Identifying Dynamic Spillovers of Crime with a Causal Approach to Model Selection

Gregorio Caetano & Vikram Maheshri*

November 15, 2016

Abstract

Does crime in a neighborhood cause future crime? Without a source of quasi-experimental variation in local crime, we develop an identification strategy that leverages a recently developed test of exogeneity (Caetano (2015)) to select a feasible regression model for causal inference. Using a detailed incident-based data set of all reported crimes in Dallas from 2000-2007, we find some evidence of dynamic spillovers within certain types of crimes but no evidence that lighter crimes cause more severe crimes. This suggests that a range of crime reduction policies that target lighter crimes (prescribed, for instance, by the "Broken Windows" theory of crime) should not be credited with reducing the violent crime rate. Our strategy involves a systematic investigation of endogeneity concerns and is particularly useful when rich data allow for the estimation of many regression models, none of which is agreed upon as causal *ex ante*. *Keywords: neighborhood crime, "Broken Windows", model selection, test of exogeneity. JEL Codes: C52, C55, K42, R23.*

1 Introduction

Does crime in a neighborhood cause future crime? When crime occurs, it may alter the physical and social environment through a variety of mechanisms. For instance, potential criminals may be influenced by their peers' behavior (Glaeser et al. (1996)), neighbors may respond by forming community watch groups (Taylor (1996)), and law enforcement may react by reallocating their resources (Weisburd and Eck (2004)). This may in turn affect future crime levels in different and ambiguous ways that could depend on the type of crime committed and its salience.

*Departments of Economics, University of Rochester and University of Houston. We thank Carolina Caetano, David Card, Aimee Chin, Scott Cunningham, Ernesto Dal Bo, Liran Einav, Frederico Finan, Alfonso Flores-Lagunes, Willa Friedman, Keren Horn, Justin McCrary, Amy Schwartz, Noam Yuchtman and various seminar and conference participants for valuable discussions. All errors are our own.

In this paper, we estimate the local effects of several different crimes (rape, robbery, burglary, auto theft, assault, and light crime) on the future levels of each of those crimes using a comprehensive database containing every police report (nearly 2 million in total) filed with the Dallas Police Department from 2000-2007. We find that robbery, burglary, auto theft and light crime cause modest increases (on the order of 5-15%) in future crimes of the same types in a neighborhood. However, we find no statistically or economically significant evidence that crimes of any type cause – directly or indirectly – more severe crimes in the future.

From a policy perspective, the idea that light crime (e.g., broken windows, graffiti, vandalism) in a neighborhood can lead to more severe crime has been particularly influential, and our results can be brought to bear on this debate. The "Broken Windows" theory of crime (Kelling and Wilson (1982), Kelling and Coles (1998)) asserts that the proliferation of visible light crime should signal to potential criminals that enforcement and punishment are lax in the area. This leads to an increase in the frequency and severity of crimes that reinforces this positive feedback mechanism. Proponents of this theory have ensured that the intense policing of light crime and the adoption of "zero-tolerance" policies remains on the agenda of law enforcement agencies in many major US cities such as New York City[1] and Chicago[2] in spite of little evidence that light crime actually leads to severe crime.[3] Similarly, "stop-and-frisk" practices have been defended on the grounds that they will deter light crime and eventually reduce more severe crimes in spite of evidence that they are racially biased, potentially contributing to the disproportionate incarceration of the poor and of racial minorities (Gelman et al. (2007)). Our findings suggest that these policies should be reconsidered.

Our empirical analysis addresses two important methodological concerns. First, the appropriate causal question is difficult to pose: in order to estimate the local effects of crime on future crime, how should we classify crimes, how local is "local", and how far in the future is "future"? With a database that is detailed in multiple dimensions (description, location and

[1] In a recent interview, NYPD Commissioner William Bratton stated, "We will continue to focus on crime and disorder." ("Inside William Bratton's NYPD: broken windows policing is here to stay", *The Guardian*, June 8, 2015).

[2] In 2013, Chicago Police Superintendent Gary McCarthy proposed "to authorize arrests for unpaid tickets for public urination, public consumption of alcohol, and gambling... 'Fixing the little things prevents the bigger things,' said McCarthy, a longtime advocate of the 'broken windows' approach to fighting crime." ("Chicago is Adopting The 'Broken Windows' Strategy", *Law Enforcement Today*, May 5, 2013). Ultimately, the Chicago city council drastically increased penalties for these misdemeanors.

[3] Levitt (2004) describes efforts by the media to attribute falling crime rates in New York City to innovative law enforcement policies, including "Broken Windows" policies, but he argues that this conclusion is premature given other confounding changes that occurred in New York City at the same time or even before such policies were implemented.

time), there are too many initially plausible ways of aggregating the data. Second, identifying the causal effects of interest is difficult because unobserved determinants of future crimes are persistent, and it is hard to conceive of instrumental variables that are both transitory and highly localized. These two concerns are inseparable because the assumptions necessary to identify causal effects depend on the level of aggregation of the data.

Our analysis starts with the conjecture that we may be able to isolate exogenous variation in past crime rates by appropriately specifying fixed effects that absorb confounding variation in past crime rates. Intuitively, social interactions occur at fine levels of geography *and* time (Bikhchandani et al. (1992); Ellison and Fudenberg (1995); Akerlof (1997)), whereas confounders only vary at fine levels of geography *or* time, but not both. For example, localized confounders of crime such as neighborhood wealth levels (Flango and Sherbenou (1976)) and family structure (Sampson (1985)) vary more slowly than crime itself, while rapidly varying causes of crime such as weather (Cohn (1990)) tend to affect nearby neighborhoods similarly. Our primary empirical obstacle is that we cannot know *ex ante* (a) whether this conjecture is correct and (b) if it is correct, what levels of temporal and geographical fixed effects (and aggregation of the data) will successfully isolate the variation of interest.

In light of this, we develop a strategy that allows us to select an empirical model from which we can identify parameters of interest without an *ex ante* known source of quasi-experimental variation. Instead, our strategy finds the levels of temporal and geographical fixed effects (and aggregation of the data) that isolate variation shown to be valid for causal inference *ex post*. We do so by exploiting a recently developed test of exogeneity (Caetano (2015)) that yields an objective statistical criterion for whether the parameters of interest in a particular empirical model can be interpreted as causal. Unlike other tests of exogeneity (e.g., Hausman (1978)), this test does not require instrumental variables; instead, it requires that unobservables vary discontinuously at a known threshold of the main explanatory variable of interest, which often happens in contexts where observations bunch at this threshold. In our context, we argue both theoretically and empirically that such discontinuities exist at the zero crime threshold.

Of course, one can never fully validate a research design: a failure to reject the null hypothesis of exogeneity for an empirical model does not imply that the model is exogenous. Thus, we systematically develop the case that our failure to reject the null hypothesis of exogeneity reasonably points to the conclusion that the null is correct and that we have successfully identified the causal effects of crime on future crime. We do so with a theoretical and empirical analysis of the statistical power of the test that enumerates the many properties that confounders must possess in order to remain undetectable by the test of exogeneity. We find that not a single observed variable that we construct from our detailed database qualifies

as an undetectable confounder, and our case is further supported by a battery of robustness checks that are designed to detect unobserved confounders that may have otherwise evaded detection.[4]

Ultimately, we arrived at the qualified conclusion that the cumulative list of properties that a confounder must possess in order to bias the results of our preferred regression model is restrictive beyond reasonable doubt; that is, our preferred model should be reasonably interpreted as causal.

Our findings contribute to a considerable literature that seeks to identify intertemporal links in criminal behavior. Jacob et al. (2007) use weekly weather shocks as instruments and find a small, negative relationship between citywide past crimes and citywide future crimes. Kelling and Sousa (2001), Funk and Kugler (2003) and Corman and Mocan (2005) analyze whether targeting less severe crimes has been effective in reducing more violent crimes in the future, but as pointed out by Harcourt and Ludwig (2006), it may be difficult to interpret these estimates as causal. Harcourt and Ludwig (2006) take advantage of a random allocation of public housing under the "Moving to Opportunity" experiment in five US cities and find that individuals assigned to neighborhoods with higher misdemeanor crime levels were as likely to commit violent crime as those who were assigned to "better" neighborhoods. In contrast, Damm and Dustmann (2014) find that refugee boys who were randomly assigned to high violent crime neighborhoods in Denmark exhibit a higher propensity for criminal behavior as young adults.[5] In addition, our findings contribute to the long standing literature on crime and geography (see, Anselin et al. (2000) for a review of this literature).

Our identification strategy may help researchers leverage detailed datasets to conduct robust causal inference in many settings that were previously difficult to explore using observational data alone, which may prove particularly valuable given the recent explosion in the availability of such datasets without a similar increase in the availability of quasi-experimental variation (Varian (2014)). These trends suggest an increasing need for empirical approaches that can exploit rich data for causal inference by identifying variation that is agreed upon to be exogenous only *ex post*. Our approach, in particular, may allow researchers to take further advantage of recent law enforcement agency efforts to maintain and release large, detailed crime databases.[6] However, we caution that the ability of our

[4]The systematic implementation of the exogeneity test for the express purpose of validating an identification strategy is a novel contribution of our paper. Caetano (2015) implements her test with the goal of rejecting a model, rather than validating it. Thus, while Caetano (2015) needs only to show that the test has some power, we instead need to show that the test is very powerful, so much so that not detecting endogeneity in a given model reasonably suggests that it is exogenous.

[5]Braga and Bond (2008) randomized police patrols in certain light crime "hot spots" of Lowell, Massachusetts and found that increased policing also reduced citizen calls for service for more severe crimes.

[6]Notably, our approach may allow researchers to allay concerns due to measurement, such as under-

approach to validate models does not center on simply the availability of a large amount of data; it is also crucial to justify empirically that the test of exogeneity in the context of interest has statistical power to detect endogeneity from all sources that researchers should be concerned about.

The remainder of the paper is organized as follows. In Section (2), we present a stylized dynamic model of crime that describes how crime may affect future crime through a simple learning process. In Section 3, we conceptualize such intertemporal linkages in a simple empirical model and provide an overview of our empirical approach. In Section 4, we offer intuition for the test of exogeneity that we use and show how it can be used as the centerpiece of our identification strategy. In Section 5, we describe our sample and explain how we address the inherent trade-offs we face in model selection. In Section 6, we present estimates of the short- and long-run intertemporal effects of light crime and the full long-run dynamic spillovers associated with various hypothetical crime reduction policies. In Section 7, we subject our results to a variety of robustness checks In Section 8, we discuss our findings in the context of the "Broken Windows" theory before concluding in Section 9. We provide further detail on the test of exogeneity, our data, and our ability to detect distinct sources of endogeneity in the Appendix.

2 A Dynamic Model of Crime

Crime may affect future crime levels through a variety of direct and indirect channels. We provide a theoretical basis for how such intertemporal linkages may arise with a simple and highly stylized dynamic model of rational criminal behavior that builds on Becker (1968). Let $\mathbb{C}$ be the set of all potential crimes. An individual i in neighborhood j would choose to commit a crime of type $y \in \mathbb{C}$ in week t if her private benefits to committing that crime exceeded her costs, or if

$$B^y_{ijt} > p^y_{jt} C^y_{ijt} \tag{1}$$

where B^y_{ijt} is the total private benefit to the individual, p^y_{jt} is the probability of punishment conditional on committing the crime, and C^y_{ijt} is the cost of punishment.

Individuals may possess imperfect knowledge of p^y_{jt}, but by observing past levels of crime (along with features of the environment that led to past crime), individuals form beliefs of p^y_{jt}, which we denote by π^y_{ijt}. It follows that the total number of crimes of type y that are committed in neighborhood j in week t is

reporting and misreporting, which have been long identified as important obstacles in empirical analyses of crime data (see, e.g., Skogan (1974, 1975); Levitt (1998b)).

$$\text{Crime}^y_{jt} = \sum_{i \in I^y_{jt}} \mathbb{I}_{\{b^y_{ijt} > \pi^y_{ijt}\}}, \tag{2}$$

where I^y_{jt} represents the pool of potential criminals, $\mathbb{I}_{\{\cdot\}}$ is the indicator function, and $b^y_{ijt} = \frac{B^y_{ijt}}{C^y_{ijt}}$ represents i's "benefit-cost factor" of commiting a crime. To keep the model tractable, we make assumptions about the individual heterogeneity within the neighborhood to facilitate aggregation: (a) $\pi^y_{ijt} = \pi^y_{jt}$ is a common prior for all individuals within the neighborhood; and (b) b^y_{ijt} is drawn from a cumulative distribution $F(\cdot; \Theta^y_{jt})$. Both the prior π^y_{jt} and the parameter of this distribution, Θ^y_{jt}, may vary by neighborhood, week and type of crime. It follows that

$$\text{Crime}^y_{jt} = I^y_{jt} \cdot \left(1 - F(\pi^y_{jt}; \Theta^y_{jt})\right) \tag{3}$$

Each of the three parameters that describe the criminogenic environment, $(I^y_{jt}, \Theta^y_{jt}, \pi^y_{jt})$, can be affected by previous crime levels. Denoting Crime_{jt-1} as the vector of crimes of all types in $t-1$ (of which the xth element is Crime^x_{jt-1}), we can express this as

$$I^y_{jt} = I^y\left(\text{Crime}_{jt-1}, \eta^I_{jt-1}\right) \tag{4}$$

$$\Theta^y_{jt} = \Theta^y\left(\text{Crime}_{jt-1}, \eta^\Theta_{jt-1}\right) \tag{5}$$

$$\pi^y_{jt} = \pi^y\left(\text{Crime}_{jt-1}, \eta^\pi_{jt-1}\right) \tag{6}$$

where η^I_{jt-1}, η^Θ_{jt-1} and η^π_{jt-1} represent other (observable and unobservable) determinants of I^y_{jt}, Θ^y_{jt} and π^y_{jt} respectively.[7]

We illustrate how these three equations encompass many of the specific intertemporal linkages in criminal behavior that have been offered by researchers by way of several concrete examples. For instance, equation (4) allows for the possibility of incapacitation effects (e.g. Levitt (1998a)), as past crimes may lead to arrests and reductions in the future pool of potential criminals. Equation (5) captures all changes to the private costs and benefits of crime to individuals induced by prior crimes. This includes learning from previous experiences (Kempf (1987)), peers' experiences (Glaeser et al. (1996)) and responses by law enforcement that increase the cost of punishment, conditional on arrest. Finally, equation (6) captures the learning process whereby previous crimes lead criminals to update their prior beliefs of the probability of punishment conditional on commiting a crime. This learning process may reflect the mechanism suggested by the "Broken Windows" theory (i.e., previous crimes

[7] Of course, past crimes from $t-2$, $t-3$ and so on may also be included in these equations. We empirically assess this possibility in Section 6.2.

signal neighborhood distress (Kelling and Wilson (1982))),[8] or it could reflect the fact that neighbors and the police may respond to crime with increased monitoring which, if observed by criminals, might deter future crime (Taylor (1996); Weisburd and Eck (2004)).

We can describe the total intertemporal relationship between crimes as

$$\begin{aligned}\frac{\partial \text{Crime}^y_{jt}}{\partial \text{Crime}^x_{jt-1}} = & \underbrace{\frac{\partial\left(I^y_{jt}\cdot\left(1-F(\pi^y_{jt};\Theta^y_{jt})\right)\right)}{\partial I^y_{jt}}\frac{\partial I^y_{jt}}{\partial \text{Crime}^x_{jt-1}}}_{\text{channel 1}} + \\ & \underbrace{\frac{\partial\left(I^y_{jt}\cdot\left(1-F(\pi^y_{jt};\Theta^y_{jt})\right)\right)}{\partial \Theta^y_{jt}}\frac{\partial \Theta^y_{jt}}{\partial \text{Crime}^x_{jt-1}}}_{\text{channel 2}} + \\ & \underbrace{\frac{\partial\left(I^y_{jt}\cdot\left(1-F(\pi^y_{jt};\Theta^y_{jt})\right)\right)}{\partial \pi^y_{jt}}\frac{\partial \pi^y_{jt}}{\partial \text{Crime}^x_{jt-1}}}_{\text{channel 3}}\end{aligned} \tag{7}$$

This equation incorporates the three different and broad channels defined in equations (4), (5) and (6) by which past crime can cause future crime. Each causal response is likely to differ depending on the types of past and future crimes. For example, light crimes such as graffiti or public urination may generate little or no incapacitation effects relative to violent crimes, but they may be more salient to criminals, police and neighbors relative to harder-to-observe crimes such as rape. Moreover, the propensity of criminals to commit certain crimes in the heat of the moment such as murder may be less affected by incapacitation effects than more professionalized crimes such as burglary and auto-theft (Blumstein et al. (1986)). Furthermore, given the generality of the model, it is premature to sign the three terms in equation (7), as they depend on the relative intensities of responses from a variety of different agents (e.g., potential criminals, police, criminal justice policymakers, and other private citizens) who have countervailing and potentially complex incentives. As such, identifying the causal effects $\frac{\partial \text{Crime}^y_{jt}}{\partial \text{Crime}^x_{jt-1}}$ for each combination of x and y is a fundamentally empirical question that we seek to answer in this paper. This question is of importance because while policy makers typically have no way of directly controlling the parameters of the model $(I^y_{jt}, \Theta^y_{jt}, \pi^y_{jt})$, they are more capable of devising policies that target the levels of certain types of crime Crime^x_{jt-1}, which may indirectly affect these parameters. Indeed, the relative

[8] In this scenario, the function $\pi^y(\cdot)$ would likely be decreasing in Crime^x_{jt-1} for $x = y$ (this would be the case if, for instance, individuals were rational and updated their priors according to Bayes Law). Hence, if Crime^y_{jt-1} is higher than expected, then individuals would revise their estimate of π^y_{jt} downward, leading to an increasing crime. Further, $\pi^y(\cdot)$ would likely decrease in Crime^x_{jt-1}, $x \neq y$ if individuals expect π^y_{jt} and π^x_{jt} to be positively correlated to each other.

benefits of various targeting policies have occupied a prominent place in the law enforcement policy debate in many large US cities. In light of this, while identifying the contribution of each of these channels is beyond the scope of this paper, identification of the full effect $\frac{\partial \text{Crime}^y_{jt}}{\partial \text{Crime}^x_{jt-1}}$ is quite valuable. This involves isolating the component of $Cov(\text{Crime}^y_{jt}, \text{Crime}^x_{jt-1})$ that is not attributable to $Cov(\eta^a_{ijt}, \eta^b_{ijt-1})$ where a and b may correspond to I, Θ or π. We now discuss our strategy to do so.

3 Identification Strategy

In order to test empirically whether crime affects future crime levels, we formally specify the intertemporal linkages described above in a system of equations of motion that summarize the co-evolution of crimes of various types in a neighborhood. The equation of motion for crime y can be written as

$$\text{Crime}^y_{jt} = \sum_{x \in \mathbb{C}} \text{Crime}^x_{jt-1} \beta^{xy} + \text{Controls}_{jt} \gamma^y + \text{Error}^y_{jt} \tag{8}$$

where β^{xy} denotes the effect of a crime of type x on a future crime of type y (we will index dependent crime variables with y and explanatory crime variables with x throughout the paper), Controls_{jt} is a vector of observed covariates, and Error^y_{jt} includes all unobserved determinants of crime. Each observation in equation (8) is uniquely indexed by j, t and y. We collect these equations and represent the system of equations of motion in matrix form as

$$\text{Crime}_{jt} = \beta \text{Crime}_{jt-1} + \Gamma \text{Controls}_{jt} + \text{Error}_{jt}, \tag{9}$$

where Crime_{jt} is a $|\mathbb{C}| \times 1$ column vector. The parameter matrix β (whose (y, x) element is equal to β^{xy}) contains the $|\mathbb{C}|^2$ treatment effects of interest: the intertemporal effects of all crimes both within and across types of crimes.

Unobserved determinants of crime - e.g., neighborhood amenities, characteristics of neighbors, and law enforcement practices in the area - are likely to persist over time. As a result, a naive estimation of (9) by ordinary least squares (OLS) will yield a biased estimator of β. The standard solution to this issue is the use of instrumental variables (IV) to identify β, but IVs are difficult to find since any candidate IV must be both transitory and vary at the neighborhood level. This difficulty is further compounded by the fact that there are $|\mathbb{C}|$ endogenous variables, so at least $|\mathbb{C}|$ separate IVs would be required.

In light of this, we take an alternative approach. β is identified under the standard exogeneity assumption:

Assumption 1. $\text{Cov}\left(\text{Crime}^{x}_{jt-1}, \text{Error}^{y}_{jt} | \text{Crime}^{-x}_{jt-1}, \text{Controls}_{jt}\right) = 0$ *for all* x, y.

where Crime^{-x}_{jt-1} is the vector containing $\text{Crime}^{x'}_{jt-1}$ for all $x' \neq x$. The plausibility of this assumption depends upon the model in question, which is the unique representation of equation (9) consisting of four objects: the classification of crimes ($\mathbb{C}$), definitions of neighborhoods (j) and time periods (t), and choice of covariates (Controls_{jt}).

Given sufficiently detailed data, Assumption 1 may be satisfied for some feasible model, but we do not know *ex ante* which model (if any) does so. Accordingly, we develop an empirically driven identification strategy that is guided by a formal test of Assumption 1 (Caetano (2015)). We outline our approach below:

1. Leveraging institutional and theoretical knowledge, as well as unique features of our data, we begin by considering a large subset of candidate models (Section 5.2).

2. For each candidate model, we test Assumption 1 using a formal test of exogeneity (Section 4). Most models do not survive, but one model does survive (Table 2 in Section 6).

3. We present the results of the model that survives the test of exogeneity (Table 3 in Section 6).

4. Because the failure to reject exogeneity does not imply exogeneity – there could be confounders undetectable by the test that bias our results – we present systematic evidence of the power of the test. We construct a large pool of observed variables from our detailed database and auxiliary datasets (691 variables in total) and show that none of these variables is undetectable by the test of exogeneity. Because our pool of observed variables may not be representative of all unobservables, this alone does not entirely rule out the existence of undetectable unobserved confounders. Thus, we also show that our pool of observed variables is representative in an important way: we observe detectable confounders that correspond to the full spectrum of potential endogeneity concerns in our application (Appendix B).

5. We perform additional robustness checks with the particular goal of detecting confounders that are undetectable by the test (Section 7). We find that the surviving model above is the only one that survives all other checks.

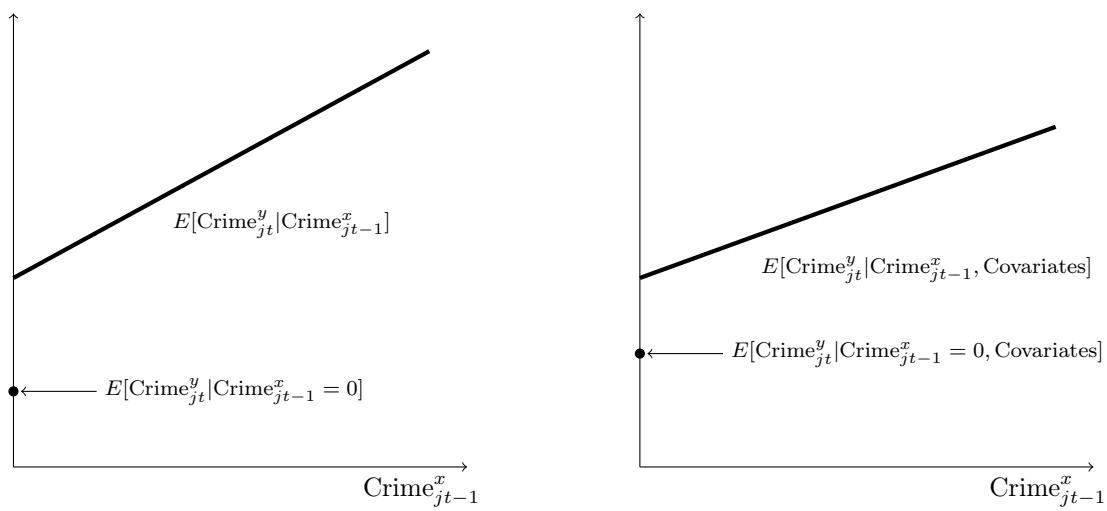
6. From our sensitivity analysis, we systematically catalog the necessary properties that any variable must possess to bias the results from our surviving model (Table 11): it (a) must be undetectable by the test of exogeneity, (b) cannot be absorbed by controls, and (c) must survive the many robustness checks performed.

