Dynamics of the COVID-19 Spread by Modified Graphical Network Analysis

Yigit Aydede (Saint Mary's Uni.)
& **Mutlu Yuksel** (Dalhousie Uni.)
Nova Scotia COVID-19 Research Coalition
Halifax, NS, Canada

This (WIP) version: 20 October 2021

Section 1

Motivation

What we aim

- Although the evidence unambiguously indicates successful mobility restrictions have the largest effect on curbing the pandemic (before vaccines), studies looking at the dynamics of these confinement policies are rare.
- Can we recover the short-term temporal dynamics in the relationship between the COVID-19 spread and mobility restrictions by using observed data with minimum assumptions (about data structure, model) for any location?
- PS: We will not predict case numbers or $R(t)$.

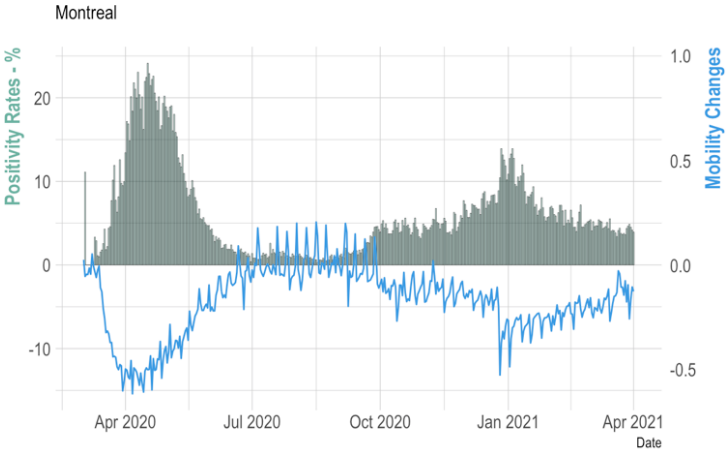
Main goal

- What we observe are the case numbers and positivity rates based on imperfect testing practices (random and selective) on symptomatic or even non-symptomatic people.
- We have imperfect proxies for mobility each reflecting different metrics of "movements" in a location.
- Due to the incubation period (estimated 1 to 21 days) and delays in testing, there are no observed data on the spread ($R(t)$-the average number of secondary cases of disease caused by a single infected individual over his/her infectious period)
- If we can develop a model that recovers the temporal relationship between NPIs and the spread from the **observed data**, we can better understand the dynamics of the relationship.

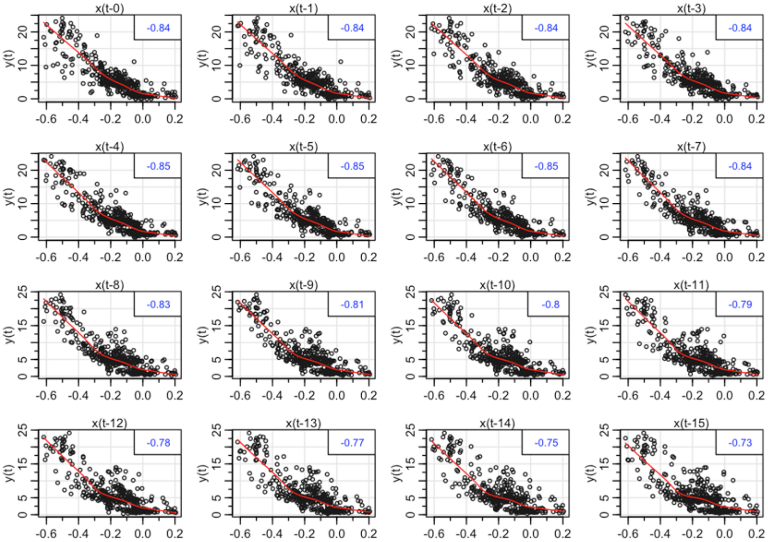
In this study

- Focusing on a specific location, we develop a method that demonstrates the number of days required to generate an intended effect on positivity rates following the mobility restriction policies implemented to curb the spread of the Covid-19 (based on the case numbers and degree of mobility restrictions).
- We use raw data with minimum restrictions.
- Our model incorporates methods from 3 different fields.
- PS: This method can be implemented to any two time-series signals (with a known direction of correlation) to recover dynamic correlations.

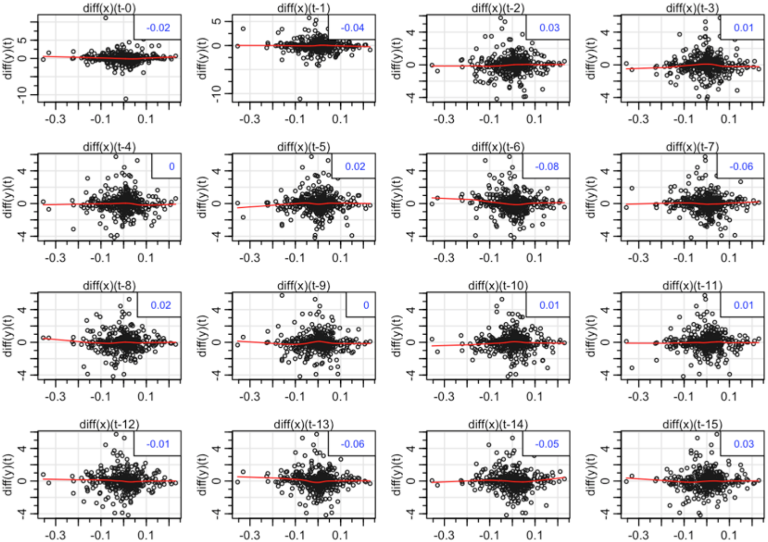
PR vs. Mobility



Cross-Correlations - Level



Cross-Correlations - (diff)



Why correlations have no meaning?

- Reverse causality: mobility reduction as a response to spikes in cases.
- Mobility shows its effect on PR dynamically (over time).
- They are zero-order cross-correlations
- There is no static relationship. For instance, 7 day-lag could be too short or too long in different windows.
- Contacts are not homogeneous across individuals and locations.

Review studies

ELSEVIER

Physics Reports
Volume 913, 23 May 2021, Pages 1-52

Non-pharmaceutical interventions during the COVID-19 pandemic: A review

Nicola Perra

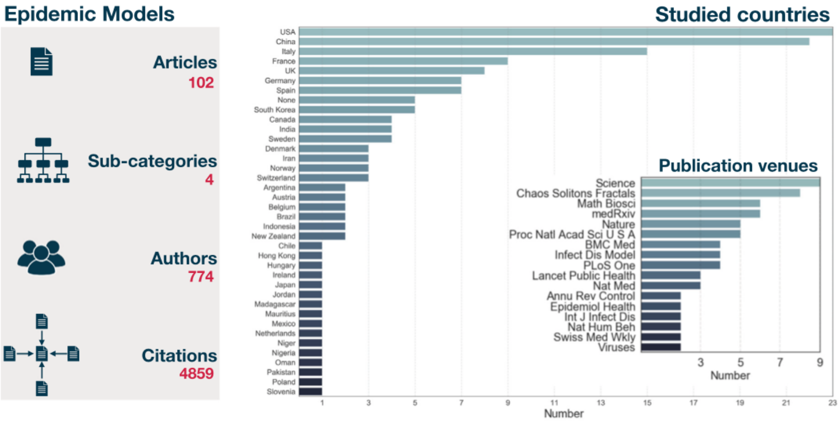
Networks and Urban Systems Centre, University of Greenwich, London, UK

Received 14 January 2021, Accepted 8 February 2021, Available online 13 February 2021.

Table 1
Number of authors, total number of citations, median number of citations, and top three papers for citation for each of the categories. Citations and authors names have been extracted via Semantic Scholar on December 19th, 2020.

