The role of active monitoring in reducing the execution times of public works. A regression discontinuity approach

Giuseppe F. Gori, Patrizia Lattarulo, Marco Mariani

CREI Seminar Series “Program Evaluation and Spatial Dimension”
Rome, 10th December 2015

IRPET

Introduction and Previous literature

- Public procurement accounts for a sizeable share of the economic activity and is an important component of public expenditure
- Attention in the literature devoted to efficiency issues (Dimitri et al, 2006), e.g. contract formation, costs,. . .
- The most frequent performance indicators for the procurement of public works are
  - Awarding price, cost savings
  - Later stages: total final cost and overruns

  How they are explained by auction formats (Bajari et al., 2009; Lewis and Bajari, 2011; Bucciol et al., 2013; Decarolis, 2014)
- Less attention devoted to understanding which factors can affect the time to completion, descending both from contracts, from the later behavior of the agents involved or from wider "environmental" issues
  - Heterogeneity among government levels (Guccio et al., 2014)
  - Systemic/institutional factors, e.g. court efficiency, corruption, (Coviello et al., 2013; Iossa and Martimort, 2011)
- Outside of a rather descriptive policy management literature, no economic analyses have focused on the specific role of monitoring that the buyer can exert on the executor of the work

The framework

- In the execution phase, the obvious idea is that buyers could combat delays and cost escalations by performing tighter monitoring which enables the timely adoption of remedies/actions against bottlenecks
- However, even assuming a benevolent public buyer, relevant monitoring costs can arise and thus reduce her efforts in fighting moral hazard ("passive waste" in Bandiera et al., 2009). Note that monitoring costs can be buyer and/or project specific
- All this can result into a suboptimal level of monitoring at the "social" level
- This may call for institutional designs or concrete measures to obtain higher monitoring levels by buyers

A simple idea is that if monitoring costs are too high for a specific buyer, they can be shared with higher-order public agents pursuing the same general interest, which may reduce costs by providing know-how or by adding their monitoring efforts on the execution of public works.

Contribution and data

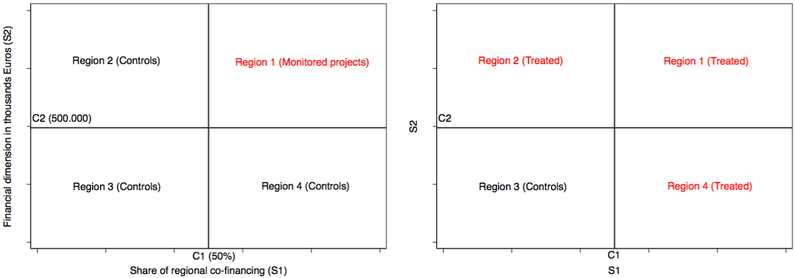
We analyze **the effect of increased monitoring on the execution times of public works**, drawing **causal inferences** within a particularly complex setting: a RD approach with two forcing variables. Since execution durations are potentially right-censored, an appropriate survival model must be specified.

This opportunity is offered by:

- **Tuscany's Regional Law 35/2011** aims at supporting local buyers in the contracts' enforcement for those public works that pass a 'financial' size of 500K euros AND benefit from co-financing by the regional government above a certain threshold (50%).
- In addition to providing guidelines on how to monitor and report advancements in execution phases (it also requires buyers to report progress updates to governmental offices) the law guarantees to the buyer active support by the regional offices in order to address administrative and legal issues whose cost is unaffordable by a specific buyer.
- The analysis is based on a rich administrative dataset of public works implemented in Tuscany from 2008 to 2014, where projects are subject to mandatory active monitoring provided that they meet the already mentioned size and co-financing conditions.
- 1.896 projects, where 74 works are subject to the monitoring and 1.822 are not.

The empirical strategy / 1

- The discontinuity is estimated in the presence of two assignment variables.
- Previous literature on multiple assignment variable in RDD focuses on cases where assignment to treatment depends on meeting just one of multiple conditions (e.g. Papay et al., 2011; Wong et al., 2013).
- We instead focus on a the less explored case in which **projects must fulfil multiple conditions at the same time** in order to receive treatment/monitoring. Following Choi and Lee (2014), the problem can be viewed as follows.



