Understanding the dynamics of a viral spread by sparse composite likelihood selection

Yigit Aydede
Saint Mary's University, Halifax, Canada

Work-in-progress. Please do not cite without permission.
This version: January 10, 2023

Abstract

We develop and apply a novel quantitative method to understand disease dynamics that help policy makers plan interventions to taper viral transmissions exemplified by the ongoing effort to stop the spread of COVID-19. Sparse composite likelihood selection is a recently developed method focusing on identifying true sparsity in complex models in case of high-dimensional data without compromising its asymptotic properties such as unbiasedness. The algorithm has a number of advantages including a rigorous sparsified solution to graphical network analysis of complex systems and the ability to efficiently handle high-dimensional data such that it is computationally efficient and statistically sound. We demonstrate that, unlike conventional shrinkage methods, when sparse composite likelihood selection is applied in a graphical network setting, it can capture the time-varying effects of mobility restrictions on the COVID-19 spread in Montreal and New York. For each case, we describe the utility of the method for surveillance and resource allocation.

1 Introduction

The recent advances in surveillance systems for infectious disease, capability of data collections and storage, and increased computational resources in the last decade have provided unprecedented tools for the scientific community to understand and, more importantly, combat the spread of infectious disease in populations. The importance of understanding the dynamics of underlying process in viral spread in response to its epidemiological factors such as weather-dependent correlates, the effects of socio-economic factors, and most importantly non-pharmaceutical interventions has become more evident with the recent COVID-19 pandemic.

On the other hand, a rise in the availability of sizable and granular data collected from non-clinical experiments such as mass testing or unconventional sources such as Google Flue

Trends has created a set of new computational challenges for analyzing large amounts of infectious disease data. This big-data regime requires data-driven analysis methods that can both mitigate the difficulties of high-dimensional measurements and maintain the fundamentally dynamic nature of disease spread. In this manuscript, we demonstrate how one such method, sparse composite likelihood selection (SCLS) applied to a graphical network analysis, can help in the analysis of infectious disease data.

Although compartmental models (such as SIR - Susceptible, Infectious, or Recovered) used in epidemiology are the workhorse of studying dynamics of an infectious disease, they are not equipped for recovering the short-term temporal dynamics in a viral spread due to challenges in the complexity and heterogeneity of the unknown underlying system. For example, since all observations in the epidemic data reflect transmission events from some time in the past, obtaining temporally accurate transmission rate (the effective reproduction number of infections, R(t), the average number of secondary cases of disease caused by a single infected individual over his infectious period) requires estimations of R(t) with the assumptions about the time delay between infections and observations. Sampling from a delay 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) (Gostic et al. 2020, Locatelli 2021). What we all observe in the COVID-19 pandemic, for example, is the case numbers and positivity rates based on imperfect testing practices (random and selective) on symptomatic or even non-symptomatic people. Due to the incubation period (estimated 1 to 21 days) and delays in testing and reporting, there are no observed data on the spread - R(t).

Our method, SCLS applied to a graphical network analysis, is fundamentally assumption-free operating solely on sliding windows in time, thus alleviating the need for a set of governing equations. Further, the required input data can be observational from historical data and need not be generated from simulations to adjust for incubation period. This is particularly true for the COVID-19 pandemic. Our research question, effects of NPI on the spread, is the most investigated question in the COVID-19 literature with more than three thousand published studies in two years (Perra, 2021). The contribution of our study is to uncover short-term temporal dynamics in the effect of NPI on the spread by using graphical network analysis applied with a new method, sparse composite likelihood selection. We aim to answer the following question: when the varying delays between the spread and mobility restrictions are identified properly, what would be the maximum possible effect of mobility restrictions? We develop three time-varying metrics to answer this question: (1) the correlation that reflects the nature of relationship between mobility and positivity rates; (2) the elasticity that measures how effectively that relationship is utilized to curb the spread; (3) the delay in the effect of these restrictions that reflects how efficient the contact tracing is.

Our findings reveal that the relationship between mobility and the spread is highly unstable even over short time intervals and there are several periods that show decoupling in the relationship. The delay in the effects of mobility restrictions ranges from 2 days to 22 days during the second wave reflecting the fact that we can observe immediate effects of mobility when contact tracing is very effective and the spread is relatively low. When there is a rise in community spread without well identified sources, however, the effect of mobility on PR stretches back to the upper bound of the incubation period, which is estimated to be 14 days, or beyond. In addition to contact tracing, logistical imperfections, such as delays in testing symptomatic people and long gaps between testing and processing, would also increase the delay. Therefore, the varying delays in the effects of mobility restrictions indicate how effective the local public policies are and how well they are implemented. This study documents these delays since the end of the first wave for New York City (NYC) and Montreal which can be used for fine-tuning the public health policies in place.

The rest of the paper is structured as follows. Section 2 summarizes the literature related to the COVID-19 pandemic and its challenges in finding the effect of mobility restrictions on the spread. Section 3 presents the data, a method that introduces lag optimization with rolling correlations, and the initial results. In Section4, we progress to network analysis from rolling correlations and discuss how sparse composite likelihood selection is different than other penalization methods. The results and robustness analysis are presented in Section 5. The last section concludes with a discussion about our application with COVID-19 data and future research.

2 COVID-19 Pandemic

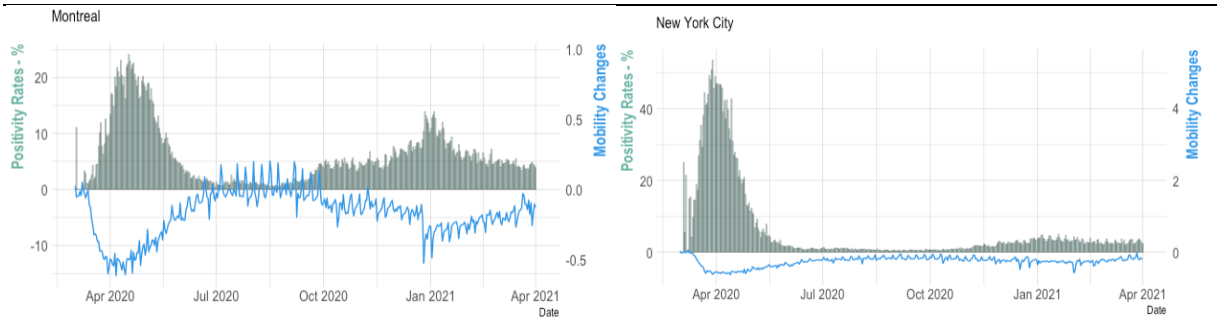
Centers for Disease Control and Prevention defines nonpharmaceutical interventions (NPI) as "actions, apart from getting vaccinated and taking medicine, that people and communities can take to help slow the spread of illnesses like pandemic influenza (flu)". Although the evidence unambiguously indicates that implementing NPIs with successful mobility restrictions have the largest effect on curbing the pandemic, studies looking at the dynamics of these confinement policies are rare. For example, recent studies (Askitas et al. 2021 and Nouvellet 2021) find the evidence of decoupling of transmission and mobility restrictions in certain periods in some countries. As well, the dynamic impact of each of the confinement policies on the spread is different due to the complexity in their association in terms of timing and intensity. While convincing reductions in case numbers have been observed in many regions after rapidly implemented mobility restrictions, spatial differences in their efficacy, even across the

neighboring regions with the same interventions, require understanding how the relationship between transmission and mobility changes over time and space.[1]

Despite the incredible effort of many scientific communities, there are still open questions and challenges for future work. For example, there is no study capturing the temporal dynamics in the relationship between mobility and the spread. Studies investigating the effect of lockdown policies on the incidence of COVID-19 have been mostly using regional - within country – or international panel observations with the assumption that the delay between mobility changes and their effect on the spread is constant over time and space. These studies assume that the delay ranges from 7 days to 14 days when the spread is measured by reproduction numbers and about 23 to 30 days when it is measured in terms of deaths (Flaxman et al. 2020, Fang et al. 2020).