Category	Authors	Total Citations	Median citations	Top three articles for citations
Epidemic models	774	4859	5	1st [22], 2nd [23], 3rd [24]
Surveys	751	775	1	1st [25], 2nd [26], 3rd [27]
Comments and/or perspectives	420	1049	4	1st [28], 2nd [29], 3rd [30]
Quantifying the effects of NPIs	405	2126	2	1st [31], 2nd [32], 3rd [33]
Reviews	131	192	3.5	1st [34], 2nd [35], 3rd [36]
Measuring NPIs with proxy data	88	159	2	1st [37], 2nd [38], 3rd [39]
Datasets	105	37	1.5	1st [40], 2nd [41], 3rd [42]

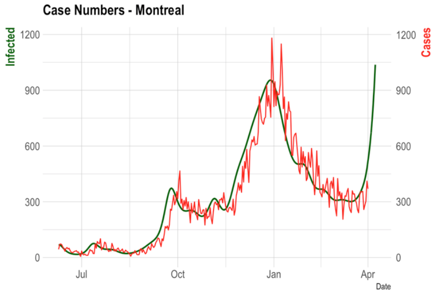
Epidemic Models

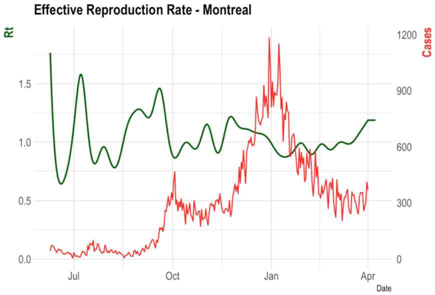


Data & Epi Models

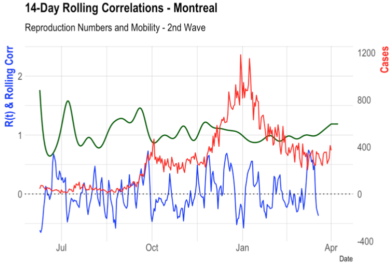
- Every study uses the same data as starting points: case/death numbers and mobility
- But, the response variable is now $R(t)$ - not observed but estimated by `EpiEstim` or `EpiNow2` based on SIR models from observed case numbers
- Estimated $R(t)$ can be defined as deviations from its base reproduction rate (R_0, which is also estimated) by the changes in mobility, $m(t)$, and other factors.
- Obtaining temporally accurate $R(t)$ estimates requires assumptions about lags from infection to observation.
- Sampling from a delay (gamma) distribution to impute individual times of infection from times of observation accounts for uncertainty
- But **blurs peaks and valleys** in the underlying incidence curve, which, in turn, compromises the ability to rapidly detect changes in $R(t)$ (Locatelli et al. 2021)

Montreal with `EpiNow2`





Naive way with $R(t)$




```
##
## Call:
## lm(formula = rtmont[, 1] ~ rtmont[, 2] + I(rtmont[, 2]^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44273 -0.09831 -0.01507  0.09672  0.72641
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.07799    0.01421  75.867   <2e-16 ***
## rtmont[, 2]       0.28283    0.13018   2.173   0.0306 *
## I(rtmont[, 2]^2) -0.28611    0.44760  -0.639   0.5232
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1738 on 294 degrees of freedom
## Multiple R-squared:  0.05512,    Adjusted R-squared:  0.04869
## F-statistic: 8.575 on 2 and 294 DF,  p-value: 0.0002402
```

With a smoothed mobility ($\text{loess}(degree = 2, span = 0.06)$), it's 0.498 and $Pr(> |t|) = 0.0538$

What we do

- TVFC with dynamically selected lags
- Partial correlations with $n > p$
- Partial correlation with $n < p$
  - GGM with Ridge - for delays
  - GGM with de-biased 2-stage Ridge - for correlations

Section 2

TVFC

Our data

- We use only observed data: Positivity Rate (cases/tests) and Mobility index (Facebook)
- Take them as two time-series signals and see if we can recover any meaningful relationship between them
- We use Montreal, as it is the most detailed COVID-19 Data (not publicly available)
- Later, we add Toronto, NYC, and Nova Scotia (confidential data)
- We also use SafeGraph data for Halifax at 6-digit postal codes (not shown in this presentation)

Now, we will go through one-by-one existing methods to find dynamic correlations, then methods' shortcomings, then fixing that using another method, and so on!

TVFC

NETWORK NEURO SCIENCE

Review

Questions and controversies in the study of time-varying functional connectivity in resting fMRI

Daniel J. Lurie[1,*], Daniel Kessler[2,*], Danielle S. Bassett[3,4,5,6], Richard F. Betzel[3], Michael Breakspear[7,8], Shella Keilholz[9], Aaron Kucyi[10], Raphaël Liégeois[11,12], Martin A. Lindquist[13], Anthony Randal McIntosh[14,15], Russell A. Poldrack[16], James M. Shine[17], William Hedley Thompson[16,18], Natalia Z. Bielczyk[19], Linda Douw[20], Dominik Kraft[21], Robyn L. Miller[22], Muthuraman Muthuraman[23], Lorenzo Pasquini[24], Adeel Razi[25,26,27], Diego Vidaurre[28], Hua Xie[29], and Vince D. Calhoun[22,30,31,*]

- Recently, **time-varying functional connectivity** (TVFC) has emerged as a major topic in the resting-state BOLD fMRI literature.
- TVFC uses running correlations between pairs of stochastic time series to identify their low-frequency evolution, which gives an idea about the functional organization of the brain
- Other fields, like Environmental Science, Behavioral Psychology, and Finance use rolling correlations as their main tool
- TVFC measures **simultaneous associations** between two series in sliding-windows

More on TVFC

$$r(\mathbf{X},\mathbf{Y},\tau) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau}\frac{\left(x_i - \overline{\mathbf{X}}\right)\left(y_{i+\tau} - \overline{\mathbf{Y}}\right)}{\mathrm{sd}(\mathbf{X})\,\mathrm{sd}(\mathbf{Y})}$$

- The sliding window technique is not new and has multiple parameters such as window function, length, and step size that must be set.
- But the appropriate settings remain unknown due to lack of "ground truth".
- The problem of "window-size" still remains as a main challenge in both methods:
  - **very long windows eventually measure static connectivity.**
  - **shorter windows can increase sensitivity for detecting short transition states but at the expense of decreasing the signal-to-noise ratio**

Modified TVFC

- **Ground truth**: the mobility changes must predict the events of infection measured by positivity rates, only if mobility changes occur before the events of PR.
- Estimate the association with dynamically selected delays
- So that the only one lag (i.e., the time difference in starting points of both series) maximizes the strength of their positive association.

Algortihm

Original Series			Lagged Series						
			Lag 1		lag 2		...	lag 21	
MOB	PR		MOB	PR	MOB	PR		MOB	PR
1	1		1	2	1	3	...	1	22
2	2		2	3	2	4	...	2	23
3	3		3	4	3	5	...	3	24
4	4		4	5	4	6	...	4	25
5	5		5	6	5	7	...	5	26
6	6		6	7	6	8	...	6	27
7	7		7	8	7	9	...	7	28
8	8		8	9	8	10	...	8	29
9	9		9	10	9	11	...	9	30
10	10		10	11	10	12	...	10	31
⋮	⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮
389	389		388	388	387	387	...	368	368

Rolling Windows	Correlation Matrix (windows x lags)			
	cor1	cor2	...	cor21
Window 1	Red	Red	...	Red
Window 2	Blue	Blue	...	Blue
Window 3	Green	Green	...	Green
Window 4	⋮	⋮	⋮	⋮
Window 5	⋮	⋮	⋮	⋮
Window 6	⋮	⋮	⋮	⋮
⋮ ⋮	⋮	⋮	⋮	⋮

Summary of each row				
Max	Lag at Max	Med	Q25	Q75
Red	Red	Red	Red	Red
Blue	Blue	Blue	Blue	Blue
Green	Green	Green	Green	Green
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

Windowed Crossed-Lagged Correlations

Journal Information
Journal TOC

Search APA PsycNet

APA PsycArticles: **Journal Article**

Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series.