The empirical strategy / 2

- Considering that the process of assignment to monitoring is a deterministic function of two specific variables, the financial size of the project and the share of regional co-financing, we choose to adopt a **sharp regression discontinuity design** (RDD, see Lee and Lemieux, 2010).
- In the presence of thresholds set out regardless of the potential outcomes of the projects, a **local causal effect** at the thresholds can be estimated exploiting the information in their immediate vicinity. In our case, exogeneity holds since the financial dimension of the project and the share of regional co-financing are set even before the work is put out to tender.
- In our dataset, durations (execution times) are expressed in days and are potentially right-censored at the end of the observation period (31-12-2014). This calls for **survival analysis** techniques. We discretise execution times, so as to handle issues of hazard non proportionality in an easier way (Caliendo et al., 2013; Kalbfleisch and Prentice, 2011)
- According to the most recent trends in the regression discontinuity literature, it is preferable to rely only on observations "close to the thresholds" so that the estimation can be carried out by means of a simply specified **local regression**, instead of resorting to complex polynomial specifications to be run on all available observations (Imbens and Lemieux, 2009; Gelman and Imbens, 2014).

▸ Figure

Notation

We split durations into three time periods ($p = 1$: six months; $p = 2$ six to twelve; $p = 3$ over twelve). After splitting, each project is repeated in the dataset as many times as are the periods its execution lasts.

This implies that:

1. each period has a specific "population at risk"
2. observations are no more independent but clustered at the project level $\rightarrow$ need to compute cluster-robust SE (Cameron and Miller, 2015)

For each p we have now a dichotomous outcome for each project i:

$$Y_{ip} = \begin{cases} 1 & \text{if the project is completed during period } p \\ 0 & \text{if the project is completed later than } p \end{cases}$$

Let

- $j = \{1, 2\}$ be the subscript for the assignment variables $S1$ and $S2$
- c_j be the cutoff value constituting the threshold over assignment variable
- T_j be the binary variable for the fulfillment of the threshold criterium
- M be the binary variable for the monitoring status: $T_1 \ x \ T_2 = 1$
- $k = \{1, ..., 4\}$ be the subscript for the region defined by the interaction between T_1 and T_2
- $d_{j,k}$ be the variable for the distance $S_{ji} - c_j$ for observations belonging to a given k

Estimands

Building on Choi and Lee (2014):

$$\tau^{p}_{SRD} = E[Y_i(1) - Y_i(0) | S^1_i = c_1, S^2_i = c_2, P = p]$$

for each p, this holds with no partial effects.
If partial effects are there...

$$\tau^{p}_{SRD} = E[Y_i(1)^{T_1=1,T_2=1} - Y_i(0)^{T_1=0,T_2=1} + \\ - Y_i(0)^{T_1=1,T_2=0} + Y_i(0)^{T_1=0,T_2=0} | S^1_i = c_1, S^2_i = c_2, P = p]$$

In the presence of a binary response ...

$$\begin{aligned}\tau^{p}_{SRD} = &\, Pr(Y_i = 1 | T_1 = 1, T_2 = 1, S^1_i = c_1, S^2_i = c_2, P = p) + \\ &- Pr(Y_i = 1 | T_1 = 0, T_2 = 1, S^1_i = c_1, S^2_i = c_2, P = p) + \\ &- Pr(Y_i = 1 | T_1 = 1, T_2 = 0, S^1_i = c_1, S^2_i = c_2, P = p) + \\ &+ Pr(Y_i = 1 | T_1 = 0, T_2 = 0, S^1_i = c_1, S^2_i = c_2, P = p)\end{aligned}$$

Under the continuity assumption of all conditional expectations we can identify, for each p, τ^{p}_{SRD} as:

$$\begin{aligned}\tau^{p}_{SRD} = &\lim_{(S^1,S^2)\to(c_1^+,c_2^+)} Pr(Y_i = 1 | P = p) - \lim_{(S^1,S^2)\to(c_1^-,c_2^+)} Pr(Y_i = 1 | P = p) + \\ &- \lim_{(S^1,S^2)\to(c_1^+,c_2^-)} Pr(Y_i = 1 | P = p) + \lim_{(S^1,S^2)\to(c_1^-,c_2^-)} Pr(Y_i = 1 | P = p)\end{aligned}$$