In total, this allows us to reach the qualified conclusion that our surviving model is appropriate for causal inference of β by OLS as it is difficult to conceive of a variable that possesses the three properties above given our empirical evidence.

4 Testing the Exogeneity Assumption

Unlike tests of exogeneity that require valid IVs (e.g., Hausman (1978)), the test we use relies on there being a known threshold value of the endogenous variable around which unobservable confounders vary discontinuously (Caetano (2015)). Figure 1 provides graphical intuition for the idea. We illustrate the expected number of crimes of type y for each level of past crimes of type x in a neighborhood. This relationship as presented constitutes a raw correlation; our goal is to determine whether any of it can be interpreted as causal.

Figure 1: Test of Exogeneity: Intuition



(a) Past Crime and Future Crime, Unconditional

(b) Past Crime and Future Crime, Conditional on Covariates

Assume that Crime^x_{jt-1} has a continuous causal effect on Crime^y_{jt} at $\text{Crime}^x_{jt-1} = 0$ (this follows trivially from the specification of equation (9)). Then the discontinuity observed in the unconditional relationship between Crime^x_{jt-1} and Crime^y_{jt} (Panel (a)) can be attributed to either observed covariates or unobserved confounders that vary discontinuously at $\text{Crime}^x_{jt-1} = 0$. Now, suppose that we condition this relationship on all observed covariates and reproduce this plot in Panel (b). Any remaining discontinuity observed at $\text{Crime}^x_{jt-1} = 0$ can only be due to unobserved confounders that were not absorbed by the controls. Thus, finding a discontinuity after controlling for all covariates is equivalent to detecting endogeneity in the specification.

This test of exogeneity is easy to implement. Let d^x_{jt-1} be an indicator variable that is equal to 1 if $\text{Crime}^x_{jt-1} = 0$, and let D_{jt-1} be the $|\mathbb{C}| \times 1$ vector whose x^{th} element is d^x_{jt-1}. To test Assumption 1, we rewrite equation (9) to include these indicator variables:

$$\text{Crime}_{jt} = \beta \text{Crime}_{jt-1} + \Gamma \text{Controls}_{jt} + \Delta D_{jt-1} + \epsilon_{jt}, \tag{10}$$

where Δ is a $|\mathbb{C}| \times |\mathbb{C}|$ matrix of parameters that represent the sizes of the discontinuities at $\text{E}\left[\text{Crime}^y_{jt} | \text{Crime}^x_{jt-1} = 0, \text{Crime}^{-x}_{jt-1}, \text{Controls}_{jt}\right]$ for all combinations of x and y. It follows that an F-test of whether all elements of Δ are equal to zero is equivalent to a test of Assumption 1.[9]

Remark 1. The test of exogeneity requires that Crime^x_{jt-1} has a continuous causal effect on Crime^y_{jt} at $\text{Crime}^x_{jt-1} = 0$, otherwise the parameters in Δ would incorporate the treatment effect. If this assumption did not hold, then all models would be rejected irrespective of whether they were endogenous or exogenous. Thus, the fact that some models survive the test is direct evidence that this assumption is valid. Conceptually, we believe that this assumption is valid in our context because every neighborhood crime is not necessarily observed by everyone (all neighbors, all potential criminals, etc), and each person does not respond to this knowledge the same way. (For instance, the behavior of some potential criminals might be affected when $\text{Crime}^x_{jt-1} = 1$, whereas the behavior of other potential criminals will be affected only when $\text{Crime}^x_{jt-1} = 2$.) This will lead the effects we want to estimate, which represent the direct or indirect responses of these individuals to their knowledge of these crimes, to be smoothed away.

4.1 Power of the Test

As with any identification strategy, we can never validate Assumption 1 for a given model beyond all doubt. Instead, we can only make the strongest attempt possible to reject Assumption 1 in candidate models and arrive at the qualified conclusion that we cannot reject the causal interpretation of a model that survives powerful tests. Hence, our empirical burden is to argue that a failure to reject the null hypothesis of exogeneity reasonably points to the conclusion that the null is correct. For our identification strategy, this amounts to carefully and systematically establishing the statistical power of the test to ensure that we can detect endogeneity from all sources.

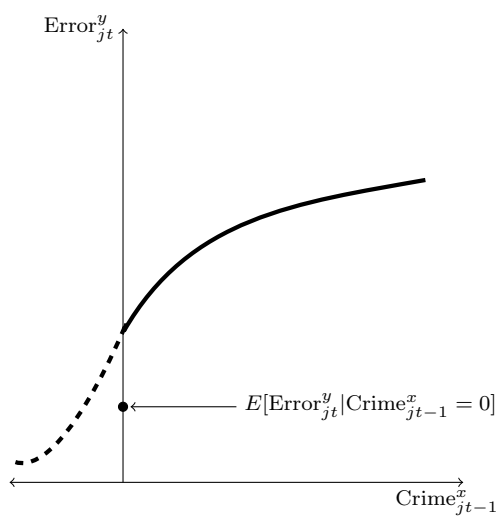
In our application, the statistical power of the test is derived from the assumption that unobserved confounders vary discontinuously at $\text{Crime}^x_{jt-1} = 0$ for some x. Although this

[9] Because the test that we implement is an extension of Caetano (2015) to a multivariate context, we provide a more formal derivation in Appendix A.

does not hold in all settings (which restricts the realm of applications of our identification strategy), there is a clear theoretical reason for why we should find such discontinuities in our setting.[10] Among the neighborhoods with $\text{Crime}^x_{jt-1} = 0$ there are those that are so wealthy (or so safe, or so heavily patrolled by police, etc.) that we would expect $\text{Crime}^x_{jt-1} = 0$ even if they were slightly poorer (or more dangerous, or less policed, etc.). The latent heterogeneity in these infra-marginal neighborhoods implies that neighborhoods with zero crime should be discontinuously different on average than the set of neighborhoods with barely positive amounts of crime. In other words, the mere fact that crime levels must be non-negative generates a bunching of neighborhoods with zero crime that may in turn lead to discontinuities in unobservable determinants of crime.

We illustrate this intuition in Figure 2, where we plot the expected value of a particular unobservable for each level of a crime of type x. Without loss of generality, we assume that this relationship is positive. The dashed line is suggestive of what the expected value of the unobservable would have been if crime was not truncated at zero. Note that this truncation mechanically generates a discontinuity in the expected value of the unobservable at zero, which provides power to detect endogeneity from this source.

Figure 2: Why are Unobservables Discontinuous at $\text{Crime}^x_{jt-1} = 0$?



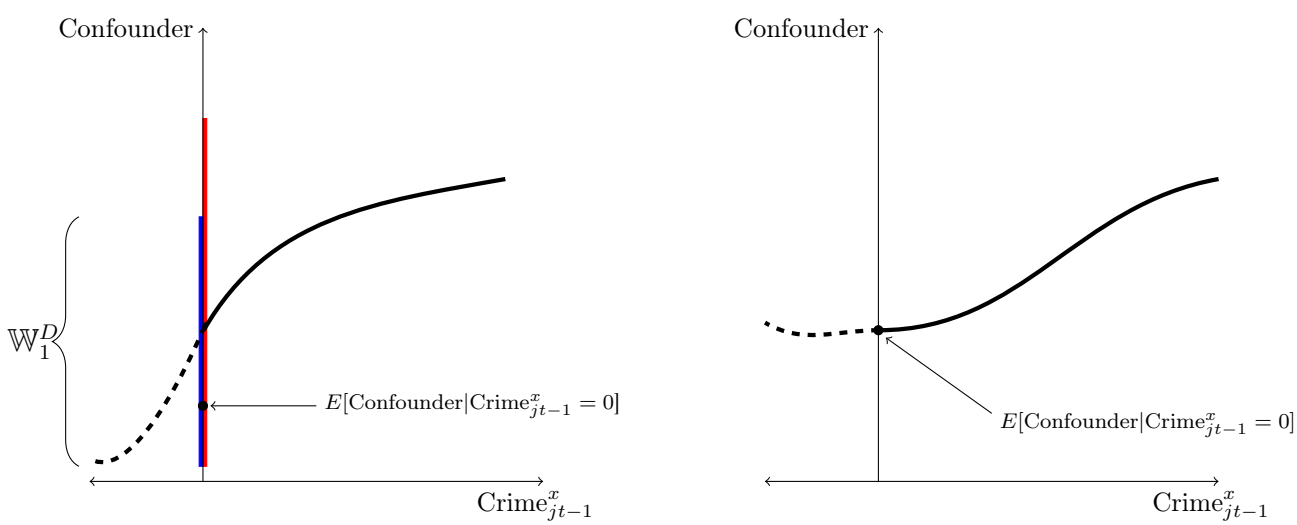
Let $\mathbb{W}$ be the set of all model confounders, defined as the set of variables w that are both correlated to Crime^y_{jt} and to Crime^x_{jt-1} for some combination of x and y. We provide an intuitive framework to understand the statistical power of our test by splitting the set $\mathbb{W}$ into two disjoint subsets: $\mathbb{W}^D$, which contains variables that vary discontinuously at $\text{Crime}^x_{t-1} = 0$ for some x, and $\mathbb{W}^C$, which contains variables that vary continuously at

[10] Caetano (2015) discusses other potential settings where this test can be applied.

$\text{Crime}^x_{t-1} = 0$ for all x. $\mathbb{W}^D$ can be further split into $\mathbb{W}^D_1$, which contains variables that are correlated to Crime^y_t when $\text{Crime}^x_{t-1} = 0$ for some combination of x and y, and $\mathbb{W}^D_2$, which contains variables that are uncorrelated to Crime^y_t when $\text{Crime}^x_{t-1} = 0$ for all combinations of x and y. These three sets, $\mathbb{W}^D_1$, $\mathbb{W}^D_2$ and $\mathbb{W}^C$ form a partition of $\mathbb{W}$. Our test of exogeneity can detect endogeneity from all confounders in $\mathbb{W}^D_1$, but it cannot detect endogeneity from confounders in $\mathbb{W}^D_2$ or $\mathbb{W}^C$. Hence, the statistical power of our test intuitively corresponds to the size of $\mathbb{W}^D_1$ relative to $\mathbb{W}$. If in a particular setting, all potential confounders belonged to $\mathbb{W}^D_1$, then the test would have "full" power, i.e., we could interpret the estimates of a surviving model as causal.[11]

In practice, our testing procedure is increasingly powerful in a multivariate application such as ours: as the number of crimes considered gets larger, the relative size of $\mathbb{W}^D_1$ grows while the relative sizes of $\mathbb{W}^D_2$ and $\mathbb{W}^C$ shrink, which increases the power of our test. We illustrate this point in Figure 3, where we present examples of confounders in $\mathbb{W}^D$ and in $\mathbb{W}^C$.

Figure 3: Types of Confounders



(a) $\mathbb{W}^D_1$ vs. $\mathbb{W}^D_2$ (b) $\mathbb{W}^C$

Notes: Red region: Support of confounder among all observations of sample. Blue region: Support of confounder among all observations of sample with zero past crime.

In the first panel, we distinguish between confounders belonging to $\mathbb{W}^D_1$ and $\mathbb{W}^D_2$. The confounder shown varies discontinuously when $\text{Crime}^x_{jt-1} = 0$. The thick, dashed line represents the average value of the confounder for each level of the propensity for past crime when it is negative, and the dot, as implied by the thin, dashed line, represents the average

[11] More formally, $\mathbb{W}^D_1$, $\mathbb{W}^D_2$ and $\mathbb{W}^C$ are all defined conditional on Crime^{-x}_{jt-1}, as we discuss in Appendix C.

value of this confounder across all observations with $\text{Crime}^x_{jt-1} = 0$. For this example, the red region along the right side of the vertical axis is the support of the confounder in the whole sample, and the blue region along the left side of the vertical axis is the support of the confounder in the subsample of observations where $\text{Crime}^x_{t-1} = 0$. To be a confounder, it must, by definition, be correlated to Crime^y_t in the red region. If it is also correlated to Crime^y_t in the blue region, then it belongs to $\mathbb{W}^D_1$. Of course, there are $|\mathbb{C}|^2$ diagrams like this, one for each of the combinations of x and y, and the confounder need only to belong to the blue region of at least one of these diagrams for it to belong to $\mathbb{W}^D_1$. The only way a confounder could belong to $\mathbb{W}^D_2$ would be if it did not belong to the union of all blue regions across any of the $|\mathbb{C}|^2$ combinations of x and y.

In the second panel, we illustrate a confounder belonging to $\mathbb{W}^C$. This confounder is correlated to Crime^x_{jt-1} but only when $\text{Crime}^x_{jt-1} > 0$. Once again, in a multivariate setting this test has greater power; as long as observations with $\text{Crime}^x_{jt-1} > 0$ are such that $\text{Crime}^{x'}_{jt-1} = 0$ for some x', and w varies discontinuously at $\text{Crime}^{x'}_{jt-1} = 0$, then w will still be detectable by the test.

Figure 3 highlights the value of implementing the exogeneity test in a multivariate context. Any confounder w is detectable by the test as long as it is detectable by at least one combination of x and y. In Appendix Section B.1, we show that these additional layers of redundancy substantially increase the power of the test.

We supplement this discussion with abundant empirical evidence of the power of the test in Appendix B (available online). We first show that no observed variables in our database belong to $\mathbb{W}^D_2 \cup \mathbb{W}^C$. For further context, we also provide examples of observed variables belonging to $\mathbb{W}^D_1$ that correspond to the various potential sources of endogeneity that might arise in our application.

5 Data

The set of potential candidate models of intertemporal effects of crime is very large, which requires us to limit our attention to a relevant subset before implementing our identification strategy. Because there is no purely empirical criterion for determining this relevant subset of models,[12] we take into account theoretical and institutional characteristics and show how the test of exogeneity can be leveraged to aid in this task.

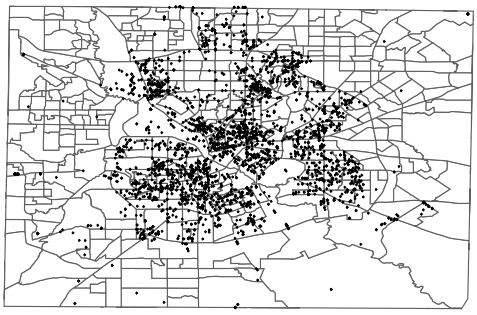
[12]The need to predetermine a relevant subset of candidate models is a requirement of all existing model selection approaches (Kadane and Lazar (2004)).

5.1 Sample

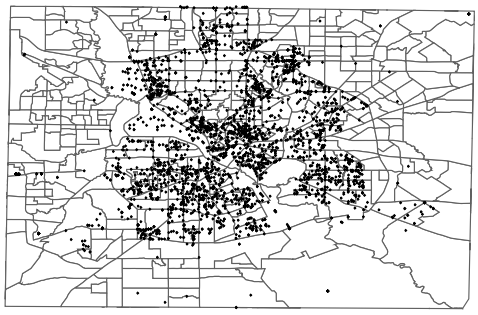
We assembled a database encompassing every police report filed with the Dallas Police Department (DPD) from January 1, 2000 to September 31, 2007.[13] Every report in our database lists the exact location (address or city block) of the crime, the exact description of the crime, and its five digit Uniform Crime Reporting (UCR) classification as given by the responding officer.[14] To offer a sense of the size and richness of this database, we plot all crimes that were reported in the first two weeks of our sample period in Figure 4. The spatial variation in crimes is immediate. The temporal variation in crime from week to week is less visually stark, which is suggestive of serially correlated determinants of crimes, and hence the difficult endogeneity problem that we face.

Figure 4: Reported Crimes in Dallas

(a) Jan. 1, 2000 - Jan. 7, 2000



(b) Jan. 8, 2000 - Jan. 14, 2000



Notes: We map all reported crimes in Dallas in the first two weeks of 2000. Census Tract boundaries are shown for geographic perspective.

A detailed description of the complainant is also provided with the exception of anonymous reports. Private companies and public officials/offices may be listed as complainants. Every report also lists a series of times from which we can construct the entire sequence of crime, neighborhood response and police response. Specifically, we observe the time (or estimate of the time) that the crime was committed, the time at which the police were dispatched, the time at which the police arrived at the scene of the crime, and the time at which the police departed the scene of the crime.

[13] A small number of police reports – sexual offenses involving minors and violent crimes for which the complainant (not necessarily the victim) is a minor – are omitted from our data set for legal reasons.

[14] If a particular complaint consists of multiple crimes (e.g., criminal trespass leading to burglary), then the report is classified only under the most severe crime (burglary) per UCR hierarchy rules from the FBI.

5.2 Aggregation Choices

Before performing any estimation, we must first define the different types of crimes of interest, neighborhood boundaries, and time periods over which we construct crime rates. These modeling choices correspond to choices of $\mathbb{C}$, j and t, respectively, which partially determine the set of possible models that we should consider. All else constant, the most disaggregated model is preferable since it better exploits the heterogeneity present in the data and contains more sharply interpretable parameters. However, the set of potential control variables is exponentially larger in more disaggregated models, requiring us to consider an exponentially larger set of candidate models. Here, we discuss the practical trade-offs we encounter in choosing $\mathbb{C}$, j, and t. As discussed in Section B.2, we use the test of exogeneity to simplify our model search by reducing the number of models that we need to consider. (This is important since the detail of our data allows us to estimate an unfeasibly large number of candidate models.) For example, if we know that we can detect endogeneity due to both under-aggregation and over-aggregation, then we need not be concerned about re-aggregating our data at another level if we arrive at a model that survives the test.

5.2.1 Classification of Crimes ($\mathbb{C}$)

In principle, crimes could be classified very coarsely (e.g., violent crimes) or finely (e.g., robberies at day time with a knife). On the one hand, as we disaggregate the types of crimes we are able to specify more treatment effects, which allows for a richer analysis of the intertemporal effects of crime. In addition, a larger set of crimes under consideration should, other things equal, increase the statistical power of our test. On the other hand, it is difficult to precisely measure local crime rates if they are too finely classified – e.g., it is harder to precisely calculate the rate of robberies with a knife relative to the overall rate of robberies with an incident based data set – and furthermore, the number of parameters to estimate grows quadratically in $|\mathbb{C}|$.

Our practical solution is to start by defining types based on the FBI's uniform categorization of crimes from which we choose a relatively heterogeneous subset of these types. We perform our analysis on six crimes: rape, robbery, burglary, motor vehicle theft, assault and light crime. Because of potential misclassification, we define assault as both aggravated and simple assault (Zimring (1998)). We classify criminal mischief, drunk and disorderly conduct, minor sexual offenses (e.g., public urination), vice (minor drug offenses and prostitution), fence (trade in stolen goods) and found property (almost exclusively cars and weapons) as light crimes.[15] Together, these six crimes comprise 55% of all police reports to the DPD

[15] We also conducted our full analysis defining only criminal mischief and found property as light crime,

during the sample period.[16]

We select this set of crimes for three reasons. First, this set includes both violent crimes and property crimes of varying levels of severity, which allows us to test for dynamic spillover effects of lighter crimes to more severe crimes. Second, these crimes occur relatively more frequently than other publicly observable crimes such as homicide and arson, which should yield more variation in our variables of interest. And third, these crimes are relatively accurately reported in comparison with crimes such as larceny and fraud.[17]

We support this choice by showing that our test can detect endogeneity stemming from the list of crimes not being exhaustive (see Section B.2.1) and from crimes being too coarsely classified (see Section B.2.2).[18]

5.2.2 Neighborhood Boundaries (j)

Neighborhoods are fundamentally difficult to define, especially when data is observed at high geographic detail, which allows for many different definitions that are equally plausible *ex ante*. Accordingly, we face a trade-off regarding our choice of j. On the one hand, we would like to define neighborhoods broadly (i.e., coarse j) in order to incorporate all spillover effects and to avoid contamination issues. For instance, defining j as a street block is likely too fine, as the intertemporal effects of crime on one street block may spill over to adjacent street blocks. On the other hand, if j is too coarsely defined, then we might not have observations with $\text{Crime}_{jt-1}^{x'} = 0$ for some or even all x's, which makes the test of exogeneity unfeasible (or at least less powerful).

Our practical solution is to start by defining neighborhoods based on the DPD's geographic classification scheme. During our sample period, the DPD geographically organized their policing area into six divisions subdivided into 32 sectors, which were further subdivided into 236 police beats.[19] Police beats range from roughly 0.5 to 2 square miles in area, while sectors range from roughly 5 to 10 square miles in area.

We define neighborhoods as either sectors or beats. We support this choice by show-

or alternatively defining criminal mischief only as light crime and obtained similar results.

[16]Roughly 25% of police reports in the database do not directly correspond to criminal acts per se (i.e., they declare lost property, report missing persons, report the failure of motorists to leave identification after auto damages, etc.) so the six crimes that we consider comprise a much larger majority of total crime in Dallas during the sample period.

[17]The accuracy of reported rape statistics is admittedly poor (Mosher et al. (2010)). As an added robustness check, we replicated our full analysis excluding rapes and obtained similar results.

[18]If we rejected all models, then we would have to redefine $\mathbb{C}$ to be more exhaustive (e.g., add larceny to the list), and disaggregated (e.g., treat burglary at night differently than burglary at day time).

[19]In October 2007, DPD added a seventh division to their classification and made slight modifications to some beat and sector boundaries. We end our sample in September 2007 to ensure that the administrative boundaries in our data set are geographically consistent over the entire sample period.

ing that our test can detect endogeneity stemming from j being too coarsely or too finely defined.[20]

5.2.3 Time Periods (t)

There is an important trade-off regarding our choice of temporal aggregation as well. We would like to choose t to be as short as possible in order to incorporate short-run intertemporal effects. However, this comes at a cost for two reasons. First, such a specification might miss longer-run intertemporal effects of crime. The obvious solution to this problem is to modify equation (9) to include additional lagged right-hand-side variables, but doing so will dramatically increase the number of parameters to be estimated. Second, if t is defined to be too short of a time period, then we will be unable to precisely measure local crime rates – e.g., the robbery rate in a neighborhood between 10AM and 10:05AM on 6/15/2003 – from an incident database.

Given these competing concerns, we define t as a week, which preserves substantial heterogeneity in neighborhood crime rates over time and provides a long time series (402 periods). We then show that our test can detect endogeneity stemming from t being overly aggregated and add lagged right-hand-side variables to equation (9) to demonstrate directly that any intertemporal effects subside fairly quickly.[21]

5.3 Choice of Controls

Finally, we must choose what to include in Controls_{jt}. Given the richness of our data set, the potential number of combinations of control variables is too large, so we need an "educated guess" of which sets of control variables have a chance of absorbing all confounding unobservables. Accordingly, we turn to the theory of social interactions to restrict the set of candidate models that we should consider.

Specifically, theory suggests that the two intertemporal links in crimes described in equation (9) – as encapsulated in β and Error_{jt} – operate at different levels of aggregation. While β is identified off variations at fine spatial *and* temporal levels, confounding effects tend to vary at fine spatial *or* temporal levels, but not both. Intertemporal effects of crimes propagate along individual and social learning networks, and social learning dissipates rapidly as social distance increases. Because social distance is strongly correlated to both spatial distance (Akerlof (1997)) and temporal distance (Ellison and Fudenberg (1995)), the bulk

[20] See Appendix Sections B.2.2 and B.2.3 available online.

[21] We opt for adding lags instead of choosing t = month because we otherwise might not have observations where $\text{Crime}_{jt-1}^{x'} = 0$ for some x', which would reduce the power of our test. In addition, adding lags allows for more heterogeneity in case the treatment effects decay over time.

of the causal response to a crime will likely remain close to the scene of the crime and be strongest in its immediate aftermath.[22] In contrast, most confounding determinants of crimes operate at more aggregated levels in at least one of these dimensions. For instance, the demographic composition of a neighborhood tends to change relatively slowly over time, and judicial institutions vary at larger geographic levels. Hence, we conjecture that these differences in aggregation should allow us to specify fixed effects that control for confounders of crime without absorbing the treatment effect that we want to measure.