© Request Permissions

Boker, S. M., Rotondo, J. L., Xu, M., & King, K. (2002). Windowed cross-correlation and

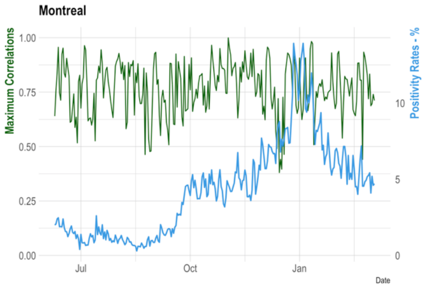
Published: 23 June 2015

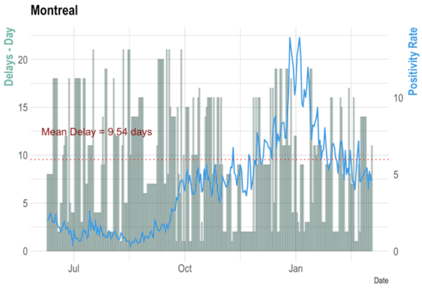
Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models

Roger T. Dean & William T. M. Dunsmuir

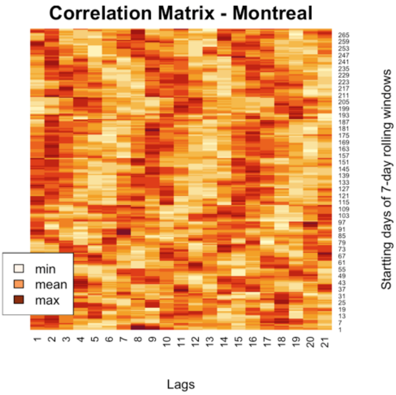
Behavior Research Methods **48**, 783–802 (2016) | Cite this article

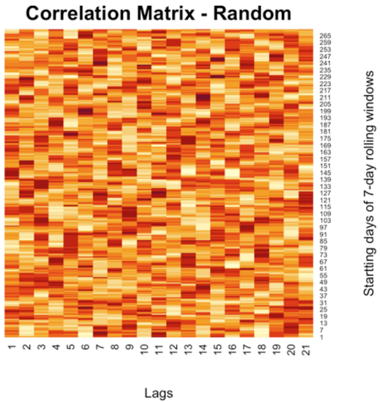
Maximums and Delays





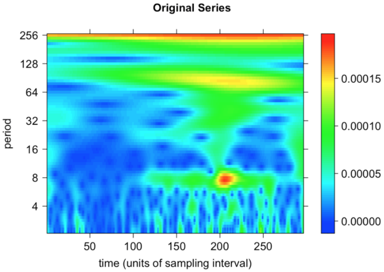
Heatmaps - Montreal Realized/Surrogate

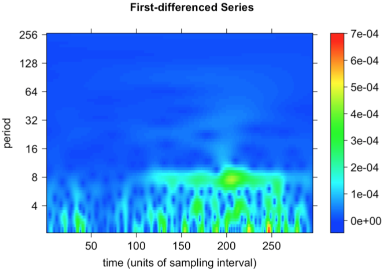




Why 7-day windows?

Wavelets allow us to study localized periodic behavior. In particular, we look for regions of high-power in the frequency-time plot.





The most suggested method is to keep the window length not shorter than the largest wavelength present in both series. This length is about 7 days in both series.

Shortcomings

- Correlations are not partial: intermediate lags are not controlled
- Even with a well-grounded epidemiological “truth” and with de-trended $\mathbf{I}(0)$ series, we need to know:
    - Whether the genuine association between two series is distinguishable from lagged synchrony that would occur by chance.
    - Whether correlations are out of 95% CI

Section 3

P.Correlations ($n > p$)

Background

- Most studies look at the **synchronous temporal correlations** among regions of interest (bivariate or multivariate). See Brain Imaging Methods.
- When it's bivariate and synchronous, zero-order correlations with sliding time-window based analysis are just fine
- When it's multivariate, $n > p$ is required for non-singular covariance matrix to obtain partial correlations.
- When $n << p$, a regularized inverse covariance (precision) matrix is needed
- Regularization leads to a network analysis that identifies the set of substantial connections (edges) between variables (nodes) and eliminates others
- Mostly used in genomics, finance, psychology, neuroscience to identify the "edges".
- With a proper visualization of the network, it's called **Gaussian Graphical Method**, if MVN.

Delay-coordinate embedding

Original Series		Lagged Series: Lag 1		...	lag 5		...	lag 21	
MOB	PR	MOB	PR		MOB	PR		MOB	PR
1	1	1	2		1	6	...	1	22
2	2	2	3		2	7	...	2	23
3	3	3	4		3	8	...	3	24
4	4	4	5		4	9	...	4	25
5	5	5	6		5	10	...	5	26
6	6	6	7		6	11	...	6	27
7	7	7	8		7	12	...	7	28
8	8	8	9		8	13	...	8	29
9	9	9	10		9	14	...	9	30
10	10	10	11		10	15	...	10	31
⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮
389	389	388	388		387	387	...	368	368

Rolling Windows	Dimension for each Partial Correlation Matrix (n x p): cor1	...	...	cor5	...	...	...	cor21
1	7x3	...	...	7x11	...	...	...	7x42
2	7x3	...	...	7x11	...	...	...	7x42
3	7x3	...	...	7x11	...	...	...	7x42
⋮	⋮		⋮	⋮	⋮	⋮	⋮	⋮

Embedded Data Matrix for Window 1 and Lag 5

Zero-order Window 1: PR5	MOB	Control Variables for Fully Partial Correlation for Lag5: PR	PR1	PR2	PR3	PR4	MOB1	MOB2	MOB3	MOB4
6	1	1	2	3	4	5	2	3	4	5
7	2	2	3	4	5	6	3	4	5	6
8	3	3	4	5	6	7	4	5	6	7
9	4	4	5	6	7	8	5	6	7	8
10	5	5	6	7	8	9	6	7	8	9
11	6	6	7	8	9	10	7	8	9	10
12	7	7	8	9	10	11	8	9	10	11

m-dimensional **reconstruction-space** vectors

$$\vec{R}(t) = [y(t), y(t-\tau), y(t-2\tau), \dots, y(t-(m-1)\tau)]$$

The standard strategy for state-space reconstruction is delay-coordinate embedding, where a series of past values of a single scalar measurement y from a dynamical system are used to form a vector that defines a point in a new space.

Possible solutions

- **Low-dimension:** $n > p$
  - Increase n and reduce p so that $n > p$ in each sliding window's data matrix
- **High-dimension with regularization:** $n << p$
  - Moore-Penrose Inverse
  - Graphical Lasso (`glasso`) or Thresholding with Ridge
  - Other methods: SIS, LPC etc.
- **Nonparametric predictive algorithms**
  - Identify which lag maximizes the predictive accuracy among the ones that lagged `mob` is in the top 3 most important predictors

Application with $n > p$

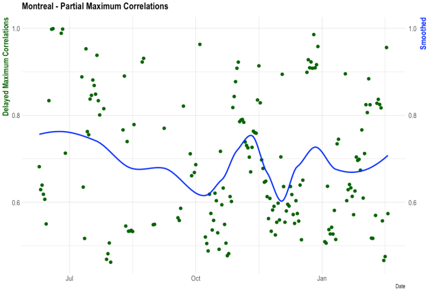
To avoid singular Σ, $w = 21$ and $Lags = \{1:18\}$ with only intermediate lags of `mob`.

