

Testing Economic Hypotheses with State-Level Data: A Comment on Donohue and Levitt (2001)

Christopher L. Foote and Christopher F. Goetz

Abstract:

State-level data are often used in the empirical research of both macroeconomists and microeconomists. Using data that follows states over time allows economists to hold constant a host of potentially confounding factors that might contaminate an assignment of cause and effect. A good example is a fascinating paper by Donohue and Levitt (2001, henceforth DL), which purports to show that hypothetical individuals resulting from aborted fetuses, had they been born and developed into youths, would have been more likely to commit crimes than youths resulting from fetuses carried to term. We revisit that paper, showing that the actual implementation of DL's statistical test in their paper differed from what was described. (Specifically, controls for state-year effects were left out of their regression model.) We show that when DL's key test is run as described and augmented with state-level population data, evidence for higher per capita criminal propensities among the youths who would have developed, had they not been aborted as fetuses, vanishes. Two lessons for empirical researchers are, first, that controls may impact results in ways that are hard to predict, and second, that these controls are probably not powerful enough to compensate for the omission of a key variable in the regression model. (Data and programs to support this comment are available on the web site of the Federal Reserve Bank of Boston.)

JEL Classifications: I18, J13

Christopher L. Foote is a Senior Economist at the Federal Reserve Bank of Boston. Christopher F. Goetz is a Research Assistant at the Federal Reserve Bank of Boston. Their email addresses are chris.foote@bos.frb.org and christopher.goetz@bos.frb.org, respectively.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at <http://www.bos.frb.org/economic/wp/index.htm>.

The views expressed in this paper are those of the authors alone and are not necessarily those of the Federal Reserve System in general or of the Federal Reserve Bank of Boston in particular.

We thank John Donohue and Steven Levitt for making their data and computer programs available via Donohue's Yale University website.

This version: November 22, 2005

Testing Economic Hypotheses with State-Level Data: A Comment on Donohue and Levitt (2001)

Christopher L. Foote and Christopher F. Goetz

1. Introduction

In a fascinating and controversial paper, Donohue and Levitt (2001, henceforth DL) suggest that the advent of legalized abortion in the 1970s is responsible for much of the steep and persistent drop in the crime rate during the 1990s. DL suggest two possible channels for a causal link between abortion and crime. First, holding current pregnancy rates constant, abortion lowers the total number of young people in populations 15-25 years later. Because young people are more likely to commit crimes than older ones, the criminal propensity of the future population falls, and crime declines. Far more controversial is DL's second suggested channel between abortion and crime. If unwanted children are more likely to commit crimes, abortion can lower crime by preventing the births of persons most likely to become criminals. Although the morality of abortion is controversial, DL correctly point out that the question of whether abortion reduces crime is purely empirical. Their paper provides evidence for a link between abortion and crime that operates via both channels outlined above. Their estimates imply that the availability of abortion reduced crime rates in the 1990s by as much as one-half.¹

DL's paper has incited a storm of criticism from people who believe that the authors are making a normative claim about the social benefits of legalized abortion. We are also critical of DL's paper, but for purely empirical reasons. We show that when examined correctly, there is no evidence in DL's own data for the second channel they outline for an abortion-crime link. In other words, there are no statistical grounds for believing that the hypothetical youths who were aborted as fetuses would have been more likely to commit crimes had they reached maturity than the actual youths who developed from fetuses carried to term. Abortion may lower crime by reducing the overall number of young persons in the population (that is, via the first channel DL specify). Yet because only one channel operates in the abortion-crime link, the overall impact of abortion on crime would be smaller than DL claim.

¹ Because the abortion-crime link plays a prominent part in Levitt and Dubner's bestselling book (2005), many people who are not regular readers of top-tier economics journals are now familiar with DL's claim.

