Consolidation of Prosecutor Offices

CLAUDIO DETOTTO, University of Corsica
BRYAN C. MCCANNON, West Virginia University

ABSTRACT

We investigate the impact of the consolidation of local agencies publicly providing services, focusing on the prosecution of crime. In many states in the United States, prosecution is a county office with a head prosecutor directing decisions. Other states merge counties into prosecutorial districts. We show that, along with the cost savings that arises, the evidence suggests that prosecutorial services benefit from increasing returns to scale. The efficiency of prosecutorial output is higher in consolidated districts. Further, we show that it is not owing to differing outcomes of the criminal justice system but to fewer inputs used.

I. INTRODUCTION

The prosecution of suspected criminals is a crucial publicly provided service. In the United States, state-level prosecutors close approximately 3 million cases a year, which make up the bulk of crimes (95% of all felonies), and exercise a substantial amount of discretion.[1] For filed charges, prosecutors make investigatorial investments and prosecution decisions, which can include dismissing the charges, plea-bargaining (through either sentence, fact, or charge bargaining; Piehl and Bushway 2007), or prosecuting cases at trial.

The decisions they make affect the victims and their families and the accused and have spillover effects through deterrence. Also, importantly, the decisions affect public finances. Prosecuting crimes not only exhausts the prosecutor's budget and labor resources but also affects the costs of the courts, jails and prisons, and probationary supervision and has opportunity costs owing to the backlog of cases. Thus, the funds available for prosecution should affect the quality of the criminal justice system.

Contact the corresponding author, Bryan C. McCannon, at bryan.c.mccannon@gmail.com.

1. The description of the data is given in Sec. III. See Simmons (2004) as well for other nationwide data.

Journal of Law and Courts (Fall 2020) © 2020 by the Law and Courts Organized Section of the American Political Science Association. All rights reserved. 2164-6570/2020/0802-0003$10.00. Electronically published August 27, 2020.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

In the majority of states in the United States, the prosecution of crimes is controlled by the counties. The county prosecutor, typically known as the district attorney, is allocated a budget from county and state funds. Given the other publicly provided services and the deadweight loss from taxation, one would like to ensure that the funds are being used effectively. Additionally, one would like to know which prosecutorial decisions and institutional rules lead to more effective use of the funds.

Some states in the United States have consolidated multiple county prosecutor offices into a single prosecutorial district to conserve on public finances. For example, North Carolina has 100 counties but only 43 offices. More heavily populated counties, such as Mecklenburg County, which contains the city of Charlotte, are a single district. Rural counties are merged and have a single district attorney directing the prosecution. When a district's population grows, it can be divided into separate offices. Between 2000 and 2006, for example, the state of North Carolina has done this four times.[2] Alternatively, when the public finances are especially tight, offices can be merged and costs saved. For example, when the most recent recession hit the state's budget in 2010, proposals were put forth in the North Carolina Assembly to explore reductions in the number of offices.[3] Twenty-three states follow this institutional design strategy.[4]

Since the redistricting is done to conserve resources, local communities are typically concerned about the quality of the public services provided.[5] For example, District Attorney Roxanne Vaneekhoven, commenting on a proposal to consolidate the prosecutor office she heads with a neighboring county, claims that "bigger government does not serve the victims of crime" and argues that "district attorneys in smaller districts can give victims of crimes bigger and louder voices, and also keep district attorneys more accountable and more accessible to the people they serve" (quoted in Smith 2011).

Here we explore this concern. Does the consolidation of prosecutorial services in fact reduce effectiveness? There is suggestive evidence that it might. Rasmusen, Raghav, and Ramseyer (2009) investigate the size of the budgets allocated to state-level prosecutor offices. They provide evidence, using a cross-sectional data set, that increased budgets expand both the extensive margin (number of crimes prosecuted) and the intensive margin (amount of resources devoted to each case). A similar finding arises when considering the impact of resources on the prosecution of corruption cases (Alt and Lassen 2014). These works show that increased resources increase prosecutorial output. While consolidation

2. See Bandyopadhyay and McCannon (2014) for a discussion of the redistricting in North Carolina. Also, see http://law.justia.com/north-carolina/ for a collection of annual changes in statutes in North Carolina.

3. See http://www.clemmonscourier.net/News/041411-DA-consolidation-and-more-qcd for details on this example. Also, see Jacoby et al. (1996) for a detailed study of the efficacy of prosecution in North Carolina.

4. Four of these 23 states have consolidated all counties into a single state office with local head prosecutors appointed, rather than elected, to their positions.

5. See http://www.clemmonscourier.net/News/041411-DA-consolidation-and-more-qcd.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

is distinct in that it need not be associated with budget cuts, the research findings would suggest that office consolidation aimed at financial savings would be expected to reduce service effectiveness. We evaluate an alternative impact of consolidation. Prosecutorial offices in the United States could be experiencing increasing returns to scale. The merging of multiple county offices into larger districts allows for more output to be achieved with disproportionately less resources. If this is the case, then consolidation can lead to both conserved public funds and improved public services.

To illustrate, consider the role of investigators. The prosecutor's office is provided with evidence justifying arrest of the suspect. Additional investigation could potentially lead to improved information helping secure a trial conviction or facilitating plea bargaining. It could also expose evidence that the accused is not guilty. For smaller prosecutorial offices, investigative services may be a luxury they cannot afford. In a nationwide comprehensive survey of local prosecutor offices in the United States (see Sec. III for details on the data), 56.4% of responding offices did not employ any investigator even though they averaged 109.8 violent crimes and 937.1 property crimes committed per year. Only 24.7% of offices can afford more than one full-time investigator. If two or more smaller offices can be merged, then supporting staff, such as investigators, can be employed and deployed on those cases in which their services can have a valuable impact on the outcome.

