Public Choice
https://doi.org/10.1007/s11127-019-00696-2

Applying regression discontinuity designs to American political development

Anna Harvey[1]

Received: 11 July 2019 / Accepted: 18 July 2019

Abstract
Students of American political development (APD) have long been interested in questions related to the development of "state capacity" in the United States. The apparent macro-level nature of those questions may appear to discourage the pursuit of micro-level causal inferences. Yet attention to causal inference is not necessarily incompatible with inquiry into macro-level phenomena. This article explores the application of a specific causal inference strategy, namely regression discontinuity design (RDD), to three questions of interest to APD scholars of state capacity. First, the article illustrates the use of a geographic RDD to estimate the causal impacts of a Reconstruction-era federal civil rights statute during the period prior to the development of significant federal state capacity. Second, it explores the possible causes of the late 19th century decline in the use of monetary rewards to motivate civil servants by using a population-based RDD to estimate the causal impacts of financial incentives on law enforcement effort and civilian compliance. Third, it illustrates an opportunity to test claims about the impacts of the growth of the "carceral state" by applying a resource-constraint RDD to estimate the causal impacts of law enforcement effort on a variety of outcomes.

Keywords Labor discrimination · Public policy · Criminal law

JEL Classification J7 · J78 · K14

1 Introduction

Scholars who study American political development (APD) often are distinguished by their interests in historical perspective and macro-level questions. In the Introduction to the recently published *Oxford Handbook of American Political Development*, Suzanne Mettler and Richard Valelly (2016, p. 4) observe that the field's "wide-angle lens" may be interpreted by some to necessitate assigning lesser weight to questions of causal inference than is typical in the social sciences: among those who study APD, "the reliability of the proposed causal inferences is not taken as utterly primary." Mettler and Valelly suggest

✉ Anna Harvey
anna.harvey@nyu.edu

[1] Department of Politics, New York University, New York, USA

that placing lesser weight on questions of causal inference may be appropriate in the APD research community, given the community's commitment to macro-level questions that may not feasibly be addressed by typical causal inference techniques.

Yet attention to causal inference is not necessarily incompatible with a commitment to answering macro-level questions. That claim is here illustrated in the context of questions asked by APD scholars about the development of "state capacity", or the growth of a salaried and merit-based federal bureaucracy capable of competently administering programs of social provision (Skowronek 1982; Evans et al. 1985; Carpenter 2001). State capacity made possible European programs of social provision, APD scholars have reasoned; without it, US federal social provision was doomed to marginality. APD scholars thus have devoted considerable attention to investigating the impacts of the comparative absence of federal state capacity during the 19th century; of the subsequent growth of a professional salaried civil service around the turn of the 20th century and beyond; and of the resulting growth of the administrative state in Americans' daily lives.

This article suggests that it is possible to support that work by giving careful attention to what we can infer about the causal nature of state capacity's effects. It illustrates the application of one well-known causal inference technique, namely regression discontinuity design (RDD), to three questions of interest to APD scholars of state capacity. First, the article illustrates the application of a geographic RDD (GRD) to estimate the causal impacts of the Civil Rights Act of 1875 on African American well-being. One claim in the APD literature has been that the potential for federal policy was constrained severely during the 19th century owing to undeveloped administrative capacity (Skowronek 1982; Carpenter 2001). Yet, using a GRD design, Harvey and West (2019) find significant policy impacts from a Reconstruction-era civil rights statute, calling into question whether large federal bureaucracies were or are a necessary condition for social policy. GRDs also have been used successfully to demonstrate the long-term effects of policies and institutions having a geographic component.

Another claim in the APD literature has been that the decline in the use of financial incentives to enforce the criminal law around the turn of the 20th century can be explained by the inefficacy of those incentives (Parrillo 2013). According to that account, the salarization of criminal justice professionals, part of the more general growth of a salaried administrative bureaucracy during the same period, better facilitated civilian compliance with the criminal law. The article explores the credibility of that claim by applying a population-based RDD to estimate the effects of financial incentives on law enforcement effort and civilian compliance (Harvey and Mungan 2019). Using that design, Harvey and Mungan (2019) find that financial incentives in fact appear to be very effective both in motivating targeted law enforcement effort and in inducing civilian compliance with the law, calling into question whether the decline in the use of those incentives in the late 19th century was in fact caused by their inefficacy. Population-based RDDs also may also be applied productively to examine other questions of interest to APD scholars.

Finally, the article illustrates the potential application of an RDD in the presence of resource constraints to the question of estimating the impacts of policing on compliance outcomes. A recent question of interest among APD scholars has been the consequences of the growth in what some call the "carceral state" (Lerman and Weaver 2016).[1] A prominent claim has been that encounters between civilians and law enforcement professionals

[1] The term "carceral state" is used in this literature to refer to the growth in the share of public resources devoted to law enforcement and criminal justice.

reduce voter turnout and other forms of civic engagement among individuals who are the subjects of such encounters (Weaver and Lerman 2010; Lerman 2014). The article suggests that that claim, currently supported only by observational survey-based evidence, could productively be tested using a strategy that attempts to account for the endogeneity of law enforcement encounters to unobservable civilian characteristics. One such strategy could be the use of an RDD strategy based on planned law enforcement deployments that eventually are limited by resource constraints, such that the threshold distinguishing jurisdictions that receive additional resources from those that do not serves as a discontinuity used to estimate the impacts of incremental police deployments on outcomes. Like geographic and population-based RDDs, resource-constraint RDDs productively could be used to tease out causal inferences about several other questions potentially of interest to APD scholars.

2 Estimating federal policy impacts in the 19th century

APD scholars have pointed to a general lack of federal "state capacity" during the 19th century as a root cause of the absence of redistributive federal social policies during that period (e.g. Skowronek 1982).[2] Carpenter (2001) maintains, for example, that state capacity is "a central variable in explaining policy change" and that the federal government's relative lack of state capacity during the 19th century "minimized the presence and reputation of national authority in local affairs" (Carpenter 2001, p. 64).

One policy area thought to have been affected particularly by the lack of federal state capacity after the Civil War was racial inequality. King and Lieberman (2016) claim, for example, that, "[b]efore the civil rights revolution...there was little bureaucratic capacity within the state to enforce racial equality." King and Lieberman (2016, pp. 249, 251) suggest that "the crux of the American federal state-building dilemma" with respect to policies addressing racial inequality is that "the American state's mechanisms or instruments for enforcement were historically underdeveloped."

But private enforcement of federal policy directives mandated by courts as a result of civil litigation, criminal complaints, or both, including but not limited to civil rights policies, may be as effective, if not more effective, than public enforcement of statutes (Farhang 2010; Lemos and Minzner 2014). Some scholars even have suggested that private enforcement of civil rights policies especially may be *more* effective than public enforcement (Selmi 1998).

The mere availability of compliance mechanisms, whether public or private, also might induce adherence to regulatory statutes even in the absence of formal law enforcement actions, if actors anticipate the possibility of such actions and change their behavior accordingly. Only a relatively minimal bureaucratic presence might be necessary to induce significant policy compliance in the presence of such foresight.

This section explores the application of the principles of causal inference to this debate. If it is correct that federal "state capacity" was a necessary but absent condition for the effective implementation of redistributive federal policies during the 19th century, then we should not expect any such policies enacted during that period to have had significant effects. Drawing on recent work by Harvey and West (2019), the section assesses the

[2] A possible exception is the provision of federal pension benefits to Civil War veterans (Skocpol 1995). Not all historians agree with the broad claim in the text; see, for example, Novak (2008).

evidence for that claim by examining the causal consequences of the Civil Rights Act of 1875.

2.1 The Civil Rights Act of 1875

The Civil Rights Act of 1875, passed by Congress on March 1, 1875, directed that "all persons within the jurisdiction of the United States shall be entitled to the full and equal enjoyment of the accommodations, advantages, facilities, and privileges of inns, public conveyances on land or water, theaters, and other places of public amusement," subject only to any limitations "established by law, and applicable alike to citizens of every race and color, regardless of any previous condition of servitude." The law provided both civil and criminal penalties for violations of its provisions; victims of discrimination could sue for damages of \$500–\$1000, or bring criminal charges with potential sentences ranging from one to 12 months in prison. In 2013 dollars, the damages available for victims of discrimination ranged from \$10,500–21,000, while the average African American family's income at the time was approximately \$5250 per year (Ng and Virts 1989). The law also provided both civil and criminal penalties for district attorneys who failed to prosecute violations; victims of discrimination whose cases were not pursued could sue for \$500 in damages (\$10,500 in 2013 dollars), or bring misdemeanor charges with potential fines ranging from \$1000 to \$5000 (\$21,000 to \$105,000 in 2013 dollars). In *The Civil Rights Cases*, 109 U.S. 3 (1883), the Supreme Court struck down the law's public accommodations provisions, ruling that Congress lacked the power under either the Thirteenth or Fourteenth Amendments to enact them.[3]

Prior to the passage of the Civil Rights Act of 1875, significant restrictions on African American mobility existed in states without state-level public accommodations statutes. At least some Southern employers, for example, preferring to restrict the mobility of African American workers rather than compete on wages, pressured interstate railroad and steamboat operators not to carry African American passengers (Franklin 1974; Painter 1977). African Americans also often were prohibited from using intracity transportation services, such as streetcars and omnibuses, important services for African Americans during this period because of their typical residence in outlying districts far from the economic opportunities available in city centers (Foner 1973a, b; Rabinowitz 1971; Wright 1985).[4] Those barriers to mobility likely raised the relative costs of pursuing economic opportunities that required travel, thereby exerting downward pressure on African Americans' wages.

