

How Much Do Elections Increase Police Responsiveness? Evidence from Elected Police Commissioners*

Tobias Nowacki,[†] *Stanford University*

Daniel M. Thompson,[‡] *UCLA*

May 24, 2021

Abstract

How much do elections increase police responsiveness to public preferences? In 2012, directly elected police commissioners replaced a committee of appointed officials overseeing local policing in the UK. We pair police force-level data on arrests, budgets, and voting behavior with a continuous difference-in-differences design to estimate the change in responsiveness. We find that, when police forces switched to directly-elected oversight, left-leaning districts reduced their drug arrest share relative to right-leaning districts, while right-leaning districts reduced administrative staff levels relative to left-leaning districts, both consistent with the different preferences held by partisans on the left and right. We also present evidence that this effect is not concentrated in places with a change in the political party responsible for oversight, suggesting that the increased responsiveness may come from encouraging the parties already overseeing the police to pursue policies the public wants rather than voters changing which party controls the office.

*For helpful discussion, the authors thank Graeme Blair, Cameron DeHart, Andy Eggers, Ted Enamorado, Saad Gulzar, Justin Grimmer, Jens Hainmueller, Andy Hall, Sandy Handan-Nader, Neil Malhotra, Will Marble, Greg Martin, Julia Payson, Michael Sances, Sarah Thompson, Matt Tyler, Jesse Yoder, and members of the Stanford Democracy Policy Lab. For survey data on drug policy preferences, the authors thank Chris Curtis and Sam FitzPatrick at YouGov and Ben Lauderdale at University College London.

[†]Ph.D. Candidate, Department of Political Science. tnowacki@stanford.edu.

[‡]Assistant Professor, Department of Political Science. dthompson@polisci.ucla.edu.

1 Introduction

Can elections make police more responsive to the public? Elections often produce policies that reflect the preferences of the electorate (Besley and Coate 2003; Erikson et al. 1993; Gailmard and Jenkins 2009; Sances 2016; Tausanovitch and Warshaw 2014). Still, elections can fail to produce responsive government when voters cannot observe the official’s actions or their consequences (Canes-Wrone, Herron, and Shotts 2001; Rogers 2017; Snyder Jr and Strömberg 2010). While some policy changes from leadership can change police practices (Mummolo 2018), public oversight of the police may be particularly hard given the limited visibility voters—and even police leadership—have into day-to-day police behavior (Brehm and Gates 1999; Lipsky 1980; Wilson 1978). Would changing police oversight from appointed to elected make police more responsive to public preferences?

We study this question by examining the case of elected police commissioners in England and Wales, a dramatic, large-scale reform to the institutions overseeing police. Prior to 2012, committees made up of local appointees oversaw the performance and budget of 41 police forces in England and Wales. In 2012, each of the 41 police districts replaced their appointed committee with a single elected official. We combine police district-level data on arrests with data on voting behavior, police budgets, police personnel, police stops, the partisan composition of the appointed oversight committees, and the party of the elected commissioners. In the UK, drug policing is a salient issue and area of partisan dispute. The parties also have a long-standing disagreement about the size of government and privatization of formerly public positions. Using a difference-in-differences design, we estimate the differential effect of switching from appointed to elected police commissioners on drug policing across places with different preferences regarding drug prohibition.

We find that left-leaning districts decreased drug arrests noticeably faster and invested more in staff relative to officers than did right-leaning districts once voters elected police oversight directly. We also find that right-leaning districts decreased administrative staff levels more than left-leaning districts after the reform. These effects are not driven by

differential changes in police budgets across places, are not preceded by differential trends, and are robust to alternative measures of preferences. Our results suggest that the elections made police more responsive to public opinion.

We also present suggestive evidence that the increase in responsiveness comes from a change in incentives rather than a change in partisan control. The two most notable mechanisms for increased responsiveness are a change in the people making policy decisions and a change in the incentives officials face (Caughey and Warshaw 2018; Feigenbaum and Hall 2015; Stimson, MacKuen, and Erikson 1995). Both may be responsible at the same time. We divide the police forces into those where party control switched and those where it did not, and we find that the differential effects on behavior are similar for switchers and non-switchers.

Looking beyond England and Wales, our results suggest that directly elected oversight may make police more responsive, but with a few important caveats. England and Wales is typical among Western democracies in the number of elections, but stands apart from the United States which has an unusually high number of elected officials. When voters already face many choices on their ballots, adding new elected positions may make it harder for the public to monitor their governments, increasing the influence of concentrated interests and highly-engaged subpopulations (Anzia 2011; Ashworth and Bueno de Mesquita 2017; Berry and Gersen 2009). Our findings suggest that policy responsiveness could still be increased along one policy dimension by adding more elected offices at the current margin in many Western democracies, but this may not extend to countries with many more elected officials. Also, while many of the barriers to police accountability are shared worldwide—e.g., limited information about police behavior with which to hold police accountable, a broad range of responsibilities that are hard to codify and standardize (Brehm and Gates 1999; Lipsky 1980; Wilson 1978)—some aspects of English and Welsh policing during this period are distinctive, including a meaningful drop in police budgets. We interpret our results as evidence that

elections can make police more responsive and encourage further work studying how these effects vary across contexts.

2 Local Elections and Policy Responsiveness

Should we expect elections to make police more responsive to public preferences? In general, elections should push government officials to be more responsive to public preferences, even when that responsiveness is not desirable. Political economy models of elections describe an agency relationship in which a representative voter intends to select candidates who have preferences similar to their own. This creates an incentive for politicians to behave as though their preferences resemble the voter’s even when they hold different views privately (Alt, Bueno de Mesquita, and Rose 2011; Ashworth 2012; Besley and Case 2003; Fearon 1999; Fournaies and Hall 2018).¹ When voters cannot observe all policy choices or outcomes, these incentives may move policy away from the representative voter’s ideal point (Canes-Wrone, Herron, and Shotts 2001). The median voter’s controlling vote can also move policy in a way that disadvantages minority groups (Meltzer and Richard 1981; Sances 2016).

Appointment processes, on the other hand, shield officials from broad public preferences to some extent. When voters select a single politician to determine policy in a large number of areas, policy in any given area will favor highly motivated interests (e.g., Besley and Coate 2003). The official making the appointment may also be better able to monitor the behavior of the appointee while possibly biasing policy toward the preferences of the appointer and away from the median voter (Gailmard and Jenkins 2009).

Still, electing local police oversight may be less effective as a tool than standard models predict given some of the particular challenges in this setting. Most notably, the public—and even police leadership—often find it difficult to monitor the choices police make on

¹It is important to distinguish policy responsiveness — political actors implementing policies closer to the voters’ ideal point (Erikson et al. 1993) — from a more general form of responsiveness in which governments are faster to handle citizen requests or are less corrupt, though the incentives generally operate in the same way (e.g. Christensen and Ejdemyr 2018; Dipoppa and Grossman 2020; Ferraz and Finan 2011; Grossman 2014; Ofosu 2019).

the beat, and attempts to change police behavior regularly fail (Brehm and Gates 1999; Lipsky 1980; Wilson 1978). Directives that clearly communicate a change in policy and how policy will be monitored can meaningfully change police behavior (Mummolo 2018). Still, efforts by police leadership to change policy in response to local preferences may be thwarted when on-the-ground officers ignore the policy. Some alternative forms of oversight, including civilian oversight boards (Ofer 2015) and court-mandated oversight institutions (Hu and Conrad 2020) have overcome barriers to police accountability in some settings, and constitute a form of ‘fire-alarm’ oversight (McCubbins and Schwartz 1984).² And in the US, consent decrees and other court orders have been used to protect resident rights and change police practice (Devi and Fryer Jr 2020; Harvey and Mattia 2019). However, in a democratic society, some believe police accountability should go further: police should be held responsible for how and where they decide to act and for the crimes they prioritize for enforcement (Bayley 1990; Sen 2010).

A number of distinctive qualities of local elections also make accountability more difficult: Limited information dampens the candidate-specific signal to the wider electorate (Gavazza, Nardotto, and Valletti 2019; Hopkins 2018; Moskowitz 2018; Rogers 2017) and low turnout increases the influence of highly-motivated voters (Anzia 2011). While elections of local criminal justice officials, like judges, prosecutors, and sheriffs, may make government more likely to choose policies the public prefers (Lim 2013; Facchini, Knight, and Testa 2020), they can also fail if bad outcomes are easier to observe than good ones (Huber and Gordon 2004; Gordon and Huber 2007). Still, local elections are often able to achieve some degree of responsiveness (Arnold and Carnes 2012; Christensen and Ejdemyr 2018; Dipoppa and Grossman 2020; Grossman 2014; Payson 2017; Tausanovitch and Warshaw 2014), and local politicians behave as though they are not free to depart from local preferences (Ferraz and Finan 2011; Ferreira and Gyourko 2009; Thompson 2020).

²Hu and Conrad (2020) find that Indian police forces that implemented court-mandated oversight bodies experienced a notable decrease in reported human rights violations.

Lastly, since attention is costly, the gains in responsiveness due to switching from appointment to direct election may decline as the number of elected offices increases (Berry and Gersen 2009). Each additional office has a smaller and smaller effect on policy responsiveness. Meanwhile, monitoring costs increase as new offices are introduced. Increasing the number of elected offices improves the fit between policy and voter preferences only to a point where the marginal cost of monitoring is below the benefit.

3 Case: Policing Under Elected Commissioners

In this section, we describe police oversight in England and Wales before and after they introduced directly elected police and crime commissioners in 2012. We then discuss how we use the creation of commissioners to study the effect of direct elections on police responsiveness, describing our measure of preferences, the data we have assembled, and our research design.

3.1 Police and Crime Commissioners, A New Elected Office

From 1964 to 2012, oversight of local police in England and Wales was conducted by large committees of appointed officials known as police authorities. The 41 authorities were made up of an odd number of appointees,³ typically 17. A bare majority of seats were given to elected officials sitting on local authorities and county councils, distributed to approximate the partisan makeup of these councils. Vacancies for seats dedicated to elected officials were filled by a vote of the existing authority members. The remaining bare minority of seats were held by local citizens. Vacancies were filled in a three-step process wherein citizens applied, a subset were approved by the national government, and the final choice was selected by the existing authority members from the approved list.

³For the purposes of our analysis, we leave out two police forces in London, as their governance structure is non-standard.

Writing in *The Times* in 2005, columnist and now Conservative Member of Parliament and Cabinet Minister Michael Gove argued that “the level of democratic accountability to which the police are subject in Britain is woeful.”⁴ This attitude extended to a broad enough group of Conservative and Liberal Democrat party members that the parties included it in their manifestos in the run up to the 2010 parliamentary election. In 2011, a bill replacing the police authorities with directly elected police and crime commissioners passed through Parliament. The first commissioner elections were held in November 2012, and commissioners took office shortly thereafter.

Scholars, politicians, and activists held a variety of expectations about the consequences of the institutional change. Many politicians and analysts were optimistic that an institutional change would increase responsiveness and improve the efficiency of the police forces (Caless and Owens 2016; Raine and Keasey 2012). Others raised issues like those discussed above that might lead to the failure of local policy responsiveness. The populations overseen by a single commissioner are large enough that most are meaningfully covered in the media. But, some advocates and scholars worried at the time of the reform that tabloid coverage would offer commissioners incentives to pander (Jones, Newburn, and Smith 2012). Other advocates and scholars worried that the local policies would be overtaken by the national party agendas of the commissioners (Lister and Rowe 2015).

After two full election cycles, evaluations of the performance of elected police and crime commissioners are still divided. Their elections have been marred by unusually low turnout (only 15% in 2012) and subsequent concerns about representativeness. This, in turn, led the then Deputy Prime Minister Nick Clegg – who pushed for implementing the reform in the first place – to brand police and crime commissioners a ‘discredited experiment’ (Kirkland 2015).⁵ Both Labour and the Liberal Democrats supported a return to an appointed board

⁴<https://www.thetimes.co.uk/article/if-crimes-on-the-up-your-police-chief-must-explain-why-or-be-sacked-s12582qsrns>

⁵<https://www.bbc.com/news/uk-politics-29172812>

at various times.⁶ Conservative Home Secretary Theresa May, on the other hand, hailed the reform’s success, highlighting the “role police and crime commissioners are playing in making policing more accountable and effective”.⁷

3.2 Left-Leaning Voters and Policing Preferences

If the direct election of police commissioners improves the match between the preferences of the electorate and the policies officials implement, we should be able to see this in police behavior. But many policing preferences may be similar across places—most people likely want to see their police arrest those who are credibly accused of serious crimes. To investigate the effect of direct elections on the match between policy and preferences, we need to study outcomes over which preferences vary across police force areas.

We propose two primary outcomes: drug arrests and staff levels. We have chosen these outcomes because they map onto party politics in the UK, with left-leaning parties more supportive of harm-reduction approaches to drug policy, and right-leaning parties often favoring privatization of many administrative aspects of police work. Drug policing and personnel are also promising as outcomes because police leadership have the authority to change policy in these areas.

In England and Wales, drug policy has been one of the most salient issues in crime and policing at the national and local level with left-leaning parties advocating for less punitive policies.⁸ Labour, Lib Dem, and Green party leadership have supported re-classification or decriminalization of cannabis since as early as 2004 while Conservative leaders have sup-

⁶<https://www.telegraph.co.uk/news/politics/labour/11111745/Labour-plan-to-scrap-Police-Crime-Commissioners.html>

⁷<https://www.publicfinance.co.uk/news/2016/02/police-and-crime-commissioners-could-run-free-schools-may-suggests>

⁸Salient policy issues concerning crime and policing might differ significantly between the US and UK. For instance, debates about police violence are less common in the UK given that many street-level police officers in the UK are not armed with guns and the number of police-caused fatalities in the UK in 2019 was 3. Our intention is not to study the reform’s impact on drug policing in particular, but rather whether it induces responsiveness to outcomes where voters have meaningfully different preferences.

ported more punitive policies.⁹ A pro-legalization activist group that publicly endorses PCC candidates endorsed no Conservative candidates, instead endorsing Labour, Lib Dem, and Green party candidates in most cases.¹⁰ Beyond the salience of the cannabis decriminalization debate, drug supply networks known as “county lines” have developed into an important topic of national policy debate. As with cannabis decriminalization, politicians from opposing parties have emphasized different policy approaches in response—the Labour mayor of London emphasized rehabilitation, while the Conservative-led Home Office announced a crackdown.¹¹ At the local level, left-party candidates for police and crime commissioner have also advocated relaxed drug policy (Austen 2016).¹² Drug policy was also an important issue in the discussion of police and crime commissioners. Some drug policy advocacy groups saw the commissioners as a pressure point to change policy.¹³ The Guardian also listed changing drug policing as one of the top five reasons the new commissioners might be a successful institution worth keeping.¹⁴

⁹<https://www.independent.co.uk/news/uk/home-news/jeremy-corbyn-cannabis-decriminalisation-ridge-sunday-labour-a8425326.html>,
<http://www.bbc.co.uk/newsbeat/article/40819720/lib-dems-under-vince-cable-still-want-to-legalise-cannabis>,
<https://web.archive.org/web/20140511000935/http://policy.greenparty.org.uk/du.html>,
<https://www.telegraph.co.uk/news/uknews/1368775/Tory-crackdown-on-cannabis.html>,
<https://www.independent.co.uk/news/uk/politics/theresa-may-cannabis-legal-uk-law-william-hague-conservatives-latest-a8406111.html>

¹⁰We present a summary of these endorsements in Section A.4.2 in the online appendix.