To formalize this idea, we can describe a city as being composed of smaller geographic units (neighborhoods) indexed by j that can be further grouped into larger geographic units (regions) indexed by J. Similarly, our sample period can be divided into shorter time periods t that can be further grouped into longer time periods T. We consider models in which $\text{Controls}_{jt} = \lambda_{Jt} + \lambda_{jT}$ for different definitions of J and T, where λ_{Jt} is the unobserved component of crime that varies at a high frequency within a region and λ_{jT} is the unobserved component of crime that varies at a low frequency within a neighborhood. The λ_{Jt} fixed effects absorb all confounding factors that do not vary within region, and similarly, the λ_{jT} fixed effects absorb any neighborhood specific confounding factors that vary at the lower frequency T.

For a given level of j and t, the choices of J and T reflect the following trade-off: Finer choices of J and T relative to j and t imply fixed effects that absorb more confounding variation, which makes it more likely that any remaining identifying variation is "as good as random." However, as J and T approach the level of refinement of j and t, the number of covariates grows exponentially, taking the model closer to saturation. Our practical solution is to choose the coarsest values of J and T for which we fail to reject Assumption 1.

5.4 Summary Statistics

We present summary statistics aggregated to the sector-week level in Table 1. Not surprisingly, light crime is the most prevalent crime reported, followed by assault, burglary, auto theft, robbery and rape. In 69% of sector-week observations, zero past crimes of at least one type are reported.[23] The high prevalence of zeros in the explanatory variables is particularly valuable for our identification strategy, as it should yield smaller standard errors in our estimates of Δ, thereby resulting in a more powerful test of exogeneity.

[22] Block (1993) surveys this topic and shows that individuals' beliefs about neighborhood crime levels and even their beliefs about their own victimization have been repeatedly found to be subject to recency bias.

[23] When we define "past crimes" more flexibly (i.e., when we consider as explanatory variables $\text{Crime}^{x}_{jt-\tau}$ for some $x \in \mathbb{C}$ and $\tau \leq 6$), zero past crimes of at least one type are reported in 100% of sector-week observations. This more flexible definition of "past crimes" increases the power of the test, as shown in Section 5.2.1.

Police respond to crimes in approximately 80 minutes on average, although they respond to reports of rape roughly an hour slower and to reports of motor vehicle theft roughly half an hour faster. On average, police spend less than half an hour at the scene of a motor vehicle theft, but they spend up to an hour at the scenes of robberies and light crimes and over an hour at the scenes of reported rapes. All types of crimes occur slightly more frequently on weekends than weekdays with the exception of burglaries, which happen less frequently on weekends than weekdays. Just over half of robberies, light crimes and motor vehicle thefts occur at night, and as expected, a majority of these crimes take place outdoors. On the other hand, burglaries, assaults and rapes tend to occur indoors, with the first two occurring predominantly during the day time. Private businesses report approximate one fifth of robberies and light crimes and one third of burglaries, but they report very few motor vehicle thefts and no rapes or assaults.

Table 1: Summary Statistics: 2000-2007

Variable	Rape	Robbery	Burglary	Auto Theft	Assault	Light Crime
Avg. reported crimes in a sector per week	0.42 (0.69)	4.58 (3.18)	13.30 (7.56)	10.71 (6.01)	21.29 (12.07)	23.15 (10.15)
Avg. police response time (hours)	2.37 (1.45)	1.29 (1.00)	1.39 (0.73)	0.88 (0.66)	1.38 (0.74)	1.41 (0.72)
Avg. police duration (hours)	1.08 (1.61)	0.97 (1.73)	0.59 (0.81)	0.41 (0.68)	0.67 (0.71)	0.62 (0.61)
Frac. of crimes committed at night	0.62	0.55	0.36	0.50	0.45	0.43
Frac. of crimes committed outdoors	0.26	0.59	0.02	0.79	0.33	0.58
Frac. of crimes committed on the weekend	0.35	0.33	0.24	0.30	0.35	0.29
Frac. of crimes reported by private businesses	0.00	0.20	0.34	0.06	0.00	0.10
Total reported crimes	5,439	59,015	171,506	138,086	274,586	298,520

Notes: Standard deviations are presented in parentheses where relevant. Average police response time is measured from dispatch time to the officer's arrival. Average police duration is measured from the officer's arrival at the crime scene to their departure. Night time is defined as 8:00PM-8:00AM.

6 Empirical Results

6.1 Main Results

We consider models with $\mathbb{C} = \{$rape, robbery, burglary, auto theft, assault, light crime$\}$, $j \in \{$beat, sector$\}$ and $t =$ week. For Controls$_{jt}$, we consider 6 specifications. In specification (1), we include no controls. In specification (2), we include type of crime fixed effects. In specification (3), we include year-type of crime fixed effects. This specification is closely related to previous attempts to identify intertemporal relationships between crimes (Funk and Kugler (2003)) and between crime and policing (Corman and Mocan (2005)), as they utilize only low-frequency control variables such as annual unemployment rates, which are likely absorbed by these fixed effects. In specification (4) we add neighborhood-type of crime fixed effects to specification (3) in order to absorb any neighborhood characteristics that did not change over the sample period. In specification (5), we include both week-type of crime and neighborhood-type of crime fixed effects. Finally, in specification (6), we include both division-week-type of crime and neighborhood-year-type of crime fixed effects.

Table 2: Tests of Exogeneity

	Specifications (p-values in parentheses)					
	(1)	(2)	(3)	(4)	(5)	(6)
$j =$ Beat, $t =$ Week	53.06	5.84	5.86	5.85	5.91	6.39
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$j =$ Sector, $t =$ Week	11.21	1.78	2.05	1.73	1.46	**0.88**
	(0.00)	(0.01)	(0.00)	(0.01)	(0.05)	**(0.66)**

Notes: This table shows the F statistic and p-value of the test of exogeneity described in Section 3 for various specifications of equation (10). Entries in bold denote "surviving models" for which we cannot reject exogeneity at typical significant levels. Each one specifies fixed effects at different levels: (1) no fixed effects; (2) fixed effects at the c level; (3) fixed-effects at the year $\times$ c level; (4) fixed-effects at the year $\times$ c and at the $j \times c$ levels; (5) fixed effects at the $t \times c$ and at the $j \times c$ levels (6) fixed effects at the $J \times t \times c$ and at the $j \times T \times c$ levels, where $J =$ division, $T =$ year. All standard errors are clustered at the $j \times$ year $\times$ c level.

Table 2 contains the F-statistics and respective p-values for each test of exogeneity performed. We are able to reject exogeneity for all but one model at standard critical levels.[24] The surviving model is

[24]Model (5) for $j =$ sector and $t =$ week is only marginally rejected at the 5% level of significance. When we look at the elements of Δ individually, we find that 3 out of 30 of them are statistically significantly different from zero at the 99% level. At this level of statistical significance, we would expect to reject 0.3 of the coefficients at random. In the interest of conducting a conservative analysis, we reject exogeneity for this model. As a comparison, in model (6) for $j =$ sector and $t =$ week, zero elements of Δ are statistically significant at the 99% level.

$$
\begin{aligned}
\mathbb{C} &= \{\text{rape, robbery, burglary, auto theft, assault, light crime}\} \\
j &= \text{sector} \\
t &= \text{week} \\
\text{Controls}_{jt} &= \{\lambda_{\text{division-week-type of crime}}, \lambda_{\text{sector-year-type of crime}}\}
\end{aligned}
$$

The division-week-type of crime fixed effects absorb all time varying determinants of each crime that vary across the six police divisions of Dallas, and the sector-year-type of crime fixed effects absorb all neighborhood specific determinants of each crime that vary on an annual basis (e.g., demographic characteristics).[25] In short, the only potential omitted variable that could bias our estimates would have to vary across weeks within a calendar year and across sectors within a division. Because we find no evidence of such omitted variables from the test of exogeneity, the estimates of this model can be interpreted as causal (provided the test is powerful enough). For the sake of exposition, we interpret our estimates as such throughout this section before presenting supporting evidence in Section 7 and Appendix B.

In Table 3 we present estimates of the intertemporal effects of crime (β) for the lone surviving model. All coefficients are precisely estimated with the exception of the rape coefficients, whose standard errors are relatively large due to the lack of variation in reported rape levels. We find small within-crime intertemporal effects for robbery, burglary, auto theft and light crime, indicating that such crimes will generate an additional 0.05-0.15 crimes of that type in the following week. We find less evidence for across-crime intertemporal effects although we do find some positive effects at the 95% confidence level, mostly in the direction of decreasing severity. Notably, we find no evidence that light crime leads to more severe crimes. With this model, we are able to explain 87% of the variation in reported weekly neighborhood crime levels.

[25] Sector specific unobservable amenities that are changing over time due to gentrification will be absorbed by these fixed effects to the extent that they vary across years in the sample.

Table 3: Intertemporal Effects of Crimes

	Rape_t	Robbery_t	Burglary_t	Auto Theft_t	Assault_t	Light Crime_t
Rape_{t-1}	0.001	0.077	-0.121	-0.083	-0.006	0.128
	(0.025)	(0.072)	(0.147)	(0.127)	(0.200)	(0.204)
Robbery_{t-1}	-0.006	0.060**	0.013	0.026	0.064	0.081**
	(0.003)	(0.013)	(0.024)	(0.016)	(0.030)	(0.029)
Burglary_{t-1}	-0.001	0.010	0.153**	0.009	0.009	0.046**
	(0.002)	(0.006)	(0.013)	(0.010)	(0.015)	(0.016)
Auto Theft_{t-1}	-0.001	-0.006	0.012	0.084**	0.029	0.025
	(0.002)	(0.006)	(0.013)	(0.012)	(0.017)	(0.019)
Assault_{t-1}	0.002	0.010	-0.004	0.003	0.016	0.012
	(0.001)	(0.004)	(0.007)	(0.007)	(0.012)	(0.011)
Light Crime_{t-1}	0.002	0.002	0.003	0.008	0.016	0.060**
	(0.001)	(0.004)	(0.008)	(0.007)	(0.010)	(0.013)
R^2			0.867			
Number of Observations:			77,184			

Notes: This table shows the estimated intertemporal effects of various crimes in week $t-1$ on crime levels in week t (i.e., the parameter matrix β). Fixed effects at the division-week-type of crime and sector-year-type of crime are included in each of the six equations, which are estimated simultaneously by seemingly unrelated regression. All errors are clustered at the sector-year-type of crime level. ** - significant at the 99% level, * - significant at the 95% level.

Table 4: Full Long-Run Reduction in Crime^y from a One-Week Elimination of Crime^x

	Effect on Crime^y					
Crime^x	Rape	Robbery	Burglary	Auto Theft	Assault	Light Crime
Eliminate 0.42 Rapes	0.1%	-6.3%	-2.0%	-0.1%	11.2%	6.5%
	[-4.1, 5.9]	[-13.0, 3.3]	[-12.6, 13.5]	[-10.8, 12.8]	[-1.0, 28.1]	[-8.8, 28.1]
Eliminate 4.58 Robberies	0.7%	6.5%	3.6%	-1.4%	5.0%	1.4%
	[-0.5, 2.6]	[3.7, 10.2]	[-0.9, 10.0]	[-4.6, 3.2]	[0.9, 10.7]	[-3.9, 9.0]
Eliminate 13.30 Burglaries	-0.4%	0.6%	18.2%	1.2%	-0.8%	1.1%
	[-1.4, 0.9]	[-1.3, 3.2]	[14.8, 22.8]	[-1.2, 4.6]	[-3.3, 2.8]	[-2.6, 6.4]
Eliminate 10.71 Auto Thefts	-0.3%	1.3%	1.6%	9.2%	0.6%	2.4%
	[-1.3, 1.0]	[-0.1. 3.4]	[-1.5, 5.9]	[6.6, 12.7]	[-2.1, 4.5]	[-1.4, 7.7]
Eliminate 21.29 Assaults	-0.0%	1.6%	0.8%	1.7%	1.8%	2.1%
	[-0.7, 1.0]	[0.2, 3.4]	[-1.3, 3.9]	[-0.1, 4.2]	[-0.6, 5.0]	[-0.5, 5.7]
Eliminate 23.15 Light Crimes	0.3%	1.6%	3.0%	1.2%	1.1%	6.0%
	[-0.3, 1.2]	[0.5, 3.2]	[1.1, 5.8]	[-0.5, 3.6]	[-0.8, 3.7]	[3.1, 9.9]

Note: Reductions are calculated by hypothetically eliminating all crime of type x in the average neighborhood in the sample for one week, computing the total number of future crimes of each type y that is reduced in that neighborhood and dividing by the average number of weekly crimes of type y in a neighborhood in the sample. Positive values correspond to long- run reductions in crime. 95% confidence intervals of these effects are presented below in brackets. For example, eliminating light crime in the average neighborhood for a week (a reduction of 23.15 light crimes) will generate a future reduction in robberies equal to 1.6% of the average number of weekly robberies in a neighborhood in the sample (about 0.07 robberies). This effect lies between a 3.2% and a 0.5% reduction in robberies with 95% probability (a reduction of between 0.02 and 0.14 robberies).

We provide more context for these effects by expressing them as semi-elasticities of one-week crime elimination in Table 4. Specifically, we compute the total spillover effect of eliminating all crimes of a given type x in an average week on each crime of type y. We express this effect as a percentage of crime of type y in a single week in an average Dallas neighborhood.[26] As expected, the point estimates of all semi-elasticities tend to be extremely small. For example, a complete elimination of light crime in week t in the average neighborhood – a reduction of 23.15 light crimes in that week – will generate a total future reduction in robberies aggregated over weeks $t+1, t+2, \ldots$ that is equal to only 1.6% of the number of robberies in a single week in the average neighborhood in Dallas (about 0.07 robberies). Even though the majority of these effects are not statistically distinguishable from zero, the precision of our estimates allows us to rule out even modest intertemporal spillovers. For example, with 95% confidence we can rule out a spillover reduction in robberies of over 3.2% from the elimination of all light crime, which corresponds to about 0.14 robberies. Hence,

[26]This calculation takes into account all cumulative effects, including indirect ones. For instance, if eliminating robberies in week t reduces burglaries in week $t+1$, then the calculation associated with the elimination of robberies in t will incorporate the reduction of assaults in week $t+2$ that was due to the corresponding reduction of burglaries in week $t+1$.

we can comfortably rule out the sizable self-sustaining reductions in severe crimes promised by proponents of the "Broken Windows" theory.

Remark 2. Our identification strategy is based on the premise that we should expect that any confounder will be absorbed as we add controls, otherwise the test of exogeneity would detect its presence. However, this might not be the case if the standard errors of $\hat{\Delta}$ also increased with the addition of controls. In that case, a discontinuity in $\mathrm{E}\left[w|\mathrm{Crime}^x_{jt-1} = d\right]$ at $d = 0$ would be wrongly interpreted as continuous; i.e., elements of $\mathbb{W}^D_2$ would be erroneously understood to be elements of $\mathbb{W}^C$. To check if this is the case, we present the distribution of the standard errors of all elements of $\hat{\Delta}$ for models (1)-(6) in Appendix Figure 14. In practice, the standard errors do not seem to increase as more detailed fixed effects are added in the models that we consider. This is not surprising since the addition of controls is simply an addition of incidental parameters to the regression, so it does not necessarily affect inference on the parameters of Δ, which remain fixed across all models.

6.2 Longer-Run Effects: Dynamic Spillovers of Crime

Criminal behavior may lead to intertemporal effects that extend to more than a single period in the future. If that were the case, then the equations of motion in (10) would be misspecified, which would affect our interpretation of β as the full intertemporal effects of crime, and which could also lead to an endogeneity problem. To explore these issues, we generalize equation (10) as

$$\mathrm{Crime}_{jt} = \sum_{\tau=1}^{\bar{\tau}} \beta_\tau \mathrm{Crime}_{jt-\tau} + \sum_{\tau=1}^{\bar{\tau}} \Delta_\tau D_{jt-\tau} + \Gamma \mathrm{Controls}_{jt} + \epsilon_{jt} \tag{11}$$

where $\bar{\tau}$ captures the maximum duration of direct, long-run effects of crime.

Table 5 presents the results of two tests for various values of $\bar{\tau}$ and specifications of fixed effects for j = sector and t = week. First, we perform the most powerful test of exogeneity available by jointly testing whether all $|\mathbb{C}|^2 \cdot \bar{\tau}$ elements of $\Delta_1, \ldots, \Delta_{\bar{\tau}}$ are equal to zero, and present the p-value of this test at the top of each cell. Only the preferred model (6) survives this test, and it does so for any choice of $\bar{\tau}$. The fact that no other specification survive for any choice of $\bar{\tau}$ can be understood as a falsification test in favor of model (6). Second, we test whether the elements of β_τ, for all $\tau \leq \bar{\tau} - 1$ are the same as their counterparts in the corresponding model with only $\bar{\tau} - 1$ lags and present the p-value of this test at the bottom of each cell in brackets. For instance, when we add a fourth lag to the model with three lags, we jointly test whether any of the 36 coefficients in each of the first three lags change. These parameters do not differ in our preferred model (6) for any values of $\bar{\tau}$, which

further restricts the set of potential confounders in model (6). For instance, any confounding omitted variable that is not detected by our test (i.e., an element of $\mathbb{W}_2^D \cup \mathbb{W}^C$) would need to be correlated to Crime_{jt-1} and to Crime_{jt} but also uncorrelated to $\text{Crime}_{jt-2}, \ldots, \text{Crime}_{jt-6}$. This is a difficult condition to meet since it rules out persistent confounders.[27]

Table 5: Sensitivity Tests: Longer Run Effects

	Specifications of Fixed Effects					
Num. of Included Lags ($\bar{\tau}$)	(1)	(2)	(3)	(4)	(5)	(6)
$\bar{\tau} = 2$	0.00	0.00	0.00	0.00	0.00	**0.64**
	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	**[1.00]**
$\bar{\tau} = 3$	0.00	0.00	0.00	0.00	0.00	**0.21**
	[0.00]	[0.00]	[0.00]	**[0.94]**	**[1.00]**	**[1.0]**
$\bar{\tau} = 4$	0.00	0.00	0.00	0.00	0.00	**0.25**
	[0.00]	[0.00]	[0.00]	**[1.00]**	**[1.00]**	**[1.00]**
$\bar{\tau} = 5$	0.00	0.00	0.00	0.00	0.00	**0.39**
	[0.00]	[0.00]	[0.00]	**[1.00]**	**[1.00]**	**[1.00]**
$\bar{\tau} = 6$	0.00	0.00	0.00	0.00	0.00	**0.18**
	[0.06]	**[0.37]**	**[0.39]**	**[1.00]**	**[1.00]**	**[1.00]**

Notes: This table shows the p-values of two tests associated with the generalized equation of motion of crime (11), with various fixed effects (corresponding to the columns) and lags (corresponding to the rows). The first p-value listed in each cell is for the test of exogeneity described in Section 3 for $\Delta_1, \ldots, \Delta_{\bar{\tau}}$ (p-values in bold denote "surviving models" for which we cannot reject exogeneity at the 5% level.) The second p-value listed in each cell is for a test of whether all $36 \times (\bar{\tau} - 1)$ elements of β_τ, $\tau = 1, ..., \bar{\tau} - 1$ in the listed model are equal to the respective elements of β_τ, $\tau = 1, ..., \bar{\tau} - 1$ when the lag $\bar{\tau}$ is excluded (p-values in bold denote we cannot reject that all parameters are the same at the 5% level.) The specifications of controls are the same as the ones described in Table 2 for j = sector and t = week. Each one specifies fixed effects at different levels: (1) no fixed effects; (2) fixed effects at the c level; (3) fixed-effects at the year $\times$ c level; (4) fixed-effects at the year $\times$ c and at the $j \times c$ levels; (5) fixed effects at the $t \times c$ and at the $j \times c$ levels (6) fixed effects at the $J \times t \times c$ and at the $j \times T \times c$ levels, where J = division, T = year. All errors are clustered at the $j \times$ year $\times$ c level.

Because the appropriate value of $\bar{\tau}$ is not obvious *ex ante*, we check for long-run effects by choosing different values of $\bar{\tau}$ and testing whether we can reject that all 36 elements of $\beta_{\bar{\tau}} = 0$. We present the results of these tests in Table 6, which shows that for the surviving model (6), $\bar{\tau} = 4$ plausibly captures all long-run effects.[28]

[27] The second test is unable to reject models (4) and (5) for $\bar{\tau} \geq 3$, which suggests that our test of exogeneity is more powerful than this standard robustness check in this particular application.

[28] Note that this test is meaningful only for model (6) since we are unable to interpret the estimates of models (1)-(5) as causal. Nevertheless, we present all estimates for completeness. The fact that the estimated spillovers from the non-surviving models last six weeks or more when the spillovers estimated from model (6) do not is additional evidence that these non-surviving models are biased by persistent confounders.

Table 6: Should the $\bar{\tau}^{\text{th}}$ lag be included?

	Specifications of Fixed Effects					
Number of Included Lags ($\bar{\tau}$)	(1)	(2)	(3)	(4)	(5)	(6)
$\bar{\tau} = 2$	0.00	0.00	0.00	0.00	0.00	0.00
$\bar{\tau} = 3$	0.00	0.00	0.00	0.00	0.00	0.01
$\bar{\tau} = 4$	0.00	0.00	0.00	0.00	0.00	0.01
$\bar{\tau} = 5$	0.00	0.00	0.00	0.00	0.00	**0.43**
$\bar{\tau} = 6$	0.00	0.00	0.00	0.00	0.00	**0.57**

Notes: This table shows the p-value for a test of whether all 36 elements of $\beta_{\bar{\tau}} = 0$ for various specifications of equation (11) (p-values in bold denote that we cannot reject that all parameters are zero at the 5% level.) These 6 specifications are the same as the ones described in Table 2 for j = sector and t = week. Each one contains fixed effects at different levels: (1) no fixed effects; (2) fixed effects at the c level; (3) fixed-effects at the year $\times$ c level; (4) fixed-effects at the year $\times$ c and at the $j \times c$ levels; (5) fixed effects at the $t \times c$ and at the $j \times c$ levels (6) fixed effects at the $J \times t \times c$ and at the $j \times T \times c$ levels, where J = division, T = year. All errors are clustered at the $j \times$ year $\times$ c level.

In Table 7 we present estimates of $\hat{\beta}_1, \ldots, \hat{\beta}_4$ of our surviving model. Our results are similar as before. We do find some statistically significant long-run within crime intertemporal effects for burglary and auto theft. However, we find no systematic evidence for long-run across crime intertemporal effects.[29] Moreover, these results indicate that (direct) intertemporal effects last at most 4 weeks. Although we cannot rule out the possibility of effects that lasts for up to four weeks, disappear in the medium run (5th and 6th weeks) and reappear after that (7th week and later), the fact that the magnitude of the estimates in Table 7 decay for larger τ suggests that such hypothetical effects, if they exist, are unlikely to be economically significant.

[29] Our finding that $\beta_1 \neq \beta_2, \beta_3, \beta_4$ also serves as additional support for t = week versus t = month.