The model

We specify the following discrete-time survival model, which can be estimated by means of a pooled logit (Kalbfleisch and Prentice, 2011)

$$Pr(Y_{i,p} = 1 | at\ the\ threshold) = \frac{exp(\cdot)}{1 + exp(\cdot)}$$

where the linear predictor $(\cdot)$ is as follows

$$(\cdot) = \sum_{p=1}^{P} \beta_{0p} P_i^h + \beta_1 M_i^h + \sum_{p=1}^{P} \beta_{2p} M_i^h(p) +$$

$$+ \beta_3 T_{j=1,i}^h + \sum_{p=1}^{P} \beta_{4p} T_{j=1,i}^h(p) + \beta_5 T_{j=2,i}^h + \sum_{p=1}^{P} \beta_{6p} T_{j=2,i}^h(p) +$$

$$+ \beta_7 d_{j=1,k=1,i}^h + ... + \beta_{10} d_{j=1,k=4,i}^h + \beta_{11} d_{j=2,k=1,i}^h + ... + \beta_{14} d_{j=2,k=4,i}^h +$$

$$+ \sum_{j=1}^{2} \sum_{k=1}^{4} \sum_{p=1}^{P} \beta_{15jkp} d_{jki}^h(p)$$

Since we do not assume odds (hazard) proportionality, note that the quantity of interest in the linear predictor is $\beta_1 + \beta_{2p}$. Therefore, $\tau_{SRD}^p = \frac{exp(\beta_1 + \beta_{2p})}{1 + exp(\beta_1 + \beta_{2p})}$.

The model

In a **single assignment variable setting** with **proportional hazard**, the argument of the exponential function could be written including only the terms in blue:

$$
\begin{aligned}
(\cdot) = {} & \underbrace{\sum_{p=1}^{P} \beta_{0p} P_i^h}_{\textit{time dummies}} + \underbrace{\beta_1 M_i^h}_{\textit{treatment (monitoring)}} + \sum_{p=1}^{P} \beta_{2p} M_i^h(p) + \\
& + \beta_3 T_{1i}^h + \sum_{p=1}^{P} \beta_{4p} T_{1i}^h(p) + \beta_5 T_{2i}^h + \sum_{p=1}^{P} \beta_{6p} T_{2i}^h(p) + \\
& + \underbrace{\beta_7 d_{11i}^h + \ldots + \beta_{10} d_{14i}^h}_{\textit{distances wrt threshold } 1} + \beta_{11} d_{21i}^h + \ldots + \beta_{14} d_{24i}^h + \sum_{j=1}^{2} \sum_{k=1}^{4} \sum_{p=1}^{P} \beta_{15jkp} d_{jki}^h(p)
\end{aligned}
$$

The model

while in a **multiple assignment variable setting** could be written including also the terms in green:

$$(\cdot) = \underbrace{\sum_{p=1}^{P} \beta_{0p} P_i^h}_{time\ dummies} + \underbrace{\beta_1 M_i^h}_{treatment\ (monitoring)} + \sum_{p=1}^{P} \beta_{2p} M_i^h(p) +$$

$$+ \underbrace{\beta_3 T_{1i}^h}_{partial\ effect\ score\ 1} + \sum_{p=1}^{P} \beta_{4p} T_{1i}^h(p) + \underbrace{\beta_5 T_{2i}^h}_{partial\ effect\ score\ 2} + \sum_{p=1}^{P} \beta_{6p} T_{2i}^h(p) +$$

$$+ \underbrace{\beta_7 d_{11i}^h \ldots + \beta_{10} d_{14i}^h + \beta_{11} d_{21i}^h + \ldots + \beta_{14} d_{24i}^h}_{distances\ wrt\ both\ thresholds} + \sum_{j=1}^{2} \sum_{k=1}^{4} \sum_{p=1}^{P} \beta_{15jkp} d_{jki}^h(p)$$

The model

Accounting for **hazard non proportionality** requires to interact variables with time

$$(\cdot) = \underbrace{\sum_{p=1}^{P} \beta_{0p} P_i^h}_{\textit{time dummies}} + \underbrace{\beta_1 M_i^h + \sum_{p=1}^{P} \beta_{2p} M_i^h(p)}_{\textit{treatment (monitoring)}} +$$

$$+ \underbrace{\beta_3 T_{1i}^h + \sum_{p=1}^{P} \beta_{4p} T_{1i}^h(p)}_{\textit{partial effect score } 1} + \underbrace{\beta_5 T_{2i}^h + \sum_{p=1}^{P} \beta_{6p} T_{2i}^h(p)}_{\textit{partial effect score } 2} +$$

$$\underbrace{+\beta_7 d_{11i}^h \ldots + \beta_{10} d_{14i}^h + \beta_{11} d_{21i}^h + \ldots + \beta_{14} d_{24i}^h + + \sum_{j=1}^{2} \sum_{k=1}^{4} \sum_{p=1}^{P} \beta_{15jkp} d_{jki}^h(p)}_{\textit{distances wrt both thresholds}}$$

Bandwidth & Cross validation procedure

The extent to which observation are to be included in the analysis should be decided using a **bandwidth selection** procedure (Imbens and Lemieux, 2008, Imbens and Kalyanaraman, 2012, Calonico et al., 2014). However, all these selectors are conceived with respect to a single assignment variable, and need to be extended to our multiple assignment variable setting.

