Crowding In with Impure Altruism: Theory and Evidence from Volunteerism in National Parks*

Matthew J. Kotchen
Yale University and NBER

Katherine R. H. Wagner
University of British Columbia

March 16, 2023

Abstract

This paper makes three contributions to the literature on private provision of public goods. First, we identify limitations of the frequently used specification test that distinguishes between the standard models of pure and impure altruism based on the extent of crowding out. While the literature takes as given the result that crowding out should be less with impure altruism than with pure altruism, we show that it can be either more or less. Second, we propose a new, more general test based on the presence of crowding in or greater than one-for-one crowding out, both of which are consistent with impure altruism only. Third, we provide empirical evidence. Using a unique panel data set on volunteerism in U.S. National Parks, we estimate the causal effect of changes in a park's public funding on the amount of within-park volunteerism. The overall finding is that each additional dollar of public expenditure crowds in 27 cents worth of volunteerism on average. We show how the estimates of crowding in, along with heterogeneity based on park and volunteer hour types, are theoretically consistent with the mainstay model of impure altruism.

*Kotchen: Yale University and NBER, 195 Prospect Street, New Haven, CT 06511 (email: matthew.kotchen@yale.edu); Wagner: University of British Columbia, 6000 Iona Drive, Vancouver, BC V6T 1L5 (email: katherine.wagner@ubc.ca). We are grateful to James Andreoni, Ted Bergstrom, Judson Boomhower, Nathan Chan, Robert Deacon, Joshua Gottlieb, Catie Hausman, Grant Jacobsen, Erin Mansur, Todd Sandler, Joe Shapiro, Lise Vesterlund, and Casey Wichman for helpful comments and discussions, along with seminar participants who commented on earlier versions of this research at Carnegie Mellon University, Columbia University, London School of Economics, McGill University, Stanford University, UC Berkeley, UC Santa Barbara, University of Connecticut, University of Pittsburgh, Yale University, and the NBER Summer Institute. Constructive feedback from the co-editor David Seim and three anonymous referees greatly improved the paper. Finally, we thank staff at the National Park Service for access to data and helpful discussions, including Karyn Ferro, Tom Medema, Kerry Olson, Joy Pietschmann, Lenny Teh, and Earnestine White.

1 Introduction

How does government provision of a public good affect private provision? The question is central to public economics. Implementing the optimal supply of public goods through government policy requires an understanding of the behavioral responses to public provision. If, for example, public supply crowds out private provision, then failing to account for the behavioral response means that the intended level of the public good will come up short. If, however, there is crowding in, then public provision can be used to leverage greater private contributions. Theoretical and empirical studies of crowding effects also provide the basis for an extensive literature that seeks to identify the underlying motives of charitable behavior. In particular, the extent of crowding out is widely used as the standard test between the models of pure and impure altruism, which provide different explanations for why individuals engage in charitable activity.

In this paper, we identify a limitation of the theoretical foundation for using crowding out as a specification test between pure and impure altruism. Thirty years-worth of theoretical and empirical literature takes as given the result that crowding out should be less with impure altruism than with pure altruism, but we show that it can be more or less. In particular, we show that the standard specification test is based on an implicit special case. As an alternative, we identify a new test based on the presence of crowding in or greater than one-for-one crowding out that does not rely on the implicit assumptions of the special case. We then illustrate the test in an empirical setting that seeks to overcome many of the challenges in previous studies that estimate crowding effects. We collect a unique panel data set on volunteerism in U.S. National Parks to estimate the causal effect of changes in a park's public funding on the amount of within-park volunteerism, along with heterogeneous effects motivated by theory.

The seminal theory on private provision of a public good establishes the basic framework for crowding out (e.g., Warr 1982; Roberts 1984; Cornes and Sandler 1985; Bergstrom, Blume and Varian 1986). With a pure public good, an individual's own provision and public provision are perfect substitutes, and the private and public goods over which individuals obtain utility are both assumed to be normal (i.e., not inferior). One consequence is that funding public goods through lump-sum taxation crowds out private provision one-for-one, assuming an interior solution. Despite the influence and theoretical appeal of these models, many of their results, including that of complete crowding out, are viewed as limited within the literature because of their empirical implausibility. Indeed, Andreoni (2006) summarizes many of the implications of complete crowding out as a classic *reducto* *ad absurdum*.[1]

Other models of privately provided public goods seek to reconcile the apparent disconnect between theoretical predictions and empirical observations. In many cases, a broader formulation of individual preferences is used to explain why crowding out might be less than complete. One of the most influential is Andreoni's (1989, 1990) model of impure altruism. In addition to enjoying the benefits of a public good, individuals are assumed to enjoy a private "warm-glow" benefit from the act of their own giving. This setup implies that private and public provision are no longer perfect substitutes, and a foundational result for much of the literature is that crowding out becomes incomplete (i.e., less than one-for-one). In a related model, Cornes and Sandler (1984, 1994) consider the provision of an impure public good, which nests impure altruism through the joint production of private and public characteristics.[2]

There exists a substantial empirical literature on crowding out that tests between the models of impure altruism and pure altruism (i.e., the pure public good model). As mentioned above, underlying much of this research is the theoretical prediction that crowding out should be less with impure altruism than with pure altruism. In an early and influential contribution, Kingma (1989) estimates incomplete crowding out of private contributions to public radio stations, and because of this, the results are generally interpreted as consistent with the model of impure altruism. Building on the same theoretical and empirical approach, numerous other papers estimate crowding out in a variety of contexts (e.g., Ribar and Wilhelm 2002; Okten and Weisbrod 2011; Borgonovi 2006; Hungerman 2005; Payne 2009; Andreoni and Payne 2003, 2011). Another strand of the literature uses controlled laboratory experiments to estimate crowding out and test between the models of pure and impure altruism (e.g., Andreoni 1993; Eckel, Grossman and Johnston 2005; Ottoni-Wilhelm, Vesterlund and Xie 2017).[3]

The majority of studies find evidence of incomplete crowding out, but many studies also estimate crowding in. A recent meta-analysis including hundreds of empirical estimates finds that crowding out is more likely to arise in experimental studies, whereas observational studies are more skewed toward a finding of crowding in (de Wit and Bekkers 2017). Figure 1 shows the probability density functions for the estimated crowding effects

[1] Specifically, Andreoni (2006) argues that, "If we are going to accept complete crowding out, we also need to believe in near complete crowding of any government gifts to charity, that only the very richest are giving, that redistributions of income are neutral as long as people are giving to charities, and that even 'distortionary' taxes may be non-distortionary" (p.1219).

[2] Joint production is also consistent with other approaches for modeling charitable behavior based on reputation (Hollander 1990; Harbaugh 1998), signaling about income (Glazer and Konrad 1986), gift exchanges (DellaVigna, List, Malmendier and Rao 2016), and environmentally friendly consumption (Kotchen 2005, 2006).

[3] We focus here on the private provision of individuals. Another strand of the public finance literature studies crowding in or out at different levels of government. In that context, crowding in is sometimes referred to as the "flypaper effect" in recognition that public money tends to stick where it hits (Hines and Thaler 1995).

separately for experimental and non-experimental studies. In total, 38 percent of the estimates have a finding of crowding in of private donations, rather than crowding out. In these cases, because the current interpretation of the standard model of impure altruism does not admit crowding in, the results are typically described as outliers or interpreted in the context of models that account for additional elements (e.g., Khanna and Sandler 2000; Payne 2001; Andreoni and Payne 2011; Andreoni, Payne and Smith 2014). These explanations include the possibility that public provision signals quality in a way that makes private contributions more attractive to donors, or perhaps increases the scale or scope of fundraising.

In the first part of this paper, we reconsider the theoretical foundation for crowding out in the standard model of impure altruism, and we come to different conclusions. We show that crowding out itself can be either more or less with impure altruism than with pure altruism, even under the standard assumptions of the original model. This means that, in general, the extent of crowding out cannot be used to distinguish between pure and impure altruism, as is typical in the literature. We find that, in general, the relative magnitudes depend on the degree of complementarity or substitutability between the jointly produced private and public characteristics (i.e., the public good and warm glow, or some other private benefit). Nevertheless, we do show that crowding out will always be less with impure altruism than with pure altruism in the special case of additive separability in the utility function and strict concavity. This means that the standard result holds if the private marginal benefits from giving (e.g., "warm glow") are independent of the public benefits (e.g., scope of charitable activity). We argue that this unstated assumption seems unlikely to hold in most settings, and yet without it, the standard specification test is inconclusive.

We also propose a new specification test that is valid under more general conditions. We show that crowding in and greater than one-for-one crowding out are consistent with impure altruism, but not pure altruism, given conventional normality assumptions about goods. We therefore propose an alternative test to distinguish between models based on the presence of crowding in or greater than one-for-one crowding out.

In the second part of the paper, we estimate the crowding effect in an empirical setting that allows us to illustrate our new specification test. To do so, we collect administrative data from the Volunteers-In-Parks (VIP) program of the U.S. National Park Service (NPS) from the years 1998 to 2013. The data set includes the total number of volunteer hours in all NPS units, along with the type of activity associated with each hour (e.g., resource conservation, indoor administration, etc). We combine these 16 years of volunteer data with data on annual park budgets and visitation. Our primary objective is to estimate the causal effect of changes in federal funding to a park on within-park volunteerism. Our preferred estimates are based on a shift-share instrumental variables approach (Bartik 1991; Autor, Dorn and Hanson, 2013; Notowidigdo 2019). Specifically, we instrument for changes in park funding with a Bartik-style interaction between individual park budgets in a base year and changes in Congressional support for conservation issues, as measured by the League of Conservation Voters.

Compared with most other studies of crowding effects, our study takes place in a setting where the impure altruism model (or equivalently the impure public good model) is more apparently the expected model *ex ante,* rather than the model deduced from the estimates *ex post*. A unique feature of the focus on volunteerism in national parks is that private provision of the public good is clearly associated with a jointly produced private benefit. Volunteerism contributes to the public good of park quality (e.g., conservation), while simultaneously creating opportunities for *in situ* park enjoyment.[4] The presence of such joint production creates the theoretical possibility for crowding in or greater than one-for-one crowding out, findings which would be consistent with our new specification test for distinguishing between the models of pure and impure altruism.

We do in fact find robust evidence of crowding in, along with heterogeneous effects by volunteer-hour and park types that support the theory. We estimate that, on average, a $1,000 increase in a park's annual funding increases volunteerism by 12 hours. Using a conventional wage rate for conversion, this translates into crowding in of 27 cents for every additional dollar of public expenditure. We also find evidence of crowding in that is greater for parks and volunteer activities that are more conservation and outdoors oriented—that is, in circumstances where the joint production of public and private benefits are most likely to arise. Overall, these estimates of crowding in substantially increase the marginal benefit of NPS funding, emphasizing the importance of considering individuals' responses to changes in public funding.

Our empirical setting also has a number of unique advantages that help to overcome challenges that arise in many of the existing studies of crowding effects. First, all of the park units in our sample are managed through the same administrative unit (i.e., the NPS), and this adds a unique degree of homogeneity across the units of analysis. Most other studies rely on cross-sections or panels of organizations that provide different public goods and operate in different ways.[5] Second, federal funding accounts for almost all of each park's operating budget, thereby reducing concerns about endogenous fundraising from individuals and corporations, and funding from other levels of government. Third, individuals generally volunteer in only one park, and so our analysis is not subject to concerns about substitution between units of study masked by aggregate data. Fourth, the

[4]Although the jointly produced private benefit of impure altruism is usually characterized as a warm glow, the model is itself general enough to account for any jointly produced private benefit, such as *in situ* park enjoyment in this case, or for both private benefits.

[5]While most studies in the literature on crowding effects focus on social services (Andreoni and Payne 2013), one distinguishing feature herein is a focus on behavior related to environmental and natural resource conservation.

breadth and length of our panel means that we can estimate within-unit crowding effects over a longer period of time than most other studies. Finally, the detailed data on volunteer hours broken down by type of activity enables a unique opportunity to examine heterogeneous effects informed by theory.

The remainder of the paper proceeds as follows. Section 2 begins with background on the model for impure altruism. Section 3 revisits the standard specification test between pure and impure altruism and puts forth our more general alternative. Section 4 describes our empirical setting and data collection. Section 5 outlines our empirical strategy. Section 6 reports the estimation results, considers robustness checks, and discusses possible alternative explanations. Section 7 provides a concluding discussion.

2 Impure Altruism: Preliminaries

This section begins with the setup and assumptions of Andreoni's (1989, 1990) seminal model of impure altruism. We then reproduce the model's well-established results on crowding out to provide points of reference for our alternative derivation and more general results in Section 3.

2.1 Setup and Assumptions

In the standard model of impure altruism, individual preferences are given by a strictly increasing and strictly quasiconcave utility function of the form $U_i(x_i, G, g_i)$, where x_i is a private good, g_i is the individual's own contribution to the public good G, and G_{-i} is the exogenously given level of the public good provided by all others. Total provision of the public good is $G = g_i + G_{-i}$. Letting w_i denote the individual's wealth endowment, and normalizing prices to unity, the individual's utility maximization problem can be written as

$$\max_{x_i, g_i} U_i(x_i, G, g_i) \text{ s.t. } x_i + g_i = w_i \text{ and } G = g_i + G_{-i}. \tag{1}$$

The inclusion of g_i as a separate argument in the utility function is the distinguishing feature of this setup, compared with the pure public good model, where preferences are given by $U_i(x_i, G)$. Utility from g_i on its own is often interpreted as a "warm glow" or "joy of giving" that one receives from voluntarily contributing to the public good. More generally, however, the separate argument for g_i can represent any private benefit that arises through joint production of an impure public good (Cornes and Sandler 1984, 1994). In our empirical context, which we motivate in detail below, the choice of g_i represents time spent volunteering in a national park, where volunteerism provides an impure public good. The joint production in this case comes from provision of a public good G in the form of park quality (e.g., conservation), along with a private benefit g_i of park enjoyment while volunteering *in situ*, of warm-glow, or both.

