Computation of the MASC Estimator

Maxwell Kellogg[*] Magne Mogstad[†]

Guillaume A. Pouliot[‡] Alexander Torgovitsky[§]

December 20, 2019

Given a treated unit treated in period T and a set of control units, The estimator $\hat{\mu}_t^{masc}$ fit to outcomes up to period $t, t < T$ is defined as:

$$\hat{\mu}_t^{masc}(\phi, m) \equiv \phi \hat{\mu}_t^{ma}(m) + (1 - \phi)\hat{\mu}_t^{sc}$$

where $\hat{\mu}_t^{ma}(m)$ is an m-nearest neighbor estimator and $\hat{\mu}_t^{sc}$ is a standard synthetic control estimator, both fit to the path of outcomes up to period t. The model averaging parameter ϕ governs the weight placed on matching versus synthetic control, $\phi \in [0, 1]$.

Our implementation uses a rolling-origin cross-validation procedure to resolve the implicit trade off between interpolation bias versus extrapolation bias in choosing the tuning parameters m and ϕ. This involves making a series of one-step ahead forecasts, each of which is estimated only using data from periods prior to the forecast date. Mathematically, the criterion is:

$$Q(\phi, m) = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} (y_{1,f+1} - \hat{\mu}_{f+1}^{masc}(\phi, m))^2$$

where $y_{1,f+1}$ is the outcome of the treated unit in period $f + 1$, and $\mathcal{F}$ is a subset of time periods taken before the treatment period. There is a natural bias-variance trade off which drives the choice of folds f to include in $\mathcal{F}$. Later periods are preferred because they will likely be more relevant to the post-treatment period and will use more data. However, including earlier periods may reduce the variance in $Q(m, \phi)$. We recommend that the analyst pick a cutoff value f^*, defining $\mathcal{F}$ to include all $f \geq f^*$ in the pre-treatment period. The analyst must also choose the set of candidate nearest neighbor estimators from which to pick m.

Computationally, we solve the cross-validation problem in two steps. First, we fix the candidate nearest neighbor estimator m. Conditional on m, the cross-validation criterion has an analytic solution for ϕ using least square algebra:

[*]Department of Economics, University of Chicago.

[†]Department of Economics, University of Chicago; Statistics Norway; NBER.

[‡]Harris School of Public Policy, University of Chicago.

[§]Department of Economics, University of Chicago. Research supported by National Science Foundation grant SES-1846832.

$$\phi^*(m) = \frac{\sum_{f\in\mathcal{F}}(\hat{\mu}^{ma}_{f+1}(m) - \hat{\mu}^{sc}_{f+1})(y_{1,f+1} - \hat{\mu}^{sc}_{f+1})}{\sum_{f\in\mathcal{F}}(\hat{\mu}^{ma}_{f+1}(m) - \hat{\mu}^{sc}_{f+1})^2}$$

where the the real solution $\hat{\phi}(m)$ is then defined to respect the bounds on ϕ:

$$\hat{\phi}(m) \equiv \begin{cases} 0, & \text{if } \phi^\star(m) \leq 0 \\ 1, & \text{if } \phi^\star(m) \geq 1 \\ \phi^\star(m) & \text{otherwise} \end{cases}$$

For the second step, we then set $\hat{m} \equiv \underset{m}{\text{argmin}} Q(\hat{\phi}(m), m)$ and set $\hat{\phi} = \hat{\phi}(\hat{m})$. The cross-validated MASC estimator is a weighted average of $\hat{\mu}^{sc}_T$ and $\hat{\mu}^{ma}_T(\hat{m})$ with weights $(1 - \hat{\phi})$ and $\hat{\phi}$ respectively.