CrossMark

How much should we trust regression-kink-design estimates?

Michihito Ando[1]

Abstract In a regression kink (RK) design with a finite sample, a confounding smooth nonlinear relationship between an assignment variable and an outcome variable around a threshold can be spuriously picked up as a kink and results in a biased estimate. In order to investigate how well RK designs handle such confounding nonlinearity, I firstly implement Monte Carlo (MC) simulations and then study the effect of fiscal equalization grants on local expenditure in Japan using an RK design. Results in both the MC simulations and the empirical application suggest that RK estimation without covariates can be easily biased, and this problem can be mitigated by adding observed covariates to the regressors. On the other hand, a smaller bandwidth or a higher-order polynomial, even a quadratic polynomial, tends to result in imprecise estimates although they may be able to reduce estimation bias. In sum, RK estimation with confounding nonlinearity often suffers from bias or imprecision, and estimates are credible only when relevant covariates are controlled for. I also examine how placebo RK estimation can effectively address these issues.

Keywords Regression kink design · Endogenous regressors · Confounding nonlinearity · Intergovernmental grants

✉ Michihito Ando
andou-michihito@ipss.go.jp

1   National Institute of Population and Social Security Research, Japan, Hibiya Kokusai Building 6th Floor 2-2-3 Uchisaiwai-cho, Chiyoda-ku, Tokyo 100-0011, Japan

🍂 Springer

JEL Classification C13 · C21 · H71 · H72 · H77

1 Introduction

To find a plausible exogenous variation which is generated by a quasi-experimental situation has become an essential element of recent econometric policy evaluation, particularly when it is not possible to randomly assign the treatment of interest.

Regression kink (RK) designs have recently been added to the methodologies available for the implementation of this trend. The basic idea of an RK design is similar to a regression discontinuity (RD) design, which is now one of the most popular approaches in applied microeconometrics. Whereas an RD design utilizes a "discontinuity" or "jump" in treatment status at a threshold of an assignment variable, an RK design exploits a "kink" in treatment status at a threshold. The applicability of RK designs may be potentially broader than that of RD designs in policy evaluation because various public policies avoid abrupt discontinuity at a threshold in the treatment and outcome status of relevant parties and result in a continuous but often kinked transition in status following introduction of the treatment in question.

Two such examples are progressive taxation and welfare transfer programs. Because marginal income tax rates and social welfare benefits are often determined by the income levels of taxpayers and welfare recipients, the budget constraints of individuals are often kinked at eligibility thresholds. The existence of such kinks in budget constraints is well known and is taken into account in empirical studies of labor supply and other relevant topics.[1]

Nonetheless, kinks generated by institutional or policy settings have not been recognized as explicit sources of identification. One notable recent exception is studies of bunching such as Saez (2010), which exploits the bunching behavior of taxpayers at a kink point to estimate labor-supply elasticity. While the study of bunching utilizes endogenous sorting at a kink for identification, RK designs exploit the seemingly exogenous variation that is generated by a kink in a manner similar to the way RD designs use a discontinuity as a source of exogenous variation.

The empirical applications of RK designs are still limited but increasing. Before the seminal work of Nielsen et al. (2010) introduced the term "regression kink design," a few papers had already exploited a kink as a means of identification (Guryan 2001; Dahlberg et al. 2008). After Nielsen et al. (2010) (working paper 2008), Card et al. (2015) [and its previous versions Card et al. (2009, 2012)] formally discussed nonparametric identification and estimation using an RK design. Other recent applications include Bravo (2011), Ek (2013), Turner (2014), Lundqvist et al. (2014), Simonsen et al. (2016) Landais (2015), and Engström et al. (2015).[2]

---

[1] See, among others, Hausman (1985) and Moffitt (1986, 1990) for econometrics issues concerning nonlinear budget sets. Studies of nonlinear budget sets have mostly appeared in the labor-supply literature, but Moffitt (1984), for example, applies these techniques to the study of the effects of federal grants on state and local expenditures.

[2] A closely related but distinctly different study is Dong (2014), which applies an RD design with a binary treatment to a situation in which the treatment probability is not jumped but kinked. A key difference between an RK design and Dong (2014)'s RD design is that treatment is continuous in the former and binary in the latter.

Although an RK design seems to be a fruitful identification strategy which can be applied to various institutional settings, there is some concern about the applicability of RK designs to real-world finite samples. That is, although an RK design tries to capture a discontinuous change in slope at a kink point, it may be incapable of providing accurate estimation with a finite sample if a kinked distribution in an outcome variable against an assignment variable is significantly confounded by noise around the cutoff point. While a similar problem arises in RD designs, RK designs may have a more severe problem due to the intrinsic subtleness of the estimation of a "kink" when compared with a "jump."

In particular, a confounding smooth nonlinear relationship, e.g., a quadratic or more complicated nonlinearity, between an assignment variable and an outcome variable around a cutoff point could be problematic in a real-world sample, because this smooth nonlinearity could be spuriously captured as a kink. This problem is pointed out and considered by Landais (2015), but has not been directly investigated in the RK literature.[3]

Another recent important contribution to this topic is the study of Ganong and Jäger (2015), which proposes a permutation test for RK estimation, assuming that the assignment of a kink is random around a threshold. The authors' method directly addresses the issue of confounding nonlinearity around the threshold, but mainly focuses on credible statistical tests for RK designs, not the issue of the degree of bias and imprecision caused by confounding nonlinearity.

Calonico et al. (2014b) (hereafter CCT) also examine a closely related issue. The authors propose robust confidence intervals for RD/RK estimation based on a bias-corrected RD/RK estimator and a corresponding standard error estimator. Their bias-correction technique, which primarily follows the standard approach in the nonparametric literature such as Fan and Gijbels (1996), can potentially reduce bias from confounding nonlinearity around the cutoff in RK designs. The performance of CCT's procedures in RK designs, however, is still largely unknown because CCT focus on RD designs in their simulation study.[4]

In this paper, therefore, I study how well an RK design can eliminate confounding smooth nonlinearity around the threshold and capture only the "kink" that is generated by a treatment variable. I first examine this problem with MC simulations and then apply an RK design to a real-world situation where I estimate the effect of a fiscal equalization grant on local expenditure with panel data from Japanese municipalities. As I discuss below, the MC simulations and the real-world application are strong complements, yielding useful and practical implications.

---

[3] In order to cancel out or reduce estimation bias caused by a quadratic relation which is spuriously picked up as a kink, Landais (2015) proposes a double-difference RK design which exploits the shifting of kink positions over time. Note that Landais (2015) mentions that the double-difference RK design is much more demanding in terms of the precision of estimates than his baseline RD design. In a difference-in-difference framework, Chetty et al. (2014) utilize a change in the slope of pension contributions with respect to income over time that arises at a certain income threshold. These approaches may be able to deal with confounding nonlinearity that is time-invariant, but I do not discuss them in this paper in which my focus is on RK designs with cross-sectional or pooled data.

[4] See also Card et al. (2015) for the application of CCT's procedures to an RK design.

In my MC simulations, I use a data-generating process (DGP) that generates both an endogeneity problem and confounding smooth nonlinear relation between an assignment variable and an outcome variable. The DGP can also be interpreted as a stylized version of a Japanese fiscal equalization scheme, which I subsequently investigate. For RK estimation in each simulation, I adopt both a conventional estimator and CCT's bias-corrected estimator.

My finding in the MC simulations is that a conventional linear RK estimate can be biased when there is a confounding nonlinearity around the kink point. This bias could be mitigated or eliminated by using a quadratic polynomial model and/or CCT's bias-corrected estimator, but either procedure comes at a cost of reduced precision and power, and this cost is often prohibitive. The simulation results also show that the performance of RK estimation is significantly improved by controlling for observed covariates.

I then examine the plausibility of an RK design using Japanese local public finance panel data. Specifically, I investigate the effect of Japanese fiscal equalization grants on local expenditure by exploiting the kinked formula of the fiscal equalization grant.

My results of the real-world application show that their implications are more or less consistent with those of the MC simulations in at least two respects. First, in both the MC simulations and the empirical application, RK estimation without covariates can be easily biased, and this problem can be mitigated by adding observed covariates to the regressors. Second, in both cases, a smaller bandwidth, a higher-order polynomial, even a quadratic polynomial, and/or CCT's bias-correction procedures tend to result in highly imprecise estimates.

Overall, my studies with MC simulations and the Japanese municipal data provide robust evidence that RK designs should be used with great caution. The MC simulations, which are based on somewhat arbitrary but known data-generating processes, indicate that RK estimation provides highly misleading results when confounding nonlinearity exists around the threshold. The empirical application, which in turn based on actual data but with an unknown data-generating process, does not provide as conclusive implications as the MC simulations. It does, however, strongly suggest that the problems of bias and imprecision in RK estimation, which are found in the MC simulations, actually exist in real-world settings.

Finally, this paper also considers an informal way to check whether an RK estimation suffers from serious bias or imprecision. The idea is akin to a permutation test proposed by Ganong and Jäger (2015), but I use placebo RK estimates with varying placebo cutoffs not for a statistical test but as a graphical examination, as is done in several RD and RK studies.[5] Although this placebo RD or RK estimation is not new, I discuss how this procedure may be particularly useful in detecting confounding nonlinearity and examining the validity of an RK design.

The rest of the paper is organized as follows: In Sect. 2, I briefly explain an RK design and discuss potential problems in this approach. Section 3 presents MC simulations with an DGP that can also be interpreted as a stylized fiscal equalization scheme. Section 4 provides an empirical application of an RK design to a real-world situation

---

[5] As for RK studies, Engström et al. (2015) use placebo RK estimation as a robustness check. Ando (2015) also uses similar placebo estimation, exploiting the same kinked variation in a grant formula as in this paper.

by studying the effects of fiscal equalization grants on local expenditures in Japanese municipalities. In Sect. 5, I examine placebo RK estimation as a tool to detect bias in RK designs. Section 6 concludes.

2 An identification problem in RK designs

2.1 Estimation in RK designs

Consider the following constant-effect and additive model that is presented by Nielsen et al. (2010):[6]

$$Y = \tau B + g(V) + \varepsilon, \tag{1}$$

where $B = b(V)$ is a deterministic and continuous function of V with a kink at $v = 0$, $g(V)$ is an unrestricted function and ε is an error term. They show that if $g(\cdot)$ and $E(\varepsilon|V = v)$ have derivatives that are continuous in v at $v = 0$, then the RK estimand τ can be expressed as follows:

$$\tau = \frac{\lim\limits_{v_0 \to 0^+} \left.\dfrac{dE(Y|V = v)}{dv}\right|_{v=v_0} - \lim\limits_{v_0 \to 0^-} \left.\dfrac{dE(Y|V = v)}{dv}\right|_{v=v_0}}{\lim\limits_{v_0 \to 0^+} \left.\dfrac{db(v)}{dv}\right|_{v=v_0} - \lim\limits_{v_0 \to 0^-} \left.\dfrac{db(v)}{dv}\right|_{v=v_0}}. \tag{2}$$

Intuitively speaking, the numerator of the RK estimand is the change in the slope of the conditional expectation function $E(Y|V = v)$ at the kink point ($v = 0$) and the denominator is the change in the slope of the deterministic assignment function $b(V)$ at the kink. As Card et al. (2015) discuss, one important feature of an RK design is that it allows for other less extreme forms of endogeneity if the density of the assignment variable is smooth and rules out deterministic sorting at the kink point.