If the law had been enforced effectively during the roughly 8.5 years of its lifespan, we might expect it to have reduced restrictions on African American mobility and, thus, to have had positive effects on African American wages. The conventional APD story about the period, however, is that enforcement of federal policy initiatives at the time necessarily

[3] The ruling consolidated five separate cases brought under the Civil Rights Act of 1875, namely *United States v. Stanley, United States v. Ryan, United States v. Nichols, United States v. Singleton, and Robinson v. Memphis & Charleston R.R. Co.* Despite its earlier ruling that the Constitution's Commerce Clause requires that prohibitions on within-state racial discrimination by common carriers operating between the states "must come from Congress and not from the States" [*Hall v. De Cuir*, 95 U. S. 485 (1877)], the Supreme Court did not evaluate the constitutionality of the Civil Rights Act of 1875 under the Commerce Clause. The statute itself did not specify a constitutional justification.

[4] Roback (1986) suggests that exclusionary streetcar practices were no longer in widespread effect after 1890.

was ineffective owing to the absence of a large federal bureaucracy staffed by salaried civil servants (Skowronek 1982; Carpenter 2001). That story then predicts that the Civil Rights Act of 1875 would have had no causal effects on African Americans' economic well-being. The question then becomes, how do we estimate those causal impacts, if any?

Historians seeking to assess the law's causal effects have looked for instances of civil litigation and criminal prosecutions pursued under the Civil Rights Act. Noting the relative infrequency of such enforcement actions, the historians have concluded that the law was not implemented effectively, and therefore that it had no causal consequences (Franklin 1974; Gillette 1979; Wright 1985; Foner 1988; Wright 2013).[5]

But, as noted earlier, the mere availability of compliance mechanisms might have induced compliance with the law's provisions even in the absence of formal enforcement actions. In that case, racial discrimination would have declined, African Americans would have become more mobile, and their wages would have increased, even absent observable enforcement. In fact anecdotal reports exist of apparently successful threats to invoke the law in order to compel access to public accommodations. In the late 1870s, for example, many African Americans (known as "Exodusters") sought to emigrate up the Mississippi River to Kansas so as to escape states that had been redeemed by the Democratic Party (Painter 1977). Under pressure from Southern employers, riverboat companies operating on the Mississippi barred passage to African American travelers. Some of those seeking passage threatened to sue under the Civil Rights Act of 1875. United States Colored Troops veteran John Solomon Lewis, for example, who sought to emigrate from Louisiana with his family, told a riverboat captain, "I am a man who was a United States soldier, and I know my rights, and if I and my family gets put off, I will go in the United States Court and sue for damages." Lewis and his family were allowed passage (Painter 1977, p. 3). The US Attorney General later issued a public statement to the effect that refusals to carry African American passengers constituted a violation of the Civil Rights Act of 1875, which "makes such refusal an offense to which considerable penalty attaches, and provides the method by which a prosecution for the penalty or for damages to the party entitled can be pursued" (Franklin 1974, p. 229). In the wake of the Attorney General's statement, operators of riverboat companies began to carry all African Americans seeking passage (Franklin 1974, p. 229; Jack 2007, pp. 50, 91–92).

As such, the proper empirical strategy for estimating the law's causal effects, then, would be not to count instances of observable enforcement actions brought under the Civil Rights Act, but rather to collect behavioral data on wages or other measures of African Americans' economic well-being, before and after the law's passage, its striking down, or both, and estimate the law's impacts on relevant outcomes. Following Fogel (2004) and Floud et al. (2011), Harvey and West (2019) measure African Americans' economic well-being during the period using measures of individual-level weight gains and losses.[6] They source weight data from the medical examinations of United States Colored Troops (USCT) veterans collected as part of the postwar pension application process, a random

[5] Historian John Hope Franklin asserted that the law "was never effectively enforced" (Franklin 1974, p. 235). William Gillette characterized the law as "the most meaningless piece of postwar legislation...the deadest of dead letters", a characterization echoed by historians George Wright and Eric Foner (Gillette 1979, pp. 271, 279; Wright 1985, p. 58, 2013; Foner 1988).

[6] Fogel (2004) and Floud et al. (2011) have found that short-term changes in body mass may proxy short-term changes in economic well-being during periods of relative scarcity.

sample of which were collected in connection with the Early Indicators Project.[7] Because many USCT veterans were examined more than once, for a subset of the Early Indicators sample Harvey and West are able to identify changes in weight at the individual level. Given the greater availability of data for later years, they focus their attention on the years bracketing the law's strike by the Supreme Court in 1883, rather than on those bracketing the law's enactment in 1875.[8]

Even with that individual-level data on African Americans' economic well-being in hand, however, Harvey and West (2019) had to consider several questions before drawing a causal inference about the law's effects. It is possible, for example, that any observed deterioration in African Americans' well-being after the Court struck federal public accommodations protections could have been explained by trends in racial conservatism, rather than to the Court's rejection of those provisions. Harvey and West address that concern by leveraging the ex ante variation in state-level public accommodations statutes. Many states, generally northern ones, enacted public accommodations statutes either before or shortly after the Court ruled in *The Civil Rights Cases* (Johnson 1919). In those states, the decision handed down on the Civil Rights Act of 1875 presumably would have had little impact, because the state-level statutes remained in place. One could thus use changes in the well-being of African Americans living in those "control" states to difference out the effects of any national-level events or trends on changes in the well-being of African Americans in the "treatment" states, namely those lacking state-level public accommodations statutes.[9]

Using control and treatment states in that way is a classic difference-in-differences (DD) design (Dunning 2012). It assumes that, conditional on a set of observed pretreatment covariates, the entire treated and control samples are comparable to each other. Yet that assumption might be violated by unmeasured confounders. In the case of the Civil Rights Act of 1875, racial hostility in the 1880s might have been rising faster in the more southern states than in the more northern states that were more likely to have enacted public accommodations statutes. Differences in the evolution of racially hostile attitudes could have led to declines in African Americans' well-being in the southern states in the 1880s, independently of the Court's 1883 rejection of federal public accommodations protections.

Harvey and West thus deploy an additional research design, one that exploits geographic proximity. If individuals sort themselves around a border between treated and control areas with error, a local treatment effect is identifiable under a geographic regression

[7] Between 1862 and 1890, Union Army and USCT veterans could apply for military pensions by claiming that a current disability, interfering with their capacity to perform manual labor, was related directly or indirectly to their wartime service. One component of the pension application process was a medical examination. Upon receiving a veteran's application, the Pension Bureau would direct the applicant to appear before a board of examining surgeons "at a location near his place of residence" (Logue and Blanck 2010, p. 30). The examining surgeons were given a standardized form to complete instructing them to report any medical evidence relevant to an applicant's claim of disability; the form also required examining surgeons to record basic intake information on all applicants, including weight, height and age ("Instructions to Examining Surgeons, 1870–1926", Department of the Interior, Pension Office, Washington D.C.). Many applicants were examined multiple times, either because they reapplied after their applications initially had beendenied, or because they applied for increases to their current pensions.

[8] Harvey and West acknowledge that the USCT sample is not a representative sample of the African American male population; relative to their counterparts in the 1900 census, USCT veterans were older, more likely to be married, and more likely to own their own homes.

[9] Several of the treatment states, including those along the primary border distinguishing control and treatment states, were not former Confederate states. These states include Maryland, West Virginia, Kentucky, Missouri, and Oklahoma.

discontinuity (GRD) framework (Keele and Titiunik 2015; Keele et al. 2015; Keele and Titiunik 2016; Keele et al. 2017). Under that design, treated and control individuals near the border may be good counterfactuals for one another, because location in the treated or control areas can be thought of as as-if random very near the border (or at least as subject to less confounding variation, relative to locations further from the border).

