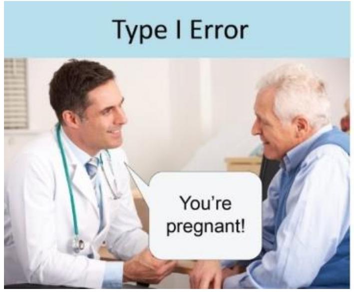
Exact Inference of Linear Dependence between Multiple Autocorrelated Time Series

Oliver Cliff, Leo Novelli, Ben Fulcher, Mac Shine, Joe Lizier

Centre for Complex Systems and Brain and Mind Centre collaboration
USYD

THE UNIVERSITY OF SYDNEY


Type I Error
You're
pregnant!
Type II Error
You're not
pregnant!



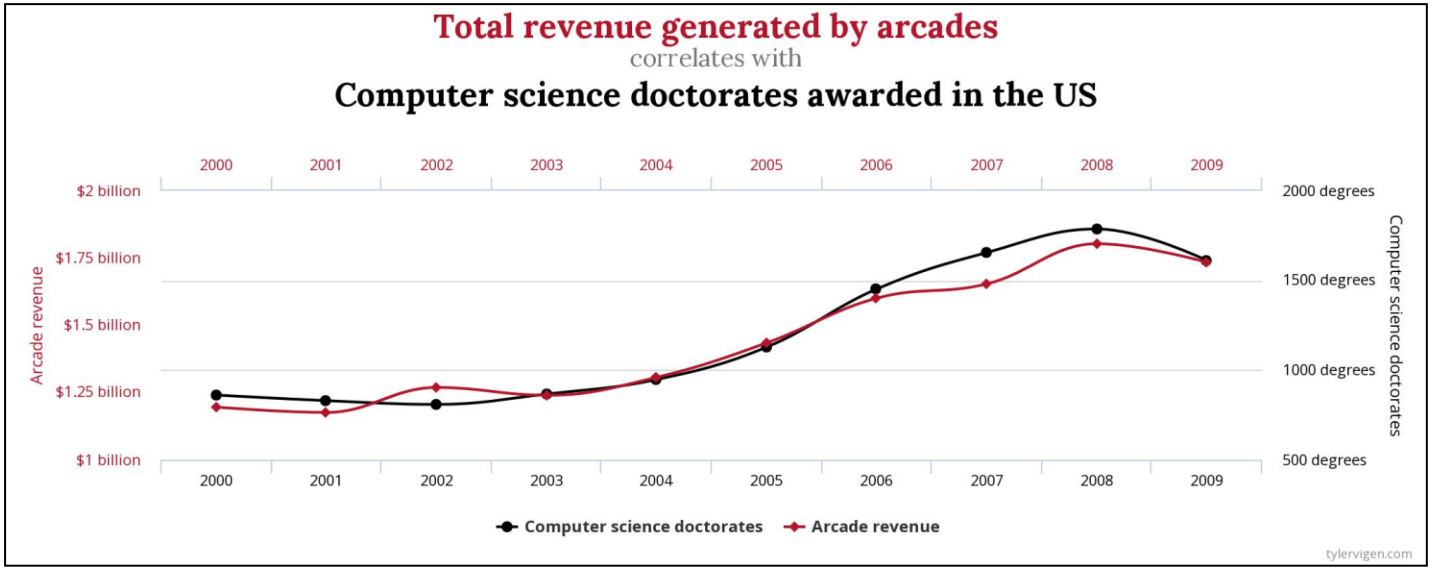
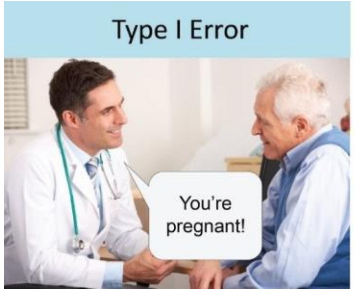


Image source: http://www.tylervigen.com/spurious-correlations



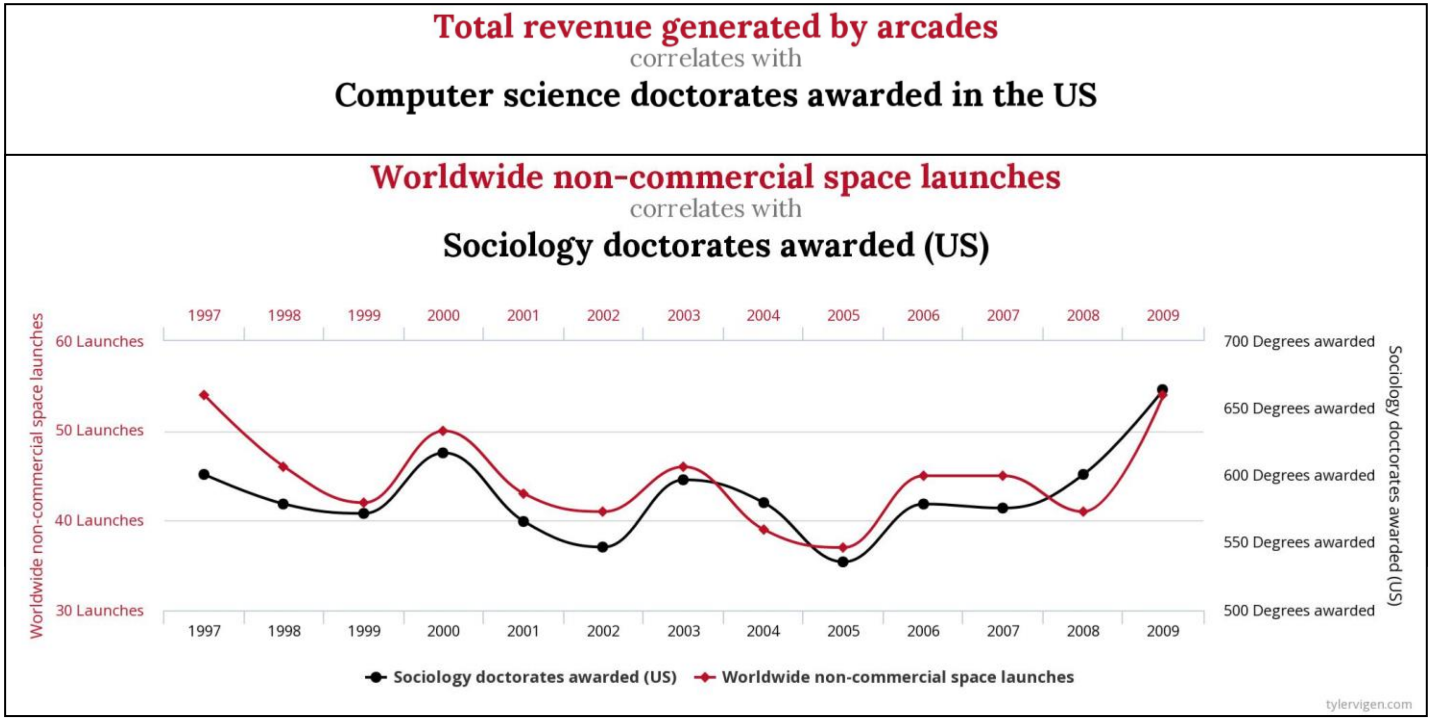


Image source: http://www.tylervigen.com/spurious-correlations

Total revenue generated by arcades
correlates with
Computer science doctorates awarded in the US

Worldwide non-commercial space launches
correlates with
Sociology doctorates awarded (US)

Per capita consumption of mozzarella cheese
correlates with
Civil engineering doctorates awarded

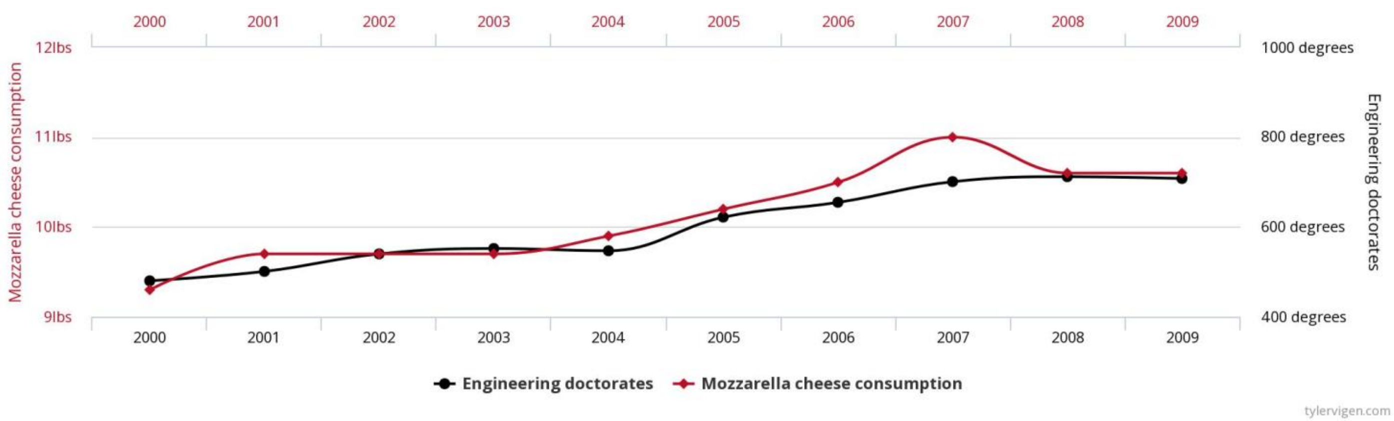
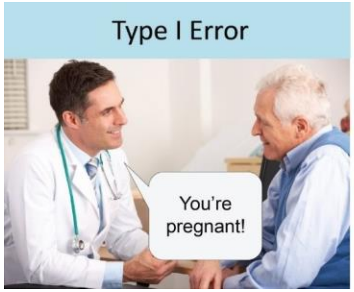


Image source: http://www.tylervigen.com/spurious-correlations



Total revenue generated by arcades
correlates with
Computer science doctorates awarded in the US

Worldwide non-commercial space launches
correlates with
Sociology doctorates awarded (US)

Per capita consumption of mozzarella cheese
correlates with
Civil engineering doctorates awarded

Math doctorates awarded
correlates with
Uranium stored at US nuclear power plants

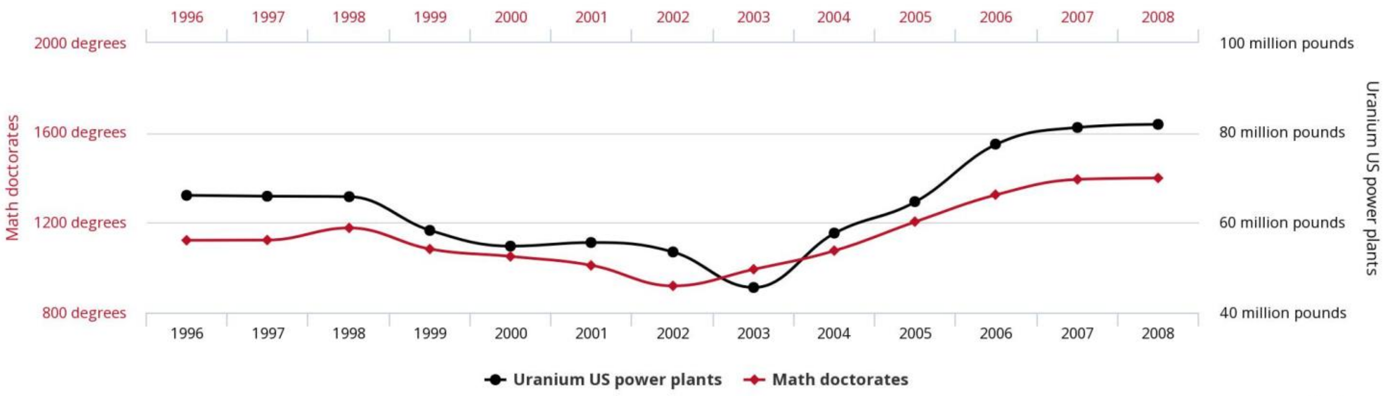


Image source: http://www.tylervigen.com/spurious-correlations

False positives in neuroscience

- Many causes for false positives in neuroscience
  - Spatial autocorrelation in the brain [Burt et al., NeuroImage (2020)]
  - Co-expression of genes (for GSEA) [Fulcher et al., bioRxiv (2020)]
  - **Temporal autocorrelation in neuroimaging** (fMRI, M/EEG, etc.) [Afyouni et al., NeuroImage (2019)]

- It is known that Pearson correlation is affected but so are others:
  - Canonical correlation analysis
  - **(Conditional) mutual information** (undirected and multivariate)
  - **Granger causality** (directed and multivariate)

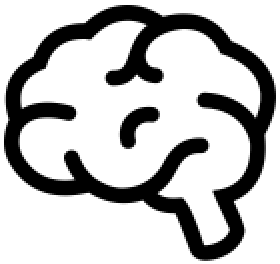
- We provide the exact null distributions for these measures under autocorrelation

Ridiculous question: Does one patient’s fMRI data influence another’s?

THE UNIVERSITY OF SYDNEY

Neuroimaging case study

Ridiculous question: Does one patient’s fMRI data influence another’s?

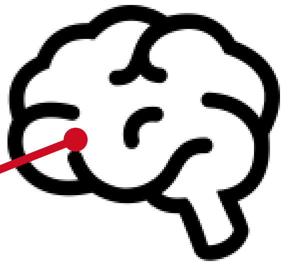


Human connectome project

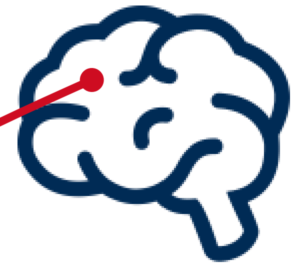
- Resting-state fMRI data
- Independent subjects

Neuroimaging case study

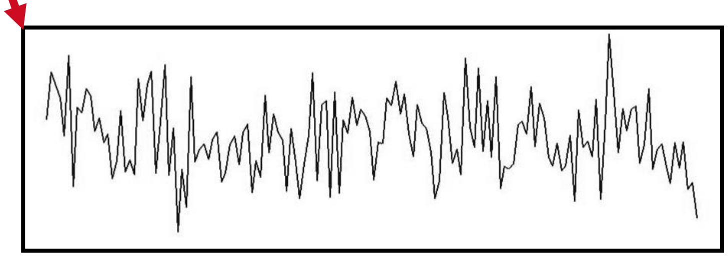
Ridiculous question: Does one patient's fMRI data influence another's?

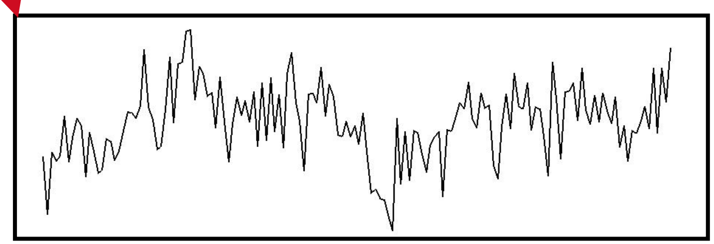


Human connectome project
- Resting-state fMRI data
- Independent subjects



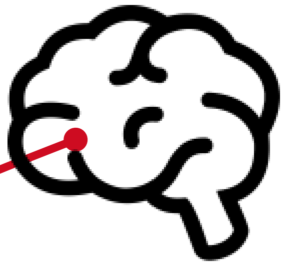
- Independent regions
- Sample size = 800
- Detrended and z-scored.