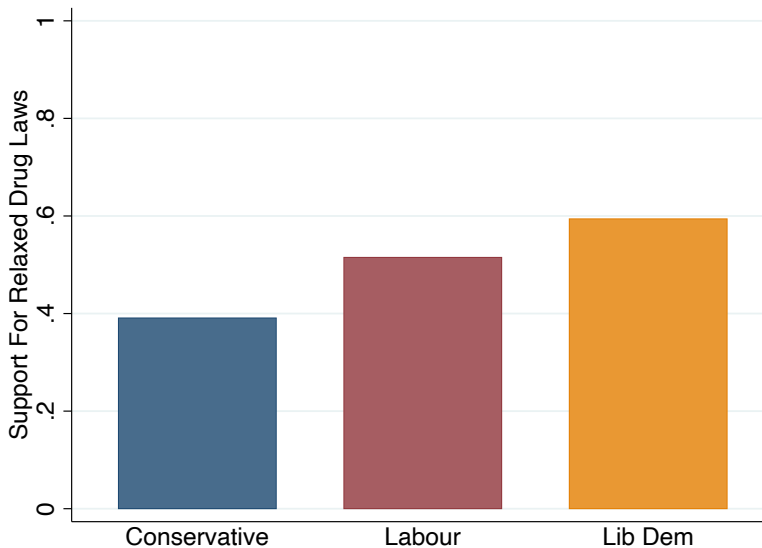
¹¹<https://www.london.gov.uk/press-releases/mayoral/young-londoners-supported-to-leave-county-lines>,
<https://www.theguardian.com/uk-news/2019/oct/01/priti-patel-unveils-county-lines-crackdown>

¹²The Labour PCC from Durham, Ron Hogg, called for a “radical change” in drug policy and emphasized that his goal was to “protect the most vulnerable and help those with a drug dependency to recover and turn their lives around.” (<https://www.bbc.com/news/uk-england-45182442>) The Labour commissioner for the West Midlands, David Jamieson, followed suit with a proposal to “divert those suffering from addiction into treatment and away from the courts” and prescribe “heroin in a medical setting to people suffering from addiction.” (<https://www.westmidlands-pcc.gov.uk/practical-proposals-to-tackle-the-scourge-of-drugs-announced-by-police-and-crime-commissioner/>) The PCCs in three more districts—Thames Valley, Avon and Somerset, and North Wales—have also changed policy that effectively decriminalize possession of most drugs in small amounts and created treatment diversion programs. (<https://www.vice.com/en/article/bvxgn5/britains-second-largest-police-force-to-stop-criminalising-drug-users>)

¹³<https://volteface.me/feature/drugs-policing-radical-changes-drugs-policy-first-great-harvest-police-crime-commissioners/>

¹⁴<https://www.theguardian.com/commentisfree/2013/nov/04/police-crime-commissioners-five-best-worst-ideas>

Figure 1: **Voters for Left Parties in Parliament Favor Relaxed Drug Laws.** Across five YouGov surveys of UK adults, those who reported voting for a Liberal Democrat or Labour candidate in the 2010 parliamentary election were more likely to support decriminalization or legalization of soft drugs such as cannabis than voters for the Conservative party.



In surveys, the typical voter holds the position of their elected leaders—Lib Dem, Labour, and Green voters are more likely than Conservative voters to support legalizing or decriminalizing cannabis. Figure 1 presents the aggregate results of five YouGov surveys of British adults between October 2013 and November 2014 on the subject of drug policy. Across the five survey waves, respondents who reported voting for a Liberal Democrat or Labour candidate in the 2010 parliamentary election were more likely to support legalization or decriminalization of soft drugs such as cannabis.¹⁵ We cannot directly test the link between preferences about drug legalization or decriminalization and preferences for relaxed policing of drugs.¹⁶ Instead, the surveys provide evidence that drug policy is one domain where we

¹⁵While only a small number of voters supported Green party candidates in the 2010 parliamentary election, these voters are even more supportive of decriminalization and legalization—nearly 80% of these voters supported relaxed drugs laws across the five surveys. This is also consistent with the party’s position on drug prohibition at the time: the party has long advocated a very liberal drug policy, and supported cannabis decriminalization and a move towards a regulated, legal market for cannabis products in the 2010 parliamentary election.

¹⁶Simonovits, Guess, and Nagler (2019) correctly point out that responsiveness can be overstated if all policymakers are equally biased toward a particular position and the analyst measures only relative positions.

might expect a divergence in preferences across places based on support for left and right parties.

Drug policy is also an appropriate outcome to study because police have authority to change local policy around drug law enforcement. With drug policing, and cannabis in particular, police forces' responses may vary from a verbal warning or simple 'notice' (akin to a parking ticket) all the way to arrest and prosecution. In this field, police forces thus have enough leeway to shift to a more or less punitive approach.

Another dimension along which left-leaning and right-leaning voters might exhibit different preferences is personnel. In particular, Conservative voters may be more willing to lose some office staff in a trade for more police officers, while left-leaning voters may prefer to prevent staff layoffs. As with drug enforcement, we see evidence of this relationship at the elite level—a majority of Labour candidates explicitly stated an anti-privatization stance prior to the election while no Conservative candidates voiced opposition to privatization (Crawford 2013; White 2015). Unfortunately, we do not know of any surveys that ask the public about police staffing policy, so we are left to infer public preferences based on the party a voter supports.

We pursue two main approaches in operationalizing voters' preferences on drug policing. First, with the above discussion in mind, we use left-party vote share in the 2010 parliamentary election as a natural proxy for drug and personnel preferences. We also produce additional measures in the form of left party vote shares in the 2012 PCC elections and local elections between 2009 and 2012. Second, we use the data from the YouGov surveys to construct a more direct measure of drug policy preferences in each police force. Among these survey-based measures, we include force-level estimates using random forest and multilevel regression poststratification.¹⁷

Our analysis, instead, evaluates whether responsiveness increases, but cannot rule out large, persistent gaps between government policy and public preferences.

¹⁷See Section A.3.5 in the online appendix for additional details and results using alternative measures.

Although we obtain broadly consistent results across all of our measures, we prefer the parliamentary vote share measure and use this for our main results throughout. Given that, as we discussed above, the parties hold different positions on these issues and citizen positions align with their party’s position on average, vote shares convey important information about the electorate’s revealed preference. Party elites and candidates are also unlikely to have access to sophisticated polling and estimation techniques, meaning the left party vote share is likely to be the closest estimate of public opinion among their electorate that police and crime commissioners have.

3.3 Election and Policing Data

In order to test the effect of police and crime commissioner elections on police behavior, we construct four datasets: left party vote shares in the 2010 parliamentary election, arrests by crime category, police force budgets, and police personnel by job type, all by police force area and fiscal year. Throughout our analyses, we drop observations from London, whose police governance structures are non-standard.

The election data originally comes from Pippa Norris’s collection of 2010 parliamentary election results at the constituency level. We gathered digitized vector boundaries for parliamentary constituencies and police force areas from the British government. Using a spatial merge, we identified the share of each parliamentary constituency that falls within each police force area. When a constituency falls entirely inside a police force area, we assign all of its votes to that police force area. Three (out of 573) constituencies do not neatly fall completely inside a single police force area. In those three cases, we assign each constituency’s votes to a police force area based on the share of the constituency’s land area that falls within that police force’s borders. Once we have assigned all of the votes to a police force area, we calculate the total votes cast for each party, and calculate vote shares by party.

Following the previous section, we define left-party vote share as the share of votes cast for a Labour, Liberal Democrat, or Green party candidate.

We gather arrest data from reports issued by the UK Home Office. The reports, typically entitled “Police Powers and Procedures, England and Wales,” contain tables listing the number or share of total arrests by offense type and police force area. These annual reports also include tables listing police stops and searches by the reason for the stop and police force area. We calculate the drug-related share of arrests and stops by totaling the arrests or stops for drugs and dividing by the number of arrests or stops for any reason.¹⁸ All of these statistics are calculated by UK fiscal year, running from April to March of the next year. Going forward, the word year will refer to the year in which the fiscal year ends.¹⁹

We also collect data on police force budgets and employed workforces. Our data on police budgets comes from the annual *Value for Money* reports by Her Majesty’s Inspectorate of Constabulary and Fire and Rescue Services (HMICFRS). These reports break down police revenue streams by fiscal year and into three different categories: central government funding (in the form of grants), locally raised revenue (through a tax precept on local council tax), and drawing from past financial reserves. Data on police personnel comes from the UK Home Office and reports the number of employees by year, rank, function, and gender. For our purposes, we aggregate the counts of employees for two main categories: police officers and staff.

3.4 Empirical Strategy: Continuous Difference-in-Differences

If elected officials are better than appointed boards at producing outcomes residents prefer, we expect the effect of elected oversight to be different in left- and right-leaning districts: elected oversight will cause a bigger reduction in drug arrests in left-leaning districts and a larger drop in administrative staff in right-leaning districts. Our research design estimates

¹⁸A reclassification of offence types in 2015/16 means that some of the reported crime groups may not be comparable across time. However, the ‘drug offence’ category remained stable over time. Using guidance from the Office of National Statistics about the changes, we constructed other crime categories that are consistent over time.

¹⁹The system used to report arrest statistics changed in 2007, leading to several police forces reporting inaccurate or incomplete data for that year. In order to guarantee consistent data quality and maintain comparable specifications across different outcomes, we restrict our data series to the years from 2008 to 2018 (drug stops, personnel, budget) or 2019 (drug arrests).

how different are the treatment effects in left- and right-leaning districts. Unlike most studies which intend to estimate the average effect of a reform, the quantity we want to estimate is the difference in the effect of the reform conditional on vote preferences, a conditional average treatment effect (CATE).

To estimate these differential effects, we estimate regression equations of the form

$$Y_{ft} = \tau LeftShare_f * (Year_t > 2012) + \gamma_f + \delta_t + \epsilon_{ft}$$

where Y_{ft} is an outcome and $LeftShare_f$ is the share of votes going to Labour, Lib Dem, or Green candidates in the 2010 elections for parliament, falling between zero and one. γ_f and δ_t are police force and year fixed effects, respectively. τ is an estimate of the effect of elected police commissioners on the share of drug arrests in further left districts as compared to further right districts.²⁰

The design is akin to a classic difference-in-differences design but with a few modifications so that we estimate a difference in treatment effects rather than an average treatment effect. In the standard difference-in-differences set-up, the analyst follows a group of units before and after they are exposed to a treatment and compare them to units never exposed to treatment. Here, we measure the change in outcomes from before to after oversight is elected, comparing districts that vote more for left parties to places that vote more for right parties, using a continuous measure of left party voting (Angrist and Pischke 2008; Card 1992).²¹ For exposition, we will pretend as though there are only two groups—the intuition generalizes to the continuous case. Since both groups are subject to the treatment, we are no longer estimating the average treatment effect on the treated. Instead, we estimate the

²⁰In this research design, classic cluster robust standard errors are known to be biased in small samples (Cameron, Gelbach, and Miller 2008). Given the limited number of police forces, we report standard errors from a clustered bootstrap procedure that tends to perform better in smaller samples, clustering on police force. Despite the concerns described in Cameron, Gelbach, and Miller (2008), the estimated standard errors are similar using the clustered wild bootstrap procedure they describe, the blocked bootstrap procedure in Bertrand, Duflo, and Mullainathan (2004), and the classic cluster robust standard errors.

²¹For examples of recent work using this research design in political science, see Feigenbaum and Hall (2015) and Lueders, Hainmueller, and Lawrence (2017).

difference in the conditional average treatment effect between these two groups. In other words, we estimate a function of multiple causal effects rather than estimating any one causal effect directly. This differential effect is identified under the usual difference-in-differences assumption, that the difference in police behavior between left- and right-leaning places would have remained the same had police and crime commissioners not been introduced. This assumption is not directly testable, but we can interrogate it by assessing whether the police behavior is moving in parallel across left and right districts prior to the introduction of commissioners. We discuss all of our checks in Section 4.4 and evaluate this assumption thoroughly in the appendix.

The significant financial austerity in the UK beginning in 2011 poses a potential threat to the assumption that left and right police forces would have been on similar trends in the absence of oversight reform. While the national government cut police budgets everywhere,²² the budget cuts were especially deep in areas with larger budgets before 2012, and many of these places have higher support for left parties (Fetzer 2019). In the same period, police forces and other local governments were highly restricted in their ability to raise additional revenue, leaving them unable to fill the gap left by the national government. Because the cuts hit left-leaning districts more than right-leaning places, we may have seen left-leaning places reduce drug arrests or changed staffing even if elected police commissioners were never introduced.

While the differential budget cuts pose a potential threat, the fact that local governments, including police commissioners, were unable to balance the shortfall with increases in revenue during this period offers a solution: We can study districts with similar budgets prior to reform and check whether, after reform, the police budgets in left-leaning districts are bigger than in the right-leaning districts with similar pre-reform budgets.²³ We find that, once we adjust for 2012 revenue, left- and right-leaning police forces had similar budgets

²²<https://www.nao.org.uk/report/financial-sustainability-of-police-forces-in-england-and-wales/>

²³For a more precise description of our view on the relationship between local preferences, police budgets, and policing outcomes, see the causal diagram in Section A.1 in the appendix.

after commissioners were introduced. We present this check in A.3.3, finding that budgets are similar after reform in left- and right-leaning districts with similar pre-reform budgets. Knowing that budgets in left- and right-leaning districts are similar after the reform once we account for the district’s pre-reform budget, we make this adjustment in every analysis throughout the paper. We do so in two ways: first, we include year-by-2012 revenue-decile fixed effects. This breaks our data into ten sets of four police forces with similar revenue in 2012, estimates the differential effect of direct elections in left- vs right-leaning forces within each group, and takes a variance-weighted average of these differential effects. As a second, alternative, approach, we estimate year-specific slopes on 2012 revenue which allows for the relationship between the outcome (personnel or arrests) and pre-reform revenue to vary over time.

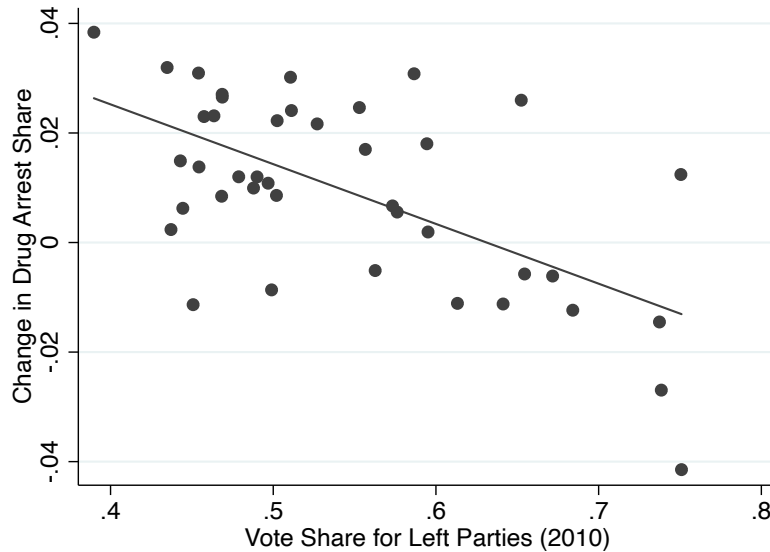
Also, our primary outcomes—policing (arrests per capita, drug arrest share, stops per capita, drug stop share) and personnel (employees per capita, officer share, and staff share)—are connected in important ways. Having more police officers increases a police force’s capacity to make stops and arrests. Also, while police can make an arrest without recording a stop—e.g., the police can arrest someone who is openly carrying a controlled substance or is accused of participating in trafficking without a stop and search—many arrests, especially drug arrests, begin with a stop.²⁴ When the treatment affects what cases are observable, classic treatment effect estimators may be biased under standard assumptions (Knox, Lowe, and Mummolo 2020). Yet, no such dependencies threaten our analysis—all outcomes are observable for all districts regardless of local preferences.

