Applying Regression Discontinuity Designs to American Political Development

Anna Harvey*
New York University
April 30, 2019

Abstract

Scholars in the subfield 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 these 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 questions. This article explores the application of a specific causal inference strategy, namely regression discontinuity (RD) design, to three questions of interest to APD scholars of state capacity. First, the article illustrates the use of a geographic RD 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 through the use of a population-based RD 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” through the use of a resource constraint RD to estimate the causal impacts of law enforcement effort on a variety of outcomes.

*Department of Politics, New York University, email: anna.harvey@nyu.edu.
1 Introduction

The subfield of American political development (APD) nurtures within the discipline of political science a commitment to a more sweeping historical perspective than is typical in the discipline’s other subfields. In the Introduction to the recently published *Oxford Handbook of American Political Development*, Suzanne Mettler and Richard Valelly observe that the subfield’s “wide-angle lens” may be interpreted by some to necessitate a lesser weight given to questions of causal inference than is typical in the discipline’s other subfields: in the APD subfield, “the reliability of the proposed causal inferences is not taken as utterly primary” (Mettler and Valelly, 2016, p. 4). Mettler and Valelly suggest that giving lesser weight to questions of causal inference may be appropriate in the APD subfield, given the subfield’s commitment to macro-level questions that may not be feasibly addressed by typical causal inference techniques.

Yet attention to causal inference is not necessarily incompatible with a commitment to macro-level questions. This 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, Rueschemeyer and Skocpol, 1985; Carpenter, 2001). This kind of state capacity made possible European programs of social provision, scholars have reasoned; without it, federal social provision in the U.S. was doomed to relative marginality. APD scholars have thus devoted considerable attention to investigating the impacts of the relative 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 presence of the administrative state in Americans’ daily lives.

This article suggests that it is possible to support this work by giving careful attention to what we can infer about the causal nature of these impacts. It illustrates the application of one well-known causal inference technique, namely regression discontinuity (RD) design, to three questions of interest to APD scholars of state capacity. First, the article illustrates the application of a geographic RD (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 impact was severely constrained during the 19th century due to undeveloped administrative capacity (Skowronek, 1982; Carpenter, 2001). Yet, using a GRD design, Harvey and West (2019)
find significant policy impacts from this Reconstruction-era civil rights statute, calling into question whether large federal bureaucracies were or are a necessary condition for social policy impact. GRDs have also been used successfully to demonstrate the long-term impacts of policies and institutions with 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 was due to the inefficacy of those incentives (Parrillo, 2013). According to this account, the salarization of criminal justice professionals, part of the more general growth of a salaried administrative bureaucracy during this time period, better achieved civilian compliance with the criminal law. The article explores the credibility of this claim through the illustration of the use of a population-based RD design to estimate the impacts of financial incentives on law enforcement effort and civilian compliance (Harvey and Mungan, 2019). Using this design, Harvey and Mungan (2019) find that financial incentives in fact appear to be very effective both at motivating targeted law enforcement effort, and at inducing civilian compliance with the law, calling into question whether the decline in the use of these incentives in the late 19th century was in fact due to their inefficacy. Population-based RD designs may also be used productively to examine other questions of interest to APD scholars.

Finally, the article illustrates the potential application of an RD design based on resource constraints to the question of estimating the impacts of policing on outcomes. A recent question of interest among APD scholars has been the impacts of the growth in what some call the “carceral state” (Lerman and Weaver, 2016). A prominent claim has been that encounters between civilians and law enforcement professionals decrease turnout and other forms of civic engagement among those who are the subjects of these encounters (Weaver and Lerman, 2010; Lerman and Weaver, 2014). The article suggests that this 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 RD strategy based on planned law enforcement deployments that are eventually limited by resource constraints, where the threshold distinguishing those 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 RD designs, resource constraint RD designs could productively be used to explore causal inferences about several
other questions potentially of interest to APD scholars.

2 Estimating Federal Policy Impact 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 relative absence of redistributive federal social policies during this period (e.g. Skowronek 1982).\footnote{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 body text; see for example Novak (2008).} 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 particularly affected by the lack of federal state capacity after the Civil War was racial inequality. King and Lieberman (2016) claim, for example, that, “before the civil rights revolution...there was little bureaucratic capacity within the state to enforce racial equality” (p. 246). King and Lieberman 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” (pp. 249, 251).