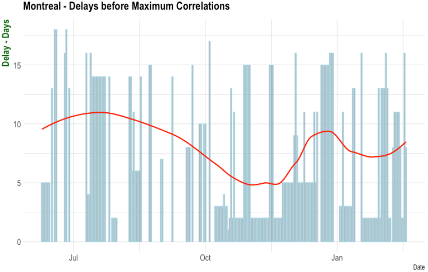
```
> round(cop[1:12,], 4)
         [,1]    [,2]   [,3]    [,4]   [,5]    [,6]    [,7]    [,8]    [,9]   [,10]  [,11]   [,12]  [,13]   [,14]   [,15]   [,16]   [,17]   [,18]
 [1,]  0.0430  0.1854 0.1172 -0.1666 0.6818 -0.4445  0.0214  0.4449 -0.4822  0.0797 0.4731  0.0136 0.3329 -0.0781  0.0386 -0.5607 -0.3274  0.9388
 [2,]  0.0273  0.1950 0.1168 -0.2464 0.6290 -0.4358  0.1839  0.5799 -0.5949  0.3408 0.3759 -0.0779 0.3874 -0.0360 -0.0634 -0.1150 -0.0300 -0.7211
 [3,]  0.0263  0.2051 0.0981 -0.2429 0.6394 -0.4093  0.1924  0.5505 -0.5190  0.3557 0.3264 -0.3067 0.5666  0.0516 -0.4090 -0.1481  0.3796 -0.9987
 [4,]  0.0348  0.1682 0.1054 -0.2408 0.6183 -0.4147  0.1920  0.4818 -0.5083  0.4405 0.0977 -0.2721 0.6181 -0.1400 -0.4769 -0.0468 -0.0991 -0.8400
 [5,]  0.0170  0.1738 0.0509 -0.2064 0.6068 -0.3197  0.1342  0.4884 -0.5749  0.0388 0.2190 -0.2815 0.6489 -0.1136 -0.7264  0.6133 -0.5066  0.7463
 [6,]  0.0484  0.1700 0.1972 -0.0978 0.5502 -0.2148  0.0674  0.4874 -0.2066 -0.0603 0.2530 -0.2540 0.6622 -0.3245 -0.8418  0.4240 -0.2177  0.8372
 [7,]  0.0309  0.0427 0.0953 -0.0404 0.4489 -0.1299 -0.0415 -0.1352  0.2047 -0.1861 0.3179 -0.2601 0.5706 -0.3920 -0.9294  0.5881 -0.5030  0.8912
 [8,] -0.1146 -0.0452 0.1321 -0.0871 0.5070 -0.1706 -0.2951  0.0868  0.1136 -0.0802 0.2956 -0.3855 0.8337 -0.5756 -0.1082  0.2359 -0.6322  0.8406
 [9,] -0.1304 -0.0368 0.1329 -0.0995 0.4681 -0.1483 -0.2072  0.0788  0.1398 -0.0832 0.2788 -0.5542 0.6826 -0.1661 -0.1850  0.4685 -0.4030  0.6218
[10,] -0.1258 -0.0353 0.1168 -0.1375 0.3252 -0.0938 -0.2070  0.1128  0.1301 -0.1426 0.4233 -0.4950 0.5419 -0.1633 -0.1301  0.6538 -0.8876  0.9979
[11,] -0.1158 -0.0586 0.1199 -0.0651 0.2593 -0.0805 -0.2727  0.1297  0.2580 -0.4276 0.3228 -0.1465 0.2671 -0.0133  0.0292 -0.1378 -0.8538  0.9991
[12,] -0.1492 -0.1117 0.1103 -0.0352 0.2608 -0.0530 -0.2693  0.0580  0.3955 -0.3470 0.1624  0.0088 0.0831 -0.0965  0.7748 -0.7896  0.1413  0.6621
>
```

We identified partial correlations with $p < 0.05$

```
> round(copp[1:12,], 4)
      [,1] [,2] [,3] [,4]   [,5] [,6] [,7]   [,8]    [,9] [,10] [,11] [,12]  [,13] [,14]   [,15] [,16] [,17]   [,18]
 [1,]    0    0    0    0 0.6818    0    0 0.0000  0.0000     0     0     0 0.0000     0  0.0000     0     0  0.0000
 [2,]    0    0    0    0 0.6290    0    0 0.5799 -0.5949     0     0     0 0.0000     0  0.0000     0     0  0.0000
 [3,]    0    0    0    0 0.6394    0    0 0.0000  0.0000     0     0     0 0.0000     0  0.0000     0     0 -0.9987
 [4,]    0    0    0    0 0.6183    0    0 0.0000  0.0000     0     0     0 0.0000     0  0.0000     0     0  0.0000
 [5,]    0    0    0    0 0.6068    0    0 0.0000  0.0000     0     0     0 0.0000     0  0.0000     0     0  0.0000
 [6,]    0    0    0    0 0.5502    0    0 0.0000  0.0000     0     0     0 0.0000     0 -0.8418     0     0  0.0000
 [7,]    0    0    0    0 0.0000    0    0 0.0000  0.0000     0     0     0 0.0000     0 -0.9294     0     0  0.0000
 [8,]    0    0    0    0 0.5070    0    0 0.0000  0.0000     0     0     0 0.8337     0  0.0000     0     0  0.0000
 [9,]    0    0    0    0 0.0000    0    0 0.0000  0.0000     0     0     0 0.0000     0  0.0000     0     0  0.0000
[10,]    0    0    0    0 0.0000    0    0 0.0000  0.0000     0     0     0 0.0000     0  0.0000     0     0  0.9979
[11,]    0    0    0    0 0.0000    0    0 0.0000  0.0000     0     0     0 0.0000     0  0.0000     0     0  0.9991
[12,]    0    0    0    0 0.0000    0    0 0.0000  0.0000     0     0     0 0.0000     0  0.0000     0     0  0.0000
>
```

Results


Montreal - Partial Maximum Correlations
Delayed Maximum Correlations
Smoothed
1.0
0.8
0.6
Jul
Oct
Jan
Date


Montreal - Delays before Maximum Correlations
Delay - Days
15
10
5
0
Jul
Oct
Jan
Date

Robustness check: simulation with surrogates

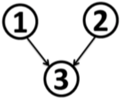
- Search for the maximum could find a random match by coincidence
- 1000 Surrogates of `PR` and `mob`.
- In each of 255 windows, about 84% of 1000 partial correlations are 0.
- Hence, about 84% of the maximum correlations (from each of the 1000 simulated correlations matrix) are zero.
- This ratio is 13% in our actual findings

Section 4

Regularization ($n < p$)

Background

- p doesn't include intermediate PRs
- We can use regularization to identify "significant" partial correlations
- **Berkson's paradox** could be an issue



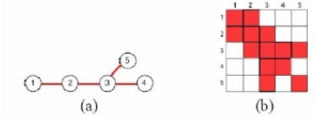
$$\begin{cases} x_1(t) = u_1(t) + noise \\ x_2(t) = u_2(t) + noise \\ x_3(t) = 0.5x_1(t) + 0.5x_2(t) + noise \end{cases}$$

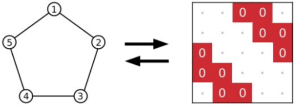
```
> x1 <- rnorm(100)
> x2 <- rnorm(100)
> x3 <- 0.5*x1+0.5*x2 + rnorm(100)
> mat <- cbind(x1, x2, x3)
> S <- cov(mat)
> -cov2cor(solve(S))
           x1         x2         x3
x1 -1.0000000 -0.4076111  0.6085341
x2 -0.4076111 -1.0000000  0.5647095
x3  0.6085341  0.5647095 -1.0000000
>
```

Although there is no link between Node 1 (PR) and Node 2 (mob), the partial correlation between these two nodes could be high and significant - **Regularization may correct the paradox and reduce the noise-to-signal ratio** (Nie et al. 2015).

Regularization for GGM

- $n > p$ or $n < p$, **we want to find a sparse graph** capturing the conditional dependence between the entries of a Gaussian random vector
- In GGM, the graph structure can be expressed only through its precision matrix, $\mathbf{\Omega}$.





Formally, let $\hat{\Omega}$ denote a generic estimate of the precision matrix and consider its transformation to a partial correlation matrix $\hat{\mathbf{P}}$. Then the following relations can be shown to hold for all pairs $\{Y_j, Y_i\} \in \mathcal{V}$ with $j \neq i$:

$$(\hat{\mathbf{P}})_{ji} = 0 \iff (\hat{\Omega})_{ji} = 0 \iff Y_j \perp Y_i \mid \mathcal{V} \setminus \{Y_j, Y_i\}$$

What do we want?

Original Series

MOB	PR
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
⋮	⋮
389	389

Lagged Series

Lag 1		...	lag 5		...	lag 21	
MOB	PR		MOB	PR		MOB	PR
1	2		1	6	...	1	22
2	3		2	7	...	2	23
3	4		3	8	...	3	24
4	5		4	9	...	4	25
5	6		5	10	...	5	26
6	7		6	11	...	6	27
7	8		7	12	...	7	28
8	9		8	13	...	8	29
9	10		9	14	...	9	30
10	11		10	15	...	10	31
⋮	⋮		⋮	⋮	⋮	⋮	⋮
388	388		387	387	...	368	368

Sparsified P. Correlation Matrix for Window 1 & Lag 5

	PR5	MOB	PR0	PR1	PR2	PR3	PR4	MOB1	MOB2	MOB3	MOB4
PR5	1	?									
MOB		1									
PR0			1								
PR1				1							
PR2					1						
PR3						1					
PR4							1				
MOB1								1			
MOB2									1		
MOB3										1	
MOB4											1