Table 7: Intertemporal Effects of Crimes: 4 Lags (1 of 2)

(a) $\hat{\beta}_1$

	Rape$_t$	Robbery$_t$	Burglary$_t$	Auto Theft$_t$	Assault$_t$	Light Crime$_t$
Rape$_{t-1}$	0.001	0.078	-0.121	-0.086	-0.006	0.108
	(0.025)	(0.074)	(0.145)	(0.127)	(0.201)	(0.206)
Robbery$_{t-1}$	-0.006	0.056**	0.005	0.025	0.061*	0.084**
	(0.003)	(0.012)	(0.024)	(0.016)	(0.030)	(0.029)
Burglary$_{t-1}$	-0.001	0.008	0.133**	0.004	0.003	0.041**
	(0.002)	(0.007)	(0.012)	(0.010)	(0.014)	(0.015)
Auto Theft$_{t-1}$	-0.001	-0.006	0.008	0.074**	0.028	0.021
	(0.002)	(0.006)	(0.013)	(0.011)	(0.017)	(0.019)
Assault$_{t-1}$	0.002	0.008*	-0.007	0.001	0.017	0.009
	(0.001)	(0.004)	(0.007)	(0.007)	(0.012)	(0.011)
Light Crime$_{t-1}$	0.001	0.001	0.001	0.006	0.015	0.058**
	(0.001)	(0.004)	(0.008)	(0.007)	(0.009)	(0.013)

(b) $\hat{\beta}_2$

	Rape$_t$	Robbery$_t$	Burglary$_t$	Auto Theft$_t$	Assault$_t$	Light Crime$_t$
Rape$_{t-2}$	-0.030	0.091	0.210	-0.041	0.113	-0.106
	(0.021)	(0.077)	(0.129)	(0.128)	(0.192)	(0.183)
Robbery$_{t-2}$	0.003	0.026*	0.018	-0.002	0.002	0.029
	(0.003)	(0.012)	(0.020)	(0.017)	(0.029)	(0.035)
Burglary$_{t-2}$	0.003	0.005	0.087**	0.010	0.020	0.019
	(0.002)	(0.006)	(0.011)	(0.010)	(0.014)	(0.015)
Auto Theft$_{t-2}$	-0.003	0.012	0.013	0.041**	-0.003	0.037
	(0.002)	(0.007)	(0.012)	(0.011)	(0.019)	(0.017)
Assault$_{t-2}$	0.002	0.001	0.019*	0.004	0.002	0.013
	(0.001)	(0.005)	(0.009)	(0.007)	(0.012)	(0.012)
Light Crime$_{t-2}$	-0.001	0.003	-0.001	0.007	0.013	0.018
	(0.001)	(0.004)	(0.008)	(0.006)	(0.010)	(0.011)

Table 7: Intertemporal Effects of Crimes: 4 Lags (2 of 2)

(a) $\hat{\beta}_3$

	Rape$_t$	Robbery$_t$	Burglary$_t$	Auto Theft$_t$	Assault$_t$	Light Crime$_t$
Rape$_{t-3}$	-0.032	0.075	0.193	-0.118	0.049	0.030
	(0.022)	(0.086)	(0.148)	(0.131)	(0.167)	(0.204)
Robbery$_{t-3}$	0.002	0.031**	0.025	-0.017	0.00	0.003
	(0.003)	(0.012)	(0.022)	(0.019)	(0.025)	(0.027)
Burglary$_{t-3}$	0.001	0.005	0.036**	0.011	0.037*	0.011
	(0.002)	(0.006)	(0.011)	(0.009)	(0.015)	(0.016)
Auto Theft$_{t-3}$	0.001	-0.006	0.006	0.032**	-0.013	-0.012
	(0.002)	(0.006)	(0.012)	(0.011)	(0.016)	(0.017)
Assault$_{t-3}$	-0.000	0.005	-0.010	0.004	-0.001	0.016
	(0.001)	(0.004)	(0.008)	(0.007)	(0.011)	(0.011)
Light Crime$_{t-3}$	0.001	0.001	0.010	0.002	0.002	0.022*
	(0.001)	(0.004)	(0.007)	(0.006)	(0.011)	(0.011)

(b) $\hat{\beta}_4$

	Rape$_t$	Robbery$_t$	Burglary$_t$	Auto Theft$_t$	Assault$_t$	Light Crime$_t$
Rape$_{t-4}$	-0.037	-0.062	0.213	-0.001	-0.094	-0.313
	(0.022)	(0.079)	(0.145)	(0.132)	(0.182)	(0.229)
Robbery$_{t-4}$	-0.004	0.008	-0.004	0.005	0.016	0.049
	(0.003)	(0.013)	(0.022)	(0.018)	(0.025)	(0.030)
Burglary$_{t-4}$	-0.000	-0.002	0.017	0.004	-0.048**	-0.014
	(0.002)	(0.006)	(0.011)	(0.009)	(0.014)	(0.016)
Auto Theft$_{t-4}$	-0.000	-0.010	0.008	0.018	0.028	0.027
	(0.002)	(0.007)	(0.013)	(0.011)	(0.016)	(0.019)
Assault$_{t-4}$	0.001	0.006	-0.005	-0.013	-0.003	0.008
	(0.001)	(0.006)	(0.008)	(0.007)	(0.012)	(0.012)
Light Crime$_{t-4}$	-0.000	-0.009*	0.004	0.001	0.010	0.010
	(0.001)	(0.004)	(0.007)	(0.006)	(0.008)	(0.011)

Notes: These tables show the estimated intertemporal effects of various crimes in weeks $t-1, \ldots, t-4$ on crime levels in week t (i.e., the parameter matrices $\hat{\beta}_1, \ldots, \hat{\beta}_4$ from equation (11)). Fixed effects at the division-week-crime type and sector-year-crime type are included in each of the six equations, which are estimated simultaneously by seemingly unrelated regression. The F-statistic for the discontinuity test over 144 indicator variables is 1.09 (p-value is 0.25). $N = 76,608$, $R^2 = 0.907$. All errors are clustered at the sector-year-crime type level. ** - significant at the 99% level, * - significant at the 95% level

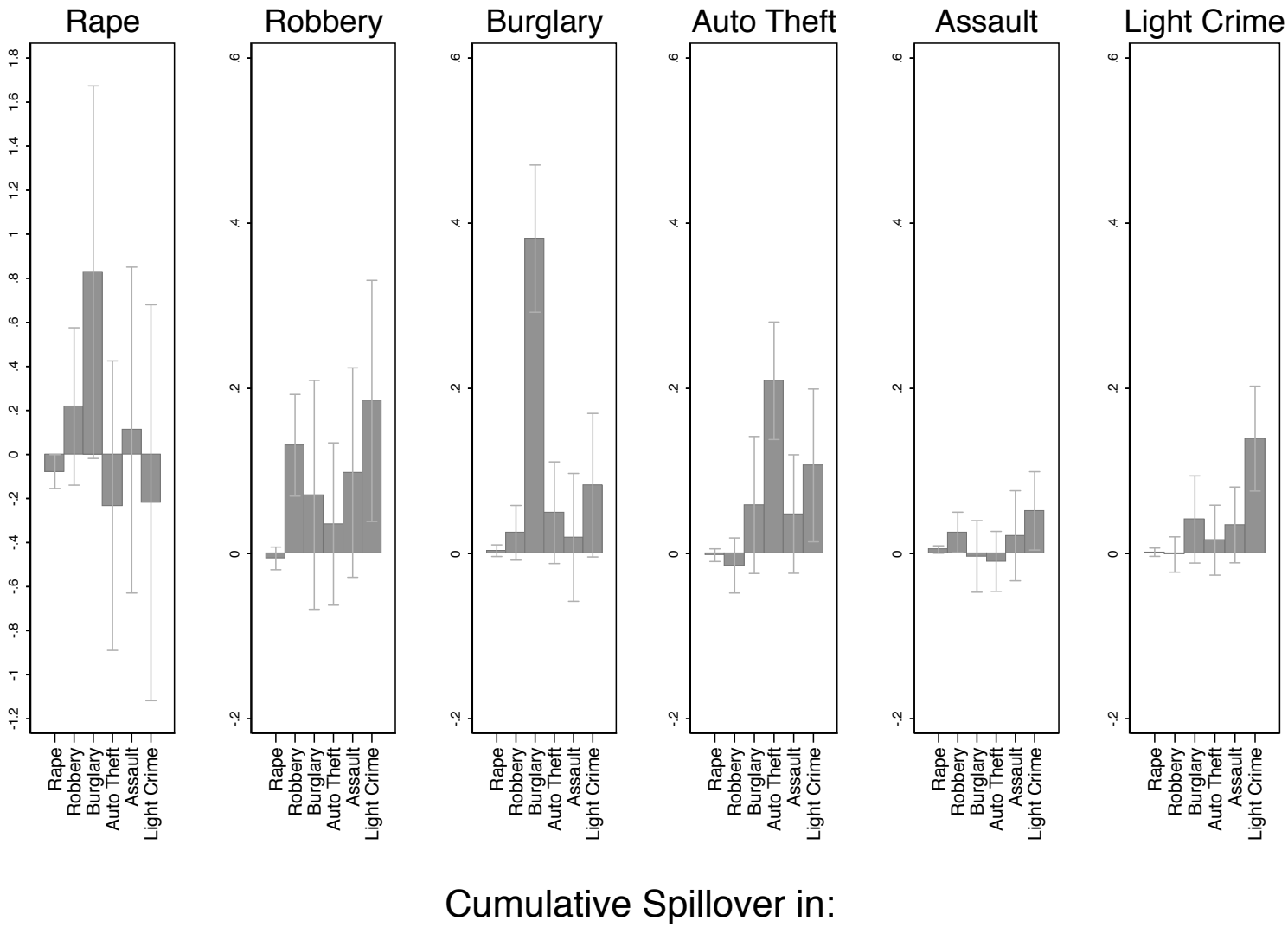
A key feature of a dynamic model of crime is that short-lived direct effects of crime may

generate longer lasting indirect effects.[30] In order to explore such richer dynamic interactions that are captured in the model described in equation (11), we use our coefficient estimates to perform a simulated experiment in which we reduce one reported crime of a given type in week 0 and then simulate the evolution of all reported crimes in weeks $1, 2, \ldots$ holding all else constant.[31] We then compute the cumulative change in the levels of all crimes relative to how they would have evolved in the absence of the counterfactual reduction. We interpret the cumulative simulated changes in future crime levels as the dynamic spillovers that are associated with reductions in current crime levels holding all else constant except the endogenous behavioral responses to crime.

[30]Indeed, Gladwell (2000) has popularized the notion that the "Broken Windows" theory implies the existence of a "tipping point" level of light crime beyond which the levels of light crime and more severe crimes are on an ever increasing trajectory. Our findings are inconsistent with this view.

[31]In particular, we hold constant the current arrest and incarceration policies used by law enforcement.

Figure 5: Long Run Cumulative Spillovers From Unit Crime Reductions



Note: Each of the six panels refers to each type of crime that was reduced by one unit. Vertical bars represent 95% confidence intervals for long run cumulative spillovers calculated via the delta method.

We present the results of this simulation exercise in Figure 5, which shows the full dynamic spillovers of unit crime reductions along with 95% confidence intervals.[32] The label above each panel refers to the type of crime that we hypothetically reduce by one unit, and the labels for each bar refer to the type of crime that experiences the spillover. Note that the y-axis for rapes is at a different scale from the y-axis for the other crimes, since the estimated spillovers for rapes as an explanatory variable are relatively imprecise. It is immediate that all within-crime dynamic spillovers are statistically significant, with the exception of assault, and these spillovers tend to be large relative to across-crime dynamic spillovers, with the exceptions of rape and assault.

Our finding of no statistically significant across-crime dynamic spillovers associated with reductions in light crime suggests that a "Broken Windows" policy will have little success reducing the levels of severe crimes. For perspective, we note that the spillover benefits associated with a policy that targets either robbery or auto theft strictly dominate the spillover benefits of a policy that targets light crime, as the across-crime effects of reducing robbery and auto theft on light crime are of the same order of magnitude as the within-crime effect of reducing light crime. A policy that targets assaults generates spillover reductions in light crime that are smaller than the within-crime spillovers associated with light crime reduction, but this policy also generates positive spillover reductions in rape and robbery levels. However, this policy generates no within-crime spillover. Even though a policy that targets burglaries does not generate across-crime spillovers, it generates the largest within-crime positive spillovers of all of the crimes. [33]

Remark 3. In Appendix D, we present the simulated impulse response functions of all crimes from unit crime reductions of each type in Figure 13. For brevity, we only present these impulse response functions using our preferred specification of the equations of motion for crime with up to four lags of each explanatory variable ($T = 4$). Six weeks after the unit crime reduction, nearly all long run spillovers are realized.

Remark 4. Our findings speak directly to the effectiveness of various crime reduction policies. In order to discuss the efficiency of crime reduction policies, we perform a rough cost-benefit analysis of various unit crime reduction policies that incorporates external information on the social benefits of reducing various types of crime and present the results in Table 8. For the social benefits of crime reduction, we use estimates from external studies that attempt to

[32] Because our equations of motion are linear in $X_{jt-1}, \ldots, X_{jt-T}$, the cumulative long run spillovers can be computed analytically. The standard errors for these spillovers are calculated using the delta method, which accounts for the correlations among the elements of β_k^c for all c and k.

[33] We are unable to assess the benefits of a hypothetical policy that targets rapes due to imprecision in our estimates of the dynamic spillovers associated with such a policy. Nevertheless, the inclusion of rape in our analysis is important because we are able to precisely estimate the intertemporal effects of other crimes on rape.

account for the physical and psychic costs to victims of crime and the psychic costs to society at large. Because external estimates of the costs of crime reduction are not available in the literature, we express all benefits of crime reduction relative to light crime reduction. This exercise reveals that targeting light crime would be an efficient policy to combat all crimes only if it was dramatically more expensive to target rape, robbery, burglary, auto theft and assault (95, 32, 7, 5 and 18 times as expensive as light crime, respectively). Of course, these conclusions should be qualified by the uncertainty surrounding the social benefits of crime and our lack of data regarding the costs of combatting crime.

Table 8: Estimated Monetary Benefits of Unit Crime Reduction

Crime	Total Benefits from Unit Crime Reduction ($)	Light Crime Monetary Equivalents
Rape	240,942	95.0
	[197,943; 301,846]	
Robbery	80,506	31.8
	[73,167; 90,779]	
Burglary	18,390	7.3
	[14,210; 24,318]	
Auto Theft	11,956	4.7
	[7,793; 17,897]	
Assault	45,583	18.0
	[42,976; 49,239]	
Light Crime	2,535	1
	[-2; 6,178]	

Notes: 95% confidence intervals for total benefits from unit crime reduction are presented in brackets. The social costs of rape, robbery, burglary and auto theft are taken from Heaton (2010). We compute the social cost of all assaults by taking an average of the social cost of aggravated assault from Heaton (2010) and the social cost of simple assault from Miller et al. (1993) weighed by the relative share of aggravated assaults in our sample (22.83%). We are unable to obtain estimates of the social cost of light crime, so we assume it to be half of the social cost of larceny as given in Heaton (2010). All monetary amounts are in 2015 dollars.

7 Sensitivity Analysis

In this section, we provide robustness checks that are complementary to the test of exogeneity, in the sense that potential sources of endogeneity that are undetectable by the test of exogeneity can still be detected by these further checks.

7.1 Alternative Specifications

We can take advantage of the richness of our data set to explore alternative specifications that build on the models estimated in the previous section. For each previously estimated model, we enrich Controls_{jt} with additional variables and then perform two tests: the test of exogeneity and a second test of whether *any* of the parameter estimates of $\hat{\beta}_\tau$ change with the enriched set of controls for $\tau \leq \bar{\tau}$. In Table 9, we present the results of these tests for $\bar{\tau} = 1$.[34] In all models in the first row, we include 180 additional control variables related to the salience of crimes. In all models in the second row, we include 72 additional control variables that attempt to proxy for unobserved police attention in the neighborhood in period t. In all models in the third row, we add the levels of each type of crime in the nearest adjacent neighborhood (36 variables) as control variables.

When we subject each of these models to our test of exogeneity, only the models specifying fixed effects as in (6) survive. Moreover, when we conduct conventional robustness checks to see if the inclusion of these control variables affects our estimates of the treatment effects of interest (p-values in brackets), we find that our initial estimates are statistically unchanged in model (6) and in some other models. This suggests that our test of exogeneity is more powerful than standard robustness checks in our application. Furthermore, we find direct evidence that the fixed effects specified in (1)-(5) are unable to absorb the endogeneity related to contamination:[35] we estimate statistically significant across-neighborhood effects in these models (with a p-value of 0.00 in each specification). This is consistent with the exogeneity test results presented in Table 2, which cautions us not to interpret those estimates as causal.[36] In contrast, the fixed effects specified in (6) successfully control for such confounders, as we cannot reject that all across-neighborhood effects are equal to zero in this model (with a p-value of 0.99).

In the fourth row, we explicitly allow for non-linear effects of crime by estimating a linear b-spline in past crimes with knots at the median levels of each type of crime. Once again, the modified model (6) is still the only survivor of the exogeneity test. Moreover, when we test for whether the coefficients corresponding to the portion of the support below the median are equal to the corresponding coefficients of the previously estimated linear specification (in brackets), we find no evidence of non-linear treatment effects in our preferred model.[37]

[34] Our conclusions are unchanged for $\bar{\tau} \leq 6$. These results are available upon request.

[35] Section B.2.3 discusses the potential endogeneity issue of contamination further.

[36] In row 3 of Table 9, the test of exogeneity we perform represents a test for whether the 72 coefficients representing spillovers (the 36 elements of β plus the other 36 analogous elements pertaining to the adjacent neighborhood) can be interpreted as causal. Thus, the exogeneity test results in this table suggest that we should not interpret the coefficients representing dynamic spillovers to adjacent neighborhoods in models (1)-(5) as causal.

[37] As discussed in Remark 9, this finding is further evidence that our test has power to detect confounders

Table 9: p-Values for Sensitivity Tests Under Alternative Specifications

	Specifications of Fixed Effects					
	(1)	(2)	(3)	(4)	(5)	(6)
Add 180 Salience Controls (Num. of each crime committed outdoors, at night time, on the weekend; average police response times and durations at crime scene in period $t-1$)	0.00	0.01	0.00	0.01	0.04	**0.58**
	[0.00]	[0.00]	[0.00]	**[0.98]**	**[0.99]**	**[1.00]**
Add 72 Contemporaneous Policing Controls (Avg. police response times and durations at crime scene in period t)	0.00	0.00	0.00	0.00	0.01	**0.62**
	[0.44]	**[0.96]**	**[0.95]**	**[1.00]**	**[1.00]**	**[1.00]**
Add levels of each crime in nearest adjacent neighborhood (36 variables)†	0.00	0.00	0.00	0.01	0.01	**0.89**
	[0.05]	**[1.00]**	**[1.00]**	**[1.00]**	**[1.00]**	**[1.00]**
Non-Linear Treatment Effects (36 variables)††	0.00	0.03	0.02	0.01	0.04	**0.54**
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	**[0.17]**

Notes: This table shows the p-values of two tests for various specifications of equation (8) as described in Table 2 for j = sector and t = week. All additional controls are specified additively. The first p-value listed in each cell is for the test of exogeneity described in Section 3 (p-values in bold denote "surviving models" for which we cannot reject exogeneity at the 5% level.) The second p-value listed in each cell (in brackets) is for a test of whether all 36 elements of β in the listed specification are equal to their respective element presented in Table 3 (p-values in bold denote that we cannot reject that all of our results do not change at the 5% level.) All errors are clustered at the $j \times$ year $\times\, c$ level.

†: The first p-value refers to a test of whether the coefficients of D_{jt-1} and of $D_{j't-1}$ are jointly equal to zero where j' is the nearest neighborhood to j.

††: Non-linear treatment effects are specified with a linear b-spline with a knot at the median level of each type of crime. The second p-value refers to a test of whether the coefficient corresponding to the portion of the support below the median is equal to the coefficient of the linear specification.

7.2 Spatially and Serially Correlated Errors

As pointed out by Harcourt and Ludwig (2006), the primary empirical issue in estimating crime spillovers is the fact that many potential confounders of crime are either spatially autocorrelated or serially correlated (or both), which generates an endogeneity problem in any specification of the equations of motion of crime that has inadequate controls. Estimated residuals that are spatially and serially uncorrelated in a particular model are evidence that included controls absorb all such sources of endogeneity. Accordingly, we conduct two standard tests of spatial autocorrelation and serial correlation on the residuals in all six

that vary discontinuously at $\text{Crime}^x_{jt-1} = 0$.

specifications of control variables.

Following the suggestion of Dube et al. (2010), we re-estimate the system of equations and cluster the standard errors at a larger geographic level than our panel (by division-year-crime type as opposed to by sector-year-crime type). By doing so, we allow ϵ^y_{jt} to be correlated with ϵ^y_{kt}, where j and k are sectors within the same division of Dallas. Any systematic differences in the standard errors is evidence of spatial autocorrelation that is not controlled for. Similarly, we follow the suggestion of Angrist and Pischke (2009) and re-cluster our standard errors at the division-week-crime type level. A comparison of the standard errors clustered at this level with the standard errors clustered at the division-year-crime type level provides a test of whether the residuals are correlated across weeks within the year. Any systematic differences in the standard errors is evidence of serial correlation that is not controlled for.

The results of these exercises are presented in Appendix Table 13. To summarize, we find no evidence of either spatial autocorrelation or serial correlation in the surviving model (6). Moreover, we find much larger differences in standard errors across different clusters in non-surviving models, although these differences diminish as we add more detailed fixed effects. These two findings show that the fundamental endogeneity problem that has been identified in this literature does in fact operate in our setting, but it can be addressed successfully with appropriately specified fixed effects.

7.3 Multiple Testing

A key requirement of our identification strategy is the implementation of the exogeneity test for every candidate model, which may raise concerns related to multiple testing. The traditional problem encountered with multiple testing is *false discovery*, i.e., the rejection of the null hypothesis by pure chance when in reality the null is correct (type I error). The standard solution to this problem is to adjust the size of the hypothesis test downwards (Bender and Lange (2001)). In our approach, however, this is only a second order concern, as false discovery could only lead us to reject a model that did not suffer from endogeneity in the first place. Any surviving model in our context thus clears a more stringent threshold for exogeneity, suggesting that our approach is conservative in interpreting the main parameters as causal. In any case, false discovery does not seem to be a concern in our application. Whenever we reject the null hypothesis for a model, we continue to reject it in all further robustness checks that we perform on that model. For example, as we add various controls and lagged explanatory variables to models (1)-(5), we continue to reject the null hypothesis of exogeneity. Moreover, we often find that these models do not survive other less-powerful

tests that we conduct. This evidence suggests that we did not reject those models by pure chance.

In contrast, for our testing procedure what we should be more concerned about is the less discussed problem of *false non-discovery* (Sarkar (2006)), i.e., the failure to reject the null hypothesis by pure chance when in reality the null is incorrect (type II error). Indeed, if we test the exogeneity assumption in enough models, we are bound to fail to reject one of them by chance even if all models are truly endogenous.

In our application, concerns about false non-discovery seem unwarranted. Once we fail to reject the null hypothesis for a model, we continue to fail to reject it in all further robustness checks that we perform on that model. For example, as we add various control variables and lagged explanatory variables to model (6), we continue failing to reject the null hypothesis of exogeneity. This suggests that our failure to reject model (6) was not due to chance.

We perform one additional robustness check that relates directly to both multiple testing concerns of false discovery and false non-discovery. We randomly split our sample of sector-years into two subsamples, estimate models (1)-(6) in each of them, and then perform our test of exogeneity on each model.[38] In Table 10, we report the p-value of the test of exogeneity in each subsample and (in brackets) the p-value of whether at least one of the 36 coefficients of interest is different from its respective one in the full sample. If our testing procedure suffered from false non-discovery (false discovery), we might expect to find different surviving (non-surviving) models in each subsample; this is not the case, as only model (6) survives at the 90% level in either subsample.

Table 10: p-Values for Sensitivity Tests on Randomly Drawn Subsamples

	Specifications of Fixed Effects					
	(1)	(2)	(3)	(4)	(5)	(6)
On Subsample 1	0.00	0.00	0.00	0.00	0.02	**0.38**
	[1.00]	**[1.00]**	**[1.00]**	**[0.93]**	**[0.93]**	**[0.99]**
On Subsample 2	0.00	0.03	0.05	0.03	0.08	**0.12**
	[1.00]	**[1.00]**	**[1.00]**	**[0.99]**	**[1.00]**	**[1.00]**

Notes: For each randomly drawn subsample, we perform an exogeneity test and report the p-value of this test for various specifications of equation (10) as described in Table 2 for j = sector and t = week. In brackets, we report the p-value of a test of whether at least one of the 36 coefficients of interest is different from the respective one in the full sample. All errors are clustered at the $j \times$ year $\times$ c level.