Following Imbens and Lemieux (2008), we discard observations from the tails (above/below the median according to which side of the thresholds) prior to performing a "leave-one-out" cross validation procedure.

Cross validation in a nutshell:

- To see how well a local linear regression with bandwidth h fits the data, a local linear regression must be run for each observation i with i left out of the sample and then use the resulting coefficient estimates to predict the value of Y_i at X_i.
- Mimicking the fact that RD estimates are based on regression estimates at a boundary, the regressions are estimated using only observations to the left of i (for i below the cutoff) or the right of i (for i above the cutoff).
- Repeating this exercise N times produces a set of predicted values of Y_i that can be compared with the actual values of Y_i.
- The final "cross-validated" bandwidth is then picked by choosing the value of h that minimizes the mean square of the difference between the predicted and the actual values of Y_i.

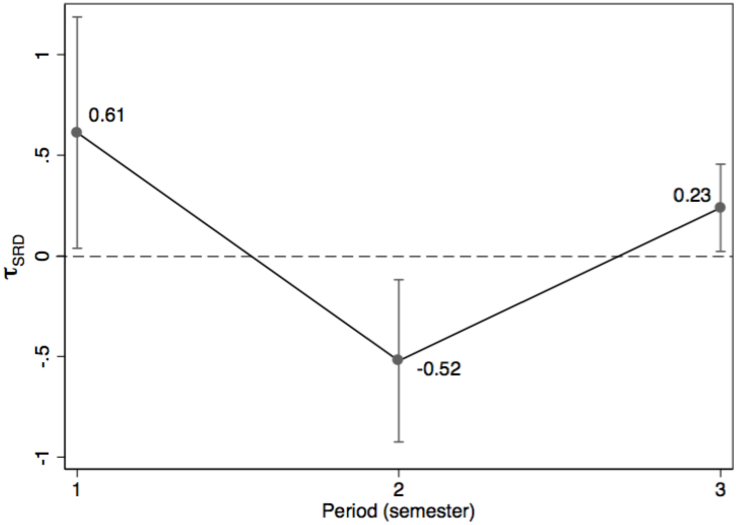
Our extension, relying on the cross-validation procedure set out in Imbens and Lemieux (2008), compares model fitting for each possible alternative combinations of the **two forcing variables** applying the procedure to **each of the four areas** defined by the interaction of the two treatment variables (T_1 and T_2). This delivers **asymmetric bandwiths** with respect to both thresholds.

Results / 1

	Coeff.	Log-Odds	C-Robust SE	p-value
p_1	β_{01}	-1.354	.78	0.08
p_2	β_{02}	-0.304	.71	0.67
p_3	β_{03}	1.250	1.13	0.27
Monitoring p_1	β_{21}	-27.968	6.01	0.00
Monitoring p_2	β_{22}	-41.515	6.96	0.00
Monitoring p_3	β_1	31.840	4.96	0.00
Partial effects (T_1 = *share*, T_2 = *total cost*)				
T_1 p_1	β_{41}	83.095	4.46	0.00
T_1 p_2	β_{42}	92.725	5.03	0.00
T_1 p_3	β_3	-85.590	3.26	0.00
T_2 p_1	β_{61}	-52.743	4.73	0.00
T_2 p_2	β_{62}	-49.425	5.15	0.00
T_2 p_3	β_5	52.183	4.15	0.00
N.Obs=541, N.Projects=257				

Results / 2

Average treatment effect at the thesholds (probability jumps) with 95% CIs



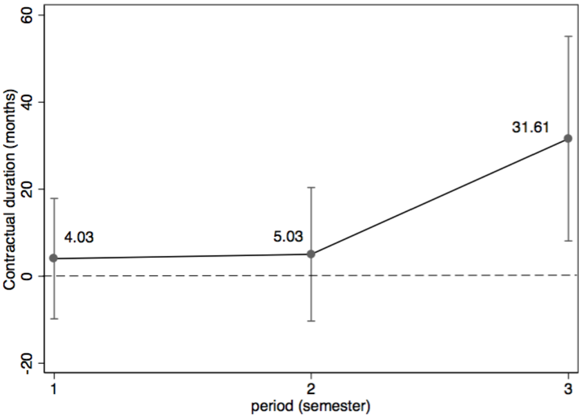
Results interpretation (hypothesis)