Substituting the constraints in (1) into the utility function, the individual's problem can be rewritten as a choice over the aggregate level of the public good:

$$\max_{G \geq G_{-i}} U_i(w_i + G_{-i} - G, G, G - G_{-i}). \tag{2}$$

The standard approach in the literature is to then express the solution as a function of the exogenous components of the maximand such that

$$G^* = f_i(w_i + G_{-i}, G_{-i}). \tag{3}$$

One reason for writing the solution in this way is to facilitate comparisons with the pure public good model (Bergstrom, Blume and Varian 1986), in which case demand for G is written more simply as $h_i(w_i + G_{-i})$, which includes only the first argument of f_i in (3). Key results for the pure public good model then follow by assuming normality of both the private and public goods, which is equivalent to assuming $0 < h_i'(w_i + G_{-i}) < 1$ for interior solutions. These results include a unique Nash equilibrium, neutrality of wealth redistributions, and crowding out.[6]

With impure altruism, the partial effects of f_i are typically intuited in the literature as follows.[7] Normality of x_i and G implies that an increase in income increases demand for the public good by less than one-for-one, i.e., $0 < f_{i1} < 1$. The intuition is that an increase in w_i must be spent on both the private and public goods. A further claim is that normality of x_i and g_i implies that demand for G increases when G_{-i} increases, i.e., $f_{i2} > 0$. Here the intuition follows from the thought experiment of simultaneously lowering G_{-i} and increasing w_i by the same amount. The argument is that some of the increase in w_i must be spent on both x_i and g_i, so demand for G must fall and hence $f_{i2} > 0$. Finally, it is also standard to assume that $0 < f_{i1} + f_{i2} < 1$, and this condition ensures uniqueness of the Nash equilibrium.[8]

2.2 Crowding Out and Specification Tests

The preceding assumptions on the partial effects of the demand function have direct implications for the model's predictions about crowding out, along with comparisons to those

[6]Neutrality refers to Warr's (1983) result showing that small redistributions of the endowments among contributors to the public good will have no effect on the total level of equilibrium provision.

[7]Andreoni (1989, 1990) provides the original analysis, and more recent applications that work through the same set of assumptions include Ribar and Wilhelm (2002) and Ottoni-Wilhelm, Vesterlund and Xie (2017).

[8]See Kotchen (2007) for a proof of equilibrium existence and uniqueness. Note that the conditions of $f_{i1} < 1$ and $f_{i1} + f_{i2} < 1$ are written with a strict inequality because of the simplifying assumption of an interior solution. Without this assumption, the inequalities would need to hold only weakly.

for pure altruism (i.e., the pure public good model). By definition, an individual's own contribution will satisfy

$$g_i^* = f_i(w_i + G_{-i}, G_{-i}) - G_{-i}, \tag{4}$$

and there are two notions of crowding out to consider based on this expression. First is the unfunded effect of an exogenous change in G_{-i} that leaves unspecified the source of the increase in funding:

$$\frac{\partial g_i^*}{\partial G_{-i}} = f_{i1} + f_{i2} - 1. \tag{5}$$

Second is a funded effect, where lump-sum taxation is used to fund the change in exogenous provision. In this case, we have

$$\left.\frac{\partial g_i^*}{\partial G_{-i}}\right|_{dG_{-i}=-dw_i} = f_{i2} - 1. \tag{6}$$

The assumptions above that $0 < f_{i1} + f_{i2} < 1$ and $0 < f_{i1} < 1$ mean that crowding out is less than one-for-one in both the funded and unfunded cases, because both (5) and (6) are between -1 and 0; the assumptions therefore eliminate the possibility for crowding in. More important, because $f_{i2} > 0$, a further prediction is that crowding out will always be less with impure altruism than with pure altruism (because in the latter case $f_{i2} = 0$). In particular, a standard theoretical result in the literature is that funded crowding out that is less than one-for-one is consistent with impure altruism but not pure altruism.

3 Specification Tests Revisited

The preceding observations underlie an extensive empirical literature that tests between pure and impure altruistic motivations for private provision of public goods. Nevertheless, we show in what follows that the standard specification test for impure altruism is not a general result and that the use of (5) and (6) to motivate it does not account for the constrained joint production between G and g_i. In particular, we show that contrary to conventional wisdom in the literature, crowding out need not be less with impure altruism than with pure altruism. We also propose an alternative specification test that focuses instead on the presence of crowding in or on greater than one-for-one crowding out.

Returning to utility maximization problem (2), the first-order condition that defines the solution is $-U_x + U_G + U_g = 0$. The distinguishing feature between impure and pure altruism is that the latter has $U_g = 0$. Following Andreoni (1989, 1990), we write the solution for impure altruism as $G^m = f_i^m(w_i + G_{-i}, G_{-i})$, and following Bergstrom, Blume and Varian (1986), we write the solution for pure altruism as $G^a = f_i^a(w_i + G_{-i})$. We introduce notation to distinguish f_i^m and f_i^a because the partial derivatives differ across the two models, as we will show.

Two assumptions are helpful to clearly state at the outset. First is the normality assumption. Following Andreoni (1989, 1990), we define normality with respect to goods rather than preferences.[9] In particular, it is assumed that regardless of whether pure or impure altruism is operational, all goods are normal with respect to a change in wealth such that $dG^*/dw_i = dg_i^*/dw_i > 0$ and $dx_i^*/dw_i > 0$. The second assumption is that U_i is additively separable in x_i. This condition is only for simplicity and not restrictive because our aim is to illustrate different possibilities, none of which depend on this assumption.

Differentiating the first-order condition with respect to w_i, we can solve for the way that a change in wealth affects demand for the public good. For pure altruism we have

$$\frac{dG^a}{dw_i} = f_{i1}^a = \frac{U_{xx}^a}{U_{xx}^a + U_{GG}^a}, \tag{7}$$

and for impure altruism we have

$$\frac{dG^m}{dw_i} = f_{i1}^m = \frac{U_{xx}^m}{U_{xx}^m + U_{GG}^m + U_{gg}^m + 2U_{Gg}^m}. \tag{8}$$

The normality assumption implies that both expressions must be strictly greater than zero and strictly less than one. Notice that the expressions are not the same, as seems to be implied in previous work, where it is assumed that $f_{i1} = f_{i1}^a = f_{i1}^m$. Indeed, even in the special case where $U_{gg}^m + 2U_{Gg}^m = 0$, the expressions will still not be equal in general because the initial solution to the utility maximization problem in (2) will be different. A second observation is that the normality assumption imposes bounds on some of the signs and relative magnitudes of the second derivatives. The denominators are simply the second-order conditions, so they must be strictly negative to ensure a maximum. Normality therefore requires $U_{xx}^a < 0$ and $U_{GG}^a < 0$ in (7), and $U_{xx}^m < 0$ and $U_{GG}^m + U_{gg}^m + 2U_{Gg}^m < 0$ in (8).

3.1 Relative Crowding Out

We can derive conditions for the relative magnitudes of crowding out for pure and impure altruism in terms of the expressions for the partial derivatives f_i^a and f_i^m. Differentiating the first-order condition with respect to G_{-i}, we have

$$\frac{dG^a}{dG_{-i}} = \frac{dG^a}{dw_i} = f_{i1}^a, \tag{9}$$

[9] Normality based on goods is a condition where demand for a good is increasing in wealth. Normality based on preferences is a condition based on primitives of the utility function. It turns out that the two definitions of normality are not equivalent with impure altruism because of the way that g_i enters in two arguments of the utility function. However, the two definitions are equivalent with the pure public good model. See Cornes and Sandler (1984, 1994) for further discussion.

which is identical to the effect of a change in exogenous wealth in (7), and

$$\frac{dG^m}{dG_{-i}} = f_{i1}^m + f_{i2}^m = \frac{U_{xx}^m + U_{gg}^m + U_{Gg}^m}{U_{xx}^m + U_{GG}^m + U_{gg}^m + 2U_{Gg}^m}. \tag{10}$$

Focusing first on unfunded crowding out, the standard result that crowding out is less with impure altruism than with pure altruism requires $f_{i1}^a < f_{i1}^m + f_{i2}^m$. Comparing (9) and (10), the condition is equivalent to requiring

$$U_{Gg}^m(U_{xx}^a - U_{GG}^a) < U_{GG}^a U_{gg}^m. \tag{11}$$

We begin to see that whether this condition holds depends on whether the public good G and private provision g_i are q-complements ($U_{Gg}^m > 0$), q-substitutes ($U_{Gg}^m < 0$), or utility is additively separable ($U_{Gg}^m = 0$).[10] The inequality holds if, for example, $U_{Gg}^m = 0$ and utility is strictly concave in g_i. But the inequality need not hold more generally. The simplest case to see alternate possibilities is to assume linearity in g_{-i}, so that $U_{gg}^m = 0$ and the right-hand side of (11) is equal to zero. This means that contradicting (11) requires only that U_{Gg}^m and $U_{xx}^a - U_{GG}^a$ have the same sign. The sign of the former is positive (negative) if G and g_i are q-complements (q-substitutes), and sign of the latter is positive (negative) if (7) is less than (greater than) $1/2$. Hence reversing the standard result can occur in two cases: $f_{i1}^a < 1/2$ and q-complements, or $f_{i1}^a > 1/2$ and q-substitutes. Intuitively, the first case implies that if the income effect with pure altruism is such that the individual wants less public good than private good, q-complements means she will want even less public good, resulting in greater crowding out with impure altruism.

The same possibilities arise with funded crowding out, where the standard result depends on the condition $f_{i2}^m > 0$, as shown in (6). Subtracting (8) from (10) and rearranging indicates that the condition is equivalent to $U_{Gg}^m < -U_{gg}^m$. But again the inequality need not hold. For example, continuing to assume $U_{gg}^m = 0$, q-complements is sufficient to violate the inequality and ensure that crowding out is greater with impure altruism than with pure altruism.

We have thus proved the following result:

Result 1 *Both unfunded and funded crowding out with impure altruism can be more or less than crowding out with pure altruism.*

[10]We employ definitions of complements and substitutes based on the sign of cross-partial derivatives of the utility function. We use these definitions, q-complements and q-substitutes, rather than the more familiar price-based definitions in order to derive direct and intuitive results that illustrate the range of possibilities. In an earlier working paper (Kotchen and Wagner 2019), we derive comparable conditions with respect to price-based relationships. The same possibilities arise, but the conditions are less intuitive because crowding effects are themselves based on quantities rather than prices. See Madden (1991) for a detailed analysis on the relationship between these different definitions of complements and substitutes.

Additionally, we have shown how the different results depend critically on whether the public good and private provision are complements or substitutes, and very plausible arguments could be made for either case. That is, the marginal utility of private provision could be increasing or decreasing in the level of the public good. One might even argue that assuming additive separability in g_i is the least plausible scenario. Nevertheless, it is worth noting that the additional assumptions of additive separability and strict concavity in g_i are sufficient conditions to rule out the different possibilities and ensure the conventional wisdom of less crowding out with impure altruism.

3.2 Crowding Out or In

Our analysis thus far has focused only on crowding out. The normality assumption admits only crowding out with pure altruism. This follows because $0 < f^a_{i1} < 1$, which means that (5) implies unfunded crowding out that is less than complete, i.e., within $(-1, 0)$, and (6) implies funded crowding out that is complete, i.e., equal to -1. We now show that the normality assumption does not imply the same bounds when it comes to impure altruism. In particular, we show that the impure altruism model admits the possibility for crowding in and for crowding out greater than one-for-one.[11] In other words, we clarify how the standard condition for impure altruism that $0 < f^m_{i1} + f^m_{i2} < 1$ is an assumption and not a result, and this differs from pure altruism.

We begin with the possibility for crowding in. Referring back to the unfunded effect in (5), it is sufficient to show the possibility for $f^m_{i1} + f^m_{i2} - 1 \geq 0$. Subtracting 1 from (10) and rearranging, we find that satisfying the needed inequality is equivalent to satisfying $U^m_{Gg} \geq -U^m_{GG}$. The required condition for funded crowding in based on (6) is $f^m_{i2} - 1 \geq 0$. Subtracting (8) and 1 from (10) and rearranging yields the condition in this case, $U^m_{Gg} \geq -U^m_{GG} - U^m_{xx}$. To see how satisfying both conditions is possible, consider the limiting case where $U_{xx} \to 0$, so that the funded condition converges to that for unfunded crowding in. It turns out that sufficiently strong q-complements is all that is required.[12]

Turning now to the possibility for greater than one-for-one crowding out, the conditions to show based on (5) and (6) are $f^m_{i1} + f^m_{i2} \leq 0$ for an unfunded effect and $f^m_{i2} \leq 0$ for a funded effect. Following the same steps, these conditions are equivalent to showing $U^m_{Gg} \geq -U^m_{xx} - U^m_{gg}$ and $U^m_{Gg} \geq -U^m_{gg}$, respectively. Continuing to assume the limiting case

[11] Cornes and Sandler (1984, 1994) also illustrate the possibility for crowding in and greater than one-for-one crowding out in the impure public good model. In doing so, however, they consider normality based on preferences rather than goods, and as noted previously, these definitions are not the same in this setting. Our analysis tracks more closely with Andreoni's (1989, 1990) original setup that discusses normality based on goods. Combining the results indicates that both crowding effects are possible under both normality assumptions.