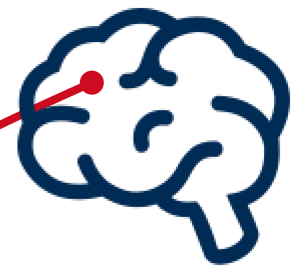




Neuroimaging case study

Ridiculous question: Does one patient's fMRI data influence another's?

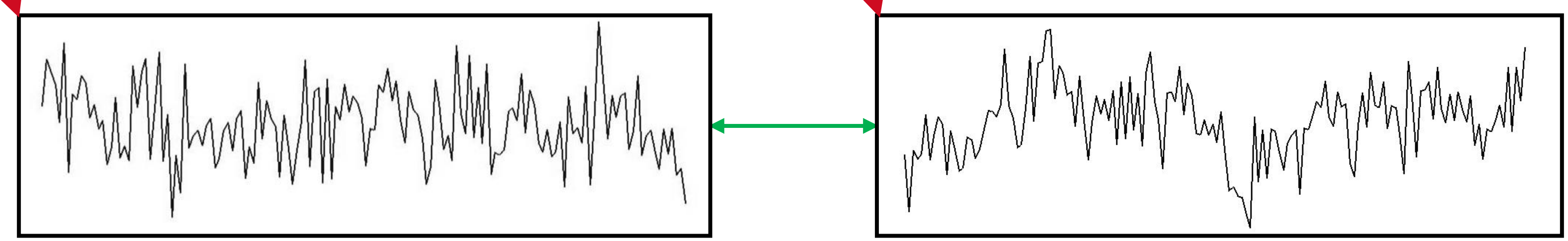




Human connectome project

- Resting-state fMRI data
- Independent subjects

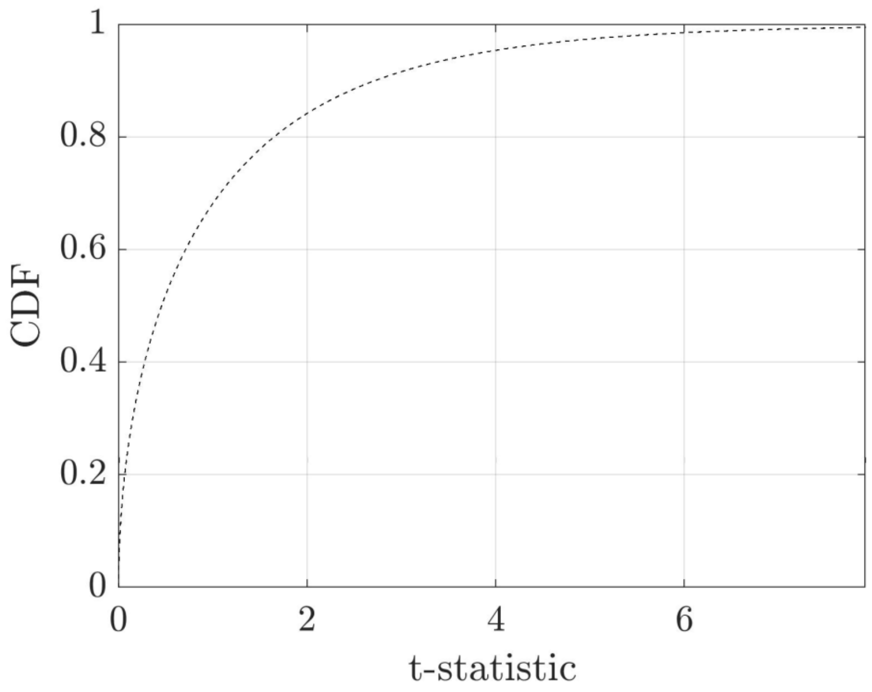
- Independent regions
- Sample size = 800
- Detrended and z-scored.

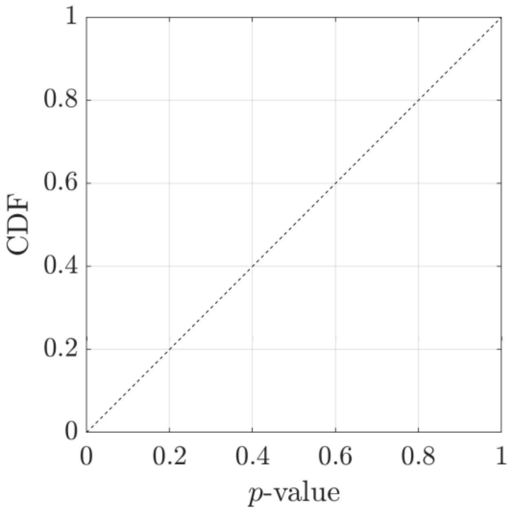


correlation, mutual information, Granger causality, etc.

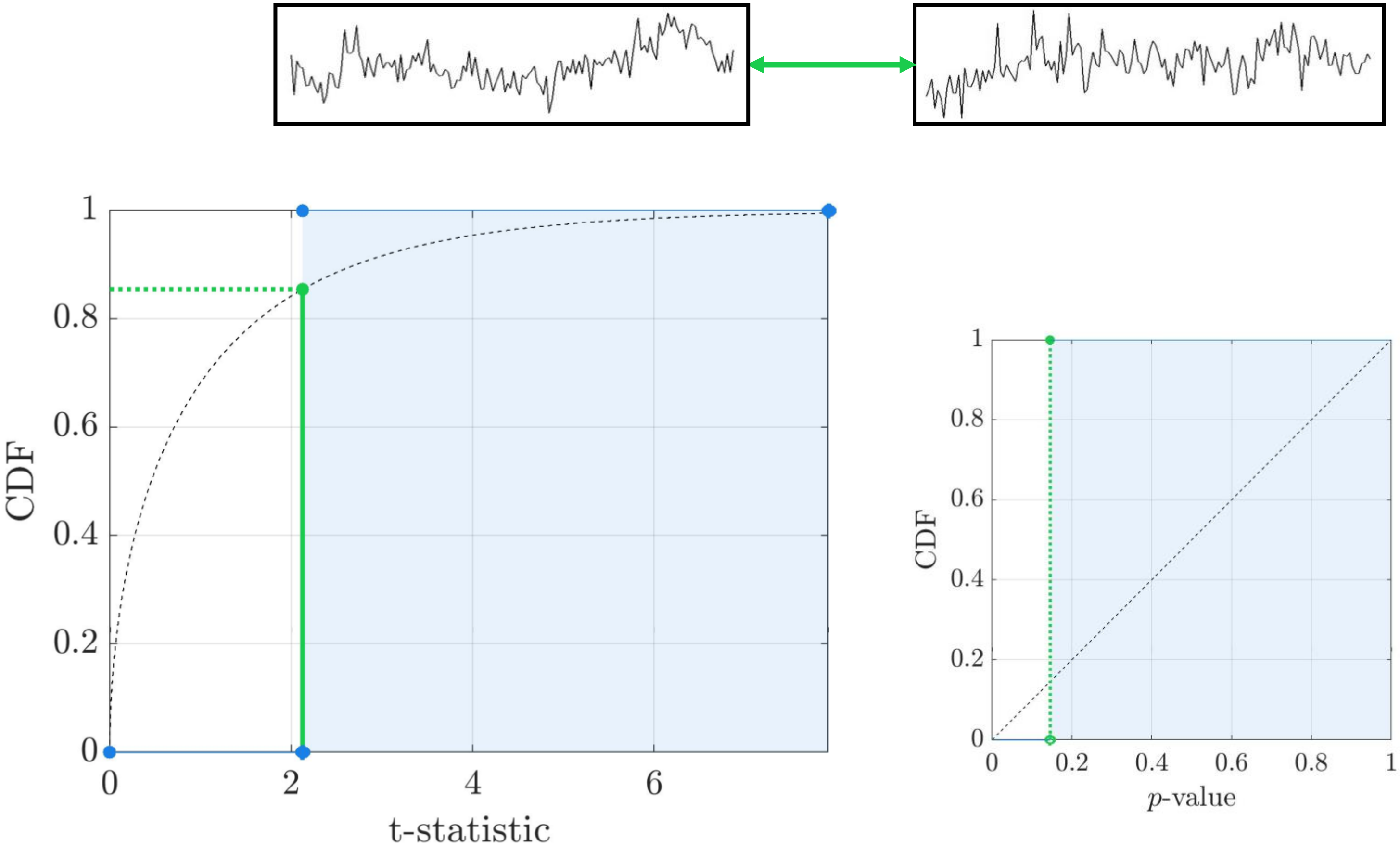
Neuroimaging FPR: Pearson correlation

› Test for bivariate correlation
  - Assumes normally distributed **i.i.d variables**
  - Null case follows Student's t-distribution