For clarification, suppose a linearly incremental treatment variable $B = b(V) = \kappa V$ where $\kappa > 0$ if $v > 0$ and otherwise $B = b(V) = 0$ in equation . Given that $g(\cdot)$ is a smooth function that is differentiable in V at $v = 0$, the slope of Y, dY/dV, changes discontinuously at $v = 0$ from $g'(0)$ to $\tau\kappa + g'(0)$. Because the change in the slope of $B = b(V)$ at $v = 0$ is κ, the treatment effect τ can be recovered as

$$\frac{g'(0) + \tau\kappa - g'(0)}{\kappa - 0} = \frac{\tau\kappa}{\kappa} = \tau,$$

---

[6] See also Card et al. (2015) for a theoretical discussion of generalized RK estimation based on a nonseparable model with a heterogeneous treatment effect. Card et al. (2015) show that an RK estimand can be interpreted as the same "local average response" parameter of Altonji and Matzkin (2005) or equivalently "treatment on the treated" parameter of Florens et al. (2008) if certain conditions are satisfied. Because the objective of this paper is to investigate the performance of an RK design in the presence of a confounding nonlinearity around a cutoff point, I exclusively rely on constant-effect and additive models in the following discussion and MC simulations.
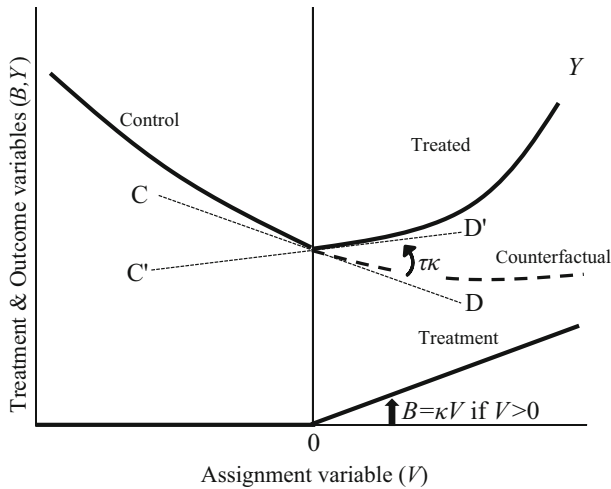
Fig. 1 Stylized features of the regression kink design

using the RK estimand in (2)

The stylized features of an RK design in this setting can also be graphically described as in Fig. 1. Here the effect of B on Y at $v = 0$ is depicted as the ratio of the change in tangent from the line CD (at $v \to 0^-$) to the line C'D' (at $v \to 0^+$) to the change in the slope of the treatment variable at $v = 0$.

For estimation with an RK design, previous studies often use a local polynomial model, which is analogous to a local polynomial model in an RD design (Lee and Lemieux 2010):

$$Y = \alpha_0 + \beta_0 \cdot D + \sum_{p=1}^{\bar{p}} [\alpha_p v^p + \beta_p v^p \cdot D] + \varepsilon \quad \text{where } |v| \leq h, \qquad (3)$$

where ε is a usual random error term and D is a dummy variable which takes one when the assignment variable V exceeds the threshold $v = 0$ and otherwise takes zero. $\bar{p}$ is the degree of a polynomial and h is the bandwidth that determines the window $[-h; +h]$ within which the sample is selected. Note that Eq. (3) has the term D, implying that this model allows discontinuity at the cutoff, but this does not affect asymptotic bias and variance of the RK estimator.[7]

In this local polynomial regression, the numerator of Eq. (2) is estimated by the OLS estimator of β_1. Thus, the RK estimator of τ can be obtained by dividing the OLS estimator of β_1 by the slope change of $B = b(V)$ at the cutoff.

---

[7] Based on a more general nonparametric kernel regression framework, Card et al. (2012) show that the bias and variance of a constrained RK estimator (i.e., imposing continuity at a cutoff) have the same asymptotic properties as those of an unconstrained RK estimator when the kernel is symmetric.
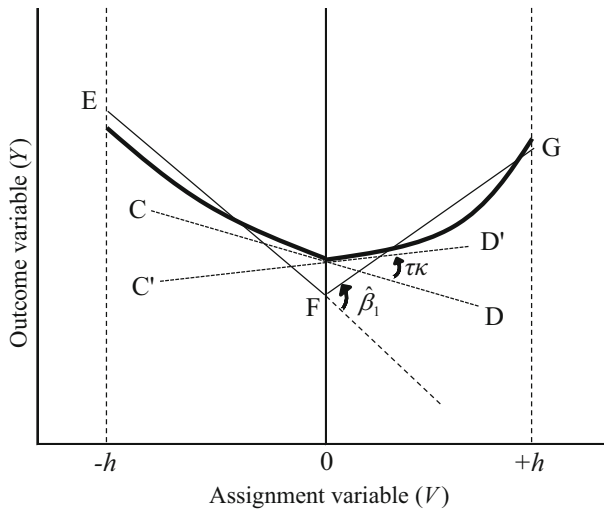
Fig. 2 Bias in an RK estimate with a linear polynomial

2.2 A potential problem in an RK design

Equation (2) holds true only at the cutoff point $v = 0$. In a real-world sample, however, we often need to include observations that are not very close to the threshold, and this inclusion of less relevant observations might lead to a biased estimate. For example, suppose $g(\cdot)$ can be expressed as a smooth nonlinear (e.g., quadratic) function as depicted in Fig. 1. In this case, it could be difficult to separate out this confounding nonlinearity from a kink generated by the treatment with a finite sample. The resulting estimate of β_1, hereafter $\widehat{\beta}_1$, can thus be biased and systematically different from $\tau\kappa$.

This is of particular concern if the first-order polynomial $\bar{p} = 1$ is used in Eq. (3) for RK estimation. To illustrate this problem, Fig. 2 replicates Fig. 1 with an arbitrary bandwidth $[-h; +h]$. The kinked line of the treatment variable is dropped for simplicity. In this graph, a kinked line EFG expresses fitted values with the piecewise linear model $Y = \alpha_0 + \alpha_1 V + \beta_1 V \cdot D + \varepsilon$ with bandwidth $[-h; +h]$, which is equivalent to Eq. (3) with $\bar{p} = 1$ and $\beta_0 = 0$. In other words, the line EF is a linear fit for observations with $v \le 0$ and the line FG is a linear fit for observations with $v > 0$, where the continuity of the two lines (kink) at $v = 0$ is imposed. Here, the estimated coefficient $\widehat{\beta}_1$ is the difference between the slopes of the lines FG and EF, which is clearly different from $\tau\kappa$.

The reason for the difference between $\widehat{\beta}_1$ and $\tau\kappa$ is intuitively quite straightforward. $\widehat{\beta}_1$ is different from $\tau\kappa$ because the linear fits EF and FG are not identical with the tangent lines CD and C'D' when $g(\cdot)$ is a nonlinear function around $v = 0$. This gap between β_1 and $\tau\kappa$ can be reduced by making bandwidth $[-h; +h]$ smaller and goes to zero as $h \to 0$. However, with a finite sample, the bandwidth may not be sufficiently narrowed to perfectly remove this discrepancy.

In this case, an RK design with a local linear regression results in a biased estimator of τ. In addition, even if a smaller bandwidth might reduce or even eliminate this bias,

it comes at the cost of less precision due to a smaller sample size and less data variation around the cutoff point. An alternative solution is to use a global or local polynomial regression. For example, in Fig. 2, quadratic fits in the both sides of the threshold seem to recover the slope change $\tau\kappa$ at $v = 0$. However, this procedure may not always work well since RK estimation with a higher-order polynomial incurs the substantial cost of larger variance in the estimator.[8]

In sum, although RK designs are as explicit and straightforward as RD designs, there is some concern about their applicability to real-world finite samples. In the following sections, I discuss the potential defects of an RK design with a finite sample using MC simulations and real-world empirical data.

3 Monte Carlo simulations

3.1 Stylized fiscal equalization scheme

In this section, I implement Monte Carlo (MC) simulations in order to examine the performance of RK estimators with a finite sample and in the presence of a confounding smooth nonlinear relation around the cutoff point. Because the empirical application I examine in the next section is the effects of Japanese fiscal equalization grants on local expenditure, I also set up a DGP based on a stylized fiscal equalization scheme in this section.

Many countries have fiscal equalization schemes that are meant to equalize or alleviate fiscal disparities among local governments through intergovernmental transfers. Although fiscal equalization schemes differ considerably across countries, it can be argued that there are in general two important components in fiscal equalization transfers from an upper-level government to a lower-level government. The first is the equalization of fiscal revenue capacity (revenue equalization), and the second is the equalization of expenditure needs (needs equalization).[9] According to Dafflon (2007), the fiscal equalization approach that takes into account both revenue equalization and needs equalization is referred to as "need-capacity gap" equalization. Because Japan has a unified fiscal equalization scheme which takes into account both revenue and needs equalization, the Japanese fiscal equalization scheme can be categorized as a need-capacity gap equalization scheme.

Suppose a stylized scheme of "need-capacity gap" equalization where fiscal equalization general (unconditional) grants are distributed to individual local bodies

---

[8] Card et al. (2012) argue that, unlike in the case of the kind of local polynomial regression for an interior point discussed by Fan and Gijbels (1996), there is a substantial cost in variance incurred by using a local quadratic model instead of a local linear model in RK designs. In addition, Gelman and Imbens (2014) argue that high-order polynomials should not be used for RD estimation because of several unfavorable properties of high-order polynomials, which are also applicable to RK estimation. Although Card et al. (2014) argue that the local linear estimator should not always dominate other higher-order local polynomial estimators in RD and RK designs, my applications of an RK design to both simulated data and real-world data in this study show that high-order local polynomial regressions often result in very imprecise estimation. In another study with an RK design, Simonsen et al. (2016) also adopt only a local linear polynomial in their baseline analysis for the same reason.

[9] See Boadway and Shah (2007) for extensive studies on fiscal equalization programs.

(hereafter municipalities) based on the following kinked assignment rule:

$$\text{GRANT}_i = \begin{cases} 0 & \text{if } V_i \leq 0 \\ V_i & \text{if } V_i > 0, \end{cases}$$

where GRANT_i is the amount of the fiscal equalization grant for municipality i at t. V_i is the "need-capacity gap" of the municipality and defined as follows:

$$V_i = \text{NEED}_i - \text{CAP}_i,$$

where NEED_i is the expenditure need, which indicates the total cost of the standard levels of local public services for municipality i. CAP_i is the revenue capacity of municipality i which represents the amount of its own tax revenues that municipality i can collect under the standard local tax system.

In short, under this fiscal equalization scheme, grants ensure that all municipalities can provide a standard level of local public services by filling the need-capacity gap in cases where NEED_i outweighs CAP_i. On the other hand, if the need-capacity gap is negative, a municipality is considered wealthy enough to cover their expenditure needs on their own and no grant is provided. I incorporate this stylized fiscal equalization scheme with a kinked grant assignment rule into my MC simulations.

3.2 Data-generating process

One primary reason an RK design is required in the analysis of grant effects on local expenditure is that endogeneity results in a biased estimate if expenditure, hereafter EXP, is regressed on GRANT with a simple OLS.[10] In this subsection, I construct a DGP which has the following three properties in order to investigate the performance of an RK design: (I) an endogeneity problem when simply regressing EXP on GRANT, (II) a deterministic assignment rule GRANT $= b(V)$ with a kink at $v = 0$, and (III) confounding nonlinearity between V and EXP through observed and unobserved covariates.

It is not difficult to include this setting in a DGP for MC simulations. First, NEED_i is defined as follows:

$$\text{NEED}_i = \text{NEED}^0 + X_i + U_i + \psi_i,$$

where NEED^0 indicates a constant basic expenditure need, X_i is an observed covariate, U_i is an unobserved covariate, and ψ_i is a random component. The coefficients of X_i

---

[10] There is an increasing number of studies which explicitly consider the endogeneity problem in grant effects. See, among others, Holtz-Eakin (1986), Knight (2002), Gordon (2004), Dahlberg et al. (2008), Feyrer and Sacerdote (2011), Bravo (2011), Chodorow-Reich et al. (2012), Wilson (2012), Conley and Dupor (2013), Litschig and Morrison (2013), Lundqvist et al. (2014), Suárez and Wingender (2016), and Lundqvist (2015).

and U_i are set as one for simplicity. Second, CAP_i is determined as:

$$CAP_i = CAP^0 + \zeta_i,$$

where CAP^0 is a constant basic revenue capacity and ζ_i is a random term. For simplicity, I assume that CAP_i is not affected by X_i and U_i.

Finally, the outcome variable EXP_i is set to be affected by CAP_i, X_i and U_i. One straightforward interpretation of the effect of CAP_i on EXP_i is that expenditure should be primarily determined by tax revenue capacity. Thus, I set the overall DGP (named as DGP 1) as follows:

$$X_i = x_i \tag{4}$$

$$U_i = u_i \tag{5}$$

$$NEED_i = NEED^0 + X_i + U_i + \psi_i \tag{6}$$

$$CAP_i = CAP^0 + \zeta_i \tag{7}$$

$$V_i = NEED_i - CAP_i \tag{8}$$

$$GRANT_i = \begin{cases} 0 & \text{if } V_i \leq 0 \\ V_i & \text{if } V_i > 0 \end{cases} \tag{9}$$

$$EXP_i = \tau GRANT_i + CAP_i + CAP_i^2 + X_i + X_i^2$$
$$+ U_i + U_i^2 + \omega_i, \tag{10}$$

where $x_i, u_i, \psi_i, \zeta_i$ and ω_i are all NID(0, 1). In (10) I include the terms X^2 and U^2 so that X and U and have linear effects on V through NEED, but nonlinear effects on EXP. In addition, I also add the term CAP^2 to this equation, which implies that an increase in CAP nonlinearly boosts EXP. As a result, there is a nonlinear smooth correlation between V and EXP through CAP, X, and U. The coefficients of CAP, CAP^2, X, X^2, U and U^2 are set as one.[11] Finally, τ is the homogeneous treatment effect of GRANT on EXP.