To explore this hypothesis, we collect data on 2,298 prosecutor offices in the United States. Rather than just focus on the correlation between budget size and convictions, as done in Rasmusen et al. (2009), we use a method known as data envelopment analysis (DEA). DEA uses data on inputs and outputs of the production of prosecution to estimate a relative production possibilities frontier. How far each office is from this frontier is a measurement of its efficiency. A second-stage analysis is done to identify the determinants of the level of efficiency. We focus on whether multiple counties being merged into a single district correlates with effective prosecution.

We find strong evidence that consolidation of county offices into prosecutorial districts coincides with more efficient production. Estimates suggest that the technical efficiency measurement improves up to three-fourths of a standard deviation. Thus, consolidation corresponds with efficiency.

Furthermore, we are able to explore how these merged districts function. To summarize, we find that there are not statistically significant differences in outcomes of the criminal justice system. These offices do not plea-bargain more or dismiss fewer cases. The difference is in the inputs used. Consolidated offices use fewer prosecutors and fewer supporting staff to achieve the same output. Merged districts are more likely, though, to have the funds to be able to hire investigators to improve the quality of the evidence than similarly matched offices. Therefore, the concern of lost output from consolidation lacks verification.

We are not the first to use DEA to analyze the functioning of the criminal justice system. Examples of areas investigated include the efficiency of policing (Drake and Simper 2005; Gorman and Ruggiero 2008), judges (Deyneli 2012), and courts

https://doi.org/10.1086/709909 Published online by Cambridge University Press

(Pedraja-Chaparro and Salinas-Jimenez 1996; Schneider 2005; Castro and Guccio 2014). In fact, Gorman and Ruggiero (2009) conduct a similar analysis on prosecutors in the United States, focusing on the relationship between socioeconomic characteristics of the districts and prosecutorial effectiveness. Finally, Detotto and McCannon (2016) also analyze the efficiency of prosecutor offices. They construct state-level aggregates of prosecution, using it as a proxy for quality of publicly provided services, and compare it to the quality of government regulation of markets, captured in economic freedom indexes.

Our work also contributes to the literature investigating the incentives and behaviors of prosecutors. While the career concerns of US federal attorneys have received attention (Glaeser, Kessler, and Piehl 2000; Boylan 2004, 2005; Boylan and Long 2005; Pavlik 2017), the attention that has been paid to state-level prosecutors focuses on the effect of election pressures and the decision to go to trial or plea-bargain (Bandyopadhyay and McCannon 2014, 2015), incarceration rates (Nadel, Scaggs, and Bales 2017), dismissals of charges (Dyke 2007), case backlogs (Bandyopadhyay and McCannon 2018), and appeals of convictions (McCannon 2013).[6] The exceptions are the DEA conducted by Gorman and Ruggiero (2009) and the public financing investigation by Rasmusen et al. (2009). We build on the former by expanding the second-stage analysis to the effect of differences in criminal justice institutions. The latter, while a valuable investigation into the consequences of public financing decisions, differs from the work presented here in that direct productive efficiency measurements are not estimated.

Our work provides important new results on the broader issue of optimal jurisdiction size. Empirical investigations into publicly provided service provision cover many applications of bureaucratic merger. Examples include health (Bates, Lafrancois, and Santerre 2011; Bloom et al. 2015) and education (Alesina, Baqir, and Hoxby 2004). These investigations do not attempt to estimate efficiency but instead focus on correlates of consolidation and spending. Important related works illustrate school consolidation's economies-of-scale benefit (Duncombe, Miner, and Ruggiero 1995; Duncombe and Yinger 2007). They do not, though, employ DEA as is done here. Ruggiero (1996) engages in a discussion of DEA methods, applying the work to school districts, but also does not consider consolidation and jurisdiction size.

Section II briefly introduces the DEA method, while Section III describes the data. Section IV presents the main results, and Section V concludes.

II. DATA ENVELOPMENT ANALYSIS

To assess the performance of the prosecutor offices, we employ the nonparametric technique known as data envelopment analysis. DEA is a linear programming formulation

6. Nelson (2014) provides complementary evidence that prosecutors respond to public opinion by changing their case dismissal practices when new information arises.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

used to measure the distance from the production possibilities frontier in multiple input and output environments.

Different from a parametric approach, DEA does not require an a priori specification of the functional form of the production function as well as an a priori hypothesis on the disturbance term. Furthermore, DEA reduces a multiple input-output productive structure into an easier virtual uni-input-output analysis, which makes the analysis more flexible than other parametric approaches.

To elaborate, suppose there are M inputs that can be used and N outputs that can be produced by the J agents. Denote the amount of input m used by agent j as x_{jm} and the amount of output n produced by agent j as y_{jn}. Thus, the production of j is denoted by the vector $Y_j = (y_{j1}, \ldots, y_{jN})$ using the input vector $X_j = (x_{j1}, \ldots, x_{jM})$. Following Gorman and Ruggiero (2009), the input-oriented programming model for efficiency is[7]

$$\text{TE} = \min \theta_j$$

such that

$$\sum_{i=1}^{J} \lambda_i x_{im} \geq \theta_j x_{jm} \qquad \forall m = 1, \ldots, M$$

$$\sum_{i=1}^{J} \lambda_i y_{in} \geq y_{jn} \qquad \forall n = 1, \ldots, N$$

$$\lambda_i \geq 0 \qquad \forall i = 1, \ldots, J.$$

The solution to this problem, for each agent, provides a measurement of efficiency, known as "technical efficiency" (TE). Hence, TE is defined as the equiproportional reduction of observed inputs consistent with existing production. Consequently, the values of the TE range between 0 and 1. A value of 1 represents no uniform reduction in inputs to achieve observed levels of output, so that the agent is producing efficiently. Values near 0 are interpreted as the situation in which the observed output could be produced with large equiproportional reductions in inputs. Such an agent is operating ineffectively. The smaller this value, the more ineffective is its production.