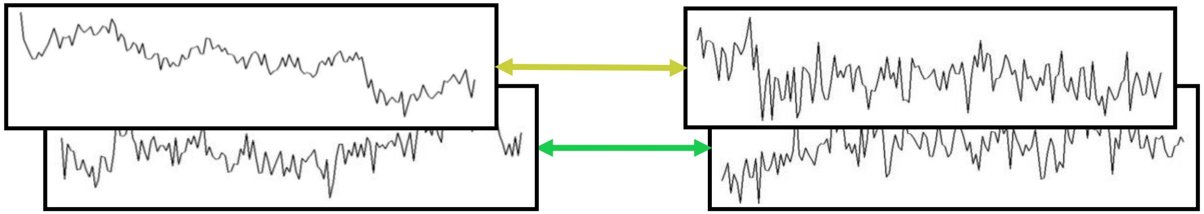


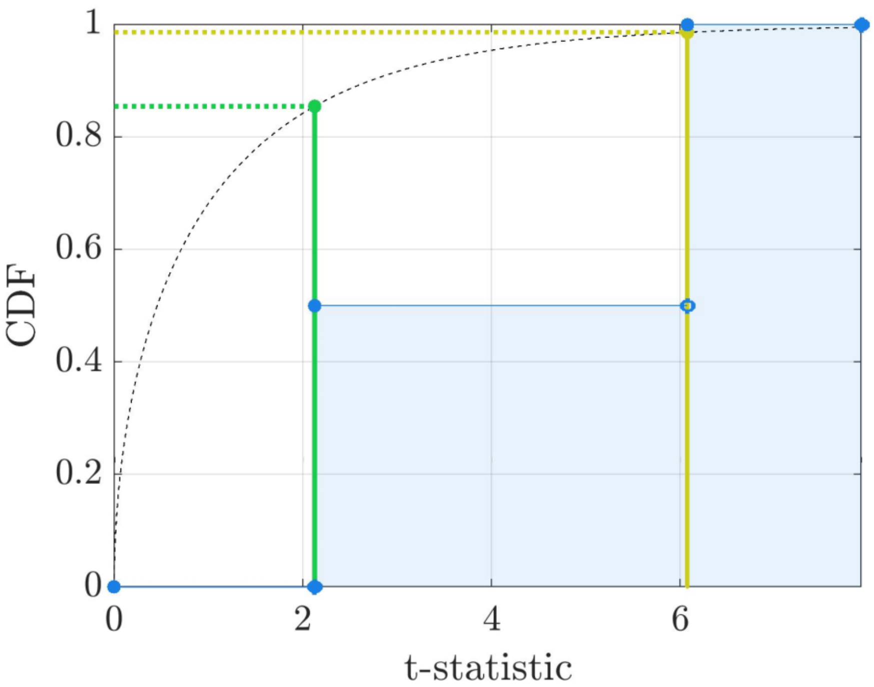


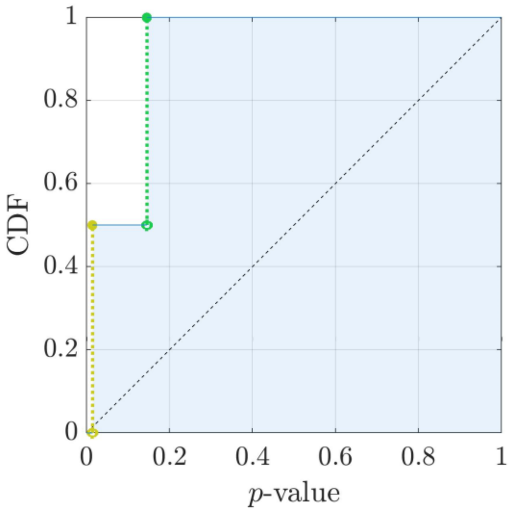
Neuroimaging FPR: Pearson correlation



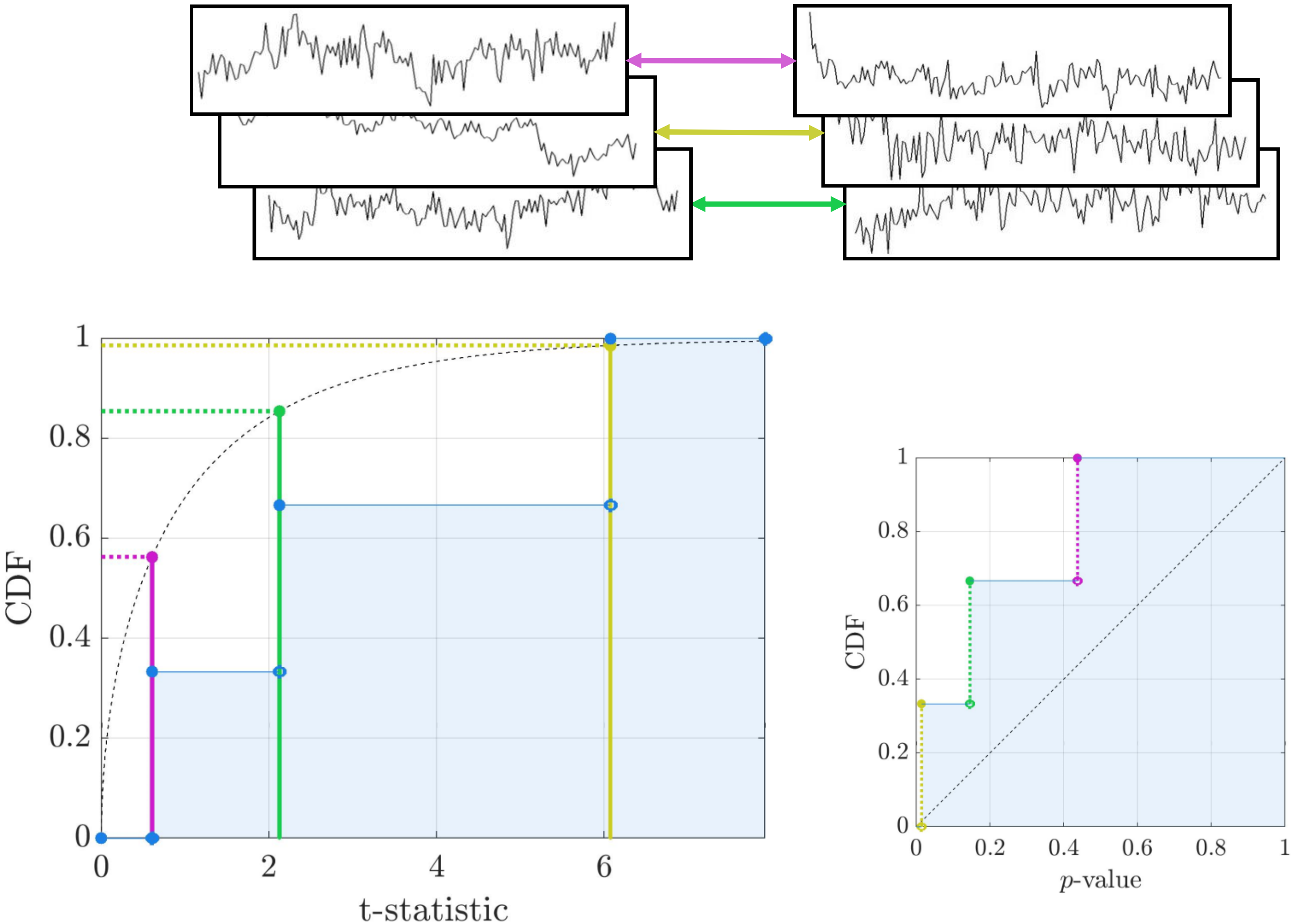
Neuroimaging FPR: Pearson correlation



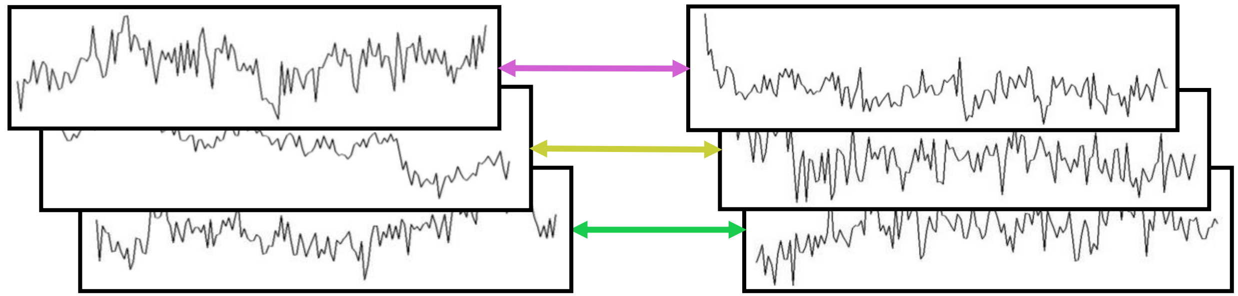


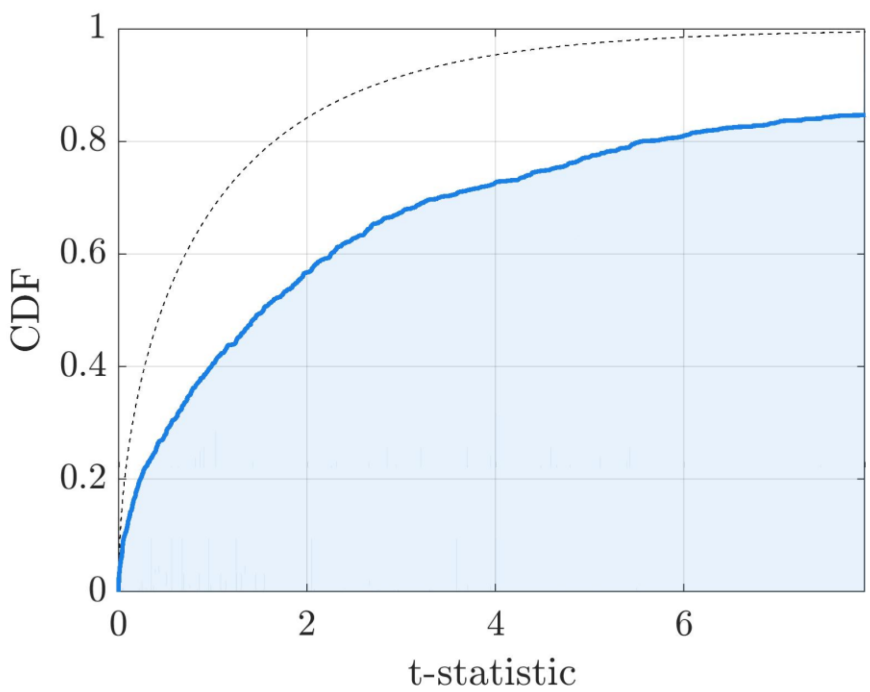


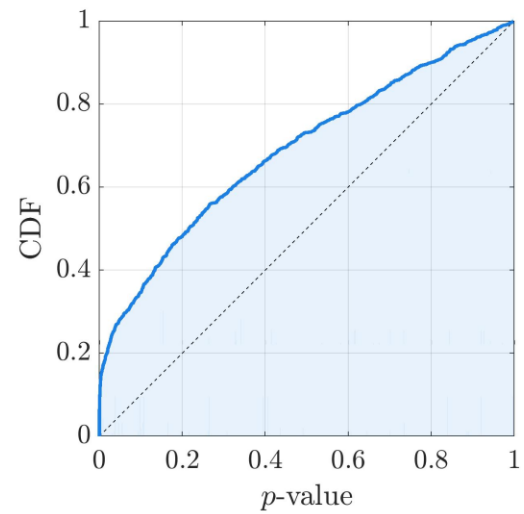
Neuroimaging FPR: Pearson correlation



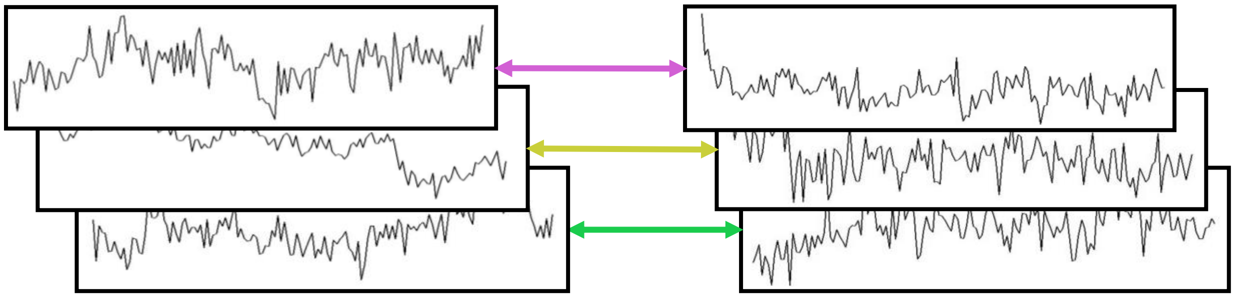
Neuroimaging FPR: Pearson correlation

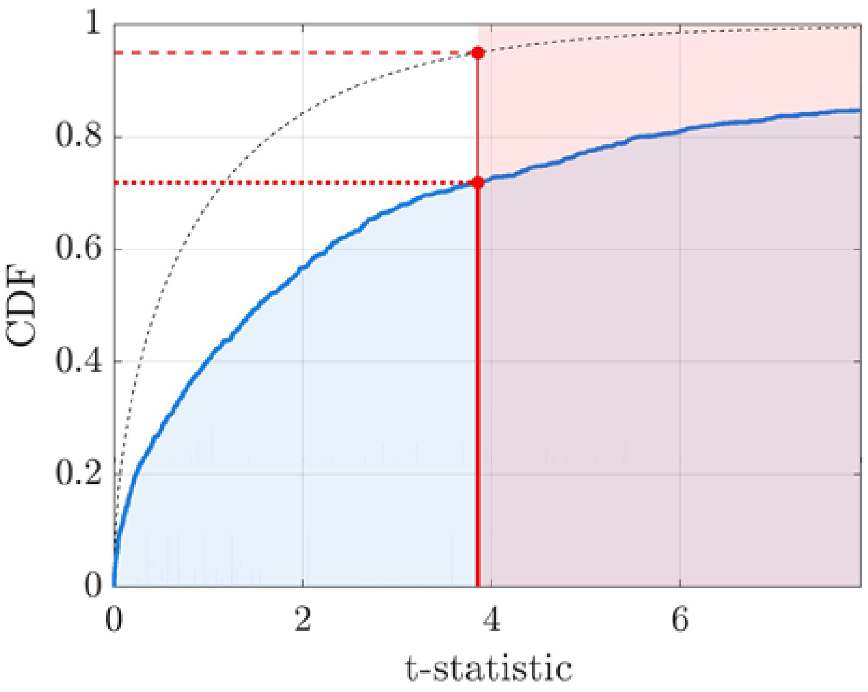




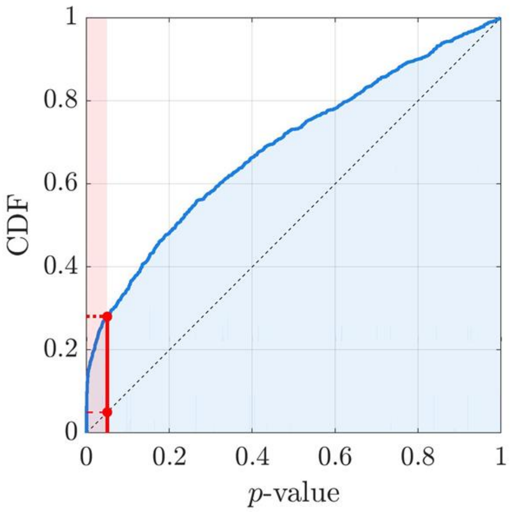


Neuroimaging FPR: Pearson correlation





34% FPR (for a nominal 5%)



THE UNIVERSITY OF SYDNEY

Why do these false positives occur?

- This doesn't just happen with fMRI data but lots of stationary time series.

- Stationary time series are defined by two things:
  - Mean
  - Autocovariance

- Pearson correlation (and other linear-dependence measures) are scale-invariant, so the mean has no effect.

- Let's investigate the autocovariance/autocorrelation…

Linear dependence under autocorrelation

- Higher autocorrelation increases the FPR of sample correlation
  - This has been known since Yule/Pearson's work in the early 20th century

- How do we fix it?
  - **Bartlett's formula** [Bartlett (1935)]
    - Correct the sample size based on autocorrelation of univariate signals
    - Does the process pass the t-test with effective sample size?
  - **Granger causality** [Granger (1969)]
    - Define process through autocovariance
    - Does adding in another process improve predictability?

Linear dependence under autocorrelation

- Higher autocorrelation increases the FPR of sample correlation
  - This has been known since Yule/Pearson's work in the early 20th century

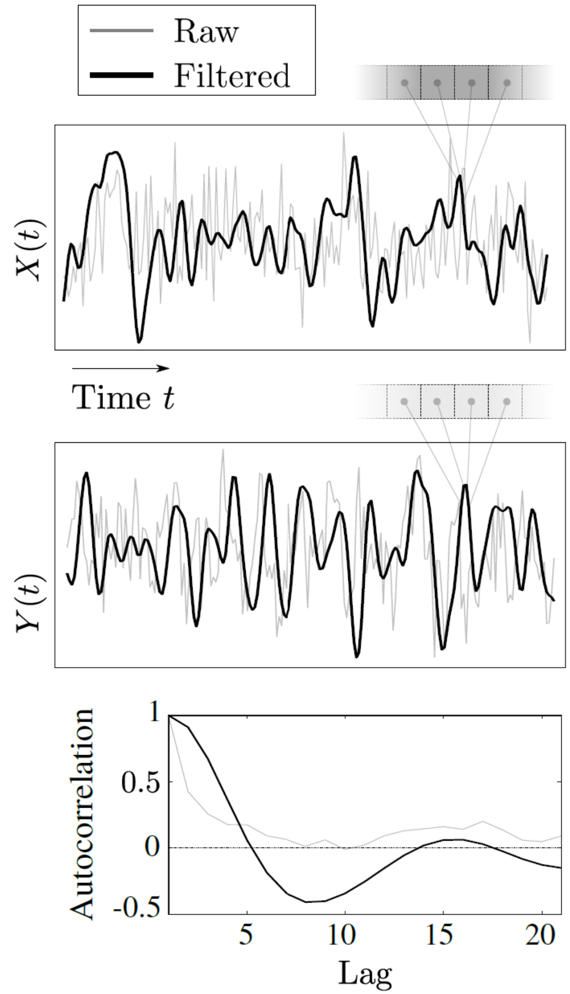
- How do we fix it?
  - **Bartlett's formula** [Bartlett (1935)]
    - Correct the sample size based on autocorrelation of univariate signals
    - Does the process pass the t-test with effective sample size?
  - **Granger causality** [Granger (1969)]
    - Define process through autocovariance
    - Does adding in another process improve predictability?

Bartlett's formula

- Compute the *effective sample size*:

$$\eta(x,y) = T\left[1 + 2\sum_{u=1}^{T-1} r_{xx}(u) r_{yy}(u)\right]^{-1}$$

- Rarely equal to sample size for time series.
  - **ESS < sample size** when:
    - both x and y are positively autocorrelated
  - **ESS > sample size** when:
    - One is positively autocorrelated and one is negatively autocorrelated



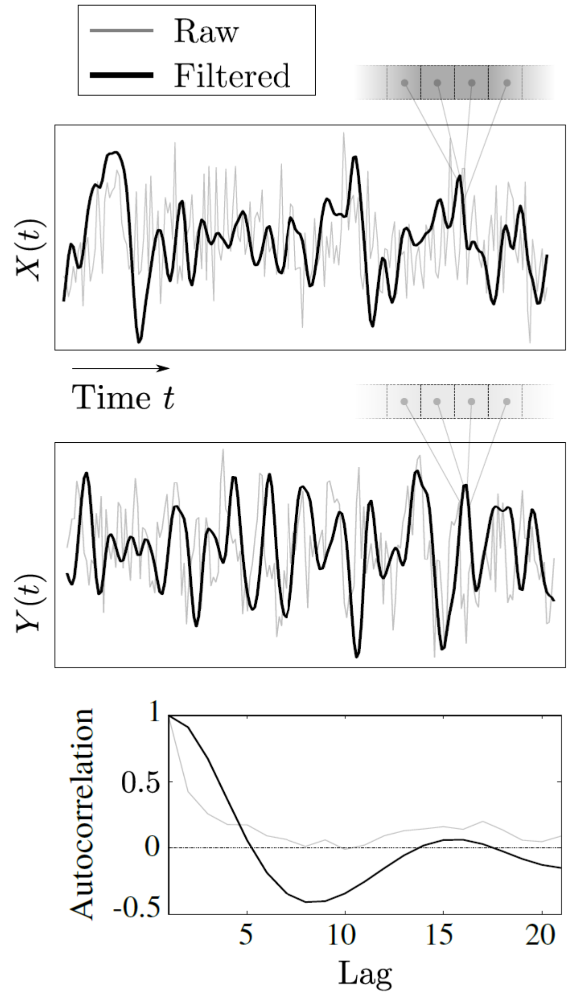
Bartlett's formula

› Test statistic against *t*-distribution with ESS:

$$r_{xy}\sqrt{\frac{\eta(x,y)-2}{1-r_{xy}^2}} \sim t(\eta(x,y)-2)$$

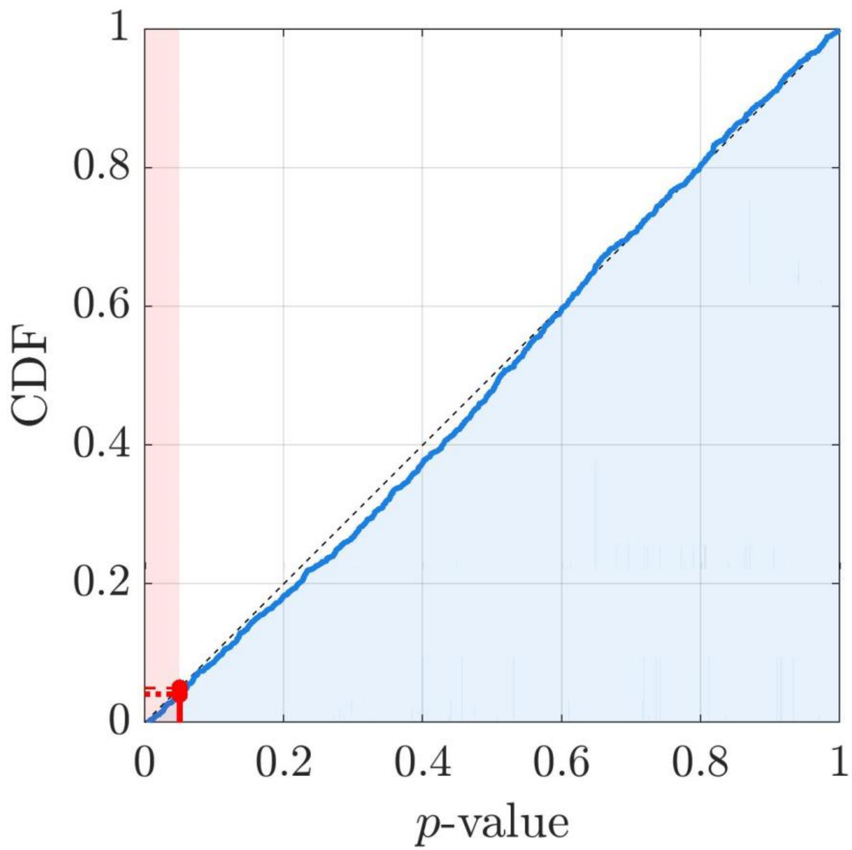
› Standard t-test Type I & II errors.

