STAT 8101: Applied Causality (Spring'22) Week 5: Feb 21, 2022 - Feb 27, 2022

Paper: Matrix Completion Methods for Causal Panel Data Models

Han Lin hl3199@columbia.edu MS in Computer Science

This paper synthesis several methods (horizontal regression, vertical regression, synthetic control, nuclear norm matrix completion, difference in difference regression) under the same structure with the same minimization objective functions and different constraints/regularizations.

Introducing horizontal and vertical regressions by first illustrating fat/thin matrix is quite good for understanding. Regarding synthetic control as a vertical regression with additional constraints on non-negative coefficients and zero intercept helps building connections between these two methods.

Staggered Adoption seems to be more general, since units could have different starting time of being treated, which should have lots of application scenarios.

The concentration bound in section 6.2 is quite illuminating. The concentration quality depends on the proportion of un-treated entries as well as the rank of the decomposition of matrix L, which is interesting and intuitive.

In section 7, the first california smoking data experiments shows that vertical regression is not good when the proportion of un-treated data is low, which is reasonable to me since there are not enough periods for the model to find a good combination of donor pools that could represent the target unit. But the second stock market data is interesting, it seems that the proposed matrix completion algorithm could be consistently perform well on both thin and fat matrices, which is a strong result in my opinion.

In section 8.2, the paper mentions that we could also do the opposite to imputing Y(0). The condition of constant treatment effect seems strong, but the condition of low rank pattern in Y(1) seems reasonable to me.

In section 8.3, the paper generalizes to the situation when the error terms epsilon(i,t) are correlated over time. As for computation (not so relavent with causal inference), the autocorrelation matrix Omega is a Toeplitz matrix, which should support fast matrix-vector multiplication.