For example, there is conflicting evidence on the influence of weather on COVID-19 transmission (Xu et al. 2021). To estimate weather-dependent signatures of the pandemic, the effects of socio-economic factors and non-pharmaceutical interventions must be controlled for. Yet, the delay between exposure and detection of infection complicates the estimation of weather impact on COVID-19 transmission, potentially explaining significant variability in results to-date. One of the first studies (Xu et al. 2020) on the subject using the county-level data with more than 3700 counties in the U.S. found that failing to control for the delay between infection and official recording of cases is the main reason for the mixed evidence. They conclude that the delay, which is a particularly understudied factor and estimated to be approximately 10 days, confounds the attempts to associate daily weather conditions with recorded new cases and may partially explain the inconsistent and inconclusive findings to date.

Figure 1: Mobility vs. Positivity Rates



[1] For example, while Born et al (2020) finds that a lag of one month after lockdown in Sweden would not differ from the actual infection dynamics, Cho (2020) with a similar approach but considering a longer lag find that lockdowns have been effective in Sweden.

Our study compares two major cities: Montreal and New York City. We use test positivity rates (PR) that reflect the spread as well as the Facebook mobility data called "all_day_bing_tiles_visited_relative_change", which measures positive or negative changes in movement relative to baseline in those three cities. The data on Montreal is from INSPQ and obtained by using an unofficial API that powers the INSPQ's own dashboard.[2] Otherwise, the data is not publicly available. The data for NYC is available in the government's website provided by daily numbers.

Most studies use the reproduction number, R, which is the average number of secondary cases of disease caused by a single infected individual over his/her infectious period estimated based on one of the epidemiological models, mostly an algorithm called "EpiEstim" or "EpiNow2".[3] This estimated statistic, which is time and situation specific, is commonly used to characterize pathogen transmissibility during an epidemic. Nevertheless, methods fitting mechanistic transmission models to incidence data (case or death numbers) are often difficult to generalize because of the context-specific assumptions often made. Instead, we choose to use the test positivity rates (PR) (Gostic et al. 2020; Locetti et al. 2020). Due to high test positivity rates led by very low and selective tests, we exclude the first wave and use the test positivity rates after June 8, 2020 in each city. The main issue in using PR is that it would be misleading when the testing is not random. However, when it comes to gauging the severity of the epidemic, the PR might serve as a good proxy for the case incidence rate specially as a day-to-day level indicator about how the spread is changing day-by-day. Hence, for a short-run analysis like ours, the positivity rate may be a plausible predictor, especially during the period when the test numbers are very high with relatively random applications in those three major cities.

3 Rolling correlations with optimal lag control

Several methods are common in different fields to model the delayed relationships between multiple variables. These methods, such as cross-correlation, autoregressive and cross–lagged structural models (Cook, Dintzer, & Mark, 1980; West & Hepworth, 1991), multivariate time series methods (Box, Jenkins, & Reinsel, 1994), cross–spectral or coherence analysis (Bloomfield, 1976; Warner, 1998), estimate some set of linear or nonlinear relations between observations separated by intervals of time while assuming that this structure remains constant over time.

A naïve application that uses the whole data with varying lags up to 21 shows that the cross correlations between mobility and PR are negative for all lags ranging from -0.84 to -0.69. When the nonstationary with seasonality is removed from both series, these cross correlations become

[2] https://www.inspq.qc.ca/covid-19/donnees

[3] https://academic.oup.com/aje/article/178/9/1505/89262

zero. This statistical outcome is not unexpected because, in both cases, one static long-term relationship would not fit the entire period. Since any change in mobility also reflects the overall social response - mixed of voluntary and government mandated behavioral changes - to the most immediate developments in pandemic, it is also likely to capture reversed causality reflected in negative correlations between the observed spread and mobility reductions: A simple visual inspection of cross correlations in A2 Appendix can confirm this fact.

The solution to this problem can be partitioning the data with sliding windows. For example, Nouvellet et al. (2021) set a parametric model to understand the relationship between transmission and mobility with 7-day sliding windows and show that the relationship is not stable and changes in each window so that, for the 43% of country-periods, the relationship was either non-significant (32%) or reversed (11%), which implies that the decline in mobility is associated with a rising transmission (measured by deaths). Although the incubation period is implicitly incorporated in their estimations, it may not capture the delay between two series in each window properly so that the cause-and-effect order is reversed.

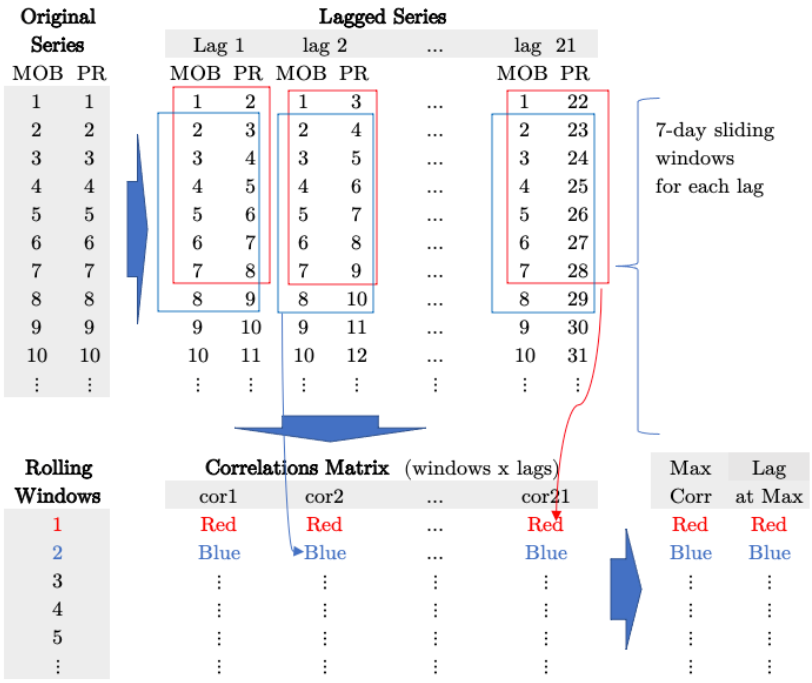
3.1 Rolling correlations

To calculate the dynamic nature of this relationship, we develop a trainable nonparametric approach inspired from two recent methods: the dynamic functional connectivity (DFC), which has emerged as a major topic in the resting-state BOLD fMRI (blood oxygenation level dependent functional MRI) literature and the windowed cross-lagged correlation (WCLC) that is used in behavioral psychology (Boker et al. 2002) to analyze the movement synchrony in nonverbal communications. The idea behind both methods is relatively simple: analogous to a moving average function, a sliding window analysis computes a succession of pairwise correlation matrices using the time series from related sources.[4] The complexity of the sliding window technique, however, comes from finding the appropriate settings of multiple parameters such as window function, length, and step size which remain unknown due to lack of "ground truth" that defines the relationship. Therefore, the problem of "window-size" remains as a main challenge in both methods: increasing a window length results in decreasing the sensitivity for identifying fast changes with very long windows eventually measuring static connectivity. On the other hand, shorter windows can increase sensitivity for detecting short transition states but at the expense of decreasing the signal-to-noise ratio leading to spurious fluctuations in the dynamic connectivity.

[4] Although DFC (and WCLC) has been long used, there are still considerable technical issues associated with both approaches. A great effort has recently been dedicated to investigating how the size of sliding windows affects DFC estimations. It is essential to determine the window length that allows reducing spurious fluctuations and at the same time capturing faster dynamic correlations. One of the most suggested methods to address spurious fluctuations in DFC is to estimate a method when the window length is not shorter than the largest wavelength present in both series. An extensive review on DFC and how it is used in neuroscience is provided by Lurie et al. (2020).

