NEW YORK CITY POLICE REFORM AND REINVENTION COLLABORATIVE

Collateral Consequences

THE EFFECTS OF JUSTICE PROCESSING FOR VIOLATIONS OF DRUG LAWS IN NEW YORK CITY

APPENDIX:
INSTRUMENTAL VARIABLES APPROACH

JOHN K. ROMAN
GREGORY HAUGAN
BENJAMIN SCHAPIRO
SOFIA RODRIGUEZ

NORC AT THE UNIVERSITY OF CHICAGO

JOHN JAY COLLEGE OF CRIMINAL JUSTICE Research and Evaluation Center

MAY 2024

Appendix: Instrumental Variables Approach

The goal of the study is to identify the effect of drug-related arrests on different outcome variables of interest, such as property tax assessments. The regression to identify this effect takes the form:

$$Y_{it} = \beta_0 + \beta_1 Drug\ Arrests_{it-1} + \boldsymbol{\beta'X_{it}} + \alpha_i + \delta_t + u_{it} \qquad (1)$$

Where Y_{it} is the outcome of interest (e.g., property tax assessments) in census tract *i* in year *t*. β_1 is the coefficient of interest to be estimated, which measures the effect of Drug Arrests in census tract *i* in the preceding year. $\boldsymbol{\beta'}$ is a vector of coefficients measuring the effects of $\boldsymbol{X_{it}}$, a vector of control variables for census tract *i* in quarter *t*, which might include also lagged variables, and α_i and δ_t are census tract and time period fixed effects, respectively. Finally, u_{it} is an error term.

The main problem with this regression is that, even when α_i and δ_t are included to control for census tract and time period-specific traits (which may be unobservable), the error term is almost certainly correlated with Drug Arrests. Some examples of how this could show up in practice:

- Specific locations are targeted for redevelopment by property developers. The number of drug arrests could factor into how they select these locations, or the decisions themselves could factor into how the locations are policed (e.g., developers or new residents push police to patrol the areas more frequently or aggressively), or the redevelopment could displace residents who are more/less likely to be arrested.
- Police do not patrol randomly, but rather respond to local dynamics and pressures. This could be related to a recent news story about local crime, a high-profile murder, a recent string of robberies, or pressure from a powerful local politician who is up for re-election. In turn, the locations and times where crime draws the attention of journalists, where homicides or robberies occur, or where elected representatives have more/less influence are not random and are likely correlated with the error term.

In short, while α_i captures the static unobservable characteristics of a census tract that may be related to crime and the outcome of interest (e.g., distance to highway, public transport access, etc.), and δ_t captures the unobservable characteristics that vary over time but are uniform across all census tracts (e.g., wholesale price of drugs, citywide tax revenue, active stop-and-frisk policy, etc.), there are likely still unobservable characteristics that vary over time within a census tract that are correlated with the error term, and result in a biased estimate of β_1.

Instrumental Variables

A potential solution to the identification problem for β_1 is to take an instrumental variables approach. This approach requires a variable, Z_{it-1}, that is correlated with $Drug\ Arrests_{it-1}$, but not correlated with u_{it}. In other words, the instrument must be related to drug arrests, and not related to Y_{it} via any other channel than its effect on drug arrests. If such an instrument can be found, Two Stage Least Squares (2SLS) can be used to estimate Equation (2) and use the predicted values for Drug Arrests from Equation (2) in Equation (3).

$$\widehat{Drug\ Arrests}_{it-1} = \gamma_0 + \gamma_1 Z_{it-1} + \boldsymbol{\gamma'X_{it}} + \alpha_i + \delta_t \qquad (2)$$

$$Y_{it} = \beta_0 + \beta_1 \widehat{Drug\ Arrests}_{it-1} + \boldsymbol{\beta'X_{it}} + \alpha_i + \delta_t + u_{it} \qquad (3)$$

Where $\widehat{Drug\ Arrests}_{it-1}$ is the predicted value of drug arrests from Equation (2). In many situations, it can be very difficult to find a valid instrument. This case is no exception, as it requires something that varies by geography and time period in a way that influences drug arrests, but not the outcomes of interest, like tax assessments or gun-related homicides. A city program that randomly selected city tracts to receive additional police patrol time might work to construct an instrument, if such a program actually existed. It is difficult to think of other good examples of what might work as an instrument. Many other plausibly exogenous factors that influence drug arrests vary over time, but not by geography (e.g., citywide stop-and-frisk policy; arrival of crack, fentanyl, K-2/Spice, or other high-potency drugs on the market; etc.), which makes them unusable.

Bartik Shock as an Instrument

One possibility could be to use a Bartik Shock as an instrument. The basic idea of a Bartik Shock is to assume that there are universal shocks that impact drug arrests (e.g., citywide stop-and-frisk policies), but that different census tracts have differential exposure to these shocks. For example, Hunts Point in the Bronx has a well-known narcotics market, high poverty rates and large presence of vulnerable populations, which means that this neighborhood may have been more exposed to stop-and-frisk, while the much wealthier neighborhoods of the Upper West Side in Manhattan would have been little affected by stop-and-frisk. While the factors that determine this differential exposure are likely endogenous to the level of the outcome variable, the strategy assumes that they are not related to changes in the outcome variable.