4 Elected Police Respond to Public Preferences

In this section, we present evidence that introducing directly elected police oversight caused more responsive policing and personnel policy. First, we show that elected oversight caused police to reduce drug arrests faster in left-leaning districts, consistent with our expectation

²⁴We describe these dependencies in more detail and plot them in Section A.1 in the appendix.

Figure 2: **More Left-Leaning Places Experienced a Bigger Drop in Drug Arrest Shares After the Reform.** Police authorities with higher vote shares for Labour, Liberal Democrats, and the Green Party in the 2010 general election experienced a smaller increase (or even decrease) in the average share of drug arrests after the reform, compared to the average share of drug arrests before the reform.



that voters in left-leaning districts prefer less drug enforcement. We then present weaker evidence that directly-elected oversight may have reduced drug stops in left-leaning districts faster than in right-leaning districts. Third, we present our finding that introducing police commissioners caused a shift in right-leaning districts away from investing in administrative staff and toward investments in officers relative to left-leaning districts. This is consistent with our expectation that left-leaning districts are more pro-labor while right-leaning districts are more open to privatization. Finally, we describe a number of robustness checks we use to validate our parallel trends assumption and our measure of voter preferences.

4.1 Arrests per Capita and Drug Arrests Reduced in Left Strongholds Relative to Right Strongholds

An initial look at the raw data on policing suggests that the most left-voting places had a steeper drop in drug policing after the introduction of police and crime commissioners than

Table 1: **Elected Police Commissioners Caused Police Forces in Left-Leaning Districts to Make Fewer Arrests per Capita and Reduce Drug Arrest Share.**

	Arrests Per Capita			Drug Arrest Share [0,1]		
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Left Share	-0.013 (0.005)	-0.021 (0.008)	-0.017 (0.007)	-0.109 (0.028)	-0.091 (0.038)	-0.061 (0.034)
Mean	0.018	0.018	0.018	0.082	0.082	0.082
# Forces	41	40	40	41	40	40
# Years	11	11	11	11	11	11
# Obs	448	437	437	448	437	437
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	No	No	Yes	No	No
Year \times Revenue Decile FEs	No	Yes	No	No	Yes	No
Year \times Revenue IFEs	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2019.

did right-voting places on average. Figure 2 captures this pattern. Each dot represents a police force; the horizontal axis captures the parliamentary vote share for left parties in 2010, while the vertical axis denotes the change in the average drug arrest share (drugs arrests as a share of all arrests) between the pre-reform period (2008 - 2012) and the post-reform period (2014 - 2019). We see that after the introduction of elected police and crime commissioners, the average share of drug arrests increased in the most right-voting places, while it decreased in the most left-voting places.

We estimate this relationship formally using difference-in-differences analyses, finding the same pattern. Table 1 reports the formal estimates of the change in total arrests per capita (columns 1-3) and drug arrest share (columns 4-6).²⁵ For both outcomes, the first column

²⁵For completeness, we also report the change in the share of other arrest categories in Section A.2.2 in the appendix.

presents the estimates from a standard two-way fixed effects specification as discussed in section 3.4. In the next columns (2 and 5), we add year-by-revenue decile fixed effects in order to limit our comparisons to districts where we expect similar levels of austerity. Finally, we also report estimates when including year-specific slopes on 2012 total revenue (columns 3 and 6). We include unit-specific fixed effects throughout.

Across all six columns, the differential effect of elected police and crime commissioners is substantively large and statistically distinguishable from zero at conventional levels. To give a sense of the magnitude, we would expect total arrests per capita to drop by about 0.002 when moving from the 25th percentile of left-voting police forces to the 75th.²⁶ Given a mean value of 0.018 arrests per capita, this represents an approximately 11% effect relative to the average arrest per capita rate. Conversely, for the same movement from the 25th to the 75th percentile in left-voting authorities, we would expect a decrease of almost 0.8 percentage points in the share of drug arrests, which, relative to the mean value (0.082), translates to an effect size of just below 10 per cent.

These linear effect estimates are simple and have the best power to detect an effect if the effect of switching to elected police oversight is approximately linear in local preferences. But, if the effect is non-linear, our linear estimators can produce substantively misleading results. Most importantly for our analysis, if drug arrests decreased most in places just to the left of the median district but dropped at the same rate in the left-most and right-most districts, this would be inconsistent with the type of responsiveness we are describing. In Appendix A.2.1, we replicate Table 1 but using a treatment variable binned into terciles following the advice in Hainmueller, Mummolo, and Xu (2019). The results from this analysis suggest our interpretation is appropriate: The most left-leaning districts saw larger declines in arrests per capita and drug arrests than districts in the middle and right. Most point estimates for districts in the middle in terms of left-party parliamentary vote share are negative, indicating a faster drop in arrests per capita and drug arrests. In many cases, we can reject the null

²⁶The 25th percentile had a left voting share of 47%, while the 75th percentile had a left voting share of 60%, leaving us with a movement of 13 percentage points.

that arrests per capita and drug arrests declined as quickly in the most right-leaning places as in the most left-leaning districts. Yet, across all specifications, the differences between the middle-of-the-road districts and the right-most districts are not large enough to reject the null hypothesis of no difference at conventional levels. Put together, these results are consistent with the type of responsiveness we are describing: we observe the largest declines in arrests in the furthest left districts, smaller declines in middle-of-the-road districts, and even smaller declines in the right-most districts.

4.2 Drug Stops May Have Declined in Left Strongholds Relative to Right Strongholds

During our study period, many drug arrests began as street or vehicle searches. Police in England and Wales are permitted to stop and search people when they have “reasonable grounds to suspect” that you are carrying illegal drugs.²⁷ This suggests one way police forces may shift priorities away from drug enforcement: limiting the number of searches or reducing the share where drugs are the focus.

We find some weak, suggestive evidence that drug stops may have declined as a share of stops in left-leaning districts relative to right-leaning districts. We do not find consistent evidence that police in left-leaning districts reduced overall searches faster than police in right-leaning districts. Table 2 reports our formal estimates. The columns use the same specifications as Table 1. Across all columns, our estimates are too noisy to rule out large effects or null effects, once we have adjusted for the potential confounding from budgets changing differently in left- and right-leaning districts. For example, Column 3 reports a point estimate that implies, relative the 25th percentile left-share district, that the 75th percentile left-share district reduced 0.65 stops per 1,000 residents per year. This is approximately 6.5% of the stop rate for the average force in the average year. Yet the 95% confidence interval ranges from a relative drop in stops in left-leaning places of 3.96 stops

²⁷<https://www.gov.uk/police-powers-to-stop-and-search-your-rights>

Table 2: **Elected Police Commissioners Caused Police Forces in Left-Leaning Districts to Make Fewer Stops per Capita and Reduce Drug Stop Share.**

	Stops Per Capita			Drug Stop Share [0,1]		
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Left Share	-0.024 (0.010)	-0.012 (0.017)	-0.005 (0.013)	-0.333 (0.109)	-0.184 (0.174)	-0.140 (0.154)
Mean	0.010	0.010	0.010	0.516	0.516	0.516
# Forces	41	40	40	41	40	40
# Years	10	10	10	10	10	10
# Obs	409	399	399	409	399	399
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	No	No	Yes	No	No
Year \times Revenue Decile FEs	No	Yes	No	No	Yes	No
Year \times Revenue IFEs	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2019.

per 1,000 residents up to a relative increase of 2.66 stops per 1,000 residents—a 40% or 26% difference over the average stop rate, respectively. Our estimates of the differential effects on drug stops are not as noisy—our point estimate in Column 6 is a relative decline in drug stop share in left-leaning districts of 3.5% of the typical drug stop share with a confidence interval including a drop of 11.1% and an increase of 4.1%—but it is still noisy relative to our estimates of on arrests.

Drug stops are just as related to our theory as arrests—there is no ex ante reason to expect voters will have meaningfully different preferences about drug arrest and drug stops. While we do not have direct evidence for this, our pattern of results is consistent with drug stops being a noisier measure of police behavior. For example, while a drug arrest is more closely connected to a charge and therefore is important to properly classify, a stop

may be harder to classify, adding noise to the measure. Further, because an arrest involves taking custody of a person and a stop does not, officers may be less likely to file the proper paperwork associated with a stop than with a more serious action like an arrest. Still, other possible differences between arrest and stop data may cut the other direction—stops records may be easier to manipulate on the margin and therefore may be an area where police can cheaply meet citizen demands without meaningfully changing their practices. We take these arguments as potentially explaining, *ex post*, why the relationship between left voting and the effect of reform on drug stops is less clear than the differential effect of reform on drug arrests.

Overall, our findings are consistent with left-leaning police forces shifting away from drug stop and searches as part of the way that they reduced drug arrests relative to right-leaning forces, but the evidence is only suggestive. Still, we cautiously read this result as consistent with our argument, understanding that this is only weak evidence given the uncertainty.

4.3 Shift Toward More Officers, Fewer Staff in Right Stronghold Relative to Left Strongholds

As we discussed in Section 3.2, Labour politicians actively campaigned against staff cuts and privatization, which would overwhelmingly affect non-officer employees. Conservative politicians did not actively speak out on this issue. While we cannot directly assess whether voters in left- and right-leaning districts have meaningfully different preferences over police personnel matters, the party stances on these issues suggest that right-leaning districts might be more open to staff cuts and privatization.

Accordingly, we measure how much the introduction of directly elected police oversight changed the gap between left- and right-leaning districts in the composition of police office employees. Table 3 reports the results using the same set of specifications as before. We find that right-leaning districts increased the share of employees who were officers and reduced the share who were non-officer staff relative to left-leaning districts. We also find that, while

Table 3: **Elected Police Commissioners Caused Police Forces in Left-Leaning Districts to Shift Personnel Away from Officers, Toward Staff.**

	Employees Per 1k Pop			Officer Share			Staff Share		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post \times Left Share	-1.054 (0.366)	-0.907 (0.467)	-0.653 (0.503)	-0.091 (0.069)	-0.119 (0.063)	-0.157 (0.075)	0.082 (0.063)	0.133 (0.064)	0.151 (0.058)
Mean	3.468	3.468	3.468	0.578	0.578	0.578	0.063	0.063	0.063
# Forces	41	40	40	41	40	40	41	40	40
# Years	10	10	10	10	10	10	10	10	10
# Obs	410	400	400	410	400	400	410	400	400
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	No	No	Yes	No	No	Yes	No	No
Year \times Revenue Tercile FEs	No	Yes	No	No	Yes	No	No	Yes	No
Year \times Revenue IFEs	No	No	Yes	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Tercile divides the police forces into terciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2018.

left-leaning places reduced the number of people they employ by a larger amount, these differences may partly be an effect of austerity as they attenuate somewhat when we adjust for employment trends based on pre-treatment revenue levels.

Our preferred point estimate, reported in Column 9, implies that, relative to a 75th-percentile left-share district, a 25th percentile district (further to the right) reduced the share of employees who are staff by nearly 2 percentage points. 6.3% of police employees are staff in the average police force in the average year; a nearly 2 percentage point change is quite substantial. The 95% confidence interval for this estimate ranges from a 0.4 percentage point differential effect to a 3.4 percentage point differential effect.

4.4 Prodding Robustness of Differential Effects of Reform

While we prod the robustness of our differential effects on each outcome above, each of these analyses relies on the assumption that left and right districts would have followed the same trajectory but for the introduction of elected police commissioners. We find that this assumption is plausible in each of the next four subsections, finding that (1) left and

right districts were on similar trajectories prior to 2013, (2) the differential effects on our primary outcomes—drug arrest share, drug stop share, and staff share—are similar even when comparing districts only to the district most similar on pre-reform outcomes, (3) the differences between left and right districts are not a consequence of differential austerity, and (4) the differences between left and right districts were not simply urban-rural differences or cross-country differences. We also discuss our finding that the differential effects are similar across a number of measures of voter preferences.

4.4.1 Left and Right Districts Were on Similar Trends Prior to 2013

One potential threat to our analyses above is that police behavior and personnel could be changing differently in left vs right districts for other reasons unrelated to the adoption of directly elected oversight. If, for example, drug policing was polarizing as an issue over time, our difference-in-differences estimator would be biased, leading us to think directly elected oversight caused more responsive policing when it would have happened regardless.

To address this concern, we plot and formally estimate how much left and right districts were moving apart prior to the adoption of police commissioners in 2013. These analyses, reported in Section A.3.1 in the appendix, suggest that our arrest and personnel measures were trending in similar ways in left- and right-leaning districts prior to the adoption of elected commissioners.

This evidence is only suggestive—the estimates of trending are imprecise. While we cannot reject the null hypothesis that left- and right-leaning districts are on the same trends prior to the reform, the confidence intervals are large in every case. We address this by pooling estimates across multiple pre-reform years, but the estimates are still relatively imprecise.

4.4.2 Estimates Are Similar After Matching on Pre-2012 Trajectories

Since the direct tests for pre-reform trending are noisy, though still consistent with no trending, we implement a matching procedure that accounts for trends explicitly. To do so, we get the best possible match of pre-treatment trajectories between relatively left-wing and right-wing districts, as in a pair blocking exercise (Imai et al. 2009).²⁸ We describe the details and formal findings from this analysis in Section A.3.2.1.

On our primary outcomes—drug arrest share, drug stop share, and staff share—our estimates are in the same direction as in our above analyses, though they are often somewhat smaller and occasionally too imprecise to confidently reject a null hypothesis of no difference. Our additional outcomes that relate to the scale of police office activities—arrests, stops, and employees per capita—are all imprecise and positive rather than negative as we saw above. Put together, we take this as evidence that, while differential trends may confound our results on arrests, stops, and employment per capita, the differential effect of elected commissioners on the outcomes over which left- and right-leaning voter preferences are most likely to be different—drug policing and privatization of the bureaucracy—are not seriously confounded by trends.

4.4.3 Differential Effects Not Driven by Austerity

As we discuss in Section 3.4, austerity poses an important potential threat to our analysis. Left-leaning places had larger budgets prior to the reform, and districts with the largest budgets faced the deepest cuts under austerity. In the preceding analyses, we find that drug arrests decreased faster in left-leaning districts even when compared to right-leaning districts with similar pre-reform budgets. We also find that the staff share of employees increases in left-leaning districts faster than in right-leaning districts with similar budgets pre-reform. As we explain in Section 3.4, districts were severely limited in their ability to

²⁸Our matching approach is closely related to Hazlett and Xu (2018); Imai, Kim, and Wang (2018) but finds matched pairs when the treatment variable is continuous.

increase local revenue during austerity, so we would only expect revenue to drop faster in left-leaning districts because they had more revenue pre-reform—the cuts from the central government were targeted at reducing budgets where they were large, not explicitly at left-leaning districts.