Unlike the applications in neuroimaging, finance, environmental studies, and behavioral psychology, we have an unquestionable epidemiological "ground truth", the fact that the spread of an infectious disease must be positively related to the intensity of contacts (contact rate x probability of transmission) in a population, which is the core idea in the so-called SIR models (Blackwood and Childs, 2018). This prior knowledge in the relationship between PR and the mobility series enables us to consider cross-correlations in sliding windows.

Figure 1: Schematic of the algorithm with zero-order correlations



Notes: Numbers represent the time in days.

3.2 Cross-correlations in sliding windows

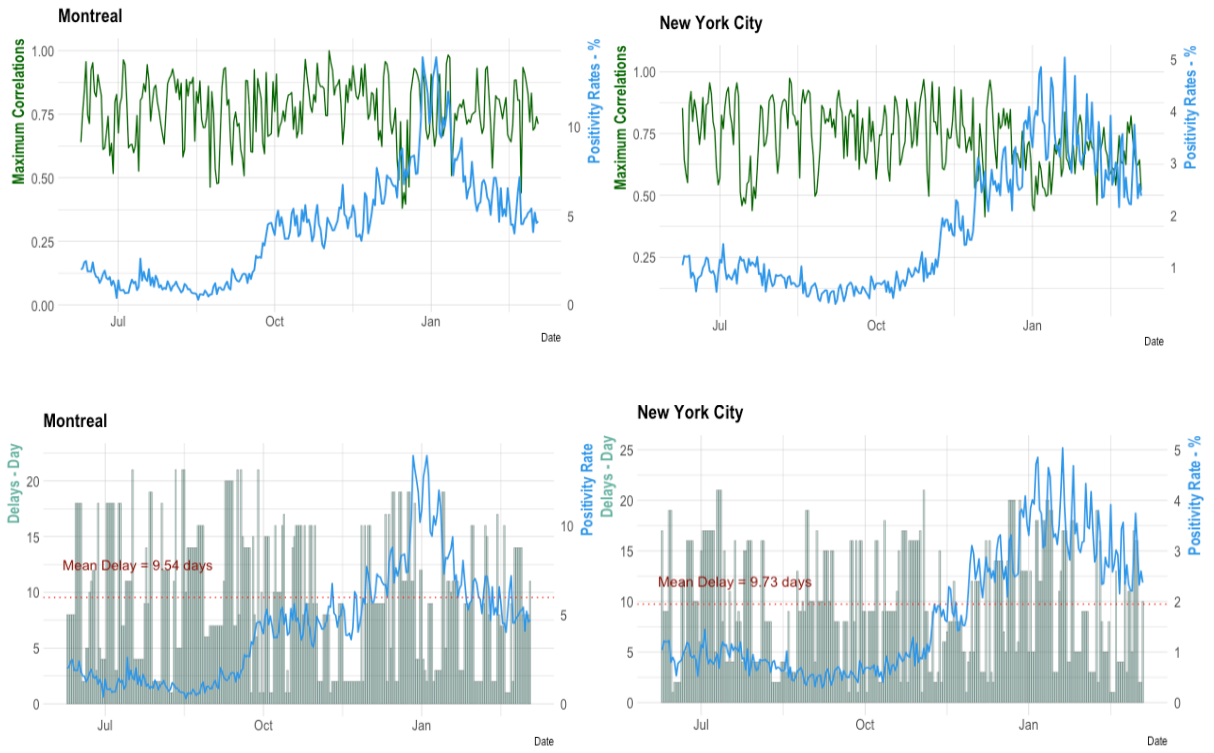
A common method for estimating the association between two time series is cross–correlation based on a vector of sequential observations selected from each time series such that both vectors contain the same number of occasions. Then, Pearson product moment correlation is calculated for these two vectors. The time interval separating the beginning of the two vectors is called as "lag" or "offset", τ. A similar method, as known as windowed cross-lagged correlation (WCLC) introduced earlier in 2002 (Boker et al. 2002) is widely used in behavioral sciences. The cross–correlation between $\mathbf{X}$ and $\mathbf{Y}$ at a lag τ can be defined as

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

However, a single measure of cross–correlation may not provide a good representation of association between two globally stationary time series, as the delayed association between two series is subject to change in small windows of observations. When a stable association between two behavioral time series is not expected, a more temporally detailed analysis of cross-correlation can be calculated using many short cross–correlational windows in which the starting time of windows of observations is incremented over the whole data set.

One way to examine how the intensities and delays of association between two time series are changing over time is to use shorter intervals of data from each time series to 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, which is illustrated in Figure 1. The justification of this objective, finding the maximum positive relationship with a dynamic lag control, is based on the epidemiological fact that the relationship between mobility restrictions and the viral spread must be positive or zero. The existence of this "grand truth" is a necessary condition for the dynamic lag control and underlines the difference between DFC and our approach. Hence, using short sliding windows that cover the time series leads to a moving estimate of association with the time-varying lags so that the correlation is maximized in each window. The results are presented in Figure 2.

Figure 2: Maximum correlations and delays - Mobility and PR

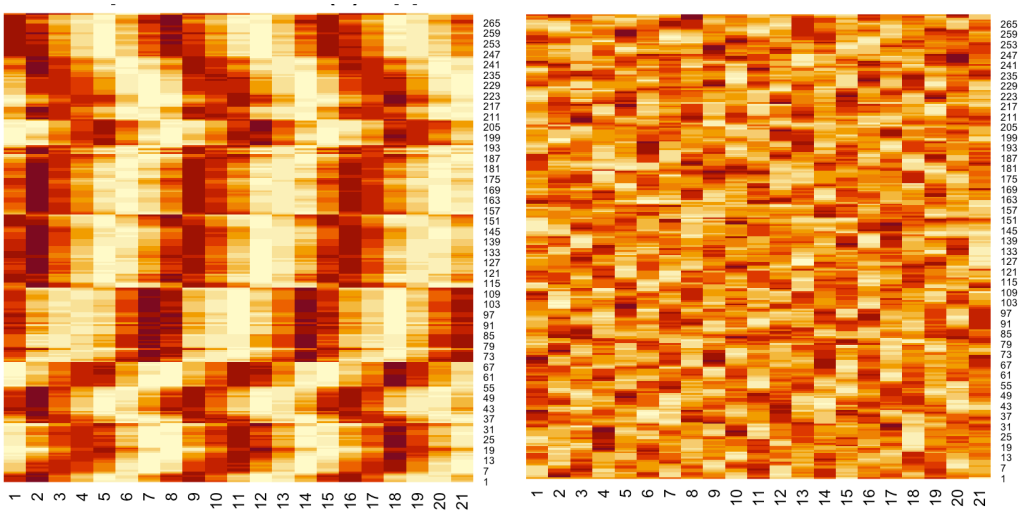


As stated before, we ignore the first 100 days and start on June 8 due to possibly misleading PR numbers. The graphs above show the estimation of maximum correlations between mobility and PR in each 7-day rolling windows calculated with optimal lag control. Note that series are first-differenced. The figures at the bottom show the time varying delays when the maximum positive correlation is obtained between mobility restrictions and PR. We can think that more immediate effects of mobility on PR happen when the effectiveness of contact tracing is high. When there is an increase in community spread without well identified sources, the effect of mobility on PR stretches back to the upper bound of the COVID-19 incubation period, which is estimated to be 14 days, or beyond.

3.3 Limitations

One of the challenges in this method is to control for coincidentally identified correlations, which has been well discussed and documented in a recent study by Dean and Dunsmuir (2015). There are two general concerns in assessing the relationship between two series: first, whether the genuine association between two series is distinguishable from lagged synchrony that would occur by chance. Second, an assessment of statistical significance is needed. A conventional approach for addressing the first concern is to establish reasonable estimates of significance limits by using surrogate data. In our case, we search the maximum positive correlation between a subset (one window) of mobility series and 21 subsets (21 subsequent rolling windows) of PR series. As described Figure 1, the algorithm finds a correlation matrix where the rows represent each rolling window, and the columns show the correlations for each of the 21 lags.