[12] It is straightforward to verify that satisfying $U^m_{Gg} \geq -U^m_{GG}$ need not contradict the second-order condition or the normality assumption. These conditions combined require that $2U^m_{Gg} < -U^m_{GG} - U^m_{gg}$, which means that U^m_{Gg} can be greater than either of the terms on the right-hand side, but not both.

of $U_{xx} \to 0$, we see again that the result follows with sufficiently strong q-complements, but this time the threshold is based on the rate of diminishing marginal utility of private provision rather than the public good.[13] If the degree of q-complementary is stronger than the rate of diminishing marginal utility of g_i, then sufficiently strong complementarity between G and g_i adds a reinforcing effect to crowding out that can cause it to be greater than one-for-one.

We summarize our results of this subsection as follows:

Result 2 *Assuming the normality of goods, evidence of crowding in or greater than one-for-one crowding out is consistent with impure altruism (i.e., provision of an impure public good), but not with pure altruism.*

The remainder of the paper focuses on empirically estimating the crowding effect on volunteerism in national parks. While a finding of either crowding out or crowding in is consistent with impure altruism, a finding of crowding in enables rejection of pure altruism.

Before turning to our empirical application, however, it is worth emphasizing how the conditions derived here can apply in any setting with joint production between the private and public goods (Cornes and Sandler 1984, 1994). The impure altruism model is a special case, but as noted previously, others include charitable giving motivated by reputation (Hollander 1990; Harbaugh 1998), signaling (Glazer and Konrad 1986), and gift exchanges (DellaVigna, List, Malmendier and Rao 2016). Environmentally friendly consumption provides another increasingly salient example (Kotchen 2005, 2006). In all of these cases, the different possibilities will depend on the degree of substitutability or complementarity between the jointly produced private and public goods.

4 Empirical Setting and Data Collection

This section begins with institutional background about the Volunteers-in-Parks (VIP) program of the U.S. National Park Service (NPS). We then discuss how VIP participation is consistent with the impure altruism motivation for private provision of a public good. This establishes a foundation for interpreting our subsequent estimates of the crowding

[13] In particular, notice that $-U^m_{gg}$ is on the right-hand side of the condition here for greater than one-for-one crowding out, whereas $-U^m_{GG}$ is on the right hand-side for crowding in. While sufficiently strong complements is required in both cases, the thresholds differ, and the conditions are mutually exclusive. To see, this, we can write the second-order condition as requiring $[U^m_{Gg} + U^m_{GG} + U^m_{xx}] + U^m_{Gg} + U^m_{gg} < 0$. The condition for crowding in is that the term in brackets is weakly positive, in which case a necessary condition is $U^m_{Gg} + U^m_{gg} < 0$. But this eliminates the possibility of satisfying the condition for greater than one-for-one crowding out, $U^m_{Gg} + U^m_{gg} \geq -U^m_{xx} > 0$, where the final inequality was shown previously to follow from the normality assumption. Of course, satisfying neither condition means that crowding out is less than one-for-one, and this is consistent with the standard model.

effect, which are based on how changes in budget appropriations to a national park affect the amount of within-park volunteerism. In this section, we also describe our data and report summary statistics.

4.1 The NPS and VIP Program

The NPS is an administrative branch of the U.S. Department of the Interior. The mission of the NPS is to preserve natural and historical landmarks for the enjoyment and education of current and future generations. The NPS system includes over 400 sites, comprises over 84 million acres, and hosts more than 330 million visitors per year. NPS sites include the iconic National Parks that prioritize environmental conservation and outdoor recreation, in addition to parks that emphasize the conservation and restoration of cultural heritage. Examples of the former include Yellowstone and Yosemite National Parks, and examples of the latter include the Statue of Liberty and the Booker T. Washington National Monuments. Throughout the paper, we use the term national park in reference to any of the NPS managed sites.[14] The NPS itself is divided into seven regional offices that manage parks and programs within their geographic jurisdiction, including partial oversight of the VIP program.[15]

Initiated in 1970, the VIP program is a NPS-wide program that facilitates the active involvement of volunteers in protecting and maintaining national parks. Volunteers through the VIP program are an integral part of the NPS workforce, as exemplified by the fact that over 300,000 volunteers contributed over eight million hours of service to the national parks annually. Volunteer recruitment occurs mainly through word-of-mouth and staff referrals, as well as through online volunteer postings on NPS websites that solicit applications for openings that require specialized skills (NPS 2007). Volunteers are employed across all parks and programs in positions that assist the NPS paid staff of approximately 22,000 employees, or that focus on tasks that could otherwise not be accomplished due to funding shortfalls (NPS 2017a). Funding for the VIP program is separate from other park operations and is regionally allocated before being distributed to individual parks.

Federal appropriations are the primary source of funding for national parks. During the time period of our study, the NPS annual budget request is about $2.5 billion on average. Monetary donations to the NPS from individuals, corporations, and non-profit organizations such as the National Park Foundation are about $30 million per year in total, or about 1% of federal appropriations (GAO 2009).[16] To formulate the annual NPS bud-

[14] NPS sites are technically classified into more than two dozen categories depending on their mandate and the types of programs they undertake, including, for example, National Parks, National Monuments, National Historic Sites, and National Battlefields (Comay, 2013).

[15] The seven NPS regions are Alaska, Intermountain, Midwest, National Capital, Northeast, Pacific West, and Southeast.

[16] The National Park Foundation is the NPS's national non-profit partner and independently solicits indi-

get request, each park begins with a baseline operational budget that is adjusted annually to reflect new projects, maintenance, and programmatic needs (Turner and Walker 2006). Individual park funding requests are then aggregated into the overall NPS budget, which is part of the overall Department of the Interior request. As with funding for all federal agencies, the budget request requires Congressional approval, and the enacted amounts typically differ from the agency requests. The budget appropriations committees in the U.S. House and Senate lead the reconciliation process of the budget between the executive and legislative branches.

The U.S. Government shutdown at the end of 2018 and beginning of 2019 illustrates the importance of public funding for park quality. When public funding for parks ceased during the shutdown, parks experienced widespread problems with vandalism, waste, litter, and damage to natural and cultural resources due to unauthorized access of sensitive areas (NPCA 2019).

4.2 Volunteerism and Joint Production

Fundamental to the design of the VIP program is that volunteers can contribute to a public good while simultaneously benefiting from time spent in national parks. The promotional materials make this explicit, as "the primary purpose of the VIP Program is to provide a vehicle through which the NPS can accept and utilize voluntary help and services from the public in such a way that is mutually beneficial to the NPS and the volunteer" (NPS 2017a). The mutual benefits arise because participation in the VIP program is associated with joint production of public and private goods: the promotion of conservation and opportunities for *in situ* park enjoyment. Volunteerism in the NPS therefore closely matches the notion of private provision of an impure public good, which nests the model of impure altruism.[17]

To see how the decision to volunteer in a national park links directly to the model of impure altruism, we need only modify the budget constraint. The composite private good, with a normalized price, must satisfy $x_i = \iota_i + \omega(\tau - v_i)$, where ι_i is the individual's wealth endowment, ω is the wage rate applied to a time budget τ, and v_i is the number of hours spent volunteering. Then, letting $w_i = \iota_i + \omega\tau$ and $g_i = \omega v_i$, we recover the original budget constraint of $x_i + g_i = w_i$, where g_i represents the monetary equivalent of time spent volunteering.[18] Nothing else needs changing about the model other than a

vidual and corporate donations, as well as donations from other non-profit entitites. Corporate donors, who comprise about 75% of the National Park Foundation's donor base, typically support specific park projects, while other funds are disbursed to individual parks at the Foundation's discretion (GAO 2004).

[17]If, for example, the private benefit were a warm glow from volunteerism rather than park enjoyment, the impure public good reduces to impure altruism. Chan and Kotchen (2014) generalize the impure public good framework to account for the joint production of multiple public and private goods, which could include, for example, both park enjoyment and a warm glow from volunteerism. Having multiple private benefits does not affect the range of possible theoretical results presented here.

[18]The relatively low level of monetary donations (which unfortunately the NPS does not track centrally

scaling for crowding effects depending on the unit of measurement. Specifically, because $\partial g_i^*/\partial G_{-i} = \omega \partial v_i^*/\partial G_{-i}$, the crowding effect on volunteer hours need only be multiplied by the wage rate to have the standard interpretation. The NPS itself places a value on volunteer time using the average, annual hourly wage of non-agricultural workers from the Bureau of Labor Statistics, plus an adjustment of 12 percent to account for fringe benefits.[19] During the time period of our study, this conventional value of volunteer time is $21.85 on average, reported in 2013 dollars.[20]

In this setting, where we study the effect of federal appropriations on volunteerism, there is also the question of whether the crowding effects should be interpreted as funded or unfunded. We contend that either is possible because any expected difference between them should be exceedingly small. While experimental studies are able to associate changes in G_{-i} with changes in lump-sum taxation, we examine the effect of changes in G_{-i} without explicit reference to its funding by volunteers. While this perspective suggests an unfunded effect, it must also be recognized that domestic volunteers are taxpayers, so changes in aggregate appropriations do not go unfunded. Moreover, even for volunteers who are not taxpayers, we would argue that their own share of the required funding would be exceedingly small. This means that any income effects would be correspondingly small, in which case the unfunded and funded crowding effects converge. This is readily seen in the limiting case of quasi-linear preferences, which we have shown with $U_{xx} \rightarrow 0$.

4.3 Data

Our primary source of data is a unique and detailed administrative data set from the NPS on annual participation in the VIP program from 1998 through 2013.[21] These data include the annual total number of volunteer hours in each park broken down by the type of volunteer activity. We also obtained data from the NPS on the annual number of full time equivalent (FTE) paid staff in each park. We use publicly available, park-specific data on

by park or on an annual basis) compared with the value of time diminishes any potential concerns about the substitutability between giving time or money in our setting (Lilley and Slonim 2014; Brown, Meer and Williams 2017).

[19]The Independent Sector, a national membership organization of nonprofits, foundations, and corporations, provides an overview of the estimates and their use at `https://independentsector.org/value-volunteer-time-methodology/`. We use the standard value of volunteer time that the NPS uses for consistency and because of its conventional use. Other wage rates could be applied if data were collected by the NPS on the demographic characteristics of volunteers in aggregate or for specific individuals (Mas and Pallais 2017). While this would affect our empirical estimates on the crowding out of dollar-equivalent volunteerism, it would not change our primitive estimates on the crowding of volunteer time.

[20]All monetary values throughout the paper are reported in 2013 dollars unless otherwise indicated.

[21]Data from 2014 through 2018, which we obtained more recently through a Freedom of Information Act request, were collected and categorized into volunteer activities through a different process and unfortunately do not have continuity with the earlier period. We exclude these data from our analysis for this reason, along with the fact that the NPS experienced unique circumstances related to visitation, volunteerism, and management during and around its centennial celebrations in 2016.

annual park visitation (1998-2013) and the federal budget appropriation to each park in each year (1995-2013).[22] We collect the League of Conservation Voters (LCV) score for each member of the U.S. House and Senate Appropriations Committees, along with the annual average for both chambers of Congress. The LCV annually scores each voting member of the House and Senate on a 0 to 100 scale based on the percent of pro-environmental and conservation legislation that each member supports.[23] We use the LCV scores to create an instrumental variable for park funding, as we describe in the next section.

Our final sample for analysis includes 326 parks among the 398 originally listed in the VIP data set. The smaller number of parks is due primarily to the way that we include only those that track annual visitation.[24] Figure 2 shows the geographic distribution of the national parks included in our analysis, along with an indicator for each park's NPS administrative region. The first column of Table 1 reports summary statistics across all 326 parks during the sample time period. On average, parks benefit from nearly 16,000 hours of volunteerism per year with an annual value of approximately $350,000, which is roughly nine percent of the average, annual park budget of approximately $3.88 million. Parks host an average of 827,000 visitors per year and employ 0.31 FTE per thousand visitors. The large standard deviations across variables indicate a large degree of heterogeneity among parks.

During the time period of our study, there is significant variation in volunteer hours and funding. Figure 3 shows the annual trends in volunteer hours and federal funding aggregated across all parks. While the total number of volunteer hours has maintained an upward trend, federal funding to the parks follows an upward trend until the recession in 2009. The aggregate trends in Figure 3 do not show a clear pattern in the relationship between park funding and volunteerism, yet our analysis focuses on estimating a causal effect of funding on volunteerism within parks (i.e., not in aggregate). Because of the substantial drop in post-recession funding, we conduct some of our analysis with and without the post-recession years to test for robustness.

One of the distinguishing features of the data set is that we can observe the specific type of activities that volunteers undertake. Figure 4 shows the distribution of volunteer hours across the different types of activities. The most common activity is interpretation, and

[22] Annual park visitation is available online at `https://irma.nps.gov/Stats/`. Annual park budget appropriations are reported in the U.S. Department of the Interior Budget Justifications for the National Park Service, referred to as the Greenbooks. Selected years are available online at `https://www.nps.gov/aboutus/budget.htm`. Other years were obtained by request.

[23] All LCV data is available online at `https://www.lcv.org/`. The website also provides a detailed description of the methodology used to generate the scores for each voting member of the U.S. Congress.