In Section A.3.3 in the appendix, we directly test our assumption that budget cuts were similar in left and right districts after accounting for pre-reform revenue. We use the same regression specifications as in all of our tables in the preceding analyses, measuring the differential change in revenue for left and right districts with similar revenue pre-reform. When we compare left and right districts with similar revenue in 2012, we find no evidence that left-leaning districts lost more revenue than right-leaning districts.

4.4.4 Differential Effects Not Driven by Country or Urban-Rural Trends

Beyond differences between left and right districts in terms of levels of austerity, left districts may be concentrated more in urban areas or in England rather than Wales. If urban and rural districts were on different trends after reform, or if England and Wales were on different trends, this, too would threaten our analysis. In Sections A.3.4.1 and A.3.4.2 we present consistent evidence that neither urban-rural differential trends nor England-Wales differential trends are driving our results.

4.4.5 Differential Effects Similar Across Alternative Measures of Preferences

In the analyses above, we use support for left-party candidates in the 2010 parliamentary election as our measure of preferences. We discuss this choice in Section 3.2. We use parliamentary vote because it captures revealed preferences over a broad set of policies and citizens are more likely to participate and be informed about parliament than local races. We also prefer parliamentary vote to local party vote because we know members of parliament toe the party line, so voters can infer their member’s policies from her party label. Lastly, we like

party vote share because it is a measure of preferences that police and crime commissioners can trust, easily observe, and use to approximate voter preferences.

Still, to be confident that our estimates are not sensitive to our measure of preferences, in Section A.3.5, we estimate the differential effects on drug arrest share using seven alternative preference measures: support for left parties in local races from 2009 to 2012, support for left parties in the 2012 PCC election, and five survey-based measures of support for legalization or decriminalization of cannabis using multilevel regression and post-stratification. We build these five more direct measures of drug policy preferences using different techniques for small area preference estimation: survey disaggregation, random forest and post-stratification using demographic predictors, random forest and post-stratification using political predictors, multilevel regression and post-stratification using demographic predictors, and multilevel regression and post-stratification using political predictors. We find a similar pattern of results across all eight measures, with one important caveat: While we consistently find that drug arrests drop faster in culturally liberal districts, the effect sizes vary considerably across survey-based measures. As we discuss in detail in Section A.3.5, this reflects differences in the techniques we use for small area estimation—the effects are small when we use disaggregation which is a high-variance estimator; the effects are large when we use MRP with few interactions because it is a low-variance estimator. We find reasonable effect sizes when we use our preference measures from a random forest and post-stratification procedure which seeks to find a error-minimizing balance between bias and variance.

Overall, we interpret the findings using these additional preference measures as consistent with our main findings. All of our estimates of differential effects of the reform on drug policing using vote share measures are very similar to our main estimates. While our survey-based estimates are noisier, we view them as consistent with our main findings.

5 Accountability or Selection?

The two primary mechanisms for policy responsiveness are a change in the ideological composition of an office (selection) and a change in the incentives the officers face (accountability). These proposed mechanisms raise a natural question: is the change in policy mostly located in places where the ideology of the officer overseeing the police changed, or was the change in policy broadly felt?

In order to tease this apart, we find all police forces for which the party of the overseeing official changed. We do this by constructing an additional dataset that notes the party affiliations of all authority members prior to the introduction of commissioners, and we pair it with data on the party of the commissioners. We then remove the authority members who are independent community members and focus on the political members. We locate the median party member for each authority, labeling Labour, Lib Dem, and Green members as -1, Conservative and UKIP members as 1, and all others as 0. This allows us to compare the party of the median authority member to the party of the commissioner.

Very few police forces go from having a left-wing majority to a right-wing commissioner or vice versa, but many switch from a partisan authority to an independent commissioner or switch from an independent authority to a partisan commissioner. Table A.14 in the appendix reports the share of police forces by their original authority median and the party of their elected commissioner. More than half of the police forces controlled by a Conservative majority in their authority elected and independent as commissioner.

We find that the differential change in drug arrest share and arrests per capita is similar in places where the party controlling oversight changed and places where it did not. Table 4 presents our formal estimates. The columns match the columns reported in Table 1, but we break out the differential change in drug arrest share for forces that switch party control and those that do not. The fact that estimates on the second row hover around zero suggests that the party in charge of overseeing the police force is not the most important factor in producing policy responsiveness.

Table 4: **Elected Police Commissioners Caused Police Forces in Left-Leaning Districts to Make Fewer Arrests per Capita and Reduce Drug Arrest Share.**

	Arrests Per Capita			Drug Arrest Share [0,1]		
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Left Share	-0.008 (0.005)	-0.013 (0.008)	-0.012 (0.006)	-0.096 (0.021)	-0.085 (0.037)	-0.063 (0.035)
Switch \times Post \times Left Share	-0.000 (0.002)	-0.002 (0.002)	0.000 (0.002)	0.016 (0.008)	0.004 (0.013)	0.014 (0.008)
Mean	0.021	0.021	0.021	0.082	0.082	0.082
# Forces	41	40	40	41	40	40
# Years	8	8	8	8	8	8
# Obs	328	320	320	328	320	320
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	No	No	Yes	No	No
Year \times Revenue Decile FEs	No	Yes	No	No	Yes	No
Year \times Revenue IFEs	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Switch is a dummy variable indicating whether partisan control of police oversight changed after PCCs were introduced. Data from 2008 to 2016.

These results do not rule out selection as an important mechanism by which voters can produce policy responsiveness. It is possible that independents have considerable ideological overlap with partisans on matters of drug policing or that there is considerable within-party ideological heterogeneity among officials. If either of these facts are true, voters may be getting the policy they prefer by selecting the particular politician within a given party that shares their preferences. Party control does not capture this type of selection.

Further, this analysis is only suggestive that partisan selection is not the main mechanism for the differential effects we observe. Districts in which the party controlling police oversight changes are not random—they may be systematically different from districts that have the same party controlling police oversight. If, for example more independents won in places

with less media coverage, party control may be important but counterbalanced by less media scrutiny and weaker incentives to be responsive.

Still, this analysis rules out one of the most commonly cited selection-based mechanisms for policy responsiveness and suggests that reelection incentives may be a more important reason for the responsiveness we observe.

6 Discussion

This article investigates an example of the relatively rare occurrence of adding a powerful elected office in a mature democracy. While this does not happen often, these cases are useful because they give us a window into whether elections are producing the responsiveness we expect. These results suggest that, at least in a context where there are relatively few powerful locally-elected officials, adding a new one still improves policy responsiveness. This does not rule out welfare-reducing reasons for improved responsiveness such as pandering, but it rules out a failure to respond to citizen preferences. And this result does not necessarily imply that the reason for the drop in drug arrest share in left-wing places relative to right-wing places was the political preferences of the electorate. This finding also does not mean imply that changing local police behavior through oversight is easy—officers may be able to prevent many new policies from taking effect on the ground, attenuating the relationship between voter preferences, district policy, and actual police behavior. Still, from the perspective of the residents living in the police districts we studied in England and Wales, the effects are observationally equivalent with reform-induced responsiveness.

This case helps to clarify something that often gets lost in work on responsiveness: the appropriate standard by which to evaluate an institution is in comparison to other viable institutions. Here, we are able to compare an appointed committee with a directly elected official. This particular comparison cannot distinguish between the effect of boards and elections. Other cases will provide different comparisons and help us learn about different

aspects of reform. Putting together studies of the consequences of adopting different institutions is necessary to test the predictions from political economy models about when elections succeed to produce responsive government.

Responsiveness was one of the core justifications for electing police commissioners that advocates articulated. Nick Herbert, the police minister during the transition to police and crime commissioners and an advocate for the change, wrote prior to the change:

“Over the years, the police have become estranged from the municipalities from which they sprang and increasingly look to the Home Office. ... From the first elections in May next year, the public will have a real say over how their area is policed.”²⁹

Herbert also highlighted the mandate of police and crime commissioners to “tackle drugs and work with local authorities and agencies.” Many of the explanations for why this goal may not be achieved were also evident at the time, including in a Guardian editorial in the months ahead of the first commissioner elections:

“So far, barely a quarter of voters even know that elections for the new commissioners will take place in November. There are serious concerns about turnout not being high enough to give the bodies legitimacy. There are worries about the caliber of candidates, and the rules by which they will fight the elections.”³⁰

Our findings suggest that this policy change had at least one its intended effects.

While responsiveness is often a desirable property of democratic institutions, policing is an area that challenges this intuition. What policing decisions should be decided by a majority? What police behavior should be limited to protect fundamental rights, regardless of majority preference? These are difficult normative questions that we cannot answer in

²⁹<https://www.telegraph.co.uk/news/uknews/crime/8410429/Its-time-for-you-to-have-a-say-on-policing.html>

³⁰<https://www.theguardian.com/commentisfree/2012/jan/08/police-crime-commissioners-ratcatcher-vote>

this paper. Further, it is important to note that not all citizens bare the same cost of police choices or reap the same benefits. In England and Wales, Black individuals are stopped at more than three times the rate of White, Asian, and other residents.³¹ Considering the varying intensity of policing in different communities and the thorny normative questions, still more work needs to be done considering whether the responsiveness we find is desirable in a democratic society.

³¹See Figure A.7 in the appendix for the data.

References

- Alt, James, Ethan Bueno de Mesquita, and Shanna Rose. 2011. “Disentangling Accountability and Competence in Elections: Evidence from US Term Limits.” *The Journal of Politics* 73(1): 171–186.
- Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- Anzia, Sarah F. 2011. “Election Timing and the Electoral Influence of Interest Groups.” *The Journal of Politics* 73(2): 412–427.
- Arnold, R Douglas, and Nicholas Carnes. 2012. “Holding Mayors Accountable: New York’s Executives from Koch to Bloomberg.” *American Journal of Political Science* 56(4): 949–963.
- Ashworth, Scott. 2012. “Electoral Accountability: Recent Theoretical and Empirical Work.” *Annual Review of Political Science* 15: 183–201.
- Ashworth, Scott, and Ethan Bueno de Mesquita. 2017. “Unified Versus Divided Political Authority.” *The Journal of Politics* 79(4): 1372–1385.
- Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. “Generalized random forests.” *The Annals of Statistics* 47(2): 1148 – 1178.
- Austen, Liz. 2016. “Police and Crime Commissioners: Emerging Drug Policy Actors?” *Safer Communities* 15(1): 4–10.
- Bayley, David H. 1990. *Patterns of Policing: A Comparative International Analysis*. Rutgers University Press.
- Berry, Christopher R, and Jacob E Gersen. 2009. “Fiscal Consequences of Electoral Institutions.” *The Journal of Law and Economics* 52(3): 469–495.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. “How Much Should We Trust Differences-in-Differences Estimates?” *The Quarterly Journal of Economics* 119(1): 249–275.
- Besley, Timothy, and Anne Case. 2003. “Political Institutions and Policy Choices: Evidence From the United States.” *Journal of Economic Literature* 41(1): 7–73.
- Besley, Timothy, and Stephen Coate. 2003. “Elected Versus Appointed Regulators: Theory and Evidence.” *Journal of the European Economic Association* 1(5): 1176–1206.
- Brehm, John O, and Scott Gates. 1999. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. University of Michigan Press.
- Caless, Bryn, and Jane Owens. 2016. *Police and Crime Commissioners: The Transformation of Police Accountability*. Policy Press.

- Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2008. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *The Review of Economics and Statistics* 90(3): 414–427.
- Canes-Wrone, Brandice, Michael C Herron, and Kenneth W Shotts. 2001. “Leadership and Pandering: A Theory of Executive Policymaking.” *American Journal of Political Science* pp. 532–550.
- Card, David. 1992. “Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage.” *ILR Review* 46(1): 22–37.
- Caughey, Devin, and Christopher Warshaw. 2018. “Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014.” *American Political Science Review* 112(2): 249–266.
- Christensen, Darin, and Simon Ejdemyr. 2018. “Do Elections Improve Constituency Responsiveness? Evidence from US Cities.” *Political Science Research and Methods* X: 1–18.
- Crawford, Adam. 2013. “The Police, Policing and the Future of the ‘Extended Policing Family’.” *The Future of Policing* pp. 173–190.
- Devi, Tanaya, and Roland G Fryer Jr. 2020. “Policing the Police: The Impact of” Pattern-or-Practice” Investigations on Crime.” Working Paper. <https://www.nber.org/papers/w27324>.
- Dipoppa, Gemma, and Guy Grossman. 2020. “The Effect of Election Proximity on Government Responsiveness and Citizens’ Participation: Evidence From English Local Elections.” 53(14): 2183–2212.
- Erikson, Robert S, Gerald C Wright, Gerald C Wright, and John P McIver. 1993. *Statehouse Democracy: Public Opinion and Policy in the American States*. Cambridge University Press.
- Facchini, Giovanni, Brian G Knight, and Cecilia Testa. 2020. “The Franchise, Policing, and Race: Evidence from Arrests Data and the Voting Rights Act.” Working Paper. <https://www.nber.org/papers/w27463>.
- Fearon, James D. 1999. “Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance.” In *Democracy, Accountability, and Representation*, ed. Bernard Mann, Adam Przeworski, and Susan C. Stokes. Cambridge University Press pp. 55–97.
- Feigenbaum, James J, and Andrew B Hall. 2015. “How Legislators Respond to Localized Economic Shocks: Evidence from Chinese Import Competition.” *The Journal of Politics* 77(4): 1012–1030.
- Ferraz, Claudio, and Frederico Finan. 2011. “Electoral Accountability and Corruption: Evidence from the Audits of Local Governments.” *American Economic Review* 101(4): 1274–1311.

- Ferreira, Fernando, and Joseph Gyourko. 2009. "Do Political Parties Matter? Evidence from US Cities." *The Quarterly Journal of Economics* 124(1): 399–422.
- Fetzer, Thiemo. 2019. "Did austerity cause Brexit?" *American Economic Review* 109(11): 3849–86.
- Fourirnaies, Alexander, and Andrew B Hall. 2018. "How Do Electoral Incentives Affect Legislator Behavior?" Working Paper. http://www.andrewbenjaminhall.com/Fourirnaies_Hall_Electoral_Incentives.pdf.
- Gailmard, Sean, and Jeffery A Jenkins. 2009. "Agency Problems, the 17th Amendment, and Representation in the Senate." *American Journal of Political Science* 53(2): 324–342.
- Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti. 2019. "Internet and Politics: Evidence from UK Local Elections and Local Government Policies." *The Review of Economic Studies* 86(5): 2092–2135.
- Gordon, Sanford C, and Gregory Huber. 2007. "The Effect of Electoral Competitiveness on Incumbent Behavior." *Quarterly Journal of Political Science* 2(2): 107–138.
- Grossman, Guy. 2014. "Do Selection Rules Affect Leader Responsiveness? Evidence from Rural Uganda." *Quarterly Journal of Political Science* 9(1): 1–44.
- Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27(2): 163–192.
- Harvey, Anna, and Taylor Mattia. 2019. "Reducing racial disparities in crime victimization." Working Paper. https://s18798.pcdn.co/annaharvey/wp-content/uploads/sites/6417/2020/07/Victimization_Revised.pdf.
- Hazlett, Chad, and Yiqing Xu. 2018. "Trajectory Balancing: A general Reweighting Approach to Causal Inference with Time-Series Cross-Sectional Data." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3214231.
- Hopkins, Daniel J. 2018. *The Increasingly United States: How and Why American Political Behavior Nationalized*. University of Chicago Press.
- Hu, Shengkuo, and Courtenay R Conrad. 2020. "Monitoring via the Courts: Judicial Oversight and Police Violence in India." *International Studies Quarterly* 64(3): 699–709.
- Huber, Gregory A, and Sanford C Gordon. 2004. "Accountability and Coercion: Is Justice Blind when It Runs for Office?" *American Journal of Political Science* 48(2): 247–263.
- Imai, Kosuke, Gary King, Clayton Nall et al. 2009. "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science* 24(1): 29–53.

