

# Lab 4: Using Targeted MLE and likelihood-based cross-validation to estimate the marginal effect of a risk factor - Due Dec. 4

November 13, 2006

This lab involves a modification of last lab (lab 3) and that is to 1) create a grid over sieves on both  $g(A | W) \equiv P(A | W)$  and  $Q(A, W) \equiv E(Y | A, W)$ , 2) use (targeted) likelihood-based cross-validation to choose the optimal combination of  $g$  and  $Q$  (for this lab, we only target the  $Q$  part of the data-generating distribution). You can do your own sieve, but I recommend sieves defined as the models of  $g(A | W)$  and  $Q(A, W)$  that start with the full models (all the  $W$ 's in) and then backwards select (i.e., take out the least significant variable and that defines the next backward step in the sieve for each). Come up with 5 or so models in this way (or another natural way) for both  $g$  and  $Q$  - this will provide 25 combinations over which to choose. Now, for each combination, do the one-step estimator to get the new  $Q_{i,j} = (A, W)(\epsilon)$ , where  $i$  and  $j$  index the initial estimators (from the sieve) of  $Q$  and  $g$ , respectively. Sum up the log-likelihood (over the validation sample) defined by  $Q_{i,j} = (A, W)(\epsilon)$  for each  $i, j$ . Pick the  $i, j$  that gives the maximum.

Include in your report:

- Your annotated code,
- the  $i, j$  chosen (of course, I need to know how you defined your sieve)
- the targeted MLE estimate at the optimal  $g$  and  $Q$ ,
- a short description of the results.