ECON 833 - Computational Methods for Economists

Fall 2021

Notes on Maximum Score Estimation

Estimation of Two-sided matching problems

- What we want to do is to be able to estimate the structural parameters of this matching model
  - We observe the matches that happen
  - We observe characteristics of the agents who are matched
  - From this we can infer the payoffs to the match for the agents
- The structure of the data
  - The structure of the data affects the asymptotic of the estimator
  - e.g., do we observe all matches in a market - so that the asymptotics depend on observing more markets?
  - Or do we observe only some of the matches in a large market - so that the asymptotics depend on observing more matches in this market?

Maximum Score Estimation

- The maximum score estimator (MSE) was introduced by Manksi (*Journal of Econometrics*, 1975)
- The maximum score estimator is similar to the maximum likelihood estimator
  - The MLE maximizes the likelihood function
  - The MSE maximizes the "score function"
- The score function is the number of observations correctly predicted by the discrete choice model:

$$Q_n(\beta) = \sum \mathbb{1}\left[f(X;\beta)\right] \tag{1}$$

  - Note: the argument inside the indicator function will vary depending upon the discrete choice model and the data
- The maximum score estimator is identified by the property that the outcomes for model agents can be rank-ordered by the deterministic part of their payoff function.
  - Unlike a maximum likelihood estimator, this stochastic structure does not need to be parameterized
  - This estimator uses the idea of revealed preference: if one option was chosen over another, then that one is revealed preferred to the others and must be ahead of those in rank order of the agent's preferences
- Maximum score estimators can be used for any discrete choice model, but maybe most helpful for models with matching since they can help to capture the equilibrium effects.
- Pros of MSE relative to MLE:
  - Do not need strong assumptions on the distribution of the error term
  - Captures equilibrium effects in a matching model
  - Simple, intuitive statistical objective function that is easy to compute

- Cons of MSE relative to MLE:
  - Convergence is slower
  - Inference is more difficult

<u>The Estimator</u>

- Notation for representing matches:
  - Let m refer to the market
  - Let i refer to the "upstream" agent in the match
  - Let j refer to the "downstream" agent in the match
  - Ordering is not important
  - A one-to-one match is then identified by the tuple $\{m, i, j\}$
  - For many-to-many matches, use $\{m, \{i_1, i_2, ...\}, \{j_1, j_2, ...\}\}$ to denote the match
- The payoff function:
  - Let $f_\beta(m, i, j)$ represent the payoff for the match of i and j in market m given parameter vector β
  - The represents the value of the match to for upstream agent i matching with downstream agent j in market m
  - $f_\beta(m, j, i)$ is the value for the agent on the other side of the match, i.e., looking at this match from the perspective of j matching with i
- The statistical objective function:
  - For one-to-one matches:

$$Q(\beta) = \sum_{m \in M} \sum_{i \in U_m} \sum_{j \in U_m \setminus i} \mathbb{1}\left[f_\beta(m, i, \mu_m(i)) + f_\beta(m, j, \mu_m(j)) > f_\beta(m, i, \mu_m(j)) + f_\beta(m, j, \mu_m(i))\right] \tag{2}$$

    - M is the set of markets
    - U_m is the set of upstream agents in market m
    - $\mu_m(i)$ is the function to reference the downstream partner of the upstream agent in market m
  - For many-to-many matches (more general formulation that can also account for 1-to-1 matches):

$$Q(\beta) = \sum_{m \in M} \sum_{\{C^{LHS}, C^{RHS}\} \in I_m} \mathbb{1}\left[\sum_{\vec{a} \in C^{LHS}} f_\beta(\vec{x}_a) > \sum_{\vec{a} \in C^{RHS}} f_\beta(\vec{x}_a)\right] \tag{3}$$

    - I_m are the sets of inequalities to be compared
    - Each element of I_m is a pair of sets of coalitions, $\{C^{LHS}, C^{RHS}\}$, one of which is observed and one of which is hypothetical
    - $\vec{x}_a$ are the covariates of coalition a
- Normalization
  - The objective function needs to be normalized for anything that requires it have an asymptotic limit
  - There are two ways to do this:

1. If complete markets are observed and new observations come from observing new markets: divide $Q(\beta)$ by M, the number of markets
2. If there is one big market and new observations come from observing more matches in this market: divide $Q(\beta)$ by $H(H-1)$, where H is the number of matches observed
   * Note $H(H-1)/2$ is the number of comparisons that will be performed in the nested sums

- The Smooth Maximum Score Estimator
  - The MSE estimator outlined above will have discrete jumps as the parameters are changed and the inequalities flip
  - This can pose problems for the numerical optimization routines used to maximize the score function and also for statistical inference
  - Thus, Horowitz (*Econometrica*, 1992) proposed a smoothed MSE
  - To do this, replace the indicator function in the score function with a kernal, $K(\cdot)$
  - $K(\cdot)$ has the following properties
    * $K(v)$ is finite valued for all v
    * $\lim_{v\to-\infty} K(v) = 0$
    * $\lim_{v\to\infty} K(v) = 1$
  - With this, we have a CDF-like kernal
  - With the smooth MSE, then with pairwise matches, inequalities that are not satisfied receive little weight and those satisfied by a large margin receive the most weight
  - Another nice property of the smoothed estimator is that bootstrapped standard errors are consistent (Horowitz (*Journal of Econometrics*, 2002))

<u>Inference</u>

- Constructing standard errors or confidence intervals for the MSE is not entirely straightforward.
- Boot-strapping is not consistent (in general) and one needs to account for the convergence properties of this estimator.
- But there are at least a couple methods:
  1. Subsampling confidence intervals
     - Pick S random subsamples, each with size $n_s <$ the number of observations in the full data set, where s denotes a particular subsample
     - Find the maximum score estimate for each s, $\hat{\beta}_s$
     - The approximate sampling distribution for the parameter vector can be computed as

$$\tilde{\beta}_s = (n_s/N)^{1/3}(\hat{\beta}_s - \hat{\beta}_{full}) + \hat{\beta}_{full} \tag{4}$$

       * $\hat{\beta}_{full}$ is the MSE from the full sample
       * This procedure accounts for the $\sqrt[3]{N}$ convergence property of the MSE
     - To compute the 95% confidence interval, take the 2.5th and 97.5th percentiles from the distribution of $\tilde{\beta}_s$
  2. Bootstrapping standard errors/confidence intervals
     - Is consistent with a smoothed MSE!

<u>An Example:</u>

- "The Determinants of Bank Mergers: A Revealed Preference Analysis" by Akkus, Cookson, and Hortaçsu (*Management Science*, 2016)
- Authors seek to understand how value is created from bank mergers (e.g. from reducing regulatory compliance costs, network effects, etc.)
- Framework - The matching model
  - The authors propose a one-to-one matching model where buyer banks look for target banks to purchase
  - The authors assume one national market per year and that markets in different years are independent of one another
  - There are M^y matches in the market in year y
  - The value of the merger is the joint value to the buyer and the target bank:

$$f(b,t) = V_b(b,t) + V_t(b,t) \tag{5}$$

  - The payoff to the buyers is the post-merger valuation less the price paid to acquire the target, $V_b(b,t) = f(b,t) - p_{bt}$
  - The target's payoff is equal to their purchase price, $V_t(b,t) = p_{bt}$
  - Each buyer maximizes $V_b(b,t)$ across targets
  - Each target maximize $V_t(b,t)$ across buyers
  - In equilibrium, all observed matches yield higher value than the counterfactual matches
    - Else we would have observed the counter factual matches
    - Note this use of revealed preference
  - This means that if buyer b bought target t and not target $t^{'}$, then:

$$\begin{aligned} V_b(b,t) &\geq V_b(b,t^{'}) \\ f_b(b,t) - p_{bt} &\geq f(b,t^{'}) - p_{bt'} \end{aligned} \tag{6}$$