We choose to employ DEA methods in our analysis rather than alternatives, such as stochastic frontier analysis. Parametric techniques such as stochastic frontier analysis, as stated, require that specific functional form assumptions are made and that the inefficiency term has a specific distribution form (e.g., normally distributed errors), which can be problematic. DEA does not suffer from these issues. It is a flexible linear technique. Furthermore, the economic variables of interest can be used jointly despite their measure scale.

7. In the literature two main DEA models are considered: the DEA-CCR model developed by Charnes, Cooper, and Rhodes (1978), which assumes that all agents are operating at constant returns to scale, and the DEA-BCC model developed by Banker, Charnes, and Cooper (1984), which is based on the variable returns to scale hypothesis. In this study, the former is adopted.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

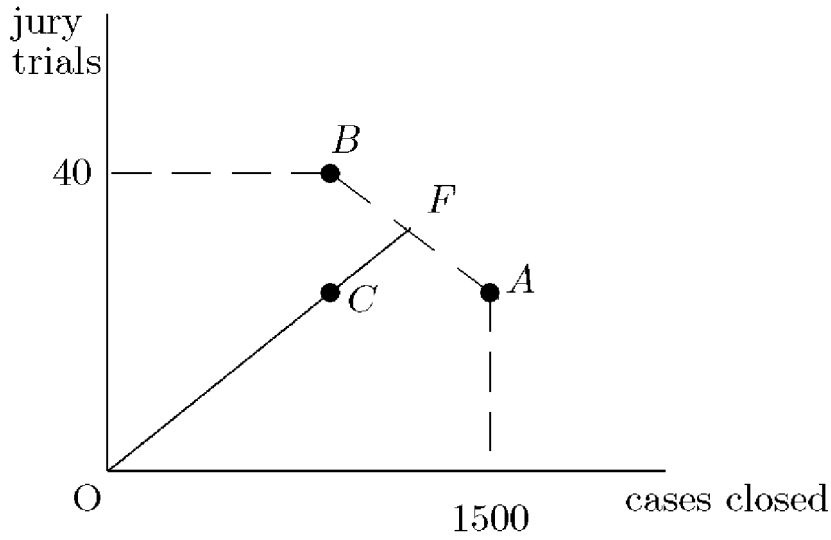


Figure 1. Technical efficiency measurement

To better understand how the DEA method captures productive efficiency, consider the following simple numerical example. Suppose there are three offices serving similar communities. To simplify the presentation, we assume each uses the same amount of inputs (e.g., 10 prosecutors and 20 supporting staff). One office, A, closes 1,500 cases in a year and prosecutes 20 cases at jury trials. A second office, B, closes 1,200 cases and prosecutes 40 at trial. Given that jury trials are resource-intensive activities and both offices are using the same input levels, we cannot say whether A or B is operating less efficiently than the other. Suppose, though, the third office, C, uses the same inputs but closes 1,200 cases and prosecutes 20 jury trials. Clearly, C is operating inefficiently. The DEA method constructs a linear production frontier using the outputs of A and B. A ray originating from the origin, passing through C's output bundle, can be drawn. How close C's point is to the frontier becomes the TE measurement for C.[8] Figure 1 graphically depicts this example. While this is a simple example with only three offices, two inputs (used in identical quantities), and two outputs, it illustrates how DEA captures efficiency.

It is important to note that DEA uses the direct inputs of the production process. Environmental constraints can be expected to influence the ability to take full advantage of the production technology. Thus, as is common in the literature, we engage in a two-stage process. We first employ the DEA method using the direct inputs and outputs to calculate a TE score. In the second stage, we consider the external influences and constraints on the efficient conversion of inputs into outputs.

8. In our numerical example, this ratio is 0.968.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

III. DATA

Data on the production of prosecution come from the Census of State Prosecutors, 2007, conducted by the Department of Justice. The survey collects basic information from every state-level prosecutor office in the United States. In 2007, the last survey produced, there were 2,330 offices in the United States. It is common in many states for prosecutor offices to be county-level offices (e.g., New York). Some states, though, organize a few counties into prosecutorial districts (e.g., North Carolina).

From this survey, four output variables and two inputs variables are calculated. The input variables are the number of prosecutors employed and the number of supporting staff reported. Part-time workers are coded as 0.5 of a worker. An office has a head prosecutor and a team of assistant prosecutors. A number of different potential supporting staff can be employed in an office. Examples of categories recorded are office managers, civil attorneys, victim advocates, investigators, and secretaries/clerical staff. Rather than disentangle the differences in labor input provided, all nonprosecutorial staff are aggregated into the supporting staff variable.

Four output variables are provided in the census and used. They are (1) the number of cases closed during the year, (2) the number of criminal convictions obtained during the year (through either guilty verdicts at trial or guilty pleas), (3) the number of jury verdicts rendered (either a conviction or an acquittal), and (4) the population of the district. Each of these measurements captures prosecutorial production. The number of closed cases is a direct measurement of how much is handled, or, rather, the extensive margin. This is an incomplete measurement, though, because it does not capture the level of effort and resources devoted to prosecution, the intensive margin. Thus, the number of convictions obtained captures some of this effect as more effort and resources devoted to prosecution should act to clear the backlog. Also, courtroom trials consume a substantial amount of resources and therefore presumably generate large expected benefits to the office. Thus, the number of jury trials measures the intensive margin as well. Finally, as is common in the literature (Gorman and Ruggiero 2009), the population of the district is used as a proxy for the number of nonprosecution services (e.g., drug awareness programs) provided.

As a robustness check, we also consider policing as inputs into the prosecution production. Data on local police staff are provided by the FBI's Uniform Crime Report.[9] County-level aggregates of both total police employed and "civilian staff" are reported for counties across the country. It can be argued that more generous staffing at local police offices provides more and better evidence. This leads to improved prosecutor output. Policing resources are a complementary input into prosecution production. Since the data coverage of policing is not as complete as for prosecution ($N = 1{,}872$), the baseline analysis will not include police inputs, but robustness checks will.

