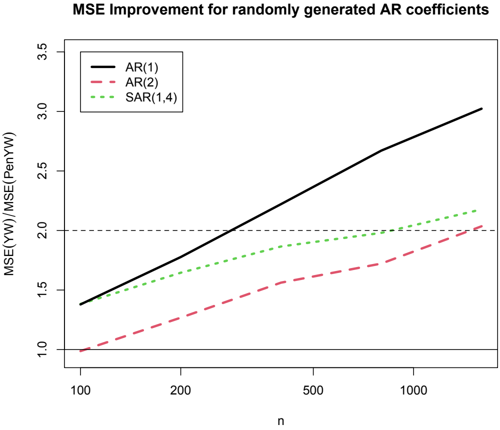


$\ell = 0.8$, $u = 0.9$, 10,000 simulations, $\rho(\{Y_t\}) > 0.9$, in 90% of simulations.

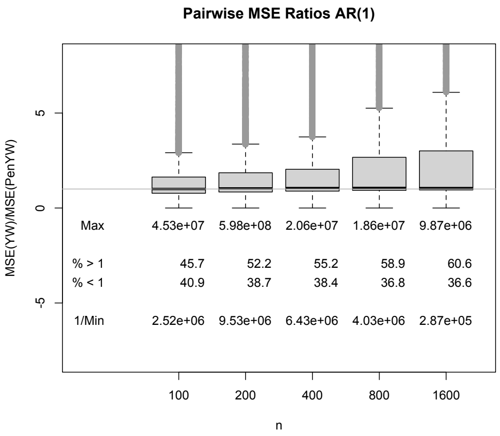


MSE for YW AR($p$) fits based on $\hat{\pi}$ vs $\tilde{\pi}$, 10,000 replications

Parameters randomly generated over parameter space.



Pairwise MSE for YW AR(1) fits based on $\hat{\pi}$ vs $\tilde{\pi}$.



# Summary



- Provided a new approach for (P)ACF estimation adaptively shrinking towards zero and target at each lag
- Improves over sample ACF, often beats Banded and Tapered too
- Computationally fast
- Adapts to a wide set of correlation structures

- Showed how PACF and AR coefficients are linked
- Used this link to describe causality regions in higher dimensions
- Visualized up to AR(4)
- Utilized the new description for simulation study generation

Available in `PenalizedCor` on Github.