But private enforcement of federal policy directives through civil litigation and/or criminal complaints, including but not limited to civil rights policies, may be as if not more effective than public enforcement (Farhang, 2010; Lemos and Minzner, 2014). Some have even suggested that private enforcement of civil rights policies in particular may be more effective than public enforcement (Selmi, 1998).

The mere availability of enforcement mechanisms, whether public or private, might also induce compliance with regulatory statutes even in the absence of 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 deterrence effects.

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 lacking condition for the effective implementation of redistributive federal policies during the 19th century, then we should not expect
any such policies enacted during this period to have had significant impacts. Drawing on recent work by Harvey and West (2019), the section assesses the evidence for this claim through an examination of the causal impacts of the Civil Rights Act of 1875.

2.1 The Civil Rights Act of 1875

The Civil Rights Act of 1875, enacted 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 Act provided for 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 1-12 months in prison. In 2013 dollars, the damages available for victims of discrimination ranged from $10,500 to $21,000, while the average African American family income of the time was only approximately $5,250 per year (Ng and Virts, 1989). The Act also provided for 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-$5000 ($21,000 to $105,000 in 2013 dollars). In The Civil Rights Cases, 109 U.S. 3 (1883), the Supreme Court struck the Act’s public accommodations provisions, ruling that Congress lacked the power under either the Thirteenth or Fourteenth Amendments to enact them.²

Prior to the passage of the Civil Rights Act of 1875, there were significant restrictions on African American mobility 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 were also often prohibited from using intracity transportation services, such as streetcars and

²This 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 interstate 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 CRA of 1875 under the interstate commerce clause. The statute itself did not specify a constitutional justification.
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). These barriers to mobility likely raised the relative costs of pursuing economic opportunities that required travel, thereby exerting downward pressure on African American relative wages.

If the Act had been effectively enforced during the approximately 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 this period, however, is that enforcement of federal policy initiatives during this period was necessarily ineffective due to the absence of a large federal bureaucracy staffed by salaried civil servants (Skowronek, 1982; Carpenter, 2001). This story then predicts that the Civil Rights Act of 1875 would have had no causal impacts on African American economic well-being. The question then becomes, how do we estimate those causal impacts?

Historians seeking to assess the Act’s causal impacts have looked for instances of civil suits and criminal prosecutions pursued under the Act. Noting the relative infrequency of such actions, these historians have concluded that the Act was not effectively enforced, and therefore that it had no causal impacts (Franklin, 1974; Gillette, 1979; Wright, 1985; Foner, 1988; Wright, 2013).3

But, as noted earlier, the mere availability of enforcement mechanisms might have induced compliance with the Act’s provisions even in the absence of enforcement actions. In this case, discriminatory actions would have decreased, African American mobility would have increased, and African American wages would have increased, even in the absence of observable enforcement actions. There are in fact anecdotal reports of apparently successful threats to invoke the Act 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 in order 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 suit under the Civil Rights Act of 1875. United States Colored Troops veteran John Solomon Lewis, for example, who sought

---

3Historian John Hope Franklin asserted that the Act “was never effectively enforced” (Franklin, 1974, p. 235). William Gillette characterized the Act as “the most meaningless piece of postwar legislation...the dearest of dead letters,” a characterization echoed by historians George Wright and Eric Foner (Gillette, 1979, pp. 271, 279; Wright, 1985, p. 58; Foner, 1988; Wright, 2013).
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 U.S. Attorney General later issued a public statement to the effect that the companies’ refusal 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).

The right empirical strategy to estimate the Act’s causal impacts, then, would be not to count instances of observable enforcement actions pursued under the Act, but rather to collect behavioral data on wages or other measures of African American economic well-being, before and after the Act’s passage and/or strike, and estimate the Act’s impacts on those outcomes. Following Fogel (2004) and Floud et al. (2011), Harvey and West (2019) measure African American economic well-being during this period using measures of individual-level weight gain and loss. They source weight data from the medical exams of United States Colored Troops (USCT) veterans collected as part of the postwar pension application process, a random sample of which were collected as part of the Early Indicators Project. Because many USCT veterans had repeated exams over time, 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 Act’s strike in 1883, rather than on those bracketing its enactment in 1875.

Even with these individual-level data on African American economic well-being, however, there

---

4Fogel (2004) and Floud et al. (2011) have found that short-term changes in body mass may proxy for short-term changes in economic well-being during periods of relative scarcity.