In this DGP, the three requirements listed above are fulfilled because the effect of GRANT on EXP cannot be estimated with a simple regression due to an omitted variable bias from unobserved U (I), GRANT is nonstochastically determined by V with a kink at $v = 0$ (II), and there is a confounding nonlinear relation between V and Y through observed covariates CAP and X and an unobserved covariate U (III).

Figure 3 shows a causal diagram for (4)–(10). In this DGP, simple regression with observed covariates cannot properly identify τ, because a confounding effect from U cannot be controlled for. The treatment variable GRANT, however, deterministically

---

[11] The introduction of X^2 and U^2 in Eq. (10) is a rather arbitrary setting for generating the confounding nonlinearity between V and EXP in my DGP. The nonlinear relationship of X and U to some fiscal variables is plausible, given the fact that observed and unobserved socioeconomic factors affect the fiscal variables of local governments through various channels. The introduction of CAP^2 is also arbitrarily set in order to increase the confounding nonlinearity in my DGP. This may be justified if I can assume that local governments increase their revenue and expenditure (EXP) as much as possible by broadening tax bases, increasing tax rates, or employing some other method when their revenue capacity (CAP) increases.

Fig. 3 Causal diagram of the stylized fiscal equalization scheme. *Note* A *solid arrow* shows causality and an *outlined arrow* represents a deterministic relation
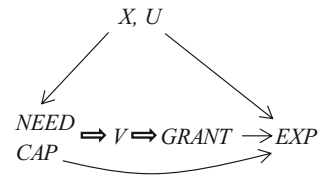
depends on V with the kinked formula (9), and I can thus exploit an RK design by using V as an assignment variable and $v = 0$ as a threshold.

I also conduct another MC simulation with a different DGP. In this second DGP (DGP 2), I use the following equations instead of (6) and (10), while the other parts of the DGP are the same as DGP 1:

$$\text{NEED}_i = \text{NEED}^0 + X_i + X_i^2 + U_i + U_i^2 + \psi_i, \tag{6'}$$

$$\text{EXP}_i = \tau \text{GRANT}_i + \text{CAP}_i + \text{CAP}_i^2 + X_i + U_i + \omega_i. \tag{10'}$$

In these models, X and U have nonlinear quadratic effects on V, but linear one-to-one effects on EXP. In this case, the confounding nonlinear relation between V and EXP is different from a straightforward quadratic U-shape relation as in DGP 1. That is, in DGP 2, a unit increase in X or U leads to a direct linear increase in EXP and quadratic increase in V, not vice versa. Therefore, the resulting distribution of $g(V)$ in Eq. (1) can be more complicated in DGP 2 than in DGP 1, which has a simple U-shape confounding nonlinearity. One important thing to note, however, is that there is still no kinked relation between V and EXP other than through $GRANT$, implying that the condition that $g(\cdot)$ has a derivative that is continuous at the cutoff is valid.

3.3 Simulation settings

When it comes to other MC simulation settings, I first impose a one-to-one constant treatment effect by setting τ as one in (10) and (10'). Second, NEED^0 and CAP^0 are set at 5. Third, the sample size is set to 10, 000, which is close to empirical application I discuss later. Finally, the number of simulations conducted is 1,000.

In simulation analysis based on the above DGPs, τ in (10) or (10') is estimated with an RK design using Eq. (3). Because Eq. (9) shows that the slope of GRANT $= b(V)$ changes from zero to one at the threshold, the denominator of the RK estimand (2) is one. The point estimate of β_1 with the RK regression model (3) can therefore be interpreted as the causal parameter of interest. Equation (10) or (10') is estimated using a linear or quadratic polynomial regression and with CCT's procedures.

In CCT's procedures, I use both regularized and nonregularized bandwidth selectors, following Card et al. (2015). Although Imbens and Kalyanaraman (2012), who propose a seminal optimal bandwidth selector, and CCT argue that "regularization" is useful to avoid very large bandwidths, Card et al. (2015) point out that bandwidths based on these regularized selectors tend to be too small in their empirical analysis
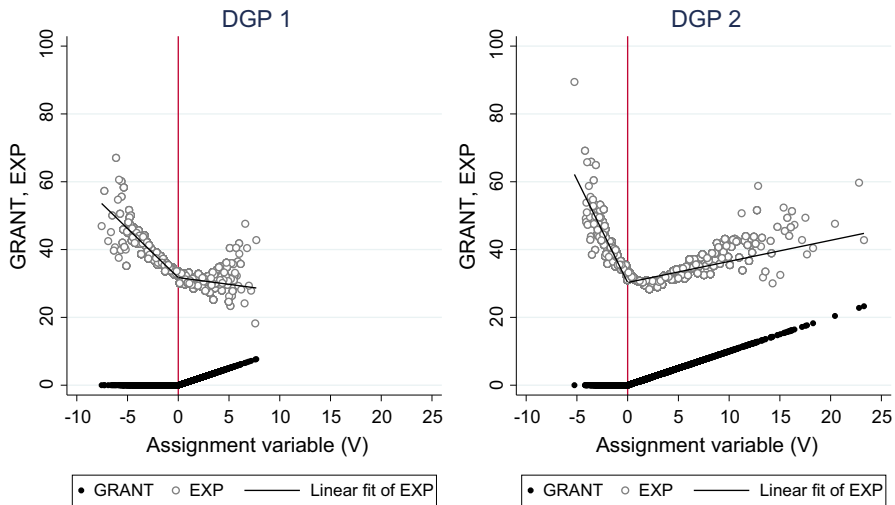
Fig. 4 Plots and linear fits with the stylized fiscal equalization. *Note* Optimal data-driven plots are generated with the default setting of `rdplot` command using Stata 13, which is developed in Calonico et al. (2015)

with an RK design. I therefore present RK estimation results with and without the regularization term in CCT's bandwidth selector. I also provide estimation results with CCT's optimal bandwidth but without CCT's bias-correction procedure.

When it comes to the dependent variable, I use not only total expenditure itself but the residuals obtained by regressing total expenditure on observed covariates CAP, CAP^2, X, and X^2 using the entire sample. The latter estimation enables me to consider how the performance of RK estimation changes if I directly control for observable confounding nonlinearity. The unobserved covariate U, which generates both endogeneity and confounding nonlinearity, is not controlled for in any of the cases.

I present the rejection rates of the null hypotheses of $\beta_1 = 1$ (size of test) and $\beta_1 = 0$ (power of test), as well as the means and standard deviations of $\widehat{\beta_1}$. Test levels are set at a conventional 5 percent, and I use both conventional and CCT's bias-corrected estimators and counterpart standard errors.

Before presenting the results of the MC simulations, Fig. 4 presents an example of the plots and the linear fits of EXP and GRANT against V with simulated data based on DGP 1 and 2. The figure shows that in both DGP 1 and 2 the slope changes of linear fits are greater than the slope changes of GRANT at the kink point, implying biases in an RK estimate with a simple linear polynomial regression.

3.4 Results

The results of MC simulations for DGP 1 are presented in Table 1. First, row 1 shows that RK estimates using a linear polynomial model are severely biased (i.e., far from one) when the entire sample is used for estimation, but RK estimates using a quadratic polynomial model with the same sample exhibit no bias. Rows 2 and 4 suggest that the

Table 1 Monte Carlo simulations with DGP 1

Bandwidth choice	Regularization	Bias correction	(I) Linear polynomial						(II) Quadratic polynomial					
			Estimates (mean)	SD	Size $H_0: \beta=1$	Power $H_0: \beta=0$	Bandwidth (mean)	Obs. (mean)	Estimates (mean)	SD	Size $H_0: \beta=1$	Power $H_0: \beta=0$	Bandwidth (mean)	Obs. (mean)
No covariates														
(1) Global	–	No	2.648	0.179	1	1	–	10,000	0.995	0.482	0.0450	0.510	–	10,000
(2) CCT	No	No	1.586	0.860	0.330	0.704	1.873	6053	1.096	1.714	0.0780	0.236	3.337	8318
(3) CCT	No	Yes	1.090	1.377	0.0410	0.205	1.873	6053	1.143	2.480	0.0540	0.0880	3.337	8318
(4) CCT	Yes	No	1.393	1.285	0.0710	0.275	1.034	3940	1.073	2.514	0.0440	0.0680	1.643	5860
(5) CCT	Yes	Yes	1.067	1.826	0.0420	0.0940	1.034	3940	1.093	3.168	0.0370	0.0460	1.643	5860
Conditioning on CAP, CAP^2, X, and X^2														
(6) Global	–	No	1.437	0.0426	1	1	–	10,000	0.994	0.121	0.0540	1	–	10,000
(7) CCT	No	No	1.127	0.194	0.302	0.986	1.594	5439	1.009	0.338	0.0640	0.915	2.903	7990
(8) CCT	No	Yes	1.003	0.294	0.0490	0.925	1.594	5439	1.017	0.487	0.0460	0.591	2.903	7990
(9) CCT	Yes	No	1.079	0.273	0.0610	0.951	0.922	3547	1.021	0.528	0.0460	0.564	1.490	5419
(10) CCT	Yes	Yes	0.994	0.383	0.0460	0.792	0.922	3547	1.026	0.668	0.0370	0.381	1.490	5419

The number of simulations is 1,000. Standard deviations (SD) of estimates are in parenthesis. The null hypothesis is tested with a 5% significance level. "Regularization" means that regularization is incorporated when CCT's optimal bandwidth is calculated. "Bias correction" means CCT's bias-correction procedure is introduced in the RK estimator. See further details in Calonico et al. (2014b) and Card et al. (2015). For RK estimation, rdrobust command (Calonico et al. 2014a) is used

estimation bias in a linear polynomial regression remains if CCT's optimal bandwidth is adopted but bias-correction procedure is not incorporated. On the other hand, the means of estimates are close to one and test sizes are around 5 percent when the bias-corrected estimator is used (rows 3 and 5, column I). An important problem, however, is that test power becomes very low if CCT's optimal bandwidth is used. This tendency is stronger when a quadratic polynomial is used for estimation, the regularization term is introduced in CCT's procedures, and/or bias-correction and robust estimation are incorporated (rows 2–5). Increased imprecision in RK estimation by these measures is theoretically expected, and the MC results suggest that this problem can indeed be very serious in practice.

Second, rows 6–10 present the same simulation results but with a different outcome variable, that is, the residuals obtained by regressing expenditure on the observed covariates X, CAP and their quadratic terms. Row 6 shows that RK estimation with a linear polynomial and the whole sample still generates biased estimates, but a quadratic polynomial regression with the sample works very well as before. Rows 7 and 9 suggest that conventional (not bias-corrected) estimates with CCT's optimal bandwidth but without regularization are still modestly biased when a linear polynomial is used, but counterpart quadratic polynomial regressions perform relatively well with much higher power than the "no covariates" case. CCT's bias-corrected estimator and robust inference in rows 8 and 10 also work fine: The test size is close to 5 percent and the test power is significantly improved compared with the "no covariates" case.

Next, the results of MC simulations for DGP 2 are provided in Table 2. Overall implications from these results are similar to those based on DGP 1. First, RK estimates with the entire sample are severely biased in the "no covariates" case (row 1). This bias is mitigated or even eliminated when CCT's optimal bandwidth and bias-correction are adopted, but at the cost of much lower test power (rows 2–5). This trade-off between bias and variance, however, almost disappears when RK estimation is implemented after conditioning on covariates (rows 6–10).

When it comes to differences between RK estimates from DGP 1 and DGP 2, quadratic RK estimation without covariates seems to work more poorly in DGP 2 with both global and CCT's bandwidths (rows 1–5, column II). This may imply that the ability of the quadratic polynomial to control for confounding nonlinearity depends on how the nonlinearity is generated. In the specific examples above, simple U-shape confounding nonlinearity in DGP 1 seems to be more easily controlled for by a quadratic polynomial.