This comment is organized as follows. Section 2 outlines the statistical issues important for testing an abortion-crime relationship. Section 3 discusses the key test that DL performed to test their claim. First, we show that this test included an apparently inadvertent but serious computer programming error. Second, we argue that DL’s choice of analyzing *total* rather than *per capita* criminal activity in this test was inappropriate, even though the population data needed for a per capita analysis are imperfect. Section 4 shows that when their error is corrected and a per capita analysis is performed, evidence for DL’s second channel vanishes. Section 5 and a data appendix conclude the paper.

2. Statistical Issues in Testing the Abortion-Crime Link

The statistical issues related to testing an abortion-crime link are the same as those that arise in any test of cause and effect, such as whether a new drug is effective in combating a certain illness. Researchers testing drugs typically use randomized trials, in which test subjects suffering from an illness are randomly separated into two groups. A “treated” group is given the drug while the “control” group is given a placebo. After a suitable period, the average incidence of the disease is compared across the two groups. If the treated group has a lower incidence than the untreated group (and this difference is “statistically significant,” or too large to have easily occurred by chance), researchers conclude that the drug is effective in treating the disease.²

Randomization in drug trials is useful because it virtually eliminates the possibility that any improvement in the treated group is due to some confounding factor. Say that people suffering the disease become healthier when the weather is warm. In a randomized drug trial, we would be sure that any improvement in the treated group was due to the drug — not the weather — because randomization would insure that the weather facing the typical control patient was the same as that facing the typical treated patient. Researchers would not even have to know about all of the potential confounding factors that could influence patient outcomes. Randomization would eliminate the effects of any confounding factor that did exist.

DL test the abortion-crime link by comparing the criminal behaviors of different groups of young people. For obvious reasons, these groups are not formed as part of randomized trials. The groups are instead defined by their birth years and states of residence. Defining

² The gold standard for these tests is the “double-blind” test, in which neither the doctors giving the drugs nor the patients taking them know which patients are in the treated group. This ensures that subjective hopes and desires for the drug’s effectiveness do not cloud the assessment of patient outcomes.

the groups in this way generates variation in “abortion exposure,” or the abortion rate relevant for a particular cohort of young people.³ DL point out that cohorts with high levels of abortion exposure are likely to contain fewer unwanted or mistimed children than groups with low abortion exposure. They test whether these high-exposure groups are also less likely to commit crime.

Confounding factors in state-level statistical tests

But just as a change in the natural climate can contaminate a non-randomized drug trial, changes in state-level criminal climates can contaminate a study of abortion and crime that is based on state-level data. Suppose we found that groups of young people with high levels of abortion exposure did not commit crimes very often. This finding would be initial evidence that abortion lowers crime, perhaps through the unwantedness or timing channels that DL discuss. But several other factors also influence crime. Some of these other factors vary on the *state-year* level, meaning that they affect everyone within a state in a particular year. In our drug-test example, a heat wave that temporarily increases the temperature in some states and years, but not others, is an example of a state-year effect. In the abortion-crime case, temporary state-specific crime waves brought about by the introduction of new, illicit drugs (like crack cocaine) can be classified as state-year effects.⁴ If not addressed explicitly, state-year effects will lead to inaccurate results if they affect crime at the same time that the abortion exposure of test subjects is changing. The researcher will link the change in crime to the abortion variable when crime is really changing because of the state-year effects.

An example of how state-year effects can affect statistical studies comes from a separate literature of interest to policymakers today, namely, the study of how the labor market changes as the working population gets older. The conventional wisdom among economists is that the unemployment rate should fall as the population ages. Young people often become unemployed as they search for the right jobs, so the economy’s unemployment rate should decline as young people make up a progressively smaller share of an aging population. Testing this hypothesis with state-level data generates a surprising result,

³ For example, the level of abortion exposure for Massachusetts 15-year-olds in 1990 is the rate of abortions per live births in Massachusetts in 1974 (= 1990-15-1). The “1” in this expression accounts for the period between the time a fetus would typically be aborted and when it would be born.