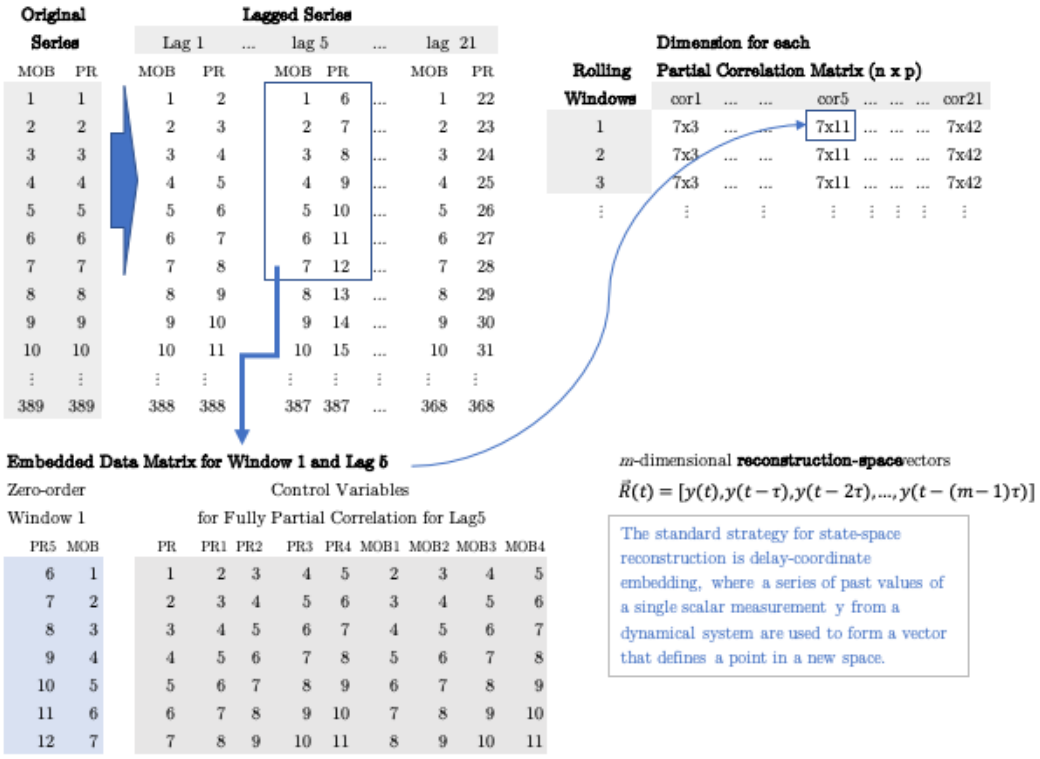
Figure 3: Heatmaps of correlation matrices



Note: Darker colors represent positive correlation. In both heatmaps, rows show the rolling windows (starting day) and the columns show the lags (delay between PR and Mobility series in days)

To address the concern of whether the correlation matrix from two completely independent random series would be distinguishable from the one that uses the original series, we show the heatmap of correlation matrices in Figure 3. The first heatmap is from the original correlation matrix after reducing the noise using a Rank 2 approximation. The second heatmap reflects the correlation matrix of a surrogate data set generated with constraints on the first and second orders. That is, we preserve the same power spectrum as in our data, with identical linear correlations and the same first order properties (variance and mean). The heatmaps show that our algorithm captures the correlations beyond a random coincidence. However, finding the delay where the maximum correlation occurs in each rolling window remains a main challenge, since it assumes that the maximums are the true correlations in each rolling window beyond a random coincidence.

Figure 4: Algorithm with partial correlations



Notes: Numbers represent the time in days.

The method described above is used to evaluate whether our search algorithm finds correlations that can be a random coincidence. Yet, to evaluate the statistical significance of each correlation coefficient, there needs to be properly calculated standard errors for each rolling window. However, due to the sample size (7) in each rolling window, a conventional normality

assumption cannot be sustained. Further, using bootstrap applications with a very small sample size (less than 20) are not generally recommended as their reliability becomes questionable.

In addition to these two problems, since the correlations are zero-order, they omit the intermediate lags of each series in their calculations. When analyzing the delayed effect of the mobility on PR with correlations, the effects of intermediate lags of both series should be controlled for by using partial correlations. These issues bring us to develop a different method so that we can address these challenges.

4 Network analysis

When the objective is to analyze the synchronous relationship between two time series signals with bivariate data, zero-order cross correlations with sliding time-windows are capable to discover the dynamics of the relationship. Zero-order correlation is robust and parameter free, but there is an intrinsic limitation to identify the direct effects of time-delayed relationship. To understand the limitation of zero-order correlations and the nature of graphical networks, we fist explain why our algorithm depicted in Figure 1 and applied above has to be restructured.

Our objective to discover the delayed effect of mobility on PR. The algorithmic described in Figure 4 takes an example of the first 7-day window for Lag 5. While a zero-order correlation between PR5 and MOB can be calculated using only these series (blue area), it ignores the indirect effects of intermediate lags shown in the grey area in the "embedded data matrix". When control for the intermediate lags between PR5 and MOB, it creates a complex network for each sliding and lagged window. As shown in the schema, this leads to high-dimensional multivariate embedded data matrices where the sample size (n) is 7 but the number of columns (p) in each matrix goes up to 42 as the delays (lags) are set between 1 and 21 in each window.

To obtain partial correlations, when the data is multivariate, its dimension must be n>p, which is required for non-singular covariance matrix. When the data become high-dimensional with 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. With a proper visualization of the network, it is called as graphical network analysis - or Gaussian Graphical Method (GGM), if the data satisfy conditions for a multivariate normal distribution - and mostly used in genomics, finance, psychology, neuroscience to identify the "edges" (David Hevey, 2018). 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 $j \neq i$:

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

A contemporary use for precision matrices is found in network reconstruction through graphical modeling (Network Analysis). Graphical modeling refers to a class of probabilistic models that uses graphs to express conditional (in)dependence relations between random variables. In a multivariate normal model, $p_{ij} = p_{ji} = 0$ if and only if X_i and X_j are conditionally independent when condition on all other variables, i.e. X_i and X_j are conditionally independent given all X_k where $k \neq i$ and $k \neq j$ if and when the ij th and ji th elements of $\mathbf{P}$ are zero. In real world applications, this means that $\mathbf{P}$ is often relatively sparse (lots of zeros). This also points to the close relationship between $\mathbf{P}$ and the partial correlations. The non-zero entries of the symmetric $\mathbf{P}$ matrix can be interpreted the edges of a graph where nodes correspond to the variables.

4.1 Ridge and gLasso for sparsified precision matrix

Interest in graphical models that combine a probabilistic description (through a multivariate distribution) of a system with a graph that depicts the system's structure (capturing dependence relationships), has surged in recent years [ref]. In its trail this has renewed the attention to the estimation of precision matrices as they harbor the conditional (in)dependencies among jointly distributed variates. With the advent of high-dimensional data, for which traditional precision estimators are not well-defined, this brought about several novel precision estimators. Generally, these novel estimators overcome the undersampling by maximization of the log-likelihood augmented with a so-called penalty. A penalty discourages large (in some sense) values among the elements of the precision matrix estimate. This reduces the risk of overfitting but also yields a well-defined penalized precision matrix estimator.

To solve the problem, penalized estimators (like Ridge or gLasso[5]) adds a so-called penalty to the likelihood functions (ℓ_2 in Ridge and ℓ_1 in lasso) that makes the eigenvalues of $\mathbf{S}$ shrink in a particular manner to combat that they "explode" when $p \geq n$ (Wieringen and Peeters, 2016). "Shrinking" is a "biased estimation" as a means of variance reduction of $\mathbf{S}$. The graphical lasso (gLasso) is the ℓ_1-equivalent to graphical ridge. A nice feature of the ℓ_1 penalty automatically induces sparsity and thus also select the edges in the underlying graph. The ℓ_2 penalty in Ridge relies on an extra step that selects the edges after the regularized precision matrix with shrunken correlations is estimated.