Table 5 in Appendix "MC Simulation results for OLS estimation" provides additional simulation results with which I show OLS estimates are "precisely biased." The online appendices of Ando (2013), the working paper version of this paper, also show that MC simulations with higher-order (third- and fourth-order) polynomials result in seriously imprecise estimates, and a larger sample leads to more precision and larger power. Ando (2013) also presents the results of MC simulations with a simpler DGP where the fiscal equalization scheme is not incorporated into the models.

Overall, it is hard to fully demonstrate the validity of RK estimation at least in these specific DGPs when we do not include any relevant covariates in the regressors. The simulation results suggest that RK estimation without covariates may generate biased estimates, particularly when a linear polynomial regression is used, because its

Table 2 Monte Carlo simulations with DGP 2

Bandwidth choice	Regularization	Bias correction	(I) Linear polynomial						(II) Quadratic polynomial					
			Estimates (mean)	SD	Size $H_0: \beta=1$	Power $H_0: \beta=0$	Bandwidth (mean)	Obs. (mean)	Estimates (mean)	SD	Size $H_0: \beta=1$	Power $H_0: \beta=0$	Bandwidth (mean)	Obs. (mean)
No covariates														
(1) Global	–	No	5.779	0.258	1	1	–	10,000	3.722	0.724	0.967	0.996	–	10,000
(2) CCT	No	No	2.115	1.586	0.331	0.572	1.276	3632	1.049	2.886	0.0780	0.173	2.429	5806
(3) CCT	No	Yes	1.011	2.303	0.0590	0.131	1.276	3632	0.976	4.218	0.0510	0.0660	2.429	5806
(4) CCT	Yes	No	1.791	2.174	0.0970	0.225	0.772	2385	0.877	4.283	0.0620	0.0650	1.216	3615
(5) CCT	Yes	Yes	1.024	3.000	0.0440	0.0670	0.772	2385	0.864	5.436	0.0540	0.0500	1.216	3615
Conditioning on CAP, CAP^2, X and X^2														
(6) Global	–	No	0.986	0.0381	0.0820	1	–	10,000	1.048	0.101	0.0830	1	–	10,000
(7) CCT	No	No	0.991	0.151	0.0590	0.993	2.054	5027	0.985	0.379	0.0680	0.880	2.599	6002
(8) CCT	No	Yes	0.995	0.276	0.0410	0.905	2.054	5027	0.968	0.639	0.0510	0.540	2.599	6002
(9) CCT	Yes	No	0.992	0.271	0.0430	0.940	0.830	2551	0.980	0.582	0.0440	0.498	1.271	3756
(10) CCT	Yes	Yes	0.997	0.390	0.0370	0.777	0.830	2551	0.974	0.748	0.0390	0.318	1.271	3756

The number of simulations is 1,000. Standard deviations (SD) of estimates are in parenthesis. The null hypothesis is tested with a 5% significance level. "Regularization" means that regularization is incorporated when CCT's optimal bandwidth is calculated. "Bias correction" means CCT's bias-correction procedure is introduced in the RK estimator. See further details in Calonico et al. (2014b) and Card et al. (2015). For RK estimation, `rdrobust` command (Calonico et al. 2014a) is used

functional form does not take into account a confounding nonlinear relation between an assignment variable and an outcome variable. This bias could be mitigated or even removed by using CCT's optimal bandwidth and bias-correction, but the resulting highly imprecise and often statistically insignificant estimates make it difficult to obtain a robust conclusion about the causal effect of interest. The inclusion of covariates may significantly improve RK estimations.

Despite these problems that may undermine the internal validity of estimation as well as inherently limited external validity as a local "treatment on the treated" parameter, it can be argued that RK designs still have some advantages over standard OLS regression as a method of causal inference. First, RK designs can potentially avoid the endogeneity bias that stems from many unobserved or unrecognized correlations by exploiting an explicit source of identification and somewhat testable identifying assumptions. Second, the availability of robustness checks using different model specifications and bandwidth choices is also an attractive feature. In the following section, I examine the usefulness of RK designs by using a sample of Japanese municipality panel data and exploiting a kinked formula in the Japanese fiscal equalization scheme.

4 An empirical application

4.1 Institutional setting

In this section, I apply an RK design to a real-world situation. I exploit the kinked formula of Japanese fiscal equalization grants to identify the effect of these grants on local expenditure.

As mentioned in the last section, the Japanese fiscal equalization scheme allocates general grants to local governments (prefectures and municipalities) in order to compensate for the "need-capacity gap" of each local government and ensure a certain standard of local services for all citizens. This fiscal equalization grant is called a Local Allocation Tax (LAT) grant.[12] The detailed allocation mechanism of LAT grants is fairly complicated, but the basic framework can be explained as follows.

First, the national-level total amount of LAT grants is determined based on the amount of central tax revenues and political and bureaucratic processes in the central government.[13] Second, the LAT grant is distributed to individual local bodies based on the same formulas (8) and (9) in Sect. 3.1. In the Japanese fiscal equalization scheme, NEED is officially referred to as "*Standard Fiscal Need*" and is calculated annually by the central government. CAP is officially referred to as *Standard Fiscal Revenue* and is also calculated annually by the central government.[14]

---

[12] The confusing name "Local Allocation Tax" comes from the fact that the total amount of the grants is, in principle, a given proportion of revenues from five national taxes.

[13] According to the legal framework in which they are described, the total amount of the LAT grants is stipulated to be a fraction of the revenues from five major national taxes. In reality, however, the total amount is also influenced by other socioeconomic and political factors.

[14] See Appendix "2" and "3" for more detailed definitions and institutional descriptions of NEED (or *Standard Fiscal Need*) and CAP (or *Standard Fiscal Revenue*).

The assumed DGP under this scheme is similar to the DGPs in the MC simulations. That is, I expect a smooth nonlinear relationship between the need-capacity gap (V) and local expenditure (EXP) because various covariates should affect both need-capacity gap and local expenditure in various, often nonlinear, ways (see also footnote 11). One difference from the DGPs in the MC simulations is that observed and unobserved variables (X and U in Fig. 3) should affect both NEED and CAP, whereas the DGPs in the MC simulations are constructed such that these variables affect only NEED. This difference, however, does not affect my empirical strategy of employing an RK design that exploits the deterministic kink in the assignment variable V.

Because the objective of this paper is to investigate the performance of an RK design with a confounding nonlinearity, it is preferable to use a clean dataset that does not contain unnecessary noise and to establish some expected bounds of the treatment effect. From this standpoint, there are at least two advantages in studying the kinked assignment of Japanese fiscal equalization grants. First, the problem of potential endogenous sorting can be almost completely ignored because my assignment variable is an indicator, which is calculated by the central government and mostly unmanipulable by local governments. Second, because it is a well-known fact that Japanese municipalities have relatively homogeneous revenue systems regardless of the relative size of their respective LAT grants, it is expected that grant effects on total expenditure will not be much different from one.[15]

4.2 Data and preliminary investigation

In estimations, I use the panel data for cities (*shi*) covering fiscal years 1980–1999.[16] I exclude from the sample the cities which experienced amalgamation during the sample period because merged municipalities follow a special fiscal equalization scheme, but a large part of the cities remain in the sample. All of the fiscal data are from *Reports on the Municipal Public Finance* (*Shichoson-betsu Kessan Jokyo Shirabe*), which are published annually by the Ministry of Internal Affairs and Communications (MIC). When it comes to observed predetermined covariates, I use revenue capacity[17], population, population density, population ratios of the elderly cohort and the young

---

[15] Because most municipalities adopted similar tax rates and similar tax bases during the sample period (MIC 2012), a marginal increase in grants should have mainly led to a marginal increase in local expenditure. At the same time, a greater than one-to-one effect may be possible because some local public services are partly financed by earmarked matching grants: One-unit increase in LAT grants can lead to one-unit direct increase in local expenditure *and* some additional increase in local expenditure by induced matching grants. See MIC (2003) on the trends in local expenditures in 1990s. MIC (2003) shows that administration, welfare, construction, and education were major expenditure items, and welfare and debt payments grew fastest during the period. Ando (2015) investigates the effects of LAT grants on disaggregated expenditures, although the data used are restricted to fiscal years 2003–2004 because of a different research design.

[16] Japanese municipalities consist of cities (*shi*), towns (*cho*), villages (*son*), and special districts in Tokyo (*ku*). Because the duties of municipalities differ depending on their type, the levels of expenditure per capita are also affected by the type of municipality in question. I therefore use only city (*shi*) data, excluding other types of municipalities.

[17] The revenue capacity variable is similar to CAP (or *Standard Fiscal Revenue*), but some modifications have been made in order to reflect real revenue capacity of municipalities. See Appendix "Description of fiscal variables" for further details.

Table 3 Summary statistics

Variable	Obs	Mean	SD	Min	Max
Outcome (Thousand yen)					
Total expenditure	12,666	308.98	96.58	155.25	1506.90
Total revenue (excluding the LAT grant)	12,623	269.06	71.56	125.33	1167.13
Treatment (Thousand yen)					
LAT grant	12,623	48.17	46.84	0.00	408.26
Assignment (Thousand yen)					
Need-capacity gap	12,666	45.59	50.29	−111.15	408.50
Covariates					
Revenue capacity (modified, thousand yen)*	12,666	117.92	40.05	24.02	320.32
Population	12,666	102,890	103,892	6178	810,482
Population density (pop/km^2)	12,666	1653.98	2306.80	20.35	14,131.37
Population ratio (age 0–15, %)	12,666	19.35	3.41	9.11	32.47
Population ratio (age 65–, %)	12,666	13.10	4.54	3.67	32.42
Sectoral ratio (primary industry, %)**	12,666	8.90	7.96	0.10	46.78
Sectoral ratio (tertiary industry, %)***	12,666	57.06	10.01	26.63	85.05

All fiscal variables are divided by population, meaning that they are per capita values. The fiscal variables are also deflated by CPI (the reference year is 2005). There are some missing values for the LAT grant. Sources: Reports on the Municipal Public Finance, Census, and CPI

* See Appendix "RK estimation validity check" for a precise definition of modified revenue capacity as a predetermined covariate.

** The primary Sector consists of agriculture, forestry, fisheries and mining.

*** The tertiary sector includes all the sectors that are not included in the primary sector and secondary sectors (construction and mining)

cohort, and the sectoral ratios of employment. All the covariates, except for revenue capacity, are from Census data. Because Census data are only available for every fifth year, I impute annual data by linear interpolation. For additional details concerning data construction, see Appendix "Description of data arrangement."

Table 3 shows the summary statistics of the variables that I use for the empirical application in this section. All the fiscal variables are expressed as per capita values and deflated by Consumer Price Index (CPI: the reference year is 2005) published by MIC. On average, the size of the LAT grant is about 16 % of total expenditure. The sum of revenue capacity and the LAT grant is smaller than total expenditure because there are other important fiscal revenues, including conditional grants from the central and prefectural governments and prefectures and debt financing.

