An Optimal Bandwidth For Difference-in-Difference Estimation with a Continuous Treatment and an Heterogeneous Adoption Design

Clément de Chaisemartin, Xavier d'Haultfoeuille

An Optimal Bandwidth For Difference-in-Difference Estimation with a Continuous Treatment and an Heterogeneous Adoption Design

Clément de Chaisemartin Xavier D'Haultfœuille

November 23, 2022

Setup We consider a design with two time periods, such that $D_1 = 0$ and $D_2 > 0$: no unit is treated at period 1, and all units receive a positive treatment dose at period 2. This corresponds to an heterogeneous adoption design, where all units adopt the treatment at period 2 but with varying treatment intensities. We assume that D_2 is continuously distributed on $\mathbb{R}_+$. Accordingly, $P(D_2 = 0) = 0$: there are no stayers such that $D_1 = D_2$. On the other hand, we assume that there are quasi-stayers: $f_{D_2}(0) > 0$ and $d \mapsto f_{D_2}(d)$ is continuous, where $f_{D_2}(.)$ is the density of the period-two treatment. The potential outcome at period t under treatment d is $Y_t(d)$, and the observed outcome is $Y_t := Y_t(D_t)$.

Target parameter Our target parameter is

$$\theta_0 := E[Y_2(D_2) - Y_2(0)],$$

the average effect, at period 2, of switching the treatment from 0 to its actual value. We conjecture that the estimation strategy we propose below would still apply, up to a normalization, if one were to consider instead

$$\frac{E[Y_2(D_2) - Y_2(0)]}{E[D_2]} = E\left[\frac{D_2}{E[D_2]} \times \frac{Y_2(D_2) - Y_2(0)}{D_2}\right],$$

a weighted average of the slopes of units' potential outcomes functions between 0 and their actual treatments, where units with a larger period-two treatment receive more weight. On the other hand, we conjecture that the estimation strategy we propose below would not readily extend if one were to consider $E[(Y_2(D_2) - Y_2(0))/D_2]$, the unweighted average of the slopes of units' potential outcomes functions between 0 and their actual treatments.

Identifying assumption and estimation strategy We assume strong exogeneity:

$$E[Y_2(0) - Y_1(0)|D_2] = E[Y_2(0) - Y_1(0)] =: \mu.$$

Then $\theta_0 = E[Y_2 - Y_1] - \mu$. μ is for instance identified by $E[Y_2 - Y_1|D_2 = 0]$, thus implying that θ_0 is also identified. However, estimating $E[Y_2 - Y_1|D_2 = 0]$ is not straightforward, as $P(D_2 = 0) = 0$: there are no stayers. Instead, we propose to use observations with D_2 lower than some bandwidth h to estimate $E[Y_2 - Y_1|D_2 = 0]$. We derive below an optimal bandwidth h, namely a bandwidth minimizing an asymptotic approximation of the mean-squared error of the resulting estimator of θ_0, in the spirit of the work of Imbens & Kalyanaraman (2012) for regression discontinuity designs.

Optimal bandwidth to estimate θ_0 We assume we have an iid sample $(D_{2i}, Y_{1i}, Y_{2i})_{i=1,...,n}$. We estimate θ_0 by:

$$\widehat{\theta}_h = \frac{1}{n}\sum_{i=1}^{n}(Y_{i2} - Y_{i1}) - \widehat{\mu}_h,$$

with $\widehat{\mu}_h$ the intercept in the local linear regression of $Y_{i2} - Y_{i1}$ on D_{i2} for the i's s.t. $D_{i2} \leq h$:

$$\widehat{\mu}_h = \frac{1}{n_h}\sum_{i:D_{i2}\leq h}(Y_{i2} - Y_{i1}) - \frac{\overline{D_{2h}}}{\widehat{V}_h(D_2)}\widehat{\text{Cov}}_h(Y_2 - Y_1, D_2),$$

where $n_h = \#\{i : D_{i2} \leq h\}$, $\overline{D_{2h}} = \sum_{i:D_{i2}\leq h} D_{i2}/n_h$ and

$$\widehat{V}_h(D_2) = \frac{1}{n_h}\sum_{i:D_{i2}\leq h}(D_{i2} - \overline{D_{2h}})^2,$$
$$\widehat{\text{Cov}}_h(Y_2 - Y_1, D_2) = \frac{1}{n_h}\sum_{i:D_{i2}\leq h}(D_{i2} - \overline{D_{2h}})(Y_{i2} - Y_{i1}).$$