Embedded Data Matrix for Window 1 and Lag 5

Zero-order Window 1		Control Variables for Fully Partial Correlation for Lag5								
PR5	MOB	PR0	PR1	PR2	PR3	PR4	MOB1	MOB2	MOB3	MOB4
6	1	1	2	3	4	5	2	3	4	5
7	2	2	3	4	5	6	3	4	5	6
8	3	3	4	5	6	7	4	5	6	7
9	4	4	5	6	7	8	5	6	7	8
10	5	5	6	7	8	9	6	7	8	9
11	6	6	7	8	9	10	7	8	9	10
12	7	7	8	9	10	11	8	9	10	11

Correlations Matrix (windows x lags)

Windows	cor1	cor2	...	cor5	...	cor21
1	0.00	0.05	...	?	...	0.51
2	⋮	⋮	⋮	⋮	⋮	⋮
3	⋮	⋮	⋮	⋮	⋮	⋮
4	⋮	⋮	⋮	⋮	⋮	⋮
5	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
368	⋮	⋮	⋮	⋮	⋮	⋮

MLE Solution to Ω

- The multivariate Gaussian distribution of a random vector $X \in \mathbf{R}^p$ is commonly expressed in terms of the parameters μ and Σ, where μ is an $p \times 1$ vector and Σ is an $p \times p$, a nonsingular symmetric covariance matrix.
- The multivariate normal distribution:

$$f(X \mid \mu, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\},$$

with mean 0 and covariance Σ, the likelihood function:

$$\ell(\boldsymbol{\Omega}; \boldsymbol{S}) = \ln|\boldsymbol{\Omega}| - \operatorname{tr}(\boldsymbol{S}\boldsymbol{\Omega})$$

where $S = \widehat{\Sigma}$ and $\Omega = S^{-1}$ (**precision matrix**), for which we seek.

Regularization - $n < p$

- For $n < p$ the empirical estimate of the covariance matrix becomes singular
- A common workaround is the addition of a penalty (ℓ_1 norm—the sum of the absolute values of the elements of Σ^{-1}) to the log-likelihood.

$$\ell(\boldsymbol{\Omega}; \boldsymbol{S}) = \ln|\boldsymbol{\Omega}| - \mathrm{tr}(\boldsymbol{S}\boldsymbol{\Omega}) - \lambda\|\boldsymbol{\Omega}\|_1$$

- This **Graphical Lasso** estimate of $\boldsymbol{\Omega}$ provide **a sparse solution**!
- But, a more accurate representations of the high-dimensional precision matrix would be an asset: **Ridge**

Regularization - Ridge

With the ℓ_2 penalty, the ridge estimation solves the following:

$$\ell(\boldsymbol{\Omega}; \boldsymbol{S}) = \ln|\boldsymbol{\Omega}| - \mathrm{tr}(\boldsymbol{S}\boldsymbol{\Omega}) - \frac{\lambda}{2}\|\boldsymbol{\Omega} - \boldsymbol{T}\|_2^2$$

Assume for now the target matrix is an all-zero matrix:

$$\log(|\boldsymbol{\Omega}|) - \mathrm{tr}(\mathbf{S}\boldsymbol{\Omega}) - \frac{1}{2}\lambda_2 \,\mathrm{tr}\left(\boldsymbol{\Omega}\boldsymbol{\Omega}^{\mathrm{T}}\right)$$

Its derivative w.r.t. the precision matrix yields the estimating equation:

$$\boldsymbol{\Omega}^{-1} - \mathbf{S} - \lambda_2\boldsymbol{\Omega} = \mathbf{0}_{p\times p}$$

Ridge

Matrix algebra then yields:

$$\widehat{\boldsymbol{\Sigma}}\left(\lambda_2\right) = \frac{1}{2}\mathbf{S} + \left(\lambda_2 \mathbf{I}_{p\times p} + \frac{1}{4}\mathbf{S}^2\right)^{1/2}$$

The derived ridge covariance estimator is positive definite, ie it's symmetric and all its eigenvalues are positive.

- For $\lambda_2 = 0$, we obtain $\widehat{\boldsymbol{\Sigma}}(0) = \mathbf{S}$.
- For large enough λ_2 : $\widehat{\boldsymbol{\Sigma}}\left(\lambda_2\right) \approx \lambda_2 \mathbf{I}_{p\times p}$

Why Ridge for Sparsity?

ELSEVIER
Computational Statistics & Data Analysis
Volume 103, November 2016, Pages 284-303

Ridge estimation of inverse covariance matrices from high-dimensional data

Wessel N. van Wieringen [a, b], Carel F.W. Peeters [a]

- The true (graphical) model need not be (extremely) sparse.
- We may prefer a regularization that shrinks the estimated elements of the precision matrix proportionally
- Wieringen & Peeters (2016) demonstrate that the alternative ridge estimators yield **more stable** networks vis-à-vis the graphical lasso, in particular **for more extreme p/n ratios**.
- They provide empirical evidence in the graphical modeling setting of what is tacitly known from regression (subset selection) problems: **ridge penalties coupled with post-hoc selection may outperform the lasso**.

Steps

- Ridge penalty shrinks the estimated elements of Ω, but cannot shoot them to zero.
- Hence, it requires a specific post-hoc **thresholding** for sparsity
- Steps:
  - Estimating the elements of Ω with the optimal penalty parameter λ^*
  - Thresholding with λ^*
  - Recovering partial coefficients from Ridge estimates
    - 2-Stage estimation
    - De-biasing
    - Re-estimations

Fist two steps gives us **the delays in the effects of** `mob`, the last step **helps us quantify the maximum effects of** `mob` **on** `PR` **in each day**

Step 1: Ridge Estimates of Partials with λ^*

In choosing λ^*: The ℓ_2-penalty does not automatically induce sparsity in the estimate, it is natural to aim to maximize predictive power (e.g., cross-validation, AIC), instead of model selection consistency (e.g., BIC, EBIC).

The K-fold CV score for a generic regularized estimate $\hat{\Omega}(\lambda)$ based on the generic fixed penalty λ can be given as:

$$\varphi^K(\lambda) = \sum_{k=1}^{K} n_k \left\{ -\ln \left| \hat{\mathbf{\Omega}}(\lambda)_{-k} \right| + \mathrm{tr} \left[\hat{\mathbf{\Omega}}(\lambda)_{-k} \mathbf{S}_k \right] \right\}$$

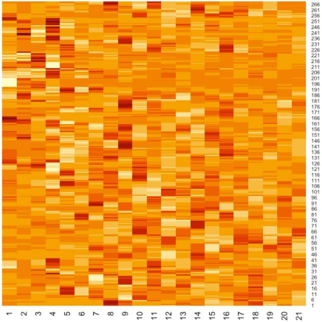
where n_k is the size of subset k, for $k = 1, \ldots, K$ disjoint subsets. Further, $\mathbf{S}_k$ denotes the sample covariance matrix based on subset k, while $\hat{\mathbf{\Omega}}(\lambda)_{-k}$ denotes the estimated regularized precision matrix on all samples not in k. Highest predictive accuracy can be obtained by choosing $n_k = 1$, such that $K = n$ a.k.a LOOCV.