Before proceeding to econometric analysis, I conduct several preliminary analyses in order to examine the validity of my identification strategy. First, Fig. 5 plots the LAT grant and total expenditure against the need-capacity gap for municipalities (only cities). This graph indicates that the LAT grant has a clear deterministic kink at the threshold and the size of the grant is not negligible for many LAT-receiving munici- palities. The linear fits of total expenditure based on RK estimation with a first-order polynomial show that the size of an estimated kink decreases when the bandwidth is
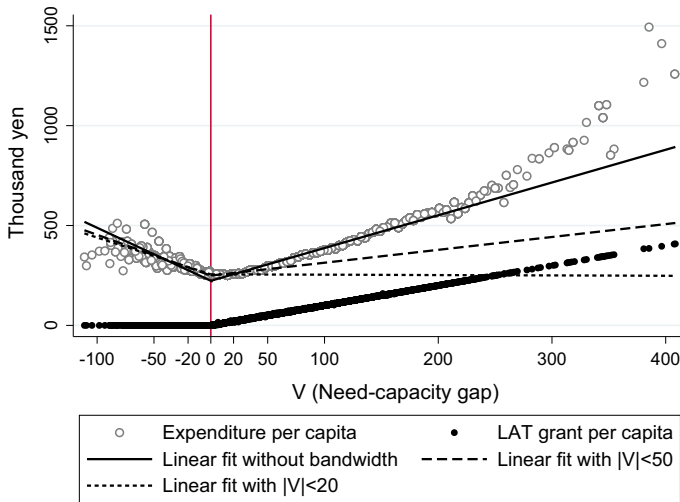
Fig. 5 Total expenditure and LAT grant against need-capacity gap. *Notes* An optimal data-driven plot is generated by the default setting of `rdplot` command with Stata 13, which is developed in Calonico et al. (2015). Linear fits of expenditure per capita are obtained by RK estimation with a linear polynomial model that imposes continuity at the cutoff. *Sources*: Reports on the Municipal Public Finance, Census, and CPI
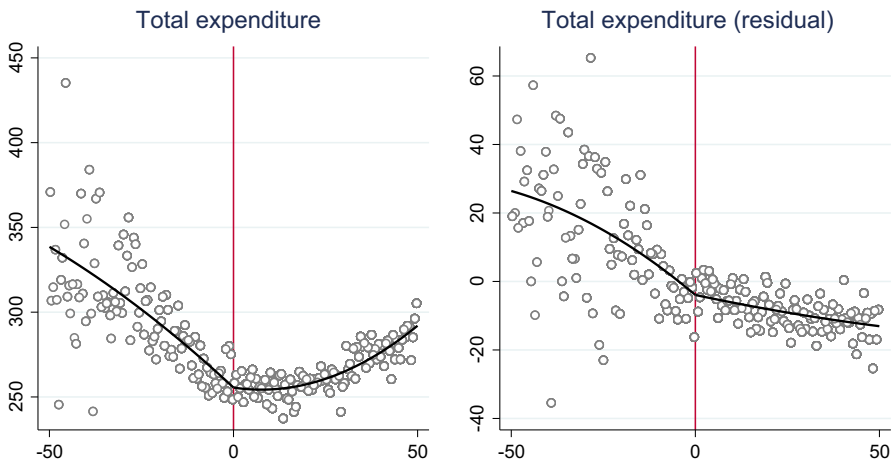

Fig. 6 Total expenditure against need-capacity gap ($bandwidth\ |V| < 50$). *Note* Optimal data-driven plots are generated by the default setting of `rdplot` command with Stata 13, which is developed in Calonico et al. (2015). Quadratic fits are based on the quadratic RK model that imposes continuity at the cutoff. The residual variable in the *right graph* is obtained by regressing total expenditure on all covariates listed in Table 3, their quadratic terms, and year dummies using the whole sample

shifted from $|V_i| < 50$ to $|V_i| < 20$, implying the existence of some confounding nonlinear relation between the assignment variable and expenditure per capita.

Second, Fig. 6 shows the plots and quadratic fits of total expenditure (my outcome variable) and its residual against the assignment variable with the bandwidth $|V_i| < 50$. The residual variable in the graph on the right is obtained by regressing total

expenditure on all covariates, their quadratic terms, and year dummies using the whole sample. This graph shows that the observed covariates effectively eliminate some nonlinearity within the bandwidth. Although a quadratic fit appears to control for this nonlinearity in the graph on the left, these two graphs also suggest that RK estimation without covariates may suffer from bias due to confounding nonlinearity, particularly when a linear RK model is used for estimation.

Third, a key identifying assumption for a valid RK design is that the density of the assignment variable is continuously differentiable at the threshold. Since the LAT grant is calculated by centrally determined uniform formulas, there is little possibility that municipalities or the central government can precisely manipulate the need-capacity gap around the threshold. It may be suspected, however, that some institutional settings or unknown factors systematically affect the determination of whether or not a given municipality near the threshold becomes an LAT-grant receiver. I therefore conduct a density test analogous to that proposed by McCrary (2008) for the RD design. Both estimation results and graphical analysis indicate that the density of the need-capacity gap is smooth at the threshold. These results are given in Appendix "Smooth density of the assignment variable."

Finally, according to Card et al. (2015), an important implication under the required conditions for a valid RK design is that any predetermined covariate should have a conditional distribution which evolves smoothly around the threshold. In other words, there should be no kink at the threshold for any predetermined covariate against the assignment variable.

However, the argument presented in Sect. 2 implies that a smooth nonlinear relation between a covariate and an assignment variable around the kink point could be mistakenly captured as a kink using RK estimation. It may thus be hard to assert that there are no kinks whatsoever at the threshold for any covariate. The plots and quadratic fits of covariates against the assignment variable within a certain bandwidth, however, at least indicate that no such kinks are visually apparent in the graphical representation of the data except for the covariate of population. Graphs are provided in Fig. 11 in Appendix "Distributions of covariates against V."

4.3 Results

Table 4 presents the results of RK estimates generated by applying linear and quadratic polynomials to the empirical model (3). I provide RK estimates with varying bandwidths and CCT's procedures. As MC simulations, I provide estimation results with and without regularization and bias correction when I adopt CCT's procedures.

First, I examine RK estimation without covariates (rows 1–8). When a linear polynomial is used (column I), estimates are around two when relatively narrow bandwidths ($|V| = 20, 30,$ or 40) are used and then become small and statistically insignificant when bandwidth is $|V| = 10$, probably because of imprecision. When it comes to RK estimation with CCT's procedures, estimates are significantly different from zero when the regularization term is not introduced (rows 5, 6), and the size of the estimate is 1.37 if the bias-correction procedure is applied (row 6). On the other hand, CCT's optimal bandwidth with the regularization term makes RK estimates not statistically

Table 4 RK estimates for total expenditure

	Bandwidth choice	Regularization	Bias correction	(I) Linear polynomial				(II) Quadratic polynomial			
				Estimates	SE	Optimal bandwidth	Obs.	Estimates	SE	Opriml bandwidth	Obs.
No covariates											
(1)	|V| <40	–	No	2.554 * **	(0.178)	–	6430	0.864	(0.615)	–	6430
(2)	|V| <30	–	No	2.259 * **	(0.229)	–	5013	0.122	(0.844)	–	5013
(3)	|V| <20	–	No	1.861 * **	(0.379)	–	3451	−1.719	(1.431)	–	3451
(4)	|V| <10	–	No	0.125	(0.941)	–	1741	−3.385	(4.006)	–	1741
(5)	CCT	No	No	1.746 * **	(0.332)	22.09	3765	−1.199	(1.318)	21.04	3628
(6)	CCT	No	Yes	1.370 * **	(0.522)	22.09	3765	−2.233	(1.984)	21.04	3628
(7)	CCT	Yes	No	0.377	(0.891)	10.42	1818	−0.405	(1.643)	17.67	3018
(8)	CCT	Yes	Yes	−0.457	(1.159)	10.42	1818	−1.193	(2.151)	17.67	3018
Conditioning on full covariates											
(9)	|V| <40	–	No	0.622 * **	(0.141)	–	6430	1.135 * *	(0.470)	–	6430
(10)	|V| <30	–	No	0.695 * **	(0.184)	–	5013	1.133*	(0.663)	–	5013
(11)	|V| <20	–	No	1.096 * **	(0.297)	–	3451	−0.436	(1.087)	–	3451
(12)	|V| <10	–	No	0.139	(0.720)	–	1741	0.706	(3.013)	–	1741
(13)	CCT	No	No	0.714 * **	(0.265)	21.91	3749	0.956	(0.642)	30.43	5079
(14)	CCT	No	Yes	0.935*	(0.481)	21.91	3749	0.549	(1.145)	30.43	5079
(15)	CCT	Yes	No	0.436	(0.534)	11.86	2061	0.013	(1.163)	19.18	3280
(16)	CCT	Yes	Yes	0.360	(0.829)	11.86	2061	−0.163	(1.538)	19.18	3280

Standard errors are clustered by the municipality level for conventional RK estimation. For RK estimation with CCT's bias correction, robust inference proposed by CCT is used for the estimation of standard errors. In the case of "conditioning on full covariates," residuals after regressing total expenditure on the covariates listed in Table 3, their quadratic terms, and year dummies with the whole sample are used for a dependent variable

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

significant, possibly due to a very narrow bandwidth of 10.42 (rows 7–8). Finally, when I adopt a quadratic polynomial (rows 1–8 in column II), RK estimates without covariates are never significantly different from zero, possibly due to imprecise quadratic RK estimation.

Second, in RK estimation with covariates (rows 9–16), linear RK estimates (column I) are around 0.6 to 1.1 and statistically significant in rows 9–11 and 13–14. These estimates are considerably smaller than their counterpart estimates in the "no covariates" case. This is suggestive evidence that RK estimates without conditioning on covariates suffer from biases from confounding nonlinearity around the kink point, and the introduction of covariates in RK estimation alleviates this problem. Linear RK estimates with narrower bandwidths ($|V| < 10$ in row 12 and $|V| < 11.86$ in rows 15–16) are again not significantly different from zero, probably because of imprecision. Finally, quadratic RK estimates (rows 9–16 in column II) are mostly not significantly different from zero as in the "no covariates" case, but rows 9, 10, and 13 show that estimates with larger bandwidths ($|V| > 30, 40$, and 30.43, respectively) are around 1.0–1.1 and their standard errors are relatively small. These results imply that the precision of quadratic RK estimation increases by conditioning on covariates. In other words, low precision may be a major reason why quadratic RK estimation without covariates or smaller bandwidths yields statistically insignificant results.

These estimation results are more or less compatible with the implications of the MC simulations. First of all, the introduction of covariates can reduce the biases in RK estimates, particularly estimates with a linear polynomial. Second, RK estimates with a quadratic polynomial fluctuate a lot and are statistically insignificant when no covariates are introduced and CCT's procedures are used. This is compatible with the MC results, which indicated that the standard deviation of RK estimates is often too high and power too low if a quadratic polynomial is used with CCT's bias-correction procedure.

In Table 7 in Appendix "OLS results from the empirical application," I also present the results of OLS estimates. OLS estimates, which are obtained by simply regressing the total expenditure on the grant, change considerably if I introduce covariates or fixed effects, implying severe endogeneity bias. It is difficult to draw useful conclusions about the magnitude of the treatment effect on the basis of these OLS estimates alone. Ando (2013) provides RK estimates with higher-order polynomials. The RK estimates with a third- and fourth-order polynomial are very unstable and mostly statistically insignificant. This could be because RK estimates with higher-order polynomials tend to be very imprecise. Ando (2013) also presents RK estimates for total revenue minus the LAT grant and shows that no effect is observed once covariates are controlled for.

Overall, relatively reliable RK estimation using a lower-order polynomial model with covariates suggests that the effect of the grant on total expenditure is roughly around one. As I have already discussed, this result appears reasonable on the basis of our institutional knowledge of Japanese local public finance. When it comes to the validity of the various RK estimations examined, the empirical results and the MC simulations suggest that an RK design employing a local linear or quadratic regression with a modestly small bandwidth and additional covariates provides arguably the most reliable estimate. CCT's procedures of bias correction and robust inference should be

used with caution because it often produces imprecise estimation, particularly when regularization is incorporated.

5 Detection of confounding nonlinearity

My MC simulations and real-world application show that RK estimation can suffer from a bias caused by confounding nonlinearity around the cutoff and some procedures for this such as a smaller bandwidth, a quadratic polynomial, and CCT's procedure instead result in highly imprecise and often impractical estimation.

In this section, I provide one complementary method to detect such bias-inducing confounding nonlinearity in RK designs. The idea is quite similar to the permutation inference with placebo RK estimates proposed by Ganong and Jäger (2015). As a complementary analysis of a formal statistical test with permutation, however, I recommend simply showing the distribution of placebo RK estimates along placebo cutoff values in order to detect bias caused by confounding nonlinearity. This method has already been applied in some studies such as Engström et al. (2015) and Ando (2015).

Referring to Ando (2013), Card et al. (2015) also conduct Ganong and Jäger (2015)'s permutation test after controlling for confounding nonlinearity with observed covariates in Appendix "Revenue capacity (modified for a predetermined covariate)" of the working paper version. This procedure is required because the original permutation test may not be informative when the curvature of the conditional expectation function $E[Y|V = v]$ changes. As a corresponding analysis, I also present the distribution of placebo RK estimates conditional on observed covariates. This analysis enables me to check how controlling for observed covariates can mitigate bias caused by confounding nonlinearity.