The typical use of a Bartik instrument is in macroeconomic settings, such as to examine the impact of job growth on real estate prices. Assume the following regression:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \boldsymbol{\alpha}' \boldsymbol{P_{it}} + u_{it} \qquad (4)$$

Where X_{it} is the employment growth rate in location i in time t, and is assumed to be endogenous, and $\boldsymbol{P_{it}}$ is a vector of control variables. The Bartik Instrument derives from the fact that X_{it} can be rewritten as:

$$X_{it} = \sum_{k=1}^{k} z_{ikt} g_{ikt} \qquad (5)$$

Where $z_{ikt}$is the share of jobs in location i in industry k in period t, and g_{ikt} is the employment growth rate in location i in industry k in period t. In other words, the employment growth rate in location i can be decomposed as a weighted average of employment growth rates across industries in i. In turn, g_{ikt} can be decomposed as:

$$g_{ikt} = g_{kt} + \tilde{g}_{ikt} \qquad (6)$$

Where g_{kt} is the industry-wide growth rate across all geographies, and $\tilde{g}_{ikt}$is an idiosyncratic variation at the level of location, industry, and time. In the conventional use, industry shares are fixed to an initial time period. Then the Bartik Instrument $\sum_{k=1}^{k} z_{ik0} g_{kt}$ serves as an instrument for X_{it} in the first stage of the 2SLS set up. The typical set up also defines Y_{it} as a change or growth rate in the outcome variable, rather than measuring the level. Goldsmith-Pinkham, Sorkin, and Swift (2020) show that the key assumption is therefore that the industry shares are exogenous to changes in the error term, which is the same as the identifying assumption in difference-in-differences models. Importantly, in situations where this assumption is not plausible, Borusyak, Hull, and Jaravel (2022) show that identification can instead come from the exogeneity of the shocks. These articles also describe ways of testing these assumptions.

Proposed Approach

The proposed approach for resolving the identification issues for the effect of drug arrests is a variation on the typical use of the Bartik Shock. Return to Equation (1).The main outcome variable of interest is defined as $Drug\ Arrests_{it-1}$, the number drug arrests in period t-1:

$$Y_{it} = \beta_0 + \beta_1 Drug\ Arrests_{it-1} + \boldsymbol{\beta' X_{it}} + \alpha_i + \delta_t + u_{it} \qquad (7)$$

Here too, $Drug\ Arrests_{it-1}$is likely endogenous. However, a Bartik-like shock may be used as an instrument:

$$\widehat{Drug\ Arrests}_{it-1} = \gamma_0 + \gamma_1 Z_{it-1} + \boldsymbol{\gamma' X_{it}} + \alpha_i + \delta_t \quad (8)$$

Where the instrument Z_{it-1} is:

$$Z_{it-1} = w_i z_{l \neq i, t-1} \qquad (9)$$

Similar to Equation (2), Equation (8) is just a regression to find the predicted value of the endogenous variable of interest in a regression that includes the exogenous instrument, Z_{it-1}. In Equation (9),the instrument is defined. Let $w_i$be a location-specific weight measuring exposure to universal shocks. One potential weight could be the share of all drug-related arrests in the city in the pre-analysis period (i.e., 2006 and 2007) that occurred in census tract *i*. Let $z_{l \neq i, t-1}$ be the number of drug arrests across all locations other than *i* (i.e., city-wide drug arrests, excluding census tract *i*). Then the key assumption for Z_{it-1} in (8) is similar to a parallel trends assumption in a difference-in-differences model: In the absence of these universal shocks, a census tract's share of arrests in the pre-analysis period must be uncorrelated with the change in the outcome. In other words:

$$Drug\ Arrests_{it-1} = Drug\ Arrests_{l \neq i, t-1} w_i + \widetilde{Drug\ Arrests}_{it-1} \qquad (10)$$

and:

$$\mathbb{E}[\Delta \widetilde{Drug\ Arrests}_{it-1} | w_i] = 0 \qquad (11)$$

Equation (10) says that $Drug\ Arrests_{it-1}$ is equal to the product of the universal Bartik Shock and the exposure weight, plus $\widetilde{Drug\ Arrests}_{it-1}$, an idiosyncratic location-level shock in *t-1*. Equation (11) is a key identifying assumption, which says that the expected value of the idiosyncratic shock $\widetilde{Drug\ Arrests}_{it-1}$ is orthogonal to the exposure weights. The assumption in Equation (11) might be violated, for example, if census tracts with more drug arrests in the pre-analysis period show faster growth in real estate prices because developers systematically target them for their building projects (for reasons other than trends in drug arrests).