- False **positives** occur when**:**
  - both x and y are positively autocorrelated
- False **negatives** occur when**:**
  - One is positively autocorrelated and one is negatively autocorrelated

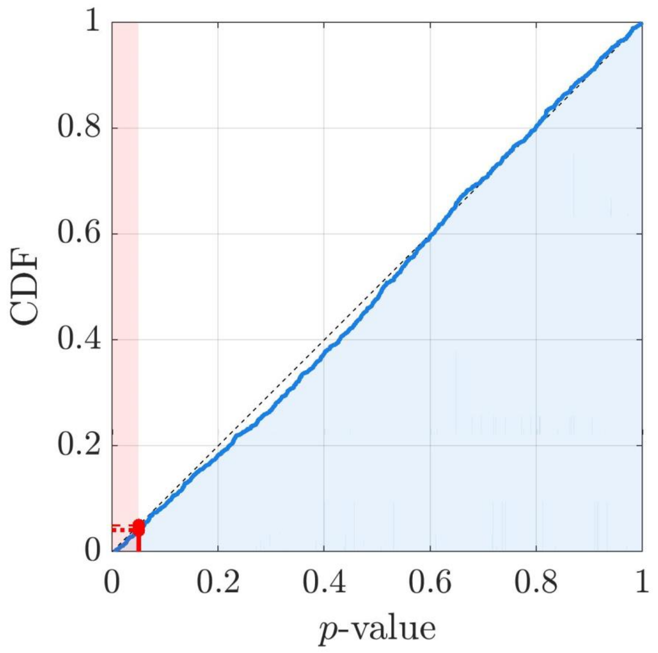


Exact Pearson correlation tests

5% FPR (for a nominal 5%)



Exact Pearson correlation tests



- Bartlett's formula corrects the false-positive bias in Pearson correlation
- This is but one linear-dependence measure, others can be directed and multivariate:
  - Granger causality
  - Mutual information
  - Canonical correlation analysis

THE UNIVERSITY OF SYDNEY

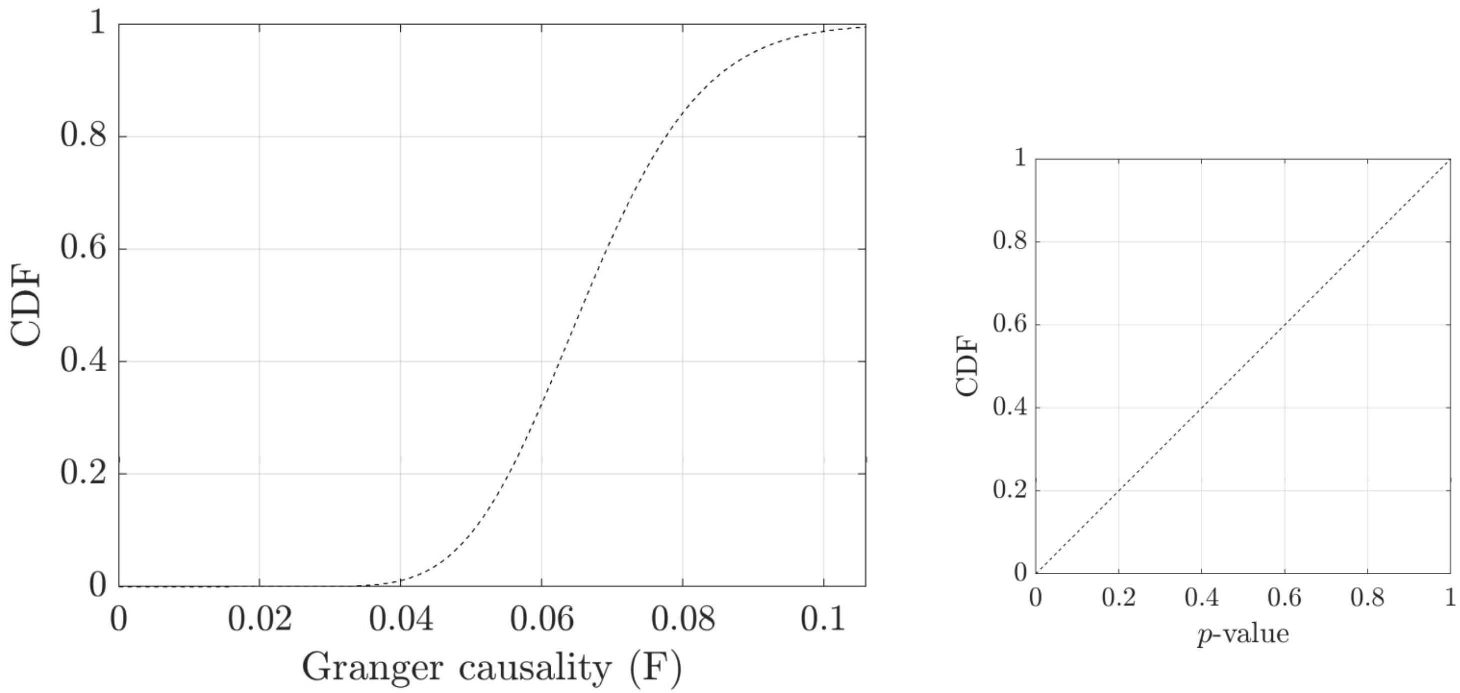
Linear dependence under autocorrelation

- Higher autocorrelation increases the FPR of sample correlation
  - This has been known since Yule/Pearson's work in the early 20th century

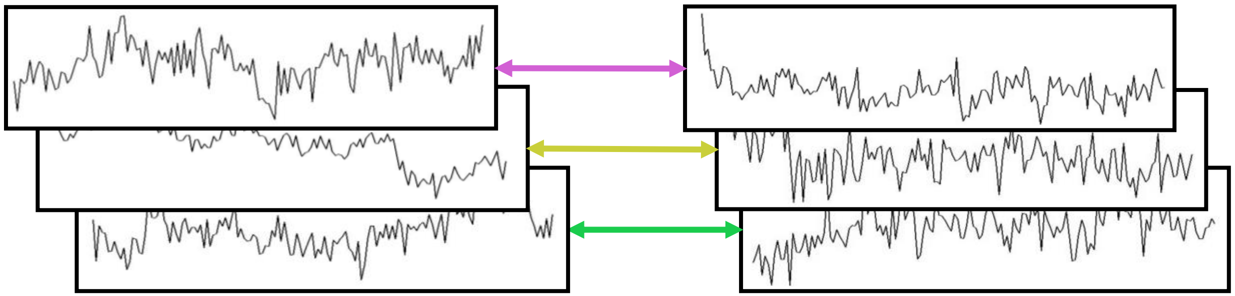
- How do we fix it?
  - **Bartlett's formula** [Bartlett (1935)]
    - Correct the sample size based on autocorrelation of univariate signals
    - Does the process pass the t-test with effective sample size?
  - **Granger causality** [Granger (1969)]
    - Define process through autocovariance
    - Does adding in another process improve predictability?

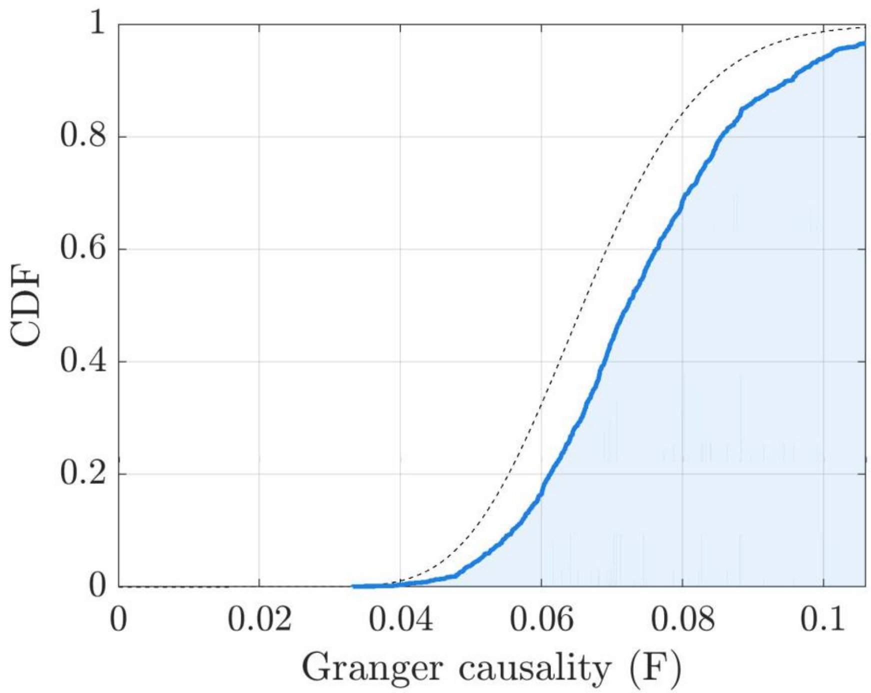
Neuroimaging FPR: Granger causality

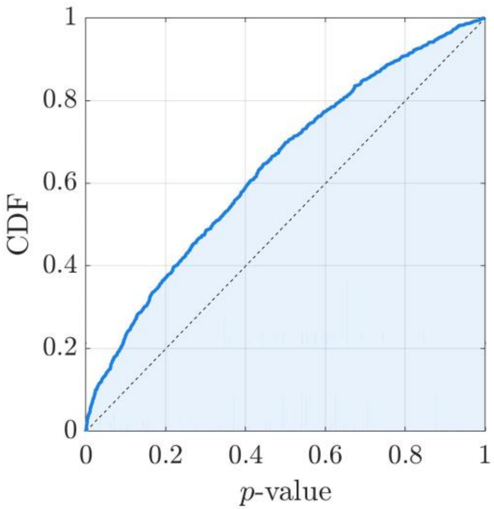
- Univariate test for directed predictability
  - Assumes Markov chain (here we use order 50)
  - Null distribution is **asymptotically** chi-square



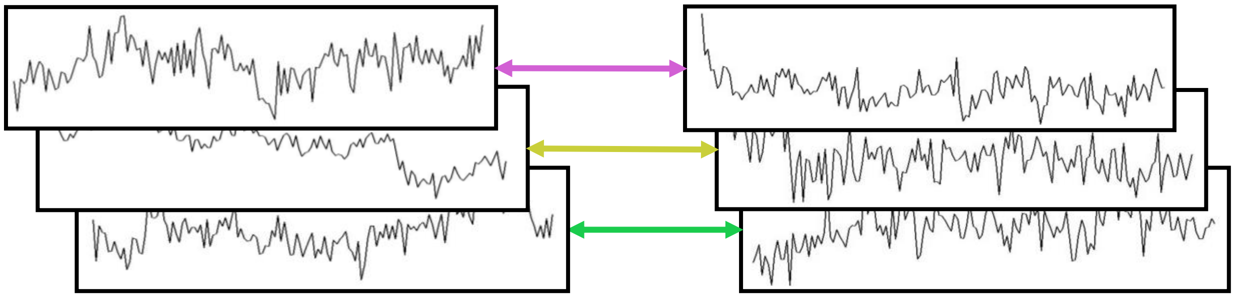
Neuroimaging FPR: Granger causality

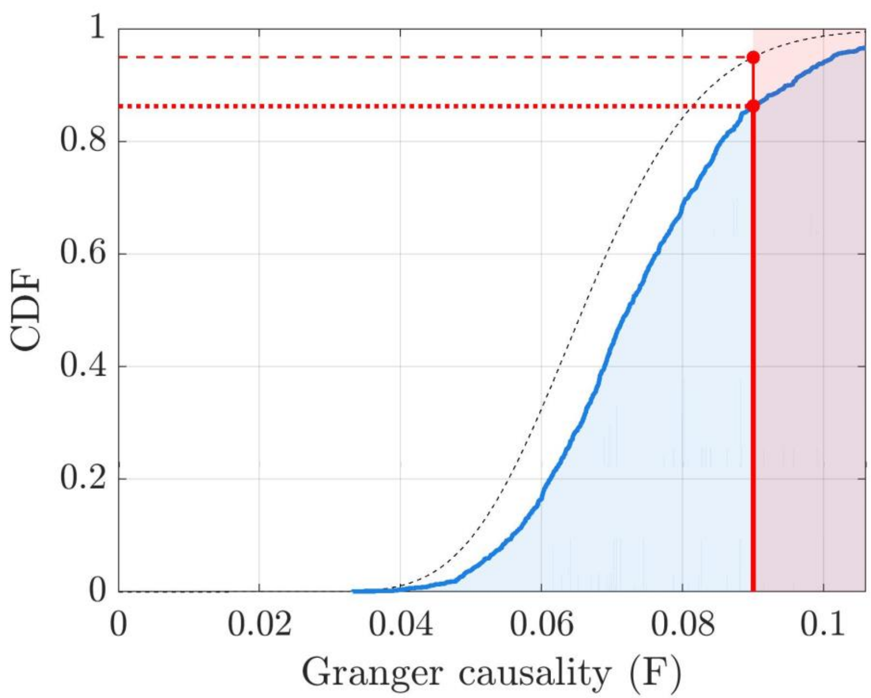




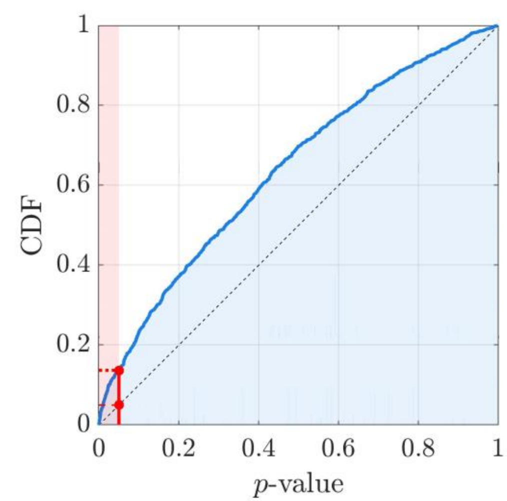


Neuroimaging FPR: Granger causality





14% FPR (for a nominal 5%)



Unifying Bartlett’s formula and multivariate measures

- Bivariate correlation
  - Undirected and univariate time series
  - Exact t-tests given due to Bartlett (1935)

- Mutual information
  - Undirected and multivariate time series
  - Asymptotic chi-square tests due to Wilks (1938)
  - Exact F-tests for iid variables