[24] Visitation counts are not recorded at 64 smaller urban parks and scenic trails, where unpredictable flows of pedestrian traffic and the existence of multiple access points make consistent counts infeasible. Additionally, we exclude two parks that do not have volunteer data, three parks that do not have their own budgets, one park that does not have information on the number of paid staff, and two parks for which we have only one year of data.

when combined with natural resource management and maintenance, the three categories comprise over 75 percent of all volunteer hours. Other categories include campground hosting, cultural resource management, and administration, for example.

We use the natural resource management category for the additional purpose of categorizing parks as primarily environmental or non-environmental in some of our analyses that examine heterogeneous effects. We conjecture that volunteering in parks and for activities with more of a natural resource focus are more likely associated with joint production and therefore impure altruism. The rationale is that environmental parks and programs typically focus on recreation, which provide different private benefits to volunteers than parks that focus on curating, for example.[25] We therefore distinguish between environmental and non-environmental parks based on whether or not volunteer hours dedicated to natural resource management are strictly positive for all years within a park. If yes, we classify the park as environmental. This procedure yields 105 environmental parks, and the last two columns of Table 1 report descriptive statistics separately for the two groups. On average, environmental parks have more volunteer hours, greater funding, more visits, and fewer paid FTE per visitor. Later in the paper, as part of our robustness checks, we use an alternative, more general classification based on official park mandates. We also categorize volunteer hours based on whether they are predicted to occur outside or inside in order to examine heterogeneity by volunteer hour type.

5 Empirical Strategy

We now turn to our econometric strategy for estimating the crowding effect and ultimately illustrating our specification test between pure and impure altruism. We focus on estimating the causal effect of changes in federal appropriations to a national park on the amount of within-park volunteerism. We first describe our estimation of the average crowding effect across all parks, followed by tests for heterogeneous effects, and our preferred instrumental variables strategy.

5.1 Average Crowding Effect

We begin with fixed effects models of the form

$$Hours_{it} = \beta Budget_{it} + \gamma X_{it} + \alpha_i + \sigma_{rt} + \varepsilon_{it}, \tag{12}$$

[25]For specific examples about how joint production through volunteerism is likely to be greater in environmental parks and outdoor hours, consider the following: on-going restoration of the Mariposa Grove in Yosemite National Park provides volunteers with additional hiking trails and boardwalks to enjoy while volunteering, whereas federally funded maintenance of the copper skin of the Statue of Liberty is less likely to materially improve the recreational opportunities available to volunteers that provide tours.

where the dependent variable $Hours_{it}$ is the total number of volunteer hours in park i and year t; $Budget_{it}$ is a park's annual budget appropriation in thousands of dollars; X_{it} is a vector of time-varying and park-specific variables; α_i is a vector of park fixed effects; σ_{rt} is a set of year-specific intercepts for each of the seven NPS regions; and ε_{it} is an error term. The variables included in X_{it} are annual park visitation and FTE per 1,000 visits in the previous year, before budgets are received. We cluster standard errors at the park level to make statistical inference robust to potential serial correlation within parks.

The coefficient of primary interest is β because it provides an estimate of the crowding effect: a positive estimate indicates crowding in, whereas a negative estimate indicates crowding out. We have shown previously that impure altruism admits the possibility for either crowding out or in, and an estimate of crowding in or greater than one-for-one crowding out would enable rejection of pure altruism as the underlying motivation for VIP participation. We use volunteer hours as the dependent variable rather than the value of volunteer hours. The reasons are that hours are the original measure of volunteerism, that the conversion to a dollar value requires additional assumptions, and that we can readily obtain a dollar-for-dollar interpretation as an *ex post* adjustment to β, as we will show.

The identifying variation for the crowding effect comes from within-park fluctuations in the annual budget, conditional on the variables in the model. The park fixed effects capture time-invariant park characteristics, such as popularity, location, type, and program scope, which could affect both funding and volunteer hours. These fixed effects also control for differences in park size and baseline budgetary needs. The region-year fixed effects control for any annual shocks that are common to all parks within each NPS region, such as changes to VIP funding and trends in the importance of environmental issues. Models that include additional covariates estimate crowding effects based on variation in funding within parks conditional on visits and FTE. These models estimate the effect of funding on volunteerism that does not operate through adjustments in these other variables. Inclusion of annual visitation controls for changes in a park's popularity over time, due perhaps to anniversary years and promotions that could simultaneously affect funding, volunteerism, and visitation. The inclusion of paid FTE per 1,000 visitors is intended to control for the way that changes in supervisory constraints within a park could potentially affect volunteerism. A positive or negative sign of this effect also suggests that volunteer hours and permanent FTE are complements or substitutes, respectively.

5.2 Heterogeneous Effects

We next examine heterogeneous crowding effects across parks and program types. We begin by testing whether the crowding effect differs between environmental and non-environmental parks. The estimating equation differs only by the inclusion of an interaction term between $Budget_{it}$ and an indicator variable $1[Envr]_i$ that takes the value of 1 for environmental parks. The full model is

$$Hours_{it} = \beta Budget_{it} + \lambda Budget_{it} \times 1[Envr]_i + \gamma X_{it} + \alpha_i + \sigma_{rt} + \varepsilon_{it}. \tag{13}$$

In this case, β and $\beta + \lambda$ provide estimates of the crowding effect in non-environmental and environmental parks, respectively. If $\lambda \neq 0$, then the effect differs between the two types of parks. Testing for the difference is of interest because of the conjecture that volunteerism in an environmental park is more likely to be associated with private provision of an impure public good. Empirical evidence of $\lambda \neq 0$ would therefore support the notion that individual behavior and therefore crowding effects differ in the extent of impure altruism. In either case, a finding of crowding in or greater than one-for-one crowding out would also enable rejection of pure altruism as the underlying motivation for volunteerism in either type of park.

We further examine heterogeneous effects in two ways. First, we examine how crowding effects differ between volunteer activities that are likely to occur outside or inside. Here again the motivating assumption is that volunteer hours focused on outside activities are more likely associated with joint production involving park enjoyment than those taking place inside. For example, the experiences of volunteers who lead guided hiking tours is different from those of volunteers who assist with office administration. Specifically, we estimate the benchmark specification (12), but replace aggregate volunteer hours as the dependent variable with volunteer hours that take place either outside or inside, and estimate the equations separately.[26] Second, we estimate the heterogeneous effects specified in equation (13) with outdoor or indoor hours separately as the dependent variable. This combines the two previous approaches to examine whether the crowding effect differs both by park type (environmental *vs.* non-environmental) and by hour type (outside *vs.* inside).

5.3 IV Strategy

Even after including the park covariates and region-year fixed effects, one might be concerned about potential endogeneity that would bias the crowding effect estimates in equations (12) and (13). The concern is that park funding and volunteer hours might be correlated with some unobserved, time-varying park characteristic. For example, a downward trend in park funding that increases the park's deferred maintenance backlog might also increase the demand for park volunteers (NPS 2019b). This would bias the estimates

[26]Outside hours consist of the categories archeology, campground hosting, interpretation, resource management, and park protection. Inside hours consist of the categories general management, curating, administration, maintenance, training, and other.

downward because unobserved maintenance needs are positively correlated with volunteerism and negatively correlated with funding. Other plausible scenarios might bias estimates of the crowding effect in the other direction. For example, the unobserved introduction of new park programs could simultaneously create additional volunteer opportunities and funding requirements.

Addressing concerns about endogeneity requires an instrument that is correlated with the annual budget for each park, but uncorrelated with other determinants of volunteerism, conditional on the model covariates. We construct an instrument that plausibly meets these two requirements by combining cross-sectional variation in historical programmatic and operational budgetary needs with aggregate shifts to public funding in a manner that is commonly used in the literature on trade and local labor markets (Bartik 1991; Autor, Dorn and Hanson, 2013; Notowidigdo 2019). Specifically, we simulate annual park funding using the interaction of each park's funding levels three years prior to the start of our panel with changes in the average LCV score for the U.S. Congressional appropriations committees, excluding members that represent the state or district where the park is located.

The instrument is:

$$Z_{it} = Budget_{i,1995} \times \frac{\overline{LCV}_{i,t-1}}{\overline{LCV}_{i,1995}}, \tag{14}$$

where $Budget_{i,1995}$ is each park's federal budget appropriation in 1995, and $\overline{LCV}_{i,t}$ is the average of the LCV scores of the two committees in year t, excluding the elected official(s) of park i.[27] The $Budget_{i,1995}$ weights capture differences in each park's base budgetary needs due to their size or historical program scope. The ratio $\overline{LCV}_{i,t-1}/\overline{LCV}_{i,1995}$ captures the year-to-year change in the extent to which the key congressional committees prioritize environmental and conservation legislation.[28] The IV strategy is therefore based on the theory that congressional members that prioritize these objectives will allocate higher levels of funding to the NPS, so that the identifying variation comes from changes in park budgets that arise from changes in general congressional support for conservation. We use the LCV scores for the House and the Senate appropriations committees because their members have the most significant influence on federal budget appropriations in each

[27] In particular, we calculate the average among members in the House and Senate committees separately, and then take the average of the two committees to estimate $\overline{LCV}_{i,t}$ for each year. In any given year, the appropriations committees consist of approximately 60 members in the House and 30 members in the Senate.

[28] We choose the base year 1995 because it is the earliest year for which we observe funding. While the choice of other possible base years changes the coefficient estimates in the first stage, it has no effect on the second stage IV estimates. There are, however, 36 parks in our sample that do not exist in 1995, so there is no $Budget_{i,1995}$ for these observations. In these cases, we use the first year of budget data available, and include observations beginning three years after the designation of these parks. Excluding or including these observations does not meaningfully change any of the results.

year. We lag the LCV score because the budget approved in year $t-1$ is for use in year t. Figure 5 shows the time series variation of the mean LCV score, and referring back to Figure 3, the trend appears to roughly track the variation in NPS funding.

We use Z_{it} to instrument for $Budget_{it}$ in specifications (12) and (13), and $Z_{it} \times 1[Envr]_{it}$ as an instrument for $Budget_{it} \times 1[Envr]_i$ in specification (13). Table 2 reports the first stage results, and we find that the instrument is highly correlated with actual park funding.[29] Column 1 shows that a 10 percent increase in LCV score increases park budgets by 2 percent relative to their budget in 1995.

The identifying assumption is that this interaction of parks' initial budgets with annual changes in congressional LCV scores is independent of any unobserved shocks to park volunteerism, conditional on the model covariates. This assumption would be violated if volunteers systematically redirect their efforts to different parks within a region in response to changes in the LCV scores of Congressional members in other districts or states, for example. The inclusion of region-year fixed effects controls for any trends in environmentalism that could be correlated with both volunteerism in national parks and political preferences. For each park, the instrument also excludes members of Congress representing their state or district to reduce any concern about factors affecting both park volunteerism and local elected officials' preferences. Reassuringly, however, 78.4 percent of the NPS volunteers report that their primary motivation is interest in specific parks or projects, along with the overall NPS mission (NPS 2007).

As part of robustness checks, we also discuss and use alternative instruments, each of which scales each park's initial funding by other variables that predict aggregate NPS funding as in (14). These other shifters of park funding include LCV scores for the all members of the U.S. House and Senate (rather than just the appropriation committees), annual funding for the Department of the Interior's Fish and Wildlife Service, and the annual NPS funding for all regions excluding the region for each park. Using these alternative instruments also allows us to conduct over-identification tests of their exogeneity, which we find support the validity of the instruments at conventional levels (Goldsmith-Pinkham, Sorkin, and Smith 2020).

6 Estimation Results

We now turn to our estimates of the effect of changes in park funding on within-park volunteerism. We consider the overall, average estimates of the crowding effect, heterogene-

[29]The cluster-robust Kleibergen-Paap rk Wald F-statistic for a weak instrument shows that while the instrument is highly correlated with park funding across all parks, it is less correlated with the non-environmental parks (column 2). Relative to the benchmark specification (12), this produces a lower test statistic for joint identification in the two first-stage regressions for specification (13) ($F = 6.4$). This means that the heterogeneous IV estimates of specification (13) may be more biased toward OLS.

ity between park and volunteer hour types, a range of robustness checks, and alternative explanations for our results on crowding in.

6.1 Crowding Out or In?

The estimates in Table 3 address the question of whether increases in park funding crowd out or crowd in volunteerism on average across parks. The first three columns report the OLS estimates of specification (12). Model (1) includes only $Budget_{it}$ along with the park and region-year fixed effects, model (2) includes the additional variables of visits and FTE per 1,000 visits in the previous year, and model (3) is the same though estimated excluding the post-recession years 2011-2013. We find positive and statistically significant effects of park budgets on volunteerism across all three models, and the estimated magnitudes are very similar. Recall that the positive coefficients are consistent with crowding in, rather than crowding out. The OLS estimates suggest that a $1,000 increase in a park's annual budget is associated with an increase in volunteerism of approximately 2.5 hours on average.