5Between 1862 and 1890, Union Army and USCT veterans could apply for military pensions by claiming that a current disability, interfering with their ability to perform manual labor, was directly or indirectly related 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). Examining surgeons were given a standardized form to complete directing them to report any medical evidence relevant to an applicant’s claim of disability; the form also directed 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 given multiple exams over time, either because they reapplied after their applications were initially denied, or because they applied for an increase to their current pension.
are still several questions Harvey and West (2019) have to consider before making a causal inference about the Act’s impacts. It is possible, for example, that any observed deterioration in African American well-being after the Act’s strike could have been due to trends in racial conservatism, rather than to the Court’s ruling striking the Act’s public accommodations provisions. Harvey and West address this concern by leveraging the ex ante variation in state-level public accommodations statutes. Many states, generally the more northern states, enacted state-level public accommodations statutes either before or shortly after the Court’s ruling in *The Civil Rights Cases* (Johnson, 1919). In these states, the strike of the Civil Rights Act of 1875 presumably would have had little impact, because these state-level statutes remained in place. One could thus use changes in the well-being of African Americans living in these “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.

Using control and treatment states in this way is a classic differences in differences (DD) design (Dunning, 2012). It assumes that, conditional on a set of observed pretreatment covariates, the entire treated and control areas are comparable to each other. Yet this 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 increasing faster in the more southern states, relative to the more northern states that were more likely to have enacted public accommodations statutes. This relative difference in the evolution of racially hostile attitudes could have led to relative declines in African American well-being in the former states in the 1880s, independently of the Court’s 1883 strike of federal public accommodations protections.

Harvey and West thus deploy an additional research design, one that exploits geographic proximity. If individuals sort around a border between treated and control areas with error, a local treatment effect is identifiable under a geographic regression discontinuity (GRD) framework (Keele and Titiunik, 2015; Keele, Titiunik and Zubizarreta, 2015; Keele and Titiunik, 2016; Keele et al., 2017). Under this design, treated and control individuals near the border may be good counterfactuals for each other, 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 location further from the border).

In the case of the Civil Rights Act of 1875, the primary border between states with and without
state-level public accommodations statutes is depicted in Figure 1. Harvey and West (2019) treat this border as a geographic discontinuity, with the identifying assumption being that variation in racial attitudes (and associated actions) was likely relatively continuous during this 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.

Any relative differences in differences in African American well-being at this border should then be the result of statutory variation between the control and treatment states, not attitudinal variation.

Figure 1: States With Public Accommodations Statutes (in Blue)

States that did not enact public accommodations statutes, however, might also have been less likely to enact statutes providing for 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 American well-being at the border distinguishing states with and without state-level public accommodations statutes.

In the GRD literature, this is known as the problem of compound treatments at the border of interest (Keele and Titiunik, 2016). There are several strategies to address this problem. First, if there is a population that should not be affected by the treatment of interest at the relevant

---

6 There is some anecdotal evidence to support this assumption. In the postwar period, African Americans reported significantly more progressive racial attitudes in the former slave states bordering former nonslave states, relative to the deeper South, while those living in cities just across this border (e.g. Philadelphia) reported attitudes less racially progressive than those found in the more northern cities (Foner, 1973a,b; Wright, 1985).
border, this population can be used as a control group. Presumably the control group should be affected by all the 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 would presumably have affected not only African American veterans near the border of interest, but also white Union Army veterans. Measuring post-ruling changes in African American well-being near the border of interest, relative to post-ruling changes in white well-being near this border, allow Harvey and West to address this possible confounder.

Another strategy to address compound treatments at a geographic border is to look for effects at the border during the period just before the treatment of interest is implemented. During this pretreatment period, if all the other treatments at the border except the treatment of interest remained in effect, one would not expect to find the predicted effect from the treatment of interest. Harvey and West (2019) replicate their DD and GRD specifications during the period that the federal public accommodations provisions remained in effect, between 1875 and 1883. If state-level variation other than that associated with public accommodations statutes were driving relative changes in African-American well-being near the border of interest during the period spanning the Act’s strike, then one would expect to see similar relative changes near this border during the pre-ruling period.