Step 1: Ridge Estimates of Partials with λ^*

```
> round(copR[1:12, ], 4)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]
[1,]	0.0005	0.0006	0.0007	-0.0017	-0.0011	0.0000	0.0004	0.0286	0.0090	-0.0033	-0.0007
[2,]	0.0002	0.0001	0.0001	-0.0017	0.0000	-0.0053	-0.0047	0.0253	0.0083	-0.0019	-0.0006
[3,]	0.0003	0.0003	0.0003	-0.0042	0.0000	0.0000	-0.0030	0.0255	0.0115	-0.0038	-0.0030
[4,]	0.0002	0.0008	0.0008	-0.0048	0.0000	0.0000	-0.0027	0.0250	0.0134	-0.0016	-0.0006
[5,]	0.0007	0.0001	-0.0008	-0.0020	0.0000	0.0000	-0.0039	0.0233	0.0010	-0.0172	0.0140
[6,]	0.0048	0.0005	-0.0032	-0.0021	0.0000	0.0000	0.0035	0.0030	-0.0326	0.0317	0.0097
[7,]	0.0052	0.0032	-0.0037	0.0000	0.0000	-0.0048	0.0142	0.0274	-0.0646	0.0325	0.0083
[8,]	0.0045	0.0025	-0.0081	0.0000	0.0000	-0.0073	0.0392	-0.0113	-0.0656	0.0344	0.0124
[9,]	0.0062	0.0054	-0.0062	0.0000	0.0000	0.0000	0.0277	-0.0063	-0.0664	0.0407	0.0059
[10,]	0.0069	0.0007	-0.0099	0.0000	0.0000	-0.0800	0.0293	-0.0098	-0.0517	0.0255	0.0125
[11,]	-0.0119	0.0430	-0.0212	-0.0343	0.1345	-0.0799	0.0296	-0.0239	-0.0380	0.0210	0.0254
[12,]	-0.0126	0.0155	-0.0368	-0.0309	0.1246	-0.0855	0.0363	-0.0262	-0.0369	0.0469	0.0011

	[,12]	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]
[1,]	-0.0072	-0.0038	0.0145	-0.0125	-0.0118	0.0063	0.0142	-0.0118	0.0121	-0.0120
[2,]	0.0012	0.0076	0.0008	-0.0114	-0.0115	0.0045	0.0144	-0.0123	0.0044	-0.0090
[3,]	-0.0008	0.0074	0.0017	-0.0128	-0.0084	0.0023	0.0155	-0.0086	-0.0010	-0.0079
[4,]	-0.0018	0.0067	0.0026	-0.0149	-0.0070	0.0013	0.0130	-0.0065	-0.0012	-0.0084
[5,]	-0.0007	0.0054	0.0058	-0.0169	-0.0047	0.0143	0.0003	0.0086	0.0004	-0.0077
[6,]	0.0029	-0.0049	0.0162	-0.0175	0.0024	0.0256	0.0199	0.0030	-0.0021	-0.0233
[7,]	0.0075	-0.0093	0.0160	-0.0222	0.0024	0.0127	0.0224	0.0043	0.0052	-0.0251
[8,]	0.0006	-0.0072	-0.0029	-0.0192	-0.0219	0.0153	0.0245	0.0142	0.0021	-0.0512
[9,]	0.0035	-0.0060	-0.0161	-0.0271	-0.0187	0.0171	0.0339	0.0139	-0.0155	-0.0415
[10,]	0.0010	-0.0296	-0.0326	-0.0198	-0.0140	0.0376	0.0319	-0.0292	0.0048	-0.0436
[11,]	0.0104	0.0013	-0.0394	-0.0253	-0.0347	0.0407	0.0720	-0.0430	0.0043	-0.0507
[12,]	0.0401	0.0031	-0.0382	-0.0244	-0.0306	0.0505	0.0614	-0.0452	-0.0031	-0.0393



Step 2 - Thresholding

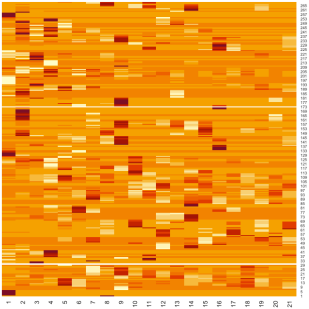
- As the Ridge estimators will not generally produce sparse estimates, they will need to rely on an additional procedure for support determination.
- The suggested method is called as the local false discovery rate (lFDR) procedure (Efron et al. 2001, Efron 2010, Schäfer and Strimmer 2005).

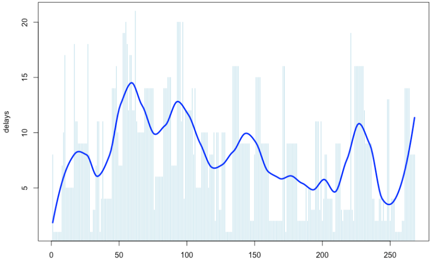
$$P\left(Y_j(no-edge)Y_{j'}\middle|\left[\hat{\mathbf{P}}(\lambda)\right]_{jj'}\right)=\frac{\hat{\eta}_0 f_0\left\{\left[\hat{\mathbf{P}}(\lambda)\right]_{jj'};\hat{\kappa}\right\}}{\hat{\eta}_0 f_0\left\{\left[\hat{\mathbf{P}}(\lambda)\right]_{jj'};\hat{\kappa}\right\}+(1-\hat{\eta}_0)\hat{f}_{\mathcal{E}}\left\{\left[\hat{\mathbf{P}}(\lambda)\right]_{jj'}\right\}}$$

which gives the empirical posterior probability that the edge between Y_j and $Y_{j'}$ is null given $\left[\hat{\mathbf{P}}\left(\lambda\right)\right]_{jj'}$.

Sparsified Ridge Estimates