9. Dataset from the FBI's Criminal Justice Information Services (CJIS) at https://www.fbi.gov/services/cjis/ucr.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

Table 1. Prosecutorial Production

	Mean	Min	Max
Input:			
No. of prosecutors	12.00	.5	926.5
No. of supporting staff	22.52	.17	1,247.5
Output:			
No. of closed cases	1,314.8	0	64,585
No. of individuals convicted	940.4	0	58,050
No. of jury trial verdicts	32.0	0	3,000
Population of the district	132,210	474	9,948,081

A number of adjustments are made to the data. First, the census survey queried each office on the total number of staff and total supporting staff and asked for a breakdown of staff by role. Owing to reporting error or intentional omissions, the greater of the total supporting staff and the sum of the breakdown in staff roles is used in the analysis. In 105 observations only the total supporting staff was reported. Also, for 26 observations only the total staff was given. For these cases the number of prosecutors is subtracted to measure the supporting staff. Second, 30 observations are eliminated owing to no responses on supporting staff questions. This represents only 1.3% of the population. The average number of prosecutors in these districts is 1.83, which is substantially less than the sample average. Third, with regard to the output variables, two observations are dropped owing to missing information. Consequently, there are 2,298 observations used in the analysis.[10]

For the United States there are a total of 2.91 million closed cases generating 2.18 million convictions (a 75% conviction rate). In 2007 there were 73,274 jury trials, which is 2.5% of all closed cases. Table 1 provides descriptive statistics on the inputs and outputs used. There is substantial variation in the size of caseloads of the districts. The low proportion of convictions that come from jury trials (3.4%) matches previous observations in the literature (Bandyopadhyay and McCannon 2014). Jury trials are a resource-intense, rarely used prosecution tool.

From the US Census, a number of county-level variables are collected. The gender and racial composition are identified. They are measured as proportions of the population. To account for economic differences, unemployment, labor force participation, the prevalence of poverty, median household income, and educational attainment are recorded. Political differentiation is identified by collecting information on the percentage of votes cast in the 2008 presidential election for John McCain, as well as the voter participation rate in that year for every county in the United States. Economic variables not only capture the characteristics of a district but also indicate the level of resources available for the

10. Gorman and Ruggiero (2009) limit their analysis to districts with populations between 100,000 and 500,000. In 2007 this range constitutes only 20.3% of the offices. As a consequence, many of these data issues are not present there.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

Table 2. Descriptive Statistics

	Mean	SD
Technical efficiency	.159	.099
Multiple	.149	.356
No. of counties in the district	1.335	.951
Prop. of the population that is white	.889	.139
Prop. of the population that is male	.498	.034
Unemployment rate (% of labor force)	4.701	1.609
Labor force participation rate (prop. of population)	.486	.057
Median household income (2007, in USD)	44,166	10,759
High school degree (% of population)	83.60	7.510
Prop. below the poverty line	.490	.355
Population	1.3×10^5	3.8×10^5
Prop. of votes for McCain in the 2008 election	.557	.138
Prop. that voted in 2008 election	.464	.221
No. of violent crimes in 2007	618.8	2,642.8
No. of property crimes in 2007	4,134.6	13,463

prosecutor's office. Furthermore, the selection of prosecutors differs across jurisdictions. While most states use popular election to select and retain the head prosecutor, four states appoint them. An indicator variable is created to account for this difference. Finally, data on the number of violent crimes and property crimes committed in 2007 are included. The level of crime is relevant for an office's costs. Also, if crime is high, it may be easier for a prosecutor to clear a given number of cases. Thus, crime is an important external factor to be included. The variables used in the second stage of the analysis are presented in table 2.

Thus, approximately 15% of the sample consists of prosecutorial districts that collect multiple counties into one office. The descriptive statistics are calculated over the sample of prosecutor offices, which are not equal in population size. Hence, the mean values of the poverty rate and electoral support for McCain, for example, do not match national values.[11]

The mean value of TE is low. DEA constructs a relative production frontier so that an observation's TE score is in comparison to the most extreme observations in the data set. Figure 2 presents a histogram with the number of districts across ranges of TE. Each bin is a range of values with interval lengths of 0.05.

As one can see, the TE scores' magnitudes are driven by 12 offices with TE values in excess of 0.75. This constitutes 0.5% of the sample. Owing to potential data misreporting or an unusual, unique event, for example, these observations may not be reliable. If they are excluded, the mean value increases by more than 90%.[12] We have also computed the

11. John McCain, e.g., received 45.7% of the votes, while Barack Obama received 52.9%. Also, as mentioned, police staffing data are also collected. The average number of police is 115.61 and supporting staff is 73.6, with standard deviations of 488.9 and 278.0 and medians of 26 and 16, respectively.

12. Reestimating the DEA method excluding the 12 outliers increases the mean value of TE to 0.307. Furthermore, the main results presented in the next section persist when the outliers are

https://doi.org/10.1086/709909 Published online by Cambridge University Press

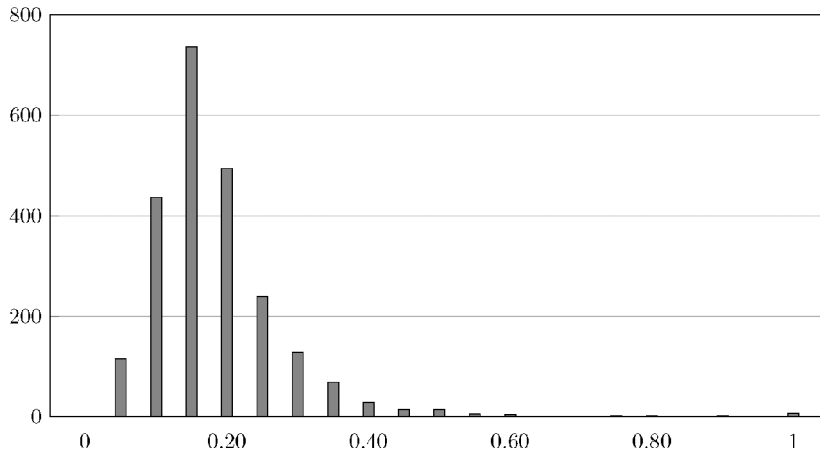


Figure 2. Distribution of technical efficiency

outliers detection procedure proposed by Wilson (1993). Only three outliers are detected. It is worth noting that they are not on the frontier, and they do not affect the main findings. This gives further support to the robustness and validity of our results.