⁴ Joyce (2004, 2005) has criticized DL (2001) for conflating the effects of abortion with arrival of crack cocaine in the 1980s and 1990s. We discuss Joyce’s work more extensively below.

however. From the late 1970s to the early 1990s, U.S. states typically enjoyed their lowest unemployment rates when their youth shares of population were high, not low.⁵ This was certainly the case for Massachusetts. Figure 1 graphs the unemployment rate for Massachusetts from 1978 to 1993, along with the share of working age population (ages 16 to 64) that is typically considered young (ages 16 to 24) by labor economists.⁶ The graph clearly suggests that a rising youth share *reduces* unemployment, in contrast to the conventional wisdom.

We should not be too quick to throw out the conventional wisdom, however. Many other factors also affected state-level unemployment during this period. In Massachusetts, the youth share was particularly high in the mid 1980s. But this was also the era of the “Massachusetts Miracle”: a boom among high-technology firms and knowledge-intensive industries that dramatically increased labor demand in the state. By chance, Massachusetts’s rising youth share *coincided* with a fall in the state’s unemployment rate in the 1980s, but it did not *cause* this fall. The low unemployment rate was instead caused by state-year effects, specifically, all of the effects that made up the Massachusetts Miracle.⁷

Accounting for state-year effects

In DL’s case, there is a way to remove state-year effects from the analysis. This is possible because each observation in their data set is defined on the *state-year-age* level. For example, for Massachusetts in 1990, the data contain individual values for abortion exposure and crime for several age cohorts (15-year-olds, 16-year-olds, etc., up to 24-year-olds).⁸ Because DL have several points of data within each state-year combination, they can test for the abortion-crime link in a way that will not be contaminated by state-year effects.

⁵ The first paper to notice this correlation is Shimer (2001).

⁶ The figure graphs Massachusetts unemployment and youth shares minus the national means of these variables. These are the relevant concepts for a state-level study of the demographic determinants of unemployment. Because of the aging of the baby boom generation, the actual youth share in Massachusetts is trending down over this period, as it is in all U.S. states.

⁷ State-year factors were also important in other states during this period. Among all states, the average correlation between unemployment and youth shares is strongly negative from 1973–1996, just as in the Massachusetts case. But this correlation essentially disappears if the sample period is extended to 2002. This disappearance suggests that the original correlation does not reflect a causal effect of the youth share, but rather a chance pattern of state-year effects across the country.

⁸ By contrast, in the unemployment study above, there is only a single unemployment rate and single youth share for Massachusetts in 1990. Hence, the unemployment study cannot eliminate confounding state-year effects.

A well-known way to do this is as follows: First, calculate the *mean* levels of both abortion exposure and criminal activity across age groups within each state and year.⁹ Then, subtract the relevant state-year means from the original, age-specific data. Finally, look for a statistical relationship between *de-meaned* crime and *de-meaned* abortion exposure. Removing state-year means from the abortion and crime data removes any confounding effect that affects all ages within a state equally (that is, that operates on the state-year level). If we continued to find a negative relationship between abortion and crime in the de-meaned data, we would be safe in surmising that this relationship was not being driven by a confounding state-year effects.

In addition to controlling for state-year effects, we can also control for confounding effects that operate on the *age-year* level. An age-year factor would affect, say, all 19-year-olds in the United States in 1989, or all U.S. 15-year-olds in 1992. Controlling for age-year effects would require “de-meaning the de-meaned data,” this time from age-year means. Finally, we might still be concerned that any relationship between abortion and crime in the twice-de-meaned data could be due to a confounding effect that operated on the *state-age* level. Such an effect would be present if, say, 15-year-olds in Ohio were more likely to commit crimes than 19-year-olds in Texas. A third and final round of de-meaning, this time from state-age means, would clean the data of state-age effects, and effectively isolate the causal impact of abortion on crime.