I implement placebo RK estimation with a varying placebo cutoff c_{plcb} using the following model:

$$Y_i = \alpha_0 + \beta_0 \cdot D_i + \sum_{p=1}^{\bar{p}} [\alpha_p(v_i - c_{plcb})^p$$
$$+ \beta_p(v_i - c_{plcb})^p \cdot D_i] + \varepsilon \quad \text{where } |v_i - c_{plcb}| \leq h, \tag{11}$$

where $D_i = 1$ if $(v_i - c_{plcb}) > 0$ and otherwise 0. I apply this placebo test to a simulated set of data with DGP 2 in Sect. 3 and the dataset used in Sect. 4. Because the motivation of this placebo test is to detect confounding nonlinearity, I do not use CCT's bias-correction procedure for RK estimation.

Results for simulated data with DGP 2 are shown in Fig. 7. Here I fix the bandwidth h at 2 and change the placebo cutoff c_{plcb} from $-3, -2.8\ldots$ to 3. For estimation, I use two outcome variables, that is, the original outcome EPX and residuals after regressing EXP on observed covariates CAP, CAP^2, X and X^2.

The two upper graphs in Fig. 7 show placebo RK estimates for EXP with local linear and quadratic regression models. Most placebo estimates with a linear RK model are significantly different from zero (or one, a true "effect" in this DGP) regardless
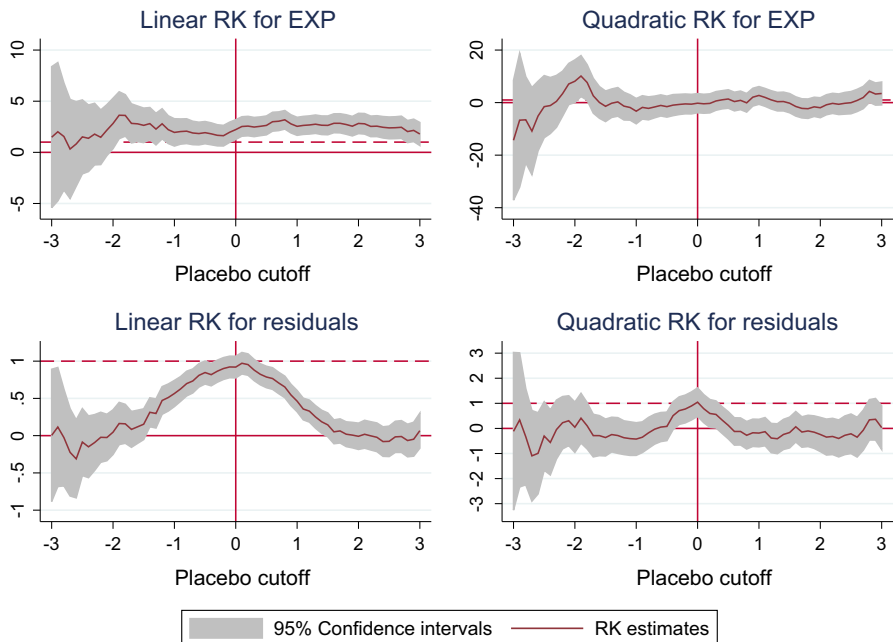
Fig. 7 Placebo RK estimates for simulated data. *Notes* Data for analysis are generated based on DGP 2 in Sect. 3. All placebo estimates and their confidence intervals are obtained by Eq. (11) with $p = 1$ or 2. The bandwidth h is fixed at 2. CCT's procedure for bias correction and a robust confidence interval is not applied to these placebo estimates. A *dashed line in each graph* indicates a value of one

of the values of the placebo cutoff, indicating that there exists severe confounding nonlinearity and that a true RK estimate with $c = 0$ suffers from bias because of it. On the other hand, placebo estimates with a quadratic RK model are almost always statistically insignificant, and their confidence intervals are too large, suggesting highly imprecise estimation.

The two lower graphs, in turn, present placebo estimates when the dependent variable is the residual. Both graphs show that placebo estimates are closer to one, a true "effect" of GRANT, when placebo cutoffs c_{plcb} are closer to the true cutoff of zero. On the other hand, placebo estimates are around zero and statistically not different from zero when placebo cutoffs are further away from the true cutoff. This strongly implies that the true RK estimate is not severely biased, although the possibility of confounding nonlinearity exactly at the true cutoff cannot be ruled out by this placebo test.[18]

Figure 8 present placebo RK estimates for total expenditure using the Japanese municipality data. I fix the bandwidth h at 20 and change the placebo cutoff c_{plcb} from -50, $-48 \ldots$ to 50. For dependent variables, I use the total expenditure and

---

[18] In addition, for linear RK estimation, placebo estimates are still significantly different from zero when placebo cutoffs are around ± 1. This, however, is presumably not due to confounding nonlinearity around these points. Rather, the placebo estimates with a linear RK model capture the nonlinearity caused by the true kink when the true cutoff is inside the bandwidth.
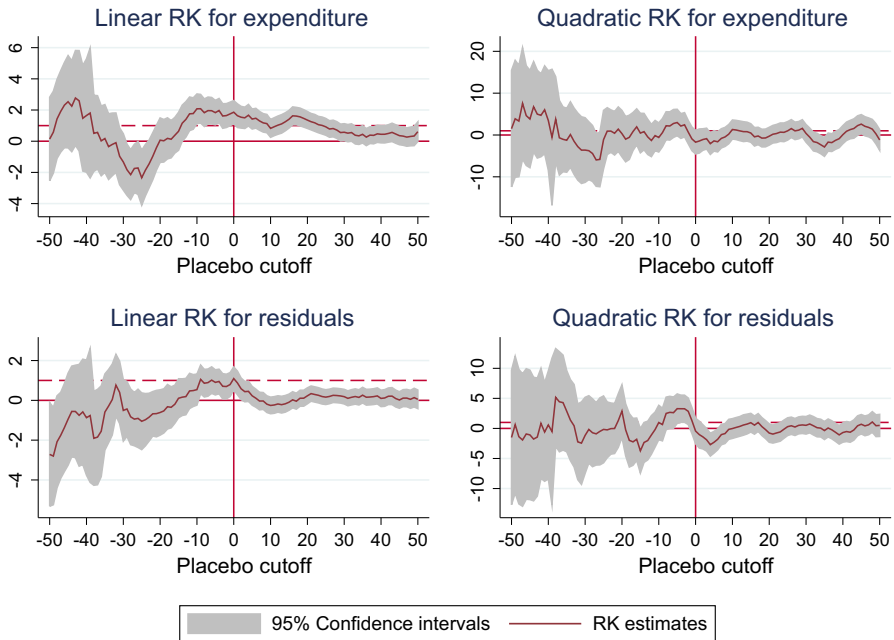
Fig. 8 Placebo RK estimates for total expenditure. *Notes* The dataset used in Sect. 4 is utilized for analysis. All placebo estimates and their confidence intervals are obtained with Eq. (11) with $p = 1$ or 2. The bandwidth h is fixed at 20. CCT's procedure for bias correction and a robust confidence interval is not applied to these placebo estimates. A *dashed line in each graph* indicates a value of one

its residual after regressing it on observed covariates listed in Table 3, their quadratic terms, and year dummies.

The implications of the graphs are similar to those of Fig. 7. The upper-left graph shows that placebo estimates for total expenditure with a linear RK model seem to be confounded by some nonlinearity because placebo estimates are often significantly different from zero when the placebo cutoffs are between -30 and 30. Placebo estimates for total expenditure with a quadratic RK model presented in the upper-right graph suggest that quadratic RK estimation yields overly wide confidence intervals due to high imprecision.

In contrast, placebo estimates for residuals with a linear RK model presented in the lower-left graph are significantly different from zero only when placebo cutoffs are close to the true cutoff of zero. In particular, placebo estimates are stably around zero when the placebo cutoffs are larger than ten, where the number of observations is relatively large. These results suggest that confounding nonlinearity around the true kink is effectively mitigated by covariates. Finally, placebo estimates with a quadratic RK model presented in the lower-right graph are not as precise as the counterpart graph in Fig. 7, possibly due to higher imprecision in quadratic RK estimation with real-world data.

In sum, the placebo test presented here provides an effective, though supplementary, graphical tool to detect the confounding nonlinearity that can cause bias in RK estima-

tion. For RK estimation to be plausible, placebo RK estimates should be reasonably close to zero at the points where true local RK estimation with a certain bandwidth may capture confounding nonlinearity if it exists.

6 Conclusion

Regression kink (RK) designs, which have several attractive features similar to those of regression discontinuity (RD) designs, are potentially quite useful, but the weakness of this approach has been largely ignored in the emerging applied studies. In order to investigate the validity of RK designs, I first examined the finite sample properties of RK estimation in the presence of a confounding smooth nonlinearity around the cutoff, using Monte Carlo (MC) simulations. Then, I applied an RK design to the study of the causal effects of fiscal equalization grants on the spending of local governments.

The results of the MC simulations suggested that RK estimation often resulted in biased estimates when there was a confounding nonlinear relation between an assignment variable and an outcome variable. Introduction of a higher-order polynomial, a smaller bandwidth, or Calonico et al. (2014b)'s bias-correction procedure could mitigate this bias, but they often resulted in an imprecise estimate and low power that might significantly undermine the results obtained. The simulations also provided evidence that the introduction of observed covariates can improve RK estimation, although the improvement may be insufficient if confounding nonlinearity is generated by unobserved covariates.

When it comes to the real-world application, I first found seemingly very large biases in the estimated effect of the equalization grant on local expenditure when I did not include any additional covariate in my RK estimation. After conditioning on observed covariates, however, relatively robust positive estimates close to one were obtained when a linear polynomial was used for RK estimation. RK estimation with a quadratic polynomial was less robust, but they also imply that the effect of the grant on local expenditure is not far from one.

In conclusion, my MC simulations and real-world application provide mixed answers concerning the usefulness of an RK design with a finite sample. On the one hand, it can be argued that RK analysis is more reliable than a simple pooled or fixed-effect OLS regression when it comes to causal interpretation because an RK design has an explicit identification strategy and transparent tools for validity and robustness checks. On the other hand, if unobserved covariates generate confounding nonlinearity around the cutoff point and the sample size around this point is relatively small, a biased or imprecise estimate may be obtained. In empirical studies employing an RK design, this possible weakness should be explicitly addressed by careful robustness checks and placebo tests. The construction of formal criteria for a sufficiently credible RK design is an important area for further study.

Appendix 1: MC simulation results for OLS estimation

See Table 5.

Table 5 MC Simulation results for OLS estimation

Obs.	Mean	SD	Size $H_0 : \beta = 1$	Power $H_0 : \beta = 0$
DGP 1				
No covariates				
1000	−1.866	0.0952	1	1
Conditioning on CAP, CAP^2, X, and X^2				
1000	1.849	0.0293	1	1
DGP 2				
No covariates				
1000	0.0125	0.0493	1	0.0640
Conditioning on CAP, CAP^2, X, and X^2				
1000	1.272	0.00801	1	1

Simulation settings are the same as the setting in Sect. 3. Coefficients are estimated with simple OLS where EXP is regressed on GRANT. Heteroskedasticity-robust standard errors are used for estimation

Appendix 2: Description of fiscal variables

Expenditure need: NEED

This index measures the cost of a "standard" level of local public services for a municipality. It is officially referred to as "*Standard Fiscal Need*" (*Kijun Zaisei Juyo Gaku*) and calculated annually by the Ministry of Internal Affairs and Communications. *Standard Fiscal Need* is calculated as follows:

$$\text{NEED}_i = \Sigma_k (\textit{Unit Cost}_k \times \textit{Measurement Unit}_{ik} \times \textit{Adjustment Coefficient}_{ik}),$$

where k expresses kth public service. *Unit Cost$_k$* is a kind of net standard cost per measurement unit for each service item. *Measurement Unit$_{ik}$* is in most cases the number or size of the beneficiaries of a particular service. *Adjustment Coefficient$_{ik}$* is a modification ratio that reflects the socioeconomic diversity of a local body and modifies the unit cost in order to make it fit the local body's socioeconomic circumstances.

Revenue capacity: CAP

This index measures the fiscal revenue capacity of a municipality before fiscal equalization. It is officially referred to as "*Standard Fiscal Revenue*" (*Kijun Zaisei Syunyu Gaku*) and calculated annually by the Ministry of Internal Affairs and Communications. CAP is calculated as follows:

$$\text{CAP}_i = \textit{Standard Tax Revenue}_i \times \frac{3}{4} + \textit{Local Transfer Tax, etc.}_i$$

where *Standard Tax Revenues* are estimated based on standard tax rates, standard tax collection rates, and estimated tax bases which are calculated using relevant statistics

or past tax revenues. *Local Transfer Tax, etc.* represents the sum of revenues from the *Local Transfer Tax* and the *Special Grant for Traffic Safety Measures*. In brief, CAP captures the potential amount of local general revenues before fiscal equalization, which cannot be manipulated by municipalities in the short run.