IV. RESULTS

The hypothesis is that there is a relationship between the measured efficiency of the prosecutorial office and its composition. Comparing the subsample with multiple counties to the subsample of single-county districts, the average value of TE is 0.224 in the former and 0.148 in the latter. A two-sided *t*-test of the difference in these means (allowing for unequal variances) has $|t| = 10.7$ ($p < .001$). Also, considering nonparametric tests of the difference in the two distributions, a Wilcoxon rank sum test confirms their distinction ($|z| = 13.8$, $p < .001$), and the Kolmogorov-Smirnov test indicates that those districts comprising multiple counties have TE measurements that stochastically dominate the single-county offices ($D = 0.35$, $p < .001$).

A noticeable difference arises. Districts containing more than one county have an average TE approximately two-thirds of a standard deviation greater than the full sample mean. This suggests that consolidation increases productive efficiency. Of course, regression analysis, controlling for differences in the composition and characteristics of the districts, is needed.

The preliminary results illustrate a relationship between the consolidation of counties and TE. Table 3 presents the second-stage results. The positive and statistically significant coefficient on Multiple provides evidence that the consolidation of counties into prosecutorial

dropped. For example, the mean value of TE for those districts that merge multiple counties together is 0.424, while the unconsolidated districts have a mean of 0.287.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

Table 3. Results

	Baseline		Excluding Population	With Police
	(1)	(2)	(3)	(4)
Multiple	.062***	.055***	.013**	.055***
	(.010)	(.010)	(.006)	(.012)
No. of counties in the prosecutorial district	.005	.005	.001	.009**
	(.010)	(.004)	(.002)	(.010)
State where prosecutors are appointed		.013***		
		(.005)		
State where some county offices are consolidated		−.034*		
		(.020)		
Controls?	Yes	Yes	Yes	Yes
Adjusted R^2	.137	.140	.034	.158
F	26.9***	24.4***	6.70***	22.5***
Akaike information criterion	−4,421	−4,429	−6,671	−3,464.6

Note.—All controls listed in table 2 are included in each specification, along with a constant term. $N = 2,294$ in all specifications except col. 4, in which it is 1,718. Unadjusted standard errors in parentheses.

* 10% level of significance.

** 5% level of significance.

*** 1% level of significance.

districts corresponds with more efficient functioning of the criminal justice system. Specifically, it is whether there is consolidation, and not the number of counties consolidated, that matters. Controlling for the characteristics of the population served, having multiple counties in a prosecutorial district increases the measured efficiency of the office by 46% at the mean, or approximately three-quarters of a standard deviation.

It is important to distinguish between across-state variation and within-state variation. As stated, some states allow for prosecutor offices to be merged, while others do not. There can be important historical reasons, for example, for this difference. The distinction in the governance structures may have effects on the population other than the direct effect of the effectiveness of the prosecutor's office. Within a state that allows for consolidation, the efficiency of those that have consolidated is expected to differ from those that have not. To account for this distinction, we create a new indicator variable that equals 1 if the office is within a state that allows for multiple counties to be consolidated into one office. This state fixed effect allows us to differentiate the state governance policies from the direct effect of the consolidation. Additionally, as mentioned previously, while most states elect their local prosecutor, a few appoint them. A state fixed effect differentiating the two selection and retention mechanisms is added.

Table 3 column 2, which includes these state fixed effects, shows that the consolidation has an important within-state effect on office efficiency. It is noteworthy that the state fixed effect also has a positive and statistically significant coefficient. Those states that allow

https://doi.org/10.1086/709909 Published online by Cambridge University Press

for office consolidation see efficiency gains overall. Column 3 reestimates the TE but does not include population as an output. We include it, as do Gorman and Ruggiero (2008), to account for nonprosecution services. Since it is not a distinct outcome, we consider its omission. When it is dropped, consolidating offices still have a statistically significant relationship. Column 4 presents the results using the TE measurements that arise when police staffing, both the number of police and the number of support staff, is included as inputs in the DEA estimation. The results are robust to what is included as inputs in the production process.[13]

Table 3 presents unadjusted standard errors. If, alternatively, heteroscedasticity-robust standard errors, bootstrapped standard errors, or standard errors clustered by state are calculated, the main result persists. Additionally, if log transformations of the dependent and control variables are considered, the significance of Multiple remains. Furthermore, the results presented are straightforward ordinary least squares (OLS) estimates. Given that the dependent variable takes values only between 0 and 1, a concern arises that OLS might be inappropriate. The main result persists when a fractional regression model is estimated (Papke and Woolridge 1996).

Our study's motivation is that some prosecutor offices have increasing returns to scale but that policy impediments do not allow these benefits to materialize. This leads to reduced levels of TE. TE measurements can be lowered owing to either unrealized returns to scale or inefficient input allocations. Hence, it is appropriate to disentangle the source of the reduced productivity measurement to ensure that returns to scale are an important driver of prosecution in the United States.

Within DEA, a method to do so is to compare estimated TE assuming a constant returns-to-scale technology to estimated TE under increasing returns to scale. The technology experiences constant returns to scale if the maximization problem presented in Section II requires $\Sigma\lambda_i = 1$. Alternatively, an observation has increasing returns to scale if $\Sigma\lambda_i < 1$.[14] By comparing the estimated TE with and without the restriction that $\Sigma\lambda_i = 1$, we can identify whether the θ parameter is minimized with an increasing returns-to-scale technology. See Banker, Charnes, and Cooper (1984) and Banker et al. (2011) for thorough discussions of returns-to-scale estimation in DEA.