In short, organizing data along the state-year-age level allows the researcher to eliminate any confounding effect that varies along any two of these levels (that is, state-year, age-year, and state-age). Virtually all confounding effects would be expected to operate in one of these three ways. Therefore, if we still detected a negative relationship between abortion and crime after three rounds of de-meaning, we could be reasonably sure that this relationship was not due to a confounding factor. Importantly, we would not need to know the source of these state-year, age-year, or state-age effects. They could be due to the emergence of crack in the 1980s and 1990s, the proliferation of video games, a general breakdown in social morals, changing unemployment rates, or anything else. As long as the confounding factors affected the appropriately defined *means* of criminal activity, they would not contaminate a study of cause and effect using de-meaned data.

While this process would clearly be tedious if done by hand, modern tools of multiple

⁹ For example, figure the average level of arrests across all age groups in Massachusetts in 1990. Do the same for abortion exposure.

regression analysis will de-mean the data appropriately when state-year, age-year, and/or state-age effects are specified in the analysis. In practice, a multiple regression is simply a mathematical statement that the variable being explained (here, criminal behavior of a group of young people) depends on our independent variable of interest (abortion exposure) and whatever “controls” we enter (state-year, age-year, and state-age). Including the controls in the regression automatically de-means the data appropriately. The regression calculates an estimated effect for each independent variable, taking into account the various controls. Typically, running a regression with modern statistical software is done with a single line of computer code.

3. DL’s Main Statistical Test: Two Criticisms

Our first criticism of DL’s paper is that in an apparent mistake, they did not include state-year controls in their regressions. Therefore, any confounding effects that vary along the state-year level potentially contaminate their estimated effects.¹⁰ This contamination is potentially serious, because the crack wave waxed and waned during the sample period, and by all accounts, crack affected different states at different times and with differing levels of severity. As a result, the crack wave probably generated effects on crime that varied on the state-year level.

Indeed, Joyce (2004) argues that DL (2001) confuse the effects of abortion on crime with the crack wave (though he does not mention the omission of state-year controls in DL’s main regression). Joyce suggests a different way of identifying the effect of abortion that uses data organized on the state-year level, not the state-year-age level, and that focuses more closely on cohorts born around 1973. In their response to Joyce, DL (2004) rebut a number of his arguments, returning at the end to their regression that uses data on the state-year-age level. They correctly point out that tests performed on these data are to be preferred to other identification strategies, including Joyce’s, because tests using state-year-age data can control for so many potentially confounding factors:

¹⁰ Specifically, DL’s Equation 3 (p. 411) is

$$\ln(\text{ARRESTS}_{stb}) = \beta_1 \text{ABORT}_{sb} + \gamma_s + \lambda_{tb} + \theta_{st} + \epsilon_{stb}.$$

In this equation, the subscript s indexes the individual state, b indexes the cohort’s birth year, and t indexes the year in which the criminal activity is being measured. ABORT_{sb} is the abortion exposure of the cohort born in state s in birth year b . DL’s error is that there is nothing in their computer code to reflect the θ_{st} term, which represents the state-year controls. The computer code in question (as well as the data DL used in their paper) can be found at Donohue’s Yale University web page: <http://islandia.law.yale.edu/donohue/pubsdata.htm>.

Because of the richness of the data (the variation is at the level of state-cohort-year, rather than simply state-year), we are able to control for national cohort effects, national year effects, and (in some specifications) state-age interactions in crime rates (DL [2004], p. 45).

Note that this paragraph omits any discussion of *state-year* effects, which, as we have seen, can also be controlled for using DL’s data and which were supposed to have been included in their original paper. This omission is puzzling. In our view, state-year effects are precisely the type of effects that a crack wave would be expected to generate. Hence, controls for state-year effects should have taken center stage in a response that argues that the link between abortion and crime is not due to the potentially confounding factor of the crack wave. Presumably, however, DL also ran the regressions in their 2004 response to Joyce without state-year controls.¹¹

