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

EUHEA 2022 - Oslo

Yigit Aydede - Saint Mary's University, Canada

This version: 07 July 2022

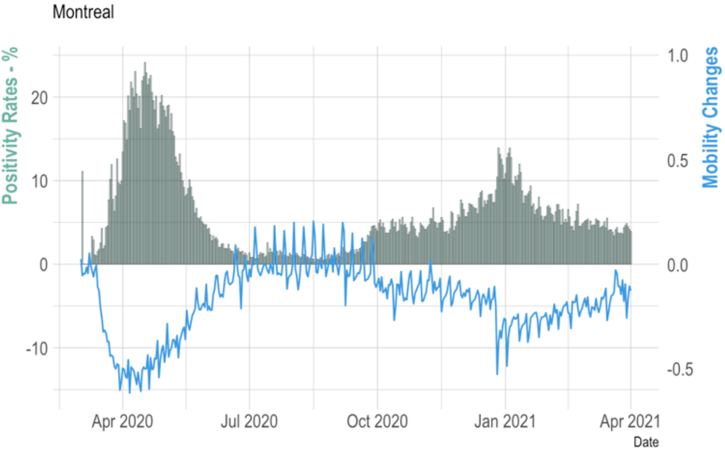
Motivation

- Can we recover the short-term temporal dynamics in the relationship between two (or more) time-series signals with minimum assumptions (about data structure, model)?
- For example, a method is needed to identify the number of days required to generate an intended effect on positivity rates (PR) following the mobility restrictions.
- 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.
- This method can be implemented to any two time-series signals (**with a known direction of correlation**) to recover dynamic correlations.

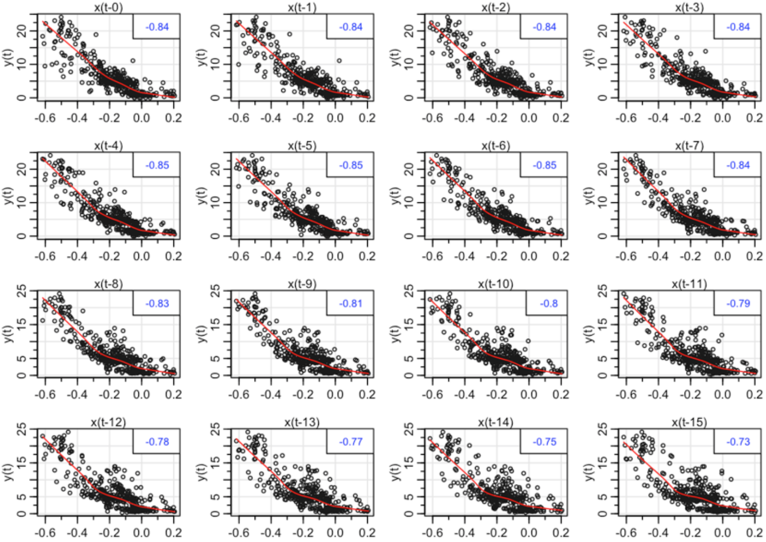
Summary

- Two time-series signals: X_t and Y_t and both are $\mathbf{I}(0)$,
- $X_t \to Y_{t+s}$ and $s \neq 0$ and $s \in \{1 : 21\}$,
- Find s^* that maximizes the partial correlation between X_t and Y_{t+s}
- Since s^* is not constant (time-varying relationship), sliding-window correlations are calculated
- Window length can be found by a wavelet analysis (as in TVFC literature)
- Since the algorithm searches for s^* in the window, we need to be sure that it represents the genuine association between two series. It must be distinguishable from lagged synchrony that would occur by chance.
- Regularization and statistical significance can solve (or reduce) this problem

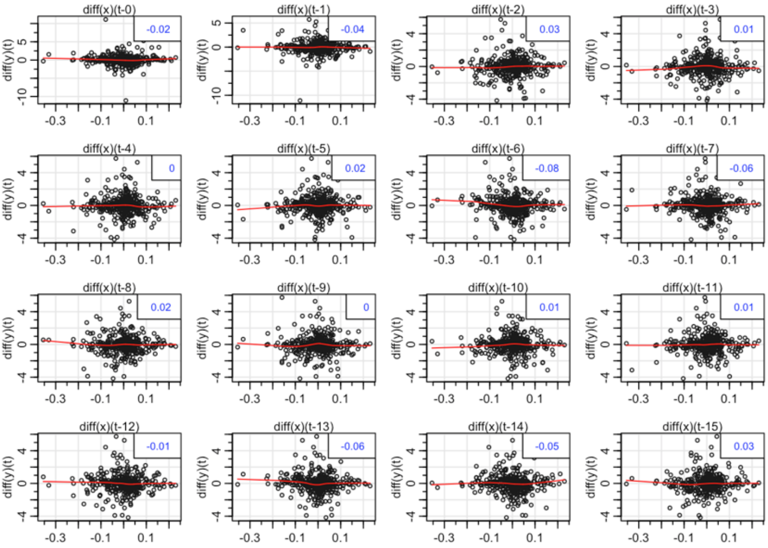
Example: PR vs. Mobility



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.