- Granger causality
  - Directed and multivariate time series
  - Asymptotic chi-square tests due to Wilks (1938)

Sampling distribution of multivariate measures

1. **Partial Correlation**
   1. Derive one-tailed test from Bartlett's formula
   2. Exact two-tailed test by squaring
2. **Conditional mutual information (between two time series)**
   1. Express as a squared partial correlation
   2. Exact test from two-tailed PC test
3. **Conditional mutual information (between multiple time series)**
   1. Use chain rule to decompose as a sum of CMI terms
   2. Exact test by summing each two-tailed PC test
4. **Granger causality** (between two/multiple time series)
   1. Equivalent to conditional mutual information (Barnett et al., 2009)
   2. Exact test from CMI (above)

Sampling distribution of multivariate measures

1. **Partial Correlation**
   1. Derive one-tailed test from Bartlett’s formula
   2. Exact two-tailed test by squaring
2. **Conditional mutual information (between two time series)**
   1. Express as a squared partial correlation
   2. Exact test from two-tailed PC test
3. **Conditional mutual information (between multiple time series)**
   1. Use chain rule to decompose as a sum of CMI terms
   2. Exact test by summing each two-tailed PC test
4. **Granger causality** (between two/multiple time series)
   1. Equivalent to conditional mutual information (Barnett et al., 2009)
   2. Exact test from CMI (above)

THE UNIVERSITY OF SYDNEY

Tests for partial correlation

› Take residuals of both processes:

$$e_{x|\boldsymbol{w}} = x - \hat{x}(\boldsymbol{w})$$
$$e_{y|\boldsymbol{w}} = y - \hat{y}(\boldsymbol{w})$$

Tests for partial correlation

› Take residuals of both processes:

$$e_{x|\boldsymbol{w}} = x - \hat{x}(\boldsymbol{w})$$
$$e_{y|\boldsymbol{w}} = y - \hat{y}(\boldsymbol{w})$$

› Compute the correlation between those residuals:

$$r_{xy\cdot\boldsymbol{w}} = \frac{\sum_t \left[e_{x|\boldsymbol{w}}(t)\, e_{y|\boldsymbol{w}}(t)\right]^2}{\sum_t e^2_{x|\boldsymbol{w}}(t) \sum_t e^2_{y|\boldsymbol{w}}(t)} = \frac{S_{xy|\boldsymbol{w}}}{S_{x|\boldsymbol{w}}\, S_{y|\boldsymbol{w}}}$$

Tests for partial correlation

- Take residuals of both processes:

$$e_{x|\boldsymbol{w}} = x - \hat{x}(\boldsymbol{w})$$
$$e_{y|\boldsymbol{w}} = y - \hat{y}(\boldsymbol{w})$$

- Compute the correlation between those residuals:

$$r_{xy\cdot\boldsymbol{w}} = \frac{\sum_t \left[e_{x|\boldsymbol{w}}(t)\, e_{y|\boldsymbol{w}}(t)\right]^2}{\sum_t e^2_{x|\boldsymbol{w}}(t) \sum_t e^2_{y|\boldsymbol{w}}(t)} = \frac{\boxed{S_{xy|\boldsymbol{w}}}}{\boxed{S_{x|\boldsymbol{w}}}\boxed{S_{y|\boldsymbol{w}}}}$$

Sum of squares

Tests for partial correlation

› Take residuals of both processes:

$$e_{x|\boldsymbol{w}} = x - \hat{x}(\boldsymbol{w})$$
$$e_{y|\boldsymbol{w}} = y - \hat{y}(\boldsymbol{w})$$

› Compute the correlation between those residuals:

$$r_{xy\cdot\boldsymbol{w}} = \frac{\sum_t \left[e_{x|\boldsymbol{w}}(t)\, e_{y|\boldsymbol{w}}(t)\right]^2}{\sum_t e^2_{x|\boldsymbol{w}}(t) \sum_t e^2_{y|\boldsymbol{w}}(t)} = \frac{S_{xy|\boldsymbol{w}}}{S_{x|\boldsymbol{w}}\, S_{y|\boldsymbol{w}}}$$

› For **i.i.d. variables**, test against Student's t-distribution: $t(\nu)$

$$\nu = \boxed{T} - \boxed{c} - 2$$

Sample size

Dimension of conditional (c-variate)

Exact one-tailed tests for partial correlation

› Take residuals of both processes:

$$e_{x|\boldsymbol{w}} = x - \hat{x}(\boldsymbol{w})$$
$$e_{y|\boldsymbol{w}} = y - \hat{y}(\boldsymbol{w})$$

› Compute the correlation between those residuals:

$$r_{xy\cdot\boldsymbol{w}} = \frac{\sum_t \left[e_{x|\boldsymbol{w}}(t)\, e_{y|\boldsymbol{w}}(t)\right]^2}{\sum_t e^2_{x|\boldsymbol{w}}(t) \sum_t e^2_{y|\boldsymbol{w}}(t)} = \frac{S_{xy|\boldsymbol{w}}}{S_{x|\boldsymbol{w}}\, S_{y|\boldsymbol{w}}}$$

› For **autocorrelated processes**, test against Student's t-distribution with ESS of residuals:

$$r_{xy\cdot\boldsymbol{w}} \sqrt{\frac{\boxed{\eta\left(e_{x|\boldsymbol{w}}, e_{y|\boldsymbol{w}}\right)} - \boxed{c} - 2}{1 - r^2_{xy\cdot\boldsymbol{w}}}} \sim t\left(\boxed{\eta\left(e_{x|\boldsymbol{w}}, e_{y|\boldsymbol{w}}\right)} - \boxed{c} - 2\right)$$

Effective sample size

Dimension of conditional

Exact two-tailed test for partial correlation

› Effective degrees of freedom:

$$N_{xy|\boldsymbol{w}} = \eta(e_{x|\boldsymbol{w}}, e_{y|\boldsymbol{w}}) - c - 2$$

› Now, squaring the statistic gives a two-tailed test:

$$r_{xy\cdot\boldsymbol{w}}\sqrt{\frac{\eta\big(e_{x|\boldsymbol{w}}, e_{y|\boldsymbol{w}}\big) - c - 2}{1 - r^2_{xy\cdot\boldsymbol{w}}}} \sim t(\eta\big(e_{x|\boldsymbol{w}}, e_{y|\boldsymbol{w}}\big) - c - 2)$$

$$\downarrow$$

$$N_{xy|\boldsymbol{w}}\,\frac{r^2_{xy\cdot\boldsymbol{w}}}{1 - r^2_{xy\cdot\boldsymbol{w}}} \sim F(1, N_{xy|\boldsymbol{w}})$$

The $\mathcal{L}$-distribution

› Let's take our (squared) statistic…

$$N_{xy|\boldsymbol{w}} \frac{r^2_{xy\cdot\boldsymbol{w}}}{1 - r^2_{xy\cdot\boldsymbol{w}}} \sim F(1, N_{xy|\boldsymbol{w}})$$

The $\mathcal{L}$-distribution

- Let's take our (squared) statistic…

$$N_{xy|\boldsymbol{w}} \frac{r^2_{xy\cdot\boldsymbol{w}}}{1 - r^2_{xy\cdot\boldsymbol{w}}} \sim F(1, N_{xy|\boldsymbol{w}})$$

- …and transform it:

$$\log\left(\frac{N_{xy|\boldsymbol{w}}}{N_{xy|\boldsymbol{w}}} \frac{r^2_{xy\cdot\boldsymbol{w}}}{1 - r^2_{xy\cdot\boldsymbol{w}}} + 1\right) = -\log\left(1 - r^2_{xy\cdot\boldsymbol{w}}\right)$$

The $\mathcal{L}$-distribution

- Let's take our (squared) statistic…

$$N_{xy|\boldsymbol{w}} \frac{r^2_{xy\cdot\boldsymbol{w}}}{1 - r^2_{xy\cdot\boldsymbol{w}}} \sim F(1, N_{xy|\boldsymbol{w}})$$

- …and transform it:

$$\log\left(\frac{N_{xy|\boldsymbol{w}}}{N_{xy|\boldsymbol{w}}} \frac{r^2_{xy\cdot\boldsymbol{w}}}{1 - r^2_{xy\cdot\boldsymbol{w}}} + 1\right) = \boxed{-\log\left(1 - r^2_{xy\cdot\boldsymbol{w}}\right)}$$

Equiv. to CMI

The $\mathcal{L}$-distribution

- Let's take our (squared) statistic…

$$N_{xy|\boldsymbol{w}} \frac{r^2_{xy\cdot\boldsymbol{w}}}{1 - r^2_{xy\cdot\boldsymbol{w}}} \sim F(1, N_{xy|\boldsymbol{w}})$$

- …and transform it:

$$\log\left(\frac{N_{xy|\boldsymbol{w}}}{N_{xy|\boldsymbol{w}}} \frac{r^2_{xy\cdot\boldsymbol{w}}}{1 - r^2_{xy\cdot\boldsymbol{w}}} + 1\right) = -\log\left(1 - r^2_{xy\cdot\boldsymbol{w}}\right)$$

- We label this statistic's null distribution *the $\mathcal{L}$-distribution*:

$$-\log\left(1 - r^2_{xy\cdot\boldsymbol{w}}\right) \sim \mathcal{L}(N_{xy|\boldsymbol{w}})$$

The $\mathcal{L}$-distribution

- This could be defined analytically from an F-distribution:

$$-\log\left(1 - r^2_{xy\cdot\boldsymbol{w}}\right) \sim \mathcal{L}(N_{xy|\boldsymbol{w}})$$

- But for our purposes it's best to sample $F \sim F(1, N)$

$$\log\left(\frac{F}{N} + 1\right) \sim \mathcal{L}(N)$$

- Because higher-dimensional measures (MI and Granger) are intractable.

$$\sum_{i=1}^{n} L_i = \sum_{i=1}^{n} \log\left(\frac{F_i}{N_i} + 1\right) \sim \mathcal{L}(\boldsymbol{N})$$

The $\mathcal{L}$-distribution

- This could be defined analytically from an F-distribution:

$$-\log\left(1 - r^2_{xy\cdot\boldsymbol{w}}\right) \sim \mathcal{L}(N_{xy|\boldsymbol{w}})$$

- But for our purposes it's best to sample $F \sim F(1, N)$

$$\log\left(\frac{F}{N} + 1\right) \sim \mathcal{L}(N)$$

- Because higher-dimensional measures (MI and Granger) are intractable.

$$\sum_{i=1}^{n} L_i = \sum_{i=1}^{n} \log\left(\boxed{\frac{F_i}{N_i}} + 1\right) \sim \mathcal{L}(\boldsymbol{N})$$

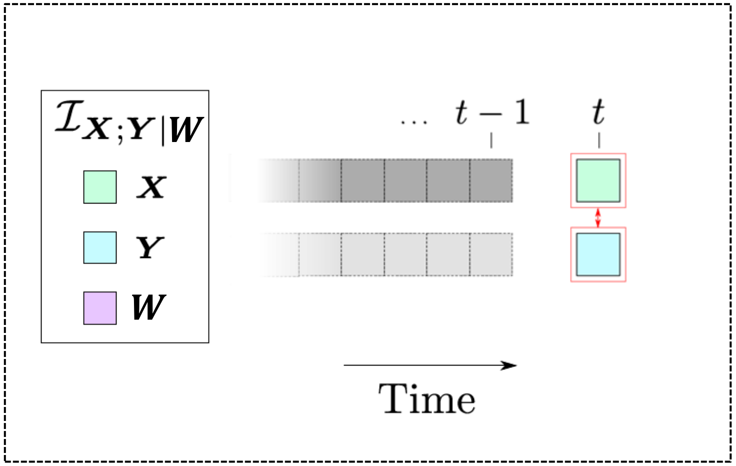
Potentially different for each L-term

Sampling distribution of multivariate measures

1. **Partial Correlation**
   1. Derive one-tailed test from Bartlett's formula
   2. Exact two-tailed test by squaring
2. **Conditional mutual information (between two time series)**
   1. Express as a squared partial correlation
   2. Exact test from two-tailed PC test
3. **Conditional mutual information (between multiple time series)**
   1. Use chain rule to decompose as a sum of CMI terms
   2. Exact test by summing each two-tailed PC test
4. **Granger causality** (between two/multiple time series)
   1. Equivalent to conditional mutual information (Barnett et al., 2009)
   2. Exact test from CMI (above)

(Conditional) mutual information

Between univariate Gaussians



- Equivalent to squared partial correlation [Davey (2013)]:

$$2\,\hat{\mathcal{I}}_{x;y|\boldsymbol{w}} = -\log(1 - r^2_{xy\cdot\boldsymbol{w}})$$

- Null distribution is asymptotically chi-square [Wilks (1938)]:

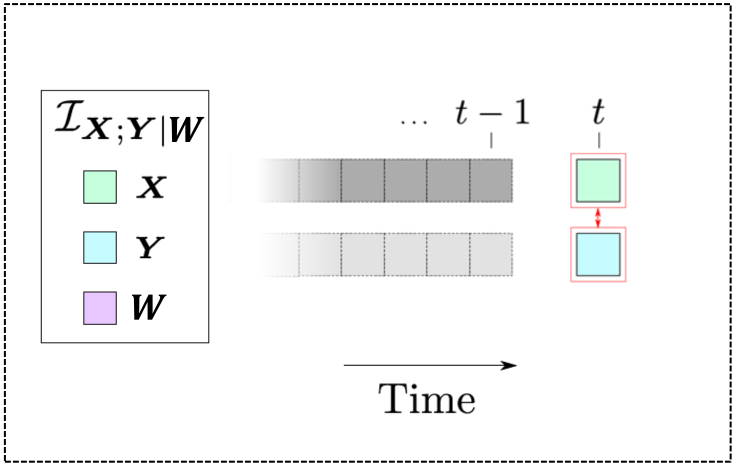
$$2T\,\hat{\mathcal{I}}_{\boldsymbol{x};\boldsymbol{y}|\boldsymbol{w}} \sim \chi^2(kl)$$

$$\begin{aligned} \dim(\boldsymbol{X}) &= k \\ \dim(\boldsymbol{Y}) &= l \\ \dim(\boldsymbol{W}) &= c \end{aligned}$$

- Exact null distribution for **i.i.d. variables** is F-distributed

$$\frac{T-(l+c+1)}{l}\left[\exp\left(2\hat{\mathcal{I}}_{x;\boldsymbol{y}|\boldsymbol{w}}\right)-1\right] \sim F(l, T-(l+c+1))$$

Between univariate Gaussians



› Equivalent to squared partial correlation [Davey (2013)]:

$$2\,\hat{\mathcal{I}}_{x;y|\boldsymbol{w}} = -\log(1 - r^2_{xy\cdot\boldsymbol{w}})$$

› Null distribution is asymptotically chi-square [Wilks (1938)]:

$$2T\,\hat{\mathcal{I}}_{\boldsymbol{x};\boldsymbol{y}|\boldsymbol{w}} \sim \boxed{\chi^2(kl)}$$

No conditional

$$\begin{aligned} \dim(\boldsymbol{X}) &= k \\ \dim(\boldsymbol{Y}) &= l \\ \dim(\boldsymbol{W}) &= c \end{aligned}$$

› Exact null distribution for **i.i.d. variables** is F-distributed

$$\frac{T-(l+\boxed{c}+1)}{l}\left[\exp\left(2\hat{\mathcal{I}}_{x;\boldsymbol{y}|\boldsymbol{w}}\right)-1\right] \sim F(l, T-(l+\boxed{c}+1))$$