Our second criticism of DL’s original (2001) paper is more basic. When comparing the arrest propensities of different groups of young people, DL use the *total* number of arrests in an age-state-year observation, not the number of arrests *per capita*. Unfortunately, only by looking at per capita crime rates can we determine conclusively that the hypothetical youths who would have developed had they not been aborted as fetuses would have been more likely to commit crimes. If abortion reduces the size of future cohorts (because some fetuses will not be born), DL’s choice of specification means that they are likely to find a negative relationship between higher abortion exposure and lower arrests, simply because the high-abortion cohorts will be smaller than low-abortion cohorts.¹²

DL (2001) state they they do not perform a per capita analysis “because of the absence of reliable measures of state population by single year of age” (p. 411). They do not mention that the Census Bureau calculates state-level estimates of population by single year of age (the data begin in 1980). Of course, like any estimates, these population estimates may

¹¹ The estimating equation on the top of p. 45 in DL’s (2004) response to Joyce is:

$$\ln(\text{Crime}_{csy}) = \beta \text{Abort}_{cs} + \theta_c + \gamma_{sa} + \gamma_y + \epsilon_{cst},$$

where c, s, y , and $[a]$, represent cohort of birth, state, year, and age, respectively. In the paper, the “ a ” in square brackets is printed as “ α ,” presumably because of a printing error. There are no state-year controls in this specification (that is, there is no term subscripted by both “ s ” and “ t ” like the θ_{st} term in the equation in Footnote 10).

¹² In the language of the introduction, the test would find a negative relationship between abortion and crime even if only the first channel between them operated (lower numbers of future young persons) and not the second (hypothetical youths, aborted as fetuses, would have been more likely to commit crimes).

be flawed. For our statistical tests, we use a modified population data set constructed by the National Cancer Institute (SEER, 1969–2002). While even these data may not be perfect, we believe that it is inappropriate to simply ignore the population data, given the importance of a per capita analysis to the controversial question at hand.

4. Re-estimating DL’s Test

Correcting DL’s programming error is straightforward, because adding state-year controls requires only that we rewrite a line of computer code for each regression.¹³ Our second modification, performing the analysis on a per capita basis, requires the addition of some population data. The data we use come from the Surveillance, Epidemiology, and End Results program of the National Cancer Institute. This program adjusts Census data to calculate county-level population by single year of age.¹⁴ The SEER data include further breakdowns by racial categories that may be predictive of particular types of cancer. We aggregate across county and racial designations to construct our state-level population estimates by single year of age. We discuss the SEER data more extensively in the Data Appendix.

Our estimates of the effects of abortion exposure and population size on arrests are presented in Table 1 (violent crime arrests) and Table 2 (property crime arrests). The variable to be explained in each regression is either the natural log of the total number of arrests for the age-state-year observation (columns 1–3) or the log of arrests per capita (column 4). All regressions in the tables include age-year controls.¹⁵ As in DL (2001), the regressions include state-age effects where designated (row 1 excludes the state-age effects; row 2 includes them). As for the crucial state-year controls, they are left out of the regressions in column 1 (to facilitate comparison with DL’s papers), then added in columns 2–4.¹⁶ The numbers in parentheses below each estimated effect are the estimated

¹³ All of our data and programs are available on the web site of the Federal Reserve Bank of Boston.

¹⁴ We obtained these data from SEER’s web site: <http://seer.cancer.gov/popdata/>.

¹⁵ In addition, the regressions always include controls for fixed state-level effects. These controls differ from the state-year controls discussed above because they account only for state-level effects that are constant throughout the sample period. For example, entering state fixed controls means that the estimated abortion-crime relationship will not be affected by the fact that crime is *always* higher in some states than others. By contrast, state-year effects sweep out effects that imply some states will have crime rates that are *temporarily* higher than others. Because factors like the crack wave are expected to wax and wane over time, their confounding effects can only be removed by state-year controls, not by fixed state controls. When state-year controls are added to the regression, the fixed state controls become redundant.