Following Charnes, Cooper, and Rhodes (1978) and Färe and Grosskopf (1994), if the θ parameter is greater with the restriction that the sum of the λ parameters is less than 1 than when the sum is equal to 1, then the observation is experiencing constant returns to

13. While not presented here, similar results arise if only police levels are included or if police and police staffing are merged into one input variable.

14. The linear programming problem has the output bundle measured as the radial distance from the frontier. The frontier is established using the parameters λ_i. When the parameters λ_i sum to 1, then the frontier is constructed from the convex combination of these points. Hence, this restriction captures a constant returns-to-scale technology. With increasing returns to scale, if these weights do not sum to 1, then an output bundle can be achieved using proportionately less inputs. Hence, increasing returns to scale exist.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

scale (or at least nonincreasing returns to scale). If the estimated value of θ is less, then it has increasing returns to scale. According to this procedure, for each prosecutor office a binary variable IRS can be derived capturing the nature of the scale inefficiency. If it has increasing returns to scale, then the variable IRS equals 1. It takes a value of 0 otherwise.

In the sample, 32.4% of the observations have IRS = 1. Interestingly, the two subsamples (IRS = 1 and IRS = 0) differ in the ratio of multiple counties, average income per capita, demographics (race and political ideology), and average population. Table 4 presents a comparison of the subsamples. Additionally, the results from a probit analysis are presented.

Districts consisting of multiple counties tend to have a higher likelihood of showing increasing returns to scale than the subsample of single-county districts. From the bottom panel of table 4, the probability that an observation is measured as experiencing increasing returns to scale is positively correlated with consolidation of multiple counties. This further supports the hypothesis that consolidation improves prosecution's efficiency.

Districts at a smaller scale do not necessarily operate under the same production function as larger ones that have merged multiple counties into a single office. Their technology has increasing returns to scale. Hence, consolidation can be beneficial.

Estimating the production frontier allowing for variable returns to scale was done to identify the presence of increasing returns to scale. Alternatively, though, one can be concerned that the primary results are sensitive to the constant returns-to-scale assumption. In effect, this presumes there is a constant marginal cost to prosecution. The offices, though,

Table 4. Increasing Returns to Scale (IRS)

	IRS = 1	IRS = 0	Δ
Multiple	.256	.097	.159***
Prop. of the population that is white	.826	.918	−.092***
Prop. of the population that is male	.492	.499	−.007
Unemployment rate (% of labor force)	4.693	4.691	.002
Labor force participation rate (prop. of population)	.493	.483	.010
Median household income (2007, in USD)	49,503	41,484	8,019***
High school degree (% of population)	84.362	83.277	1.085
Prop. below the poverty line	.522	.476	.046**
Population	353,894	27,375	326,519***
Prop. of votes for McCain in the 2008 election	.512	.580	−.068***
Prop. that voted in 2008 election	.458	.469	−.011
	Marginal Effect	SE	
Multiple	.066***	.020	
Adjusted R^2	.803		

Note.—The asterisks in the top panel's final column indicate the statistical significance of the difference in means. The bottom panel includes all control variables presented in the top panel.

* 10% level of significance.

** 5% level of significance.

*** 1% level of significance.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

Table 5. Sources of Efficiency

	Coefficient	SE
Input:		
No. of prosecutors	−2.037***	.742
No. of supporting staff	−1.271	1.690
Investigators? (yes = 1)	.864***	.152
Output:		
No. of closed cases	328.4**	135.2
No. of convictions	160.6*	92.6
No. of jury trial convictions	6.75	5.46
Jury/closed	−.055	.090
Closed/staff	42.86*	22.40

Note.—Coefficient for Multiple presented, along with its standard error. Each row represents the results with a different dependent variable. The specification in col. 1 of table 3 is used, omitting the number of counties in the prosecutorial district. The dependent variable Jury/closed is the number of jury trial convictions divided by the number of closed cases. The dependent variable Closed/staff is the number of closed cases divided by the total staff (prosecutors plus supporting staff).

* 10% level of significance.

** 5% level of significance.

*** 1% level of significance.

can be expected to have mostly fixed costs and nonconstant variable costs. Thus, the TE allowing for variable returns to scale can be used as an alternative dependent variable. If it is used as a replacement in column 1 of table 3, the coefficient on Multiple is 0.216 (SE = 0.16, $p < .001$), and in column 2 of table 3 it is 0.243 (SE = 0.015, $p < .001$).[15] Thus, the results are not sensitive to the model used.

While the previous analysis focuses on productivity measurement via DEA, the sources of the improved efficiency need exploration. The specification presented in column 1 in table 3 is replicated, but the inputs and outputs are used as dependent variables. Additionally, an indicator variable identifying whether the office employs at least one investigator is used. Table 5 presents these additional results. All control variables included in table 3 are included in each specification but are not reported. Only the estimated coefficient for Multiple and its standard error are given in table 5.

The results indicate that consolidation is correlated with reduced input use. Specifically, fewer prosecutors are employed. Interestingly, it is also more likely for an office to hire investigative staff. The marginal impact at the mean is an increase in the likelihood of employing at least one investigator by 23.6 percentage points.

There is also a greater output by districts consisting of multiple counties. The coefficients on the number of closed cases and the number of convictions are all positive and statistically significant. This suggests that more output is being obtained. Additionally, the number of cases closed per staff member is higher, which is an independent

15. The consolidation of multiple counties corresponds to an increase in TE by approximately four-fifths of a standard deviation. Thus, if one allows for a nonconstant marginal cost, the efficiency gain may be greater.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

identification of the efficiency gains from consolidation. It is noteworthy that the use of jury trials does not change. There is no statistically significant relationship between office consolidation and either the total number of jury convictions or the proportion of closed cases disposed of through a jury trial. Thus, the improved efficiency is not obtained through an increased rate of plea bargaining.[16]

V. CONCLUSION

Does the consolidation of local agencies publicly providing services hurt their effectiveness? We address this question by considering prosecutor offices. In many US states, the prosecution of crime occurs in a county office with an elected head prosecutor directing decisions. Other states merge counties into a prosecutorial district. We show that along with the cost savings that arise, the evidence suggests that prosecutorial services benefit from increasing returns to scale. The efficiency of prosecutorial output is higher in consolidated districts. Further, we show that it is not due to differing outcomes of the criminal justice system but to fewer inputs used.