[38] We intentionally split our sample randomly into sector-years as opposed to sector-weeks and do so only after we are confident that treatment effects last less than one year and do not spillover outside of a sector (otherwise the splitting of the sample itself might artificially generate or eliminate endogeneity from the model.) The p-values in brackets show that the randomized samples are sufficiently representative of the full sample.

7.4 Summary

We summarize the implications of our sensitivity analysis in Table 11, where we describe all of the properties that a variable w must concurrently possess in order to bias the main estimates of the surviving model in Section 6. This can be understood as a way of keeping track of the confounders that may bias our preferred estimates. Briefly, for w to bias our preferred estimates, it must (i) belong to $\mathbb{W}$, but not be detected by the exogeneity test (i.e., it needs to belong to $\mathbb{W}_2^D \cup \mathbb{W}^C$); (ii) not be absorbed by sector-year-crime type and division-week-crime type FEs; (iii) be uncorrelated to prior crimes that occurred up to 6 weeks in the past; (iv) be uncorrelated to 180 controls related to specific features of crimes that reflect their salience; (v) be uncorrelated to 72 controls related to the response of the police to crimes; (vi) be uncorrelated to 36 variables related to crime rates in adjacent neighborhoods; (vii) be spatially uncorrelated across neighborhoods within the six divisions of Dallas; and finally (viii) be serially uncorrelated across weeks within a calendar year. In the next Section, we provide empirical evidence that the restriction imposed by (i) should eliminate most, if not all, of the set of potential confounders. Any of the remaining undetectable confounders must possess properties (ii) through (viii) to bias our estimates.

Table 11: Necessary Properties of a Confounder in the Surviving Model

	Property	Justification
i	Undetectable by the exogeneity test	Models survived the exogeneity test even in its strongest form ($\bar{\tau} = 6$). (See Sections 6 and 7.)
ii	Not be absorbed by sector-year-crime type or division-week-crime type FEs	Model contains sector-year-crime type and division-week-crime type FEs. (See Section 5.3.)
iii	Uncorrelated to $\text{Crime}_{jt-\tau}$ for $\tau = 2, ..., 6$	Model still survives, yet estimates do not change when $\text{Crime}_{jt-\tau}$, $\tau = 2, ..., 6$ are added as controls. (See Section 6.2.)
iv	Uncorrelated to 180 salience control variables	Model still survives, yet estimates do not change when these 180 variables are added. (See Section 7.1.)
v	Uncorrelated to 72 police response control variables	Model still survives, yet estimates do not change when these 72 variables are added. (See Section 7.1.)
vi	Uncorrelated to 36 crime variables in nearby sectors	Model still survives, yet estimates do not change when these 36 variables are added. (See Section 7.1.)
vii	Not spatially autocorrelated within division	Model still survives, yet standard errors do not change when clustering at division-year-type of crime level (instead of sector-year-type of crime level). (See Section 7.2.)
viii	Serially uncorrelated within year	Model still survives, yet standard errors do not change when clustering at division-week-type of crime level (instead of division-year-type of crime level). (See Section 7.2.)

Note: The surviving model corresponds to specification of FEs (6) (i.e., division-week-type of crime and sector-year-type of crime FEs) for j = Sector, t = Week, and $\mathbb{C}$ = {Rape, Robbery, Burglary, MVT, Assault, Liught Crime}.

Remark 5. This discussion is useful for considering how one should select a model when there are multiple survivors of the test of exogeneity. If two models survived but yielded different estimates, then it must be the case that undetectable confounders were confounding at least one of the models. That would be evidence that the test was not powerful enough. Because our identification strategy is based upon the premise that the exogeneity test is sufficiently powerful, in this case we would have to conclude that it is infeasible to identify the parameters

of interest with this identification strategy and dataset alone.[39] In our application, the estimates of all treatment effects of interest are indistinguishable (both statistically and economically) among all surviving models. Intuitively, if the test of exogeneity was not powerful enough, then many confounders would clear restriction (i) in the summary table but might not clear at least one of the restrictions (ii)-(viii). We find that once a model survives the test of exogeneity, it survives all other tests, which suggests that the test is sufficiently powerful for our application.

8 Do "Broken Windows" Matter?

The "Broken Windows" theory leaves room for interpretation due to the informal way in which it was introduced. Because of its impact on policy, we attempt to provide a more structured discussion of its hypotheses and their relationship with our findings.

According to the theory, a law enforcement policy that reduces the perception of light crime in a neighborhood will reduce more severe crimes in the future (all else being equal).

$$\text{Policy} \longrightarrow \text{(Perceived/Actual) Light Crime}_{t-1} \Longrightarrow \text{Severe Crime}_t \tag{12}$$

This "Broken Windows" effect is shown as a double arrow in diagram (12). Such a policy can be focused on either reducing actual light crime (e.g., preventing broken windows) or by simply removing salient signs of light crimes (e.g., fixing broken windows). Although we do not directly observe such signs of light crime, we can bring our findings to bear on whether reductions in actual light crimes (and in particular, the most salient light crimes) have effects on future crimes. We gather these findings in Table 12, which summarizes the evidence on the "Broken Windows" effect for all surviving models in our analysis.[40]

In the first row, we report the p-values from tests of whether light crime has intertemporal spillovers *across* other types of crimes in the future (i.e., F-tests of $\beta_\tau^{xy} = 0$ for all $\tau = 1, \ldots, \bar{\tau}$, $x =$ Light Crime and $x \neq y$). We find no evidence of such "Broken Windows" effects at any point over a six week period.

In the second row, we report the p-values from tests of whether intertemporal spillovers from light crime vary by the salience of the crime. For each x, we interact Crime_{jt-1}^x with the police's speed of arrival and their duration of stay at the crime scene along with the number of crimes that occurred outdoors, at day time and during the weekend in all periods $t - \tau$,

[39] Of course, the value of the test as discussed by Caetano (2015) – to discard certain models – is still intact in this context. In order to reduce even further the set of candidate models, the test of exogeneity could in principle be implemented with additional robustness checks aimed at detecting these undetectable confounders, as is done in our sensitivity analysis.

[40] As shown in Table 9, all these models survive the test of exogeneity.

$\tau = 1, \ldots, \bar{\tau}$ to the right-hand-side of the equations of motion ($180 \times \bar{\tau}$ additional control variables in total).[41] These descriptions of crimes are plausibly associated with their salience from the perspective of police, neighbors or potential criminals. We then test for whether the coefficients on all variables that capture the across-crime intertemporal effects of light crimes (irrespective of their salience) are equal to zero ($30 \times \bar{\tau}$ coefficients in total). We find no evidence that any of these kinds of light crimes generate "Broken Windows" effects.

One might worry that light crimes become more noticeable only if many of them occur. In the third row, we show p-values from tests of non-linear intertemporal effects by estimating a linear b-spline with a knot at the median number of weekly crimes for each type of crime. This allows for intertemporal effects to differ depending on whether a neighborhood experienced more or less than the median number of crimes of a given type in a week. We find no evidence of "Broken Windows" effects in neighborhoods with high or low levels of light crime.

Table 12: Is There a "Broken Windows" Effect? P-values for Various Tests of its Existence

	Number of Included Lags ($\bar{\tau}$)					
	$\bar{\tau} = 1$	$\bar{\tau} = 2$	$\bar{\tau} = 3$	$\bar{\tau} = 4$	$\bar{\tau} = 5$	$\bar{\tau} = 6$
Baseline (Across-Crime Intertemporal Effects of Light Crime)	0.40	0.62	0.85	0.71	0.78	0.84
Allowing for Heterogeneous Effects by Salience	0.89	0.87	0.58	0.69	0.81	0.84
Allowing for Non-Linear Light Crime Effects	0.97	0.92	0.98	0.98	1.00	0.97
Controlling for a Contemporaneous Police Responses	1.00	1.00	1.00	1.00	1.00	1.00

Notes: This table presents p-values for a variety of tests of the "Broken Windows" effect for models with $1, \ldots, 6$ lagged crimes of each type on the right hand side. All models shown survive the exogeneity test. The first row contains p-values for F-tests of whether the $5 \times \bar{\tau}$ coefficients representing across-crime intertemporal effects of light crime are equal to zero. The second row contains p-values for F-tests of whether the $30 \times \bar{\tau}$ coefficients representing across-crime intertemporal effects of light crime, stratified by their salience, are all equal to zero. Salience in this context refers to features of crime which may be associated (positively or negatively) with its perception: whether crime occurs on the weekend, at day time, outdoors, and whether the police arrive quickly to the crime scene and stay longer at the crime scene. The third row contains p-values for F-tests of whether the $10 \times \bar{\tau}$ coefficients representing across-crime intertemporal effects of light crime, stratified by whether light crimes was higher or lower than the median level in a given week, are all equal to zero. The fourth row contains p-values for F-tests of whether across-crime intertemporal effects of light crime are still equal to zero when we include 36 variables referring to the average speed of police arrival to crime scenes in t for each crime type and 36 variables referring to the average durations of the police at crime scenes in t for each crime type.

One potential concern with these findings is that light crime may generate a secondary

[41] This model survives at standard levels of significance for all values of $\bar{\tau}$.

effect – an institutional response by law enforcement – that offsets what some might consider the true "Broken Windows" effect (see diagram (13)).[42] To address this concern, we report in the fourth row of Table 12 the p-values from tests of whether the observed speed of arrival to and duration of stay at the crime scene by law enforcement in period t mitigates our estimates of the intertemporal effects of light crime. These variables partially proxy for neighborhood- and week-specific changes in the presence of law enforcement (e.g., if the police is patrolling near a reported crime scene, it will respond more quickly). We do so by including these variables as controls for each type of crime (a total of 72 variables) and testing whether the estimates of across-crime intertemporal spillovers of light crime are equal to zero. We find no evidence that our estimates are contaminated by an institutional response.

$$\begin{tikzcd}
\text{Policy} \arrow[r] \arrow[rrr, bend left, dashed] & \text{(Perceived/Actual) Light Crime}_{t-1} \arrow[rr, Rightarrow] \arrow[dr, dotted] & & \text{Severe Crime}_t \\
& & \text{Inst. Response}_t \arrow[ur, dotted] &
\end{tikzcd} \tag{13}$$

We supplement the evidence presented in Table 12 with our earlier empirical findings. First, to the extent that our controls do not fully absorb all differences between actual light crime and perceived light crime, our test of exogeneity has the statistical power to detect whether any resulting measurement error biases our estimates (Section B.2.2), yet we find no evidence of such bias.[43] This further allays the concern that our estimates may not include the potentially salutary effects of, say, fixing broken windows. Second, we find no evidence that our results depend on which kinds of non-severe crimes are categorized as light crimes (Panel (f) of Figure 7 and Panel (e) of Figure 9). Moreover, given the many robustness checks that we performed (summarized in Table 11), it is difficult to conceive of a source of potential bias in our estimates that is consistent with all of evidence provided, let alone a source of *negative* bias that would lead us to underestimate the "Broken Windows" effect. Taken all together, our findings lead us to conclude that this hypothesized effect does not exist in practice, at least in Dallas during our sample period.

[42] To illustrate this point, consider two otherwise similar neighborhoods, one of which experienced one additional light crime in the past week. This additional crime may lead the police to monitor more intensely this neighborhood in the current week relative to the other neighborhood, which helps deter severe crime in the current week. Thus, past light crime may appear to have no effect on current severe crimes purely because of the increase in police monitoring in the current week. The policy maker may also have control over the police response to previous crimes, so it is also useful to estimate the "Broken Windows" effect net of institutional responses.

[43] Given that the Dallas Police Department did not have a policy that separately targeted the perception of light crime during our sample period, it is not surprising that this source of measurement error does not bias our results. Anyway, to the extent that such policy may exist, as long as it varies at a broader geographical level (division), or rather varies yearly at the neighborhood (sector) level, our fixed effects in the surviving model should have absorbed it.

Remark 6. In the absence of a "Broken Windows" effect, any observed causal link between a so-called "Broken Windows" law enforcement policy on severe crime in the future is a direct effect, as depicted in the dashed arrow in diagram (13). This distinction is important for policy because a direct effect is not subject to a dynamic multiplier, whereas an indirect "Broken Windows" effect is. To be sure, policies focused on combating light crime may directly affect the levels of severe crimes (e.g., in the process of cleaning up streets, a sanitation worker may deter a murder that was about to occur). However, it would be misleading to interpret falling rates of severe crimes that accompany such a policy as support for the "Broken Windows" theory. Indeed, when it comes to *directly* reducing severe crimes, it is probably better to target severe crimes.

9 Conclusion

Researchers typically approach causal inference problems with a source of variation that is already known to be "as good as random." But isolating such variation is difficult in practice, which limits the scope of questions that can be credibly answered this way. In this paper, we pose one such question of long-standing importance: "What is the local effect of crime on future crime?". To address the obstacles to causal inference that play a central role in this literature, we develop an identification strategy that does not require an *ex ante* known source of quasi-experimental variation and instead demonstrate how a careful consideration of alternative models, informed by the systematic use of a recently developed test of exogeneity, successfully leads us to an *ex post* known source of quasi-experimental variation. More generally, our identification strategy can be applied to isolate variation that is "as good as random" in any setting where this test of exogeneity can be used and where data is sufficiently rich.

We find evidence that robberies, burglaries and auto thefts cause modest increases in those crimes in the future. However, we find no evidence that light crime in a neighborhood will cause more severe crimes to proliferate. This stands in conflict with a simple, intuitive idea that has influenced law enforcement policy in a number of cities over the past three decades. Our analysis indicates that law enforcement policies based on the "Broken Windows" theory that feature the disproportionate targeting of lighter crimes, which include zero-tolerance and stop-and-frisk policies, are not empirically sound strategies to reduce more severe crimes in the future. Put simply, a policymaker aiming to reduce severe crimes ought to target severe crimes.

Methodologically, our work complements the existing empirical literature on model selection and inference. Sophisticated methods in statistical learning such as model averaging

and lasso have been successfully applied to uncover important insights in economics of crime (e.g., Durlauf et al. (2008), Cohen-Cole et al. (2009), Durlauf et al. (2010), Durlauf et al. (2014), Belloni et al. (2014)) and many other topics in the social sciences.[44] Our work illustrates how, when applicable, the exogeneity test developed by Caetano (2015) might be used as a new criterion for assessing the plausibility of models (in the case of model averaging) or for choosing tuning parameters (in the case of lasso), when the primary goal is causal identification as opposed to prediction.

References

Akerlof, G., 1997. Social distance and social decisions. Econometrica: Journal of the Econometric Society, 1005–1027.

Altonji, J., Elder, T., Taber, C., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. Journal of Political Economy 113 (1), 151–184.

Angrist, J., Pischke, J., 2009. Mostly harmless econometrics: An empiricist's companion. Princeton Univ Pr.

Anselin, L., Cohen, J., Cook, D., Gorr, W., Tita, G., 2000. Spatial analyses of crime. Criminal justice 4 (2), 213–262.

Becker, G., 1968. Crime and punishment: An economic approach. The Journal of Political Economy 76 (2), 169–217.

Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies 81 (2), 608–650.

Bender, R., Lange, S., 2001. Journal of clinical epidemiology 54 (4), 343–349.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of political Economy, 992–1026.

Block, R., 1993. A cross-national comparison of victims of crime: victim surveys of twelve countries. International Review of Victimology 2 (3), 183–207.

Blumstein, A., et al., 1986. Criminal Careers and" Career Criminals,". Vol. 2. National Academies.

[44]See Hastie et al. (2009) for a general survey and many examples of applications of these methods.

Braga, A., Bond, B., 2008. Policing crime and disorder hot spots: A randomized controlled trial*. Criminology 46 (3), 577–607.

Caetano, C., July 2015. A test of exogeneity without instrumental variables in models of bunching. Econometrica 83 (4), 1581–1600.

Card, D., Dobkin, C., Maestas, N., 2008. The impact of nearly universal insurance coverage on health care utilization: Evidence from medicare. American Economic Review 98, 2242–2258.

Cohen-Cole, E., Durlauf, S., Fagan, J., Nagin, D., 2009. Model uncertainty and the deterrent effect of capital punishment. American Law and Economics Review 11 (2), 335–369.

Cohn, E. G., 1990. Weather and crime. British Journal of Criminology 30 (1), 51–64.

Corman, H., Mocan, N., 2005. Carrots, sticks, and broken windows. Journal of Law and Economics, 48.

Damm, A. P., Dustmann, C., 2014. Does growing up in a high crime neighborhood affect youth criminal behavior? The American Economic Review 104 (6), 1806–1832.

Dube, A., Lester, T., Reich, M., 2010. Minimum wage effects across state borders: Estimates using contiguous counties. The review of economics and statistics 92 (4), 945–964.

Durlauf, S., Navarro, S., Rivers, D., 2008. On the interpretation of aggregate crime regressions. Crime Trends.

Durlauf, S. N., Navarro, S., Rivers, D. A., 2010. Understanding aggregate crime regressions. Journal of Econometrics 158 (2), 306–317.

Durlauf, S. N., Navarro, S., Rivers, D. A., 2014. Model uncertainty and the effect of shall-issue right-to-carry laws on crime. Tech. rep., University of Western Ontario, CIBC Centre for Human Capital and Productivity.

Ellison, G., Fudenberg, D., 1995. Word-of-mouth communication and social learning. The Quarterly Journal of Economics 110 (1), 93–125.

Flango, V. E., Sherbenou, E. L., 1976. Poverty, urbanization, and crime. Criminology 14 (3), 331–346.

Funk, P., Kugler, P., 2003. Dynamic interactions between crimes. Economics Letters 79 (3), 291–298.

Gelman, A., Fagan, J., Kiss, A., 2007. An analysis of the new york city police department's 'stop-and-frisk' policy in the context of claims of racial bias. Journal of the American Statistical Association 102 (479).

Gladwell, M., 2000. The tipping point: How little things can make a big difference. Little, Brown and Company.

Glaeser, E., Sacerdote, B., Scheinkman, J., 1996. Crime and social interactions. The Quarterly Journal of Economics, 507–548.

Harcourt, B., Ludwig, J., 2006. Broken windows: New evidence from new york city and a five-city social experiment. The University of Chicago Law Review, 271–320.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning. Vol. 2. Springer.

Hausman, J. A., 1978. Specification tests in econometrics. Econometrica: Journal of the Econometric Society, 1251–1271.

Heaton, P., 2010. Hidden in Plain Sight. RAND Corporation.

Jacob, B., Lefgren, L., Moretti, E., 2007. The dynamics of criminal behavior: Evidence from weather shocks. Journal of Human Resources 42 (3), 489–527.

Kadane, J. B., Lazar, N. A., 2004. Methods and criteria for model selection. Journal of the American statistical Association 99 (465), 279–290.

Kelling, G., Coles, C., 1998. Fixing broken windows: Restoring order and reducing crime in our communities. Free Press.

Kelling, G., Sousa, W., 2001. Do Police Matter?: An Analysis of the Impact of New York City's Police Reforms. CCI Center for Civic Innovation at the Manhattan Institute.

Kelling, G. L., Wilson, J. Q., March 1982. Broken windows. The Atlantic Monthly.

Kempf, K., 1987. Specialization and the criminal career. Criminology 25 (2), 399–420.

Lee, D., McCrary, J., 2005. Crime, punishment, and myopia. Tech. rep., National Bureau of Economic Research.

Levitt, S., 1998a. Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error? Economic Inquiry 36 (3), 353–372.

Levitt, S., 2004. Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. The Journal of Economic Perspectives 18 (1), 163–190.

Levitt, S. D., 1998b. The relationship between crime reporting and police: Implications for the use of uniform crime reports. Journal of Quantitative Criminology 14 (1), 61–81.

Miller, T., Cohen, M., Rossman, S., 1993. Victim costs of violent crime and resulting injuries. Health Affairs 12 (4), 186–197.

Mosher, C., Hart, T., Miethe, T., 2010. The mismeasure of crime. Sage Publications, Inc.

Sampson, R. J., 1985. Neighborhood and crime: The structural determinants of personal victimization. Journal of Research in Crime and Delinquency 22 (1), 7–40.

Sarkar, S. K., 2006. False discovery and false nondiscovery rates in single-step multiple testing procedures. The Annals of Statistics, 394–415.

Skogan, W., 1974. The validity of official crime statistics: An empirical investigation. Social Science Quarterly 55 (1), 25–38.

Skogan, W., 1975. Measurement problems in official and survey crime rates. Journal of Criminal Justice 3 (1), 17–31.

Taylor, R., 1996. Neighborhood responses to disorder and local attachments: The systemic model of attachment, social disorganization, and neighborhood use value. In: Sociological Forum. Vol. 11. Springer, pp. 41–74.

Varian, H. R., 2014. Big data: New tricks for econometrics. The Journal of Economic Perspectives, 3–27.

Weisburd, D., Eck, J., 2004. What can police do to reduce crime, disorder, and fear? The Annals of the American Academy of Political and Social Science 593 (1), 42–65.

Zimring, F., 1998. Youth violence epidemic: Myth or reality, the. Wake Forest L. Rev. 33, 727.

A Appendix: A Test of Exogeneity (For online publication only.)

Here we present a more formal and general derivation of the test of exogeneity than presented in the text. Consider the estimating equation

$$y_j = X_j'\beta + Z_j'\gamma + \epsilon_j \tag{14}$$

where X_j is an $n^X \times 1$ vector of explanatory variables of interest, Z_j is a vector of control variables, and ϵ_j is an error term that may be conditionally correlated to X_j. We can rewrite this equation as

$$y_j = X_j'\beta + Z_j'\gamma + \underbrace{\xi_j + \mu_j}_{\epsilon_j}. \tag{15}$$

where we split the error into two terms: ξ_j, which contains all unobserved determinants of y_j that are correlated to X_j conditional on Z_j irrespective of their source, and μ_j, which is the remainder. Taking conditional expectations of both sides of equation (15), we obtain

$$\mathrm{E}\left[y_j|X_j, Z_j\right] = X_j'\beta + Z_j'\gamma + \mathrm{E}\left[\xi_j|X_j, Z_j\right] \tag{16}$$

We can decompose the source of endogeneity in this equation as

$$\mathrm{E}\left[\xi_j|X_j, Z_j\right] = \left(D_j'\pi_D + X_j'\pi_X + Z_j'\pi_Z\right)\sigma_{\xi,X|Z,} \tag{17}$$

where $\sigma_{\xi,X|Z} \equiv \mathrm{Cov}\left(\xi_j, X_j|Z_j\right)$ is a n^X-vector, and D_j is an n^X-vector of indicator variables that are each equal to 1 if the corresponding element of X_j is equal to zero.[45]

Our goal is to design the following hypothesis test:[46]

$$\begin{aligned} H_0 &: \sigma_{\xi,X|Z} = \vec{0} \\ H_1 &: \sigma_{\xi,X|Z} \neq \vec{0} \end{aligned}$$

To implement this hypothesis test, we substitute equation (17) into equation (16), which we rewrite as

[45] More generally, equation (17) can be written as $\mathrm{E}\left[\xi_j|X_j, Z_j\right] = \left(D_j'\pi_D + f\left(X_j, Z_j\right)\right)\sigma_{\xi,X|Z}$ where f is continuous in X_j at all points where some element of X_j is equal to zero but otherwise unrestricted.