¹⁶ The estimates in column 1 of our tables differ slightly from the corresponding regression in Table VII

standard errors. These are inverse indicators of the statistical precision of each estimated effect; small standard errors signal a high degree of confidence that we have estimated the corresponding effect precisely.¹⁷ If the absolute value of the estimated effect is larger than twice the value of the standard error, it is unlikely that the estimated effect is being generated by chance, so the estimate is said to be statistically significant.¹⁸

Results

The wide variety of estimates in Table 1 and Table 2 can seem bewildering at first glance. Depending on the specification, the estimated abortion effect is sometimes significantly negative, sometimes significantly positive, and sometimes very close to zero. This subsection applies some basic tenets of regression analysis to choose our preferred estimates by a process of elimination.

First, *the state-age interactions ought to be included in the analysis; therefore, we should focus on the second row of each table.* Note that the estimated standard errors in the second row are generally smaller than the corresponding errors in the row immediately above.¹⁹ This means that the inclusion of the state-age effects is reducing the overall prediction error of the regression by a great deal. The reduction is large enough, in fact, to shrink the standard errors even though the number of independent variables in the analysis is much larger when state-age controls are added.²⁰ A reduction of the overall error of the variance means that there are several effects on arrests that vary by state and age and that

of DL (2001), because we are using an updated data set made available by DL. Our sample period extends though 1998; theirs extends through 1996.

¹⁷ Standard errors tend to be small when the number of explanatory variables in the analysis (for example, abortion exposure, population, and the various sets of controls) do a good job in explaining the dependent variable (arrests or arrests per capita). Standard errors also tend to be small when there are only a few explanatory variables in the analysis. Adding more explanatory variables to a regression with a given amount of data often means that the effect of each variable will be estimated less precisely. Only if the additional variables greatly reduce the overall prediction error of the analysis will they shrink the standard errors.

¹⁸ To be specific, the estimated effect is statistically significant at the 5 percent level when the ratio is equal to or greater than 1.96. This means that if the relationship were generated by chance, the data would generate a ratio greater than 1.96 only 5 percent of the time. Larger values of the ratio mean that the estimate is significant at higher levels of significance (for example, 1 percent).

¹⁹ The single exception is the first column of Table 1.

²⁰ As noted in Footnote 17, standard errors tend to increase when more variables are added to a regression analysis. The effect of each independent variable is typically estimated less precisely when more independent variables are added.

may be hindering a clean analysis of the abortion effect. Using state-age controls means that we can eliminate the first row of each table and focus on the second.

Second, *we ought to include state-year controls in the analysis, so we should focus on columns 2–4*. The inclusion of the state-year controls is warranted by their large impact on the estimated abortion coefficient. Specifically, the absolute value of the abortion effect declines by more than half for both violent crime and property crime when the state-year controls are entered.²¹ This implies that the regressions without state-year controls are confusing some state-year effects as abortion effects. Entering the state-year controls eliminates this confusion.

Third, *the number of arrests in an observation rises with the number of persons in the observation, so we should focus on columns 3 and 4, where population is included in the analysis*. When entered by itself (column 3), the population variable is always significant at a very high level, confirming the commonsense notion that arrests for a group of young people will rise if the number of people in the group goes up. Columns 3 and 4 include the population data in some way (either as an explanatory variable or in the construction of per capita arrests). Our preferred estimates will be in one of these two columns.

Finally, *the per capita analysis in column 4 is to be preferred to column 3, where population is entered as a separate explanatory variable*. This decision warrants some discussion. After all, the coefficient on log population is only 0.670 in Table 1 and 0.603 in Table 2. If arrests were strictly proportional to population, so that a per capita analysis is appropriate, the population effects would theoretically equal 1.00.²² How can we justify using per capita data when the estimated population effects in column 3 are so low?