We now derive asymptotic approximations of the bias and variance of $\widehat{\theta}_h$, conditional on $\boldsymbol{D} := (D_{12}, ..., D_{n2})$, when $n \to +\infty$ and $h \to 0$. Let $f(d) = E[Y_2 - Y_1|D_2 = d]$, so that $\mu = f(0)$ and assume that f is twice differentiable at 0. Then:

$$\begin{aligned}
-\text{Bias}(\widehat{\theta}_h|\boldsymbol{D}) &= \text{Bias}(\widehat{\mu}_h|\boldsymbol{D}) \\
&= \frac{1}{n_h}\sum_{i:D_{i2}\leq h} f(D_{i2}) - \frac{\overline{D_{2h}}}{\widehat{V}_h(D_2)}\frac{1}{n_h}\sum_{i:D_{i2}\leq h} f(D_{i2})(D_{i2} - \overline{D_{2h}}) - f(0) \\
&\simeq f(0) - f(0) + f'(0)\overline{D_{2h}} - f'(0)\frac{\overline{D_{2h}}}{\widehat{V}_h(D_2)}\widehat{V}_h(D_2) + \frac{f''(0)}{2}\overline{D_{2h}^2} \\
&- \frac{f''(0)\overline{D_{2h}}}{2\widehat{V}_h(D_2)}\widehat{\text{Cov}}_h(D_{i2}^2, D_{i2}) \\
&= \frac{f''(0)}{2}\left[\overline{D_{2h}^2} - \frac{\overline{D_{2h}}}{\widehat{V}_h(D_2)}\widehat{\text{Cov}}_h(D_{i2}^2, D_{i2})\right],
\end{aligned}$$

where the approximation follows from a Taylor expansion of order 2.

Now, remark that as $h \to 0$,

$$\begin{aligned} E[D_2^k|D_2 \leq h] =& \frac{\int_0^h u^k f_{D_2}(u)du}{\int_0^h f_{D_2}(u)du} \\ \sim& \frac{f_{D_2}(0)\int_0^h u^k du}{f_{D_2}(0)\int_0^h du} \\ \sim& \frac{h^k}{k+1}, \end{aligned}$$

where the equivalence follows by continuity of f_{D_2} at 0. This implies that $E[D_2^2|D_2 \leq h]/E[D_2|D_2 \leq h]^2 \to 4/3$ as $h \to 0$. Therefore, as $h \to 0$,

$$\frac{E[D_2|D_2 \leq h]^2}{V(D_2|D_2 \leq h])} \to 3. \tag{1}$$

This implies after some algebra that

$$\text{Bias}(\widehat{\theta}_h|\boldsymbol{D}) \sim \frac{f''(0)h^2}{12}.$$

Now, we consider the conditional variance of $\widehat{\theta}_h$. Let $\sigma^2(d) := V(Y_2 - Y_1|D_2 = d)$, assumed to be continuous at $d = 0$, and let $\sigma^2 := E[\sigma^2(D_2)]$. We have:

$$\begin{aligned} V(\widehat{\theta}_h|\boldsymbol{D}) =& \frac{1}{n^2}\sum_{i=1}^n \sigma^2(D_{2i}) + V(\widehat{\mu}_h|\boldsymbol{D}) - \frac{2}{n}\text{Cov}\left(\sum_{i:D_{i2}\leq h}(Y_{i2} - Y_{i1}), \widehat{\mu}_h|\boldsymbol{D}\right) \\ =& \frac{\sigma^2 + o_p(1)}{n} + \frac{V\left(\sum_{i:D_{i2}\leq h} Y_{i2} - Y_{i1}|\boldsymbol{D}\right)}{n_h}\left(\frac{1}{n_h} - \frac{2}{n}\right) + \left(\frac{\overline{D_{2h}}}{\widehat{V}_h(D_2)}\right)^2 \frac{1}{n_h^2}\sum_{i:D_{i2}\leq h}(D_{i2} - \overline{D_{2h}})^2 \\ & \times \sigma^2(D_{i2}) - \frac{2\overline{D_{2h}}}{n_h\widehat{V}_h(D_2)}\left(\frac{1}{n_h} - \frac{1}{n}\right)\sum_{i:D_{i2}\leq h}\left(D_{2i} - \overline{D_{2h}}\right)\sigma^2(D_{2i}) \\ =& \frac{\sigma^2}{n} + (\sigma^2(0) + o(1))\left(\frac{1}{n_h} - \frac{2}{n}\right) + \frac{\sigma^2(0) + o(1)}{n_h}\frac{\left(\overline{D_{2h}}\right)^2}{\widehat{V}_h(D_2)} + o_p\left(\frac{1}{n}\right) \\ =& \frac{4\sigma^2(0)}{nhf_{D_2}(0)}\left(1 + o_p(1)\right). \end{aligned}$$

The second equality follows from the law of large numbers. The third follows from Taylor expansions of order 0. The fourth follows from $n_h/n = o_p(1)$, $n_h = nP(D_2 \leq h)(1 + o_p(1)) = nhf_{D_2}(0)(1 + o_p(1))$ (where the last equality follows again by continuity of f_{D_2} at 0), and (1). Hence,

$$\text{MSE}(\widehat{\theta}_h|\boldsymbol{D}) \simeq \frac{4\sigma^2(0)}{nhf_{D_2}(0)} + \frac{f''(0)^2h^4}{144}.$$

The optimal h^* of the approximation satisfies

$$0 = \frac{-4\sigma^2(0)}{nh^{*2} f_{D_2}(0)} + \frac{f''(0)^2 h^{*3}}{36}.$$

Hence,

$$h^* = \left[\frac{144\sigma^2(0)}{nf''(0)^2 f_{D_2}(0)}\right]^{1/5}.$$

Estimation of the optimal bandwidth Estimating h^* requires estimating $\sigma^2(0)$, $f''(0)$, and $f_{D_2}(0)$. $\sigma^2(0)$ may be estimated as the difference between the intercept in a local linear regression of $(Y_{i2} - Y_{i1})^2$ on D_{i2} and the square of the intercept in a local linear regression of $Y_{i2} - Y_{i1}$ on D_{i2}. $f''(0)$ may be estimated as the coefficient on D_{i2}^2 in a local quadratic regression of $Y_{i2} - Y_{i1}$ on D_{i2} and D_{i2}^2. Finally, $f_{D_2}(0)$ may be estimated using a kernel density estimator.

Next steps This project is still at a very preliminary stage, and many important steps still need to be conducted. First, in lieu of the heuristic derivation of the optimal bandwidth above, a formal derivation will have to be provided. Second, the asymptotic distribution of $n^{2/5}(\widehat{\theta}_{h^*} - \theta_0)$ will have to be provided. This distribution will have a first-order bias, which will have to be estimated to construct valid confidence intervals, in the spirit of the work of Calonico et al. (2014) for regression discontinuity designs. Third, in regression discontinuity designs, the treatment effect estimator is the difference between two estimators converging at the non-parametric $n^{2/5}$ rate. Here, $\widehat{\theta}_{h^*}$ is the difference between an estimator converging at the parametric rate and an estimator converging at the non-parametric rate. Therefore, the asymptotic distribution of $\widehat{\theta}_{h^*}$ only depends on that of the second estimator. This is slightly unusual, and may reduce the ability of this asymptotic distribution to replicate the finite sample distribution of $\widehat{\theta}_{h^*}$, an issue that we shall investigate via simulations.

References

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014), 'Robust nonparametric confidence intervals for regression-discontinuity designs', *Econometrica* **82**(6), 2295–2326.

Imbens, G. & Kalyanaraman, K. (2012), 'Optimal bandwidth choice for the regression discontinuity estimator', *The Review of Economic Studies* **79**(3), 933–959.
URL: *http://dx.doi.org/10.1093/restud/rdr043*