Conditional

Exact (conditional) mutual information tests

Between univariate Gaussians

› Equivalent to squared partial correlation

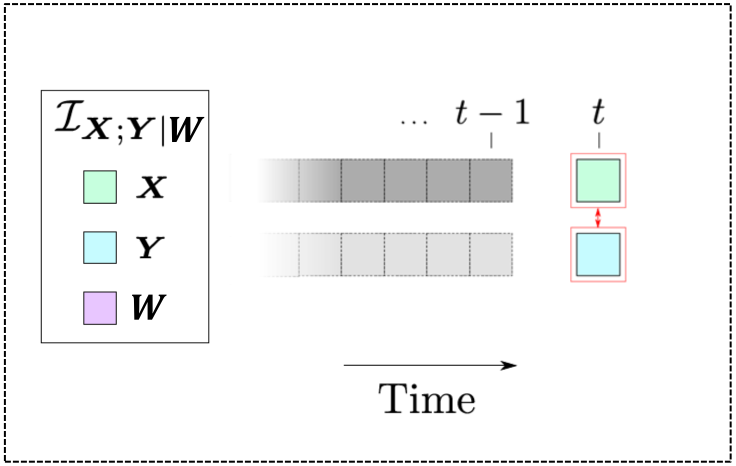
[Davey (2013)]

$$2\,\hat{\mathcal{I}}_{x;y|\boldsymbol{w}} = -\log(1 - r^2_{xy\cdot\boldsymbol{w}})$$

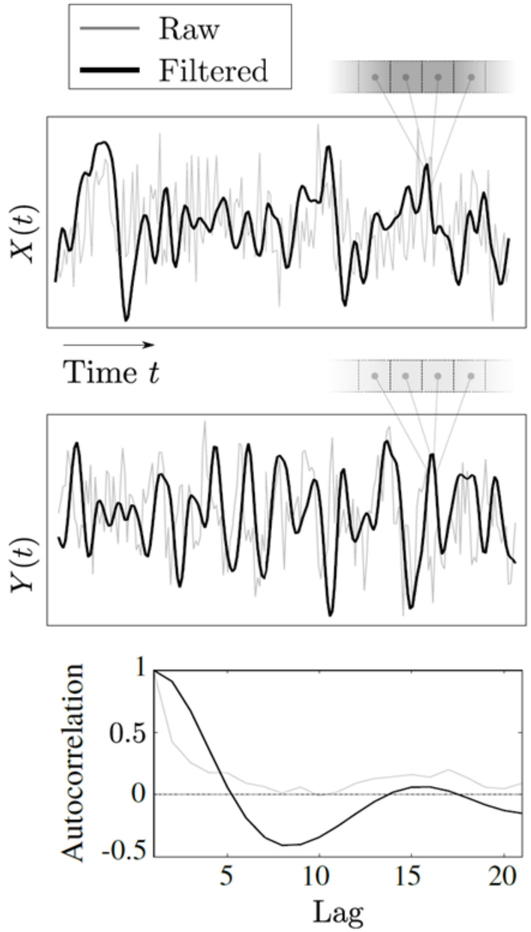
› So, we can use the $\mathcal{L}$-distribution:

$$-\log\left(1 - r^2_{xy\cdot\boldsymbol{w}}\right) \sim \mathcal{L}(N_{xy|\boldsymbol{w}})$$

$$2\,\hat{\mathcal{I}}_{x;y|\boldsymbol{w}} \sim \mathcal{L}(N_{xy|\boldsymbol{w}})$$

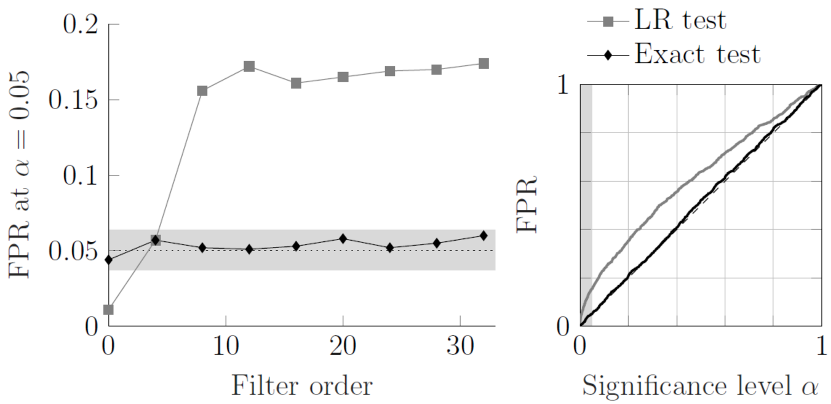


Numerical simulations

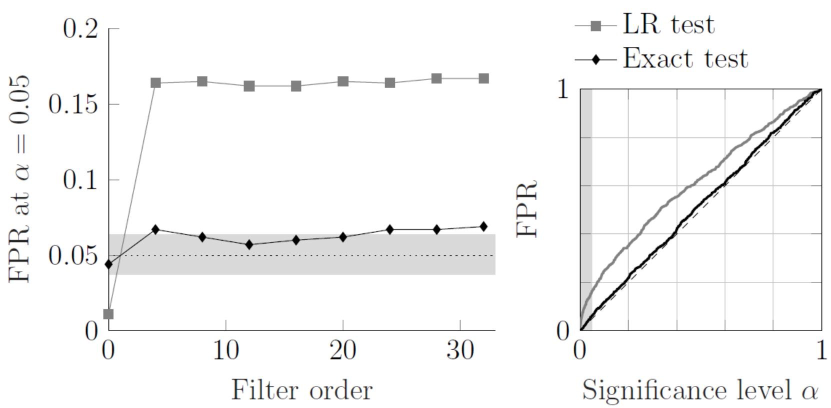


- Generate independent first-order autoregressive processes $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}$
  - 512 samples
  - $\dim(\boldsymbol{X}) = k,\ \dim(\boldsymbol{Y}) = l,\ \dim(\boldsymbol{W}) = c$
- Digitally filter with IIR/FIR filter
  - Increase filter order to increase AC
  - Common preprocessing step in neuro
- Perform 1000 trials of each configuration

Numerical simulations: MI for two time series



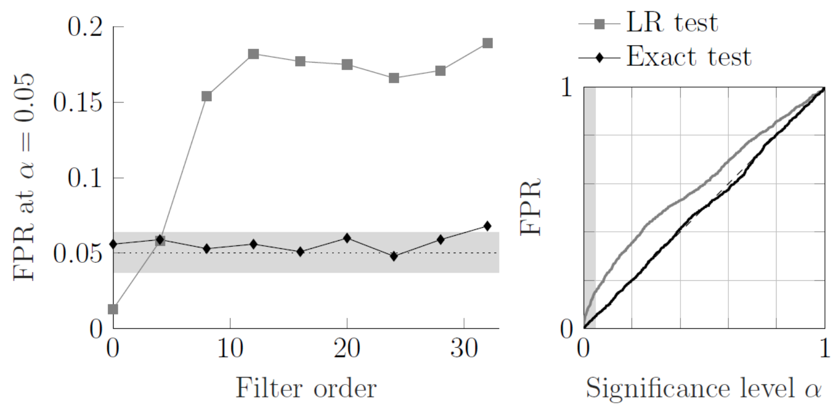
(a) FIR filter



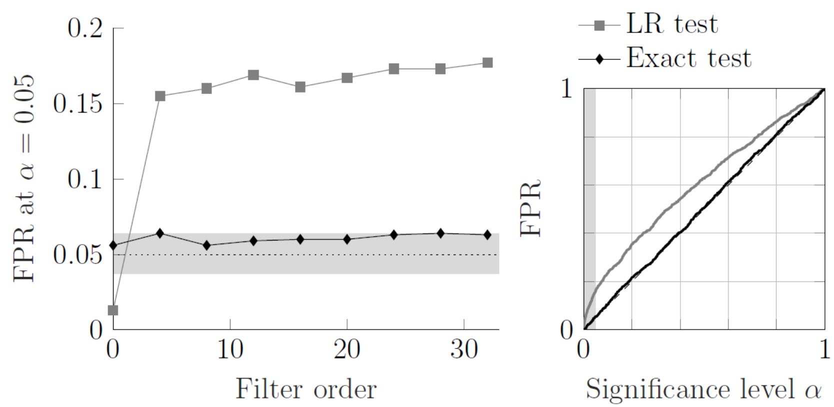
(b) IIR filter.

- Simulation:
  - X, Y are independent univariate processes ($k = l = 1$)
  - no conditional process W ($c = 0$)
  - Variable filter order
- Perform two tests:
  - **LR tests** are asymptotically valid
  - **Exact tests** are ours
- Both FIR and IIR increase the FPR dramatically to > 15% (3-times nominal value)
- Exact tests stay within (binominal) confidence intervals

Numerical simulations: CMI for two time series



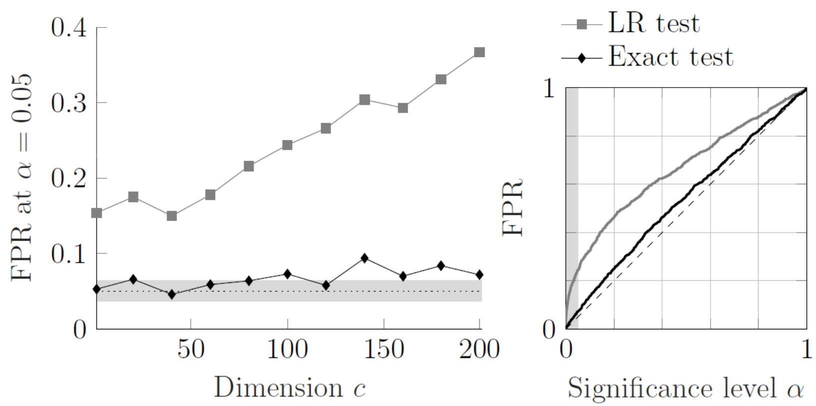
(a) FIR filter with univariate conditional.



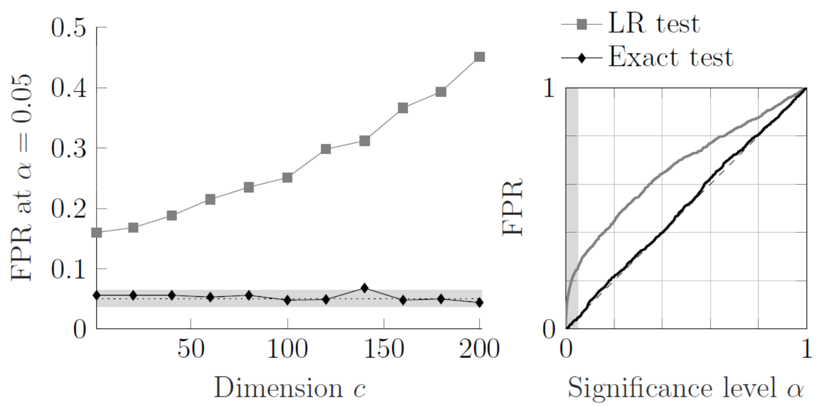
(b) IIR filter with univariate conditional.

- Simulation:
  - X, Y are independent univariate processes ($k = l = 1$)
  - Conditional W is univariate ($c = 1$)
  - Variable filter order
- Slight increase in LR test due to conditional not being taken into account

Numerical simulations: CMI for two time series



(a) FIR filter with c-variate conditional.



(b) IIR filter with c-variate conditional.

- Simulation:
  - X, Y are independent univariate processes ($k = l = 1$)
  - $\boldsymbol{W}$ is a higher-order conditional ($c = [1,200]$)
  - Fixed 8th order filter
- Approx. linear increase in FPR of LR test due to conditional
  - We stopped it at > 40% FPR (c=200)

Sampling distribution of multivariate measures

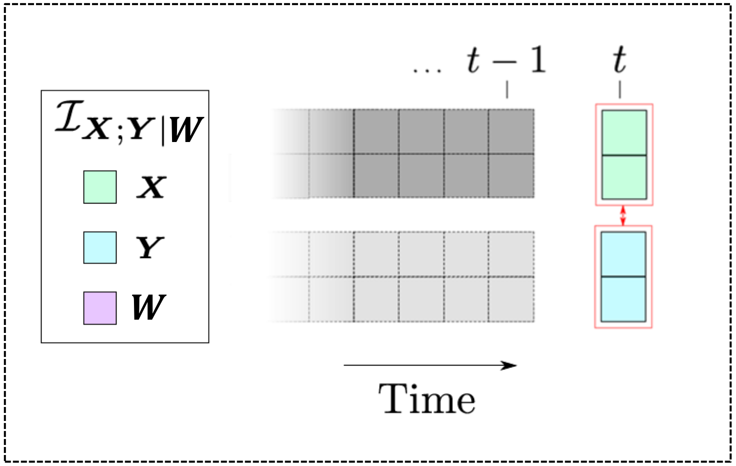
1. **Partial Correlation**
   1. Derive one-tailed test from Bartlett's formula
   2. Exact two-tailed test by squaring
2. **Conditional mutual information (between two time series)**
   1. Express as a squared partial correlation
   2. Exact test from two-tailed PC test
3. **Conditional mutual information (between multiple time series)**
   1. Use chain rule to decompose as a sum of CMI terms
   2. Exact test by summing each two-tailed PC test
4. **Granger causality** (between two/multiple time series)
   1. Equivalent to conditional mutual information (Barnett et al., 2009)
   2. Exact test from CMI (above)

(Conditional) mutual information

Between multivariate Gaussians

› Expressed as a log-likelihood of sum-of-squared residuals

$$\hat{\mathcal{I}}_{\boldsymbol{x};\boldsymbol{y}|\boldsymbol{w}} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{y}|\boldsymbol{w}}|}{|\boldsymbol{S}_{\boldsymbol{x}|\boldsymbol{w}}||\boldsymbol{S}_{\boldsymbol{y}|\boldsymbol{w}}|}\right)$$



› Tested against the chi-square distribution:

$$2T\,\hat{\mathcal{I}}_{\boldsymbol{x};\boldsymbol{y}|\boldsymbol{w}} \sim \chi^2(kl)$$

Exact mutual information test

Multiple time series

$$\boldsymbol{v}^{\{gh\}}_{\boldsymbol{xy}|\boldsymbol{w}} = \begin{bmatrix} \boldsymbol{x}_{1:g-1} \\ \boldsymbol{y}_{1:h-1} \\ \boldsymbol{w} \end{bmatrix}$$

› Decompose mutual information by the chain rule:

$$\hat{\mathcal{I}}_{\boldsymbol{x};\boldsymbol{y}|\boldsymbol{w}} = \sum_{g=1}^{k}\sum_{h=1}^{l} \hat{\mathcal{I}}_{x_g;y_h|\boldsymbol{v}^{\{gh\}}_{\boldsymbol{xy}|\boldsymbol{w}}}$$

› Gives expression as squared partial correlations:

$$2\hat{\mathcal{I}}_{\boldsymbol{x};\boldsymbol{y}|\boldsymbol{w}} = -\sum_{g=1}^{k}\sum_{h=1}^{l} \log\left(1 - r^2_{x_g y_h \cdot \boldsymbol{v}^{\{gh\}}_{\boldsymbol{xy}|\boldsymbol{w}}}\right)$$

› Which is $\mathcal{L}$-distributed (under the null):

$$2\hat{\mathcal{I}}_{\boldsymbol{x};\boldsymbol{y}|\boldsymbol{w}} \sim \mathcal{L}(\boldsymbol{N}_{\boldsymbol{xy}|\boldsymbol{w}})$$