In the case of the Civil Rights Act of 1875, Harvey and West (2019) treat the primary border distinguishing states with and states without state public accommodations statutes as a geographic discontinuity, with the identifying assumption being that variation in racial attitudes (and associated actions) likely was continuous during the period, while variation in statutes was discontinuous. Within small intervals just north and south of the border separating the control and treatment states, they expect less variation in racial attitudes, relative to areas further away from the border.[10] Any difference in differences in African American well-being at the relevant border should then be the result of statutory variation between the control and treatment states, not attitudinal variation.

States that did not enact public accommodations statutes, however, also might have been less likely to enact statutes providing other public goods, such as schools, hospitals, sanitation and roads, during the period of interest. The variation in the enactment of such statutes might generate confounding variation in changes in African Americans' well-being at the borders distinguishing states with and without state-level public accommodations statutes.

In the GRD literature, the foregoing empirical issue is known as the problem of compound treatments at the border of interest (Keele and Titiunik 2016). Several strategies have been proposed to address that problem. First, if a population exists that should not be affected by the treatment of interest at the relevant border, that population can be assigned to a control group. Presumably, the control group should be affected by all other treatments at the relevant border, other than the treatment of interest. Harvey and West (2019) use white Union Army veterans as a control group, sourcing data on their weight gains and losses over the period of interest from the same Early Indicators sample. Variation in racially neutral public goods provision over the period of interest presumably would have affected not only African American veterans near the border of interest, but also white Union Army veterans. Measuring post-Supreme-Court-ruling changes in African Americans' well-being near the border of interest, relative to post-ruling changes in whites' well-being near that border, allow Harvey and West to address a possible confounder.

Another strategy for addressing compound treatments at a geographic border is to look for local effects during the period just before the treatment of interest is implemented. During that pretreatment period, if all the other treatments at the border except the treatment of interest remain in effect, one would not expect to find the treatment of interest to have the predicted effect. Harvey and West (2019) replicate their RDD and GRD specifications during the period when federal public accommodations provisions were in force (1875–1883), during which state-level variation in public accommodations statutes would have had little impact. This strategy allows them to isolate the impacts of state-level variation in public goods provision, other than public accommodations statutes, on African American well-being.

[10] Some anecdotal evidence supports that assumption. In the postwar period, African Americans reported significantly more progressive racial attitudes in the former slave states bordering former nonslave states than in the deeper South, while those living in cities just across the border (e.g. Philadelphia) reported attitudes less racially progressive than those found in the more northern cities (Foner 1973a, b; Wright 1985).

2.2 Applying RDD and GRD designs to the Civil Rights Act of 1875

To implement their RDD and GRD strategies, Harvey and West (2019) geo-locate US Colored Troops (USCT) and Union Army veterans' medical exams to precise latitude and longitude coordinates, computing the shortest distance from each identified set of coordinates to the border distinguishing states enacting their own public accommodations statutes from those that did not. They identify veterans examined during an eight-year window bracketing the Court's ruling in *The Civil Rights Cases*.

They first assess pre-ruling covariate balance on individual-level measures of pre-ruling weight, height, age, the incidence of several systemic medical conditions, and pre-ruling trends in weight. They also assess covariate balance on county-level measures from the 1880 Census, including the percentage of the population living in towns larger than 2500 persons, per capita manufacturing output, the fraction of a county's acreage dedicated to farming, and the percentage of the population that was African American. If racial attitudes did in fact vary smoothly near the border of interest, then one would expect to find covariate balance near this border (Keele and Titiunik 2015; Keele et al. 2015, 2017; Keele and Titiunik 2016).

Harvey and West in fact do not find, for the USCT or the placebo Union Army samples, any significant cross-border differences in veteran-level pre-ruling characteristics, within either the full sample or within windows of 300 miles or 200 miles from the border of interest. However, they do observe significant cross-border differences on all county-level pre-ruling census covariates in both the USCT and Union Army samples, at generally the same magnitudes and significance levels; their placebo exercises using Union Army veterans allow them to test whether the county-level cross-border differences are driving outcomes. They also see that the cross-border differences in pre-ruling census covariates generally decline as the border is approached, suggesting that focusing on veterans closer to the border may reduce the influence of possible confounders (Keele and Titiunik 2016).

Harvey and West (2019) then estimate both difference-in-differences (DD) and GRD models of weight gain/loss in both the full sample and amongst both Union Army and USCT veterans living near the border of interest. For the DD models with veteran and year fixed effects and pre-ruling covariates, they estimate that USCT veterans located in states without public accommodations statutes lose 2.2 pounds post-ruling, relative to USCT veterans located in states with public accommodations statutes. Point estimates do not diminish appreciably as samples are narrowed around the border of interest, suggesting that the results are not being driven by progressively more racial hostility in the more southern states. They do not find comparable effects in the placebo sample of white Union Army veterans, suggesting that their findings for USCT veterans are not being driven by cross-border differences common to all veterans.

Harvey and West then implement a GRD at the border of interest, specifying the outcome variable as the change in a veteran's post-ruling weight from his pre-ruling weight. The running variable is defined as the average distance from a veteran's exam location to the border distinguishing states with and without state-level public accommodations statutes, and the treatment variable as whether the veteran is or is not located on the side of the border without state-level public accommodations statutes. They include pre-ruling covariates and fixed effects for border segments. The coefficient on the treatment variable is estimated using local linear regression with a triangular kernel within the MSE-optimal bandwidths to the north and south of the border. Standard errors are clustered in 20-mile intervals on either side of the border, with the interval-identifiers increasing in value with distance south of the border (Calonico et al. 2014, 2018a, b). Harvey and West also estimate models that implement local quadratic bias correction of the local linear point estimates, and local quadratic bias correction with robust variance estimation (Calonico et al. 2018b). They restrict their sample to veterans whose medical exams all took place on one side of the border of interest throughout the 8-year window.

The GRD estimates indicate that USCT veterans located just over the border in states without state-level public accommodations statutes experienced an estimated weight loss of 3.7–3.8 pounds in the full sample, and from 4.2 to 4.5 pounds in the subsample for which pre-ruling trends in weight gain/loss are available (including those trends as a covariate). Those point estimates, which are estimated within MSE-optimal bandwidths ranging from 72 to 102 miles north of the border and from 135 to 182 miles south of the border, are significant at levels ranging from 90 to 99%. By contrast, in all models, white Union Army veterans located just south of the border distinguishing states with and without public accommodations statutes gain weight in the four years following the Court's ruling, relative to their Union Army veteran neighbors located just north of the border. The authors find further that their estimates are robust to restricting the samples to veterans reported in the 1860 Census to be living on the same side of the border as they did during the period of interest.

Harvey and West then replicate their RDD and GRD specifications during the period that the Civil Rights Act of 1875 remained in effect, assuming, counterfacutually, that the law's public accommodations provision was struck down in 1879. They find in both models that USCT veterans just over the border in states without state-level public accommodations statutes experienced no differences in weight gain/loss during this pre-ruling period, relative to USCT veterans just on the other side of the border.

Finally, Harvey and West estimate RDD models on a series of measures of the incidence of systemic medical conditions over their primary period of interest, on the theory that changes in longer-term systemic medical conditions should have been uncorrelated with short-term changes in body mass associated with the Court's ruling in *The Civil Rights Cases*. They find that USCT veterans located in states without state-level public accommodation statutes were no more likely to experience a systemic medical condition after the Court's ruling, relative to their USCT neighbors located in states with public accommodation statutes.

Collectively, the findings in Harvey and West (2019) suggest that the Civil Rights Act of 1875 in fact mitigated at least some of the harms suffered by victims of racial discrimination in public accommodations prior to its passage, and that the Supreme Court's strike of the law's public accommodations provisions had the opposite effect. More generally, their findings suggest that federal regulatory policies could have had substantial impacts on outcomes in the nineteenth century, even prior to the development of significant federal "state capacity." To the extent that the "presence and reputation of national authority in local affairs" was minimal during the period studied by Harvey and West (Carpenter 2001 p. 64), the absence of federal state capacity may not have been the root cause.

2.3 Using GRDs to estimate policy impacts

Geographic boundaries have been adopted in a number of other contexts to estimate the impacts of policies that change discontinuously at those boundaries; many of the applications may be of interest to APD scholars. For example, GRDs have been used in studying the effects of school quality on parental choices (Black 1999), the effects of foreclosure laws on loan sizes (Pence 2006), the effects of campaign activities on voting behavior (Middleton and Green 2008; Gerber et al. 2011), the effect of school choice on student performance (Lavy 2010), the effects of centralized bargaining agreements on employment (Magruder 2012), the effect of air pollution on life expectancy (Chen et al. 2013), the effects of national institutions on economic development in Africa (Michalopoulos and Papaioannou 2014), the effects of post-WWII resettlement programs on subsequent population levels (Schumann 2014), the effects of highway location on geographic polarization (Nall 2015), the effects of a fruit-fly eradication program on agricultural outcomes (Salazar et al. 2016), the effects of state tax rates on the migration of high-income earners (Young et al. 2016), the effects of private police forces on crime rates (MacDonald et al. 2016), the effects of alternative military strategies on post-conflict outcomes (Dell and Querubin 2017), and the effects of political advertising on election results (Spenkuch and Toniatti 2018).