The work has important policy implications. Given constrained public finances, prosecutor office consolidation does not necessarily compromise the service quality. This is important because there are strong spillover effects from prosecution (e.g., deterrence). Furthermore, an important difference between the control and treatment groups is the use of investigator services. While measurements of the quality of evidence are not available in the data, one would expect that increased investments in investigation should improve the decision-making quality. Not only would one expect better evidence to be obtained to facilitate the conviction of those who cause harm, but wrongful convictions would be reduced. Suggestive evidence exists, though. The proportion of cases closed that end in conviction decreases by approximately 8 percentage points. This is consistent with improved screening, where the innocent (or at least nonculpable) defendants have their cases dismissed.

By estimating productive efficiency, this work does not address what optimal prosecutorial decision making is. For example, disparities by race, gender, and ethnicity (Starr 2015); mistakes in convictions (McCannon 2013); and the distortions caused by reelection concerns (Bandyopadhyay and McCannon 2014, 2015) are important considerations to include in debates of institutional design. To fully appreciate the implications of merging county-level offices into prosecutorial districts, future research needs to further investigate the decision-making differences across offices.

Limitations in the data constrain our work. First, one should still hesitate before making full causal claims. The decision to merge offices is endogenous. One should be

16. Also, one can argue that the results presented in table 5 provide independent evidence of the value of office consolidation, as they do not rely on DEA. Thus, our results are robust to the manner in which inputs and outputs are considered.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

concerned that unobservable confounders are important in driving the office's efficiency and are related to its composition. The full set of factors that drive the decision is not fully measureable and identifiable. Also, while prosecutors may be expected to have "justice" as their objective, it is hard to know whether all outputs are being included in our analysis. Four reasonable ones are considered here. Relatedly, other inputs may matter. We show, for example, that the findings are robust to the inclusion of the complementary inputs provided by police. Furthermore, given the availability of cross-sectional data only, we are unable to fully explore the dynamic nature of the offices. This would be valuable for multiple reasons. For one, the case backlog is unknown. Thus, if a district entered the year with a larger caseload, those cases will be completed, and measured prosecutorial output expands, as compared to a similarly functioning office without the entering backlog. For another, using a cross-sectional data set, we are only able to compare consolidated districts to unconsolidated ones. Temporal variation would allow us to compare the same area over time as its office's scope changes. Given these data limitations, though, the strong results suggest an avenue for a meaningful policy reform.

To quantify the potential impact of consolidation, the results presented an estimate that 32.4% of the prosecutor offices in the United States are experiencing increasing returns to scale. Taking the most conservative estimate that all consolidated offices have increasing returns to scale (and no more consolidation is feasible), this leaves 552 offices in the United States with potential efficiency gains from merger. We estimate that the average savings in the budget of the merged districts is over $3 million per year,[17] and so a pairwise merger of these offices with a neighbor (taking the conservative estimate assuming no improvement in the neighbor's expenses) could potentially reduce the direct costs to the prosecution by approximately $1.7 billion. Hence, this public policy should be given further consideration.

While we focus here on the prosecution of crime, the work points to broader efficiency gains from local service consolidation. County lines, for example, were originally drawn for convenience. The consequence of the formation of the jurisdictions has been shown to still have spillover effects today (Libecap and Lueck 2011). While the debate regarding decentralization focuses primarily on the national government versus state/provincial governments and the incentives on spending (Oates 1972), the results presented here point to a new dimension of the argument—cost and effectiveness of publicly provided services are not always traded off. There is no reason to believe that other publicly provided services could not benefit from consolidation. We provide strong evidence that consolidation benefits the prosecution of crime, and future research should investigate the possibility in other services.

17. This calculation is provided in the appendix, available online. We use propensity score matching to identify the proper control group to estimate the average cost savings.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

REFERENCES

Alesina, Alberto, Reza Baqir, and Caroline Hoxby. 2004. "Political Jurisdictions in Heterogeneous Communities." *Journal of Political Economy* 112 (2): 348–96.

Alt, James E., and David Dreyer Lassen. 2014. "Enforcement and Public Corruption: Evidence from the American States." *Journal of Law, Economics, and Organization* 30 (1): 306–38.

Bandyopadhyay, Siddhartha, and Bryan C. McCannon. 2014. "The Effect of the Election of Prosecutors on Criminal Trials." *Public Choice* 161 (1): 141–56.

———. 2015. "Prosecutorial Retention: Signaling by Trial." *Journal of Public Economic Theory* 17 (2): 219–56.

———. 2018. "Queueing Up for Justice: Prosecutor Elections and Case Backlogs." Working paper, University of Birmingham.

Banker, Rajiv D., Abraham Charnes, and William W. Cooper. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 30 (9): 1078–92.

Banker, Rajiv D., William W. Cooper, Lawrence M. Seiford, and Joe Zhu. 2011. "Returns to Scale in DEA." In *Handbook on Data Envelopment Analysis*, ed. William W. Cooper, Lawrence M. Seiford, and Joe Zhu, 44–70. Boston: Springer.

Bates, Laurie J., Becky A. Lafrancois, and Rexford E. Santerre. 2011. "An Empirical Study of the Consolidation of Local Public Health Services in Connecticut." *Public Choice* 147 (1): 107–21.

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John van Reenen. 2015. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." *Review of Economic Studies* 8 (2): 457–89.