Our preferred estimates of the average crowding effect are the results of the IV estimation in the last three columns of Table 3. These models correct for potential endogeneity of park budgets using the IV strategy described above. The three models differ in parallel fashion with those in the first three columns. Here again the results are consistent with crowding in, and the magnitudes are larger and more precisely estimated. Focusing on the results of model (5), which includes the covariates, we find that a $1,000 increase in federal funding causes an average increase in volunteerism of approximately 12.1 hours. We also find that the results are economically meaningful. To quantify the average value of crowding in, we multiply the estimated change in hours by the average hourly valuation of volunteerism over the years 1998 through 2013. As described previously, this estimate is $21.85 per hour. Accordingly, based on our preferred model (5), a $1,000 increase in a park's federal appropriation crowds in volunteerism that is worth $265 per year on average. This means that an additional dollar of federal funding within a park crowds in roughly 27 cents worth of volunteerism, where the benefit of crowding in occurs over and above the direct benefit of the marginal dollar of park funding. As noted previously, one reason why the magnitudes of the IV estimates might be larger than the OLS estimates is that lower levels of funding add to the cumulative deferred maintenance in parks, and this increases the set of potential tasks for volunteers.[30]

The number of visits has no statistically significant effect on volunteerism, but the FTE per 1,000 visits does. With greater paid staff per visitor, there is less volunteerism, and the

[30] In 2017, the deferred maintenance backlog in the NPS was estimated at $11.61 billion in current dollars (NPS 2019b).

magnitude is such that one additional FTE is associated with roughly 50 fewer hours of volunteerism. The sign of this effect is important because it suggests that paid staff and volunteerism are substitutes rather than complements when it comes to park management, or that some volunteers are eventually hired as paid staff. This is a subject to which we return below, but it is worth noting here that the finding helps to rule out an alternative explanation for crowding in, whereby greater budgets result in more paid staff, who are then able to recruit and manage more volunteers.[31]

In sum, we find evidence consistent with a mechanism whereby greater funding results in better parks that are more enjoyable places in which to volunteer. We have shown how the presence of such joint production readily admits the possibility for crowding in, whereas the model for private provision of a pure public good does not. We therefore interpret the results of crowding in, along with the differences between the OLS and IV estimates, as consistent with the private provision of an impure public good (i.e., impure altruism), but not a pure public good.

6.2 Heterogeneous Effects

We now test for heterogeneous crowding effects across park and hour types. We begin with potential differences between parks designated as either environmental or non-environmental. Recall our conjecture that the joint production of volunteerism is greater in the environmental parks, where individuals are more likely to experience *in situ* benefits of park enjoyment (i.e., recreating in a natural environment) while contributing to the public good.

Table 4 reports the estimates of specification (13), and we again show the OLS and IV results. In all models, we find positive and statistically significant coefficient estimates on the interaction $Budget_{it} \times 1[Envr]_i$, which indicates more crowding in for environmental parks than for non-environmental parks. Focusing on the IV estimates, particularly those for model (5), which includes park covariates, the magnitude of the difference indicates that a $1,000 increase in park funding crowds in 12.1 more hours in the environmental parks on average. The overall effect in environmental parks, which is the sum of the first two coefficients, is 12.8 hours, which has an equivalent monetary value of $279. That is, each additional dollar of park funding crowds in roughly 28 cents worth of volunteerism. In contrast, the crowding effect in non-environmental parks is statistically indistinguishable from zero. Based on our theoretical results for distinguishing between models based on a finding of crowding in, the results suggest that volunteerism in environmen-

[31] The substitutability, rather than complementarity, between paid staff and volunteers is also echoed in survey responses as part of the 2007 VIP Program Assessment Report, where one staff member states in response to budget shortfalls that, "In my 16-plus years, I have seen the volunteer role evolve to replace many of the functions that National Park Service staff previously performed" (NPS 2007).

tal parks is consistent with private provision of an impure public good, but not a pure public good. However, no distinction between the models is possible for volunteerism in non-environmental parks.[32]

Turning now to heterogeneous effects by hour type, Table 5 reports estimates of specification (12) separately for outside and inside hours in panels A and B, respectively.[33] Focusing on the IV estimates, we find evidence of crowding in for hours of both types, although the magnitudes and precision are greater for outside hours. Based on the preferred models in column (4), we find crowding in of 9.1 outside hours, compared to crowding in of 3.0 inside hours. These translate to monetary equivalent measures of 20 cents and 7 cents worth of volunteerism per dollar of federal budgeting, respectively. The difference between the two estimates—and the fact that both indicate crowding in—remains consistent with the general notion that environmental parks and outdoor hours create circumstances where joint production of volunteerism is likely to be greatest.

The final set of heterogeneous effects that we examine is a two-way analysis, where we estimate the environmental versus non-environmental park effect separately for outside and inside hours. Table 6 reports estimates of specification (13) for outside and inside hours separately in panels A and B, respectively. Focusing on the preferred estimates in column (4), we find evidence that environmental parks differ from non-environmental parks in the direction of more crowding in for both outside and inside hours. Moreover, as expected, the magnitudes of the point estimates suggest greater crowding in for outside hours than for inside hours, at 8.0 versus 4.1 hours per $1,000 of funding. Nevertheless, among the four different cases of the two-way analysis, the overall estimate of crowding in is statistically significant at conventional levels in only one case: outside hours in environmental parks, with crowding in of 9.6 hours per $1,000 of funding, and a 95 percent confidence interval that ranges between 1.6 and 17.7 hours.

6.3 Robustness Checks

We estimate a series of alternative models to demonstrate robustness of our results on crowding in and heterogeneity. In particular, we consider lags of explanatory variables, the addition of a linear time trend that differs between large and small parks, a broader definition of environmental parks, and alternative instruments in our IV strategy.[34] We report the full set of results in Appendix tables and summarize the key findings here.

[32]Separately estimating equation (12) for environmental and non-environmental parks yields estimates that are similar in sign and magnitude to the estimates implied by the heterogeneous effects regression (13), though less precisely estimated as expected due to the split sample.

[33]We no longer report estimates that exclude the post-recession years for two reasons: the results do not change in any meaningful way, and we prefer to focus on results that take advantage of the full data set.

[34]Note that the results reported earlier where we exclude the post-recession years (2011-2013) should also be considered robustness checks. We found that dropping these years did not meaningfully change the results.

We first consider a one-year lag of the instrumented annual park budgets. Central to our analysis is the idea that greater funding improves parks in ways that may affect volunteerism, but many funded initiatives take time to complete or implement. Accordingly, one could argue that funding in previous years is a preferable measure of improvements in park quality, and we therefore estimate specifications (12) and (13) using an instrumented version of the lagged budget, $Budget_{i,t-1}$, on the right-hand side rather than the contemporaneous year's budget (Table A.1, columns 1 and 2). One consequence of using the lagged variable is having fewer observations upon which to estimate the model. Nevertheless, we again find statistically significant crowding in, and the magnitudes are greater than those estimated previously. For the overall effect across parks, we find crowding in of 13.4 hours per $1,000 of funding on average, or equivalently 30 cents worth of volunteerism for an additional dollar of funding. The estimated effect for environmental parks is similar in sign, magnitude and precision to the overall estimate of crowding in, and we again find no statistically significant crowding effect in non-environmental parks.

We also estimate models that include the additional controls of two-year lags of park visitation. Visitation from prior years is the only information available to Congress when the NPS budget must be approved.[35] Accounting for the popularity of parks in prior years might also capture how popularity affects volunteerism, which in many cases may require planning far in advance. It turns out that the visitation variables themselves have statistically insignificant effects, and we find that the estimates of crowding in remain very similar to the main estimates (see Table A.1, columns 3 and 4).[36]

A further robustness check is to assess whether the crowding effect is driven primarily by large parks. Our first test of this hypothesis is to include separate linear time trends for parks with budgets above and below the median in 1995. This accounts for the possibility that unobserved funding trends may differ over time in large and small parks. Here again we find that the results are very similar to those already reported (see Table A.1, columns 5 and 6). We also test this hypothesis more directly by including an interaction term in equation (12) that allows the effect of park budgets to differ in parks with above or below median budgets. This interaction term is small in magnitude and only marginally statistically significant, and suggests if anything a slightly smaller crowding effect in large parks (Table A.2, column 7).

Our tests for heterogeneous effects based on park type relied on a particular definition for environmental parks (i.e., whether the park engages volunteers in natural resource management every year). This definition yields 105 environmental parks that, on average,

[35]In addition, the way that our instrument exploits Congressional voting behavior the year before volunteerism is observed also reduces any concern about reverse causality.

[36]We also find quantitatively similar estimates in models that exclude visits entirely (i.e., control only for lagged FTE) and models that include two-period lags of both visits and FTE.

have more volunteers, larger budgets, greater visitation, and fewer FTE per visitor than non-environmental parks (see Table 1). As a robustness check, we consider an alternative definition of an environmental park based on each park's official mandate. We reviewed the descriptions of each park's activities on its official webpage and classified parks as environmental if they list natural resource management as part of their mission, regardless of whether conservation is a primary or secondary objective.[37] The result is a broader definition of environmental parks, accounting for 233 of all 326 parks, and nesting 99 of the 105 included in the previous definition. Table A.2 includes the descriptive statistics for parks of both types based on this alternative definition. We find that testing for heterogeneous effects using this definition of environmental parks in specification (13) results in a similar estimate of crowding in for environmental parks, at 12.5 hours compared to 12.1 hours previously, and still no statistically significant effect in non-environmental parks (Table A.1, column 8).

Finally, we consider alternative instruments, while still using the basic shift-share approach in equation (14). All results are reported in Table A.3, though we note that our preferred instrument is a better predictor of funding in the heterogeneous effects models. In each case, we scale a park's base year funding with the change in a variable that predicts aggregate changes in NPS funding. First, we use the annual shift in the LCV score for all members of Congress in other states and districts, averaged across the whole House and Senate, rather than only the appropriations committees. We find very similar estimates of crowding in for both specifications (12) and (13). Second, in place of LCV scores, we use the shift in the annual budget of the U.S. Fish and Wildlife Service, which is another Bureau within the U.S. Department of the Interior with a similar environmental orientation as the NPS.[38] This approach is based on actual funding decisions that are plausibly related to funding levels in the NPS, and we again find results that are similar in sign, magnitude, and precision. Finally, for each park i, we use the total NPS budget in each year, exclusive of funding for the region in which park i is located. With this approach, the scaling of each park's budget differs among the seven NPS regions, and the magnitude of the crowding in estimates are somewhat smaller but have the same qualitative interpretation.

[37] For example, National Battlefields, such as Antietam National Battlefield, are established with the primary goal of commemorating specific historic events. However, Antietam National Battlefield is classified as an environmental park based on its mandate because it additionally records and preserves local wildlife and vegetation, monitors invasive species and other pests, and reforests the land on which it is established (NPS 2019a).

[38] Data on the total budgets for all Bureaus within the Department of the Interior are available in the annual Greenbooks. In this case, the ratio in equation (14) does not differ by park.

6.4 Alternative Explanations

Might there be alternative explanations for crowding in of volunteerism in national parks? One possibility for crowding in that has been raised in the literature more broadly is a signaling explanation. For example, Payne (2001) finds that greater public funding can signal greater governmental approval of a recipient organization, which, in turn, promotes an increase in private contributions. In a related finding, Khanna and Sandler (2000) show how private contributions might be greater for institutions perceived as subject to more stringent governmental oversight. We argue that a signaling explanation is unlikely to explain the crowding in of volunteerism in national parks for several reasons. First, the park fixed effects control for any persistent quality signal among parks, which is likely to be more relevant given the opacity of annual budgeting. We find it easier to argue that volunteers are more likely to observe park or program quality than annual fluctuations in a park's budget. Second, the inclusion of park visitation in our regression models controls for quality signaling that would plausibly affect both visitors and volunteers. Finally, and perhaps most important, all of the parks in our analysis are subject to NPS authority and thereby face a similar set of oversight standards for management and accountability. Indeed, it is the relative homogeneity across our units of analysis that provides a distinct advantage compared to many of the other studies in the literature that seek to estimate crowding effects across a range of heterogeneous charitable organizations.

Another candidate explanation relates to the possibility of endogenous fundraising, whereby charitable organizations adjust their own solicitations of individual and corporate donations in response to changes in public funding (Andreoni and Payne 2001, 2003). While fundraising plays a small role in the NPS, concerns might arise in our setting about the ability of park staff to recruit and manage volunteers. In particular, might the estimated crowding in be the result of greater funding relaxing a binding constraint on the ability of park managers to recruit and supervise volunteers? Here again, there are several reasons why we believe this mechanism is not explaining our results. First, we include the permanent FTE per visitor as an explanatory variable in the regression models, which controls for variation in staff availability to recruit volunteers. In fact, we find robust evidence that more FTE per visitor is negatively associated with the number of volunteer hours within a park. Second, an evaluation of the VIP program provides broad survey evidence that supervisory capacity is not generally a binding constraint (NPS 2007).[39] Third, the VIP program is regionally funded before allocations are made to individual parks, and this

[39]In particular, the survey finds that 88 percent of the volunteers "do not need more of their supervisor's time" and 89 percent are at least moderately "satisfied with the leadership, management, and support that [they] receive as volunteers," with 77 percent reporting strong satisfaction. Moreover, only 16 percent of the staff agree with the statement that "volunteers are not given the necessary attention and direction throughout their assignments" (NPS 2007).

means that the region-year fixed effects would control for any common, regional funding constraints for other complementary inputs (NPS 2005).[40] Fourth, even the largest parks employ only a handful of designated VIP program administrators, and this suggests that increases in operational funding are unlikely to be used for expanding the VIP program staff within parks (NPS 2017b). Finally, our estimates of heterogeneous effects would mean that supervisory constraints or constraints on other inputs would need to apply differently to environmental and non-environmental parks, along with outside and inside hours within the same park. While possible, we see no compelling reason why such differences would arise.