Of particular interest to APD scholars may be work on the historically persistent effects of geographic borders. For example, Dell (2010) estimates the long-run impacts of the geographic boundaries of a forced labor system (the *mita*) implemented by the Spanish government in colonial Peru and Bolivia. The *mita* determined whether indigenous communities were required to send adult males to work in silver and mercury mines. The impact of being just within the *mita* boundary, as opposed to just outside of it, persists to the present day: regions just within the *mita* boundary now are less connected to road networks, consume less food, experience more stunted growth, and are more likely to be populated by subsistence farmers, relative to regions just outside the *mita* boundary.

Geographic boundaries also have influenced the long-run development of property values in the United States. In the nineteenth century, all of Ohio was settled under a rectangular system (RS) of land demarcation, except for the Virginia Military District (VMD), which was settled under a metes and bounds (MB) land demarcation regime (Libecap and Lueck 2011). The VMD comprised about 16% of Ohio, running along the Ohio River between the Scioto and Little Miami rivers. Libecap and Lueck (2011) use the VMD boundary as a geographic discontinuity to estimate the long-run impact of the RS and MB land demarcation regimes on property values. They hypothesized that the RS system would have increased property values for relatively flat land by reducing the transaction costs of trading plots in land markets. Where terrain was more rugged, however, land values could have been enhanced under the more flexible MB system to divide properties by value and productivity. They find that the values of relatively flat properties located just outside the historic VMD boundary, settled under the RS land demarcation regime, are approximately 31% higher between 1850 and 1997, relative to similar properties settled just inside that boundary. As land becomes more rugged, however, properties settled just inside the VMD boundary become more valuable, relative to properties settled just outside it; that reversal appears to occur when the land has a 6% slope (Libecap and Lueck 2011).

3 Estimating the impact of financial incentives on enforcement

Another claim that has been made in the APD literature on state building during the 19th century is that the rise of the salarization of public employees around the century's close was driven by the failure of financially incentivized civil servants to induce compliance with the law (Parrillo 2013). Providing financial incentives to law enforcers, for example, once was a common practice in the United States. In many jurisdictions, police officers were rewarded financially for making arrests, and prosecutors were rewarded for indictments or convictions.

Such financial incentives for the enforcement of criminal laws largely had been eliminated by the late nineteenth century. (Parrillo 2013, p. 4) makes the argument that financial rewards for enforcement of the criminal law had "disappointing and perverse results" by undermining civilian compliance with the law. Parillo cites as evidence anecdotal reports gathered from approximately 20 jurisdictions that hired "tax ferrets" to look for assets subject to property taxes, but not visible to local assessors; the ferrets were permitted to keep a share of the revenue they collected. Finding few documented instances of assets detected and reported by the ferrets, Parillo claims that decisionmakers in these jurisdictions "learned from experience that the tax ferrets, in seeking to achieve compliance through coercion and deterrence, in fact yielded at best a modest increase in compliance and perhaps even a reduction." Parillo goes on to make a more general claim about the inefficacy of financial incentives in motivating effective law enforcement effort: "The effective implementation of legislative will depended (and still depends) on a large degree of mass voluntary cooperation by the affected individuals, and bounties turned out to undermine such cooperation. The officer's monetary incentive to impose sanctions on laypersons placed him in such an adversarial position toward them as to vitiate their trust in government and elicit from them a mirror-image adversarial response" (Parrillo 2013, p. 4).

Yet an alternative hypothesis could explain Parillo's finding of few documented instances of assets detected and reported by the tax ferrets, namely that knowledge of the ferrets' presence in fact induced greater voluntary compliance with the relevant jurisdictions' tax laws. With more voluntary compliance, triggered by the greater probability of detection of tax evasion in the jurisdictions employing the tax ferrets, few cases of tax fraud would have been committed for the ferrets to find.[11] In fact, using financial incentives for enforcement of the criminal law has been criticized by others not for being ineffective, but rather for being *too* effective (Landes and Posner 1975).[12]

As noted in the previous section, relying on reports of observable enforcement actions to assess the efficacy of policy enforcement may be misleading, because the mere existence of enhanced enforcement mechanisms may induce greater citizen compliance. The same issue also arises in studies of the effects of financial incentives for law enforcers that rely on nonanecdotal data, for example studies estimating the impacts of variation in states' civil asset forfeiture revenue sharing rules on drug arrests and seizures (Baicker and Jacobson 2007; Kelly and Kole 2016), along with studies estimating the effects of municipality-level fiscal stress on traffic citations (Makowsky and Stratmann 2009, 2011). By using observable enforcement actions rather than civilian compliance as the outcome variable, such studies may miss the deterrent effects of law enforcement effort.

3.1 Financial incentives for traffic safety enforcement in Saskatchewan

Harvey and Mungan (2019) exploit data on the frequency, severity and costs of accidents in Saskatchewan, Canada, to study the effects of financial incentives on traffic

[11] See McCormick and Tollison (1984) for an application to college basketball.

[12] Indeed, Helland and Tabarrok (2004) examine the effectiveness of bounty hunters in criminal case proceedings and note that individuals who are similar on all observable margins except that they are released on bond surety are 28% less likely to fail to appear, relative to individuals released from custody on their own recognizance.

safety enforcement. In Saskatchewan, the provincial government contracts with the Royal Canadian Mounted Police (RCMP) to enforce traffic laws. Under that contract, the RCMP's provincial commanding officer "acts under the direction of the [Provincial or Territorial] Minister" (2014 RCMP Provincial and Territorial Companion Document).

RCMP provincial contracts are signed for periods of 20 years in duration. Under the Saskatchewan RCMP contract in effect from April 1, 1992 through March 31, 2012, the RCMP was delegated responsibility to police jurisdictions with populations of less than 1500 in the 1991 census. Towns with populations of more than 500 but less than 1500 in the 1991 census initially were required by the province either to establish and fund their own municipal police services, or to enter into agreements with the province under which the province would delegate to the small towns both financial and administrative responsibilities for its share of the province's RCMP costs (The Police Act 1990, effective January 1, 1992). Under those agreements, the RCMP would be directed in its policing by local authorities and towns would keep the revenue from citations issued within their boundaries.

That arrangement was modified in 1997 (The Police Act Amendments 1997, effective January 1, 1999). As of January 1, 1999, towns of at least 500 but less than 1500 in 1991 population would be policed under the province's RCMP contract in the same manner as towns with populations below 500 (The Police Act Amendments 1997). However, apparently to assuage opposition to that change, the province allowed towns with populations of at least 500 in the 1996 census to retain 75% of the revenue from citations issued within their boundaries, even though the citations would be issued by RCMP officers reporting to the province, rather than to the towns themselves (The Police Regulations 1998). That allocation of citation revenues persisted throughout the remainder of the duration of the RCMP contract and into the contract currently in effect.

The 1997 Amendments to The Police Act thus created two distinct revenue opportunities for the province under its RCMP contract. After January 1, 1999, the province received 100% of the revenue from citations written by RCMP officers within the boundaries of 39 towns with 1996 populations of less than 500, but only 25% of the revenue from citations written within the boundaries of 80 towns with 1996 populations of 500 or more. RCMP detachments were responsible for policing both categories of towns. Harvey and Mungan (2019) ask whether the post-1998 revenue allocation rule created incentives for provincial officials to direct RCMP officers to devote greater enforcement effort to the towns wherein the province received a larger share of the fine revenue, and whether the possibly greater enforcement effort in those towns induced more driver compliance with traffic safety laws. If the Parrillo (2013) story is correct, then we would not expect the province's financial incentives to have motivated greater driver compliance in the high-revenue-sharing towns. In fact, according to that story, the province's financial incentives would have put an RCMP officer patrolling a high-revenue town "in such an adversarial position toward [drivers] as to vitiate their trust in government and elicit from them a mirror-image adversarial response" (Parrillo 2013, p. 4).

Instead of pursuing the cross-sectional variation across the two categories of towns, which would require assuming comparability on both observables and unobservables for all towns over time, Harvey and Mungan (2019) instead leverage the fact that the post-1998 revenue allocation policy depended discontinuously on a population threshold, and pursue a population-based RDD to estimate the impact of the policy change on driver behavior.