- Imai, Kosuke, In Song Kim, and Erik Wang. 2018. “Matching methods for causal inference with time-series cross-section data.” Working Paper. <https://imai.fas.harvard.edu/research/tscs.html>.
- Jones, Trevor, Tim Newburn, and D Smith. 2012. “Democracy and Police and Crime Commissioners.” *Policing: Politics, Culture and Control*, Hart Publishing, Oxford pp. 219–244.
- Kirkland, Christopher. 2015. “PCC Elections as a ‘Failed Experiment’: What Lessons Can Be Learned?” 86(3): 403–410.
- Knox, Dean, Will Lowe, and Jonathan Mummolo. 2020. “Administrative Records Mask Racially Biased Policing.” *American Political Science Review* 114(3): 619–637.
- Lim, Claire SH. 2013. “Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges.” *American Economic Review* 103(4): 1360–97.
- Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. Russell Sage Foundation.
- Lister, Stuart, and Michael Rowe. 2015. “Electing Police and Crime Commissioners in England and Wales: Prospecting for the Democratisation of Policing.” *Policing and Society* 25(4): 358–377.
- Lueders, Hans, Jens Hainmueller, and Duncan Lawrence. 2017. “Providing Driver’s Licenses to Unauthorized Immigrants in California Improves Traffic Safety.” *Proceedings of the National Academy of Sciences* 114(16): 4111–4116.
- McCubbins, Mathew D, and Thomas Schwartz. 1984. “Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms.” *American Journal of Political Science* pp. 165–179.
- Meltzer, Allan H, and Scott F Richard. 1981. “A Rational Theory of the Size of Government.” *Journal of Political Economy* 89(5): 914–927.
- Montgomery, Jacob M, and Santiago Olivella. 2018. “Tree-Based Models for Political Science Data.” *American Journal of Political Science* 62(3): 729–744.
- Moskowitz, Daniel J. 2018. “Local News, Information, and the Nationalization of US Elections.” Working Paper. <https://bit.ly/2W15rjp>.
- Mummolo, Jonathan. 2018. “Modern Police Tactics, Police-Citizen Interactions, and the Prospects for Reform.” *The Journal of Politics* 80(1): 1–15.
- Ofer, Udi. 2015. “Getting it Right: Building Effective Civilian Review Boards to Oversee Police.” *Seton Hall L. Rev.* 46: 1033.
- Oforu, George Kwaku. 2019. “Do Fairer Elections Increase the Responsiveness of Politicians?” *American Political Science Review* 113(November): 963–979.

- Payson, Julia A. 2017. "When Are Local Incumbents Held Accountable for Government performance? Evidence from US School Districts." *Legislative Studies Quarterly* 42(3): 421–448.
- Raine, John W, and Paul Keasey. 2012. "From Police Authorities to Police and Crime Commissioners: Might Policing Become More Publicly Accountable?" *International Journal of Emergency Services* 1(2): 122–134.
- Rogers, Steven. 2017. "Electoral Accountability for State Legislative Roll Calls and Ideological Representation." *American Political Science Review* 111(3): 555–571.
- Sances, Michael W. 2016. "The Distributional Impact of Greater Responsiveness: Evidence from New York Towns." *The Journal of Politics* 78(1): 105–119.
- Sandford, Mark. 2021. "Council Tax: Local Referendums." (05682).
- Sen, Sankar. 2010. *Enforcing Police Accountability Through Civilian Oversight*. SAGE Publications India.
- Simonovits, Gabor, Andrew M Guess, and Jonathan Nagler. 2019. "Responsiveness without Representation: Evidence from Minimum Wage Laws in US States." *American Journal of Political Science* 63(2): 401–410.
- Snyder Jr, James M, and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118(2): 355–408.
- Stimson, James A, Michael B MacKuen, and Robert S Erikson. 1995. "Dynamic Representation." *American Political Science Review* 89(3): 543–565.
- Tausanovitch, Chris, and Christopher Warshaw. 2014. "Representation in Municipal Government." *American Political Science Review* 108(3): 605–641.
- Thompson, Daniel M. 2020. "How Partisan Is Local Law Enforcement? Evidence from Sheriff Cooperation With Immigration Authorities." *American Political Science Review* 114(1): 222–236.
- White, Adam. 2015. "The Politics of Police "Privatization": A Multiple Streams Approach." *Criminology & Criminal Justice* 15(3): 283–299.
- Wilson, James Q. 1978. *Varieties of Police Behavior*. Harvard University Press.

Online Appendix

Intended for online publication only.

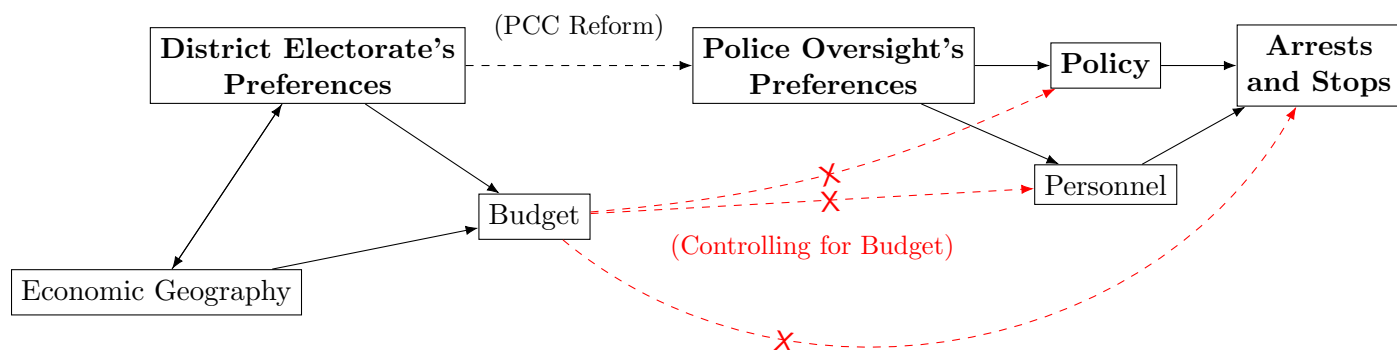
Contents

A.1	Connecting the Theory and Case Data	38
A.2	Additional Arrest-Related Results	40
A.2.1	Arrest Results with Binned Treatment Variable	40
A.2.2	Differential Effects on Arrest Shares by Crime Type	42
A.2.3	Drug Arrests and Drug Stops Per Capita	43
A.3	Prodding Robustness of Differential Effects of Reform	44
A.3.1	Left- and Right-Leaning Police Forces on Similar Trends Before PCCs Introduced	44
A.3.2	Matching Procedure and Results	50
A.3.3	Differential Effects Not Driven by Austerity	54
A.3.4	Differential Effects Not Driven by Country or Urban-Rural Trends .	57
A.3.5	Differential Effects Similar Across Alternative Measures of Preferences	60
A.4	Descriptive Results	66
A.4.1	Map of Parliamentary Voting by Police Force Area	66
A.4.2	Cannabis Interest Group Endorsements	67
A.4.3	Partisan Control of Police Forces Before And After Reform	68
A.4.4	Prevalence of Police Stops, By Ethnicity	69

A.1 Connecting the Theory and Case Data

In this section, we discuss how the case of elected police and crime commissioners fits our theoretical framework, and explain the hypothesized mechanisms in detail.

Figure A.1: **Police Force Responsiveness to District Preferences: Illustration of Causal Mechanisms.**



In Figure A.1, we illustrate the theoretical mechanism in graphical form. Police oversight bodies implement their preferences over policy and personnel decisions, both of which affect subsequent police behavior. The introduction of police and crime commissioners in 2012 strengthens the link between district ideology and police oversight's preferences. We expect that, as the result of direct elections, police oversight's preferences become more closely aligned with the district electorate, as do policy outcomes and police behaviors at the end of the mechanism chain. Our main task in this paper is therefore to rigorously examine whether police behavior has become more congruent with voters' preferences after the introduction of elected police and crime commissioners. While we would, ideally, also measure the intermediate outcomes along the mechanism's path (e.g., implemented policy), data availability and operationalization concerns prevent us from doing so quantitatively.

One notable aspect of our case is that police forces in England and Wales have reduced control over their revenue streams: a majority of their funding comes from the central government. Local revenue, which, on average, constitutes about a third of funding, is raised in the form of a precept on a local property tax. Police forces' ability to increase this

precept is limited, as the central government incentivized local governments to freeze their precept levels in response for additional central government grants. Moreover, any ‘excessive’ raise in the precept would either be restricted by the central government, or, in England since 2011, trigger a local referendum to approve it.³² Increases in the precept are therefore, at best, keeping in line with inflation. Consequently, police forces have very limited means of adjusting their revenue to meet district voters’ preferences.

A potential concern is that police forces’ dependence on central government grants may confound the mechanism running from district preferences to policy outcomes. Against the backdrop of the 2008-09 Great Recession, the central government implemented harsh cuts to public funding of local government units starting in 2010-11. Between 2011 and 2015, central government funding to police forces decreased, on average, by 25% in real terms.³³ These cuts were not distributed uniformly across the country: cities and poorer, left-leaning areas were affected most severely (Fetzer 2019). Any change in outcomes that we attribute to increased responsiveness might instead be an impact of disproportionately severe austerity measures. This concern is illustrated by the bottom part of Figure A.1, where economic geography not only affects voters’ preferences, but also police forces’ budgets, and through it downstream policy outcomes.

It follows that any endeavor to estimate whether the PCC reform increased police forces’ responsiveness to the electorate’s preferences will have to control for this confounding path. We do so throughout our analyses and discuss our strategy for doing so in the main text..

³²Typically, the threshold above which a referendum is required, was set at a 2% or 3% increase of the nominal precept amount. In practice, the threat of a referendum means that police forces (and local governments) do not attempt local tax increases beyond this threshold. Only one police force, Bedfordshire in 2015, attempted to raise local precepts by 15.8%, which triggered a referendum (Sandford 2021). The proposal was defeated by a two-thirds majority of voters. Note that while the referendum requirement does not apply to the four police forces in Wales, the devolved Welsh government still steps in if police forces were to hike up local tax rates. As in England, a majority of Welsh police forces’ funding comes from the central (Westminster) government.

³³<https://www.nao.org.uk/report/financial-sustainability-of-police-forces-in-england-and-wales/>

A.2 Additional Arrest-Related Results

A.2.1 Arrest Results with Binned Treatment Variable

For the main analysis, we present results using a continuous measure of local preferences. This provides a simple summary statistic and is powerful test if the effect of switching to elected police oversight is approximately linear in local preferences. But, if the effect is non-linear, this linear estimator can produce substantively misleading results.

Following Hainmueller, Mummolo, and Xu (2019), we have binned the treatment variable into three equally-sized bins as a robustness check. The results, reported in Table A.1, are in line with the main results using a linear estimator. As we discussed in the draft, the post-reform drop in drug arrests is biggest in the left-most districts, then the middle districts, and finally in the right-most district.

Table A.1: **Elected Police Commissioners Caused Police Forces in Left-Leaning Districts to Make Fewer Arrests per Capita and Reduce Drug Arrest Share.**

	Arrests Per Capita			Drug Arrest Share [0,1]		
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Mid Left Share	-0.000 (0.001)	-0.002 (0.002)	-0.000 (0.001)	-0.005 (0.005)	0.000 (0.008)	0.001 (0.006)
Post \times High Left Share	-0.002 (0.001)	-0.003 (0.002)	-0.002 (0.002)	-0.021 (0.007)	-0.015 (0.009)	-0.007 (0.007)
Mean	0.018	0.018	0.018	0.082	0.082	0.082
# Forces	41	40	40	41	40	40
# Years	11	11	11	11	11	11
# Obs	448	437	437	448	437	437
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	No	No	Yes	No	No
Year \times Revenue Decile FEs	No	Yes	No	No	Yes	No
Year \times Revenue IFEs	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Mid Left Share is a binary variable with value one if the police force's total vote share for Labour, LibDems, and Green candidates in the 2010 parliamentary election fell into the middle tercile. High Left Share is a binary variable with value one if the police force's total vote share for Labour, LibDems, and Green candidates in the 2010 parliamentary election fell into the top tercile. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2019.

A.2.2 Differential Effects on Arrest Shares by Crime Type

In the table below, we report the differential effect of the introduction of elected police and crime commissioners on shares of different arrest categories in addition to drug arrests. Here, we have to limit our analysis to a much shorter post-reform period because many non-drug offenses changed categories between 2015 and 2016, making overtime comparisons outside of drug arrests less clear. We find that the drop in drug arrests share was offset by increases in burglary and theft arrests as well as a category labeled “other” that is primarily disorder-related crimes, often near or inside bars.

Because we did not have any expectations *ex ante* as to how arrest priorities might shift beyond drug enforcement, especially over a very short timespan, we interpret these estimates with caution.

Table A.2: **Elected Police Commissioners Caused Police Forces in Left-Leaning Districts to Reduce Drug Arrest Share, Increase Disorder-Related Arrest Share (Other), 2008 to 2015.**

	Arrest Share [0,1]								
	Drug (1)	Violent (2)	Sex (3)	Robbery (4)	Burglary (5)	Theft (6)	Damage (7)	Fraud (8)	Other (9)
Post × Left Share	-0.049 (0.041)	-0.064 (0.101)	-0.013 (0.011)	0.009 (0.007)	0.038 (0.020)	0.047 (0.040)	-0.027 (0.031)	-0.004 (0.019)	0.063 (0.116)
Mean	0.082	0.336	0.030	0.018	0.071	0.220	0.090	0.019	0.134
# Forces	40	40	40	40	40	40	40	40	40
# Years	7	7	7	7	7	7	7	7	7
# Obs	280	280	280	280	280	280	280	280	280
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Revenue IFEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Year × Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2015.

A.2.3 Drug Arrests and Drug Stops Per Capita

As we discuss in the body of the paper, one concern with our analysis on drug arrests and stop shares is that changes in priorities around other arrest and stop types might influence the drug arrest and stop outcomes. We address this by estimating the differential effects on drug arrests and stops per capita. We find a similar pattern of results. In some cases these estimates are not statistically different from zero. We prefer to study the effect on shares because, if a police budget is exogenous, police leadership have much more control of priorities rather than arrests per capita. This may explain in part why the estimated differential effects are somewhat less precise. Still, we take this evidence as consistent with our main findings.