7 Conclusion

This paper makes theoretical and empirical contributions to the literature on private provision of public goods. Understanding the ways in which public provision affects private provision—through either crowding in or crowding out—is fundamental for evaluating the positive and normative consequences of policies that affect the supply of public goods. Moreover, within the literature on privately provided public goods, estimates of the extent of crowding out are the foundation for tests between the candidate models of underlying behavior based on pure or impure altruism. This paper identifies limitations of the standard specification test based on crowding out and proposes a more general alternative test based on the presence crowding in or greater than one-for-one crowding out. We also provide empirical evidence consistent with the new test in a setting where impure altruism appears to apply *ex ante*, because of joint production of private and public benefits.

Our theoretical analysis revisits the crowding out conditions in Andreoni's (1989, 1990) seminal model of impure altruism. In contrast with the pure public good model, we show that crowding out with impure altruism is an assumed result, rather than a consequence of well-understood properties of individual utility functions. In particular, we show how assuming normality of goods readily admits possibilities that are implicitly assumed away: crowding in and greater than one-for-one crowding out.

These results have especially important implications for the specification test between pure and impure altruism that is employed extensively in the literature. Contrary to the conventional wisdom that crowding out must be less for impure altruism than for pure altruism, we show that this need not be the case. The standard results will always hold under the assumptions of additive separability and strict concavity of utility functions,

[40] We have sought to obtain detailed data on park-level VIP budgets and other spending in each year, but unfortunately the NPS does not keep track of these data in regional offices. Moreover, VIP program funding is only about 0.1% of the NPS budget, and volunteer activities are typically not tied to specific programs or personnel funded by the VIP annual budgets.

but more generally, and even with the seminal model's original setup, crowding out with impure altruism can be greater than or less than with pure altruism. We show that the different cases depend in part on the degree of substitutability or complementarity between the public good and the private benefit associated with one's own provision. As an alternative, we propose a new specification test based on crowding in or greater than one-for-one crowding out: evidence of either is consistent with impure altruism, but not with pure altruism.

The empirical portion of the paper focuses on estimating the causal effect of changes in public funding in U.S. national parks on within-park volunteerism. We find robust evidence of crowding in across a range of specifications and identification strategies, relying on OLS and IV estimates, and heterogeneous effects. The overall finding is that each additional dollar of public expenditure within a park crowds in 27 cents worth of volunteerism on average. Our empirical setting and data set have unique advantages for estimating crowding effects because of the relatively long panel of homogeneously managed units, volunteers that tend to donate their time at only one site, and federal funds that account for almost all of park budgets. From a policy perspective, our findings suggest that greater budgets for the NPS meaningfully leverage greater private contributions, and this may be an important consideration for a government agency that suffers from perennial budgetary shortfalls and a growing backlog of deferred maintenance. More generally, the theoretical possibility of crowding in and our empirical evidence of it suggest that accounting for crowding in is important for fully measuring the marginal benefit of public provision of a public good.

Finally, we have argued that volunteerism in national parks, which involves giving time in parks rather than money, is consistent with private provision of an impure public good and therefore the model of impure altruism. The argument is based on how volunteerism jointly contributes to park conservation and management (a public good) and *in situ* park enjoyment (a private benefit). In the context of our theoretical results and new specification test, the findings of crowding in support the notion that volunteerism is driven by the underlying model of impure altruism rather than pure altruism. The empirical findings thus provide well-identified, quasi-experimental evidence of crowding in that is theoretically consistent with the standard model for private provision of a public good.

References

Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy* 97 (6): 1447-1458.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal*, 100 (401): 464-477.

Andreoni, James. 1993. "An Experimental Test of the Public Goods Crowding-Out Hypothesis." *American Economic Review* 83 (5): 1317-1327.

Andreoni, James. 2006. "Philanthropy." In *Handbook of Giving, Reciprocity and Altruism*, Vol. 2, edited by Serge-Christophe Kolm and Jean Mercier Ythier. North-Holland: Elsevier.

Andreoni, James and A. Abigail Payne. 2003. "Do Government Grants to Private Charities Crowd Out Giving or Fund-raising?" *American Economic Review* 93 (3): 792-812.

Andreoni, James and A. Abigail Payne. 2011. "Is Crowding Out Due Entirely to Fundraising? Evidence from a Panel of Charities." *Journal of Public Economics* 95 (5-6): 334-343.

Andreoni, James, A. Abigail Payne and S. Smith. 2014. "Do Grants to Charities Crowd Out Other Income? Evidence from the U.K." *Journal of Public Economics* 114: 75-86.

Andreoni, James and A. Abigail Payne. 2013. "Charitable Giving." In *Handbook of Public Economics*, Vol. 5, edited by Alan J. Auerbach, Raj Chetty, Martin Feldstein,and Emmanuel Saez. Amsterdam: Elsevier.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121-2168.

Bartik, Timothy. 1991. "Who Benefits from State and Local Economic Development Policies?" W.E. Upjohn Institute for Employment Research: Kalamazoo, MI.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian. 1986. "On the Private Provision of Public Goods." *Journal of Public Economics* 29 (1): 25-49.

Borgonovi, Francesca. 2006. "Do Public Grants to American Theaters Crowd-out Private Donations?" *Public Choice* 126 (3/4): 429-451.

Brown, Alexander, Jonathan Meer, and J. Forrest Williams. 2017 "Why Do People Volunteer? An Experimental Analysis of Preferences for Time Donations," *Management Science* 65 (4): 1455-1468.

Chan, Nathan W. and Matthew J. Kotchen. 2014. "A Generalized Impure Public Good and Linear Characteristics Model of Green Consumption." *Resource and Energy Economics* 37: 1-16.

Comay, Laura B. 2013. "National Parks System: What Do the Different Park Titles Signify?" Report 41816. Washington, DC: Congressional Research Service.

Cornes, Richard and Todd Sandler. 1984. "Easy Riders, Joint Production, and Public Goods." *Economic Journal* 94 (375): 580-598.

Cornes, Richard and Todd Sandler. 1985. “The Simple Analytics of Pure Public Good Provision,” *Economica* 52(205): 103-116.

Cornes, Richard and Todd Sandler. 1994. “The Comparative Static Properties of the Impure Public Good Model.” *Journal of Public Economics* 54 (3): 403-421.

Cornes, Richard and Todd Sandler. 1996. *The Theory of Externalities, Public Goods, and Club Goods.* Cambridge University Press.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2016. “Estimating Social Preference and Gift Exchange at Work.” NBER Working Paper No. 22043.

de Wit, Arjen and Rene Bekkers. 2017. “Government Support and Charitable Donations: A Meta-Analysis of the Crowding-out Hypothesis.” *Journal of Public Administration Research and Theory* 27 (2): 301-319.

Eckel, Catherine C., Philip J. Grossman, and Rachel M. Johnston. 2005. “An Experimental Test of the Crowding Out Hypothesis.” *Journal of Public Economics* 89 (8): 1543-1560.

Government Accountability Office (GAO). 2004 “National Park Foundation: Better Communication of Roles and Responsibilities Is Needed to Strengthen Partnership with the National Park Service.” Report GAO-04-541, United States Government Accountability Office.

Government Accountability Office (GAO). 2009. “Donations and Related Partnerships Benefit Parks, but Management Refinements Could Better Target Risks and Enhance Accountability.” Report GAO-09-386, United States Government Accountability Office.

Glazer, Amihai and Kai A. Konrad. 1996. “A Signaling Explanation for Charity.” *American Economic Review* 86 (4): 1019-1028

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. “Bartik Instruments: What, When, Why, and How.” *American Economic Review* 110 (8): 2586-2624.

Harbaugh, William T. 1998. “The Prestige Motive for Making Charitable Transfers.” *American Economic Review* 88 (2): 277-282.

Hines, James R. and Richard H. Thaler. 1995. “The Flypaper Effect.” *Journal of Economic Perspectives* 9 (4): 217-226.

Hollander, Heinz. 1990. “A Social Exchange Approach to Voluntary Cooperation.” *American Economic Review* 80 (5): 1157-1167.

Hungerman, Daniel M. 2005. “Are Church and State Substitutes? Evidence from the 1996 Welfare Reform.” *Journal of Public Economics* 89 (11-12): 2245-2267.

Independent Sector. 2019. “The Value of Volunteer Time.” `https://www.independentsector.org/resource/the-value-of-volunteer-time/` (accessed February 25, 2019).

Khanna, Jyoti and Todd Sandler. 2000. “Partners in Giving: The Crowding in Effects of U.K. Government Grants.” *European Economic Review* 44 (8): 1543-1556.

Kingma, Bruce R. 1989. "An Accurate Measurement of the Crowd-out Effect, Income Effect, and Price Effect for Charitable Contributions." *Journal of Political Economy* 97 (5): 1197-1207.

Kotchen, Matthew J. 2005. "Impure Public Goods and the Comparative Statics of Environmentally Friendly Consumption." *Journal of Environmental Economics and Management* 49 (2): 281-300.

Kotchen, Matthew J. 2006. "Green Markets and the Private Provision of Public Goods." *Journal of Political Economy* 144 (4): 816-834.

Kotchen, Matthew J. 2007. "Equilibrium Existence and Uniqueness in Impure Public Good Models." *Economic Letters* 97: 91-96.

Kotchen, Matthew J. and Katherine R. H. Wagner. 2019. "Crowding In with Impure Altruism: Theory and Evidence from Volunteerism in National Parks." NBER Working Paper 26445. Cambridge, MA.

League of Conservation Voters (LCV). 2019. "National Environmental Scorecard Methodology." `http://scorecard.lcv.org/methodology/` (accessed January 16, 2019).

Lilley, Andrew and Robert Slonim. 2014 "The Price of Warm Glow," *Journal of Public Economics*, 114: 58-74.

Madden, Paul. 1991. "A Generalization of Hicksian q Substitutes and Complements with Application to Demand Rationing." *Econometrica*, 59(5) 1497-1508.

Murdoch, James C. and Todd Sandler. 1984. "Complementarity, Free Riding, and the Military Expenditures of NATO Allies," *Journal of Public Economics*. 25(1-2) 83-101.

Murdoch, James C. and Todd Sandler. 1990. "Nash-Cournot or Lindahl Behavior? An Empirical Test for the NATO Allies," *Quarterly Journal of Economics*. 105(4) 875-894.

National Park Service (NPS). 2005. "Director's Order #7: Volunteers-In-Parks." `https://www.nps.gov/policy/DOrders/DO-7.htm/` (accessed May 5, 2017).

National Park Service (NPS). 2007. "VIP Program Assessment Report." United States Department of the Interior, Washington, DC.

National Park Service (NPS). 2017a. "NPS Essentials." `https://www.nps.gov/training/essentials/html/volunteers.html` (accessed May 6, 2017).

National Park Service (NPS). 2017b. "Yosemite National Park Volunteers in Parks Handbook." `https://www.nps.gov/yose/getinvolved/supportyourpark/upload/viphandbook.pdf` (accessed May 4, 2017).

National Park Service (NPS). 2019a. "Antietam Nature and Science," `https://www.nps.gov/anti/learn/nature/index.htm/` (accessed February 26, 2019).

National Park Service (NPS). 2019b. "What is Deferred Maintenance?" `https://www.nps.gov/subjects/infrastructure/deferred-maintenance.htm/` (February 28, 2019).

National Parks Conservation Association (NPCA). 2019. "How Is the Partial Government Shutdown Affecting National Parks?" https://www.npca.org/articles/2029-how-is-the-partial-government-shutdown-affecting-national-parks (accessed on March 11, 2019).

Notowidigdo, Matthew. 2019. "The Incidence of Local Labor Demand Shocks." Forthcoming, *Journal of Labor Economics*.

Okten, Cagla and Weisbrod, Burton A. 2000. "Determinants of Donations in Private Non-Profit Markets." *Journal of Public Economics* 75 (2): 255-272.

Ottoni-Wilhelm, Mark, Lise Vesterlund, and Huan Xie. 2017. "Why Do People Give? Testing Pure and Impure Altruism." *American Economic Review* 107 (11): 3617-3633.

Mas, Alexandre and Amanda Pallais. 2017. "Measuring the Value of Alternative Work Arrangements." *American Economic Review* 107(12): 3722-3759.

Payne, A. Abigail. 2001. "Measuring the Effect of Federal Research Funding on Private Donations at Research Universities: Is Federal Research Funding More Than a Substitute for Private Donations?" *International Tax and Public Finance* 8 (5-6): 731-51.

Payne, A. Abigail. 2009. "Does Government Funding Change Behavior? An Empirical Analysis of Crowding Out." In *Tax Policy and the Economy*, Vol. 23, edited by Jeffrey R. Brown and James M. Poterba. University of Chicago Press.

Ribar, David C. and Mark Ottoni-Wilhelm. 2002. "Altruistic and Joy-of-Giving Motivations in Charitable Behavior." *Journal of Political Economy* 110 (2): 425-457.

Roberts, Russell D. 1984. "A Positive Model of Private Charity and Public Transfers." *Journal of Political Economy* 92 (1): 136-148.

Turner, Robert W. and Reed Walker. (2006). "Explaining the Determinants of Operating Budgets at U.S. National Parks." *Public Finance Review* 34 (5): 551-573.

Warr, Peter G. 1982. "Pareto Optimal Redistribution and Private Charity." *Journal of Public Economics* 19: 131-138.