3.2 Applying a population-based RDD to traffic law enforcement in Saskatchewan

The post-1998 revenue allocation policy incentivized provincial officials to direct more RCMP traffic safety enforcement effort to towns wherein the province received 100% of citation revenues, relative to towns wherein the province received only 25% of such revenue. If drivers responded to the increase in the probability of being ticketed by driving more safely, we would expect to see fewer and less severe accidents in towns just below the 1996 population threshold of 500, relative to towns just above that threshold. However, we might not see any effects of the population threshold on citation rates, depending on the elasticity of drivers' responses to observable changes in law enforcement effort: if drivers drove sufficiently more safely in the presence of more active traffic safety enforcement, citation rates might not have risen (and might even have fallen).

Eggers et al. (2018) identify two potential threats to causal inference that may arise when population thresholds are used in RDD designs. The first is that of compound treatments at the threshold, or multiple policies using the same population cutoff. The second is that of manipulative sorting at the cutoff by jurisdictions with incentives to report population totals above or below the cutoff.

A virtue of the Saskatchewan institutional context is that compound treatments do not appear to be an issue at the 1996 population threshold of 500. The threshold originated in a Canada-wide RCMP requirement that municipalities of populations of less than 500 must be policed under a provincial contract, while municipalities at or above the threshold could sign municipal contracts with the RCMP. The province adopted this RCMP rule in its 1990 Police Act. It rescinded that rule in the 1997 Amendments to the Police Act, keeping the 1996 population threshold of 500 only for the purposes of allocating citation revenue. The province does not use this population threshold for other policies. Moreover, the populations of Saskatchewan administrative units generally span the 1996 population threshold. Towns, the administrative unit that is the focus of the design in Harvey and Mungan (2019), range between 89 and 4679 persons. In order explicitly to address the possibility of compound treatments, however, Harvey and Mungan (2019) replicate their discontinuity models for the period prior to the introduction of the revenue allocation rule on January 1, 1999.

In theory, one might be concerned that towns would have had incentives to manipulate the 1996 population data so as to be above the cutoff of 500, that provincial officials would have had incentives to doctor the census data so as to produce more towns below the cutoff, or both (McCrary 2008). Harvey and Mungan (2019) test for sorting at the threshold using the density manipulation test in Cattaneo et al. (2018); they find no discontinuous jump in the density of 1996 town populations near the cutoff of 500.

Because sorting still could have occurred in both directions and be unobservable in the density manipulation test, Harvey and Mungan (2019) also explore the behavior of town-level pretreatment covariates near the population cutoff. In the absence of sorting, compound treatments, or both, towns just below and just above the 500-population cutoff should have been similar to one another on observed pretreatment covariates. They test for continuity in a variety of town-level covariates sourced from the 1996 Canadian Census, including median income, percentage married, percentage nonwhite ("visible minorities"), and fractions of the population without a high school degree and of men who drive themselves to work. They also checked for smoothness near the cutoff in average populations over their period of interest, namely January 1, 1999, through December 31, 2016.

In these RDD models, Harvey and Mungan (2019) specify the treatment variable as whether the province receives only 25% of a town's traffic-citation revenue, and the running variable as the distance of each town's 1996 population from the population cutoff of 500. They estimate the coefficient on the treatment variable using local linear regression with a triangular kernel, within the MSE-optimal bandwidths to the left and right of the population cutoff. They estimate errors using heteroscedasticity-robust nearest neighbor variance estimation with a minimum of three neighbors, reporting models that implement local quadratic bias correction of the local linear point estimates, and local quadratic bias correction with robust variance estimation (Calonico et al. 2014, 2018a, b). Examining the set of 119 towns subject to the citation revenue rules enacted in the 1997 Amendments to The Police Act, Harvey and Mungan (2019) find no discontinuities at the population cutoff in either 1996 pretreatment covariates or average population over their period of interest.

Another possible concern in the design is the possibility of displacement or substitution effects (Marceau 1997). Drivers may respond to greater enforcement effort by driving less in the towns with more enforcement effort, driving more in the towns with less enforcement effort, or opting to use mass transit more frequently in the towns with more enforcement effort. The consequences of such behavior might be a lower frequency of accidents in the latter towns simply because of a smaller number of drivers on the roads, not because the drivers were deterred from speeding.

Harvey and Mungan (2019) suggest, however, that Saskatchewan's geography renders such spillover effects highly implausible. First, essentially no public transportation is available in towns of less than 500 in population (or, really, any provincial towns). Second, each of the towns wherein the province receives 100% of ticket revenue is surrounded by considerable open space. The province also receives 100% of the revenue in that open space, so RCMP patrol effort should be more or less constant in and near these towns, undermining the possibility of displacement effects. Finally, the towns below the population cutoff generally are not located near the towns above the population cutoff. The average distance between a town below the cutoff and all towns above the cutoff ranges from 201 to 383 miles. The shortest distance between a town below the cutoff and all towns above the cutoff ranges from 9 to 64 miles. Those distances again undermine the plausibility of displacement effects.[13]

Harvey and Mungan (2019) then estimate their RDD model on annual and fully collapsed citation and accident data between January 1, 1999, through December 16, 2016. They consistently find no evidence of discontinuities in traffic citation rates between towns just below and just above the 1996 population threshold of 500. However, accident rates, accident-involved vehicle rates, accident cost rates, and accident-involved injury rates all display discontinuous jumps at the threshold, above which the province receives only 25% of citation revenue. The magnitudes of the estimates are striking. For example, the estimated effect of moving just above the population cutoff is to increase accidents by about 76%, to increase the number of accident-involved vehicles by about 83%, to increase accident costs by about 117%, and to increase accident-related injuries by approximately 250%.

Harvey and Mungan (2019) also estimate their RDD model at each of a series of placebo cutoffs on either side of the actual population cutoff of 500; each placebo

[13] Harvey and Mungan also address external validity concerns arising from the rural and sparsely populated nature of their empirical setting, noting that of the 35,789 municipal and town governments in the United States in 2012, fully 49.5% had populations of less than 1000 persons; 69.6% had populations of less than 2499 persons (U.S. Census Bureau, 2012 Census of Governments).

cutoff creates a new grouping of "treated" and "control" jurisdictions within the 119 towns in their sample. They find significant increases in accident outcomes only at the actual cutoff of interest, not at any of the placebo cutoffs, for each of their outcomes of interest. They also estimate their RDD model for the pretreatment period of April 1, 1995, to December 31, 1998, finding no discontinuities in either accidents or citations during that pretreatment period. Estimating annual RDD models between 1995 and 2000, they show that the effect of the population threshold on accident outcomes appears only after the introduction of the new citation revenue regime on January 1, 1999. They also estimate the effect of the population threshold on traffic stops and accidents flagged in their data as having occurred "near" the towns of interest, or stops and accidents that occurred sufficiently close to a town that an officer could see the town's grain elevator from the traffic stop or accident location. The province receives 100% of the revenue from citations issued in all areas "near" towns, irrespective of whether a town collects citation revenue from ticket issued within the town's boundaries. They therefore expected to see no discontinuities in accidents at the population threshold in these nearby areas and, in fact, found no such discontinuities at the population threshold in their RDD design.

Harvey and Mungan (2019) also deployed local randomization inference at the population threshold as an alternative estimation strategy (Cattaneo et al. 2015). That strategy requires identifying the largest window around the population cutoff within which all pretreatment covariates demonstrate covariate balance in differences in means tests with p values of at least 0.15, simulating test statistic distributions by assuming repeated random assignment of observations within this window to treatment and control samples, and then conducting differences in means tests on outcome variables within that window. All five pretreatment covariates demonstrated balance with p values of at least 0.15 within a window of 34 persons in 1996 population on either side of the cutoff of 500; the window contains 12 towns to the left of the cutoff and 10 towns to the right of the cutoff. Within that window, they obtained p values below 0.01 for differences in means tests on all accident variables.

Finally, Harvey and Mungan (2019) found that cited drivers just below the threshold identifying the high-revenue towns were given approximately 14 fewer days (or 22% fewer days) to pay their fines over their period of study, relative to drivers just over the threshold in the low-revenue towns. They also found that drivers in high-revenue towns just below the population threshold experienced late fines as a percentage of original fines that were approximately 3% points higher (or 100% higher), relative to the late fines as a percentage of original fines incurred by drivers in low-revenue towns just above the population threshold. Those findings are indicative of the negative economic consequences of the use of law enforcement resources to extract revenue.

In short, the findings reported by Harvey and Mungan (2019) indicate that the financial incentives enacted in the 1997 Amendments to Saskatchewan's Police Act appear to have incentivized provincial leaders to direct the RCMP's law enforcement effort to towns wherein the province received a larger share of citation revenue, causing drivers to drive more safely and to be involved in fewer, less serious and costly accidents. Their findings suggest that financial incentives to enforce the criminal law can in fact effectively motivate both enforcement effort and civilian compliance. If financial incentives to motivate law enforcement effort declined in the late nineteenth century, it may not have been because that effort did not intensify compliance with the law.