Table A.3: **Elected Police Commissioners Caused Police Forces in Left-Leaning Districts to Make Fewer Drug Arrests and Drug Stops per Capita.**

	Drug Arrests per 100k Pop			Drug Stops per 100k Pop		
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Left Share	-0.008 (0.003)	-0.009 (0.004)	-0.005 (0.004)	-0.028 (0.016)	-0.039 (0.027)	-0.025 (0.022)
Mean	0.008	0.008	0.008	0.054	0.054	0.054
# Forces	41	40	40	41	40	40
# Years	11	11	11	10	10	10
# Obs	448	437	437	410	400	400
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	No	No	Yes	No	No
Year \times Revenue Decile FEs	No	Yes	No	No	Yes	No
Year \times Revenue IFEs	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2019 (2018 for stops).

A.3 Prodding Robustness of Differential Effects of Reform

A.3.1 Left- and Right-Leaning Police Forces on Similar Trends Before PCCs Introduced

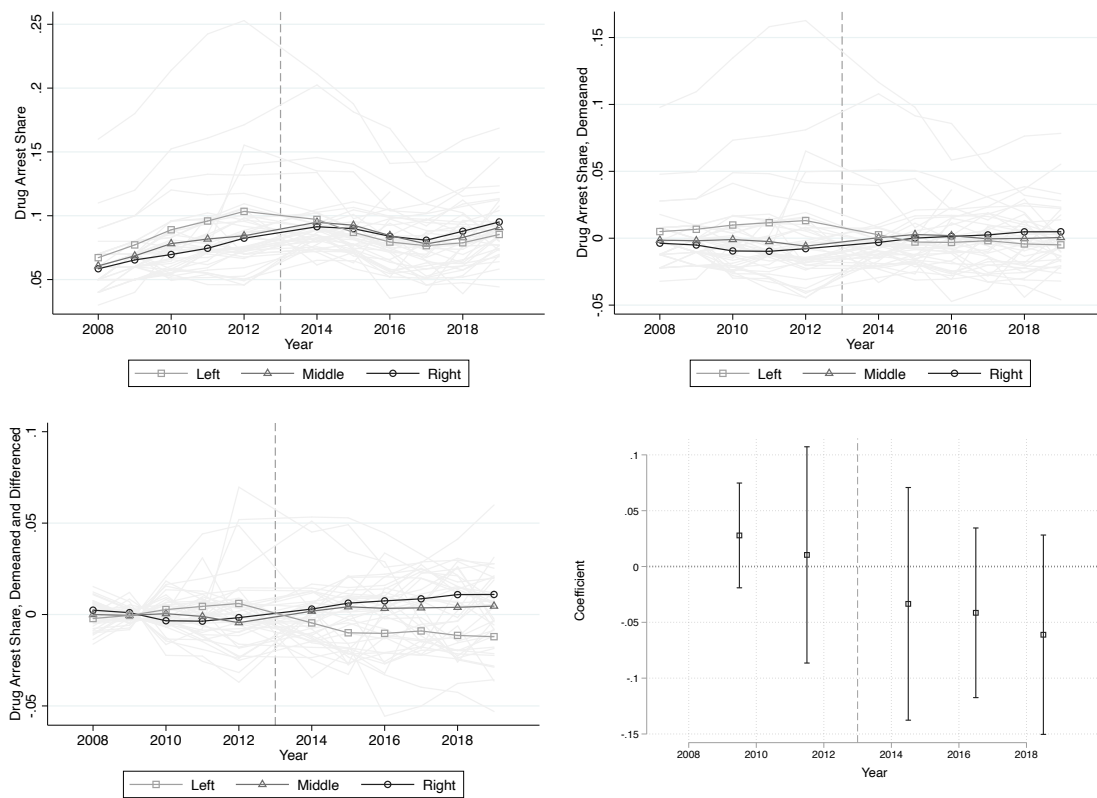
As we discuss in the main text of the paper, one important threat to the validity of our analysis is that left- and right-leaning districts may be already trending apart in terms of arrests and personnel prior to the introduction of elected police and crime commissioners. We assess this possibility visually and formally finding that left- and right-leaning districts are on approximately the same trends prior to 2013. Unfortunately, as we discuss in the main text, all of these analyses are imprecise and only suggestive.

In Figure A.2 we present the average drug arrest share for police forces grouped into terciles by the share of votes for left parties. Starting with the raw data in the top left quadrant and ending with regression coefficients in the bottom right quadrant, we present visual evidence that the change in the relationship between drug arrest share and support for left parties in the 2010 parliamentary election begins after the introduction of police and crime commissioners in 2013 rather than before.

In Figure A.3 we present plots that mirror Figure A.2 but using the average number of arrests per capita as the outcome. As with drug arrests, we find that total arrests per capita on similar trends in left- and right-leaning districts prior to the adoption of elected police and crime commissioners. This is hard to see in the top left quadrant because of the swift decline in arrests across all police forces. The other three quadrants make clearer that left- and right-leaning districts are on the approximately the same trajectories prior to 2013 and split apart following the introduction of PCCs.

In Table A.4 we formally test for trending in our preferred regression specification. We find a small, imprecisely estimated increase in arrests per capita before 2013 in left-leaning

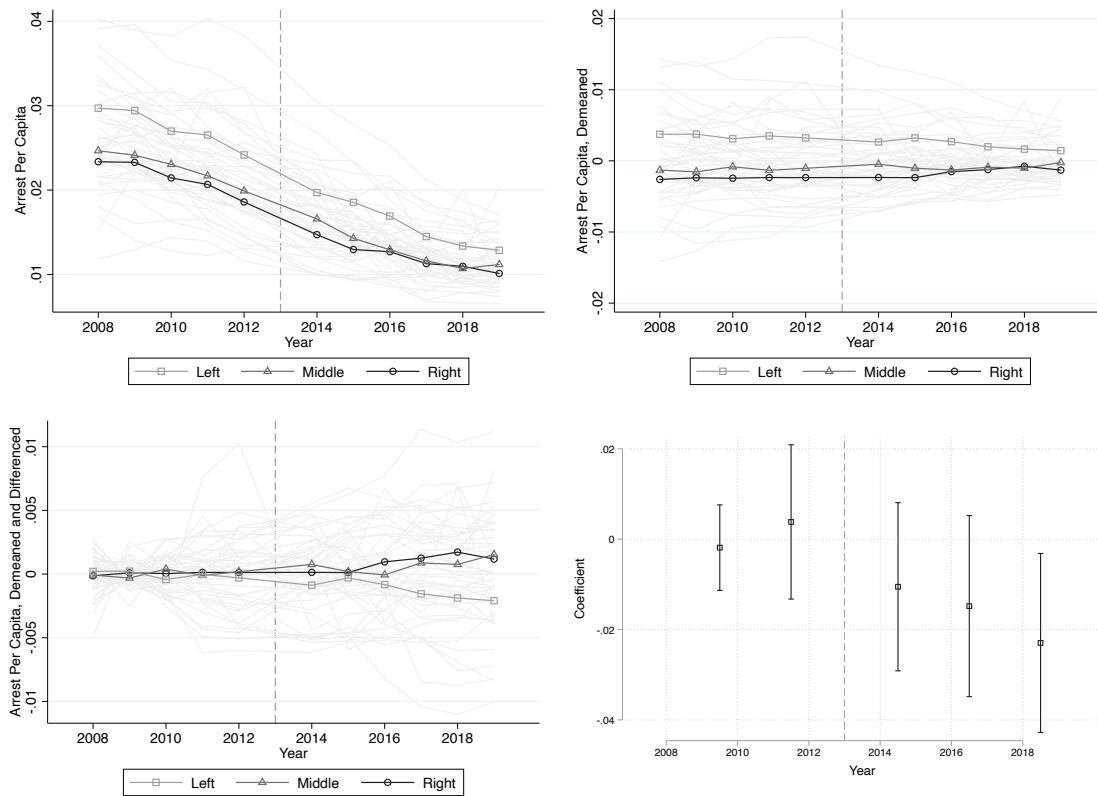
Figure A.2: **Drug Arrest Share on Similar Trend in Left- and Right-Leaning Districts Before PCCs Introduced, Split After PCCs Introduced..** The average drug arrest shares from each tercile of left-party parliamentary voting are on the same trend prior to 2013, the introduction of police and crime commissioners. The left-most tercile decreases drug arrests faster than the middle and right terciles following the introduction of PCCs. Top left quadrant: raw drug arrests shares by year for every police force and the average by left-party vote share tercile. Top right quadrant: subtracting the average drug arrest share by year. Bottom left quadrant: subtracting average drug arrest share by year then subtracting the average time-demeaned drug arrest share from 2008 to 2010. Bottom right quadrant: coefficients from drug arrest on left vote share interacted with two-year period dummies and including police force and year-by-2012-revenue interacted fixed effects.



places relative to right-leaning places. We also find a small, imprecisely estimated decrease in drug arrest share prior to 2013.

In Figure A.4 we present plots that mirror Figures A.2 and A.3 but using the staff share of police department employees as the outcome. As with drug arrests and arrests per capita, we find that staff shares are on similar trends in left- and right-leaning districts prior to the adoption of elected police and crime commissioners. This is again hard to see in the top left

Figure A.3: **Arrests per Capita on Similar Trend in Left- and Right-Leaning Districts Before PCCs Introduced, Split After PCCs Introduced.** The average arrests per capita from each tercile of left-party parliamentary voting are on the same trend prior to 2013, the introduction of police and crime commissioners. The right-most tercile increases arrests per capita faster than the left tercile following the introduction of PCCs. Top left quadrant: raw arrests per capita by year for every police force and the average by left-party vote share tercile. Top right quadrant: subtracting the average arrests per capita by year. Bottom left quadrant: subtracting average arrests per capita by year then subtracting the average time-demeaned arrests per capita from 2008 to 2010. Bottom right quadrant: coefficients from arrests per capita on left vote share interacted with two-year period dummies and including police force and year-by-2012-revenue interacted fixed effects.



quadrant, but the other three quadrants make clearer that the staff share of employment drops faster in right-leaning districts after 2013.

In Table A.5, we present the formal estimates of trending on our personnel outcomes. We find some imprecise evidence of trending in terms of total employees. Because this trending is nearly almost the magnitude of the estimated differential effect on total police department employment, we interpret this as consistent with there being no differential effect

Table A.4: Arrests and Drug Arrests on Similar Trends in Left and Right Districts Before Reform.

	Arrests Per Capita			Drug Arrest Share [0,1]		
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Left Share	-0.017 (0.007)	-0.015 (0.008)	-0.015 (0.008)	-0.061 (0.034)	-0.064 (0.037)	-0.064 (0.037)
2011/2012 \times Left Share		0.005 (0.006)			-0.008 (0.042)	
2011 \times Left Share			0.004 (0.007)			-0.025 (0.068)
2012 \times Left Share			0.006 (0.006)			0.009 (0.026)
Mean	0.018	0.018	0.018	0.082	0.082	0.082
# Forces	40	40	40	40	40	40
# Years	11	11	11	11	11	11
# Obs	437	437	437	437	437	437
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year \times Revenue IFEs	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after fiscal year 2013. 2011/12 is a binary variable taking the value one for the fiscal years ending in 2011 or 2012. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2019.

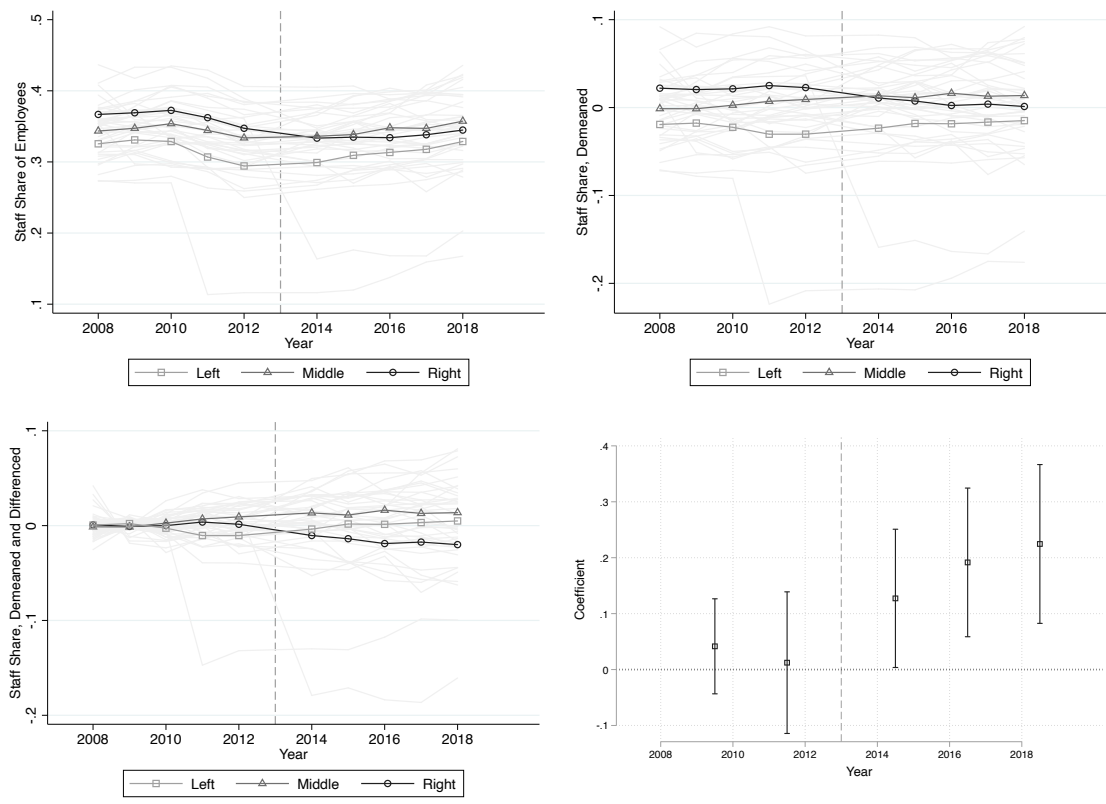
on employees post-reform. We do not find meaningful evidence of trending in terms of officer or staff share.

Table A.5: **Staff and Officer Share on Similar Trends in Left and Right Places Before Reform, While Total Employment Declines in Left-Leaning Places in 2012.**

	Employees Per 1k Pop		Officer Share		Staff Share	
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Left Share	-0.909 (0.572)	-0.909 (0.572)	-0.151 (0.080)	-0.151 (0.080)	0.145 (0.063)	0.145 (0.063)
2011/2012 \times Left Share	-0.638 (0.342)		0.015 (0.038)		-0.015 (0.044)	
2011 \times Left Share		-0.444 (0.328)		0.023 (0.039)		-0.025 (0.045)
2012 \times Left Share		-0.833 (0.401)		0.006 (0.043)		-0.006 (0.047)
Mean	3.468	3.468	0.578	0.578	0.337	0.337
# Forces	40	40	40	40	40	40
# Years	10	10	10	10	10	10
# Obs	400	400	400	400	400	400
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year \times Revenue IFEs	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after fiscal year 2013. 2011/12 is a binary variable taking the value one for the fiscal years ending in 2011 or 2012. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2019.