```
> round(copRs[1:12, ], 4)
         [,1]   [,2]    [,3]    [,4]   [,5]    [,6]   [,7]   [,8]    [,9]  [,10] [,11]  [,12] [,13]   [,14] [,15]   [,16]  [,17]  [,18]   [,19]   [,20]   [,21]
 [1,]  0.0000 0.0000  0.0000 -0.0017 0.0000  0.0000 0.0000 0.0286  0.0000 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0000  0.0000  0.0000  0.0000
 [2,]  0.0002 0.0000  0.0000  0.0000 0.0000  0.0000 0.0000 0.0000  0.0000 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0000  0.0000  0.0000  0.0000
 [3,]  0.0003 0.0000  0.0000  0.0000 0.0000  0.0000 0.0000 0.0000  0.0000 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0000  0.0000  0.0000  0.0000
 [4,]  0.0002 0.0000  0.0000  0.0000 0.0000  0.0000 0.0000 0.0000  0.0000 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0000  0.0000  0.0000  0.0000
 [5,]  0.0007 0.0000  0.0000  0.0000 0.0000  0.0000 0.0000 0.0000  0.0000 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0000  0.0000  0.0000  0.0000
 [6,]  0.0048 0.0000  0.0000  0.0000 0.0000  0.0000 0.0000 0.0000  0.0000 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0000  0.0000  0.0000  0.0000
 [7,]  0.0052 0.0032  0.0000  0.0000 0.0000 -0.0048 0.0000 0.0000 -0.0646 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0000  0.0000  0.0000  0.0000
 [8,]  0.0045 0.0025  0.0000  0.0000 0.0000  0.0000 0.0392 0.0000 -0.0656 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0000  0.0000  0.0000 -0.0512
 [9,]  0.0062 0.0054  0.0000  0.0000 0.0000  0.0000 0.0000 0.0000 -0.0664 0.0407     0 0.0000     0  0.0000     0  0.0000 0.0000 0.0339  0.0000 -0.0155 -0.0415
[10,]  0.0069 0.0007  0.0000  0.0000 0.0000 -0.0800 0.0000 0.0000 -0.0517 0.0000     0 0.0000     0  0.0000     0  0.0000 0.0376 0.0319 -0.0292  0.0000 -0.0436
[11,] -0.0119 0.0430 -0.0212 -0.0343 0.1345 -0.0799 0.0000 0.0000  0.0000 0.0000     0 0.0000     0 -0.0394     0 -0.0347 0.0407 0.0720 -0.0430  0.0000 -0.0507
[12,] -0.0126 0.0155 -0.0368 -0.0309 0.1246 -0.0855 0.0000 0.0000 -0.0369 0.0469     0 0.0401     0 -0.0382     0 -0.0306 0.0505 0.0614 -0.0452  0.0000 -0.0393
```

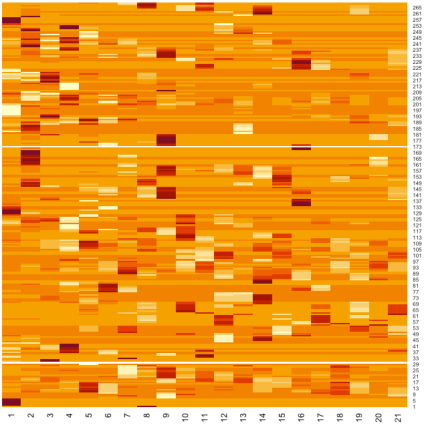


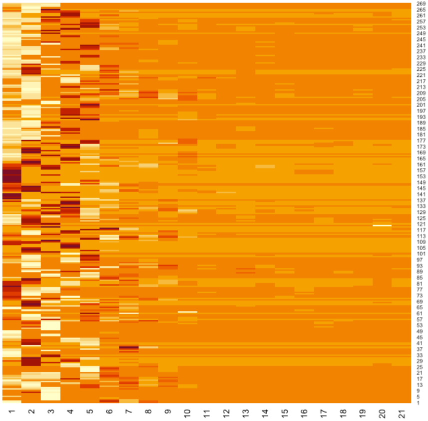


Random coincidence? Developing Surrogates

- **Constrained on the 1^{st} order**
  - Same `PR` and `mob` series with random (shuffled) temporal order
  - 2^{nd} order characteristics are not preserved.
- **Constrained on the 1^{st} and 2^{nd} order:**
  - Same power spectrum as our data, i.e. identical linear correlations
  - Same first order properties (variances and means),
  - Otherwise random
- **Constrained only on the 2^{nd} order:**
  - Same power spectrum as our data, i.e. identical linear correlations
  - Otherwise random

1000 Surrogates





Step 3: 2-Stage Partial Correlations

Original Series / **Lagged Series**

Original Series		Lag 1		...	lag 5		...	lag 21	
MOB	PR	MOB	PR		MOB	PR		MOB	PR
1	1	1	2		1	6	...	1	22
2	2	2	3		2	7	...	2	23
3	3	3	4		3	8	...	3	24
4	4	4	5		4	9	...	4	25
5	5	5	6		5	10	...	5	26
6	6	6	7		6	11	...	6	27
7	7	7	8		7	12	...	7	28
8	8	8	9		8	13	...	8	29
9	9	9	10		9	14	...	9	30
10	10	10	11		10	15	...	10	31
⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮
389	389	388	388		387	387	...	368	368

Sparsified P. Correlation Matrix for Window 1 & Lag 5

	PR5	MOB	PR	PR1	PR2	PR3	PR4	MOB1	MOB2	MOB3	MOB4
PR5	1	✓	x	x	x	✓	x	✓	✓	x	x
MOB		1									
PR0			1								
PR1				1							
PR2					1						
PR3						1					
PR4							1				
MOB1								1			
MOB2									1		
MOB3										1	
MOB4											1

Embedded Data Matrix for Window 1 and Lag 5

Zero-order Window 1 | Control Variables for Fully Partial Correlation for Lag5

PR5	MOB	PR	PR1	PR2	PR3	PR4	MOB1	MOB2	MOB3	MOB4
6	1	1	2	3	4	5	2	3	4	5
7	2	2	3	4	5	6	3	4	5	6
8	3	3	4	5	6	7	4	5	6	7
9	4	4	5	6	7	8	5	6	7	8
10	5	5	6	7	8	9	6	7	8	9
11	6	6	7	8	9	10	7	8	9	10
12	7	7	8	9	10	11	8	9	10	11

2nd Stage Partial Correlation Matrix

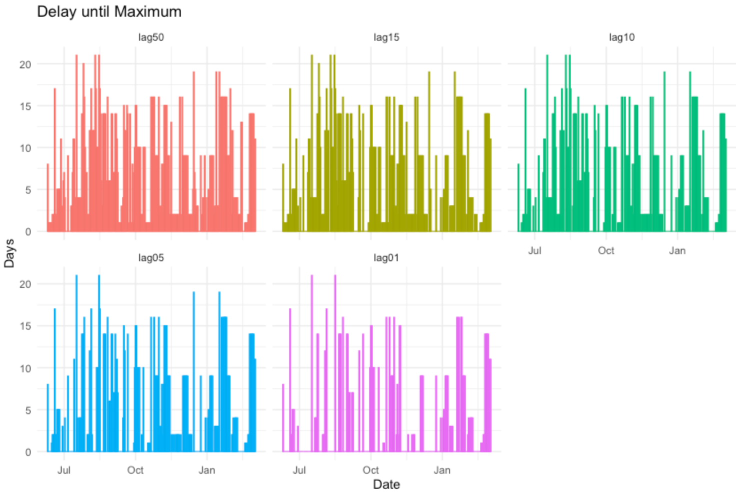
	PR5	MOB	PR3	MOB1	MOB2
PR5	1	?			
MOB		1			
PR3			1		
MOB1				1	
MOB2					1

0 if p > 0.05

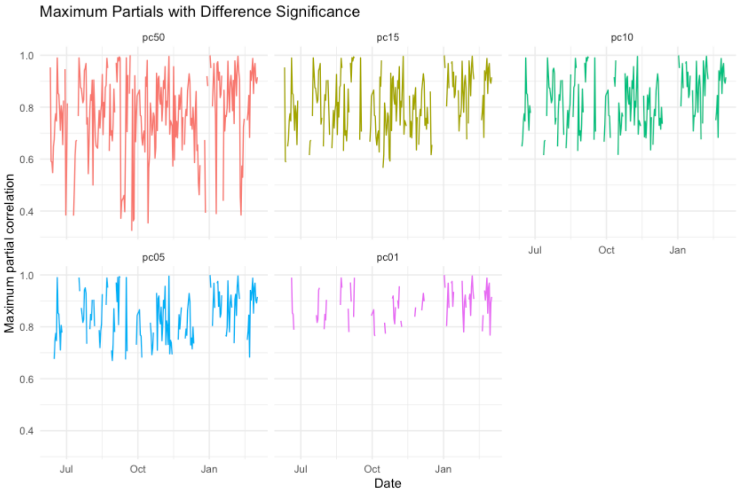
Correlations Matrix (windows x lags)

Windows	cor1	cor2	...	cor5	...	cor21
1	0.00	0.05	...	?	...	0.51
2						
⋮	⋮	⋮	⋮	⋮	⋮	⋮
269	⋮	⋮	⋮	⋮	⋮	⋮

Delays until maximum



Progressive decoupling



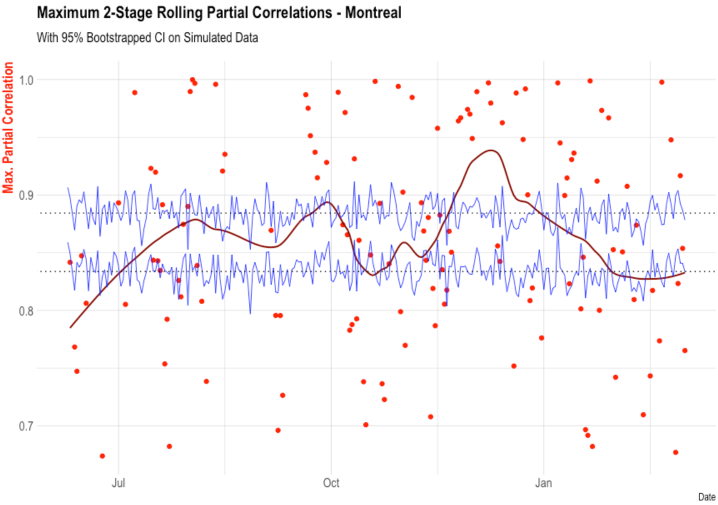
Descriptives for maximum P.Corr

p <	0.5	0.01	0.05	0.10	0.15
Min.	0.324	0.765	0.670	0.615	0.566
1st Qu.	0.689	0.844	0.768	0.748	0.732
Median	0.792	0.888	0.848	0.841	0.821
Mean	0.774	0.887	0.844	0.829	0.816
3rd Qu.	0.897	0.940	0.919	0.909	0.904
Max.	1.000	1.000	1.000	1.000	1.000
NA's	25	169	107	82	66
%.NA's	9.3	63.1	39.9	30.6	24.6

Descriptives for Delays

p <	0.5	0.01	0.05	0.10	0.15
Min.	1.00	1.00	1.00	1.00	1.00
1st Qu.	2.00	4.00	2.00	2.00	2.00
Median	7.00	9.00	9.00	9.00	8.00
Mean	7.67	8.60	7.89	8.05	7.84
3rd Qu.	12.00	14.00	12.00	14.00	12.75
Max.	21.00	21.00	21.00	21.00	21.00
NA's	25	169	107	82	66
%.NA's	9.3	63.1	39.9	30.6	24.6

Spuriousness Check



Re-Estimation (De-biasing)

Suppose that the covariance matrix $\boldsymbol{\Sigma}$ and the concentration matrix $\boldsymbol{\Omega}$ are partitioned according to random variables $\boldsymbol{X}_a$ and $\boldsymbol{X}_{-a}$, where $\boldsymbol{X}_{-a}$ is a $(p-1) \times 1$ random vector except for a random variable $\boldsymbol{X}_a$.

$$\tilde{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{a,a} & \boldsymbol{\Sigma}_{a,-a} \\ \boldsymbol{\Sigma}_{-a,a} & \boldsymbol{\Sigma}_{-a,-a} \end{pmatrix}, \quad \tilde{\boldsymbol{\Omega}} = \begin{pmatrix} \boldsymbol{\Omega}_{a,a} & \boldsymbol{\Omega}_{a,-a} \\ \boldsymbol{\Omega}_{-a,a} & \boldsymbol{\Omega}_{-a,-a} \end{pmatrix}$$

Using the symmetry in $\Omega_{-a,a}$ and $\Omega_{a,-a}$ and