There are two main reasons that *Standard Tax Revenue* is multiplied by 3/4.[19] First, the remaining 1/4 of *Standard Tax Revenue* is excluded from the fiscal equalization process and left for municipalities so that they can cover some remaining fiscal needs that are not taken into account by the *Standard Fiscal Needs* calculation. Second, this portion of tax revenue is excluded from the fiscal equalization process so that municipalities have some incentive to increase their local tax revenues by enhancing local economic growth. In other words, if the exact amount of *Standard Tax Revenue* were taken into account in CAP, LAT-receiving local bodies would have less incentive to enhance local economic growth because the increase in *Standard Tax Revenue* caused by this economic growth would be completely canceled out by the decrease in the LAT grant.

Revenue capacity (modified for a predetermined covariate)

As is explained above, CAP itself does not represent "real" pre-equalization revenue capacity as it takes into account some policy objectives of the fiscal equalization scheme such as providing economic incentives to municipalities. We can, however, easily recover real pre-equalization revenue capacity by simply replacing 3/4 for 1 in the above definition of CAP.

When I use pre-equalization revenue capacity as a control variable in Sect. 4, I use this modified version of revenue capacity that reflects the real pre-equalization revenue capacity of municipalities. However, because available statistics are only CAP and *Local Transfer Tax*, I have to assume that revenue from *Special Grant for Traffic Safety Measures* is negligible. This assumption should not be a major problem because the amount of the *Special Grant for Traffic Safety Measures* is in general much smaller than the sum of *Standard Tax Revenues* and *Local Transfer Tax*.

I therefore estimate this "real" pre-equalization revenue capacity as follows:

$$
\begin{aligned}
&\mathrm{RealCAP}_i \\
&= Standard\,Tax\,Revenues_i \times 1 + Local\,Transfer\,Tax, etc._i \\
&= (\mathrm{CAP}_i - Local\,Transfer\,Tax, etc._i) \times 4/3 + Local\,Transfer\,Tax, etc._i \\
&\simeq (\mathrm{CAP}_i - Local\,Transfer\,Tax_i) \times 4/3 + Local\,Transfer\,Tax_i.
\end{aligned}
$$

Appendix 3: Stylized description of Japanese fiscal equalization

In this appendix, I explain the stylized features of the Japanese fiscal equalization scheme and describe how the kink based on the LAT grants is generated in more detail.

---

[19] See also Appendix "stylized description of Japanese fiscal equalization," where I describe the stylized features of the Japanese fiscal equalization grants with some equations and graphs.

In order to make this description as concise as possible, throughout this appendix I redefine CAP as follows:

$$\mathrm{CAP}_i = Standard\ Tax\ Revenue_i \times \frac{3}{4}.$$

In other words, compared with the actual definition in Appendix "Description of fiscal variables," CAP is simplified by dropping the second term (*Local Transfer Tax* and some miscellaneous revenues), which is actually much smaller than the first term (*Standard Tax Revenue*) in the majority of municipalities.

Then, further assuming that there are no additional revenues other than local tax revenues and LAT grants, the relation between pre-equalization standard revenue (denoted as *PreRev*) and post-equalization standard revenue (denoted as *PostRev*) can be expressed as follows:

$$PostRev_i = \begin{cases} PreRev_i, & \text{if } V_i \leq 0 \\ PreRev_i + \mathrm{GRANT}_i, & \text{if } V_i > 0. \end{cases}$$

By inserting $PreRev_i = Standard\ Tax\ Revenue_i$, $V_i = \mathrm{NEED}_i - \mathrm{CAP}_i$, $\mathrm{GRANT}_i = V_i = \mathrm{NEED}_i - \mathrm{CAP}_i$ (if $V_i > 0$), and the above definitions of CAP_i into these equations, they can be rewritten as

$$PostRev_i = \begin{cases} \mathrm{CAP}_i + Standard\ Tax\ Revenue_i \times \dfrac{1}{4}, & \text{if } \mathrm{CAP}_i \geq \mathrm{NEED}_i \\ \mathrm{NEED}_i + Standard\ Tax\ Revenue_i \times \dfrac{1}{4}, & \text{if } \mathrm{CAP}_i < \mathrm{NEED}_i. \end{cases} \tag{12}$$

These two equations represent an essential function of the Japanese fiscal equalization scheme. First, if CAP is larger than NEED, no LAT grant is distributed and post-equalization standard revenue is identical to the sum of CAP and *Standard Tax Revenue* × 1/4. Second, when CAP is smaller than NEED, the LAT grant ensures that municipalities receive the sum of NEED and *Standard Tax Revenue* × 1/4. In both cases, this additional amount, *Standard Tax Revenue* × 1/4, exists due to the fact that CAP is calculated by *Standard Tax Revenue* × 3/4 and the other 1/4 of *Standard Tax Revenue* is excluded from the fiscal equalization formula. Because of this excluded part of *Standard Tax Revenue*, which is officially referred to as "*Reserved Revenues*," a wealthier municipality is always wealthier even after fiscal equalization. Figure 9 presents actual scatter plots and local polynomial smoothing of *PreRev*, *PostRev*, and NEED against the assignment variable V. It illustrates how the LAT grant phases in at the cutoff point $V = 0$.

Notice that in this graph *PostRev* is well above NEED around $V = 0$. This implies that municipalities just after $V > 0$ have ample additional fiscal resources in excess of NEED. These additional fiscal resource come from the term *Standard Tax Revenue* × 1/4 in the second equation of (12). This fact benefits our empirical analysis because LAT grants, which phase in after $V > 0$, can be plausibly considered as "general" and "lump-sum" around the threshold, without any difficulties
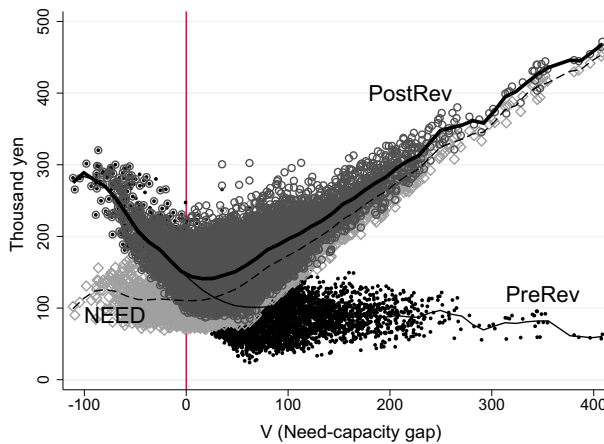
Fig. 9 Scatter plots of *PreRev*, *PostRev*, and NEED). *Notes* The same sample that is described in Sect. 4 is used for this scatter plot. The local polynomials are obtained using the `lpoly` command in Stata 13 with the default setting. *Sources*: reports on the Municipal Public Finance, Census, and CPI

caused by complicated institutional settings of these grants. I conclude this appendix by examining this issue in greater detail.

In this paper, I implicitly assume that LAT grants are "general" and "lump-sum" and local bodies have full discretion in their decision making on spending and taxation. In other words, I presuppose that an estimated coefficient can be straightforwardly interpreted as the effect of general lump-sum grants on local spending under the full discretion of local municipalities.

But it could be misleading to simply assume that LAT grants are completely general and lump-sum as the so-called Bradford-Oates equivalence theorem, and some previous empirical studies in the literature of local public finance have done. The LAT grant is nominally a general grant that a local body can spend on whatever it wants, but at the same time the LAT grant is the grant that guarantees every single municipality a sufficient amount of revenues to cover centrally determined "standard costs" for local public services, which is referred to as *Standard Fiscal Needs* and denoted as NEED in this paper. It is sometimes pointed out that the central government takes advantage of LAT grants to control local spending by arbitrarily adjusting NEED. In addition to these possibly "centralized" aspects of LAT grants, the provision of local public services is often strongly regulated by the central government through various centralized legal frameworks.

In sum, although local bodies do not have to strictly follow these centrally determined standards, they quite often cannot control their expenditures on some local public services because the basic legal and provisional frameworks of these local services are centrally determined. Hayashi (2000, 2006) provides critical reviews of empirical studies on flypaper effects in Japan and points out that these previous studies do not consider these institutional settings of the Japanese general grant and naively treat it as a general and lump-sum grant as studies of similar schemes in the USA do.

In fact, this obligatory and centralized feature of local public services is part of the institutional basis of LAT grants: Since the central government forces all local bodies to provide particular levels of local public services, fiscal resources for these services have to be guaranteed by the intergovernmental fiscal transfer which reflects the expected costs of these services. This feature of the LAT grant is officially referred to as a function of "*fiscal resource guarantee*."

According to Fig. 9, however, I would argue that LAT grants can be considered to be general and lump-sum around the threshold $V = 0$ regardless of the centralized features of these grants and local administration. In other words, around the threshold, $PostRev$ is well above NEED, and therefore, the relatively obligatory local public services that are reflected in the calculation of NEED can be financed even without the LAT grant. It is thus possible to assume that the marginal increase in the LAT grant around the threshold affects local bodies' expenditure in exactly the same way that standard general and lump-sum grants do.

Appendix 4: Description of data arrangement

Japan is a unitary state which has three tiers of administrative authorities: the central government, 47 prefectures, and 1750 municipalities (as of 2010). Municipalities are classified into four categories: cities (*shi*), towns (*cho*), villages (*son*), and special districts (*ku*). Cities are generally larger than towns and villages, and in principle the minimum population required to become a city is 50,000. Even if the population of a city becomes less than 50,000, however, it does not have to become a town or village. The 23 special districts are all located in Tokyo Metropolis (prefecture) and have similar duties to other municipalities but follow a different vertical fiscal equalization scheme managed by the prefecture. Cities and towns/villages have similar duties under the LAT grant fiscal equalization scheme, but cities have more responsibilities in some areas.

In this paper, I only use the datasets for cities, but not all cities are included in my analysis. First, I exclude so-called designated cities, which consisted of the 12 largest cities in Japan during the sample period. I drop these cities from the sample because their response to the marginal increase in their LAT grants might be institutionally different from other cities as a result of the fact that some of the duties normally assigned to prefectures are delegated to them and their administrative responsibility is thus larger than that of normal cities. Second, I also remove the cities that experienced amalgamation between 1975 and 1999 because the calculation of the LAT grants for these merged cities was affected by special measures. Because this special measure was in effect for 5 years after amalgamation, municipalities which merged before 1975 were not affected by this measure after 1980. Finally, there are some LAT-receiving municipalities whose need-capacity gap is apparently different from the amount of their LAT grant, possibly due to measurement errors or typos. Therefore, I drop 18 observations in which |Need-capacity gap − LAT grant per capita| is larger than 10,000 yen.

Appendix 5: RK estimation validity check

Smooth density of the assignment variable

Following a density test applied to an RD design in McCrary (2008), Card et al. (2015) presents a density test applied to an RK design using collapsed data with equal-sized bins based on an assignment variable. Two required variables in this collapsed data set are the number of observations in each bin and the midpoint values of the assignment variable in each bin. I use bins with width 2 and bandwidth [−50, +50], which is a benchmark bandwidth for local regressions in this paper. Table 6 shows that in each sample an RK estimate is statistically insignificant if the order of polynomial is equal to or larger than two. The value of Akaike information criteria (AIC) is smallest when the order of polynomial is two. Figure 10 graphically illustrates that there seems to be no kink at the threshold.