- Active monitoring boosts the execution speed of "short" projects.
- As a result, the population at risk is soon constituted by a mixture of short, but unmonitored, projects and of long and persistent projects of both kinds.
- In the second semester, short but unmonitored projects tend to come to an end, which explains the negative sign of the discontinuity.
- In a longer time horizon, we have only more persistent projects waiting for completion. Here, monitoring turns to work again in reducing execution times, although it has to be noted that the magnitude of the positive discontinuity is smaller than previously.

We can test this interpretation by specifying, on the same bandwidth, a descriptive "pooled" linear model for the contractual duration using the same predictors of the main treatment effects model.

Results interpretation (testing the hypothesis)

Differential contractual duration at the thesholds between monitored and unmonitored projects constituting the population at risk in each p (with 95% CIs)



N.Obs = 541,   N.Projects = 257

Robustness and sensitivity checks

1. Evaluating the identification assumption using covariates as pseudo-outcomes
2. Improving precision by adding covariates to the set of predictors
3. Evaluating sensitivity to bandwidth choice

Covariates as pseudo-outcomes

- Imbens and Lemieux (2008) suggest to test the null hypothesis of a zero average effect on covariates (pseudo outcomes) that are, by definition, unaffected by the treatment (e.g. award criterium, type, contractual duration). If we find such a discontinuity, it typically casts doubt on the assumptions underlying the RD design.
- This analysis is non-trivial, given our multiple-assignment setting.
- Accordingly, for each of the covariates, we have specified the following model, which is equivalent to the one specified in order to estimate the causal effect of monitoring, except for the time dummies and the time-interactions:

$$X_i^h = \beta_1 M_i^h + \beta_2 T_{1i}^h + \beta_3 T_{2i}^h + \beta_4 d_{11i}^h + \ldots + \beta_7 d_{14i}^h + \beta_8 d_{21i}^h + \ldots + \beta_{11} d_{24i}^h$$

- As in the case of the main model, this specification has been used to perform our (extended) cross-validation procedure in order to select the correct bandwidth for **each of the covariates**.

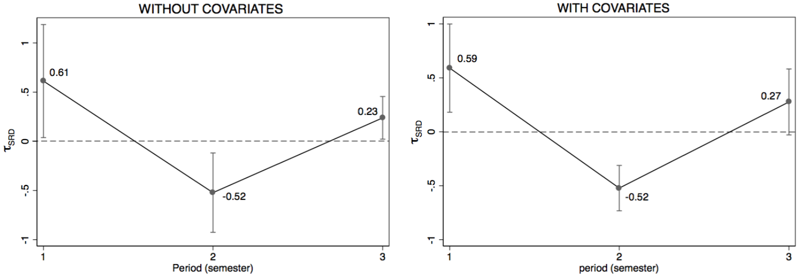
Covariates

Covariates as pseudo-outcomes / 2

	Coeff. (monitoring)	SE (monitoring)	p-value (monitoring)	N.Obs
dependent variable (pseudo-outcome):				
Tendering proc. (=1 if competitive)	31.8	6280	0.99	93
Award Criterium (=1 if lowest price)	-1.4	4.40	0.75	169
Type (=1 if buildings)	38.2	7204	0.99	64
New (= 1 if new construction)	0.6	4.06	0.88	81
Contractual duration (cont.)	-3.9	7.39	0.59	163

Adding covariates to the set of predictors

Average treatment effect at the thesholds (probability jumps) with 95% CIs



N.Obs = 541,   N.Projects = 257

Sensitivity to bandwidth choice

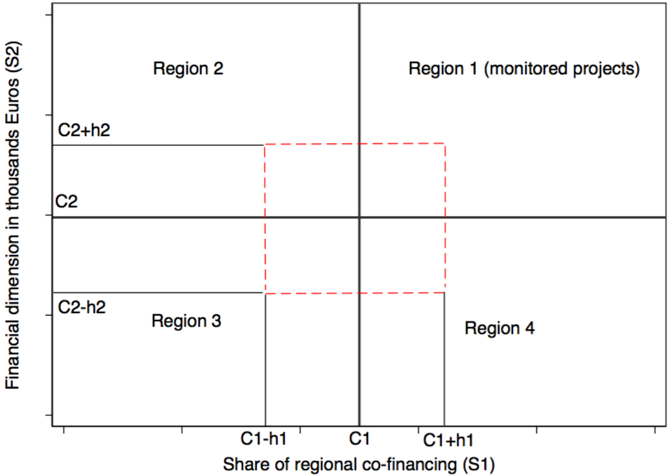
Average treatment effect at the thesholds (probability jumps) in the presence of alternative bandwidths