3.3 Using population-based RDDs to estimate policy effects

Population-based RDDs have been used in a variety of other contexts that may be of interest to scholars of American political development. The applications include estimating the effects of the size of a municipal council on the extent of municipal spending (Egger and Koethenbuerger 2010; Pettersson-Lidbom 2012); the effects of proportional representation systems on the number of effective parties (Fujiwara 2011), the performance of clientistic parties (Pellicer and Wegner 2013), and voter turnout (Eggers et al. 2018); the effects of Spanish-language ballots on turnout and voter choices (Hopkins 2011); the effects of a runoff election system on turnout (Barone and de Blasio 2013); the effect of governmental transfers on local corruption (Brollo et al. 2013) and local educational and poverty outcomes (Litschig and Morrison 2013); the effects of mayoral wages on incumbent performance (Gagliarducci and Nannicini 2013) and turnout (De Benedetto and De Paola 2015); the effects of gender quotas on parties' electoral performances (Casas-Arce and Saiz 2015); and the effects of signature requirements on the frequency of ballot initiatives (Arnold and Freier 2015), municipal expenditures (Asatryan et al. 2017a), and municipal tax rates (Asatryan et al. 2017b).

4 Estimating the growth of the "carceral state"

A growing area of interest among APD scholars has been the expansion of the "carceral state", or the expansion of state capacity in law enforcement and criminal justice. Weaver and Lerman (2010) and Lerman (2014); Lerman and Weaver (2016) suggest that the expansion of policing has reduced voting and other forms of civic engagement among those subjected to policing encounters. The evidence for that claim is drawn from survey data. Weaver and Lerman (2010) and Lerman (2014), for example, report negative cross-sectional associations between self-reported encounters with law enforcement personnel and various forms of self-reported civic engagement; using panel survey data, a negative association between self-reported encounters with law enforcement and subsequent acts of self-reported civic engagement is found.

Yet this work does not fully address nonrandom selection into police contact. Police officers choose whom to stop, using selection criteria that are unobservable to researchers. It is entirely possible that the selection criteria include correlates of low civic engagement. For example, if officers are more likely to make pedestrian stops in lower income neighborhoods than in higher income neighborhoods, the individuals they stop already will be less likely to vote than those who are not stopped. Weaver and Lerman (2010) attempt to address such omitted variable bias by conditioning on several observable individual-level demographic characteristics that may be correlated both with the likelihood that an individual is stopped by the police and with the likelihood that the stopped person engages in acts of civic participation. The conditioning characteristics include standard predictors of voter turnout, such as race, income, age, education and gender. However, that strategy requires the strong (and perhaps untenable) assumption that no unobserved behavioral or attitudinal correlates determine both selection into a police stop and the probability of participating in civic affairs. The panel approach similarly is potentially confounded by unobserved trends affecting both the probability of a police stop and the probability of voting, such as "falling in with a bad crowd" (Gerber et al. 2017).

Gerber et al. (2017) leverage the as-if random assignment of judges to cases to obtain credible causal estimates of the impacts of discretionary prison sentences on voting, finding few to no causal effects. However, that causal inference strategy is not available for questions involving the impacts of police contact on civic engagement.

A variety of quasi-experimental approaches have been used to identify the causal effects of policing on outcomes other than civic engagement. For example, MacDonald et al. (2016) leverage a geographic RD to estimate the impacts of incremental university police officers on crime rates at the border of the area patrolled by university police. As discussed above in some detail, Harvey and Mungan (2019) leverage the discontinuity induced by a revenue allocation rule to estimate the impact of financially motivated law enforcement effort on driver behavior. Several other quasi-experimental strategies estimate treatment effects of policing by leveraging anomalous events creating shocks to police deployments, such as election cycles (Levitt 1997; McCrary 2002; Levitt 2002) and terrorist attacks (Di Tella and Schargrodsky 2004; Klick and Tabarrok 2005; Draca et al. 2011). Scholars seeking to estimate the impacts of the "carceral state" productively could leverage any of those empirical designs.[14]

While such quasi-experimental strategies credibly have isolated the causal impacts of policing on a variety of outcomes, however, their focus on relatively anomalous events limits their applicability to understanding the causal impacts of everyday policing. To address that shortcoming in existing designs, this section suggests an RDD strategy based on police officer deployments that are limited by resource constraint thresholds, a scenario that may be common across policing agencies.

4.1 Using a resource-constraint RDD to estimate policy effects

Agencies often do not have resources sufficient to carry out all of the duties assigned to them. Often such responsibilities have geographic components: a set of locations is identified as deserving of incremental resource allocations, but the agency is budget-constrained. A common practice adopted in such scenarios is to rank-order locations by degree of need, determine the number of locations for which the agency has sufficient resources to make incremental allocations, and then deploy resources to the locations whose need rankings are above this resource-constraint threshold. Where agencies follow such a procedure to allocate scarce resources, the resource constraint threshold can create the opportunity for an RDD design.

Greenstone and Gallagher (2008) initiated the application of a resource-constraint RDD to estimate the impact of Superfund sponsored cleanups of hazardous waste sites on housing prices. The 1980 Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), or the Superfund Act, directed the Environmental Protection Agency (EPA) to identify "at least" 400 sites posing imminent and substantial dangers to public welfare and the environment, to place those sites on a National Priorities List (NPL), and to initiate remedial cleanups at those sites (CERCLA Section 105(8)(B)). Approximately 15,000 candidate sites initially were referred to the EPA for consideration as possible Superfund sites. After winnowing that list to the 690 most hazardous sites, the EPA

[14] Of particular interest for the question of the effects of policing on civic engagement, Cohen et al. (2019) use a novel design leveraging the randomness of fatal versus nonfatal police shootings and find no average causal effect of fatal police shootings on patterns of emergency (911) or nonemergency (311) call behavior.

developed a Hazardous Ranking System (HRS), assigning to each site a risk score from 0 to 100, with 100 representing the highest risk. Budgetary constraints then induced the EPA to set a goal of placing only exactly 400 sites on the NPL, the minimum number of sites required by CERCLA. The 400 sites with the highest HRS scores (scores exceeding 28.5) were placed on the initial NPL in 1983, making them eligible for Superfund remedial cleanups. Greenstone and Gallagher (2008) use the HRS threshold of 28.5, distinguishing those sites that just made it on the 1983 NPL from sites that just failed to be included on the list, in a resource constraint RDD estimating the impact of Superfund site identification on housing prices in census tracts adjacent to the sites. The authors find few effects, suggesting that the designation of polluted areas as Superfund cleanup sites did not increase local property values. Extending that design, however, Gamper-Rabindran and Timmins (2011) examine more finely grained census data from areas lying closer to the same set of proposed Superfund sites, finding that, between 1990 and 2000, site cleanups increased population density, housing unit density, mean household income, shares of college-educated individuals, and the shares of minorities.

One could imagine deploying a resource-constraint RDD to examine the impact of additional law enforcement effort on a variety of outcomes. In the mid-2000s, for example, Saskatchewan found itself in the undesirable position of experiencing upwardly trending traffic fatality rates, while other provinces were experiencing reductions in traffic fatalities. Between 2007 and 2012, Saskatchewan experienced, on average, 142 annual traffic-related deaths and 3200 annual traffic-related injuries, giving it the highest fatality and injury rates per capita in Canada. In response to traffic safety concerns, in 2013 the province approved a Combined Traffic Services Saskatchewan (CTSS) program, designed to reduce traffic-related fatalities and injuries by increasing the number of highway patrol officers in areas identified as having particularly pressing traffic safety needs. The program would add 120 highway patrol officers to the existing complement of 48 officers. Thirty of those 120 new positions would be created from currently employed RCMP officers redeployed to highway safety; the remaining 90 positions would be incremental positions funded by Saskatchewan Government Insurance (SGI). A working group was created to identify the areas of need and to make allocational decisions (Board of Decision, Traffic Enforcement Program).

The working group produced a detailed plan to allocate the 120 additional positions across Saskatchewan's 113 RCMP detachments, ranking detachments by degree of need. The 38 neediest RCMP detachments were designated for the new positions. Prior to implementation, however, SGI announced that it would fund only 30 incremental positions, reducing the CTSS program to 60 new positions (30 consisting of redeployed RCMP officers and 30 consisting of incremental officers). Those 60 positions were filled by making the planned allocations to the 13 detachments ranked highest on traffic safety need (Traffic Expansion Implementation Update, May 15, 2015).

The initial ranking of detachments by degree of traffic safety need, and the subsequently announced budgetary constraint for the CTSS program, created the potential for estimating a resource-constraint RDD. A discontinuity exists at the threshold above which a set of ranked locations were treated with additional law enforcement resources, and below which another set of ranked locations were not. One could imagine exploiting this discontinuity, looking at a variety of outcomes. For example, one could use the accident and citation data used in Harvey and Mungan (2019) to look at the impact of additional officers on traffic safety outcomes.