Figure A.4: **Staff Share of Employment on Similar Trend in Left- and Right-Leaning Districts Before PCCs Introduced, Split After PCCs Introduced.** The staff share of police department employees from each tercile of left-party parliamentary voting are on the same trend prior to 2013, the introduction of police and crime commissioners. The right-most tercile decreases staff share faster than the left tercile following the introduction of PCCs. Top left quadrant: staff share of police employees by year for every police force and the average by left-party vote share tercile. Top right quadrant: subtracting the average staff share by year. Bottom left quadrant: subtracting average staff share by year then subtracting the average time-demeaned staff share from 2008 to 2010. Bottom right quadrant: coefficients from staff share on left vote share interacted with two-year period dummies and including police force and year-by-2012-revenue interacted fixed effects.



A.3.2 Matching Procedure and Results

A.3.2.1 Description of the Matching Procedure

The matching procedure we describe in Section 4.2 proceeds in three steps:

1. *Calculate the distance between each unit and every other unit.* We calculate this distance by calculating the difference between the two units in every year between 2006 and 2012, squaring these differences, and adding them. This yields a single distance measure between units, in this case the Euclidean distance. In a separate version, we add a weight to the squared differences in which 2006 received a weight of $1/7$, 2012 receives a weight of 1, and the weight increases linearly with year over the intervening period. This penalizes distances closer to the policy change more.
2. *Drop the unit with the largest minimum distance it and any other unit.* Since there are an odd number of units, we drop the unit with the worst ability to match from the sample.
3. *Run nonbipartite matching algorithm to find pairs that minimize the total of within-pair distances.* Following the recommendation of described in Ryan T Moore and Keith Schnakenberg’s discussion of blockTools, we use the nbpMatching package to find the pairs that minimize this total distance.

One technical note about estimation: throughout the paper we use block bootstrapping to estimate our standard errors following best practice. Because best practice is not yet established for matching estimators, the bootstrap is not appropriate for matching estimators, and we cannot devise a feasible bootstrap procedure even if the bootstrap were appropriate, we report standard cluster-robust standard errors throughout this section.

A.3.2.2 Formal Matched Analysis Results

Tables A.6, A.7, and A.8 present the formal results from our matching analysis. Across all three tables, we find differential effects on all three of the outcomes for which we made ex ante predictions—drug arrest shares, drug stop shares, and staff shares—that are in line with our expectations. On the other hand, the differential effects on arrests, stops, and employment per capita switch signs but are statistically indistinguishable from zero. The first row presents estimates analogous to those in the first row of Table 1. The second row removes the linearity assumption and estimates the average change in the drug arrest share gap between the relatively left-wing force and the relatively right-wing force within each pair. Looking at the fourth row, we can compare the estimates in the first and second row. For concreteness, we will focus on Table A.6. The average distance between the relatively left-wing pair member and the relatively right-wing member is 8.4 percentage points in terms of left party voting. If we multiply that by the changed gap of 0.044, we get an expected average within-pair difference of about 0.0037. This suggests that, while the linear estimator increases our power, the substantive interpretation of the effects is similar across both approaches. We see a similar pattern of results across the other two tables.

Table A.6: Elected Police Commissioners Caused a Decrease in the Share of Arrests for Drugs in Left-Leaning Districts, Comparisons Within Matched Pairs.

	Arrests Per Capita				Drug Arrest Share [0,1]			
Post * Left Share [0,1]	0.007 (0.004)	0.007 (0.004)			-0.060 (0.021)	-0.044 (0.017)		
Post * Most Left in Pair {0,1}			0.0009 (0.0005)	0.0010 (0.0006)			-0.0060 (0.0034)	-0.0034 (0.0033)
Outcome Mean	0.018	0.018	0.018	0.018	0.080	0.080	0.080	0.080
Avg Diff btwn Left and Right in Pair	0.082	0.084	0.082	0.084	0.082	0.084	0.082	0.084
# Forces	40	40	40	40	40	40	40	40
# Years	11	11	11	11	11	11	11	11
# Obs	434	434	434	434	434	434	434	434
Force FE	Y	Y	Y	Y	Y	Y	Y	Y
Year × Matched Pair FE	Y	Y	Y	Y	Y	Y	Y	Y
Weighted Matching	N	Y	N	Y	N	Y	N	Y

Cluster robust standard errors clustered by police force in parentheses. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the vote share for Labour, Lib Dem, and green candidates in the 2010 election for parliament. Most Left is a flag taking the value one for each police force area that has a higher value of Left Share within matched pair. Matched pairs are constructed to minimize the total within-pair distance between selected pairs of police forces in terms of drug arrests share from 2008 to 2012. Weighted Matching indicates whether the matching penalizes distances closer to 2012 more than distances earlier in the time series. The first commissioner election year, 2012, is held out of the analysis. Data from 2008 to 2019.

Table A.7: Elected Police Commissioners Caused a Decrease in the Share of Arrests for Drugs in Left-Leaning Districts, Comparisons Within Matched Pairs.

	Stops Per Capita				Drug Stop Share [0,1]			
Post * Left Share [0,1]	0.009 (0.008)	0.008 (0.009)			-0.434 (0.085)	-0.279 (0.101)		
Post * Most Left in Pair {0,1}			0.0006 (0.0009)	0.0002 (0.0009)			-0.0426 (0.0112)	-0.0193 (0.0131)
Outcome Mean	0.010	0.010	0.010	0.010	0.513	0.513	0.513	0.513
Avg Diff btwn Left and Right in Pair	0.093	0.086	0.093	0.086	0.093	0.086	0.093	0.086
# Forces	40	40	40	40	40	40	40	40
# Years	10	10	10	10	10	10	10	10
# Obs	400	400	400	400	400	400	400	400
Force FE	Y	Y	Y	Y	Y	Y	Y	Y
Year × Matched Pair FE	Y	Y	Y	Y	Y	Y	Y	Y
Weighted Matching	N	Y	N	Y	N	Y	N	Y

Cluster robust standard errors clustered by police force in parentheses. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the vote share for Labour, Lib Dem, and green candidates in the 2010 election for parliament. Most Left is a flag taking the value one for each police force area that has a higher value of Left Share within matched pair. Matched pairs are constructed to minimize the total within-pair distance between selected pairs of police forces in terms of drug arrests share from 2008 to 2012. Weighted Matching indicates whether the matching penalizes distances closer to 2012 more than distances earlier in the time series. The first commissioner election year, 2012, is held out of the analysis. Data from 2008 to 2019.

Table A.8: **Elected Police Commissioners Caused a Decrease in the Share of Arrests for Drugs in Left-Leaning Districts, Comparisons Within Matched Pairs.**

	Employees Per 1k Residents				Staff Share [0,1]			
Post * Left Share [0,1]	0.830 (0.703)	0.830 (0.703)			0.074 (0.107)	0.064 (0.084)		
Post * Most Left in Pair {0,1}			0.1027 (0.0583)	0.1027 (0.0583)			0.0121 (0.0129)	0.0170 (0.0123)
Outcome Mean	0.018	0.018	0.018	0.018	0.080	0.080	0.080	0.080
Avg Diff btwn Left and Right in Pair	0.077	0.081	0.077	0.081	0.077	0.081	0.077	0.081
# Forces	40	40	40	40	40	40	40	40
# Years	10	10	10	10	10	10	10	10
# Obs	400	400	400	400	400	400	400	400
Force FE	Y	Y	Y	Y	Y	Y	Y	Y
Year × Matched Pair FE	Y	Y	Y	Y	Y	Y	Y	Y
Weighted Matching	N	Y	N	Y	N	Y	N	Y

Cluster robust standard errors clustered by police force in parentheses. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the vote share for Labour, Lib Dem, and green candidates in the 2010 election for parliament. Most Left is a flag taking the value one for each police force area that has a higher value of Left Share within matched pair. Matched pairs are constructed to minimize the total within-pair distance between selected pairs of police forces in terms of drug arrests share from 2008 to 2012. Weighted Matching indicates whether the matching penalizes distances closer to 2012 more than distances earlier in the time series. The first commissioner election year, 2012, is held out of the analysis. Data from 2008 to 2019.

A.3.3 Differential Effects Not Driven by Austerity

During the period when police and crime commissioners were coming into office for the first time, the UK government was making large budget cuts. Police were not immune. In this section, we follow up on the concerns we discussed in Section A.1 and demonstrate that left and right-leaning districts experienced similar levels of austerity after adjusting for pre-reform revenue. Budget cuts could affect police by reducing their capacity for making more arrests on more serious crimes. Or, perhaps, it could mean that police cut back on the community policing that often results in drug arrests. Either way, if the effects of these cutbacks are not constant across places, they could produce a change in drug arrest shares unrelated to the introduction of police and crime commissioners.³⁴ Because commissioners could not meaningfully change their budgets during this period, we interpret these changes as an exogenous potential time-varying confounder.

Table A.9 reports formal estimates of whether left-leaning police authorities suffered from a greater shortfall of funding after the introduction of elected police and crime commissioners. Columns 1 to 3 report the differential trends in total revenue following the adoption of elected commissioners; columns 4 to 6 estimate differential trends in revenue from the central government; and columns 7 to 9 estimate differential trends in local tax revenue. All three outcomes are measured per 1,000 residents. Across all three outcomes, the specifications are the same as in Table 1. First, we report the estimate with a simple two-way fixed effect. Next, we include revenue-decile-by-year fixed effects. Finally, we swap in revenue-specific yearly slopes.

We find that left- and right-leaning police forces have meaningfully different budgets after the reform, but these differences attenuate to nearly zero when we adjust for pre-reform revenue. In the unadjusted, two-way fixed effects specification (column 1), moving from the 25th percentile to the 75th percentile in terms of left voting share predicts a

³⁴For recent reporting on how these cuts influenced police, see Fewer Officers, More Calls: U.K. Police Are Stretched by Austerity in the New York Times (<https://www.nytimes.com/2019/02/01/world/europe/uk-police-crime-austerity.html>).

roughly 2% decrease in total revenue relative to the mean of £189.9 per 1,000 population. The estimates become statistically indistinguishable from and shrink towards zero when we include either of our two revenue-adjusting fixed effects specifications (columns 2 and 3). The same pattern holds true for central government revenue: here, too, the coefficients decrease in magnitude and become statistically insignificant at conventional levels when including revenue fixed effects. When we use local tax revenue as an outcome, the adjustment appears to be somewhat less effective, although the coefficient is still statistically indistinguishable from zero at conventional levels, even when including year-specific slopes on 2012 revenue.

These results suggest that the inclusion of year-specific slopes on 2012 total revenue is an effective way of adjusting for differential exposure to austerity-related budget cuts. When we include our revenue adjustment in the regression specification, the share of left votes no longer predicts a meaningful difference in a police forces' post-reform revenue. The specification with year-by-revenue decile fixed effects and, better still, year-specific slopes on 2012 revenue is therefore a plausible strategy to account for any post-reform changes in crime outcomes that are the consequence of harsher austerity measures in left-leaning places, rather than increased responsiveness.

Table A.9: **Police Forces in Left and Right Districts Had Similar Budget Cuts Under Elected Police Commissioners After Adjusting for 2012 Revenue.**

	Revenue Per 1k Pop			Central Gov Rev Per 1k Pop			Local Tax Rev Per 1k Pop		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post \times Left Share	-33.7 (15.2)	-16.4 (27.0)	2.3 (19.1)	-25.8 (11.8)	-6.0 (17.0)	10.6 (12.3)	-4.9 (8.1)	-18.8 (14.3)	-16.1 (10.9)
Mean	189.9	189.9	189.9	134.1	134.1	134.1	56.1	56.1	56.1
# Forces	40	40	40	40	40	40	40	40	40
# Years	10	10	10	10	10	10	10	10	10
# Obs	397	397	397	397	397	397	397	397	397
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	No	No	Yes	No	No	Yes	No	No
Year \times Revenue Decile FEs	No	Yes	No	No	Yes	No	No	Yes	No
Year \times Revenue IFEs	No	No	Yes	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2018.

A.3.4 Differential Effects Not Driven by Country or Urban-Rural Trends

A.3.4.1 Arrest Analysis with Population Density Adjustments

One concern with our existing analysis is that left districts are more urban than right districts and that there are some secular changes in urban places that relate to changes in policing, unrelated to reform or responsiveness. We address this in Table A.10 by making comparisons between left and right districts with similar density. Our results are similar to those we find in Table 1, suggesting that urban-rural trending is not responsible for our main results on arrests.

Table A.10: **Main Results Hold When Adjusting for Population Density Instead of Police Force Revenue.**

	Arrests Per Capita			Drug Arrest Share [0,1]		
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Left Share	-0.013 (0.005)	-0.010 (0.007)	-0.013 (0.005)	-0.109 (0.028)	-0.087 (0.034)	-0.083 (0.028)
Mean	0.018	0.018	0.018	0.082	0.082	0.082
# Forces	41	41	41	41	41	41
# Years	11	11	11	11	11	11
# Obs	448	448	448	448	448	448
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	No	No	Yes	No	No
Year \times Pop Density Decile FEs	No	Yes	No	No	Yes	No
Year \times Pop Density IFEs	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Pop Density Decile divides the police forces into deciles based on aggregated population density. Year \times Population Density IFEs are year-specific slopes on police forces' population density. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2019.

A.3.4.2 Arrest Analysis with Separate Trends by Country

Another concern is that the four police forces in Wales (Dyfed-Powys, Gwent, North Wales, South Wales) are on a different trend than English police forces. We replicate our main results with our fixed effects specification interacted with a dummy for observations in Wales, thus modeling diverging trends.

Table A.11 reports the results.

Table A.11: **Results Robust To Modelling Separate Trends For England And Wales.**

	Arrests Per Capita			Drug Arrest Share [0,1]		
	(1)	(2)	(3)	(4)	(5)	(6)
Post \times Left Share	-0.013 (0.005)	-0.021 (0.010)	-0.018 (0.007)	-0.109 (0.028)	-0.111 (0.038)	-0.060 (0.036)
Mean	0.018	0.018	0.018	0.082	0.082	0.082
# Forces	41	39	40	41	39	40
# Years	11	11	11	11	11	11
# Obs	448	426	437	448	426	437
Force FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year \times Country FEs	Yes	No	No	Yes	No	No
Year \times Country \times Revenue Decile FEs	No	Yes	No	No	Yes	No
Year \times Country \times Revenue IFEs	No	No	Yes	No	No	Yes

Standard errors reported in parentheses estimated using 1,000 bootstrap samples blocked by police force. Post is a binary variable taking the value one for years after 2013. Left Share is a share variable falling between zero and one that reports the total vote share for Labour, Lib Dem, and Green candidates in the 2010 election for Parliament in the police force area. Revenue Decile divides the police forces into deciles based on total revenue as of 2012. Country is a binary indicator indicating whether an observation belongs to Wales (rather than England). Year \times Revenue IFEs are year-specific slopes on 2012 total revenue. The first commissioner election year, fiscal year 2013, is held out of the analysis since the treatment starts in the middle of the year. Data from 2008 to 2019.