COVID-19: Effect of mobility on PR

- 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)

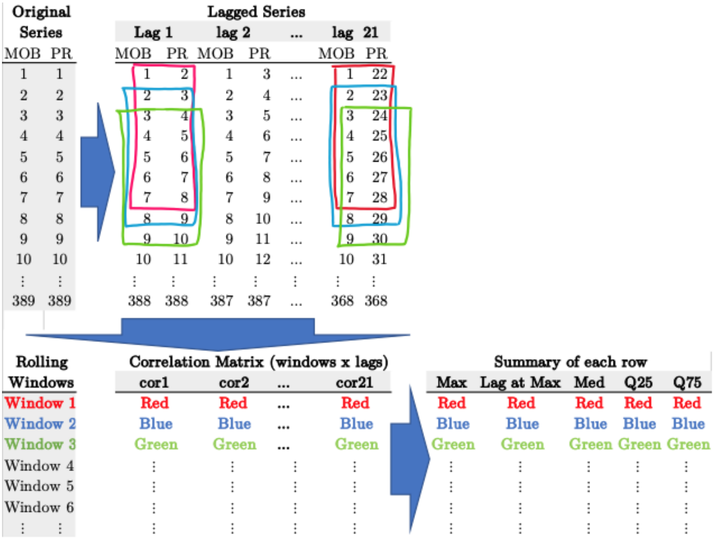
TVFC

- 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
- 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 PR, 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.

Algorithm

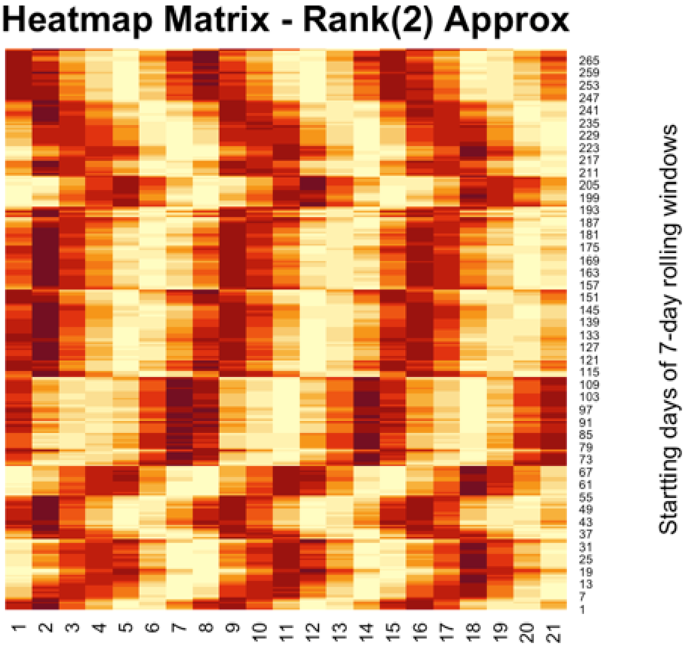


Original Series		Lagged Series						
		Lag 1		lag 2		...	lag 21	
MOB	PR	MOB	PR	MOB	PR		MOB	PR
1	1	1	2	1	3	...	1	22
2	2	2	3	2	4	...	2	23
3	3	3	4	3	5	...	3	24
4	4	4	5	4	6	...	4	25
5	5	5	6	5	7	...	5	26
6	6	6	7	6	8	...	6	27
7	7	7	8	7	9	...	7	28
8	8	8	9	8	10	...	8	29
9	9	9	10	9	11	...	9	30
10	10	10	11	10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

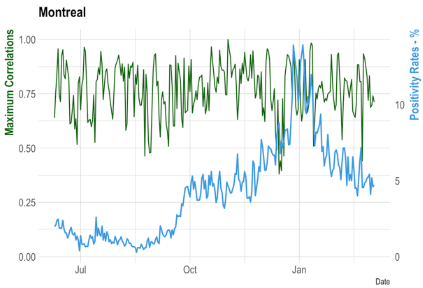
Summary of each row				
Max	Lag at Max	Med	Q25	Q75
Red	Red	Red	Red	Red
Blue	Blue	Blue	Blue	Blue
Green	Green	Green	Green	Green
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

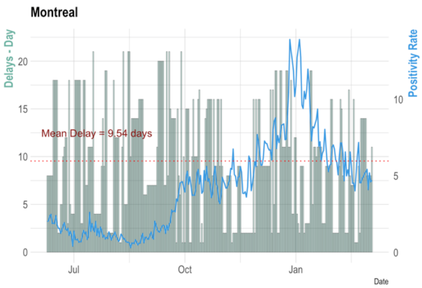
Heatmap



Heatmap Matrix - Rank(2) Approx
265
259
253
247
241
235
229
223
217
211
205
199
193
187
181
175
169
163
157
151
145
139
133
127
121
115
109
103
97
91
85
79
73
67
61
55
49
43
37
31
25
19
13
7
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Startting days of 7-day rolling windows

Maximums and Delays





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

Partial Correlations

- 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		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

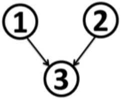
m-dimensional **reconstruction-space** vectors

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

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.