The accident and citation data used in Harvey and Mungan (2019), however, are geographically identified to the level of a municipality (at or near). RCMP detachments in Saskatchewan police multiple municipalities. The treatment in the CTSS program (additional positions dedicated to traffic safety) was administered at the RCMP detachment level. All of the municipalities within a detachment (at or near) were treated or not treated with additional CTSS officers. The policy scenario creates complications for an RDD design, as the running variable (detachment rank) is discrete rather than continuous, with masses of municipalities at each level of the running variable.

To address that issue, one could deploy two strategies. First, a researcher could use local randomization inference at the treatment threshold, finding the window of detachments on either side of the threshold within which municipality-level pre-treatment covariates are balanced, and then estimating differences in post-treatment means for a variety of accident-related outcomes (Cattaneo et al. 2015). As a placebo test, the design also could replicate this randomization inference strategy using pre-treatment means in accident-related outcomes.

One could deploy a further strategy combining local randomization inference and matching on municipality-level pre-treatment covariates (Keele et al. 2015, 2017). Using local randomization inference to identify the window within which pre-treatment census covariates are balanced on average across treated and control detachments, one could match municipalities within the treated and control detachments on pre-treatment covariates, including both census and geographic covariates. Matching municipalities within the randomization inference bandwidth would provide an even more precise comparison of the effects of additional law enforcement resources on post-treatment accident-related outcomes. As a placebo test, one also could replicate the blended randomization inference and matching strategy using pre-treatment accident-related outcomes.

A resource constraint RDD design potentially could be applied in other contexts of interest to scholars in the APD subfield. For example, the United States Department of Agriculture's Natural Resources Conservation Service administers the federal Agricultural Conservation Easement Program (ACEP), through which federal funds can be leveraged for conservation purposes. Under the ACEP, state conservation offices are directed to "[d]evelop a weighted ranking process to prioritize all eligible applications, conduct ranking on eligible applications, and recommend prioritized eligible applications for funding" (Title 440, Conservation Programs Manual). Each state's ranking procedure is to be implemented by creating a ranking worksheet assigning to each application a score of between 0 and 400. State conservation offices are further told that, once applications have been ranked, "parcels [are to be] funded in order of ranking priority unless inadequate funds are available to fund the next highest ranked parcel. If adequate funds are not available, the State may select the next-highest-ranked parcel for which funding is available" (United States Department of Agriculture, Title 440). The score within each state, above which ACEP applications are funded, and below which they are not, could establish the threshold in a resource constraint RDD. Areas adjacent to lands that are just above the threshold over which ACEPs are granted could be compared to areas adjacent to lands just below that threshold on a variety of outcomes.

5 Conclusion

The field of American political development sustains a research community committed to a focus on the evolution of political institutions over time and the consequences of those institutions for policies and outcomes. Many social scientists who may not identify as members of the APD research community nonetheless share an interest in such research questions.

Some scholars may believe that the primacy given by most contemporary social scientists to the reliability of causal inferences is inappropriate for the APD research community, given its members' interest in macro-level questions about institutional development. Yet making micro-level causal inferences with care is not necessarily incompatible with inquiry into macro-level questions.

This article has illustrated the potential value of causal inference strategies for work in American political development with respect to three questions of interest to the APD research community: the extent to which "state capacity" was a limiting factor in federal social policy development during the 19th century; the extent to which financial incentives for law enforcement officers induce noncompliance; and the extent to which police contact fosters civic disengagement. In each case, the article also has illustrated how regression discontinuity designs can productively be used to make reliable causal inferences that advance our knowledge about those questions. In short, causal inference strategies are perhaps not only compatible with work in the field of American political development, but they also may serve to advance our understanding of the causal processes underlying institutional development.

References

Arnold, F., & Freier, R. (2015). Signature requirements and citizen initiatives: Quasi-experimental evidence from Germany. *Public Choice*, *162*(1), 43–56.

Asatryan, Z., Baskaran, T., Grigoriadis, T., & Heinemann, F. (2017a). Direct democracy and local public finances under cooperative federalism. *The Scandinavian Journal of Economics*, *119*(3), 801–820.

Asatryan, Z., Baskaran, T., & Heinemann, F. (2017b). The effect of direct democracy on the level and structure of local taxes. *Regional Science and Urban Economics*, *65*, 38–55.

Baicker, K., & Jacobson, M. (2007). Finders, keepers: Forfeiture laws, policing incentives, and local budgets. *Journal of Public Economics*, *91*, 2113–2136.

Barone, G., & de Blasio, G. (2013). Electoral rules and voter turnout. *International Review of Law and Economics*, *36*, 25–35.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics*, *114*(2), 577–599.

Brollo, F., Nannicini, T., Perotti, R., & Tabellini, G. (2013). The political resource curse. *American Economic Review*, *103*(5), 1759–96.

Calonico, S., Cattaneo, M., Farrell, M., & Titiunik, R. (2018a). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, Forthcoming.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2018b). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, *113*, 1–13.

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

Carpenter, D. P. (2001). *The forging of bureaucratic autonomy: Reputations, networks, and policy innovation in executive agencies, 1862–1928*. Princeton: Princeton University.

Casas-Arce, P., & Saiz, A. (2015). Women and power: Unpopular, unwilling, or held back? *Journal of Political Economy*, *123*(3), 641–669.

Cattaneo, M. D., Frandsen, B. R., & Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate. *Journal of Causal Inference*, *3*(1), 1–24.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Simple local polynomial density estimators. Working Paper.

Chen, Y., Ebenstein, A., Greenstone, M., & Li, H. (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences*, *110*(32), 12936–12941.

Cohen, E., Gunderson, A., Jackson, K., Zachary, P., Clark, T. S., Glynn, A. N., et al. (2019). Do officer-involved shootings reduce citizen contact with government? *The Journal of Politics*, *81*(3), 1111–1123.

De Benedetto, M. A., & De Paola, M. (2015). Candidates' education and turnout: Evidence from Italian municipal elections. *German Economic Review*, *18*, 22–50.

Dell, M. (2010). The persistent effects of peru's mining 'mita'. *Econometrica*, *78*(6), 1863–1903.

Dell, M., & Querubin, P. (2017). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics*, *133*(2), 701–764.

Di Tella, R., & Schargrodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, *94*(1), 115–133.

Draca, M., Machin, S., & Witt, R. (2011). Panic on the streets of London: Police, crime, and the July 2005 terror attacks. *American Economic Review*, *101*(5), 2157–81.

Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach. Strategies for social inquiry*. Cambridge: Cambridge University Press.

Egger, P., & Koethenbuerger, M. (2010). Government spending and legislative organization: Quasi-experimental evidence from germany. *American Economic Journal: Applied Economics*, *2*(4), 200–212.

Eggers, A. C., Freier, R., Grembi, V., & Nannicini, T. (2018). Regression discontinuity designs based on population thresholds: Pitfalls and solutions. *American Journal of Political Science*, *62*(1), 210–229.

Evans, P. B., Rueschemeyer, D., & Skocpol, T. (Eds.). (1985). *Bringing the state back in*. Cambridge: Cambridge University Press.

Farhang, S. (2010). *The Litigation State: Public regulation and private lawsuits in the US*. Princeton: Princeton University Press.

Floud, R., Fogel, R. W., Harris, B., & Hong, S. C. (2011). *The changing body: Health, nutrition and human development in the Western World Since 1700*. New York: Cambridge University Press.

Fogel, R. W. (2004). *The escape from hunger and premature death, 1700–2100: Europe. America and the Third World*. New York: Cambridge University Press.

Foner, E. (1988). *Reconstruction: America's Unfinished Revolution 1863–1877*. New York: Harper and Row.

Foner, P. S. (1973a). The battle to end discrimination against negroes on philadelphia streetcars (part 1): Background and beginning of the battle. *Pennsylvania History*, *40*(3), 260–290.

Foner, P. S. (1973b). The battle to end discrimination against negroes on philadelphia streetcars (part ii): The victory. *Pennsylvania History*, *40*(4), 354–379.

Franklin, J. H. (1974). The enforcement of the civil rights act of 1875. *Prologue Magazine*, *6*(4), 225–235.

Fujiwara, T. (2011). A regression discontinuity test of strategic voting and Duverger's law. *Quarterly Journal of Political Science*, *6*, 197–233.

Gagliarducci, S., & Nannicini, T. (2013). Do better paid politicians perform better? Disentangling incentives from selection. *Journal of the European Economic Association*, *11*(2), 369–398.

Gamper-Rabindran, S., & Timmins, C. (2011). Hazardous waste cleanup, neighborhood gentrification, and environmental justice: Evidence from restricted access census block data. *American Economic Review*, *101*(3), 620–24.