Regularization helps us find the sparsified partial correlation matrix for each window and lag. Since this matrix reveals the significant "edges" in the network, it shows whether the link between mobility series in each 7-day window with a specific lag and PR series is significant or not. However, since the penalization introduces bias, the correlation coefficient could be useless with their unknown asymptomatic features. The overcome this problem, one can apply

[5] https://cran.r-project.org/web/packages/glasso/index.html

the 2-stage estimations as shown in Figure 5. This process can be defined with the following four steps:

1. The sparsified correlation matrix in each window and lag identifies whether the link ("edge") between PR and mobility (in the example above PR5 and MOB) is zero or not.
2. If it is zero (i.e., no link between PR5 and MOB), it puts zero in the final correlation matrix. If it is not zero, it identifies other control variables (intermediate lags of PR and MOB) that are significant.
3. If the number of "unsparsified" control variables is less than 5, it calculates the partial correlation for that specific window and the lag and checks its statistical significance. If it is not statistically significant it replaces its correlation with zero in the correlation matrix.
4. If the number of "unsparsified" control variables is more than five, the algorithm selects the most "significant" four control variables based on their shrunken ridge estimates.

Thus, this procedure enables a two-layer robustness check: (1) with the sparsification of each precision matrix, it identifies the nonexistent edges between PR and MOB for each window and lag; (2) if the link exists, it identifies whether it is statistically significant or not. If it is not, it replaces it with zero in the correlation matrix.

Figure 5: Algorithm with Ridge & gLasso

Notes: Numbers represent the time in days.

Recovered partial correlations after a sparsification could be problematic as the 2-step estimation described above ignores the symmetry in each correlation matrix. There are some recent discussions about how to recover de-biased estimates (Ha and Sun, 2014) but their asymptotic properties are not established yet. An alternative method to these parametric penalization methods (Ridge and gLasso) can be a solution to these issues, which we describe below.

4.2 Sparse composite likelihood selection (SCLS)

In this section we will summarize the key characteristics of using SCLS that distinguish its approach from classic shrinkage methods such as Ridge and gLasso described and applied pervious sections.

Let Y be a $d \times 1$ random vector with density $f(y;\theta)$ indexed by the parameter $\theta \in \Theta \subseteq \mathbb{R}^p$. Suppose that the full d-dimensional density of Y is difficult to specify or compute but we can identify p densities $f_j(y;\theta)(j = 1, \dots, p)$ defined on low-dimensional subsets of Y, such as marginals Y_j, pairs (Y_j, Y_k), or conditionals $Y_j \mid Y_k = y_k (j \neq k)$. Given independent observations $Y^{(1)}, \dots, Y^{(n)}$ on Y, the composite likelihood estimator maximizes the composite log-likelihood function

$$\ell\big(\theta; Y^{(1)}, \dots, Y^{(n)}\big) = \sum_{j=1}^{p} \ell_j\big(\theta; Y^{(1)}, \dots, Y^{(n)}\big)$$

where $\ell_j\big(\theta; Y^{(1)}, \dots, Y^{(n)}\big) = \sum_{i=1}^{n} \log\ f_j\big(Y^{(i)}; \theta\big)$ denotes the sub-likelihood associated with the jth data subset. The composite likelihood estimator has become popular in many areas of statistics due to the simplicity in defining the objective function and computational advantages compared to the maximum likelihood estimator. At the same time, it has the same desirable first-order properties as maximum likelihood, such as consistency; see Varin et al. (2011) for a comprehensive survey.

The composite likelihood framework naturally suits problems where the parameter dimension p is allowed to diverge with the sample size. Since composite likelihood selection is a form of regularization but not in the parameter space, it provides a crucial tool as it solves many issues in both statistical properties and computing cost of the resulting estimator. Recent works (Caterina and Ferrari, 2021; Huang and Ferrari, 2021) introduce a flexible and computationally convenient method to build a composite likelihood function starting from a very large number of potential sub-likelihood candidates[6]. The main idea is to minimize a convex criterion representing statistical efficiency with the addition of a weighted L_1-penalty to avoid selection of too many noisy terms. Each sublikelihood is assumed to contain distinct elements of θ in its

[6] See two most recent working papers:

setting; while this simplification has further computational advantages when p is large, it also enables one to conduct model selection. The key point here is that the penalty focuses on selection of sub-likelihood functions rather than of elements of θ. Unlike classic shrinkage methods (Ridge gLasso), this strategy has the advantage to retain unbiasedness of the final estimating equations and consistency of the related parameter estimator. Therefore, in our case, the sparsity of our precision matrix can be obtained without compromising consistency of the partial correlation estimates.

To summarize the method, let us look at a case where the parameter vector $\theta = (\theta_1, \dots, \theta_p)^\top$ is sparse, in the sense that a large fraction of its elements is exactly zero, and p is allowed to grow with the sample size n. Let $\mathcal{A} = \{j: \theta_j \neq 0\}$ be the index set for the p^* nonzero elements in θ indicating respective sub-vectors and sub-matrices when used as a subscript. It is assumed here that each sub-likelihood $\ell_j(\theta)$ depends only on the specific component θ_j; this simplification has computational advantages when p is large. The marginal scores are defined by $u_j(\theta_j; y) = \partial \log\ f_j(y; \theta_j)/\ \partial \theta_j (j = 1, \dots, p)$, whilst $u(\theta; y) = \{u_1(\theta_1; y), \dots, u_p(\theta_p; y)\}^\top$ denotes the vector collecting all these scores. The approach presented next is also valid for the more general setting where each sub-likelihood depends on a finite number of parameters, in which case the jth score equals $u_j(\theta; y) = \sum_{k=1}^{p} \partial \log\ f_k(y; \theta)/\ \partial \theta_j$.

The main goal is to reduce the model dimension by dropping all the zero elements of θ while estimating the remaining elements. To this end, we take the estimator $\hat{\theta}$ with jth element defined by $\hat{\theta}_j = \tilde{\theta}_j I(\hat{w}_j \neq 0)$, where θ_j is the jth marginal estimator

$$\theta_j = \left\{ \theta_j : 0 = \sum_{i=1}^{n}\ u_j(\theta_j; Y^{(i)}) \right\} (j = 1, \dots, p)$$

and $\hat{w} = (\hat{w}_1, \dots, \hat{w}_p)^\top$ is the selection rule obtained by minimizing the penalized objective

