Posted by PER PETTERSSON-LIDBOM

Posted on 6/3/2014 1:17 PM

When I was preparing my lecture notes on RD kink design for my PhD students, I discovered that this paper misspecifies the functional form of the forcing variable, f(x-c), where x is the forcing variable and c is the cut-off., in their kink design The standard procedure is to interact the function f(x-c) with the indicator variable for treatment D where D=1 if x>c and D=0 if x<c. They have omitted the interactions terms for the second and third order polynomial functions, i.e., the terms (D(x-c)(x-c) and D(x-c)(x-c)(x-c) . This implies that they are simultaneously using information on both sides of the cut-off to estimate f(x-c), which does not sit well with the idea of a RD (kink or not) design, namely local identification at the cut-off without any parametric assumptions. Imposing this type of functional form restrictions is therefore not correct and may therefore yield biased estimate unless the restrictions are true. In order to test whether that is the case, I have re-estimated all their specifications in their Table 2 together with the interaction terms. Although, the two estimates for second-order specification are broadly similar to their estimates (even somewhat larger), the two estimates from the third-order specification are completely different as can be seen by comparing the results from Columns 1 and 2 in my Table with the corresponding results from their Table 2. My estimates are 0.84 for the full sample and -1.44 for the sample with a bandwidth of 15. These estimates should be compared with 3.35 and 4.08 in Table 2, respectively. This large decrease in their estimates clearly illustrates that their results are extremely sensitive to the degree of the polynomial specification. As a result, this new finding together with my previous comment on the extreme sensitivity to choice of bandwidth, the stark conclusion must be that there is no policy kink that can be exploited for identification (it is also noteworthy that only 5 out of 27 estimates in my table are positive and around 3 while 17 are negative, namely those specifications that are more credible since they rely on smaller bandwidths and more flexible functional forms of f(.).

Table