Figures and Tables

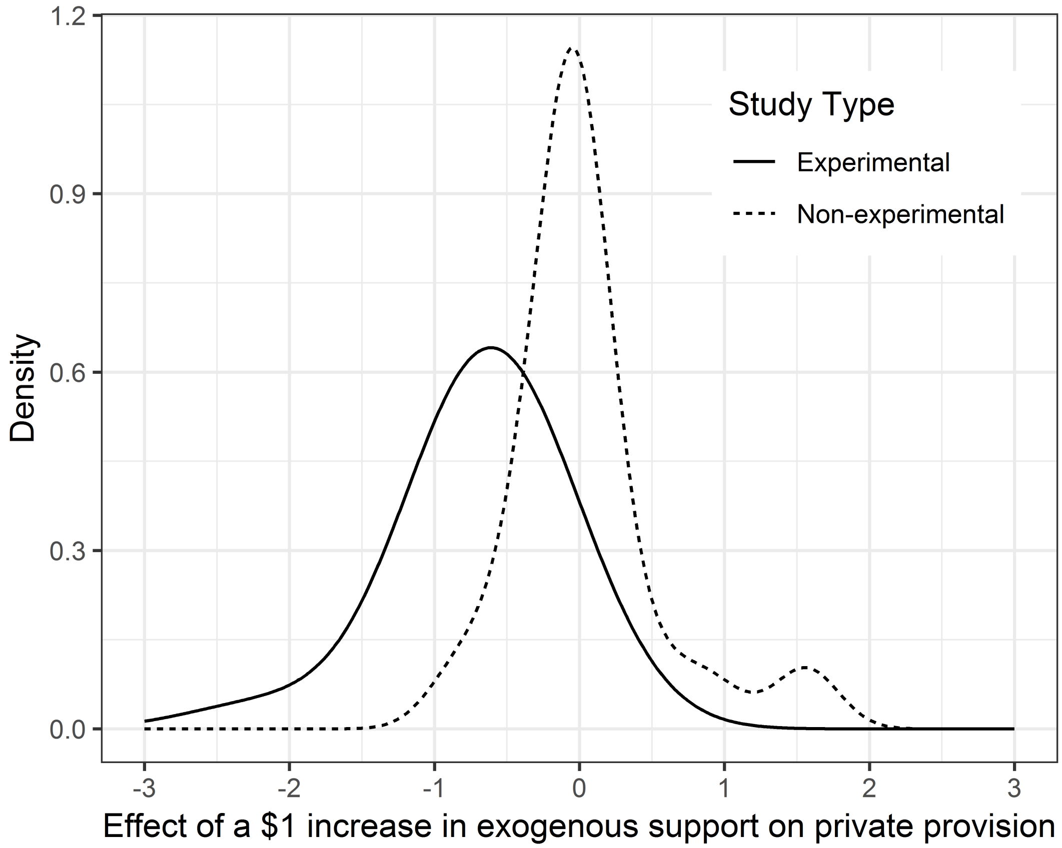

Figure 1: Kernel density function of crowding effect estimates (negative is crowding out, positive is crowding in) from the literature, shown separately for experimental and non-experimental studies, and based on 325 estimates from 54 studies collected by de Wit and Bekkers (2017).

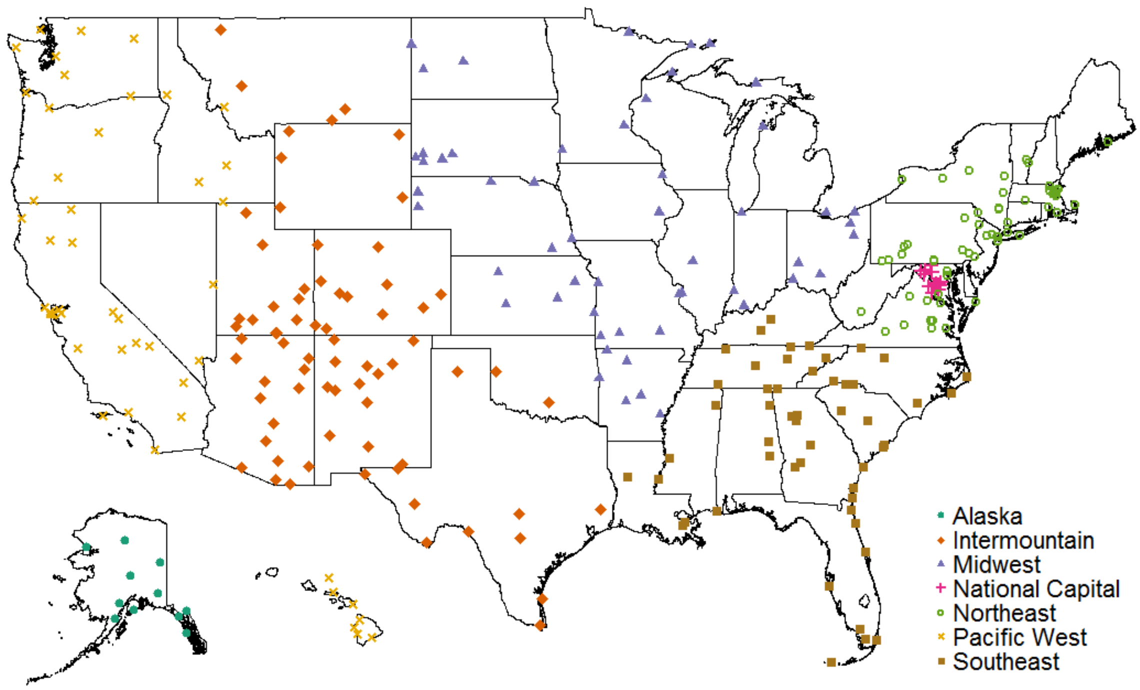

Figure 2: Geographic distribution of the 326 parks used in the analysis and managed by the National Park Service, by administrative region, excluding 5 parks in American Samoa, the Mariana Islands, and Puerto Rico.

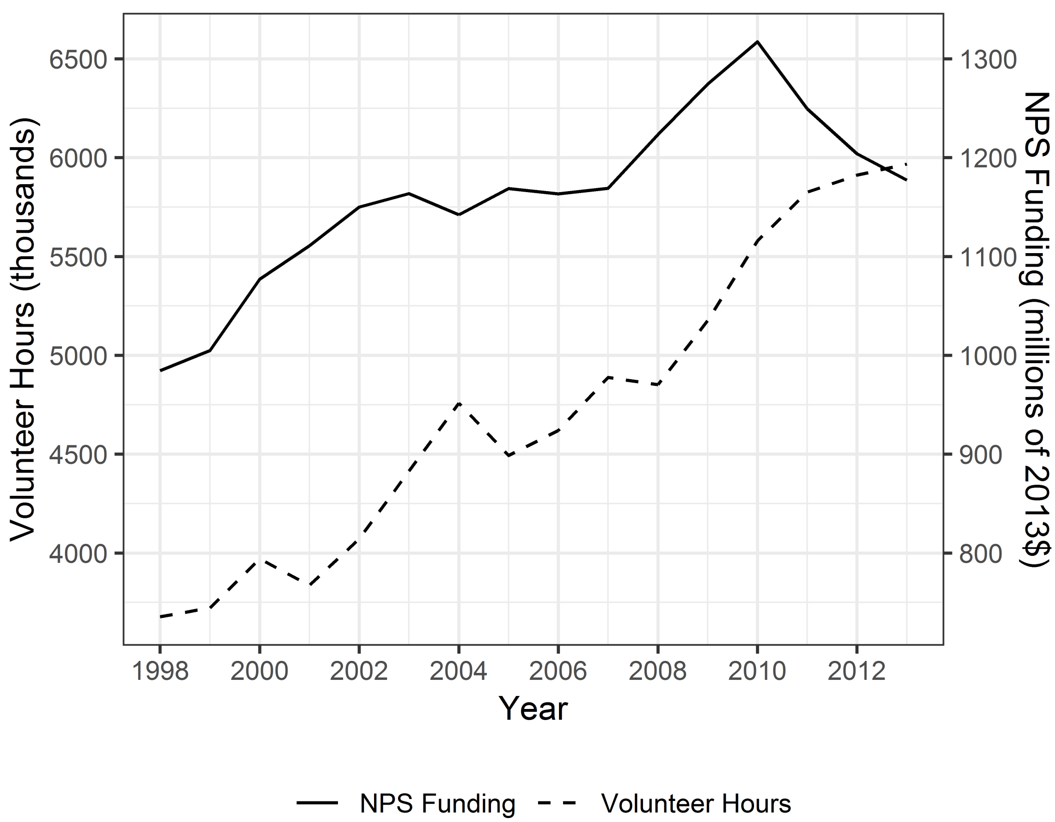

Figure 3: Trend in annual National Park Service volunteer hours and funding to parks

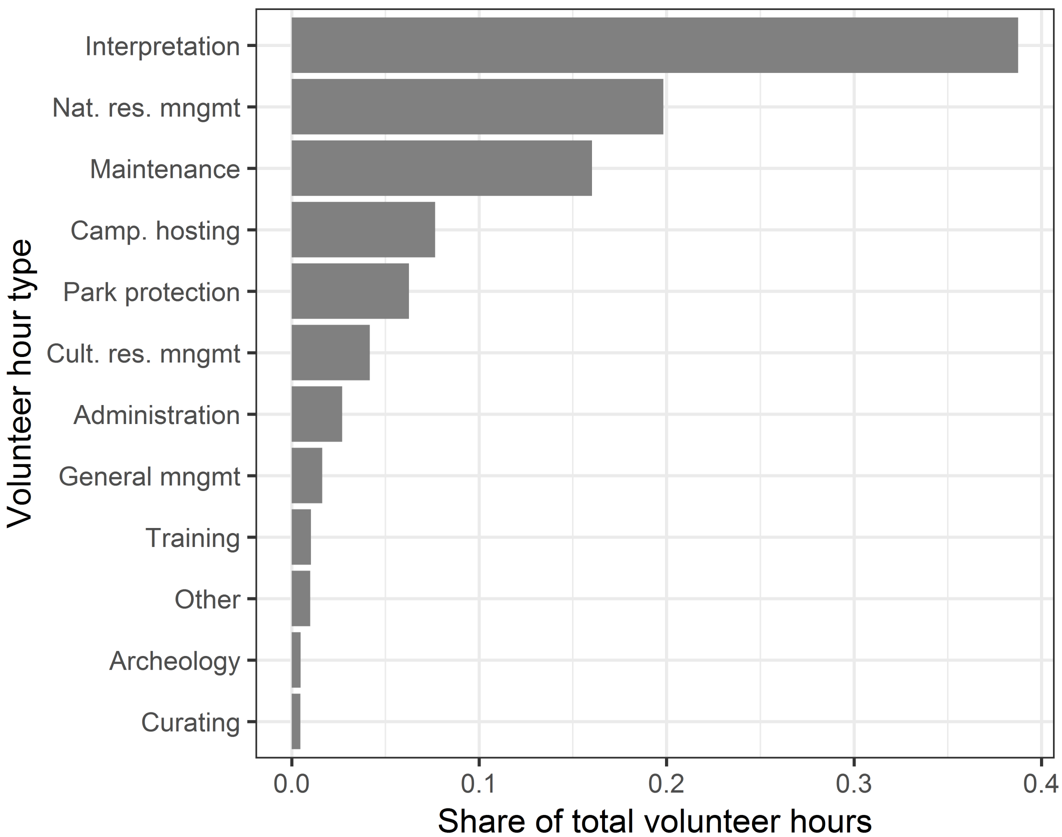

Figure 4: Breakdown of volunteer hours by type for all parks, 1998-2013

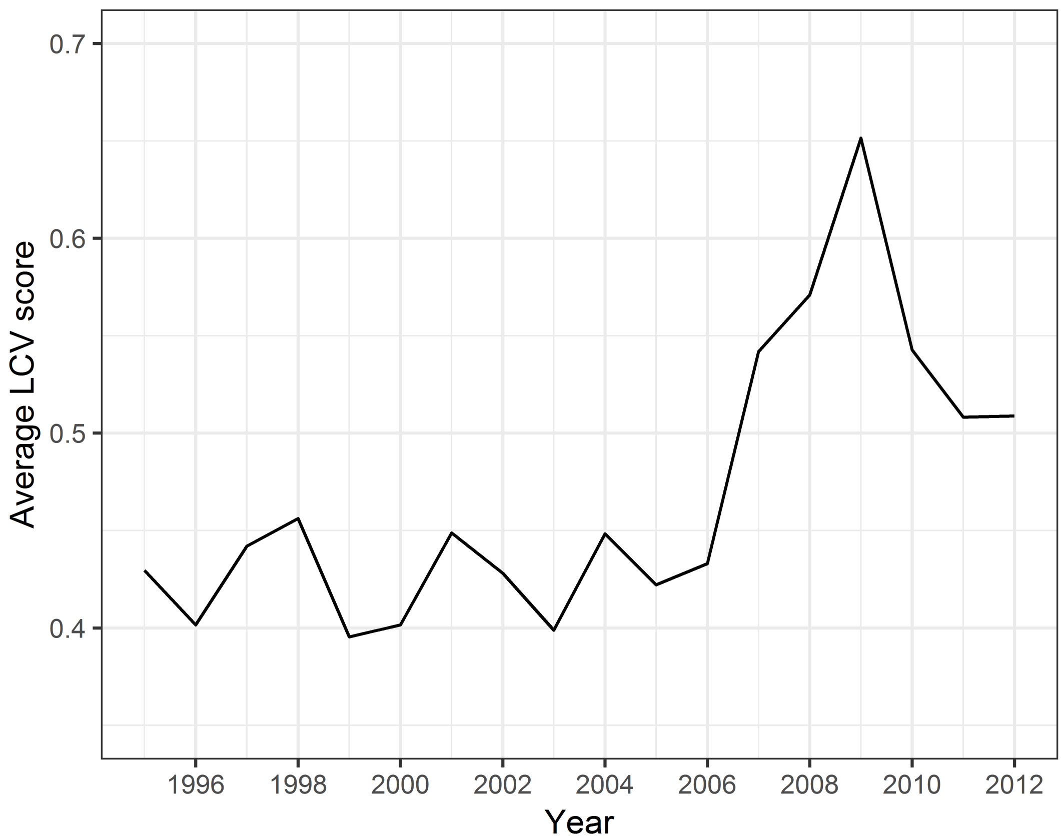

Figure 5: Time series of the average annual LCV score for members of the U.S. House and Senate Appropriations Committees

Table 1: Summary statistics

	All Parks	Environmental Parks	
		Yes	No
$Hours_{i,t}$: Volunteer Hours per Year	15,857	29,639	8,521
	(30,463)	(46,161)	(11,544)
Value of Annual Volunteer Hours ($ 1,000s)	350	655	188
	(673)	(1,020)	(255)
$Budget_{i,t}$: Annual Budget ($ 1,000s)	3,884	6,407	2,541
	(5,029)	(6,516)	(3,310)
$Visits_{i,t}$: Annual Visits (1,000s)	827	1,355	545
	(1,768)	(2,117)	(1,475)
$FTE_{i,t}$: Annual Paid Staff FTE (per 1,000 visits)	0.307	0.280	0.321
	(0.915)	(0.758)	(0.989)
Number of parks	326	105	221
Observations	4,721	1,640	3,081

Notes: Standard deviations are reported in parentheses. Parks are classified as environmental according to our primary definition of whether volunteerism in the category of natural resource management is strictly positive for every year. The value of volunteer hours is calculated using the annual value of volunteer time from the Bureau of Labor Statistics (see text). The value of volunteer hours and the annual budget are reported in 2013 dollars.