  - ISSUE: counterfactual transactions don't happen, so never observe $p_{bt'}$
  - Solution: In equilibrium $p_{b't'} = p_{bt'}$
    - i.e., the counterfactual price is the price the bank who did buy that target pays for the target
  - The same inequalities apply to buyer $b^{'}$ and we thus get the following set of inequalities:

$$\begin{aligned} f(b,t) - f(b,t^{'}) &\geq p_{bt} - p_{b't'} \\ f(b^{'},t^{'}) - f(b^{'},t) &\geq p_{b't'} - p_{bt} \end{aligned} \tag{7}$$

  - We can use the above inequalities to construct a score function, but this is problematic if we don't observe the purchase prices
  - In this case, add the two inequalities in Equation 7 together to obtain:

$$f(b,t) + f(b^{'},t^{'}) \geq f(b^{'},t) + f(b,t^{'}) \tag{8}$$

  - In words, this inequality says that the total value from any observed matches is at least as great as the value from any counterfactual matches.
- Parameterization of the payoff function:
  - The value of buyer bank b purchasing target t is given by:

$$f(b,t) = \beta_1 W_b W_t + \gamma_1^{'} X_t + \gamma_2^{'} X_{bt} + \varepsilon_{bt} \tag{9}$$

        * W_b are buyer characteristics and W_t are these same characteristics for the target
            · e.g., measures of bank size such as assets or number of branches
        * X_t are target specific covariates
            · e.g., market concentration in target market
        * X_{bt} are buyer-target specific characteristics (i.e., characteristics specific to the match)
            · e.g., market overlap
        * ε_{bt} is a match-specific error term that is independent across matches

- The max score estimator
    - With this, we can write the maximum score estimator as:

$$\hat{\beta} = \arg\max Q(\beta) = \sum_{y=1}^{Y} \sum_{b=1}^{M_y-1} \sum_{b'=b+1}^{M_y} \mathbb{1}\left[f(b,t|\beta) + f(b^{'},t^{'}|\beta) \geq f(b^{'},t|\beta) + f(b,t^{'}|\beta)\right] \tag{10}$$

    - A few notes about this:
        * Perfectly fine estimator to use in general
        * But:
            1. Leaving information on the table (if have pricing data)
            2. One of coefficients have to be normalize to one - all other coeffs interpreted relative that that effect
                · This is because the model is only identified up to a scale parameter.
                · When use prices, normalize the "coefficient" on price to one - so other coefficients interpreted as effect in dollars of the merger value
            3. If use data without prices *can't include buyer or target specific factors* – they cancel out as they are on both side of the inequality (see specifications above)
                · Note that acquirer specific information is differenced out whether or not price is used (see Equation 6).
    - The authors thus propose another max score estimator that does use the merger price information to overcome this.
    - The score function for this estimator is:

$$Q(\beta) = \sum_{y=1}^{Y} \sum_{b=1}^{M_y-1} \sum_{b'=b+1}^{M_y} \mathbb{1}\left[f(b,t|\beta) - f(b,t^{'}|\beta) \geq p_{bt} - p_{b't'} \wedge f(b^{'},t^{'}|\beta) - f(b^{'},t|\beta) \geq p_{b't'} - p_{bt}\right] \tag{11}$$

- Findings:
    - Bank mergers appear to be motivated by branching efficiency and competitive concerns
    - Not so much driven by pre-merger performance of the target or the threat of antitrust regulation
    - Through some counterfactual simulations, they look at the effects of dual charters on efficiency in the banking industry
        * The MSE suggests that banks are more likely to merge if they have the same charter
            · Two types of charters - national or state
            · National → federal regulator is the Federal Reserve
            · State → federal regulator is the FDIC
        * So some mergers don't happen because buyer doesn't want target with different charter
        * They use their estimated coefficients and the payoff functions to look at how merger value would change if all had same charter
        * Find that average merger value (i.e., welfare) increases in this case. Suggesting gains to deregulating (i.e., doing away with dual charters)