2.2 Applying DD and GRD Designs to the CRA of 1875

To implement their RD and GRD designs, Harvey and West (2019) geo-locate USCT and Union Army veterans’ medical exams to precise latitude and longitude coordinates, computing the shortest distance from each identified set of latitude and longitude coordinates to the border distinguishing those states with state-level public accommodations statutes from those without. They identify those veterans with both pre-ruling and post-ruling exams within an 8-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, and age, the incidence of several systemic medical conditions, and pre-ruling trends in weight; and county-level measures from the 1880 census of the percent of the population living in
towns greater than 2,500 in population, per capita manufacturing output, the percent of a county’s acreage dedicated to farming, and the percent 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 increasing balance in measures of individual-level African American well-being taken prior to the Supreme Court’s ruling, during the period that the Civil Rights Act of 1875 remained in effect, as one approached the border, and balance in pre-ruling covariates near the border (Keele and Titiunik, 2015; Keele, Titiunik and Zubizaretta, 2015; Keele and Titiunik, 2016; Keele et al., 2017).

Harvey and West in fact find that, for both the USCT and the placebo Union Army samples, there are no significant cross-border differences in veteran-level pre-ruling characteristics, within either the full sample or within windows of 300 miles or 200 miles around the border of interest. They do see significant cross-border differences on all county-level pre-ruling census covariates in both the USCT and Union Army samples, at generally the same relative magnitudes and significance levels; their placebo tests using Union Army veterans allow them to test whether these county-level cross-border differences are driving outcomes. They also see that the cross-border differences in pre-ruling census covariates are generally decreasing as 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 DD and GRD models of weight gain/loss in both the full sample and near the border of interest, for both UA and USCT veterans. For the DD models with veteran and year fixed effects and pre-ruling covariates, they find that USCT veterans located in states without public accommodations statutes are estimated to 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 results are not being driven by progressively greater 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 as the average distance in a veteran’s exam locations to the border distinguishing states
with and without state-level public accommodations statutes; the treatment variable as whether the veteran is located on the side of the border without state-level public accommodations statutes; including pre-ruling covariates and fixed effects for border segments; estimating 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 border; and clustering errors in 20 mile intervals on either side of the border, with the interval-identifiers increasing in value with distance south of the border (Calonico, Cattaneo and Titiunik, 2014; Calonico, Cattaneo and Farrell, 2018; Calonico et al., 2018). They 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., 2018). 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.

They find 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 – 4.5 pounds in the subsample for which pre-ruling trends in weight gain/loss are available (including these trends as a covariate). These point estimates, which are estimated within MSE-optimal bandwidths ranging from 72 – 102 miles north of the border and from 135 – 182 miles south of the border, are significant at the 99% – 90% levels. 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 this border. They further find that these estimates are robust to the restriction of the samples to veterans who are reported in the 1860 census to be living on the same side of the border as during their period of interest.

Harvey and West then replicate their RD and GRD specifications during the period that the Civil Rights Act of 1875 remained in effect, assuming a placebo ruling striking this Act 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 DD 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 due to the Courts ruling in *The Civil Rights Cases*. They find that USCT veterans located in states without state-level public accommodation statutes do not become 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 endured by victims of discrimination in public accommodations prior to its passage, and that its strike 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 there was only a minimal “presence and reputation of national authority in local affairs” during this period (Carpenter, 2001, p. 64), the relative absence of federal state capacity may not have been the root cause.

### 2.3 Using GRDs to Estimate Policy Impacts

Geographic boundaries have been used in a number of other contexts to estimate the impacts of policies that change discontinuously at those boundaries; many of these applications may be of interest to APD scholars. For example, GRDs have been used to study 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, Kessler and Meredith, 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 earners (Young et al., 2016), and the effect of private police forces on crime rates (MacDonald, Klick and Grunwald, 2016).

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*) that was implemented by the Spanish government in colonial Peru and Bolivia. The *mita* determined whether indigenous communities were required to send adult male community members to work in silver and mercury mines. The impact of being just within the *mita* boundary, as opposed to just outside the boundary, persists to the present day: regions just within the *mita* boundary are today less integrated with 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 have also impacted 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) land demarcation regime, 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 percent 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 decreasing the transaction costs of exchanging plots in land markets. Where terrain was more rugged, however, land values could have been enhanced through the use of 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 percent higher between 1850 and 1997, relative to similar properties settled just inside the historic VMD boundary. As land becomes more rugged, however, properties settled just inside the VMD boundary become more valuable, relative to properties settled just outside the VMD boundary; this reversal appears to occur when the land has a 6 percent slope (Libecap and Lueck, 2011).