Boylan, Richard T. 2004. "Salaries, Turnover, and Performance in the Federal Criminal Justice System." *Journal of Law and Economics* 47 (1): 75–92.

———. 2005. "What Do Prosecutors Maximize? Evidence from Careers of U.S. Attorneys." *American Law and Economics Review* 7 (2): 379–402.

Boylan, Richard T., and Cheryl X. Long. 2005. "Salaries, Plea Rates, and the Career Objectives of Federal Prosecutors." *Journal of Law and Economics* 48 (2): 627–51.

Castro, Massimo F., and Calogero Guccio. 2014. "Searching for the Source of Technical Inefficiency in Italian Judicial Districts: An Empirical Investigation." *European Journal of Law and Economics* 38 (3): 369–91.

Charnes, Abraham, William W. Cooper, and Edwardo L. Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2 (6): 429–41.

Detotto, Claudio, and Bryan C. McCannon. 2016. "Economic Freedom and Public, Non-market Institutions: Evidence from Criminal Prosecution." *Economics of Governance* 18 (2): 107–28.

Deyneli, Fatih. 2012. "Analysis of Relationship between Efficiency of Justice Services and Salaries of Judges with Two-Stage DEA Method." *European Journal of Law and Economics* 34 (3): 477–93.

Drake, Leigh M., and Richard Simper. 2005. "Police Efficiency in Offences Cleared: An Analysis of English Basic Command Units." *International Review of Law and Economics* 25 (2): 186–208.

Duncombe, William D., Jerry Miner, and John Ruggiero. 1995. "On the Measurement and Causes of Technical Inefficiency in Local Public Service: With an Application to Public Education." *Journal of Public Administration Research and Theory* 5 (4): 403–28.

Duncombe, William D., and John Yinger. 2007. "Does School District Consolidation Cut Costs?" *Journal of Education Finance and Policy* 2 (4): 341–76.

Dyke, Andrew. 2007. "Electoral Cycles in the Administration of Criminal Justice." *Public Choice* 133 (3/4): 417–37.

https://doi.org/10.1086/709909 Published online by Cambridge University Press

Färe, Rolf, and Shawna Grosskopf. 1994. "Estimation of Returns to Scale Using Data Envelopment Analysis: A Comment." *European Journal of Operational Research* 79 (2): 379–82.

Glaeser, Edward L., Daniel P. Kessler, and Anne Morrison Piehl. 2000. "What Do Prosecutors Maximize? An Analysis of the Federalization of Drug Crimes." *American Law and Economics Review* 2 (2): 259–90.

Gorman, Michael F., and John Ruggiero. 2008. "Evaluating US State Police Performance Using Data Envelopment Analysis." *International Journal of Production and Economics* 113 (2): 1031–37.

———. 2009. "Evaluating U.S. Judicial District Prosecutor Performance Using DEA: Are Disadvantaged Counties More Inefficient?" *European Journal of Law and Economics* 27 (3): 275–89.

Jacoby, Joan E., Edward C. Ratledge, Ronald J. Taylor, and Noddie Barrion. 1996. "An Analysis of the North Carolina District Attorneys' Resources." Technical report, Jefferson Institute for Justice Studies.

Libecap, Gary D., and Gary Lueck. 2011. "The Demarcation of Land and the Role of Coordinating Property Institutions." *Journal of Political Economy* 119 (3): 426–67.

McCannon, Bryan C. 2013. "Prosecutor Elections, Mistakes, and Appeals." *Journal of Empirical Legal Studies* 10 (4): 697–715.

Nadel, Melissa R., Samuel J. A. Scaggs, and William D. Bales 2017. "Politics in Punishment: The Effect of State Attorney Election Cycle on Conviction and Sentencing Outcomes in Florida." *American Journal of Criminal Justice* 42 (4): 845–62.

Nelson, Michael J. 2014. "Responsive Justice?" *Journal of Law and Courts* 2 (1): 117–52.

Oates, Wallace E. 1972. *Fiscal Federalism*. New York: Harcourt Brace Jovanovich.

Papke, Leslie A., and Jeffrey M. Woolridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics* 11 (6): 619–32.

Pavlik, Jamie Bologna. 2017. "Political Importance and Its Relation to the Federal Prosecution of Public Corruption." *Constitutional Political Economy* 28 (4): 346–72.

Pedraja-Chaparro, Francisco, and Javier Salinas-Jimenez. 1996. "An Assessment of the Efficiency of Spanish Courts Using DEA." *Applied Economics* 28 (11): 1391–403.

Piehl, Anne Morrison, and Shawn D. Bushway. 2007. "Measuring and Explaining Charge Bargaining." *Journal of Quantitative Criminology* 23 (1): 105–25.

Rasmusen, Eric, Manu Raghav, and Mark Ramseyer. 2009. "Convictions versus Conviction Rates: The Prosecutor's Choice." *American Law and Economics Review* 11 (1): 47–78.

Ruggiero, John. 1996. "Efficiency of Educational Production: An Analysis of New York School Districts." *Review of Economic Statistics* 78 (3): 499–509.

Schneider, Martin R. 2005. "Judicial Career Incentives and Court Performance: An Empirical Study of the German Labour Courts of Appeal." *European Journal of Law and Economics* 20 (2): 127–44.

Simmons, Ric. 2004. "Election of Local Prosecutors." Moritz College of Law e-Book. http://moritzlaw.osu.edu/electionlaw/ebook/part7/elections_prosecutors.html.

Smith, Shelley. 2011. "One DA May Handle Rowan, Cabarrus." *Salisbury (NC) Post*, April 3. http://issuu.com/salisburypost/docs/04142011-sls-a01/1.

Starr, Sonja. 2015. "Estimating Gender Disparities in Federal Criminal Cases." *American Law and Economics Review* 17 (1): 127–59.

Wilson, Paul. 1993. "Detecting Outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs." *Journal of Business and Economic Statistics* 11 (3): 319–23.

https://doi.org/10.1086/709909 Published online by Cambridge University Press