Exact mutual information test

Multiple time series

› Which is $\mathcal{L}$-distributed (under the null):

$$2\,\hat{\mathcal{I}}_{\boldsymbol{x};\boldsymbol{y}|\boldsymbol{w}} \sim \mathcal{L}(\boldsymbol{N}_{\boldsymbol{x}\boldsymbol{y}|\boldsymbol{w}})$$

$$\boldsymbol{v}^{\{gh\}}_{\boldsymbol{x}\boldsymbol{y}|\boldsymbol{w}} = \begin{bmatrix} \boldsymbol{x}_{1:g-1} \\ \boldsymbol{y}_{1:h-1} \\ \boldsymbol{w} \end{bmatrix}$$

› (Each term could have a different ESS)

$$N^{\{gh\}}_{\boldsymbol{x}\boldsymbol{y}|\boldsymbol{w}} = \boxed{\eta\left(e^{\{gh\}}_{\boldsymbol{x}|\boldsymbol{v}}, e^{\{gh\}}_{\boldsymbol{y}|\boldsymbol{v}}\right)} - \boxed{\dim\left(\boldsymbol{v}^{\{gh\}}_{\boldsymbol{x}\boldsymbol{y}|\boldsymbol{w}}(t)\right)} - 2$$

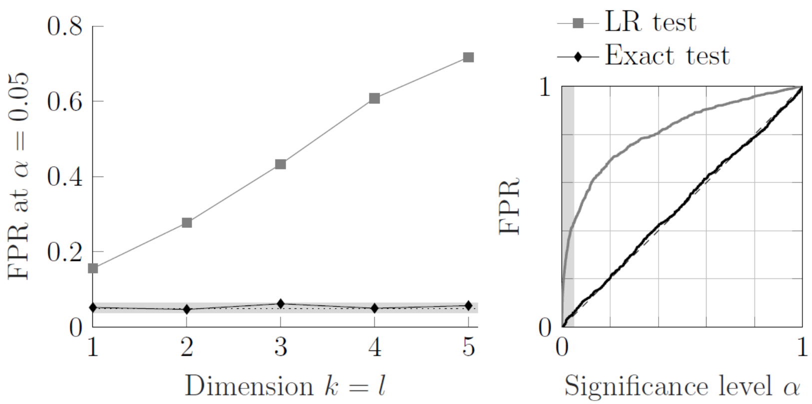
$$e^{\{gh\}}_{\boldsymbol{x}|\boldsymbol{v}} = x_g - \hat{x}_g\left(\boldsymbol{v}^{\{gh\}}_{\boldsymbol{x}\boldsymbol{y}|\boldsymbol{w}}\right)$$

$$e^{\{gh\}}_{\boldsymbol{y}|\boldsymbol{v}} = y_h - \hat{y}_h\left(\boldsymbol{v}^{\{gh\}}_{\boldsymbol{x}\boldsymbol{y}|\boldsymbol{w}}\right)$$

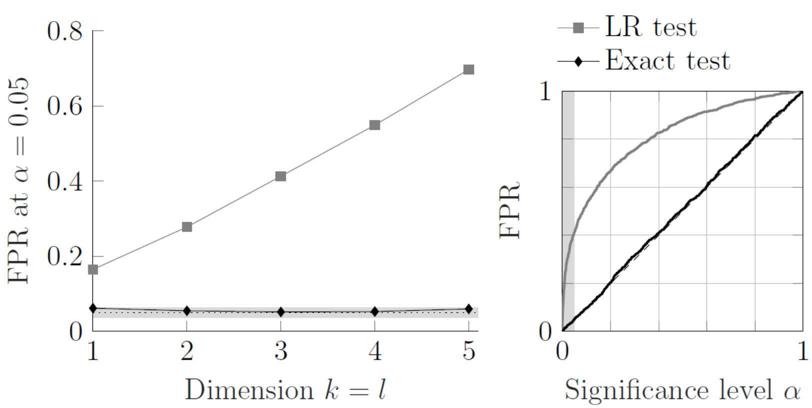
Effective sample size

Dimension of conditional

Numerical simulations: MI for multiple time series



(a) FIR filter.



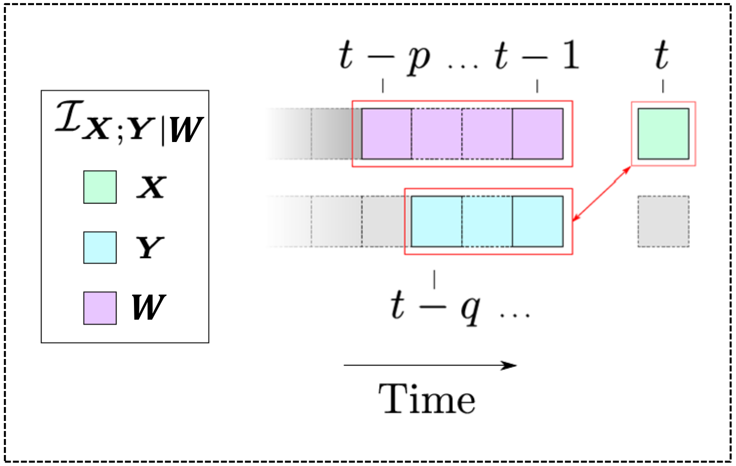
(b) IIR filter.

- Simulation:
  - X, Y are independent multivariate processes ($k = l \geq 1$)
  - No conditional process $\boldsymbol{W}$ ($c = 0$)
  - Fixed 8$^{\text{th}}$ order filter
- Approx. linear increase in FPR of LR test
  - Each term's Bartlett-correction and conditional not taken into account
  - Approaches 100% FPR

Sampling distribution of multivariate measures

1. **Partial Correlation**
    1. Derive one-tailed test from Bartlett's formula
    2. Exact two-tailed test by squaring
2. **Conditional mutual information (between two time series)**
    1. Express as a squared partial correlation
    2. Exact test from two-tailed PC test
3. **Conditional mutual information (between multiple time series)**
    1. Use chain rule to decompose as a sum of CMI terms
    2. Exact test by summing each two-tailed PC test
4. **Granger causality** (between two/multiple time series)
    1. Equivalent to conditional mutual information (Barnett et al., 2009)
    2. Exact test from CMI (above)

Granger causality



- Define relevant history of processes:
  - AIC, BIC, first minimum partial AC, etc.

$$\boldsymbol{X}^{(p)}(t) = \begin{bmatrix} \boldsymbol{X}(t-1) \\ \vdots \\ \boldsymbol{X}(t-p) \end{bmatrix}, \quad \boldsymbol{Y}^{(q)}(t) = \begin{bmatrix} \boldsymbol{Y}(t-1) \\ \vdots \\ \boldsymbol{Y}(t-q) \end{bmatrix}$$

- Expressed as a mutual information…

$$\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{W}}(p,q) = 2\,\mathcal{I}_{\boldsymbol{X};\boldsymbol{Y}^{(q)}|\boldsymbol{X}^{(p)}\boldsymbol{W}}$$

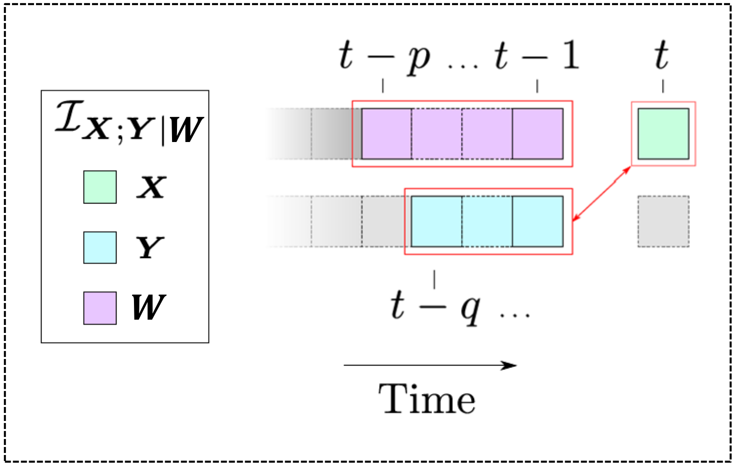
- …or equiv. as a log-ratio of the sum-of-squares

$$\hat{\mathcal{F}}_{\boldsymbol{y}\to\boldsymbol{x}|\boldsymbol{w}}(p,q) = \log\left(\frac{\left|\boldsymbol{S}_{\boldsymbol{x}|\boldsymbol{x}^{(p)}\boldsymbol{w}}\right|}{\left|\boldsymbol{S}_{\boldsymbol{x}|\boldsymbol{x}^{(p)}\boldsymbol{y}^{(q)}\boldsymbol{w}}\right|}\right)$$

Granger causality

› Asymptotic chi-square test:

$$T\,\hat{\mathcal{F}}_{\boldsymbol{y}\to\boldsymbol{x}|\boldsymbol{w}}(p,q) \sim \chi^2(klq)$$



› Exact test for **i.i.d. variables** with a univariate predictee/target:

$$\frac{T-(p+lq+c+1)}{lq}\left[\exp\left(\hat{\mathcal{F}}_{\boldsymbol{y}\to x|\boldsymbol{w}}(p,q)\right)-1\right]$$

$$\sim F(lq, T-(p+lq+c+1))$$

Exact Granger causality tests

Two time series

› Decomposes into q CMI terms via the chain rule:

$$\hat{\mathcal{F}}_{y \to x | \boldsymbol{w}}(p, q) = 2 \sum_{j=1}^{q} \hat{\mathcal{I}}_{x; y^j | \boldsymbol{v}^{\{j\}}_{y \to x | \boldsymbol{w}}}$$

$$\boldsymbol{v}^{\{j\}}_{y \to x} = \begin{bmatrix} \boldsymbol{x}^{(p)} \\ \boldsymbol{y}^{(j-1)} \\ \boldsymbol{w} \end{bmatrix}$$

› Gives the $\mathcal{L}$-distribution (under the null):

$$\hat{\mathcal{F}}_{y \to x | \boldsymbol{w}}(p, q) \sim \mathcal{L}(\boldsymbol{N}_{y \to x | \boldsymbol{w}})$$

Exact Granger causality tests

Two time series

$$\boldsymbol{v}^{\{j\}}_{y\to x} = \begin{bmatrix} \boldsymbol{x}^{(p)} \\ \boldsymbol{y}^{(j-1)} \\ \boldsymbol{w} \end{bmatrix}$$

› Decomposes into *q* CMI terms via the chain rule:

$$\hat{\mathcal{F}}_{y\to x|\boldsymbol{w}}(p,q) = 2\sum_{j=1}^{q} \hat{\mathcal{I}}_{x;y^j|\boldsymbol{v}^{\{j\}}_{y\to x|\boldsymbol{w}}}$$

› Gives the $\mathcal{L}$-distribution (under the null):

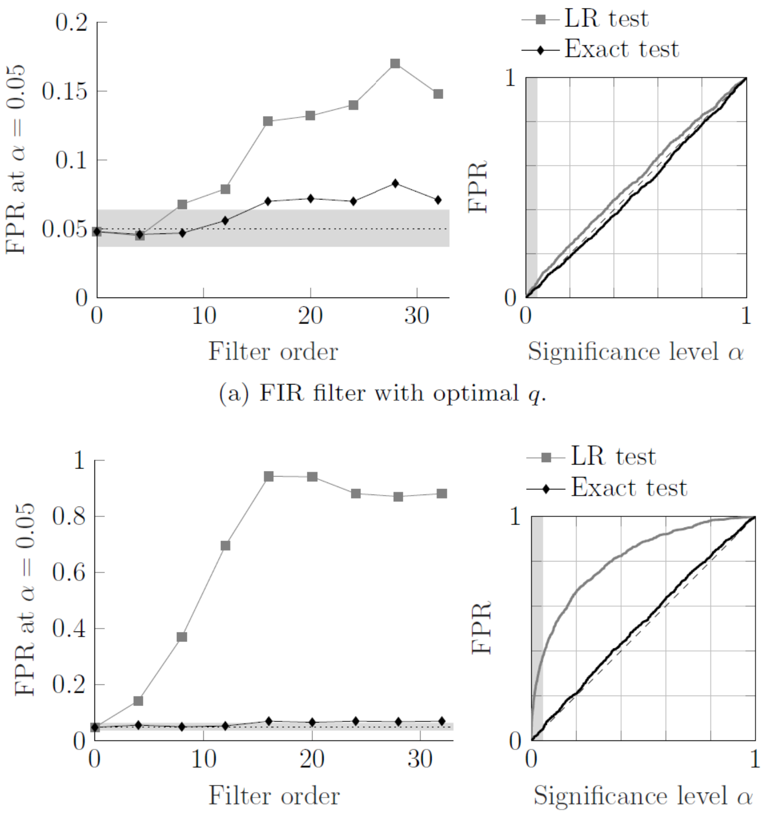
$$\hat{\mathcal{F}}_{y\to x|\boldsymbol{w}}(p,q) \sim \mathcal{L}(\boldsymbol{N}_{y\to x|\boldsymbol{w}})$$

› (Each term could have a different ESS)

$$N^{\{j\}}_{y\to x|\boldsymbol{w}} = \eta\left(e^{\{j\}}_{x|\boldsymbol{v}_{y\to x|\boldsymbol{w}}}, e^{\{j\}}_{y|\boldsymbol{v}_{y\to x|\boldsymbol{w}}}\right) - \dim\left(\boldsymbol{v}^{\{j\}}_{y\to x|\boldsymbol{w}}(t)\right) - 2$$

$$e^{\{j\}}_{x|\boldsymbol{v}_{y\to x|\boldsymbol{w}}} = x - \hat{x}\left(\boldsymbol{v}^{\{j\}}_{y\to x|\boldsymbol{w}}\right)$$
$$e^{\{j\}}_{y|\boldsymbol{v}_{y\to x|\boldsymbol{w}}} = y^j - \hat{y}^j\left(\boldsymbol{v}^{\{j\}}_{y\to x|\boldsymbol{w}}\right)$$

Numerical simulations: GC for two time series



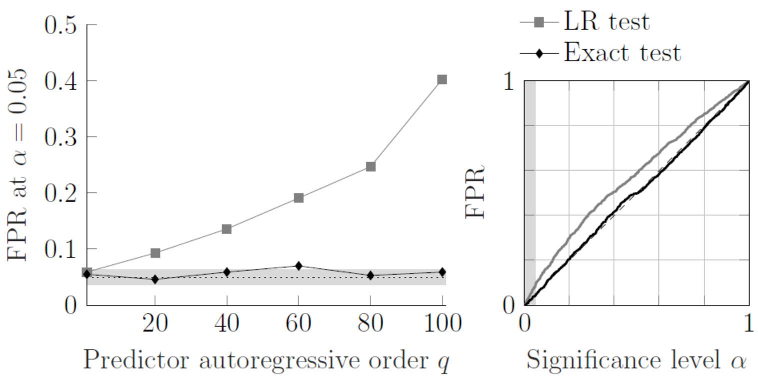
(a) FIR filter with optimal q.

(b) IIR filter with optimal q.

