Discussion of
Job Mobility Networks
and Endogenous Labor Markets
By Jan Sebastian Nimczik

Isaiah Andrews

Harvard Economics

July 26, 2018

Overview

- The paper estimates endogenous labor markets based on worker mobility patterns
- Based on heterogenous stochastic block model (SBM)
    - Firms are in the same market if they tend to send workers to (and receive workers from) similar firms
    - Firms assigned to markets by (approximately) maximizing a joint Poisson likelihood
- Paper derives SBM model from simple model of firm choice

Plan for Discussion

- A concern: endogenous labor markets are (at best) noisily estimated
  - Estimation errors may be correlated with outcomes
  - Model misspecification
  - Hard optimization problem
- Potential solution: use endogenous labor markets estimated with data from previous periods
  - If assume enough independence over time, may address concerns with estimation error
  - Spoiler: this is already done in both applications!
    - Think this is important part of proposal, and best to do throughout (e.g. for comparisons to benchmark labor markets as well)

Consistent Estimation?

- Asymptotic results (Zhao et al 2012) imply consistency of SBM ML estimates
  - But require

    $$\text{\# of movers per firm} \to \infty \text{ or } \frac{\text{\# of movers per firm}}{\log\left(\text{\# of firms}\right)} \to \infty$$

    depending on desired result
  - # of movers per firm in data ≈ 10. Are we in asymptopia?
- Without consistency, have to worry about estimated labor markets
  - Concern: same factors may drive both worker moves and outcomes of interest
  - For studying outcomes of ultimate interest, estimated endogenous labor markets may be... endogenous

Other Concerns

- Consistency results also assume SBM model is right
  - Parametric assumptions $+$ k distinct local labor markets
  - If only approximate, what happens?
- Estimation of SBM model hard
  - Discrete optimization- $k^n/k!$ possibilities
  - Consistency results assume have found global max
  - Results in paper use stochastic greedy algorithm. Seems reasonable choice, but we'd need a lot of luck to find global max

Possible Solution

- Potential solution: use estimates based on lagged mobility
- Suppose we have a panel, and assume m-dependence
  - Observations whose time indices differ by $m + 1$ or more are independent
  - Independence across t: $m = 0$
- Suppose that for regression with data from t, use labor markets estimated with data from $\tau \leq t - m - 1$
  - A form of sample splitting
- Under m dependence, condition on estimates and treat as fixed
  - Once condition, analogous to treatment of local labor markets based on e.g. geographical boundaries

Using Lagged Estimates

- Advantage: allows estimation errors in labor markets
  - ...but subsequent estimates conditional on estimated labor markets
  - Analagous to how conventional LLM estimates conditional on chosen LLM definition
- Advantage: model agnostic
  - If (when) SBM estimates incorrect, recover projection coefficients in same sense as traditional LLMs with incorrectly specified boundaries
- Disadvantage: m-dependence is a strong assumption
  - More restrictive on time-dependence than e.g. clustering by firm

An Alternative Approach

- To take SBM model seriously, rather than using lagged estimates, we could go full Bayes
  - Estimate SBM model for labor markets alongside outcome equations
- Advantage:
  - Do inference on "true" model parameters
  - Automatic summaries of uncertainty
  - Uncertainty in SBM estimates propagated to uncertainty for parameters of interest
- Disadvantages:
  - Depends heavily on model
  - Thanks to joint estimation, labor market estimates may be driven by outcomes of interest, rather than mobility

Implications for Paper

- Paper already uses lagged data to estimate endogenous labor markets in applications
  - Great!
  - May be worth thinking about how many lags want
- Would be good to use throughout
  - e.g. in comparison of endogenous labor markets to existing alternatives. Two advantages:
    - Mitigates concern about estimation error
    - Shows that good performance carries over to likely use-case

Conclusion

- Paper makes interesting proposal
- SBM model based on worker mobility seems an appealing way to estimate labor market boundaries
    - ... but introduces concern about estimation error
- Possible solution: use lagged values
    - Already done in applications