3 Estimating the Impact of Financial Incentives for 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 close of this century was driven by the failure of financially incentivized civil servants to induce compliance with the law (Parrillo, 2013). Providing financial incentives to law enforcement actors, for example, used to be 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.

These financial incentives for the enforcement of criminal laws were largely 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 twenty 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 collections they enabled. Finding relatively 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 to motivate effective law enforcement: “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 there is an alternative hypothesis that could explain Parillo’s finding of few documented instances of assets detected and reported by the ferrets, namely that knowledge of the ferrets’ presence in fact induced increased voluntary compliance with these jurisdictions’ tax laws. With increased voluntary compliance, due to the increased probability of detection of tax evasion in the jurisdictions employing the tax ferrets, there would have been few cases of tax fraud for the ferrets to find. 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).

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 increased compliance, even in the absence of observable in-

---

7Indeed, 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.
stances of enforcement actions. This is also an issue in studies of the effects of financial incentives for enforcement that use 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) and studies estimating the impacts 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, these studies may miss the deterrent effects of enforcement effort.

3.1 Financial Incentives for Traffic Safety Enforcement in Saskatchewan

Harvey and Mungan (2019) instead use data on the frequency, severity, and costs of accidents in Saskatchewan, Canada to study the effects of financial incentives on traffic safety enforcement. In Saskatchewan, the provincial government contracts with the Royal Canadian Mounted Police (RCMP) for traffic safety enforcement. Under this contract, the RCMP 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 contracted to police jurisdictions with populations less than 1500 in the 1991 census. Towns with populations above 500 but less than 1500 in the 1991 census were initially required by the province either to establish and fund their own municipal police services, or to enter into agreements with the province through which the province would delegate to the town both financial and administrative responsibilities for the town’s share of the province’s RCMP costs (The Police Act 1990, effective January 1, 1992). Under these agreements between the province and towns between 500 and 1500 in 1991 population, the RCMP would be directed in its policing by local authorities, and towns would keep the revenue from citations issued within their boundaries.

In 1997 this arrangement was modified (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 now 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 this 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 municipal boundaries, even though these citations
would now be issued by RCMP officers reporting to the province, rather than to the towns (The
Police Regulations 1998). This allocation of citation revenues persisted throughout the remainder
of the duration of this RCMP contract, and into the contract currently in effect.

Figure 2 displays a subsection of the map of Saskatchewan; RCMP detachment districts are
bounded by dark black lines; rural municipalities (the equivalent of counties) are bounded by thin
red lines. The locations of RCMP detachments are plotted in green. The 119 towns to which the
1997 Amendments to The Police Act applied are plotted in black and blue. The 39 towns within
which the province received 100% of citation revenue between 1999 and 2017 are plotted in black;
the 80 towns within which the province received only 25% of citation revenue during this period
are plotted in blue. As is evident in Figure 2, during this period RCMP detachments were often
responsible for policing areas within which the province received different proportions of revenue
from citations issued by RCMP officers.

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 fine
revenue from citations written by RCMP officers within the boundaries of towns with less than 500
in 1996 population, but only 25% of the fine revenue from citations written within the boundaries
of towns with 500 or greater in 1996 population. 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 greater share of
fine revenue, and whether the possibly greater enforcement effort in these towns induced greater
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 increased driver compliance in
the high-revenue towns. In fact, according to this 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
Figure 2: Citation Revenue Distribution, Subsection

Green: RCMP Detachments; Black: Towns within which province receives 100% of fine revenue;
Blue: Towns within which province receives 25% of fine revenue
depends discontinuously on a population threshold, and pursue a population-based RD design to estimate the impact of the policy on driver behavior.

3.2 Applying a Population-Based RD to Traffic Safety Enforcement in SK

If the post-1998 revenue allocation policy incentivized provincial officials to direct greater RCMP traffic safety enforcement effort to towns wherein the province received 100% of citation revenue, relative to towns wherein the province received only 25% of citation revenue, and if drivers responded to the increased probability of citation by driving more safely, we would expect to see fewer and less severe accidents in towns just below the threshold of 500 in 1996 population, 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 enforcement effort: if drivers drove sufficiently more safely in the presence of increased traffic safety enforcement effort, citation rates might not have increased in towns with greater enforcement effort (and might even have decreased).