[46] Altonji et al. (2005) describe a different approach to measure the importance of ξ_j relative to total explanatory power of X_j and Z_j. In addition to different assumptions, the notable distinction between their approach and ours is that this test yields a statistical criterion that we can use for the comparison and selection of competing models.

$$E\left[y_j|X_j, Z_j\right] = X_j'\left(\beta + \pi_X \cdot \sigma_{\xi,X|Z}\right) + Z_j'\left(\gamma + \pi_Z \cdot \sigma_{\xi,X|Z}\right) + D_j'\underbrace{\left(\pi_D \cdot \sigma_{\xi,X|Z}\right)}_{\delta}. \tag{18}$$

According to equation (18), β is identified by OLS under H_0, but not under H_1. In general, we cannot identify $\sigma_{\xi,X|Z}$ in order to test H_0. However, we can identify the parameter vector $\delta \equiv \pi_D \cdot \sigma_{\xi,X|Z}$ by simply including D_j in the system of equations and modifying equation (14) to

$$y_j = X_j'\beta + Z_j'\gamma + D_j'\delta + \epsilon_j. \tag{19}$$

The test of exogeneity is easily implemented as a joint F-test of whether all elements of δ are equal to zero. In order to connect our estimate of $\hat{\delta}$ to our hypothesis test, we make one additional assumption:

Assumption 2. *If $\sigma_{\xi,X|Z} \neq \vec{0}$, then $\pi_D.\sigma_{\xi,X|Z} \neq \vec{0}$ among j such that $D_j \neq \vec{0}$.*

This assumption provides power to the test. Under Assumption 2, if $\sigma_{\xi,X|Z} \neq \vec{0}$, then our estimate of δ should contain at least one non-zero element and we should reject H_0. In contrast, if we find that $\delta = \vec{0}$, then we can conclude that $\sigma_{\xi,X|Z} = \vec{0}$. Thus, if in practice we fail to reject that $\delta = \vec{0}$ there are two possible cases: either $\sigma_{\xi,X|Z} = \vec{0}$ and there is no endogeneity in the model or Assumption 2 is invalid. We can naturally extend this test to the estimation of a system of n^Y equations. We rewrite equation (14) as

$$Y_j = X_j'\beta + Z_j'\Gamma + \epsilon_j \tag{20}$$

where Y_j is an $n^Y \times 1$ vector of dependent variables and β is an $n^X \times n^Y$ parameter matrix. Z_j is a vector of control variables with an associated coefficient matrix Γ, and ϵ_j is now an $n^Y \times 1$ vector of errors. To implement the test, we modify this system of equations to

$$Y_j = X_j'\beta + Z_j'\Gamma + D_j'\Delta + \epsilon_j. \tag{21}$$

where D_j is defined above and Δ is an $n^X \times n^Y$ parameter matrix that is an analog to δ. Similarly, Π_D and $\Sigma_{\xi,X|Z}$ can be defined as $n^X \times n^Y$ and $n^Y \times n^X$ matrix analogs to π_D and $\sigma_{\xi,X|Z}$, respectively. Under the modified assumption below, the test is implemented as a joint F-test of whether all elements of Δ equal zero.

Assumption 2′. *If $\Sigma_{\xi,X|Z} \neq \mathbf{0}$, then $\Pi_D.\Sigma_{\xi,X|Z} \neq \mathbf{0}$ among j such that $D_j \neq \vec{0}$.*

Remark 7. It is useful to relate this discussion to the sets $\mathbb{W}$, $\mathbb{W}_1^D$, $\mathbb{W}_2^D$ and $\mathbb{W}^C$ as defined in Section 4.1. For simplicity in the exposition, consider the case where X and Y have each one dimension. As seen in equation (18), the OLS bias is represented by Bias $= \pi_X \cdot \sigma_{\xi,X|Z}$. OLS estimates will be biased only if $\pi_X \neq 0$ and $\sigma_{\xi,X|Z} \neq 0$, i.e., if $\xi \in \mathbb{W}$. There are three potential cases for $\xi \in \mathbb{W}$: (a) $\pi_D.\sigma_{\xi,X|Z} \neq 0$ among observations such that $X = 0$ (i.e., $\xi \in \mathbb{W}_1^D$); (b) $\pi_D.\sigma_{\xi,X|Z} \neq 0$ only among observations such that $X \neq 0$ (i.e., $\xi \in \mathbb{W}_2^D$); and (c) $\pi_D.\sigma_{\xi,X|Z} = 0$ (i.e., $\xi \in \mathbb{W}^C$). In cases (b) or (c), the test of exogeneity has no power; we would estimate $\hat{\delta} = 0$ even if $\sigma_{\xi,X|Z} \neq 0$. In contrast, in case (a) the test does have power. Assumption **2′** rules out cases (b) and (c), so it guarantees that either ξ is a detectable confounder or ξ is fully absorbed by controls (so that $\sigma_{\xi,X|Z} = 0$).[47] These two conditions refer to restrictions (i) and (ii) of the Summary Table 11, respectively. The fact that a model that survived the test of exogeneity also survived all other robustness checks (restrictions (iii)-(viii)) strongly suggests that Assumption **2′** holds in our application.

Remark 8. As discussed in Appendix B, we find abundant evidence in favor of Assumption **2′**. However, even if the test has power, one should be careful when adding controls to regression models. For intuition, consider the case where X and Y are each one dimensional. Then we can rewrite the OLS bias as Bias $= \delta \cdot \frac{\pi_X}{\pi_D}$. Under the null hypothesis ($\delta = 0$), there is no OLS bias. Moreover, for a given $\frac{\pi_X}{\pi_D}$, as the true value of δ converges to 0 the bias should converge to 0 as well. The scaling factor $\frac{\pi_X}{\pi_D}$ represents the extent to which endogenous confounders are correlated to X when $X > 0$ (conditional on Z), relative to the extent to which this endogeneity can be observed as a discontinuity at $X = 0$. If this ratio increased as we added controls to our specification, that would be a source of concern. Intuitively, this means that controls should not be used to fit the discontinuity at $X = 0$; rather, they should be used to absorb endogeneity *everywhere* in the support of X; when all endogeneity is absorbed, a natural implication is the lack of discontinuity at $X = 0$. In our application, we find ample evidence that this is not a concern. First, the controls that we specify consist only of fixed effects (which are theoretically motivated), none of which mechanically absorb discontinuities at $\text{Crime}_{jt-1}^x = 0$. Second, once a model survives the test of exogeneity, it always survives further independent tests that attempt to detect additional confounders. If controls were added with the aim of only fitting the discontinuity at $\text{Crime}_{jt-1}^x = 0$, rather than solving the endogeneity problem everywhere in the support of $\text{Crime}_{jt-1}^x = 0$, then surviving models would perform poorly in further robustness checks designed to find non-local sources of endogeneity. This idea is analogous to the idea that over-fitting a model in-sample can lead to poorer performance out-of-sample (here we interpret $\text{Crime}_{jt-1}^x = 0$ as "in-sample" and $\text{Crime}_{jt-1}^x > 0$ as "out-of-sample").

[47] Here, we could interpret ξ as a weighted average of all confounders $w \in \mathbb{W}$.

B Statistical Power: Empirical Evidence (For online publication only.)

We leverage the rich data environment of our application to present evidence that our test has statistical power in two complementary, systematic exercises.

B.1 Detectable vs Undetectable Endogeneity: Classifying Observed Confounders

We indirectly measure the statistical power of our test with an intuitive diagnostic procedure. First, we augment our dataset with data from the 2010 US Census to construct a set of 691 distinct observed variables. For each of these candidate confounders, we test whether it belongs to the initial pool of confounders by testing whether our estimates of $\hat{\beta}$ change when we include the variable versus when we exclude the variable in model (1) (i.e., no Controls_{jt} included). The subset of actual confounders, which we denote as $\hat{\mathbb{W}}$, can be understood as an observed analog to $\mathbb{W}$, since $\hat{\mathbb{W}} \subset \mathbb{W}$. Each variable in this subset can then be assigned to the observed analog $\hat{\mathbb{W}}_1^D$, $\hat{\mathbb{W}}_2^D$ or $\hat{\mathbb{W}}^C$ on the basis of its (dis)continuity at $\text{Crime}_{t-1}^x = 0$ for some x and whether it is a confounder when $\text{Crime}_{t-1}^x = 0$ for some x. Of course, $\hat{\mathbb{W}}_1^D \subset \mathbb{W}_1^D$, $\hat{\mathbb{W}}_2^D \subset \mathbb{W}_2^D$ and $\hat{\mathbb{W}}^C \subset \mathbb{W}^C$.[48]

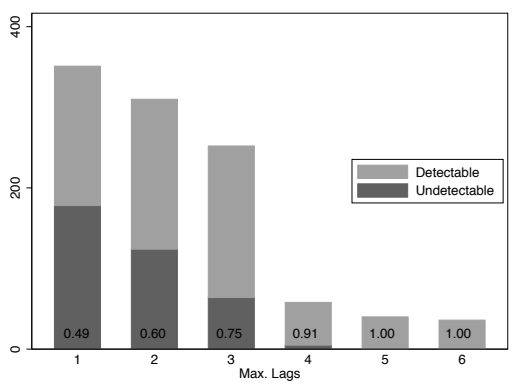
This informal evidence is similar to the evidence often shown in a regression discontinuity design (RDD). For comparison, in an RDD, researchers argue that $\mathbb{W} = \mathbb{W}^C$ by first considering a large set of observed variables that are correlated to the running variable and are potential determinants of the outcome variable (i.e., variables that are plausibly in $\hat{\mathbb{W}}$) and then showing empirically that all such variables are continuous at the relevant threshold (i.e., $\hat{\mathbb{W}} = \hat{\mathbb{W}}^C$). Intuitively, if the initial set of variables is sufficiently large and representative of the true set $\mathbb{W}$, then the RDD is plausibly validated (i.e., $\mathbb{W} = \mathbb{W}^{\mathbb{C}}$). Analogously, in our case we argue indirectly that $\mathbb{W} = \mathbb{W}_1^D$ by showing that $\hat{\mathbb{W}} = \hat{\mathbb{W}}_1^D$. If $\mathbb{W} = \mathbb{W}_1^D$, then all confounders will be detectable by our test of exogeneity, hence any surviving model can be interpreted as causal.

In Figure 6, we present the results of this diagnostic procedure for models with an increasing number of lagged explanatory variables (i.e., $\bar{\tau} = 1, \ldots, 6$) in graphical form. The height of each bar corresponds to the number of elements in $\hat{\mathbb{W}}$ for a specification with up to $\bar{\tau}$ lags, and each bar is subdivided into a darkly shaded component corresponding to the num-

[48] In Appendix C, we list the 691 candidate confounders that are used in this exercise and describe in detail the tests that enable us to classify them into $\hat{\mathbb{W}}$, $\hat{\mathbb{W}}_1^D$, $\hat{\mathbb{W}}_2^D$ and $\hat{\mathbb{W}}^C$.

ber of undetectable elements of $\hat{\mathbb{W}}$ (i.e., confounders in $\hat{\mathbb{W}}_2^D$ or $\hat{\mathbb{W}}^C$)[49] and a lightly shaded component corresponding to the number of detectable elements of $\hat{\mathbb{W}}$ (i.e., confounders in $\hat{\mathbb{W}}_1^D$). For $\bar{\tau} = 1$, over half of the 691 variables are empirically found to be confounders, which suggests that our pool of candidate confounders contains plausible determinants of neighborhood crime *ex ante*. Moreover, roughly half of the elements of $\hat{\mathbb{W}}$ are detectable. As $\bar{\tau}$ increases, $\hat{\mathbb{W}}$ diminishes, because the additional variables on the right hand side (i.e., $\text{Crime}_{jt-\tau}$ for $\tau = 2, \ldots, \bar{\tau}$) absorb an increasing number of candidate confounders. Importantly, this reduction in $\hat{\mathbb{W}}$ is disproportionately driven by a reduction in undetectable confounders; for $\bar{\tau} > 4$, all remaining elements of $\hat{\mathbb{W}}$ are detectable ($\hat{\mathbb{W}} = \hat{\mathbb{W}}_1^D$). Intuitively, this result follows from the fact that as $\bar{\tau}$ increases, we are able to perform a more powerful test (recall that we test for whether $\Delta_\tau = 0$ jointly for all $\tau \leq \bar{\tau}$.) This exercise highlights the increase in the power of the test from being performed in a multivariate context. Some confounders w that are undetectable by the test for low values of $\bar{\tau}$ end up being detected for higher values of $\bar{\tau}$. Of course, all these confounders are absorbed by the fixed effects in specification (6), which is why that model for $\bar{\tau} = 1$ survives the test. Overall, this exercise provides evidence that the property in row (i) of Table 11 substantially restricts the subset of confounders, at least in terms of observables.

Figure 6: Classifying Observed Confounders



Notes: This figure shows the total number of elements in the observed analog to $\mathbb{W}$ in a specification of equation (11) with no Controls_{jt} (i.e., specification (1) as described in Table 5 for j = sector and t = week) for various choices of $\bar{\tau} = 1, \ldots, 6$. The number in each bar represents the proportion of observed confounders that are detectable by the test of exogeneity. A detailed list of potential observed confounders is presented in Appendix C.

[49] In practice, we do not find elements of $\hat{\mathbb{W}}_2^D$ for any value of $\bar{\tau}$. This may be due to the fact that for a significant fraction of observations (69%) at least one of the six crimes is equal to zero.

B.2 Sources of Detectable Endogeneity

Even though we do not find a single undetectable confounder from our pool of observed variables, this does not entirely rule out the possibility of an undetectable confounder since our pool of observed variables may not be representative of the full set of omitted variables, both observable and unobservable. Here, we show that our pool of observed variables is representative in an important sense: it contains concrete examples of detectable confounders that correspond to a comprehensive set of endogeneity concerns in our application.

Consider the general version of the true equation of motion for crime y that we seek to estimate:

$$\text{Crime}_{jt}^{y} = f^{y}(\widetilde{\text{Crime}}_{jt-1}, \text{Other}_{jt}) \tag{22}$$

where $\widetilde{\text{Crime}}_{jt}$ is a very large row vector of *actual* (as opposed to reported) crimes in the set $\tilde{\mathbb{C}}$ that may be defined with great detail (e.g., "robbery at gunpoint outside of the main library at 5pm on a Monday, purse and \$200 taken"). Other_{jt} refers to any other determinant of Crime_{jt}^{y}, and $f^{y}(\cdot)$ is a flexible function. This is a "general" equation in the following sense: if we observed all elements of $\widetilde{\text{Crime}}_{jt-1}$ and Other_{jt}, and if we were able to estimate a non-restrictive $f^{y}(\cdot)$ for each y, then we would be able to identify the causal partial effect of $\widetilde{\text{Crime}}_{jt-1}$ on Crime_{jt}^{y} for each y. This is a good benchmark, as it allows us to discuss all of the sources of endogeneity that might show up in our application as we deviate from this ideal scenario.

In practice, we estimate models of the form:

$$\text{Crime}_{jt}^{y} = \text{Crime}_{jt-1}^{x}\beta^{xy} + \text{Crime}_{jt-1}^{-x}\beta^{-xy} + \text{Controls}_{jt}\gamma^{y} + \text{Error}_{jt}^{y} \tag{23}$$

where β^{-xy} is a column vector whose x'^{th} element is equal to $\beta^{x'y}$, for all $x' \neq x$.

The exogeneity assumption (Assumption 1) combines all of the simplifying assumptions that are required to transform the general model in equation (22) into our estimating models in equation (23) (e.g., linearity, whether Controls_{jt} are capable of proxying for Other_{jt}, absence of measurement error, etc.) A failure of any of these simplifying assumptions to hold implies the existence of a variable w. If w belongs to $\mathbb{W}$, then it will bias our estimates unless it is absorbed by Controls_{jt}. If w belongs to $\mathbb{W}_1^D \subset \mathbb{W}$, then the exogeneity test will detect its presence.

Below, we discuss each of these simplifying assumptions in turn and show examples of corresponding observed $w \in \mathbb{W}_1^D$ that are likely to be in Error_{jt}^{y} when that simplifying assumption does not hold. We present these examples in the form of discontinuity plots in Figures 7-12 (these are analogous to the *continuity* plots that validate a regression disconti-

nuity design).[50] Each point in these plots represents the mean of the variable on the vertical axis (some potential confounder) conditional on a given level of Crime^x_{jt-1} for some x. (For $\text{Crime}^x_{jt-1} = 0$, we represent the mean value of the variable on the vertical axis as a hollow dot.) The dashed curve represents a third order local polynomial regression for observations such that $\text{Crime}^x_{jt-1} > 0$, and the shaded region represents the 95% confidence region for this regression, with an out of sample prediction at $\text{Crime}^x_{jt-1} = 0$.[51]

B.2.1 Regular Omitted Variables (i.e., Other_{jt})

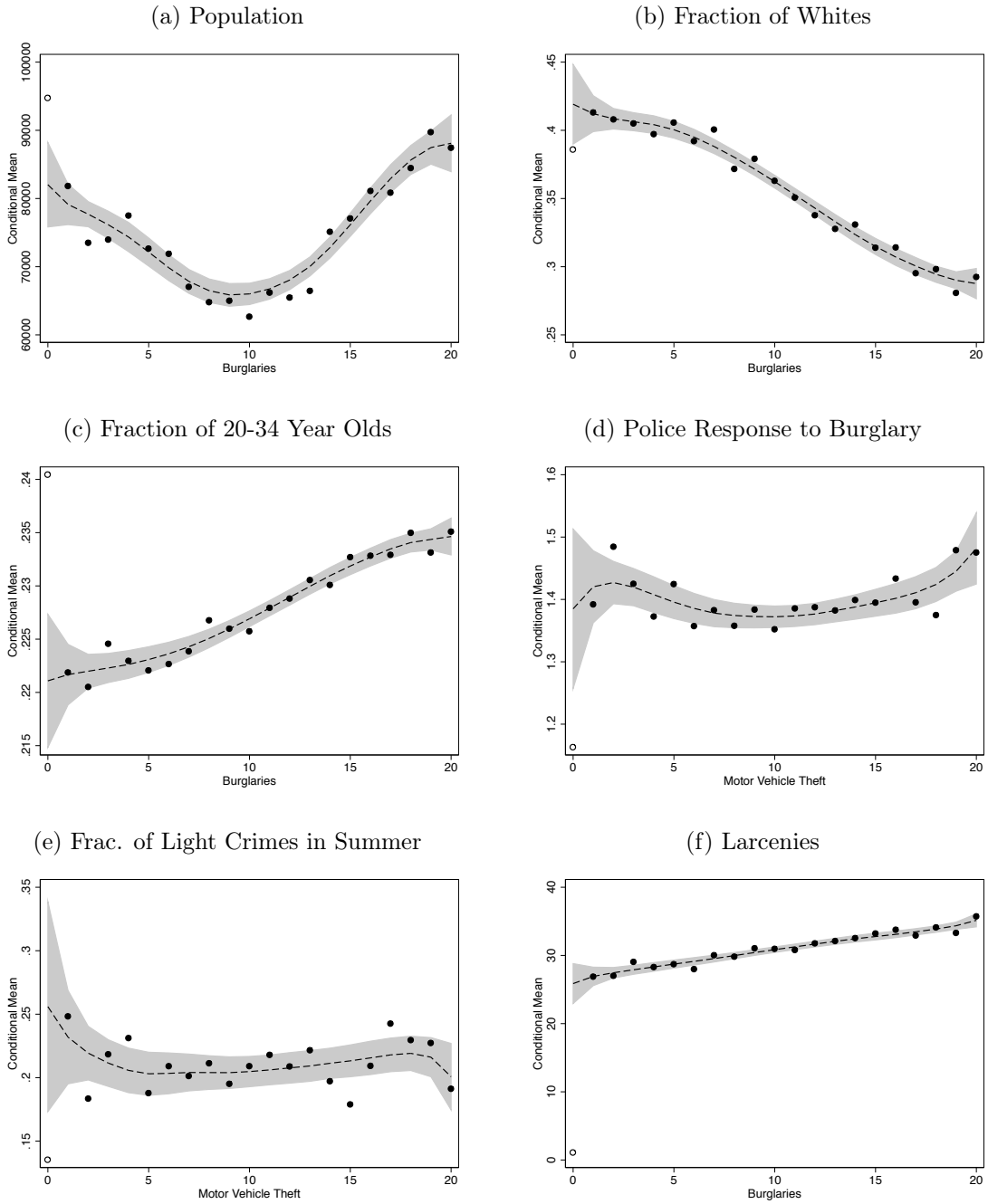
Other_{jt} in equation (22) might not be fully absorbed by the covariates in equation (23), which may lead to endogeneity. Figure 7 shows a few examples of key variables w_{jt} that are potential elements of Other_{jt} such as characteristics of neighborhoods, the timing of police responses to crimes, and types of crimes not included in $\mathbb{C}$. Panels (a)-(c) show that sectors with zero burglaries tend to have discontinuously more residents, fewer Whites, and more young adults (ages 20-34). These discontinuities would only arise if the spatial distribution of crimes was discontinuous at $\text{Crime}^x_{jt-1} = 0$ across all sector-week observations, which suggests that we can detect endogeneity from slowly varying omitted sector characteristics.[52] Panel (d) shows that the police arrive discontinuously faster to the scene of burglaries in sector-weeks with no auto thefts (such variable may proxy for the presence of police patrolling nearby in that particular week). This suggests that we can detect endogeneity from omitted sector characteristics that vary more rapidly from week to week. Panel (e) shows that light crimes are discontinuously less likely to occur in the summer in sector-weeks with no auto thefts, which suggests that we can detect endogeneity from seasonality. Panel (f) shows that the expected number of larcenies is discontinuously lower in sector-weeks with no burglaries. Because larceny is not included in $\mathbb{C}$, this suggests that we can detect endogeneity from not defining $\mathbb{C}$ exhaustively.

[50] To be sure, our running variables, Crime^x_{jt-1}, are discrete. Given the fact that they take on a wide variety of values, we treat them as continuous in order to test for discontinuities. This approach is commonly taken in regression discontinuity design studies (e.g., Lee and McCrary (2005), Card et al. (2008)). See Caetano (2015) for further discussion.

[51] The local polynomial regression and its pointwise confidence interval are estimated using the disaggregated dataset. For each regression, we use the Epanechnikov kernel with bandwidths of five for the kernel and the standard error calculation. Results are robust to different choices of bandwidths.

[52] These plots are constructed with data from the 2010 Census at the block level as follows. First, we calculate the average demographics across all census blocks within each sector. Then $\text{E}\left[w_{jt}|\text{Crime}^x_{jt-1}\right]$ is estimated as the weighted average of this measure across all weeks for sector j with weights corresponding to the frequency that Crime^x_{jt-1} takes on each value. This assures that a discontinuity can be found only if there is a discontinuity at $\text{Crime}^x_{jt-1} = 0$ in the spatial distribution of crimes across sector-weeks. For instance, consider two sectors, j and j'. If $\text{Crime}^x_{jt-1} = 0$ for 300 weeks, and $\text{Crime}^x_{j't-1} = 0$ for 150 weeks, but $\text{Crime}^x_{jt-1} = 1$ and $\text{Crime}^x_{j't-1} = 1$ for 50 weeks, then the demographics of sector j will be doubly weighted relative to the demographics of sector j' when that crime in both neighborhoods is equal to zero, but it will have the same weight when it is equal to one.

Figure 7: Discontinuity Plots: Omitted Variables



Notes: The scatter plot in each panel represents $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}\right]$ for each value of $\mathrm{Crime}^{x}_{jt-1}$, where w_{jt} is described in the title, and $\mathrm{Crime}^{x}_{jt-1}$ is described on the horizontal axis. The hollow dot represents $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}=0\right]$. A cubic local polynomial fit with its 95% confidence interval is also presented along with an outside-of-the-sample prediction at $x=0$. Panels (a)-(c) use block level data from the 2010 Census.

B.2.2 Measurement Error ($\widetilde{\textbf{Crime}}_{jt-1} \neq \textbf{Crime}_{jt-1}$)

$\widetilde{\text{Crime}}_{jt-1}$ in equation (22) may differ from Crime_{jt-1} in equation (23) for a variety of reasons, and this difference may not be fully absorbed by Controls_{jt} leading to endogeneity. Here, we discuss the two main sources of measurement error.