A.3.5 Differential Effects Similar Across Alternative Measures of Preferences

While our measure of voting for left-leaning parties in the 2010 parliamentary election captures support for parties that endorse more relaxed drug policing, drug policy is only one aspect of party competition, and may be more important at the local level than in a parliamentary race. To address these concerns, we constructed seven additional measures of support for more culturally liberal policies that should correlate with a larger decline in drug arrests under directly elected commissioners if elections increase responsiveness. Among those, we constructed direct measures of drug policy preferences building on respondent-level data from YouGov surveys. We replicate columns 4 through 6 of Table 1 for each of these additional measures and report our preferred measure of parliamentary vote for comparison.

In the remainder of this section, we first describe how we construct our alternative measurements before reporting and discussing the results. Table A.12 reports the summary statistics for our eight different measures, whereas Table A.13 reports the results of our main specification with each of the measures.

A.3.5.1 Measuring Left-Leaningness and Drug Policy Preferences

Vote share measurements. We use left voting shares from three different election contexts. The first row in Table A.13 mirrors the first row of Table 1, using the vote share for left-leaning parties in the 2010 parliamentary election, with column 1 reporting the simple two-way fixed effects estimate and columns 2 and 3 making adjustments to flexibly account for potentially differential trends in drug arrests due to austerity. The second row uses left party vote share in the 2012 commissioner elections as the measure of culturally liberal preferences, drawing on votes for parties on the exact issues relevant in police chief elections. The third row uses left party vote share in all local elections held between 2009 and 2012 in the police force area.

Drug policy preference measurements. We construct measures of voters’ drug policy preferences for the remaining rows. To do so, we draw on respondent-level data from five YouGov surveys, which ask respondents whether they are in favor of decriminalizing or legalizing cannabis. In the fourth row, we use the weighted share of survey respondents in favor of more relaxed drug laws in every police force area. Because the number of respondents in each police force, even when pooled across surveys, is small, we also estimate drug policy preferences using prediction and poststratification (MRP and RFP) approaches, which we discuss in greater detail.

We construct our preference measures by implementing the following general recipe. First, we predict voters’ preferences on more relaxed cannabis laws using a set of either demographic or vote choice variables, as well as the police force and region in which they live. We then post-stratify the results to match either demographic crosstabs as reported by the 2011 census, or the parliamentary election results in 2010.³⁵

In order to predict each individual’s drug policy preferences, we use two approaches. Rows 5 and 6 produced a random forest prediction using demographics and party choice, respectively. We run the algorithm with default starting values and use crossvalidation for tuning.³⁶ The advantage of this approach is that the algorithm may select meaningful interactions between predictors (e.g. age-by-education-by-region) without overfitting (Montgomery and Olivella 2018). Rows 7 and 8, by contrast, use a multilevel regression approach (classic MRP). Unlike the random forest method, we do not include interactions between individual-level variables and geographic variables because we do not know which interactions are reasonable *ex ante*, and including all possible interactions would yield a vastly overdetermined regression specification.

We prefer the random forest-based estimates because they offer a principled way to balance error arising from variance and bias in the prediction stage of the prediction and

³⁵Because there is no reliable data on vote choice by demographics and police force, we are unable to include both sets of predictors in the same poststratification frame.

³⁶See Athey, Tibshirani, and Wager (2019) and their `grf` package in R for implementation details.

post-stratification exercise. This approach also improves on simple disaggregation given the number of observations we have in each police force area.

It is important to note that each of these different methods yields measurements with different variances. These differences have implications for how to interpret effect sizes. Table A.12 lists basic summary statistics for our measures. Note, in particular, that poststratification techniques reduce the variance in the measure — especially when, as in the case of the simplest MRP measures, the set of individual-level variables is only modestly predictive of preferences.

Table A.12: **Summary Statistics of Cultural Liberalism Measurements**

Pref Choice	Summary Statistic				
	Min	Max	Mean	SD	IQR
Left Share (Parliamentary)	0.390	0.751	0.545	0.098	0.127
Left Share (PCC)	0.112	0.660	0.385	0.133	0.180
Left Share (Local)	0.240	0.751	0.500	0.107	0.134
Pref Share (Raw Survey)	0.280	0.605	0.473	0.072	0.098
Pref Share (Demog RFP)	0.403	0.476	0.447	0.017	0.021
Pref Share (Party RFP)	0.378	0.520	0.449	0.031	0.041
Pref Share (Demog MRP)	0.410	0.545	0.483	0.031	0.029
Pref Share (Party MRP)	0.464	0.514	0.480	0.012	0.018

SD stands for standard deviation. IQR stands for interquartile range.

A.3.5.2 Results

Table A.13 reports the results from estimating the differential effect of directly elected commissioners across our measures of culturally liberal preferences. Each row is a separate regression using only one of the eight measures. The first three measures rely on shares of votes cast for left parties in different elections, while the remaining five measures estimate the share of voters in favor of more relax drug laws in each police force.

We find a similar pattern of results, with drops in drug arrests largest in culturally liberal places across all measures of preferences. Comparing the first and second rows, the estimates in the second row are smaller but substantively similar. The distance between the 25th and 75th percentile of left-party vote for commissioner in 2012 is approximately 0.18 and associated with a 0.7 percentage point drop in drug arrest share according to column 3 in the second row. This is approximately the same effect size estimated using the parliamentary left-party vote measure, approximately 0.8 percentage points. This higher interquartile range reflects the fact that multiple 2012 commissioner races featured independent candidates who received substantial vote shares and left-party vote shares varied more across districts. Accordingly, the interquartile range of the 2010 parliamentary left-party vote share (0.127) is smaller than the interquartile range of the commissioner left-party vote measure (0.180).

Comparing the third and first rows of Table A.13, we find quite similar estimates in each. The distribution of left-party voting for local and parliamentary elections across police force areas that the substantive interpretations of these estimates in terms of the interquartile range are quite similar.³⁷

In rows 4 through 8, we present estimates using our more direct measures of drug policy preferences. The estimates vary considerably across measures—in some cases implying relatively modest differential effects, while in others implying very large differential effects. Still, they are broadly consistent with our main finding that electing police oversight causes a larger drop in drug arrests in culturally liberal districts.

Among our direct preference measures, we prefer our random forest-based measures. As we discussed above, we use cross-validation to select the random forest tuning parameters in the prediction step minimizing prediction error and removing the choice of which interactions to include from our hands. According to these estimates, drug policing would decline by 2 or 3 percentage points more in a district with 60% support of relaxed drug laws relative to a district with 40% support of relaxed drug laws.

³⁷The interquartile range of left party vote share for local elections and the 2010 parliamentary election are both approximately 0.13.

Most of the variation across rows 4 through 8 arises from differences in the variance of the preference measures. The raw survey data is the noisiest measure of police-force-level preferences because we have relatively few observations in each police force. The differential effect estimates are, accordingly, relatively small with column 3 implying that the drop in drug arrests associated with a 20 percentage point difference in drug policy preferences is less than one percentage point. On the opposite end is our party-based MRP estimate which has a much smaller variance. There we find differential effects that are quite large in our preferred specification—a 20 percentage point difference in drug policy preferences is associated with a 8 percentage-point difference in the drop in drug arrest share. Because we predict preferences based only on the party each respondent supported, this method may understate differences in drug policy preferences across places if supporters of the same party hold different views across the UK.

Reviewing Table A.13 in total, we find evidence for increased responsiveness across many measures of preferences following the introduction of direct elections. While our more direct preference-based estimates are noisy, we read the full table as suggesting a relationship between cultural liberalism and the decline in drug preferences after adopting direct elections.

Table A.13: **Differential Effects on Drug Arrest Share Similar in Effect Size Across Multiple Measures of Preferences.**

Measurement Choice	Drug Arrest Share [0,1]		
	(1)	(2)	(3)
Post × Left Share (Parliamentary)	-0.109 (0.028) [-0.014]	-0.091 (0.038) [-0.012]	-0.061 (0.034) [-0.008]
Post × Left Share (PCC)	-0.074 (0.022) [-0.013]	-0.058 (0.025) [-0.010]	-0.040 (0.024) [-0.007]
Post × Left Share (Local)	-0.098 (0.027) [-0.013]	-0.069 (0.029) [-0.009]	-0.059 (0.026) [-0.008]
Post × Pref Share (Raw Survey)	-0.012 (0.033) [-0.001]	-0.029 (0.035) [-0.003]	-0.036 (0.026) [-0.003]
Post × Pref Share (Demog RFP)	-0.035 (0.159) [-0.001]	-0.136 (0.179) [-0.003]	-0.176 (0.133) [-0.003]
Post × Pref Share (Party RFP)	-0.067 (0.083) [-0.003]	-0.121 (0.084) [-0.005]	-0.127 (0.070) [-0.005]
Post × Pref Share (Demog MRP)	-0.009 (0.093) [-0.000]	-0.082 (0.118) [-0.002]	-0.110 (0.066) [-0.003]
Post × Pref Share (Party MRP)	-0.783 (0.233) [-0.014]	-0.613 (0.293) [-0.011]	-0.408 (0.247) [-0.007]
Mean	0.082	0.082	0.082
# Forces	41	40	40
# Years	11	11	11
# Obs	448	437	437
Force FEs	Yes	Yes	Yes
Year FEs	Yes	No	No
Year × Revenue Decile FEs	No	Yes	No
Year × Revenue IFEs	No	No	Yes

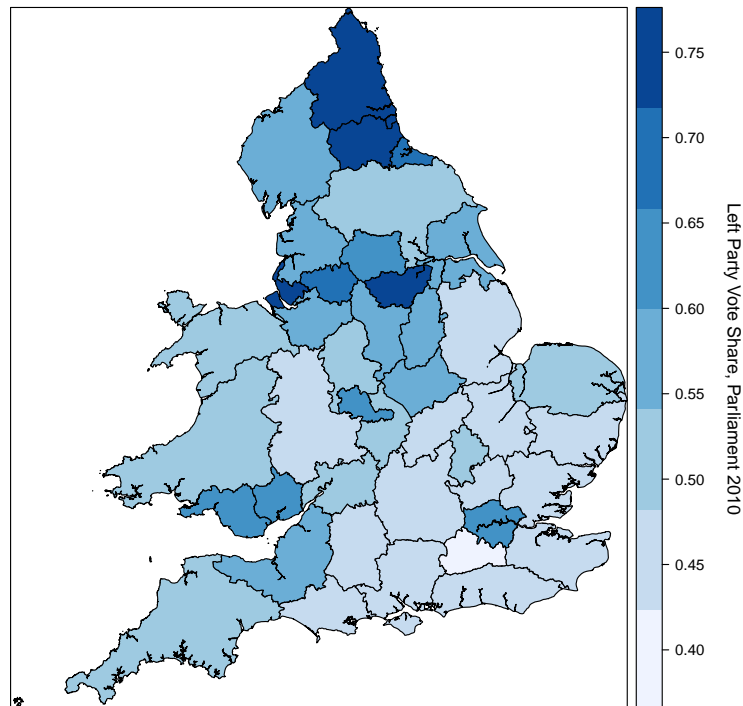
Standard errors reported in round parentheses estimated using 1,000 bootstrap samples blocked by police force. Square parentheses indicate magnitude of effect for a change along the interquartile range for each variable. Post is a binary variable taking the value one for years after 2013. Left Share variables measure the vote share for Labour, Lib Dem, and Green candidates in the 2010 Parliamentary Election, 2012 PCC, and local elections held between 2009 and 2012, respectively. Pref Share variables measure the share supporting relaxed drug laws based on survey disaggregation (raw survey), random forest and post-stratification (RFP), and multilevel regression and post-stratification (MRP).

A.4 Descriptive Results

A.4.1 Map of Parliamentary Voting by Police Force Area

Our primary measure of preferences comes from parliamentary election data. We aggregate the votes to the police force area level using a geographic merge. Since nearly all parliamentary constituencies fit neatly inside a police force area, this is mostly just a technique for finding which constituencies go with which police forces efficiently. Below, we plot a map of police force boundaries, with police forces shaded by the left party vote share.

Figure A.5: Map of Left Party Voting for Parliament in 2010 by Police Force Area.



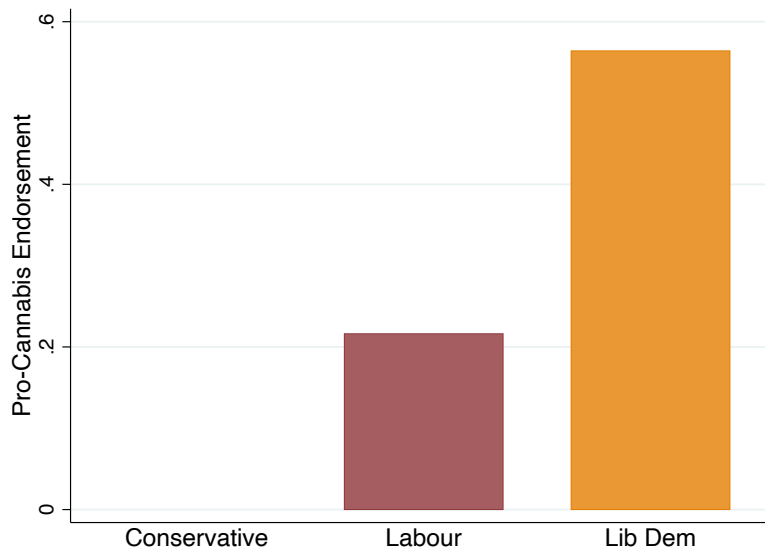
A.4.2 Cannabis Interest Group Endorsements

Access Kaneh, a cannabis industry interest group advocating for medical cannabis, contacted all candidates for police and crime commissioner candidates in 2021. Access Kaneh endorsed candidates based on their positions on drug policy. We compute the endorsement rate by party as a measure of how supportive each party is of the culturally liberal stance on drug policy. We report the results in Figure A.6.

Access Kaneh did not endorse any Conservative party candidates. In every district where a Green party candidate stood for election, Access Kaneh endorsed that candidate. Access Kaneh endorsed a higher share of Lib Dem candidates than Labour candidates, but they supported many Labour candidates nevertheless.

This result provides supplementary evidence for the partisan alignment of drug policy stances.

Figure A.6: **Pro-Cannabis Endorsement Rate by Party.**



A.4.3 Partisan Control of Police Forces Before And After Reform

Below, we report the share of police forces by party controlling the police authority in 2012 (pre-reform) and by party controlling the police authority in 2013 (post-reform, after the first PCC election).

Table A.14: **Share of Police Forces by Partisan Control, Police Authority and Police and Crime Commissioners.**

		Commissioner Party		
		Left	Middle	Right
Authority	Left	0.24	0.03	0.11
Party	Middle	0.03	0.03	0.11
	Right	0.00	0.26	0.21

Each cell reports the share of police forces by the party controlling the police authority in 2012 and the party of the police and crime commissioner in 2013. The party controlling the police authority is the median party of elected members with Conservatives defined as 1, Labour members, Greens, and Liberal Democrats defined as -1, and all others defined as 0.

A.4.4 Prevalence of Police Stops, By Ethnicity

Police stops are not uniformly distributed across communities. In England and Wales, Black residents are nearly three times as likely to be stopped by the police. With the data we have, we cannot easily determine the level of police presence in some neighborhoods versus others. Still, this provides suggestive evidence that any decisions police make will affect some residents much more than others. This raises important questions about how we can respect the majority's will while protecting the rights of electoral minorities.

Figure A.7: **Frequency of Police Stops (Per 1,000 Population), By Ethnicity.** Data from all Police Forces in Analysis Sample, 2014-2018.