Gerber, A. S., Huber, G. A., Meredith, M., Biggers, D. R., & Hendry, D. J. (2017). Does incarceration reduce voting? Evidence about the political consequences of spending time in prison. *The Journal of Politics*, *79*(4), 1130–1146.

Gerber, A. S., Kessler, D. P., & Meredith, M. (2011). The persuasive effects of direct mail: A regression discontinuity based approach. *The Journal of Politics*, *73*(1), 140–155.

Gillette, W. (1979). *Retreat from reconstruction, 1869–1879*. Baton Rouge: Louisiana State University Press.

Greenstone, M., & Gallagher, J. (2008). Does hazardous waste matter? Evidence from the housing market and the superfund program*. *The Quarterly Journal of Economics*, *123*(3), 951–1003.

Harvey, A. & Mungan, M. (2019). Policing for profit: The political economy of law enforcement. *Working Paper*.

Harvey, A. & West, E. A. (2019). Discrimination in public accommodations. *Working Paper*.

Helland, E., & Tabarrok, A. (2004). The fugitive: Evidence on public versus private law enforcement from bail jumping. *Journal of Law and Economics*, *47*, 93–122.

Hopkins, D. J. (2011). Translating into votes: The electoral impacts of Spanish-Language Ballots. *American Journal of Political Science*, *55*(4), 814–830.

Jack, B. M. (2007). *The St. Louis African American community and the exodusters*. Columbia, MO: University of Missouri Press.

Johnson, F. (1919). *The development of State Legislation concerning the free Negro*. New York: Columbia University Press.

Keele, L., Lorch, S., Passarella, M., Small, D., & Titiunik, R. (2017). An overview of geographically discontinuous treatmentassignments with an application to children's health insurance. In M. D.

Cattaneo & J. C. Escanciano (Eds.), *Regression discontinuity designs: Theory and applications; advances ineconometrics* (Vol. 38). Bingley: Emerald Publishing Limited.
Keele, L., & Titiunik, R. (2016). Natural experiments based on geography. *Political Science Research and Methods*, *4*(1), 65–95.
Keele, L., Titiunik, R., & Zubizaretta, J. R. (2015). Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society, Statistics in Society, Series A*, *178*(1), 223–239.
Keele, L. J., & Titiunik, R. (2015). Geographic boundaries as regression discontinuities. *Political Analysis*, *23*, 127–155.
Kelly, B. D., & Kole, M. (2016). The effects of asset forfeiture on policing: A panel approach. *Economic Inquiry*, *54*(1), 558–575.
King, D., & Lieberman, R. (2016). The American State. In S. Mettler, R. Valelly, & R. Lieberman (Eds.), *The Oxford handbook of American political development*. Oxford: Oxford University Press.
Klick, J., & Tabarrok, A. (2005). Using terror alert levels to estimate the effect of police on crime. *The Journal of Law and Economics*, *48*(1), 267–279.
Landes, W. M., & Posner, R. A. (1975). The private enforcement of law. *The Journal of Legal Studies*, *4*(1), 1–46.
Lavy, V. (2010). Effects of free choice among public schools. *The Review of Economic Studies*, *77*(3), 1164–1191.
Lemos, M. H., & Minzner, M. (2014). For-profit public enforcement. *Harvard Law Review*, *127*, 854–913.
Lerman, A. E., & Weaver, V. M. (2014). *Arresting citizenship: The democratic consequences of American crime control*. Chicago: Chicago University Press.
Lerman, A. E., & Weaver, V. M. (2016). The Carceral State and American political development. In R. Valelly, S. Mettler, & R. Lieberman (Eds.), *The Oxford handbook of American political development*. Oxford: Oxford University Press.
Levitt, S. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review*, *87*(3), 270–290.
Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. *American Economic Review*, *92*(4), 1244–1250.
Libecap, G. D., & Lueck, D. (2011). The demarcation of land and the role of coordinating property institutions. *Journal of Political Economy*, *119*(3), 426–467.
Litschig, S., & Morrison, K. M. (2013). The impact of intergovernmental transfers on education outcomes and poverty reduction. *American Economic Journal: Applied Economics*, *5*(4), 206–40.
Logue, L. M., & Blanck, P. D. (2010). *Race, ethnicity and disability: Veterans and benefits in post-civil war America*. New York: Cambridge University Press.
MacDonald, J. M., Klick, J., & Grunwald, B. (2016). The effect of private police on crime: Evidence from a geographic regression discontinuity design. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*(3), 831–846.
Magruder, J. R. (2012). High unemployment yet few small firms: The role of centralized bargaining in south africa. *American Economic Journal: Applied Economics*, *4*(3), 138–166.
Makowsky, M. D., & Stratmann, T. (2009). Political economy at any speed: What determines traffic citations? *American Economic Review*, *99*(1), 509–27.
Makowsky, M. D., & Stratmann, T. (2011). More tickets, fewer accidents: How cash-strapped towns make for safer roads. *The Journal of Law and Economics*, *54*(4), 863–888.
Marceau, N. (1997). Competition in crime deterrence. *Canadian Journal of Economics*, *30*(4a), 844–854.
McCormick, R. E., & Tollison, R. D. (1984). Crime on the court. *Journal of Political Economy*, *92*(2), 223–235.
McCrary, J. (2002). Using electoral cycles in police hiring to estimate the effect of police on crime: Comment. *American Economic Review*, *92*(4), 1236–1243.
McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698–714. (The regression discontinuity design: Theory and applications).
Michalopoulos, S., & Papaioannou, E. (2014). National institutions and subnational development in Africa. *The Quarterly Journal of Economics*, *129*(1), 151–213.
Middleton, J. A., & Green, D. P. (2008). Do community-based voter mobilization campaigns work even in battleground states? Evaluating the effectiveness of moveon's 2004 outreach campaign. *Quarterly Journal of Political Science*, *3*(1), 63–82.
Nall, C. (2015). The political consequences of spatial policies: How interstate highways caused geographic polarization. *Journal of Politics*, *77*(2), 394–406.
Ng, K., & Virts, N. (1989). The value of freedom. *Journal of Economic History*, *49*, 960–961.

Novak, W. J. (2008). The myth of the “weak” American State. *The American Historical Review*, *113*(3), 752–772.

Painter, N. (1977). *Exodusters: Black migration to Kansas after reconstruction*. New York: Knopf.

Parrillo, N. R. (2013). *Against the profit motive: The salary revolution in American Government, 1780–1940*. New Haven: Yale University Press.

Pellicer, M., & Wegner, E. (2013). Electoral rules and clientelistic parties: A regression discontinuity approach. *Quarterly Journal of Political Science*, *8*(4), 339–371.

Pence, K. M. (2006). Foreclosing on opportunity: State laws and mortgage credit. *The Review of Economics and Statistics*, *88*(1), 177–182.

Pettersson-Lidbom, P. (2012). Does the size of the legislature affect the size of government? Evidence from two natural experiments. *Journal of Public Economics*, *96*(3), 269–278.

Rabinowitz, H. N. (1971). *Race relations in the Urban South 1865–1890*. Athens: University of Georgia Press.

Roback, J. (1986). The political economy of segregation: The case of segregated streetcars. *The Journal of Economic History*, *46*(4), 893–917.

Salazar, L., Maffioli, A., Aramburu, J., & Adrianzen, M. A. (2016). Estimating the impacts of a fruit fly eradication program in Peru: A geographical regression discontinuity approach. *Inter-American Development Bank Working Paper Series; 677*.

Schumann, A. (2014). Persistence of population shocks: Evidence from the occupation of West Germany after world war ii. *American Economic Journal: Applied Economics*, *6*(3), 189–205.

Selmi, M. (1998). Public vs. private enforcement of civil rights: The case of housing and employment. *UCLA Law Review*, *45*, 1401.

Skocpol, T. (1995). *Protecting soldiers and mothers: The political origins of social policy in United States*. Harvard: Harvard University Press.

Skowronek, S. (1982). *Building a New American State: The expansion of national administrative capacities, 1877–1920*. Cambridge: Cambridge University Press.

Spenkuch, J. L., & Toniatti, D. (2018). Political advertising and election results*. *The Quarterly Journal of Economics*, *133*(4), 1981–2036.

Weaver, V. M., & Lerman, A. E. (2010). Political consequences of the Carceral State. *American Political Science Review*, *104*(4), 817–833.

Wright, G. (2013). *Sharing the prize: The economics of the civil rights revolution in the American South*. Cambridge, MA: Harvard University Press.

Wright, G. C. (1985). *Life behind a Veil: Blacks in Louisville, Kentucky 1865–1930*. Baton Rouge: Louisiana State University Press.

Young, C., Varner, C., Lurie, I. Z., & Prisinzano, R. (2016). Millionaire migration and taxation of the elite: Evidence from administrative data. *American Sociological Review*, *81*(3), 421–446.