The answer lies in DL’s very criticism of the population data, namely, that these data are likely to be measured with error. Standard econometric theory implies that under classical conditions, the estimated effect for a mismeasured variable will be biased, or “attenuated,” toward zero. Thus, if the true coefficient were 1.00, we should *expect* an estimated effect that is smaller than 1.00 if population is mismeasured. Given the likely presence of measurement error, DL’s choice is to ignore the population data and leave them out of the regression. To us, this seems ill-advised. Not only is there undoubtedly

²¹ Specifically, the effect moves from -.0271 to -.0094 in Table 1 and from -.0283 to -.0096 in Table 2.

²² An estimated effect of 1.00 would be expected for population because the data have been transformed into natural logarithms. Mathematically, running the regression in logs performs the analysis in percentage terms. Thus, an estimated effect of 1.00 means that a 1 percent increase in population would raise arrests by 1 percent. This is the definition of strict proportionality.

some relationship between arrests and population, the only way to determine whether the youths who would have developed had they not been aborted as fetuses would have been more likely to commit crimes is to see whether per capita propensities for criminal behavior decline as abortion exposure rises. Moreover, even the mismeasured population data enter very significantly in the regression. The relevant population effects in column 3 are each more than eight times the size of their corresponding standard errors.

The good news is that there is a way to reduce the effect of measurement error on our analysis. Well-known results in econometrics suggest that, again under classical conditions, measurement error does not result in attenuation bias when it appears in the variable to be explained (arrests) rather than in a variable doing the explaining (population).²³ Performing the analysis on a per capita basis puts the measurement error in the least damaging place, because it imports it into the arrests measure. The new variable to be explained will be (the natural log) of arrests divided by population, or arrests per capita. Therefore, beginning with a null hypothesis that arrests should be proportional to population, the per capita results in column 4 will be unbiased tests of the abortion-crime relationship. In our view, this is a better way to proceed than DL's choice to omit the population data altogether.

This process of elimination suggests that we choose the estimates in the second rows, fourth columns, for our preferred estimates. These estimates (-.0002 for violent crime and -.0004 for property crime) are small and statistically indifferent from zero. This is the source of our claim that DL do not provide evidence that legalized abortion has reduced the individual propensity to commit crime.

Robustness checks

We conducted three sets of robustness checks on our regressions. First, we re-estimated all regressions with population using the Census Bureau's original intercensal state-age-year estimates, rather than the data as corrected by SEER.²⁴ This made no difference

²³ The intuition for this fact is technically difficult, but it involves the fact that no regression analysis will explain all of the variation in the dependent variable (here, arrests or arrests per capita). The presence of a general error in the regression will not be a problem as long as the explanatory variables are well-behaved in a statistical sense. When classical measurement error appears in the variable to be explained, the regression folds this error into this general error term, where it does not bias the estimated coefficients. By contrast, when measurement error appears in an explanatory variable, this variable will no longer be statistically well-behaved, and attenuation bias will result.

²⁴ See the Data Appendix for a discussion of the SEER data.

to our results. For example, our preferred estimate of $-.0002$ in the violent crime table becomes $-.0004$ (standard error of $.0031$) when the original Census data are used. The preferred property-crime estimate changes from $-.0004$ to $-.0006$ (standard error of $.0021$). All that these similarities tell us, however, is that the SEER corrections are not driving our results. The similarities cannot tell us whether the original population data are so good as to need few corrections, or so bad as to be beyond hope. Still, the high degree of statistical significance for the population effects in column 3 suggests that some population measure should be incorporated into a study of abortion exposure and arrests. It turns out not to matter which population data are used.

Second, we re-estimated our preferred regressions without the state population weights that both we and DL (2001) employ in our main work. These population *weights* are to be distinguished from our population *effects*, in that the weights do not allow population to help explain arrests. Rather, the weights instruct the regressions to give more credence to data from large states when explaining the total arrests of a state-age-year observation. Dropping the weights results in an estimated abortion effect of $-.00001$ (s.e. $.00436$) for violent crime and a positive but insignificant $.0041$ ($.0038$) for property crime. Hence, dropping the population weights makes our results stronger. Finally, we re-estimated the per capita regressions with population weights but *without* the crucial state-year effects. The results become significantly negative again.²⁵ This suggests that both of our corrections — the state-year effects and the addition of population data — are important to understanding why our results are different than those of DL (2001).