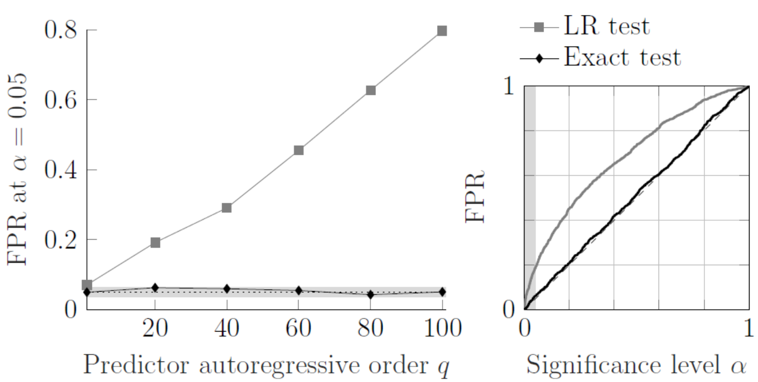
› Simulation:

- X, Y are independent univariate processes ($k = l = 1$)
- No conditional process $\boldsymbol{W}$ ($c = 0$)
- Variable filter order
- Optimal history lengths (p, q) chosen from partial ACF

Numerical simulations: GC for two time series



(a) FIR filter with varying q.



(b) IIR filter with varying q.

- Simulation:
  - X, Y are independent univariate processes ($k = l = 1$)
  - No conditional process $\boldsymbol{W}$ ($c = 0$)
  - Fixed 8th order filters
  - Variable predictor history length (q)
- LR test FPR increases approx. linearly with q due to no conditional
- Exact tests maintains correct FPR

THE UNIVERSITY OF SYDNEY

Exact Granger causality tests

Multiple time series

$$\boldsymbol{v}^{\{ghj\}}_{\boldsymbol{y}\to\boldsymbol{x}|\boldsymbol{w}} = \begin{bmatrix} \boldsymbol{x}^{(p)}_{1:g} \\ \boldsymbol{y}^{(q)}_{1:h-1} \\ \boldsymbol{y}^{(j-1)}_{h} \\ \boldsymbol{w} \end{bmatrix}$$

- Decomposes into *q x k x l* CMI terms

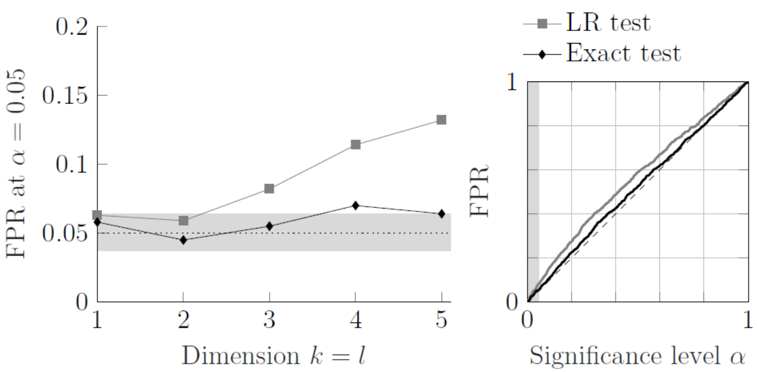
$$\hat{\mathcal{F}}_{\boldsymbol{x}\to\boldsymbol{y}|\boldsymbol{w}}(p,q) = 2\sum_{g=1}^{k}\sum_{h=1}^{l}\sum_{j=1}^{q}\hat{\mathcal{I}}_{x_g y_h^j | \boldsymbol{v}^{\{ghj\}}_{\boldsymbol{y}\to\boldsymbol{x}|\boldsymbol{w}}}$$

- Gives the $\mathcal{L}$-distribution (under the null):

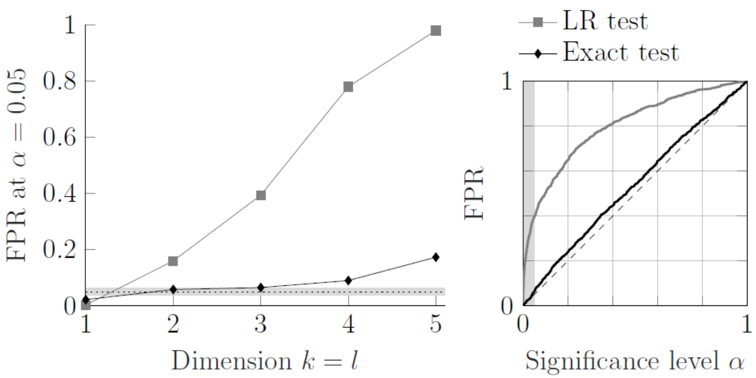
$$\hat{\mathcal{F}}_{\boldsymbol{y}\to\boldsymbol{x}|\boldsymbol{w}}(p,q) \sim \mathcal{L}(\boldsymbol{N}_{\boldsymbol{y}\to\boldsymbol{x}|\boldsymbol{w}})$$

- (Each term could have a different ESS)

Numerical simulations: GC for multiple time series



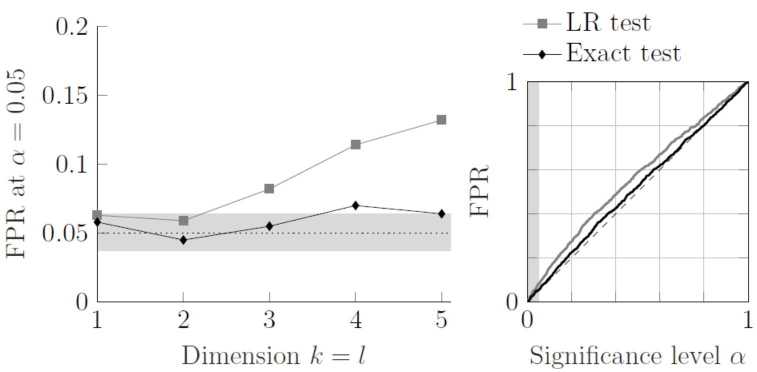
(a) FIR filter.



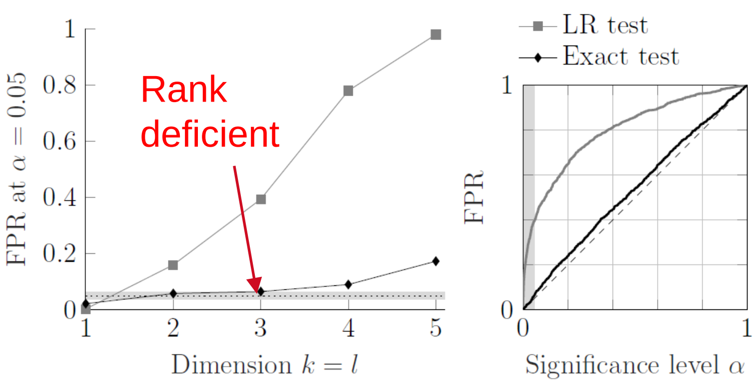
(b) IIR filter.

- Simulation:
  - $\boldsymbol{X}, \boldsymbol{Y}$ are independent multivariate processes ($k = l \geq 1$)
  - No conditional process $\boldsymbol{W}$ ($c = 0$)
  - Fixed 8th order filters
  - Optimal p but fixed $q = 1$
- LR test increases towards 100% FPR for IIR filter

Numerical simulations: GC for multiple time series



(a) FIR filter.



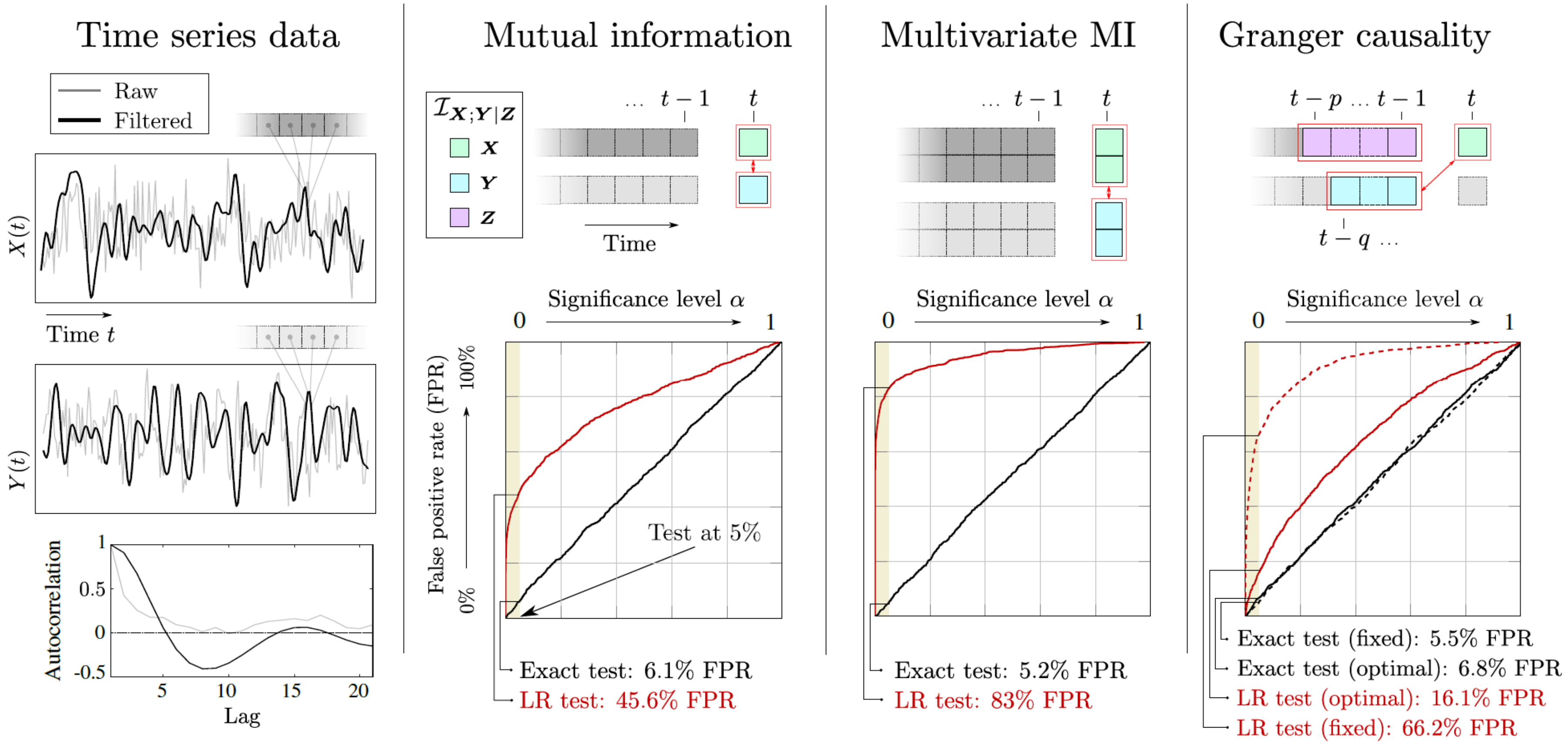
(b) IIR filter.

› Simulation:
  - $\boldsymbol{X}, \boldsymbol{Y}$ are independent multivariate processes ($k = l \geq 1$)
  - No conditional process $\boldsymbol{W}$ ($c = 0$)
  - Fixed 8th order filters
  - Optimal p but fixed $q = 1$

› LR test increases towards 100% FPR for IIR filter

› Regressions become rank deficient, so exact test begins to fail

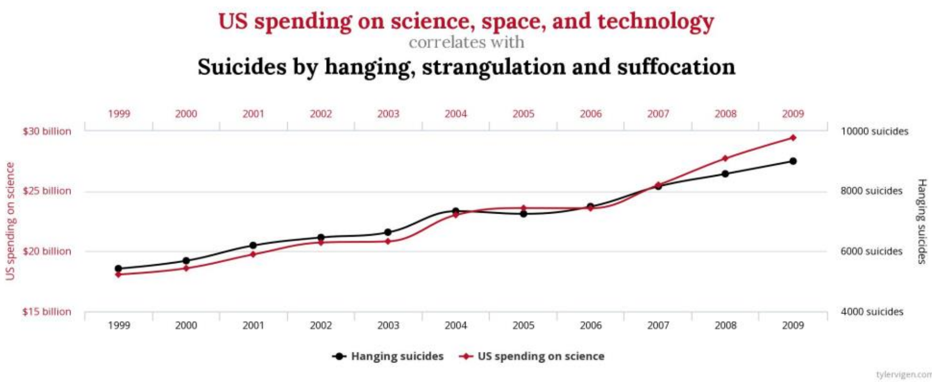
Case study: Human connectome project

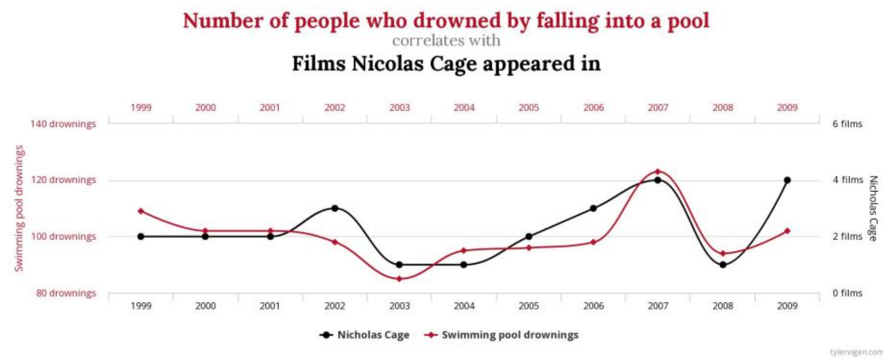


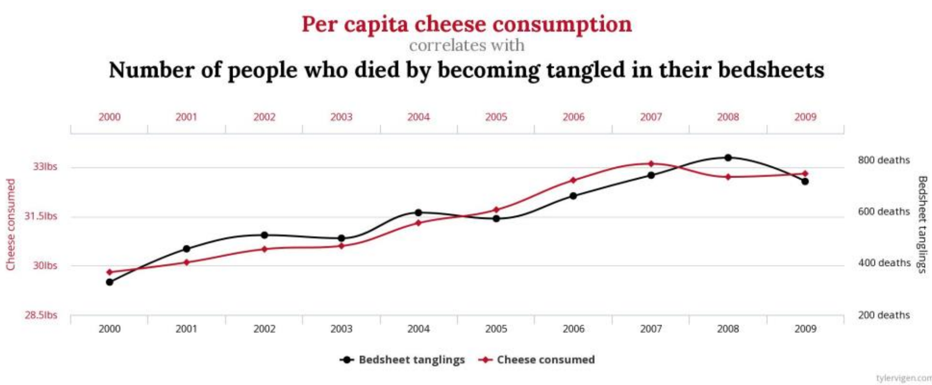
Take home messages

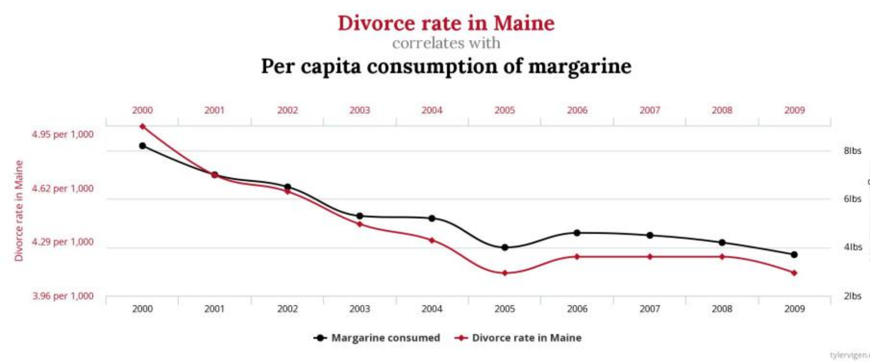
- Commonly-used linear-dependence measures exhibit bias for autocorrelated time series
- These measures can be represented as sums of squared partial correlations
- This representation allows us to derive their exact sampling distribution
- Before our work, these distributions were only valid asymptotically

Thank you!









Correlation image source: http://www.tylervigen.com/spurious-correlations