Misreporting ($\widetilde{\textbf{Crime}}^{x}_{jt-1} \neq \textbf{Crime}^{x}_{jt-1}$)

A large literature has found non-classical measurement error in reported crime levels (e.g., Levitt (1998b)). To consider this, assume the true model is

$$\text{Crime}^{y}_{jt} = \widetilde{\text{Crime}}^{x}_{jt-1}\beta^{xy} + \text{Crime}^{-x}_{jt-1}\beta^{-xy} + \text{Controls}_{jt}\gamma^{y} + \widetilde{\text{Error}}^{y}_{jt} \tag{24}$$

where Crime^{x}_{jt} is mismeasured, i.e., $\widetilde{Crime}^{x}_{jt-1} = \text{Crime}^{x}_{jt-1} + \eta_{jt}$. Re-writing the equation above yields

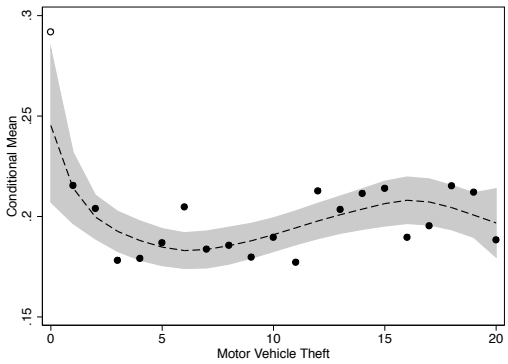
$$\text{Crime}^{y}_{jt} = \text{Crime}^{x}_{jt-1}\beta^{xy} + \text{Crime}^{-x}_{jt-1}\beta^{-xy} + \text{Controls}_{jt}\gamma^{y} + \underbrace{\widetilde{\text{Error}}^{y}_{jt} + \eta_{jt}\beta^{xy}}_{\text{Error}^{y}_{jt}} \tag{25}$$

where Error^{y}_{jt} is defined as in equation (23).

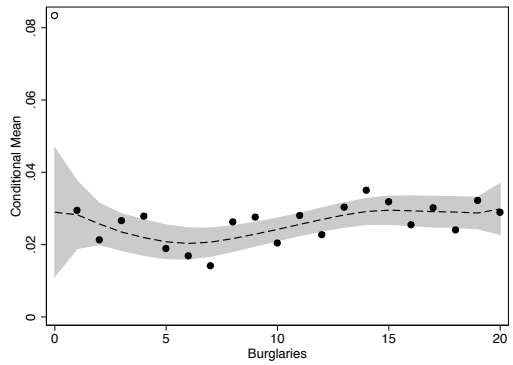
Determinants of η_{jt} that belong to $\mathbb{W}$ generate endogeneity stemming from measurement error unless they are absorbed by the covariates. Misreporting should depend on who the reporting party is. For instance, commercial businesses are obligated to report certain crimes for insurance purposes, whereas individuals are not. In Figure 8, we show plots of $\text{E}\left[w_{jt}|\text{Crime}^{x}_{jt}\right]$ that are discontinuous at zero as examples of $w \in \mathbb{W}^{D}_{1}$ that reflect who filed the report. These are likely determinants of η_{jt}. Panel (a) shows that robberies are discontinuously more likely to be reported by businesses in sector-weeks with no auto thefts. Similarly, robberies are discontinuously more likely to be anonymously reported in sector-weeks with no burglaries. Note that the test is agnostic to the particular model of measurement error.

Figure 8: Discontinuity Plots: Measurement Error

(a) Frac. Commercial Robbery



(b) Frac. Anonymous Robbery



Notes: The scatter plot in each panel represents $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}\right]$ for each value of $\mathrm{Crime}^{x}_{jt-1}$, where w_{jt} is described in the title, and $\mathrm{Crime}^{x}_{jt-1}$ is described on the horizontal axis. The hollow dot represents $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}=0\right]$. A cubic local polynomial fit with its 95% confidence interval is also presented along with an outside-of-the-sample prediction at $x=0$.

Over-aggregation ($|\mathbb{C}| < |\tilde{\mathbb{C}}|$)

Over-aggregation of crimes may also lead to endogeneity since $\widetilde{\mathrm{Crime}}_{jt-1}$ may contain many more elements than Crime_{jt-1}. Suppose that the true model is more disaggregated than (23), in the sense that $\mathrm{Crime}^{x}_{jt-1} = \mathrm{Crime}^{Ax}_{jt-1} + \mathrm{Crime}^{Bx}_{jt-1}$. Then the true model is given by

$$\mathrm{Crime}^{y}_{jt} = \mathrm{Crime}^{Ax}_{jt-1}\beta^{Axy} + \mathrm{Crime}^{Bx}_{jt-1}\beta^{Bxy} + \mathrm{Crime}^{-x}_{jt-1}\beta^{-xy} + \mathrm{Controls}_{jt}\gamma^{y} + \widetilde{\mathrm{Error}}^{y}_{jt} \quad (26)$$

Denoting $w_{jt} = \frac{\mathrm{Crime}^{Ax}_{jt-1}}{\mathrm{Crime}^{x}_{jt-1}}$, we can rewrite equation (26) as

$$\begin{aligned}\mathrm{Crime}^{y}_{jt} &= \mathrm{Crime}^{x}_{jt-1}\beta^{xy} + \mathrm{Crime}^{-x}_{jt-1}\beta^{-xy} + \mathrm{Controls}_{jt}\gamma^{y} \\ &\quad + \underbrace{\left[w_{jt}\left(\beta^{Axy}-\beta^{xy}\right) + (1-w_{jt})\left(\beta^{Bxy}-\beta^{xy}\right)\right]\mathrm{Crime}^{x}_{jt-1} + \widetilde{\mathrm{Error}}^{y}_{jt}}_{\mathrm{Error}^{y}_{jt}}\end{aligned} \quad (27)$$

where Error^{y}_{jt} is defined as in equation (23). In general, $\beta^{Axy} \neq \beta^{Bxy}$ since many features of crimes trigger different responses by police, neighbors and potential criminals, which in turn may lead to different dynamic spillovers. For instance, weekend crime may be more salient to some of these agents (relative to weekday crime). If this is the case, then aggregating the data weekly may lead to endogeneity if certain crimes occur more frequently on weekends. In Panels (a)-(e) of Figure 9, we present several plots where $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}\right]$ is discontinuous

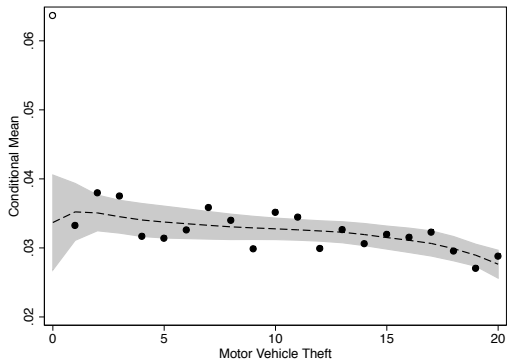
at $\text{Crime}^x_{jt-1} = 0$ for different definitions of A and B (crimes in the center versus at the edges of a neighborhood, day time versus night time crimes, indoor versus outdoor crimes, crimes committed on weekdays versus weekends, light crimes of a specific sub-type versus of all other sub-types). These w_{jt} all belong to $\mathbb{W}_1^D$.

Panels (a)-(b) show that burglaries (light crimes) are discontinuously more likely to occur at the center of a sector (at day time) in sector-weeks with no auto thefts. Panels (c)-(d) show that assaults are discontinuously more likely to occur indoor and during weekends in sector-weeks with no burglaries. Similarly, Panel (e) shows that the fraction of light crimes that are sub-classified as drunk and disorderly behavior is discontinuously lower in sector-weeks with no burglaries. Panel (f) provides an example of over-aggregation that is not binary: the police remain at the scene of a burglary discontinuously longer in sector-weeks when no assaults occur. Because the duration of the police at crime scenes might affect the perception of how seriously the police respond to crime, β might differ depending on this variable.

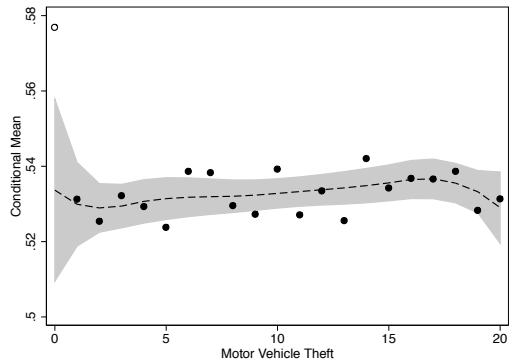
These plots show that our test has power to detect endogeneity from spatial over-aggregation (j), temporal over-aggregation (t), and over-aggregation of crime types ($\mathbb{C}$).

Figure 9: Discontinuity Plots: Over-Aggregation

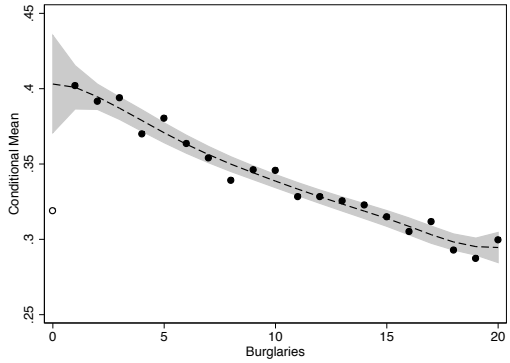
(a) Frac. Burglaries in the Center of the Sector



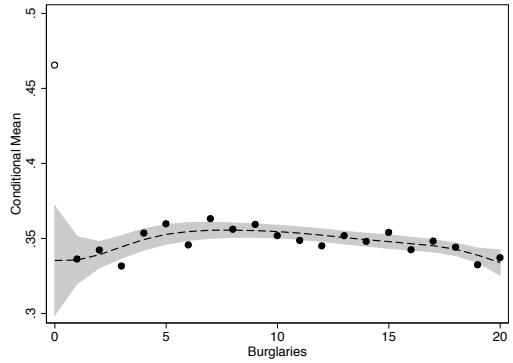
(b) Frac. Day Time Light Crimes



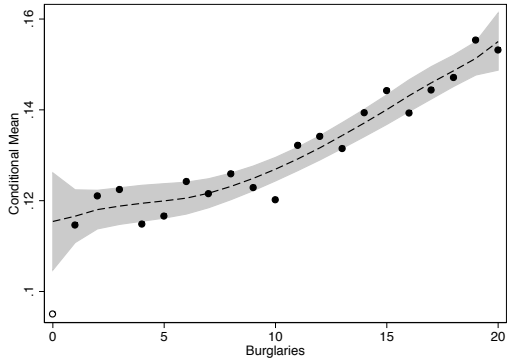
(c) Frac. Outdoor Assaults



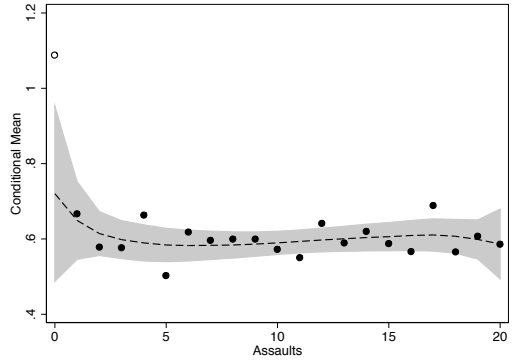
(d) Frac. Weekend Assaults



(e) Frac. of Light Crimes that are Drunk and Disorderly Offenses



(f) Police Duration at Burglaries



Notes: The scatter plot in each panel represents $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}\right]$ for each value of $\mathrm{Crime}^{x}_{jt-1}$, where w_{jt} is described in the title, and $\mathrm{Crime}^{x}_{jt-1}$ is described on the horizontal axis. The hollow dot represents $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}=0\right]$. A cubic local polynomial fit with its 95% confidence interval is also presented along with an outside-of-the-sample prediction at $x=0$.

B.2.3 Contamination (i.e., j is Too Fine)

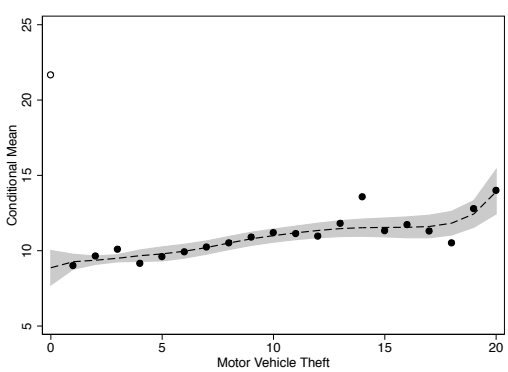
If neighborhoods are too narrowly defined, then crime in one neighborhood may have spillover effects on adjacent neighborhoods (Anselin et al. (2000)). Because nearby neighborhoods are effectively used as control groups for causal inference, we can understand this as a contamination problem. Assume that in the true model, current crime of type y is affected by past crime of type x from nearby neighborhoods, which we denote as w_{jt}. Then we can formalize the problem as follows:

$$\text{Crime}^{y}_{jt} = \text{Crime}^{x}_{jt-1}\beta^{xy} + \text{Crime}^{-x}_{jt-1}\beta^{-xy} + \text{Controls}_{jt}\gamma^{y} + \underbrace{w_{jt}\beta^{wy} + \widetilde{\text{Error}}^{y}_{jt}}_{\text{Error}^{y}_{jt}} \tag{28}$$

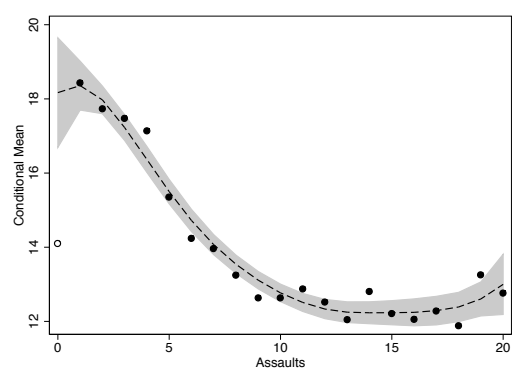
Figure 10 shows examples of w_{jt} that belong to $\mathbb{W}^{D}_{1}$. Sector-weeks with zero auto thefts (assaults) are near sectors that had a discontinuously higher number of assaults (burglaries) in that week. This is true whether we define nearby sectors as those within one mile or five miles of sector j. These plots provide direct evidence that we can detect whether j is too disaggregated.

Figure 10: Discontinuity Plots: Under-Aggregation (Contamination)

(a) Assaults in Sectors within 1 Mile



(b) Burglaries in Sectors within 5 Miles



Notes: The scatter plot in each panel represents $\mathrm{E}\left[w_{jt}|\text{Crime}^{x}_{jt-1}\right]$ for each value of Crime^{x}_{jt-1}, where w_{jt} is described in the title, and Crime^{x}_{jt-1} is described on the horizontal axis. The hollow dot represents $\mathrm{E}\left[w_{jt}|\text{Crime}^{x}_{jt-1} = 0\right]$. A cubic local polynomial fit with its 95% confidence interval is also presented along with an outside-of-the-sample prediction at $x = 0$.

B.2.4 Non-linearities (i.e., $f^{y}\left(\mathbf{Crime}_{jt-1}, \mathbf{Other}_{jt}\right) \neq \mathbf{Crime}^{x}_{jt-1}\beta^{xy} + \mathbf{Crime}^{-x}_{jt-1}\beta^{-xy} + \mathbf{Controls}_{jt}\gamma^{y}$)

The true equations of motion of crime may be non-linear, which may lead to endogeneity in our estimating equations. Here, we consider two types of non-linearities.

Non-separability

If past crime of type x affects current crime of type y differently for different values of Crime_{jt}^{-x} or Other_{jt}, then our assumption of additive separability may lead to endogeneity. Suppose the true model is given by

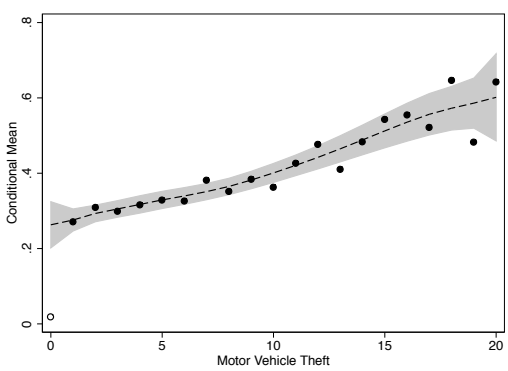
$$\text{Crime}_{jt}^{y} = \text{Crime}_{jt-1}^{x}\beta_{jt}^{xy} + \text{Crime}_{jt-1}^{-x}\beta^{-xy} + \text{Controls}_{jt}\gamma^{y} + \widetilde{\text{Error}}_{jt}^{y} \tag{29}$$

where $\text{E}\left[\beta_{jt}^{xy}\right]$ is equal to β^{xy} from equation (23). Then $\text{Error}_{jt}^{y} = \left(\beta_{jt}^{xy} - \beta^{xy}\right)\text{Crime}_{jt-1}^{x} + \widetilde{\text{Error}}_{jt}^{y}$. It is immediate that endogeneity might arise if $\text{Cov}\left(w_{jt}, \beta_{jt}^{xy}\right) \neq 0$ for any x, x', where $w_{jt} = \{\text{Crime}_{jt-1}^{x'}, \text{Other}_{jt}\}$. Thus, determinants of β_{jt}^{xy} which are correlated to other crimes or elements of Other_{jt} might be elements of $\mathbb{W}$.

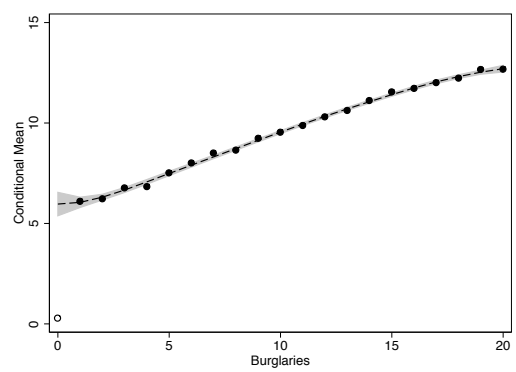
Figure 11 shows that past crimes of type x' are discontinuous at $\text{Crime}_{jt-1}^{x} = 0$, $x' \neq x$, and Figure 7 above shows that elements of Other_{jt} are discontinuous at $\text{Crime}_{jt-1}^{x} = 0$. These variables are all examples of w belonging to $\mathbb{W}_1^D$, which suggests that we can detect endogeneity from non-separability assumptions both within Crime_{jt-1} and between Crime_{jt-1} and Other_{jt}.

Figure 11: Discontinuity Plots: Nonlinear Effects

(a) Rapes



(b) Motor Vehicle Theft



Notes: The scatter plot in each panel represents $\text{E}\left[w_{jt}|\text{Crime}_{jt-1}^{x}\right]$ for each value of Crime_{jt-1}^{x}, where w_{jt} is described in the title, and Crime_{jt-1}^{x} is described on the horizontal axis. The hollow dot represents $\text{E}\left[w_{jt}|\text{Crime}_{jt-1}^{x} = 0\right]$. A cubic local polynomial fit with its 95% confidence interval is also presented along with an outside-of-the-sample prediction at $x = 0$.

Non-linear Treatment Effects

In principle, the spillovers that we want to estimate might be non-linear. For instance, it may take three or more weekly robberies to trigger a response by police, criminals or neighbors. Assume that the true model is given by

$$\text{Crime}_{jt}^{y} = f^{xy}\left(\text{Crime}_{jt-1}^{x}\right) + \text{Crime}_{jt-1}^{-x}\beta^{-xy} + \text{Controls}_{jt}\gamma^{y} + \widetilde{\text{Error}}_{jt}^{y} \tag{30}$$

where $f^{xy}(\cdot)$ is continuous at 0. Then $\text{Error}_{jt}^{y} = f^{xy}(\text{Crime}_{jt-1}^{x}) - \text{Crime}_{jt-1}^{x}\beta^{xy} + \widetilde{\text{Error}}_{jt}^{y}$, so Crime_{jt-1}^{x} is itself an element of $\mathbb{W}$.

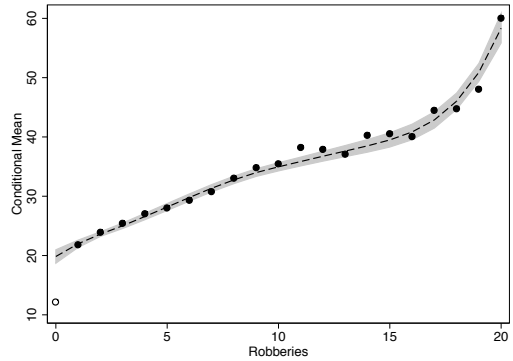
In Figure 12 we show that we can detect endogeneity from the misspecified functional form of $f^{xy}(\cdot)$. Panel (a) shows that sector-weeks with no robberies have discontinuously less light crimes on average, a drop from about 21 to about 12 light crimes per sector-week on average. For simplicity, assume for a moment that there was no heterogeneity across observations with the same value of Crime_{jt-1}^{x}, so that Light Crime = 21 for all sector-weeks with no robberies, and Light Crime = 12 for all sector-weeks with one robbery. Then, as long as $f^{xy}(21) - f^{xy}(12) \neq 9\beta^{xy}$, we should find $\text{Error}_{jt}^{y} = f^{xy}(\text{Crime}_{jt-1}^{x}) - \text{Crime}_{jt-1}^{x}\beta^{xy} + \widetilde{\text{Error}}_{jt}^{y}$ to vary discontinuously at $\text{Crime}_{jt-1}^{x} = 0$.

In reality, there could be heterogeneity across observations with the same value of Crime_{jt-1}^{x}. Panel (b) shows the cumulative distribution of light crimes across all sector-weeks with Robbery=$0, 1, 2, 3$. The distribution itself – not only its first moment – is discontinuous at 0. At the 20th percentile the horizontal difference between the solid curve (Robbery = 0) and the dashed curve (Robbery = 1) is about 18 (from 18 to 0), which suggests that if $f^{xy}(18) - f^{xy}(0) \neq 18\beta^{xy}$, then this discontinuity at 0 for the 20th percentile will show up in Error_{jt}^{y}. Similarly, the difference between the curves at the 80th percentile is roughly 5 (from 27 to 22), which suggests that if $f^{xy}(27) - f^{xy}(22) \neq 5\beta^{xy}$, then this discontinuity at 0 for the 80th percentile will also show up in Error_{jt}^{y}. It is unlikely that a non-linear $f^{xy}(\cdot)$ will behave like a linear function ($f^{xy}(d_1) - f^{xy}(d_2) = \beta^{xy}(d_1 - d_2)$) for the ranges of d_1 and d_2 suggested by Panel (b). Thus, these plots provide direct evidence that we can detect endogeneity from potential non-linear treatment effects.[53]

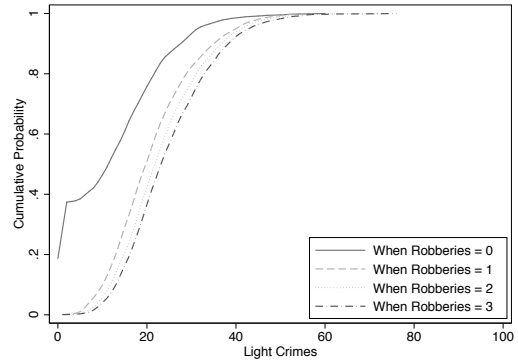
[53] This discussion also implies that if there is *any* non-linearity in the true model, then our test may have much more power to detect endogeneity than we have shown because we have only assessed discontinuity in the first moments of elements of $\mathbb{W}$.

Figure 12: Discontinuity Plots: Nonlinear Effects

(a) First Moment of Distribution of Light Crimes



(b) Cumulative Distribution of Light Crimes



Notes: Panel (a) shows the scatter plot of $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}\right]$ for each value of $\mathrm{Crime}^{x}_{jt-1}$, where w_{jt} is described in the title, and $\mathrm{Crime}^{x}_{jt-1}$ is described on the horizontal axis. The hollow dot represents $\mathrm{E}\left[w_{jt}|\mathrm{Crime}^{x}_{jt-1}=0\right]$. A cubic local polynomial fit with its 95% confidence interval is also presented along with an outside-of-the-sample prediction at $x=0$. Panel (b) shows the cumulative probability (i.e., $\Pr\left(w_j \leq w\right)$ for each value of w in the support) of the empirical distributions of w_j conditional on $x_j = x$, for selected values of x around $x=0$.

Misspecification of Controls$_{jt}$

Finally, a misspecification of Controls$_{jt}$ may also generate endogeneity. If w_{jt} is discontinuous at $\mathrm{Crime}^{x}_{jt-1}=0$, then $w'_{jt}=g(w_{jt})$ will also be discontinuous at $\mathrm{Crime}^{x}_{jt-1}=0$ for most functions $g(\cdot)$. Hence, the test will have power to detect endogeneity due to a misspecification of the functional form of observed controls w_{jt}.