	Period				
	1	2	3	N.Obs	N.Projects
Bandwidth					
Optimal	0.61	-0.52	0.23	541	257
Narrow (2%)	0.38	-0.66	0.35	174	78
Wide (2%)	0.61	-0.5	0.23	549	263
Narrow (5%)	0.47	-0.66	0.36	171	76
Wide (5%)	0.67	-0.47	-0.61	567	273

Conclusions

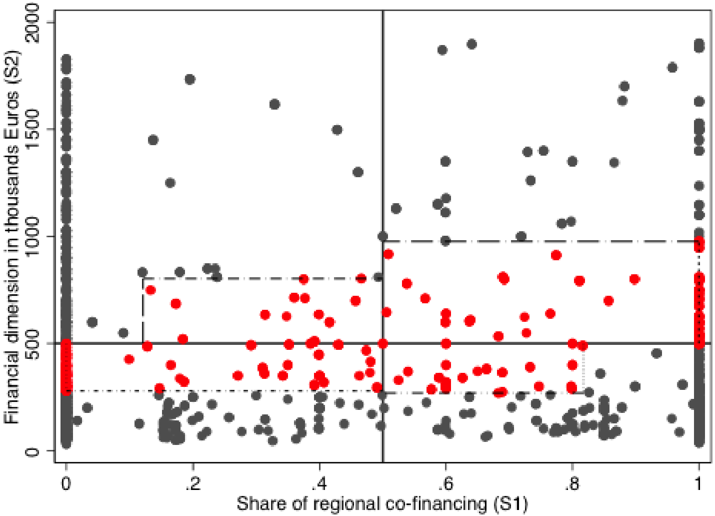
- Our estimation strategy relies on a sharp regression discontinuity design and is suitable to identify a causal relationship.
- Our estimation problem is made particularly challenging by the facts that
  1. the outcome of interest is potentially right censored, which calls for the adoption of survival analysis techniques that have so far been unusual with regression discontinuity designs
  2. the assignment of project to increased active monitoring is jointly determined by two exogenous assignment variables, which requires the design of a very novel adaptation of regression discontinuity techniques to this particular setting.
- The results suggest that, at the threshold values of the two assignment variables, the causal effect of active monitoring on time-to-completion is positive on very short projects that would have not lasted long anyway and, more interestingly, on the subset of projects that are very persistent in time.
- To the best of our knowledge, this kind of result is completely new in the public procurement literature and suggests that there is room for public buyers to increase effort and improve their action during the execution stage of projects.

Bandwidth



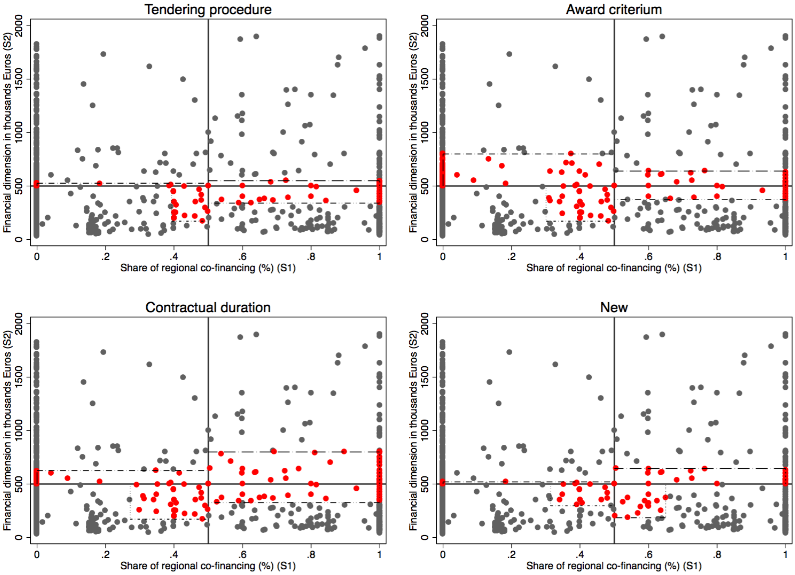
Back

Bandwidth



Back

Covariates as pseudo-outcomes - Bandwidths



Back