$\hat{\boldsymbol{\Omega}}_{a,a} = (n - |\hat{\mathbf{ne}}_a|) \,/\, \left\| \mathbf{X}_a - \mathbf{X}_{\mathrm{ne}_a} \hat{\beta}^{a,\hat{\mathrm{ne}}_\mathrm{a}} \right\|_2^2$ where

$\tilde{\boldsymbol{\Omega}}_{-a,a} = -\tilde{\boldsymbol{\Omega}}_{a,a} \hat{\beta}^{a,\hat{\mathrm{ne}}_a}$ and $\hat{\beta}^{\boldsymbol{a},\hat{\mathrm{ne}}_a} = \left(\mathbf{X}_{\mathrm{ne}_a} \mathbf{X}_{\hat{\mathrm{n}\hat{\mathrm{e}}}_a}^{\mathrm{T}} \right)^{-1} \mathbf{X}_{\mathrm{ne}_a} \mathbf{X}_a$

$$\hat{\boldsymbol{\Omega}}_{a,b} = \hat{\boldsymbol{\Omega}}_{b,a} = \mathrm{sign}\left(\hat{\beta}_b^{a,\hat{\mathrm{ne}}_a} \right) \sqrt{\left| \tilde{\boldsymbol{\Omega}}_{a,b} \tilde{\boldsymbol{\Omega}}_{b,a} \right|} \quad \text{for } a \neq b$$

We obtain the estimate the partial correlation coefficients from $-\,\mathrm{scale}(\hat{\boldsymbol{\Omega}})$

Re-Estimation (De-biasing) - Intiution

Original Series			Lagged Series							
		Lag 1		...	lag 5		...	lag 21	
MOB	PR	MOB	PR		MOB	PR		MOB	PR
1	1	1	2		1	6	...	1	22
2	2	2	3		2	7	...	2	23
3	3	3	4		3	8	...	3	24
4	4	4	5		4	9	...	4	25
5	5	5	6		5	10	...	5	26
6	6	6	7		6	11	...	6	27
7	7	7	8		7	12	...	7	28
8	8	8	9		8	13	...	8	29
9	9	9	10		9	14	...	9	30
10	10	10	11		10	15	...	10	31
⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮
389	389	388	388		387	387	...	368	368

Embedded Data Matrix for Window 1 and Lag 5

Zero-order Window 1		Control Variables for Fully Partial Correlation for Lag5								
PR5	MOB	PR	PR1	PR2	PR3	PR4	MOB1	MOB2	MOB3	MOB4
6	1	1	2	3	4	5	2	3	4	5
7	2	2	3	4	5	6	3	4	5	6
8	3	3	4	5	6	7	4	5	6	7
9	4	4	5	6	7	8	5	6	7	8
10	5	5	6	7	8	9	6	7	8	9
11	6	6	7	8	9	10	7	8	9	10
12	7	7	8	9	10	11	8	9	10	11

Sparsified P. Correlation Matrix for Window 1 & Lag 5

	PR5	MOB	PR	PR1	PR2	PR3	PR4	MOB1	MOB2	MOB3	MOB4
PR5	1	✓	x	x	x	✓	x	✓	✓	x	x
MOB	✓	1	x	✓	x	✓	✓	x	x	x	✓
PR0			1								
PR1				1							
PR2					1						
PR3						1					
PR4							1				
MOB1								1			
MOB2									1		
MOB3										1	
MOB4											1

$$\tilde{\Omega} = \begin{pmatrix} \Omega_{a,a} & \Omega_{a,-a} \\ \Omega_{-a,a} & \Omega_{-a,-a} \end{pmatrix} \quad \tilde{\Omega} = \begin{pmatrix} \Omega_{b,b} & \Omega_{b,-b} \\ \Omega_{-b,b} & \Omega_{-b,-b} \end{pmatrix}$$

$$\tilde{\Omega}_{-a,a} = -\tilde{\Omega}_{a,a}\hat{\beta}^{a,\hat{ne}_a} \qquad \tilde{\Omega}_{-b,b} = -\tilde{\Omega}_{b,b}\hat{\beta}^{b,\hat{ne}_b}$$

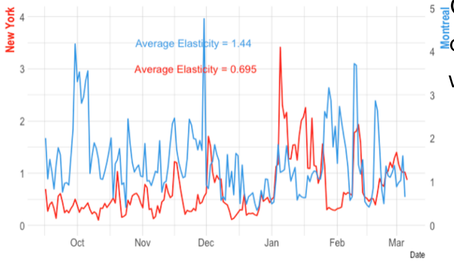
$$\hat{\Omega}_{a,b} = \hat{\Omega}_{b,a} = \text{sign}\left(\hat{\beta}_b^{a,\hat{ne}_a}\right)\sqrt{|\,\tilde{\Omega}_{a,b}\tilde{\Omega}_{b,a}\,|}$$

The 2-stage partials slightly overestimate the de-biased estimates

Elasticities

- **Zero-order correlations:**
  - We first use the full partial-correlation (delay-coordinate embedding) matrix
  - Apply the ridge-sparsity to see if `mob` is not "sparsified"
  - Use non-sparsified `mob`s for zero-order correlations (i.e., remove all intermediate lagged `PR` and `mob` columns)
  - Apply the significance test to identify the significant correlations in each window/lag: keep the significant ones.
- **Elasticities:**
  - $\epsilon = \frac{\partial PR/PR}{\partial R/R} = r \frac{s_{pr}}{s_r} \frac{\bar{R}}{\overline{PR}}$
  - When r is in the neighborhood of 1, the spread will be more sensitive or less (i.e., $\epsilon \lesseqgtr 1$) depending on two facts: the spread of COVID-19 is more or less variable than the mobility $\left(\frac{S_{PR}}{S_R}\right)$ and the magnitude of restrictions relative to how widespread PR is $\left(\frac{\bar{R}}{\overline{PR}}\right)$

Counterfactual Elasticities



Counterfactuals for Montreal are calculated in each rolling window with a dynamic lag optimization:

$$r^M \left[\frac{s_{PR}}{s_R}\right]^M \left[\frac{\bar{R}}{\overline{PR}}\right]^{NYC}$$

Differences between NYC and Montreal

	NYC	Montreal
Sensitivity = sd(PR)/sd(R)	11.9525200	18.3261807
Significance = mean(R)/mean(PR)	0.1112559	0.0291587
Beta = cov(PR,R)/var(R)	7.9307361	14.0195609
Correlation	0.7082259	0.7758325
Elasticity = Beta x Significance	0.6953200	0.4207255
Counterfactual Elasticity	**0.6953200**	**1.4404136**

What it tells us ...

In order to have this much jump in the elasticity for Montreal, two things have to be true in NYC relative to Montreal:

1. the magnitude of the decline in mobility should be much higher relative to the rise in spread ($\bar{R}/\overline{PR}$);
2. the mobility should have a much higher temporal variation relative to positivity rates (S_{PR}/S_R).

Given that the mobility metrics rather measure the people's behavioral response to the spread, these differences imply the following possibilities in Montreal:

1. the average reduction in mobility relative to the spread might not have been enough in terms of its magnitude and speed;
2. a significantly lower public sensitivity to the COVID-19 spread.

Section 5

Remarks

Concluding remarks

- We develop a method that can be used to capture the spatiotemporal dynamics of the relations between two variables (if the direction of correlations are known!)
- We show that the effect of (same) mobility restrictions on positivity rates vary by time and location
- We measure this dynamic relationship by correlation (nature of relationship) and elasticity (utilization of the relationship) for Montreal, NYC, Toronto, and Nova Scotia
- We show the main results for Montreal and compare it with NYC.
- We apply a counterfactual simulation to show why Montreal is different than NYC

Thank you

- Codes will be available on my GitHub repo: https://github.com/yaydede
- Presentation will be available on my website
- I'll be here at UniBZ for several months if you are interested in developing new ideas: **yigit.aydede@smu.ca**