Regularization

- Even if n > p, we can use regularization to identify “significant” partial correlations
- **Berkson’s paradox** could be an issue



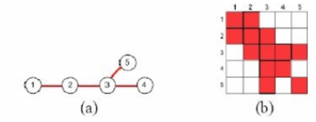
$$\begin{cases} x_1(t) = u_1(t) + noise \\ x_2(t) = u_2(t) + noise \\ x_3(t) = 0.5x_1(t) + 0.5x_2(t) + noise \end{cases}$$

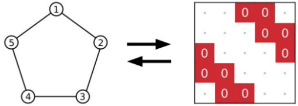
```
> x1 <- rnorm(100)
> x2 <- rnorm(100)
> x3 <- 0.5*x1+0.5*x2 + rnorm(100)
> mat <- cbind(x1, x2, x3)
> S <- cov(mat)
> -cov2cor(solve(S))
           x1         x2         x3
x1 -1.0000000 -0.4076111  0.6085341
x2 -0.4076111 -1.0000000  0.5647095
x3  0.6085341  0.5647095 -1.0000000
> 
```

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, $\boldsymbol{\Omega}$.





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 $j \neq i$:

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

What do we want?

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

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									
PR0			1								
PR1				1							
PR2					1						
PR3						1					
PR4							1				
MOB1								1			
MOB2									1		
MOB3										1	
MOB4											1

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	⋮	⋮	⋮	⋮	⋮	⋮

Ridge or GLasso?

- 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^* (False Discovery Rate - Efron)
  - Recovering partial coefficients from Ridge estimates
    - 2-Stage estimation
    - De-biasing
    - Re-estimations

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

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

$$\tilde{\Omega}_{-a,a} = -\tilde{\Omega}_{a,a}\hat{\beta}^{a,\hat{n}e_a} \qquad \tilde{\Omega}_{-b,b} = -\tilde{\Omega}_{b,b}\hat{\beta}^{b,\hat{n}e_b}$$

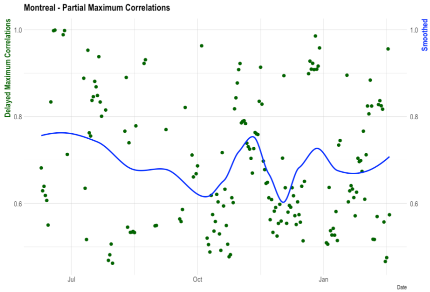
$$\hat{\Omega}_{a,b} = \hat{\Omega}_{b,a} = \mathrm{sign}\left(\hat{\beta}_b^{a,\hat{n}e_a}\right)\sqrt{|\ \tilde{\Omega}_{a,b}\tilde{\Omega}_{b,a}\ |}$$

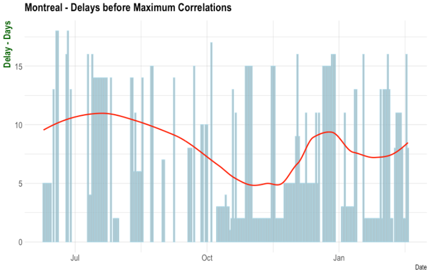
The 2-stage partials slightly overestimate the de-biased estimates

Why Composite Likelihood method (CLM)?

- Ridge and Glasso require de-biasing, but it doesn't have an established literature
- Or 2-step partials are similar to 2-step LASSO and may not be reliable in terms of their asymptotic properties.
- CLM is the perfect fit that removes the need for de-biasing and provides reliable asymptotic properties

Results

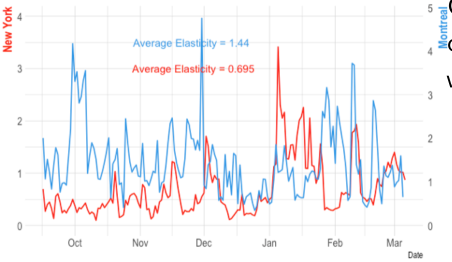

Montreal - Partial Maximum Correlations
Delayed Maximum Correlations
Smoothed
1.0
0.8
0.6
Jul
Oct
Jan
Date


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

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:

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

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.

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