Table 2: First stage results for IV estimation

	$Budget_{i,t}$ (1)	$Budget_{i,t}$ (2)	$Budget_{i,t} \times 1[Envr]_i$ (3)
$Z_{i,t}$	0.216***	0.185**	-0.043
	(0.059)	(0.077)	(0.041)
$Z_{i,t} \times 1[Envr]_i$		0.037	0.298***
		(0.083)	(0.059)
$Visits_{i,t-1}$	-0.108	-0.117	-0.150
	(0.319)	(0.318)	(0.285)
$FTE_{i,t-1}$	0.566	0.573	0.535
	(1.172)	(1.183)	(0.837)
Park FE	✓	✓	✓
Region-Year FE	✓	✓	✓
F-stat on excluded instruments	13.35	6.82	12.91
Observations	4,721	4,721	4,721

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Column 1 is the first stage regression of specification (14); the dependent variable is annual park funding in thousands of 2013 dollars. Columns 2 and 3 are the first stage regressions of specification (15); the dependent variable is, respectively, annual park budget in thousands of 2013 dollars, and the interaction of budget with the indicator variable for an environmental park. $Z_{i,t}$ is the instrumental variable described in equation (16). Robust standard errors clustered by park are reported in parentheses, and we report cluster-robust Kleibergen-Paap F-statistics on the excluded instruments.

Table 3: Overall estimates of the average crowding effect

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
$Budget_{i,t}$	2.501*	2.487*	2.461**	12.565***	12.109**	9.007**
	(1.504)	(1.492)	(1.165)	(4.745)	(4.770)	(3.583)
$Visits_{i,t-1}$		-1.070	1.979		0.296	1.411
		(2.873)	(1.604)		(3.057)	(1.892)
$FTE_{i,t-1}$		-46.345**	-49.267*		-63.456**	-55.289**
		(21.280)	(26.354)		(26.260)	(25.423)
Park FE	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓
Observations	4,721	4,721	3,784	4,721	4,721	3,784

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is total volunteer hours by park-year. Columns 1-3 are estimated with OLS, and columns 4-6 are estimated using the IV approach described in the main text. Columns 1, 2, 4, and 5 include the full sample; columns 3 and 6 are estimated on the 1998-2010 time period, excluding the post-recession years. Robust standard errors clustered by park are reported in parentheses.

Table 4: Heterogeneous crowding effects by environmental and non-environmental parks

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
$Budget_{i,t}$	-0.217	-0.224	0.676	2.939	0.649	0.052
	(0.418)	(0.432)	(0.536)	(3.877)	(3.578)	(3.039)
$Budget_{i,t} \times 1[Envr]_i$	4.473***	4.466***	2.406*	10.193**	12.104***	9.391**
	(1.554)	(1.543)	(1.361)	(4.034)	(4.251)	(3.726)
$Visits_{i,t-1}$		-0.968	2.068		-0.011	1.932
		(2.521)	(1.572)		(2.457)	(1.698)
$FTE_{i,t-1}$		-48.780**	-52.185*		-62.748***	-64.847**
		(20.894)	(27.358)		(23.037)	(29.124)
Park FE	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓
Observations	4,721	4,721	3,784	4,721	4,721	3,784

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is total volunteer hours by park-year. Columns 1-3 are estimated with OLS, and columns 4-6 are estimated using the IV approach described in the main text. Columns 1, 2, 4, and 5 include the full sample; columns 3 and 6 are estimated on the 1998-2010 time period, excluding the post-recession years. $1[Envr]_i$ is an indicator variable for an environmental park. Robust standard errors clustered by park are reported in parentheses.

Table 5: Crowding effects separately for outside and inside volunteer hours

	OLS		IV	
	(1)	(2)	(3)	(4)
Panel A: Outside Hours				
$Budget_{i,t}$	1.729	1.713	9.120**	9.110**
	(1.165)	(1.125)	(3.943)	(3.998)
$Visits_{i,t-1}$		-1.303		-0.253
		(2.850)		(2.532)
$FTE_{i,t-1}$		-50.678**		-63.833***
		(19.643)		(24.363)
Panel B: Inside Hours				
$Budget_{i,t}$	0.771	0.774	2.960*	2.999*
	(0.748)	(0.751)	(1.774)	(1.794)
$Visits_{i,t-1}$		0.233		0.549
		(1.261)		(1.549)
$FTE_{i,t-1}$		4.333		0.377
		(9.281)		(9.190)
Park FE	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓
Observations	4,721	4,721	4,721	4,721

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is volunteer hours devoted to outside tasks in Panel A and inside tasks in Panel B. Outside hours include archeology, campground hosting, interpretation, resource management, and park protection. Inside hours include general management, curating, administration, maintenance, training, and other. Columns 1-2 are estimated with OLS, and columns 3-4 are estimated using the IV approach described in the main text. Robust standard errors clustered by park are reported in parentheses.

Table 6: Heterogeneous crowding effects by park and hour types

	OLS		IV	
	(1)	(2)	(3)	(4)
Panel A: Outside Hours				
$Budget_{i,t}$	-0.201	-0.211	1.740	1.569
	(0.343)	(0.361)	(3.197)	(3.418)
$Budget_{i,t} \times 1[Envr]_i$	3.178**	3.169**	7.810**	7.966**
	(1.430)	(1.387)	(3.696)	(3.493)
$Visits_{i,t-1}$		-0.481		-0.455
		(1.825)		(2.239)
$FTE_{i,t-1}$		-48.341***		-63.367***
		(18.187)		(21.923)
Panel B: Inside Hours				
$Budget_{i,t}$	-0.015	-0.013	-1.096	-0.919
	(0.180)	(0.180)	(1.272)	(1.310)
$Budget_{i,t} \times 1[Envr]_i$	1.295	1.297	4.293**	4.138**
	(0.994)	(0.990)	(1.774)	(1.775)
$Visits_{i,t-1}$		0.263		0.444
		(1.252)		(1.416)
$FTE_{i,t-1}$		3.626		0.619
		(9.322)		(9.398)
Park FE	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓
Observations	4,721	4,721	4,721	4,721

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is volunteer hours devoted to outside tasks in Panel A and inside tasks in Panel B. Outside hours include archeology, campground hosting, interpretation, resource management, and park protection. Inside hours include general management, curating, administration, maintenance, training, and other. Columns 1-2 are estimated with OLS, and columns 3-4 are estimated using the IV approach described in the main text. $1[Envr]_i$ is an indicator variable for an environmental park. Robust standard errors clustered by park are reported in parentheses.

Appendix Tables

Table A.1: Alternative specifications to test for robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Budget_{i,t-1}$	13.438**	1.548						
	(5.279)	(3.814)						
$Budget_{i,t-1} \times 1[Envr]_i$		12.246***						
		(4.482)						
$Budget_{i,t}$			11.475**	-1.431	10.235**	-1.286	15.998**	1.368
			(4.878)	(3.881)	(4.723)	(3.838)	(6.799)	(3.716)
$Budget_{i,t} \times 1[Envr]_i$				13.382***		12.594***		12.460**
				(4.507)		(3.953)		(5.572)
$Budget_{i,t} \times 1[Large]_i$							-3.668*	
							(2.089)	
$Visits_{i,t-1}$	-4.189	-3.229	-0.337	-1.769	0.342	0.012	0.360	-0.239
	(3.689)	(3.567)	(3.275)	(2.636)	(2.808)	(2.274)	(3.103)	(2.678)
$FTE_{i,t-1}$	-336.163	-221.136	-263.026	-136.888	-73.014**	-70.226***	-51.586**	-61.796***
	(209.798)	(135.236)	(164.482)	(170.963)	(29.459)	(25.478)	(20.771)	(22.698)
Park FE	✓	✓	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Lagged Budget	✓	✓						
Include Lagged Visits			✓	✓				
Include Time Trends for Large/Small Parks					✓	✓		
More General Definition of Envr. Parks								✓
K-P F-stat	15.61	6.91	14.25	7.10	13.53	7.68	6.34	6.35
Observations	4,362	4,362	4,022	4,022	4,721	4,721	4,721	4,721

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is total volunteer hours by park-year. All specifications are estimated using the IV approach described in the main text. Models (1) and (2) include the park budgets lagged one year. Models (3) and (4) include two years of lagged visitation. Models (5) and (6) include separate linear time trends for parks with above and below median 1995 budgets. Model (7) includes a separate interaction term between park budget and an indicator for above or below median budgets. Model (8) estimates the heterogeneous park effects with our secondary definition of an environmental park based on its stated mandate. Robust standard errors clustered by park are reported in parentheses. We report the cluster-robust Kleibergen-Paap rk Wald F-statistic in the first stage.

Table A.2: Summary statistics with the alternative definition of an environmental park

	All Parks	Environmental Parks	
		Yes	No
$Hours_{i,t}$: Volunteer Hours per Year	15,857	18,344	8,796
	(30,463)	(34,280)	(12,628)
Value of Annual Volunteer Hours ($ 1,000s)	350	405	194
	(673)	(752)	(279)
$Budget_{i,t}$: Annual Budget ($ 1,000s)	3,884	4,260	2,818
	(5,029)	(5,181)	(4,396)
$Visits_{i,t}$: Annual Visits (1,000s)	827	896	631
	(1,768)	(1,898)	(1,312)
$FTE_{i,t}$: Annual Paid Staff FTE (per 1,000 visits)	0.307	0.277	0.391
	(0.915)	(0.638)	(1.432)
Number of parks	326	233	93
Observations	4,721	3,491	1,230

Notes: Standard deviations are reported in parentheses. Parks are classified as environmental according to our secondary definition of whether a park's official mandate includes activities related to environmental conservation. The value of volunteer hours is calculated using the annual value of volunteer time from the Bureau of Labor Statistics (see text). The value of volunteer hours and the annual budget are reported in 2013 dollars.

Table A.3: Alternative instruments

	(1)	(2)	(3)	(4)	(5)	(6)
$Budget_{i,t}$	10.846**	-5.424	12.207***	2.659	7.983***	1.375
	(4.644)	(4.590)	(3.954)	(3.486)	(1.965)	(2.482)
$Budget_{i,t} \times 1[Envr]_i$		16.734***		10.344***		7.237***
		(4.550)		(3.771)		(2.548)
$Visits_{i,t-1}$	0.117	-0.369	0.310	0.082	-0.290	-0.439
	(3.327)	(2.182)	(3.582)	(2.456)	(2.778)	(2.112)
$FTE_{i,t-1}$	-61.210**	-59.470***	-63.631**	-63.459***	-56.119**	-56.127**
	(25.533)	(22.438)	(26.383)	(23.153)	(23.441)	(21.671)
Park FE	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓
All Congressional LCV Scores IV	✓	✓				
Fish and Wildlife Service IV			✓	✓		
NPS Budget For Other Regions IV					✓	✓
K-P F-stat	14.02	5.42	10.42	1.89	16.69	1.55
Observations	4,721	4,721	4,721	4,721	4,721	4,721

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is total volunteer hours by park-year. All specifications are estimated using the IV approach described in the main text. Models (1) and (2) are estimated using the instrumental variable based on the LCV scores for all U.S. Congressional members, excluding those representing each park. Models (3) and (4) are estimated using the instrumental variable based on the budget for the U.S. Fish and Wildlife Service. Models (5) and (6) are estimated using the instrumental variable based on the total NPS budget, exclusive of the region where each park is located. Robust standard errors clustered by park are reported in parentheses. We report the cluster-robust Kleibergen-Paap rk Wald F-statistic in the first stage.