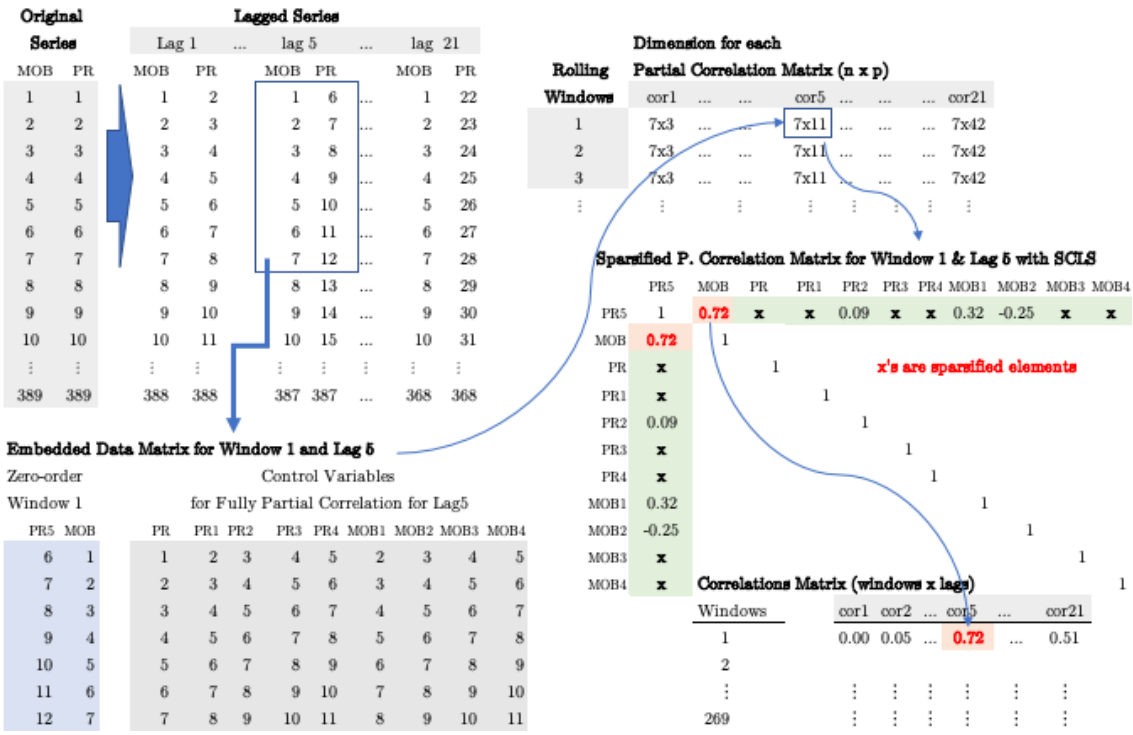
$$\hat{d}_\lambda(w) = \tfrac{1}{2} w^\top \hat{C} w - w^\top \text{diag}\ (\hat{C}) + \tfrac{\lambda}{n} \textstyle\sum_{j=1}^{p}\ \frac{|w_j|}{\theta_j^2} \tag{1}$$

for some user-specified constant $\lambda \geq 0$. Here $\text{diag}(A)$ is the diagonal vector of the square matrix A, where $\hat{C}$ is an estimator of the $p \times p$ score covariance matrix $C(\theta) = \text{var}\{u(\theta; Y)\} = E\{u(\theta; Y) u(\theta; Y)^\top\}$.

In our case, a natural choice considered here is the empirical covariance matrix

$$\hat{C} = \frac{1}{n} \sum_{i=1}^{n}\ u(\theta; Y^{(i)}) u(\theta; Y^{(i)})^\top$$

Figure 6: Algorithm with SCLS



Sparse sub-likelihood selection occurs through the minimization of the convex objective (1): the jth sub-likelihood $\ell_j(\theta)$ is included in the composite likelihood function if $\widehat{w}_j \neq 0$, else $\ell_j(\theta)$ is dropped and the corresponding parameter estimate is set as $\widehat{\theta}_j = 0 (j = 1, \ldots, p)$. The selected composite likelihood function is interpreted as one that maximizes statistical accuracy given a desired level of sparsity. In particular, when $\lambda = 0$ the objective $\widehat{d}_0(w)$ corresponds to the so-called finite-sample optimality criterion, a benchmark to find minimum variance estimators for unbiased estimating equations. The last term in (1) is a sparsity-inducing penalty discouraging overly complicated composite log-likelihoods.

The geometric properties of the L_1-penalty imply that several elements in $\widehat{w}$ are exactly zero for sufficiently large values of λ, which also induces sparsity in the estimator $\widehat{\theta}$. The considered penalty is adaptive in the sense that when θ_j is near 0 the jth sublikelihood receives a large penalty. Adaptive weighting is a fundamental feature of this method ensuring consistent model selection. The penalty is inspired by the adaptive Lasso penalty introduced by Zou (2006) in the context of sparse regression; however, the role of the adaptive penalty here is completely different because it focuses on the coefficients $w_j s$ associated with entire sub-likelihoods, rather than on the parameter elements θ_js. Penalization on the score space enables one to separate the task of model selection from that of parameter estimation.

Hence, differently from existing penalized composite likelihood procedures, the selected estimating equations remain unbiased and lead to consistent parameter estimators when the sub-likelihoods are corrected selected. In the next sections, we will report the results obtained from the application of SCLS.

5 Results

The application of SCLS provides us three vital features that we cannot have them with thresholding Ridge or gLasso: (1) since entries in the precision matrix are unbiased, partial correlation coefficients can be calculated without any auxiliary application (de-biasing or 2-stage partial correlations); (2) since each precision matrix is sparsified, we can also see whether the partial coefficient between MOB and PR in each rolling window with different lags is sparsified or not. This provides a unique opportunity to skip the conventional test for statistical significance, which is otherwise impossible with other methods.

Below, we only present selected results based on correlation matrices estimated by SCLS as described in Figure 6. The maximum correlation for each row in the matrix containing the partial correlations is selected as the "most diverse" and highest positive correlation among 21 candidates for each window (row). We can calculate 95% confidence intervals for each unsparsified correlations. After eliminating the negative correlations, the "most diverse" maximum correlation is the one that has the lowest intersections with others based on all confidence intervals in any given row. This application is discussed in the Appendix.

Figure 6: Delay - Montreal

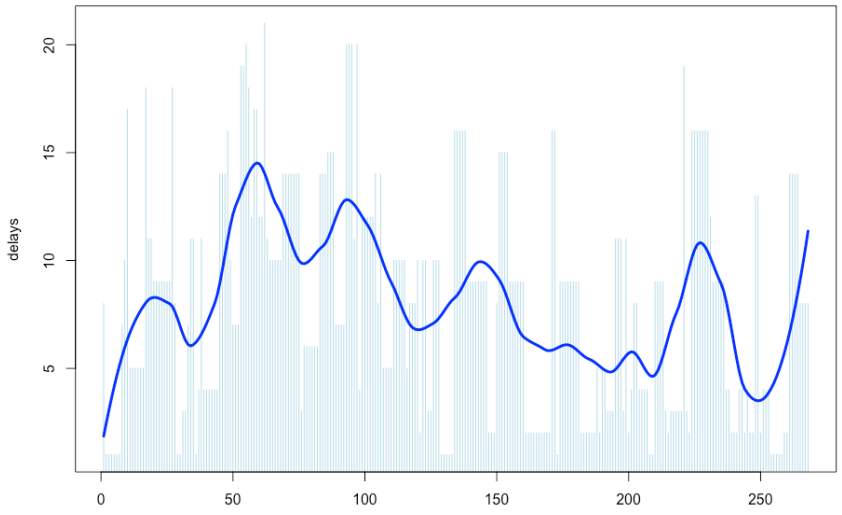


Figure 6 shows the delay in the maximum effect of mobility on PR for Montreal. The dark blue line is obtained by a simple smoothing showing that delays in the effect of mobility restrictions are relatively shorter when during the 2nd wave. The table below summarizes the correlations for both cities. The number of sparsified correlations are indicated by NA's indicating decoupling around 24 percent of the time between PR and mobility.