Table 6 RK estimates for the density of the assignment variable (*bin size = 2, bandwidth* |V| < 50)

Variables	Order of polynomial			
	(1) One	(2) Two	(3) Three	(4) Four
RK estimates	−2.217***	−1.191	−1.103	−1.120
	(0.667)	(1.164)	(2.934)	(5.066)
Observations	50	50	50	50
R^2	0.903	0.980	0.981	0.981
AIC	493.4	419.0	420.2	423.2

Heteroscedasticity-robust SE are in parenthesis
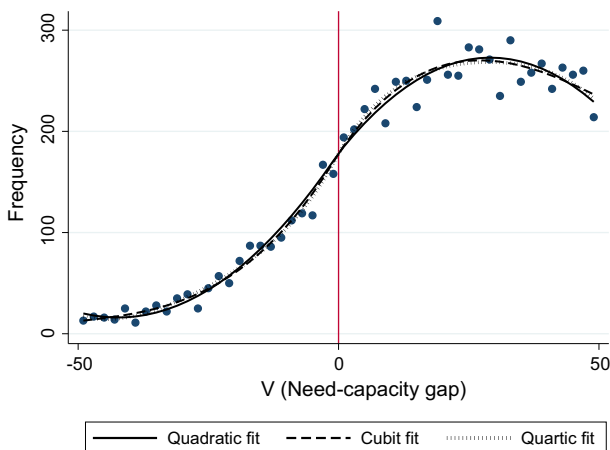*** $p < 0.01$

Fig. 10 Density of the assignment variable (*bin size = 2, bandwidth* |V| < 50). *Notes* Bin size is 2 and fitted *curves* are based on the quadratic, cubic, and quartic RK models that impose continuity at the cutoff

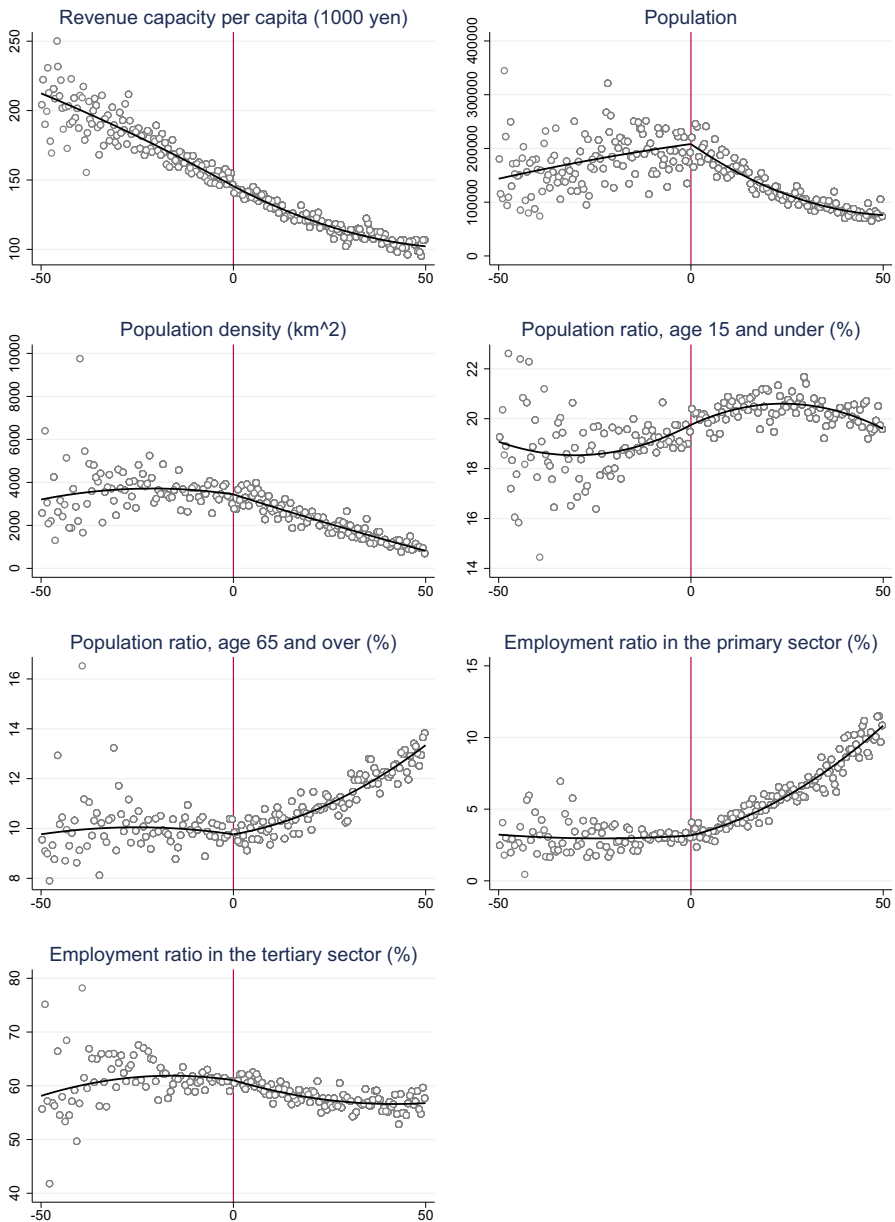## Distributions of covariates against *V*

See Fig. 11.

Fig. 11 Covariates against need-capacity gap (*bandwidth* |*V*| < 50). *Notes* Optimal data-driven plots are generated by the default setting of `rdplot` command with Stata 13, which is developed by Calonico et al. (2015). Quadratic fits are based on the quadratic RK model that imposes continuity at the cutoff

7 Appendix 6: OLS results from the empirical application

See Table 7.

Table 7 OLS estimates for total expenditure

	Estimate	SE	Covariates	Fixed effects	Obs.
(1)	1.321***	(0.084)	No	No	12,623
(2)	2.328***	(0.098)	Yes	No	12,623
(3)	1.582***	(0.100)	Yes	Yes	12,623

SE are clustered by the municipality level. "Covariates" used are those listed in Table 3 and their quadratic terms. "Fixed effects" are municipality fixed effects. Year fixed effects are incorporated in all specifications

*** $p < 0.01$

References

Altonji JG, Matzkin RL (2005) Cross section and panel data estimators for nonseparable models with endogenous regressors. Econometrica 73(4):1053–1102

Ando M (2013) How much should we trust regression-kink-design estimates? Department of Economics, working paper, Uppsala University, vol 2013, pp 22

Ando M (2015) Heterogeneous effects of fiscal equalization grants on local expenditures: evidence from two formula-based variations. IPSS working paper series (E), no. 25

Boadway R, Shah A (2007) Intergovernmental fiscal transfers: principle and practice. The World Bank, Washington

Bravo J (2011) The Effects of intergovernmental grants on local revenue: evidence from Chile. UC Economics Institute working paper, vol 393

Calonico S, Cattaneo MD, Titiunik R (2014a) Robust data-driven inference in the regression-discontinuity design. Stata J 14(4):909–946

Calonico S, Cattaneo MD, Titiunik R (2014b) Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82(6):2295–2326

Calonico S, Cattaneo MD, Titiunik R (2015) Optimal data-driven regression discontinuity plots. J Am Stat Assoc 110(512):1753–1769

Card D, Lee D, Pei Z (2009) Quasi-experimental identification and estimation in the regression kink design. Princeton University Industrial Relations Section working paper, no. 553

Card D, Lee D, Pei Z, Weber A (2012) Nonlinear policy rules and the identification and estimation of causal effects in a generalized regression kink design. NBER working paper series, no. 18564

Card D, Lee DS, Pei Z, Weber A (2015) Inference on causal effects in a generalized regression kink design. Econometrica 83(6):2453–2483

Card D, Pei Z, Lee DS, Weber A (2014) Local polynomial order in regression discontinuity designs. Working paper

Chetty R, Friedman JN, Leth-Petersen S, Nielsen T, Olsen t (2014) Active vs. passive decisions and crowd-out in retirement savings accounts: evidence from Denmark Q J Econ 129(3):1141–1219

Chodorow-Reich G, Feiveson L, Liscow Z, Woolston WG (2012) Does state fiscal relief during recessions increase employment? Evidence from the American Recovery and Reinvestment Act. Am Econ J Econ Policy 4(3):118–145

Conley TG, Dupor B (2013) The American Recovery and Reinvestment Act: Solely a government jobs program? J Monet Econ 60(5):535–549

Dafflon B (2007) Fiscal capacity equalization in horizontal fiscal equalization programs. In: Intergovernmental fiscal transfers. The World Bank, pp 361–396

Dahlberg M, Mörk E, Rattsø J, Ågren H (2008) Using a discontinuous grant rule to identify the effect of grants on local taxes and spending. J Public Econ 92(12):2320–2335

Dong Y (2014) Jump or kink? Identification of binary treatment regression discontinuity design without the discontinuity. Working paper

Ek S (2013) Gaining from local benefits? Unemployment insurance and job quality. In: Essays on unemployment insurance design. Doctoral thesis, Uppsala University

Engström P, Nordblom K, Ohlsson H, Persson A (2015) Tax compliance and loss aversion. Am Econ J Econ Policy 7(4):132–164

Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Chapman and Hall, London

Feyrer J, Sacerdote B (2011) Did the stimulus stimulate? Real time estimates of the effects of the American Recovery and Reinvestment Act. NBER working paper series, no. 16759

Florens J-P, Heckman JJ, Meghir C, Vytlacil EJ (2008) Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. Econometrica 76(5):1191–1206

Ganong P, Jäger S (2015) A permutation test for the regression kink design. Working paper

Gelman A, Imbens G (2014) Why high-order polynomials should not be used in regression discontinuity designs. NBER working paper series, no. 20405

Gordon N (2004) Do federal grants boost school spending? Evidence from Title I. J Public Econ 88(9–10):1771–1792

Guryan J (2001) Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts. NBER working paper series, no. 8269

Hausman JA (1985) The econometrics of nonlinear budget sets. Econometrica 53(6):1255–1282

Hayashi M (2000) Distortionary effects of seemingly lump-sum intergovernmental transfers in Japan: a note. Meiji Gakuin Rev Pap Proc Econ 118:63–71

Hayashi M (2006) Chihokohuzei no Keizai Bunseki Genjo to Kadai (The economics studies of the local allocation tax in Japan: a critical survey). J Econ Policy Stud 3(2):6–24 (in Japanese)

Holtz-Eakin D (1986) Unobserved tastes and the determination of municipal services. Natl Tax J 39(4):527–533

Imbens G, Kalyanaraman K (2012) Optimal bandwidth choice for the regression discontinuity estimator. Rev Econ Stud 79:933–959

Knight B (2002) Endogenous federal grants and crowd-out of state government spending: theory and evidence from the federal highway aid program. Am Econ Rev 92(1):71–92

Landais C (2015) Assessing the welfare effects of unemployment benefits using the regression kink design. Am Econ J Econ Policy 7(4):243–278

Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. J Econ Lit 48:281–355

Litschig S, Morrison KM (2013) The impact of intergovernmental transfers on education outcomes and poverty reduction. Am Econ J Appl Econ 5(4):206–240

Lundqvist H (2015) Granting public or private consumption? Effects of grants on local public spending and income taxes. Int Tax Public Financ 22:41–72

Lundqvist H, Dahlberg M, Mörk E (2014) Stimulating local public employment: do general grants work? Am Econ J Econ Policy 6(1):167–192

McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: a density test. J Econom 142(2):698–714

MIC (2003) White paper on local public finance, 2003 illustrated. Ministry of internal affairs and communications (MIC). http://www.soumu.go.jp/menu_seisaku/hakusyo/pdf/chihou/2003e_c_visual.pdf. Accessed 13 Dec 2015

MIC (2012) Zeiritsu ni Tsuite no Kazei Jishuken no Kakudai ni Tsuite (On the expansion of tax autonomy for local tax rates). Ministry of internal affairs and communications (MIC). http://www.soumu.go.jp/main_content/000174389.pdf. Accessed 13 Dec 2015 (in Japanese)

Moffitt R (1984) The effects of grants-in-aid on state and local expenditures: the case of AFDC. J Public Econ 23(3):279–305

Moffitt R (1986) The econometrics of piecewise-linear budget constraints: a survey and exposition of the maximum likelihood method. J Bus Econ Stat 4(3):317–328

Moffitt R (1990) The econometrics of kinked budget constraints. J Econ Perspect 4(2):119–139

Nielsen HS, Sørensen T, Taber C (2010) Estimating the effect of student aid on college enrollment: evidence from a government grant policy reform. Am Econ J Econ Policy 2(2):185–215

Saez E (2010) Do taxpayers bunch at kink points? Am Econ J Econ Policy 2(3):180–212

Simonsen M, Skipper L, Skipper N (2016) Price sensitivity of demand for prescription drugs: exploiting a regression kink design. J Appl Econom 31:320–337

Serrato JCS, Wingender P (2016) Estimating local fiscal multipliers. NBER working paper series, no. 22425

Turner LJ (2014) The road to pell is paved with good intentions: the economic incidence of federal student grant aid. Working paper

Wilson DJ (2012) Fiscal spending jobs multipliers: evidence from the 2009 American Recovery and Reinvestment Act. Am Econ J Econ Policy 4(3):251–282