Remark 9. We have argued that discontinuities in observables at the zero past crime threshold provide power to detect non-linear treatment effects. However, our test may be capable of detecting non-linearities in $f^{xy}(\cdot)$ that lead to biased estimates even in the absence of discontinuities at zero. If $f^{xy}(\cdot)$ is non-linear away from zero, then our linear model may incorrectly predict a discontinuous impact of past crimes on future crimes at zero even if there is no discontinuity in confounders at that point. In this case, Δ will be significantly different from zero because it corresponds to a misspecified model. Thus, in our linear specification, the exogeneity test has additional power to detect endogeneity stemming from functional form misspecification. In any case, when we performed the test of exogeneity on non-linear models in Section 7.1, we found no evidence of such non-linearity.

Remark 10. The examples of w shown in this section are observed yet omitted from the model. Thus, this is *direct* evidence that we can detect endogeneity from omitting that variable (in case our controls do not absorb them). This evidence is also useful because

it is suggestive *indirect* evidence of the kinds of omitted and unobserved variables that we can detect. If, for example, w is an observed confounder and w' is a correlated unobserved confounder, then finding a discontinuity in w implies a discontinuity in w' for most joint distributions of (w, w'). For instance, in Figure 7 we find that the population, the fraction of Whites and the fraction of 20-34 year olds in the neighborhood vary discontinuously when burglaries are zero. This suggests that many other characteristics of the neighborhood will also vary discontinuously when burglaries are zero, including ones that we cannot observe.

C Classifying Observed Variables as Detectable and Undetectable Confounders (For online publication only.)

Here we provide details of how we implemented the exercise described in Appendix B.1. Consider the baseline model below:

$$\text{Crime}_{jt} = \sum_{\tau=1}^{\bar{\tau}} \beta_\tau \text{Crime}_{jt-\tau} + \sum_{\tau=1}^{\bar{\tau}} \Delta_\tau D_{jt-\tau} + \epsilon_{jt} \tag{31}$$

for $\bar{\tau} = 1, ..., 6$. For each $\bar{\tau}$, we assess whether every potential confounder w belongs to $\hat{\mathbb{W}}_1^D$, $\hat{\mathbb{W}}_2^D$ or $\hat{\mathbb{W}}^C$ in the respective model. The full list of observed variables w is described in Section C.4.

C.1 Testing for whether w is a confounder (i.e. $w \in \hat{\mathbb{W}}$)

To assess whether w is a confounder, we test if its inclusion as a control in equation (31) affects the OLS estimates of $\hat{\beta}_\tau$ for some $\tau \leq \bar{\tau}$. If it does, then we can classify w as a confounder in this model since its omission contributes to a biased estimate. To allow for w to enter non-linearly in equation (31), we let $\boldsymbol{W} = (w, w^2, w^3, w^4, w^5, \log(1+w), \frac{1}{1+w}, \sqrt{w})$ be a vector of specific deterministic functions of w, all of which are continuous in the full support of w. We then jointly test for all $\tau \leq \bar{\tau}$ whether $\hat{\beta}_\tau^{OLS}$ changes when $\boldsymbol{W}$ is included in equation (31). If we reject the null hypothesis that it does not change, then we can conclude that $w \in \hat{\mathbb{W}}$, i.e., w is a confounder in this baseline model.[54]

[54] We implement these tests at the 5% significance level, with standard errors clustered at the sector × year × c level. Results are robust to other standard choices of significance level (1% and 10%) and to the other clustering levels discussed in Section 7.2.

C.2 Testing for whether $w \in \hat{\mathbb{W}}^D$

Next, for all variables $w \in \hat{\mathbb{W}}$, we assess whether there exists an x such that w varies discontinuously with Crime^x_{jt-1} at $\text{Crime}^x_{jt-1} = 0$ conditional on Crime^{-x}_{jt-1}. Specifically, for each $\bar{\tau}$, we test whether $\hat{\Delta}^w_\tau = 0$, jointly for all $\tau \leq \bar{\tau}$ in the following regression:

$$w = \sum_{\tau=1}^{\bar{\tau}} \beta^w_\tau \text{Crime}_{jt-\tau} + \sum_{\tau=1}^{\bar{\tau}} \Delta^w_\tau D_{jt-\tau} + \epsilon^w_{jt} \tag{32}$$

If we reject the null hypothesis, then we conclude that $w \in \hat{\mathbb{W}}^D$.

C.3 Testing for whether $w \in \hat{\mathbb{W}}^D_1$

Finally, for all variables $w \in \hat{\mathbb{W}}^D$, we assess whether w happens to be a confounder when $\text{Crime}^x_{jt-1} = 0$ for some x by repeating the test in C.1 on the restricted subsample of observations where $\text{Crime}^x_{jt-1} = 0$ for some x. If $\hat{\beta}^{OLS}_\tau$ does change for some $\tau \leq \bar{\tau}$ on this restricted subsample, then we can conclude that w belongs to $\hat{\mathbb{W}}^D_1$.

C.4 List of Potential Confounders w

The variables that we consider as potential confounders are constructed from two sources: the 2010 Census and the universe of all Dallas police reports from which our main estimating sample is constructed. None of these variables are deterministic functions of any other potential confounders (e.g., $w^2, \log w$, $1 - w$, etc). In total there are 691 such variables.

From the 2010 Census, we obtain sector level measures (w_j) of total population, its gender composition, its age distribution (broken down into the following bins: under 5, 5-9, 10-14, 15-17, 18-19, 20-34 and 35-59 years of age), and its racial and ethnic composition (subdivided into non-Hispanic Whites, non-Hispanic Blacks, Hispanics and Asians). In total, we obtain 13 observed variables from this data source.

From Dallas police reports, we include the following sets of sector-week level variables $w_{jt-\tau}$ (a total of 678 variables from this data source):

1. Crimes not included in $\mathbb{C}$ (60 variables): murder, larceny, carrying a concealed weapon, sex offenses, drugs, gambling, illegal liquor sales, DWI, other traffic offenses, and a more general variable that classifies any reported crime that does not belong to $\mathbb{C}$. In total there are 10 such variables for each $\tau = 1, ..., 6$.

2. Variables potentially omitted due to over-aggregation of crimes (210 variables): fractions of light crimes sub-classified as criminal mischief, drunk and disorderly conduct,

fence and vice; fraction of assaults deemed "aggravated"; and the number of crimes of type $x \in \mathbb{C}$ committed on a weekend, outdoor, at day time (8am-8pm), in a private residence, or in the beat at the center of the corresponding sector.[55] In total there are 35 such variables for each $\tau = 1, ..., 6$.

3. Variables potentially omitted due to under-aggregation of crimes (252 variables): Number of crimes of type $x \in \mathbb{C}$ committed in the corresponding division, in sectors within 1, 3 and 5 miles from corresponding sector,[56] in the sector in the center of corresponding division,[57] in winter weeks, and in summer weeks. In total there are 42 such variables for each $\tau = 1, ..., 6$.

4. Policing variables (84 variables): average response speed to the scene of the crime for each crime of type $x \in \mathbb{C}$ (police arrival time to crime scene minus time of call for service), and average duration at the scene of the crime for each crime of type $x \in \mathbb{C}$ (police departure time from crime scene minus police arrival time to crime scene). In total there are 12 such variables for each $\tau = 0, 1, ..., 6$.[58]

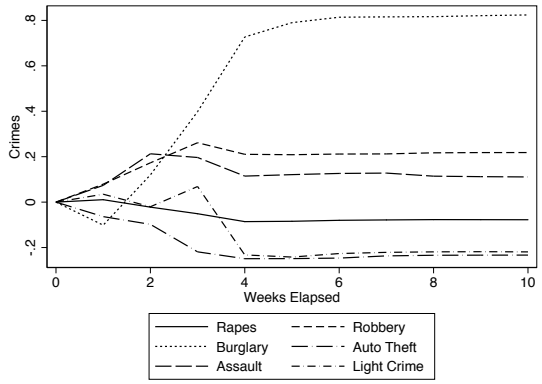
5. Variables potentially omitted due to misreporting (72 variables): fractions of crimes of type $x \in \mathbb{C}$ that are reported anonymously and by commercial establishments. In total there are 12 such variables for each $\tau = 1, ..., 6$.

D Supplemental Tables (For online publication only.)

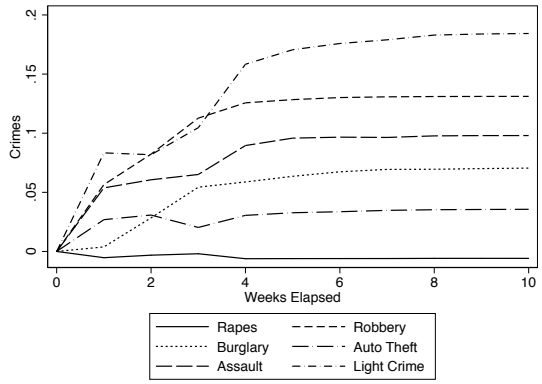
[55] This is defined as the beat whose centroid is closest to the centroid of the sector.
[56] These distances are based on the centroids of the sectors.
[57] This is defined as the sector whose centroid is closest to the centroid of the division.
[58] For the policing variables we also consider $\tau = 0$ as discussed in Section 8.

Figure 13: Cumulative Dynamic Spillovers From Unit Crime Reductions

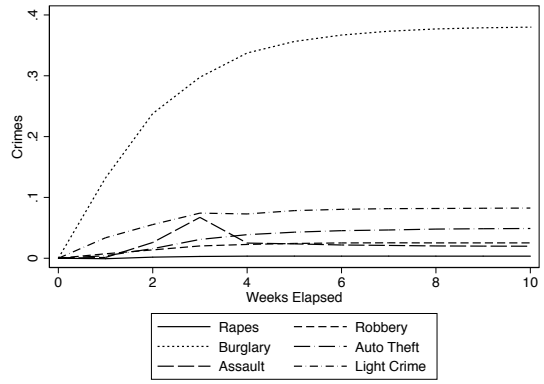
(a) Reduction in Rape



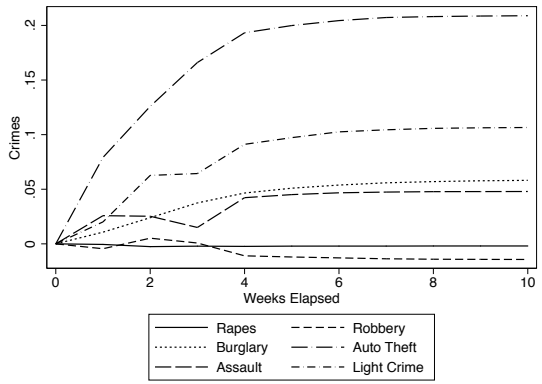
(b) Reduction in Robbery



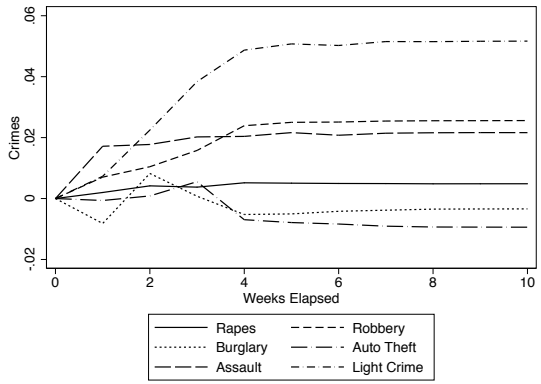
(c) Reduction in Burglary



(d) Reduction in Auto Theft



(e) Reduction in Assault



(f) Reduction in Light Crime

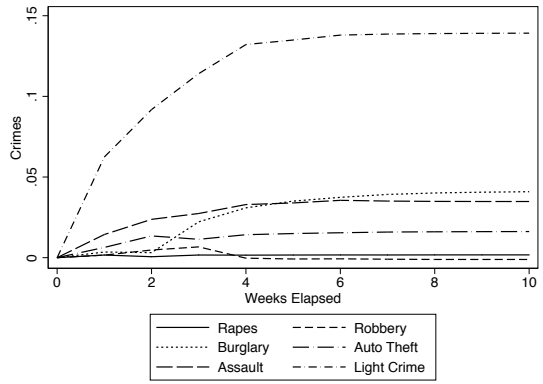
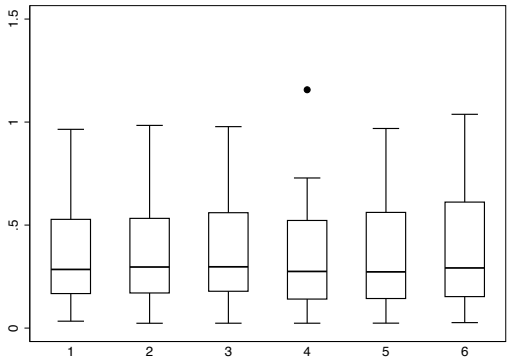


Figure 14: Standard Errors of Elements of $\hat{\Delta}$ for Models (1)-(6), j = Sector and t = Week



Notes: In this figure, we present a box plot of the standard errors of all elements of $\hat{\Delta}$ (the coefficients of the indicator variables reflecting the statistic of the test of exogeneity) for each of the six specifications of fixed effects. These are obtained from the models tested in Table 2 with j = Sector and t = Week. The box contains the 25th-75th percentile range, and the whiskers correspond to the upper and lower adjacent values. Dots represent outliers from the adjacent values.

Table 13: Standard Error Estimates for Specifications in Table 2 at Various Levels of Clustering (1 of 3)

Dep. Var. (t)	RHS Var. ($t-1$)	(1)	(2)	(3)	(4)	(5)	(6)
Rape	Rape	**0.067****	**0.019***	**0.019***	**0.020**	**0.021**	**0.025**
		0.083**	0.020	0.021	0.021	0.021	0.031
		*0.049***	*0.021**	*0.021*	*0.021*	*0.021*	*0.027*
	Robbery	**0.003****	**0.003**	**0.003**	**0.003**	**0.003**	**0.003**
		0.003**	0.003	0.003	0.003	0.003	0.003
		*0.003***	*0.002*	*0.002*	*0.003*	*0.003*	*0.003*
	Burglary	**0.002**	**0.001***	**0.001**	**0.001**	**0.001**	**0.002**
		0.002	0.001*	0.001	0.001	0.001	0.002
		0.001	*0.001**	*0.001*	*0.001*	*0.001*	*0.002*
	Auto Theft	**0.002**	**0.002***	**0.002***	**0.002**	**0.002**	**0.002**
		0.002	0.002	0.002	0.002	0.002	0.001
		0.002	*0.001**	*0.001***	*0.002*	*0.002*	*0.002*
	Assault	**0.001****	**0.001****	**0.001****	**0.001****	**0.001****	**0.001**
		0.001**	0.001**	0.001**	0.001**	0.001**	0.001
		*0.001***	*0.001***	*0.001***	*0.001***	*0.001***	*0.001*
	Light Crime	**0.001****	**0.001****	**0.001****	**0.001**	**0.001**	**0.001**
		0.001**	0.001**	0.001**	0.001	0.001	0.001
		*0.001***	*0.001*	*0.001***	*0.001*	*0.001*	*0.001*
Robbery	Rape	**0.093***	**0.082**	**0.082**	**0.074**	**0.073**	**0.073**
		0.112	0.062*	0.062	0.043	0.046	0.061
		*0.070***	*0.079*	*0.080*	*0.074*	*0.073*	*0.086*
	Robbery	**0.016****	**0.016****	**0.051****	**0.012****	**0.011****	**0.013****
		0.013**	0.013**	0.014**	0.011**	0.011**	0.014**
		*0.010***	*0.010***	*0.010***	*0.010***	*0.010***	*0.013***
	Burglary	**0.007****	**0.007****	**0.007****	**0.005****	**0.006****	**0.006**
		0.008**	0.008**	0.008**	0.005**	0.006**	0.008**
		*0.005***	*0.005***	*0.005***	*0.005***	*0.005***	*0.006*
	Auto Theft	**0.008****	**0.008****	**0.008****	**0.006****	**0.007***	**0.006**
		0.008**	0.008**	0.005**	0.006**	0.006**	0.007**
		*0.006***	*0.006***	*0.006***	*0.006***	*0.006***	*0.008*
	Assault	**0.005****	**0.005****	**0.005****	**0.004****	**0.005****	**0.004***
		0.004**	0.004**	0.004**	0.004**	0.005**	0.004*
		*0.003***	*0.003***	*0.003***	*0.004***	*0.004***	*0.005**
	Light Crime	**0.004**	**0.004****	**0.004***	**0.004****	**0.004****	**0.004**
		0.004*	0.004**	0.004**	0.004**	0.004**	0.005
		*0.003**	*0.003***	*0.003***	*0.003***	*0.004***	*0.004*

Table 13: Standard Error Estimates for Specifications in Table 2 at Various Levels of Clustering (2 of 3)

Dep. Var. (t)	RHS Var. ($t-1$)	(1)	(2)	(3)	(4)	(5)	(6)
Burglary	Rape	**0.140**	**0.148**	**0.145**	**0.134**	**0.132**	**0.147**
		0.165	0.153	0.152	0.134	0.133	0.152
		0.110	*0.151*	*0.015*	*0.145*	*0.142*	*0.163*
	Robbery	**0.025****	**0.025****	**0.025****	**0.021****	**0.022****	**0.024**
		0.024**	0.023**	0.024**	0.024**	0.026**	0.031
		*0.019***	*0.019***	*0.019***	*0.019***	*0.019***	*0.023*
	Burglary	**0.019****	**0.019****	**0.017****	**0.018****	**0.020****	**0.013****
		0.020	0.020**	0.018**	0.019**	0.021**	0.017**
		*0.010***	*0.010***	*0.010***	*0.011***	*0.011***	*0.013***
	Auto Theft	**0.015****	**0.015****	**0.015****	**0.013****	**0.014***	**0.013**
		0.020*	0.020**	0.009*	0.016	0.015*	0.017
		*0.006***	*0.011**	*0.011***	*0.012**	*0.012**	*0.014*
	Assault	**0.010****	**0.010****	**0.009****	**0.008****	**0.008***	**0.007**
		0.011**	0.010**	0.010**	0.006**	0.007*	0.008
		*0.006***	*0.006***	*0.006***	*0.008***	*0.008**	*0.010*
	Light Crime	**0.005****	**0.009****	**0.009****	**0.007****	**0.007****	**0.008**
		0.010**	0.011**	0.010**	0.008**	0.007**	0.008
		*0.006***	*0.006***	*0.006***	*0.007***	*0.007***	*0.008*
Auto Theft	Rape	**0.128**	**0.128**	**0.124**	**0.107**	**0.112**	**0.127**
		0.148	0.113	0.110	0.097	0.100	0.118
		0.099	*0.133*	*0.130*	*0.121*	*0.120*	*0.141*
	Robbery	**0.020****	**0.020****	**0.021****	**0.015****	**0.015****	**0.016**
		0.021**	0.021**	0.022**	0.013**	0.013**	0.015
		*0.016***	*0.016***	*0.016***	*0.016***	*0.016***	*0.019*
	Burglary	**0.011****	**0.012****	**0.011****	**0.010****	**0.010****	**0.010**
		0.014**	0.014**	0.013**	0.009**	0.008**	0.012
		*0.007***	*0.007***	*0.007***	*0.008***	*0.008***	*0.010*
	Auto Theft	**0.016****	**0.016****	**0.015****	**0.015****	**0.015****	**0.012****
		0.015**	0.015**	0.015**	0.012**	0.012**	0.014**
		*0.009***	*0.009***	*0.009***	*0.010***	*0.010***	*0.012***
	Assault	**0.008****	**0.008****	**0.008****	**0.006****	**0.007****	**0.007**
		0.009**	0.009**	0.010**	0.007**	0.007**	0.007
		*0.005***	*0.005***	*0.005***	*0.006***	*0.006***	*0.008*
	Light Crime	**0.007****	**0.008****	**0.008****	**0.007****	**0.007****	**0.007**
		0.009**	0.010**	0.009**	0.008**	0.008**	0.008
		*0.005***	*0.005***	*0.005***	*0.006***	*0.006***	*0.007*

Table 13: Standard Error Estimates for Specifications in Table 2 at Various Levels of Clustering (3 of 3)

Dep. Var. (t)	RHS Var. ($t-1$)	(1)	(2)	(3)	(4)	(5)	(6)
Assault	Rape	**0.181**	**0.214**	**0.212**	**0.202**	**0.194**	**0.199**
		0.226	0.217	0.210	0.207	0.213	0.232
		0.148	*0.208*	*0.207*	*0.200*	*0.190*	*0.210*
	Robbery	**0.036****	**0.036****	**0.034****	**0.031****	**0.030****	**0.030***
		0.033**	0.033**	0.033**	0.036**	0.036**	0.030
		*0.025***	*0.025***	*0.025***	*0.025***	*0.024***	*0.030**
	Burglary	**0.017****	**0.017****	**0.016****	**0.014****	**0.014****	**0.015**
		0.019**	0.019**	0.015**	0.013**	0.013**	0.012
		*0.012***	*0.012***	*0.012***	*0.013***	*0.012***	*0.015*
	Auto Theft	**0.020****	**0.020****	**0.020**	**0.019****	**0.019****	**0.017**
		0.020**	0.021	0.023	0.018**	0.018**	0.018
		*0.014***	*0.014***	*0.014*	*0.015***	*0.014***	*0.017*
	Assault	**0.017****	**0.017****	**0.015****	**0.014****	**0.015****	**0.012**
		0.027**	0.027**	0.021**	0.018**	0.020**	0.013*
		*0.010***	*0.010***	*0.009***	*0.011***	*0.010***	*0.013*
	Light Crime	**0.012****	**0.012****	**0.012****	**0.011****	**0.011****	**0.010**
		0.010**	0.010**	0.009**	0.010**	0.010**	0.009
		*0.009***	*0.009***	*0.009***	*0.009***	*0.009***	*0.010*
Light Crime	Rape	**0.184****	**0.210**	**0.208**	**0.185**	**0.193**	**0.204**
		0.211**	0.224	0.222	0.207	0.226	0.234
		*0.151***	*0.211*	*0.210*	*0.197*	*0.196*	*0.225*
	Robbery	**0.032****	**0.034****	**0.023****	**0.025****	**0.027****	**0.029****
		0.030**	0.032**	0.031**	0.026*	0.028**	0.029**
		*0.025***	*0.025***	*0.025***	*0.026***	*0.026***	*0.032***
	Burglary	**0.019****	**0.020****	**0.019****	**0.015****	**0.015****	**0.016****
		0.019**	0.021**	0.019**	0.014**	0.013**	0.016**
		*0.012***	*0.012***	*0.012***	*0.013***	*0.014***	*0.017**
	Auto Theft	**0.021****	**0.022****	**0.023****	**0.020****	**0.021****	**0.019**
		0.025**	0.027**	0.026**	0.023**	0.021**	0.026
		*0.015***	*0.015***	*0.015***	*0.016***	*0.016***	*0.020*
	Assault	**0.013****	**0.014****	**0.013****	**0.011****	**0.012****	**0.011**
		0.016**	0.015**	0.015**	0.012**	0.012**	0.010
		*0.008***	*0.008***	*0.009***	*0.010***	*0.011***	*0.013*
	Light Crime	**0.014****	**0.015****	**0.015****	**0.014****	**0.014****	**0.013****
		0.014**	0.015**	0.016**	0.013**	0.013**	0.014**
		*0.009***	*0.009***	*0.009***	*0.010***	*0.010***	*0.013***
Discontinuity test P-value		**0.00****	**0.00****	**0.00****	**0.01****	**0.05**	**0.66**
		0.00**	0.00**	0.00**	0.00**	0.02*	0.08
		*0.00***	*0.00***	*0.00***	*0.02**	*0.06*	*0.85*

Notes: Standard error estimates are presented for models estimated in Table 2. The standard errors in **bold** are clustered at the sector-year-crime type level. The standard errors in normal font are clustered at the division-year-crime type level. The standard errors in *italics* are clustered at the division-week-crime type level. *(**): corresponding coefficient (not shown in this Table) is significantly different from zero at 5% (1%) level.