Table 1: Summary of maximum correlations

	Montreal	NYC
Min.	0.566	0.495
1st Qu.	0.732	0.709
Median	0.821	0.799
Mean	0.816	0.823
3rd Qu.	0.904	0.911
Max.	0.998	1
NA's	66	61
%.NA's	24.6	23.2

5.1 Elasticies

The correlation captures the degree of relatedness between PR and mobility but cannot reveal the responsiveness of PR or to degree of which PR changes in response to changes in mobility. This is also called elasticity and can be defined as follows:

$$x\text{-elasticity of } y\text{: } \epsilon = \frac{\partial y/y}{\partial x/x},$$

which is the ratio of the percentage change in one variable to the percentage change in another variable, when the latter variable has a causal influence on the former. It is a useful tool for measuring the sensitivity of one variable to changes in another, causative variable. It also has the advantage of being a unitless ratio, independent of the type of quantities being varied. The concept of elasticity is also related to two other statistics that measure the linear association between two series, a slope coefficient in a single-variable regression, $y_i = \alpha + \beta x_i + \varepsilon_i$, and correlation coefficient that we calculated in the previous sections. The slope coefficient can be expressed as

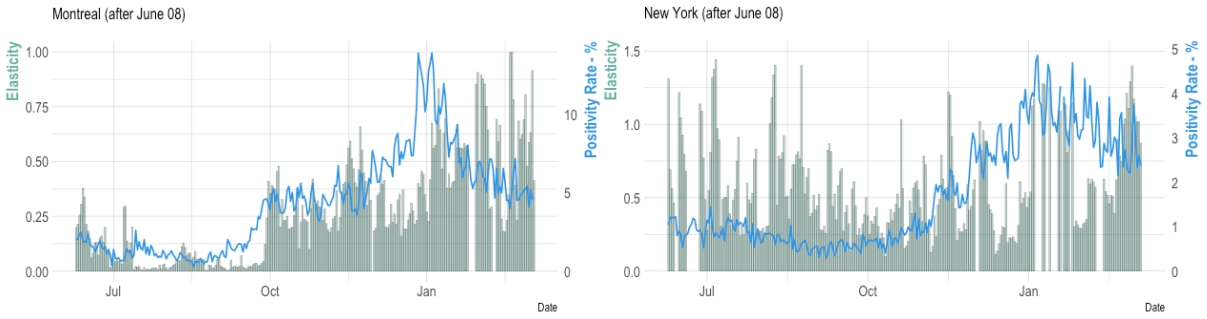
$$\begin{aligned}\beta &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\ &= \frac{s_{x,y}}{s_x^2} \\ &= r\frac{s_y}{s_x}\end{aligned}$$

where $s_{x,y}$ is a simple covariance between x and y, s_x^2 is the variance of x, r is the coefficient of correlation between the y and x, and s_y and s_x are their standard deviations, respectively. Consequently, the "mean" elasticity, ϵ, may be written with $y = PR$ (Positivity Rate) and $x = R$ (Restrictions) as follows:

$$\epsilon = \frac{\partial PR/PR}{\partial R/R} = r\frac{s_{pr}}{s_r}\frac{\overline{R}}{\overline{PR}}$$

It follows, therefore, that when r is in the neighborhood of 1, the spread will be more sensitive or less (i.e., $\epsilon \lessgtr 1$) depending on two facts: the spread of COVID-19 is more or less variable than the mobility ($\frac{S_{PR}}{S_R}$) and the magnitude of restrictions relative to how widespread PR is ($\frac{\overline{R}}{\overline{PR}}$). Figure 7 shows the elasticities for two cities.

Figure 7: Elasticities



The mean elasticity, 0.29, shows that PR is not sensitive to mobility changes on average when we consider the whole period after June 8. At the beginning of the second wave, however, it changes and becomes 0.40. which implies that, when the mobility goes down 10%, PR falls, on average, 4% during the second wave. Although New York City have lower correlation coefficients around 0.75, its elasticity is higher about 0.72.

5.2 Counterfactual Elasticies

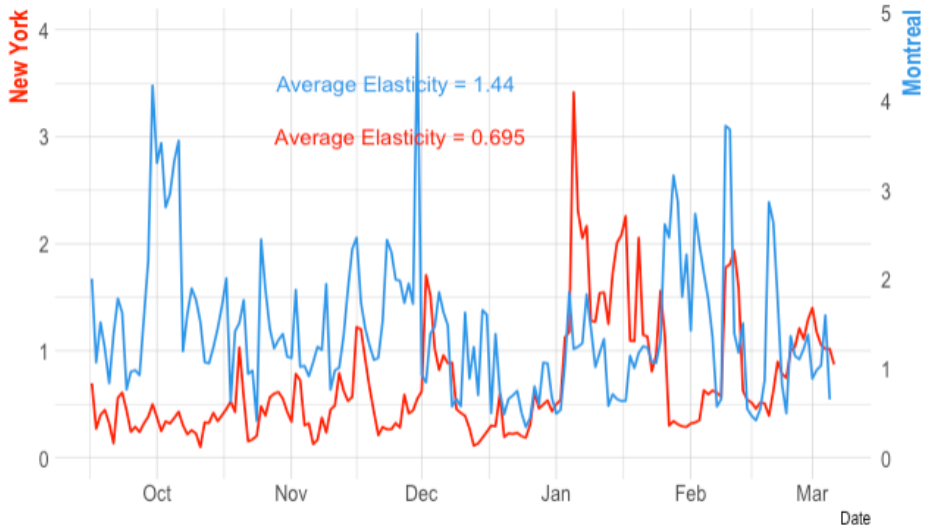
We use counterfactual elasticities to compare the effectiveness of mobility restrictions in Montreal against NYC. We create a hypothetical case where we calculate elasticities for Montreal using the data for NYC. This shows how effective the mobility restrictions would have been had Montreal had the same PR and mobility measures as NYC during the 2$^{\text{nd}}$ wave between September 16, 2020 and March 23, 2021.

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 ($\overline{R}/\overline{PR}$); (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. These facts can be verified by tables provided in Figure 8.

Figure 8: Elasticity of Montreal's PR in NYC

(Between September 16, 2020 & March 23, 2021)



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**

Measures are calculated for each 7-day sliding window and averaged over the entire period, Counter-factual elasticity, 1.44, is [Beta (Montreal) X mean(R)/mean(PR) (NYC)] calculated for each sliding window and averaged over the entire period.

5.3 Robustness Check

The main concern is whether the correlation matrix from two completely independent random series would be distinguishable from the one that we obtain using our mobility and PR series. We generated three different surrogate data sets with 5000 runs in each. The results and the codes will be available upon request

6 Conclusion

Multiple studies found that weather components are associated with the transmission of the virus. However, there is conflicting evidence on the influence of weather on COVID-19 transmission. To estimate weather-dependent signatures of the pandemic, the effects of socio-economic factors and non-pharmaceutical interventions must be controlled for. The delay between exposure and detection of infection complicates the estimation of weather impact on COVID-19 transmission, potentially explaining significant variability in results to-date. One of the first studies (Xu et al. 2020) on the subject using the county-level data with more than 3700 counties in the U.S found that failing to control for the delay between infection and official recording of cases is the main reason for the mixed evidence. They conclude that the delay, which is a particularly understudied factor and estimated to be approximately 10 days, confounds the attempts to associate daily weather conditions with recorded new cases and may partially explain the inconsistent and inconclusive findings to date.

Although the evidence unambiguously indicates that NPIs with successful mobility restrictions are the most effective tool in curbing the pandemic, studies looking at the dynamics of these confinement policies are rare. While convincing reductions in case numbers have been observed in many regions after rapidly implemented mobility restrictions, spatial differences in their efficacy, even across the neighboring regions with the same interventions, require understanding how the relationship between transmission and mobility changes over short time and space.

Despite the incredible effort of many scientific communities, studies capturing the short-term temporal dynamics in the relationship between mobility and the spread are absent. Most studies investigating the effect of lockdown policies on the incidence of COVID-19 assume that the delay between mobility restrictions and their effect on the spread is constant over time and space. Although compartmental epidemic models are able to incorporate the incubation period in their analyses by modeling the associated uncertainty based on some parameterized probability distributions, the results cannot reveal the dynamics of the time-varying relationship between mobility and the spread.

We develop and apply a novel quantitative method to understand disease dynamics that help policy makers plan interventions to taper viral transmissions exemplified by the ongoing effort to stop the spread of COVID-19. Sparse composite likelihood selection is a recently developed method focusing on identifying true sparsity in complex models in case of high-dimensional data without compromising its asymptotic properties such as unbiasedness. The algorithm has a number of advantages including a rigorous sparsified solution to graphical network analysis of complex systems and the ability to efficiently handle high-dimensional data such that it is computationally efficient and statistically sound. We demonstrate that, unlike conventional shrinkage methods, when sparse composite likelihood selection is applied in a

graphical network setting, it can capture the time-varying effects of mobility restrictions on the COVID-19 spread in Montreal and New York. For each case, we describe the utility of the method for surveillance and resource allocation.