Eggers et al. (2018) identify two potential threats to causal inference that may arise when population thresholds are used in RD 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 there do not appear to be compound treatments at the threshold of 500 in 1996 population. The threshold originated in a Canada-wide RCMP requirement that municipalities of less than 500 in population must be policed under a provincial contract, while municipalities at or above the threshold of 500 in population may sign municipal contracts with the RCMP. The province adopted this threshold in its 1990 Police Act to require that municipalities above 500 in population sign the equivalent of municipal contracts with the province. It later rescinded this rule in the 1997 Amendments to the Police Act, keeping the threshold of 500 in 1996 population only for the purposes of allocating citation revenue. The province does not use the population threshold of 500 in 1996 population to define rules for other policies. Moreover, populations of Saskatchewan administrative units generally span the threshold of 500 in 1996 population. Towns, the administrative unit that is the focus of the design in
Harvey and Mungan (2019), range between 89 and 4679 in 1996 population. In order to explicitly address the possibility of compound treatments, however, Harvey and Mungan (2019) replicate their discontinuity models during the period prior to the introduction of the revenue allocation rule on January 1, 1999.

There also does not appear to have been manipulative sorting at the threshold of 500 in 1996 population. 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, and/or that provincial officials would have had incentives to manipulate the census data so as to produce more towns below the cutoff (McCrary, 2008). Harvey and Mungan (2019) test for sorting at the threshold using the density manipulation test in Cattaneo, Jansson and Ma (2018); they find no discontinuous jump in the density of 1996 town populations near the cutoff of 500.

Because sorting could still 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 and/or compound treatments, towns just below and just above this cutoff should have been relatively similar to each other on observed pretreatment covariates. They test for continuity in a variety of town-level covariates sourced from the 1996 Canadian census, including: median income, percent married, percent nonwhite (“visible minorities”), percent of the population without a high school degree, and percent of the population comprising 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 March 31, 2017.

Specifying the treatment variable as whether the province receives only 25% of a town’s citation revenue; the running variable as the distance of each town’s 1996 population from the population cutoff of 500 in 1996 population; estimating 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; estimating errors using heteroskedasticity-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, Cattaneo and Titiumik, 2014; Calonico, Cattaneo and Farrell, 2018; Calonico et al., 2018); and using 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 this design is the possibility of displacement or substitution effects (Marceau, 1997). Drivers may respond to greater enforcement effort by driving less in the towns that receive more enforcement effort, driving more in the towns that receive less enforcement effort, and/or opting to use mass transit more frequently in the towns with more enforcement effort. The result might be a relatively lower frequency of accidents in the latter towns simply because of a relatively lower volume of drivers, not because these drivers were deterred from speeding.

Harvey and Mungan (2019) suggest, however, that Saskatchewan’s geography renders such spillover effects highly implausible. First, there is essentially no public transportation in towns of less than 500 in population (or, really, any 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 this 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 are generally not 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. These distances again undermine the plausibility of displacement effects.

Harvey and Mungan (2019) then estimate their RD model on monthly and fully collapsed citation and accident data between January 1, 1999 through March 31, 2017. They consistently find no evidence of discontinuities in stop rates between towns just below and just above the threshold of 500 in 1996 population. However, accident rates, accident-involved vehicle rates, accident cost rates, accident-involved injury rates, accident-involved fatality rates, vehicles involved per accident, cost per accident, injuries per accident, and fatalities per accident all display discontinuous jumps at this population threshold, above which the province receives only 25% of citation revenue. The magnitudes of the effect estimates are striking. For example, the estimated effect of moving just above the population cutoff is to increase accidents by by about 80%, to increase the number of accident-involved vehicles by about 90%, to increase accident costs by by about 100%, and to increase accident-related injuries and fatalities by even larger magnitudes. Increases in accident severity on the intensive margin were of similarly large magnitudes.
Harvey and Mungan (2019) also estimate their RD model at each of 12 placebo cutoffs on either side of the actual cutoff of 500 in 1996 population, where each placebo cutoff creates a new grouping of “treated” and “control” towns 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 RD model for the pretreatment period of April 1, 1995 – December 31, 1998, finding no discontinuities in either accidents or citations during this pretreatment period. Estimating annual RD estimates between 1995 and 2000, they show that the effect of the population threshold on accident outcomes appears only after the introduction of the citation revenue regime on January 1, 1999. They also estimate the effect of the population threshold on stops and accidents flagged in their data as having occurred “near” their towns of interest, or stops and accidents that occurred sufficiently near to a town that an officer could see the town’s grain elevator from the stop or accident location. The province receives 100% of the fine revenue from citations issued in all areas “near” towns, irrespective of whether a town receives citation revenue from citations issued within town boundaries. They therefore expected to see no discontinuities in accidents at the population threshold in these “near” areas, and in fact found no discontinuities at the population threshold in stops or accidents “near” the towns in their RD design.

