**Evaluating instrumental variable methods as a benchmark for fMRI connectivity analysis**

Rotation advisor: Konrad Kording

Ethan Blackwood

Fall 2019

**Question**

Instrumental variable analysis is a way to avoid confounding in estimates of causality between observed variables when an experimental manipulation is infeasible. Could this method improve the accuracy of regional functional connectivity estimates in resting-state functional magnetic resonance imaging (fMRI)?

**Background**

The method of instrumental variables (IVs) has its origins in econometrics, but is now more widely used to infer information about causality from observational data (Angrist and Pischke, 2009). When applicable, it is much simpler and requires less information than full causal inference analysis. For IV analysis, there must be an observable factor (the “instrument”, Z) that correlates with an explanatory variable of interest (X). When X is held constant, Z must also be independent of the effect of interest (Y) and any variable that correlates with Y. If these constraints hold, Z can be used to estimate the causal (though not necessarily direct) effect of X on Y. So far, a method using IVs has been developed to disambiguate the results of optogenetic stimulation with single unit recording (Lepperød et al., 2018), but no study to our knowledge has explored their use in lower-resolution modalities such as fMRI.

In fMRI studies, a frequent goal is to find which pairs or networks of brain regions communicate in a particular setting. This “functional connectivity,” may formally be defined as correlation between neural events in different parts of the brain, but in practice denotes a family of related measures (Li et al., 2009). The simplest such measure is cross-correlation of the blood-oxygen-level-dependent (BOLD) time series between regions, either with or without a time lag. The interpretation of these correlations, however, is fraught with uncertainty, due to both the low temporal resolution and precision of the BOLD signal and the fact that spurious correlations can easily arise from confounding variables. Although the resolution problem is currently unavoidable, a method like IVs that is designed to avoid the pitfalls of correlations when analyzing causality might be able to reduce confounding.

**Hypothesis**

If true instrumental variables for the (in)activation of regions of interest are available, estimates of functional connectivity using IV analysis will be closer (better correlated) with the real connectivity structure of a network than those obtained by cross-correlation. Due to omitted variables bias, methods based on correlation alone will be more susceptible to spurious connectivity as the proportion of unobserved variables in the network increases.

**Methods**

*Network generation:* For preliminary assessment of connectivity measures, we generated networks of 6-120 “brain regions” with known connectivity. For each trial with *k* regions, we:

* Drew *k*2 entries from the uniform random distribution over the range [-0.5, 0.5);
* Masked this matrix with a random *k* x *k* directed small-world matrix (Watts and Strogatz, 1998; Song and Wang, 2014) with edge density and rewiring probability equal to 0.5;
* Rescaled the result to have a spectral radius (maximum eigenvalue magnitude) of 0.75.

*Network simulation:* For each of 20 network sizes *k* (6-120, stepping by 6) and 2000 trials, we simulated 2500 steps of the first-order multivariate autoregressive process defined by:

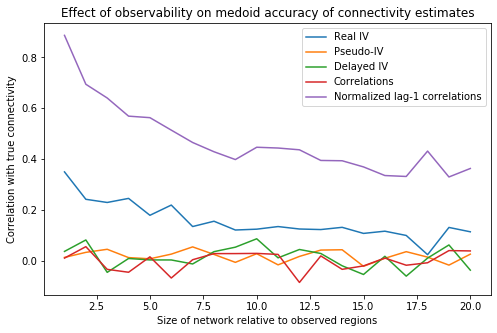
Here, *A* is the network’s connectivity matrix, the noise term , and the instrumental variable for . ( represents the natural logarithm of each region’s activity, but this logarithm was used directly in all analyses.) The instrumental variable term stands in for a stochastically recurring event that depresses one region but is independent of other regions until the next timepoint. To simulate temporal and spatial subsampling, only every fifth timepoint and the first six regions were kept observable during analysis.

*Connectivity estimation:* We used five methods to recover estimates of connectivity:

* Real IV: Use to estimate the effect of on .
* Delayed real IV: Use to estimate the effect of on .
* Pseudo-IV: Define , where and are the mean and standard deviation respectively of . Then use as an IV to estimate the effect of on .
* Correlation: Pearson correlation coefficient between each pair of regions with zero lag.
* Lag-1 correlation: Normalize the covariance of with by the covariance of .

Finally, we correlated connectivity estimates with the expected causal effect of on ,

**Results**



**Figure 1.** The real IV estimated connectivity less accurately than lag-1 correlations, but better than other tested IV and correlation methods.

A parameter exploration with 10 fully observable regions showed that real IV estimates correlated better with the expected effect when the IV had a stronger influence. The network’s signal-to-noise ratio (SNR; variance of the term relative to all variance) had a minor effect. Lag-1 correlations reproduced the expected effect much more accurately than other methods.

Since the omitted variables bias only causes confounding when there are unobserved regions, next we tried varying the fraction of regions that were observable during analysis, while keeping the number observable at six. We fixed the effect of the IV at and controlled the spectral radius rather than SNR of connectivity matrices (due to computational complexity). Unexpectedly, the lag-1 correlation method continued to outperform real IV according to median results, even when only 5% of regions were observable, although there does appear to be a steeper loss of accuracy as observability decreases for lag-1 correlation (Figure 1).

**Conclusions**

If the simulation was reasonably able to approximate activation dynamics of brain regions, our results suggest that IV analysis of BOLD timeseries would be less accurate than lagged correlation. Furthermore, it is not clear whether any measurement from fMRI data would be a valid instrument. However, this finding is inconclusive due to several assumptions in the model (e.g., that regional connectivity is small-world and the resulting dynamics are first-order autoregressive, only up to 95% unobservable, and evolve relatively little between observations).

Though the results here used no real fMRI data, we did identify a publicly available dataset complete with regional activation timeseries, and future research could focus on assessing the candidate methods on these data. Although the true regional connectivity is unknown with real data, one strategy would be to penalize results that assign high weights to connections that are anatomically unlikely. One could also construct a prior on resting state connectivity from the consensus results of several methods. Despite the advantages of the modality for human studies, fMRI datasets continue to be a challenge to analyze rigorously.

**References**

Angrist JD, Pischke J-S (2009) Mostly Harmless Econometrics: An Empiricist’s Companion. Princeton, NJ: Princeton University Press.

Lepperød ME, Stöber T, Hafting T, Fyhn M, Kording KP (2018) Inferring causal connectivity from pairwise recordings and optogenetics. bioRxiv Available at: http://biorxiv.org/lookup/doi/10.1101/463760 [Accessed June 1, 2019].

Li K, Guo L, Nie J, Li G, Liu T (2009) Review of methods for functional brain connectivity detection using fMRI. Computerized Medical Imaging and Graphics 33:131–139.

Song HF, Wang X-J (2014) Simple, distance-dependent formulation of the Watts-Strogatz model for directed and undirected small-world networks. Physical Review E 90 Available at: https://link.aps.org/doi/10.1103/PhysRevE.90.062801 [Accessed December 30, 2019].

Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. Nature 393:440–442.