References

Abbott S, Hellewell J, Sherratt K, Gostic K, Hickson J, Badr HS, DeWitt M, Thompson R, EpiForecasts, Funk S (2020). EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters. doi: 10.5281/zenodo.3957489.

Askitas, N., Tatsiramos, K. & Verheyden, B. 2021. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. Sci Rep 11, 1972 (2021). https://doi.org/10.1038/s41598-021-81442-x

Banerjee, O., El Ghaoui, L., d'Aspremont, A., 2008. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. J. Mach. Learn. Res. 9, 485–516.

Blackwood, Julie, and Lauren Childs. 2018. An Introduction to Compartmental Modeling for the Budding Infectious Disease Modeler. Letters in Biomathematics 5 (1), 195-221. https://doi.org/10.1080/23737867.2018.1509026.

Boker SM, Xu M, Rotondo JL, King K. 2002. Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. Psychol Methods. 2002 Sep;7(3):338-55. doi: 10.1037/1082-989x.7.3.338. PMID: 12243305.

Born B, Dietrich AM, Müller GJ. The lockdown effect: A counterfactual for Sweden. PLoS One. 2021. 16(4). doi:10.1371/journal.pone.0249732

Caterina, C.D., and Ferrari, D. 2021. Sparse composite likelihood selection. https://arxiv.org/abs/2107.09586

Cho, S.-W. S. 2020. Quantifying the impact of nonpharmaceutical interventions during the covid-19 outbreak: The case of Sweden. The Econometrics Journal. 23(3), 323-344.

Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S. 2013. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. American Journal of Epidemiology, 178(9), 1505–1512.

David Hevey (2018) Network analysis: a brief overview and tutorial, Health Psychology and Behavioral Medicine, 6:1, 301-328, DOI: 10.1080/21642850.2018.1521283

Dean RT, Dunsmuir WT. 2016. Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. Behav Res Methods. 2016 Jun;48(2):783-802. doi: 10.3758/s13428-015-0611-2. PMID: 26100765.

Efron, B. 2004. Large-scale simultaneous hypothesis testing. Journal of the American Statistical Association 99, 96–104.

Efron, B., 2010. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Cambridge University Press, Cambridge.

Efron, B., Tibshirani, R., Storey, J.D., Tusher, V., 2001. Empirical Bayes analysis of a microarray experiment. J. Amer. Statist. Assoc. 96, 1151–1160. Friedman, J., Hastie, T.,

Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441.

Friedman, J., Hastie, T., Tibshirani, R., 2011. Glasso: Graphical lasso-estimation of Gaussian graphical models, R package, version 1.7. URL: http://CRAN.R-project.org/package=glasso.

Fang, H., Wang, L., Y, Yang. 2020. Human Mobility Restrictions and the Spread of the Novel Coronavirus (2019-nCoV) in China. NBER Working Paper Series
26906 http://www.nber.org/papers/w26906

Flaxman, S., Mishra, S., Gandy, A. et al. 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 584, 257–261.

Fu, W.J., 1998. Penalized regressions: The bridge versus the lasso. J. Comput. Graph. Statist. 7, 397–416.

Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, et al. 2020. Practical considerations for measuring the effective reproductive number, Rt. PLOS Computational Biology 16(12): e1008409. https://doi.org/10.1371/journal.pcbi.1008409

Ha, M.J., Sun, W. 2014. Partial Correlation Matrix Estimation Using Ridge Penalty Followed by Thresholding and Re-estimation. Biometricss, 70, 765-773.

Herby, Jonas. 2021. A First Literature Review: Lockdowns Only Had a Small Effect on COVID-19
SSRN: https://ssrn.com/abstract=3764553 or http://dx.doi.org/10.2139/ssrn.3764553.

Huang, Z., & Ferrari, D. (2021). Fast construction of optimal composite likelihoods. https://arxiv.org/abs/2106.05219

Locatelli I, Trächsel B, Rousson V. 2021. Estimating the basic reproduction number for COVID-19 in Western Europe. PLOS ONE 16(3):
e0248731. https://doi.org/10.1371/journal.pone.0248731

Nicola Perra, 2021. Physics Reports, Non-pharmaceutical interventions during the COVID-19 pandemic: A review. 913, 1-52.

Nouvellet, P., Bhatia, S., Cori, A. et al. 2021. Reduction in mobility and COVID-19 transmission. Nat Commun 12, 1090

Schäfer, J., Strimmer, K., 2005a. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat. Appl. Genet. Mol. Biol. 4, art. 32.

Schäfer, J., Strimmer, K., 2005b. An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics 21, 754–764.

Xu R, Rahmandad H, Gupta M, DiGennaro C, Ghaffarzadegan N, Jalali MS. Weather Conditions and COVID-19 Transmission: Estimates and Projections. https://scholar.harvard.edu/jalali/publications/weather-conditions-and-covid-19-transmission-estimates-and-projections

Xu R, Rahmandad H, Gupta M, DiGennaro C, Ghaffarzadegan N, et al. 2021. Weather, air pollution, and SARS-CoV-2 transmission: a global analysis. Lancet Planetary Health. 5(10): 671-680.

Varin, C., Reid, N., & Firth, D. 2011. An overview of composite likelihood methods. Statistica Sinica, 21(1), 5–42. http://www.jstor.org/stable/24309261

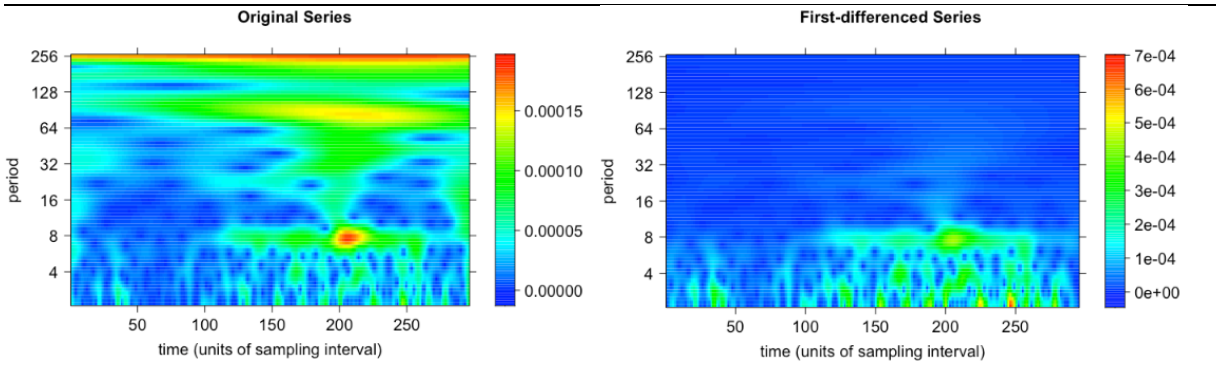
Wieringen, W.N.v, Peeters, C.F.W, 2016. Ridge estimation of inverse covariance matrices from high-dimensional data. Computational Statistics and Data Analysis, 103, 284-303

Appendix

Window size

Wavelets allow us to study localized periodic behavior. In particular, we look for regions of high-power in the frequency-time plot. The intensity of the colormaps in Figure 13 represents the variance of the time series that is associated with particular frequencies (y-axis) through time (x-axis). Our wavelet analysis is able to detect frequencies that are localized in time, and therefore if the dominant period of a time series changes over time, wavelets can be used to detect this transition. The map shows that around days 7 and 8, the second wave shows a dominant variation. We ignore the higher variations around day 100 which captures the increasing variations during the first and second waves of the epidemic. This is also captured by the spectral analysis applied on the first-differenced PR and mobility series after Day 100, which indicates the same frequency, 7 days, in both series.

Figure A1: Wavelet heatmaps



We also run a spectral analysis that supports the finding of the wavelet analysis. The results and the codes can be found in the same html file