Harvey and Mungan (2019) also deployed local randomization inference at the population threshold as an alternative estimation strategy (Cattaneo, Frandsen and Titiunik, 2015). This strategy requires identifying the largest window around the population cutoff within which all pre-treatment covariates demonstrate covariate balance in differences in means tests with p-values of at least .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 this window. All five pretreatment covariates demonstrated balance with p-values of at least .15 within a window of 139 in 1996 population on either side of the population cutoff of 500; this window contains 28 towns to the left of the cutoff and 21 towns to the right of the cutoff. Within this window, they obtained p-values for differences in means tests on accident variables on the extensive margin ranging from .028 for accident cost rates to .076 for accident rates.

Finally, Harvey and Mungan (2019) found that cited drivers just below the threshold identifying
the high-revenue towns were given approximately 14 or 22% fewer days to pay back 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 percentage points or 100% higher than the late fines as a percentage of original fines incurred by drivers in low-revenue towns just above the population threshold. These 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 RCMP enforcement effort to towns wherein the province received a larger share of citation revenue, causing drivers to drive more safely and to get into fewer and less injurious and costly accidents. These findings suggest that financial incentives to enforce the criminal law can in fact effectively motivate both enforcement effort and civilian compliance. If the use of financial incentives to motivate law enforcement effort declined in the late nineteenth century, it may not have been because that effort did not increase compliance with the law.

3.3 Using Population-Based RDs to Estimate Policy Impact

Population-based RDs have been used in a variety of other contexts that may be of interest to scholars of American political development. These 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 (Fujimura, 2011), the performance of clientistic parties (Pellicer and Wegner, 2013), and 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 (Alberto De Benedetto and De Paola, 2015); the effects of gender quotas on parties’ electoral performance (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., 2017), and municipal tax rates (Asatryan, Baskaran and
4 Estimating the Impacts of 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 and Weaver (2014, 2016) suggest that the expansion of policing has decreased voting and other forms of civic engagement among those subjected to policing encounters. Their evidence for this claim is drawn from survey data. Weaver and Lerman (2010) and Lerman and Weaver (2014), for example, report negative cross-sectional associations between self-reported encounters with law enforcement and various forms of self-reported civic engagement, and, using panel survey data, negative associations between self-reported encounters with law enforcement and subsequent acts of self-reported civic engagement.

Yet this work does not fully address the nonrandom selection into police contact. Police officers choose whom to stop, using selection criteria that are unobservable to researchers. It is entirely possible that these selection criteria include correlates of low civic engagement. For example, if officers are more likely to make pedestrian stops in lower income neighborhoods, relative to higher income neighborhoods, those whom they stop will already be less likely to vote, relative to those whom they do not stop. Weaver and Lerman (2010) attempt to address this 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 he engages in acts of civic participation. These include standard predictors of turnout, such as race, income, age, education, and gender. However, this strategy requires the strong (and perhaps untenable) assumption that there are no unobserved behavioral or attitudinal correlates of both selection into a police stop and the probability of engaging in acts of civic participation remaining. For example, perhaps those who have decided to disengage from participation in both legal employment markets and the democratic election process are more likely to be observed spending time on street corners during weekdays. The panel approach is likewise 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 sentences of incarceration on voting, finding few to no causal effects. However, this 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, as noted earlier, MacDonald, Klick and Grunwald (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 here 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 terror attacks (Di Tella and Schargrodsky, 2004; Klick and Tabarrok, 2005; Draca, Machin and Witt, 2011). Those seeking to estimate the impacts of the “carceral state” could productively leverage any of these empirical designs.

While these quasi-experimental strategies have credibly isolated 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 this shortcoming in existing designs, this section suggests an RD strategy to estimate the causal impacts of policing based on deployment allocations that are limited by resource constraint thresholds, a scenario that may be a relatively common occurrence across policing agencies.

4.1 Using a Resource Constraint RD to Estimate Policy Impact

Agencies often do not have sufficient resources to meet all of the needs over which they have jurisdiction. Often these needs have a geographic component: a set of locations is identified as deserving of incremental resource allocations, but the agency does not have sufficient resources to make allocations to all locations with need. A common practice employed by agencies 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 lie above this resource constraint threshold. Where agencies follow this procedure to allocate scarce resources, the resource constraint threshold can create the
opportunity for an RD design.

Greenstone and Gallagher (2008) initiated the use of a resource constraint RD to estimate the impact of Superfund sponsored clean-ups 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 an imminent and substantial danger to public welfare and the environment, to place those sites on a National Priorities List (NPL), and to initiate remedial clean-ups at those sites (CERCLA Section 105(8)(B)). Approximately 15,000 candidate sites were initially referred to the EPA for consideration as possible Superfund sites. After winnowing this list to the 690 most hazardous sites, the EPA 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 (those scores exceeding 28.5) were placed on the initial NPL in 1983, making them eligible for Superfund remedial clean-ups. Greenstone and Gallagher (2008) use the HRS threshold of 28.5, distinguishing those sites that just made the 1983 NPL, and those sites that just did not make the list, in a resource constraint RDD estimating the impact of Superfund site identification on housing prices in census tracts adjacent to the sites. They find few effects, suggesting that the designation of polluted areas as Superfund clean-up sites did not increase local property values. Extending this design, however, Gamper-Rabindran and Timmins (2011) examine more finely grained census data lying closer to the same set of proposed Superfund sites, finding that site clean-up increased population density, housing unit density, mean household income, shares of college-educated individuals, and the shares of minorities between 1990 and 2000.

One could imagine deploying a resource constraint RD 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 decreases in traffic fatalities. Between 2007-2012, 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. 30 of these 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 created 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 to receive 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). These 60 positions were filled by making the planned allocations to the 13 detachments ranked highest in 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, create the potential for a resource constraint RD. 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 the enforcement of traffic safety.

The accident and citation data used in Harvey and Mungan (2019), however, are geographically identified to the level of a municipality (at or near). As depicted in Figure 2, RCMP detachments in Saskatchewan police multiple municipalities. The treatment in the CTSS program (additional positions dedicated to traffic safety) was administered at the level of an RCMP detachment. All the municipalities within a detachment (at or near) were treated or not treated with additional CTSS officers. This creates complications for an RD 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 this 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 there is balance on municipality-level pre-treatment covariates, and then estimating differences in post-treatment means for a variety of accident-related outcomes (Cattaneo, Frandsen and Titiumik, 2015). As a placebo test, the design could also replicate this randomization inference strategy using pre-treatment means in accident-related outcomes.

One could further deploy a strategy combining local randomization inference and matching on municipality-level pre-treatment covariates (Keele, Titiunik and Zubizaretta, 2015; Keele et al., 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 these 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 impact of additional law enforcement resources on post-treatment accident-related outcomes. As a placebo test, one could also replicate this blended randomization inference and matching strategy using pre-treatment accident-related outcomes.

A resource constraint RD design could potentially be used 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 are available for the purchase of lands 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 through a ranking worksheet that assigns to each application a score lying between 0 and 400. State conservation offices are further directed, once applications have been ranked, that “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 ACE applications are funded, and below which they are not, could form the threshold in a resource constraint RD. Areas adjacent to lands that are just above the threshold over which ACEs are
5 Conclusion

The subfield of American political development sustains a research community committed to a focus on the evolution of political institutions over time, and on the impacts of these institutions on policies and outcomes. Many in the discipline of political science who may not identify as a member of the subfield nonetheless share an interest in these research questions.

Some may believe that the primacy given to the reliability of causal inferences in most of the discipline’s other subfields is inappropriate for the APD subfield, 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 utility of causal inference strategies for work in American political development, with respect to three questions of interest in 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 induce non-compliance; and the extent to which police contact induces civic disengagement. In each case, the article has also illustrated how regression discontinuity designs can productively be used to make reliable causal inferences that advance our knowledge about these questions. In short, causal inference strategies are perhaps not only compatible with work in the subfield of American political development, but also they may serve to further enhance our understanding of the causal processes underlying institutional development.
References


Evans, Peter B., Dietrich Rueschemeyer and Theda Skocpol, eds. 1985. Bringing the State Back In. Cambridge University Press.


Harvey, Anna and Emily A. West. 2019. “Discrimination in Public Accommodations.” *Revise and Resubmit, Political Science Research and Methods*.