Before concluding, we should note that like DL (2001), we did not perform any regressions for murder arrests. DL state that there are too few observations with non-zero values for murder arrests for such an analysis to be useful.²⁶ We take their point and suggest that the less-powerful tests they do perform on murders are probably influenced by the same factors as tests for property crime and violent crime.

5. Conclusion

Donohue and Levitt's 2001 paper is an excellent illustration of the power of state-level data to test controversial hypotheses. Our comment has criticized DL (2001) on empirical

²⁵ The coefficient in the violent crime regression becomes $-.0102$ ($.0036$), and the property-crime estimate is $-.0115$ ($.0024$).

²⁶ Because the analysis is in logs, and because it is impossible to take a log of zero, an observation with no murders cannot be included in the analysis.

grounds. We argue that controlling for state-year effects is crucial for a study of abortion and crime, as is performing the analysis on a per capita basis. With these modifications, we have shown that DL's own data reveal no evidence for a difference in criminal propensities between actual youths and the hypothetical individuals who would have developed into youths had they not been aborted as fetuses. Thus, any effect of abortion would arise by reducing the total number of young people, not by reducing the number of persons who are most likely to commit crime. We would not want this finding to be taken as evidence for a particular view on the morality of abortion. Rather, we would hope that this comment is viewed, along with DL's pioneering paper, as a useful contribution to social science research.

6. Data Appendix

Virtually all of the data used in this paper come from the the updated data set (*dlab 2.1.dta*) supplied by John Donohue via his Yale University web site. The only new data that we supply are the state-age-year population data, which come from the Surveillance, Epidemiology, and End Results program of the National Cancer Institute (SEER, 2005). We aggregate SEER’s county-, and race-specific data to state-year-age estimates.

The SEER data are based on “postcensal” estimates of population, which are originally constructed by the Census Bureau. To calculate postcensal estimates, the Census takes as a base a full decennial census count (for example, the 1990 Census). Then it uses birth rates, death rates, aging rates, and estimates of interstate migration to estimate state-by-age population for each of the 10 years after the full census. These estimates are of interest to a wide variety of researchers in social science and public health.

The SEER population data differ non-trivially from the original postcensal census estimates for 1990–1999. Due to a discrepancy of 6.8 million persons between the 1990-based postcensal estimate for the year 2000 and the actual 2000 Census, the National Center for Health Statistics (NCHS) chose to revise the numbers for all of the 1990s.²⁷ According to the NCHS, this “error of closure” was unevenly distributed across sex, race, and age groups.²⁸ The revised estimates, known as “bridged intercensal estimates,” attempt to eliminate these errors of closure, so that the intercensal estimates for the year 2000 match up with the full Census 2000 counts.

Ingram, et al. (2003) describe the method for deriving the bridged intercensal estimates. The method is based on distributing the error of closure over the entire decade. For any subgroup, the new population estimate is expressed as a function of time, the 2000 Census figure, and the original postcensal estimate for that group. The mathematical expression is:

$$P_t = Q_t [P_{3653}/Q_{3653}]^{t/3653},$$

where t is the time elapsed since April 1, 1990 (in days), P_t is the bridged intercensal estimate at time t , Q_t is the original postcensal estimate for time t , P_{3653} is the April 1,

²⁷ This paragraph and the next draw heavily from Ingram, et al. (2003).

²⁸ The Census 2000 count exceeded the postcensal numbers by a greater margin for men (3.8 million, or 2.7 percent) than for women (3.0 million, or 2.1 percent). With regards to age data, the census enumeration exceeded the postcensal numbers for all age categories under 85. However, nearly 70 percent of the error of closure was concentrated in ages 5–34.

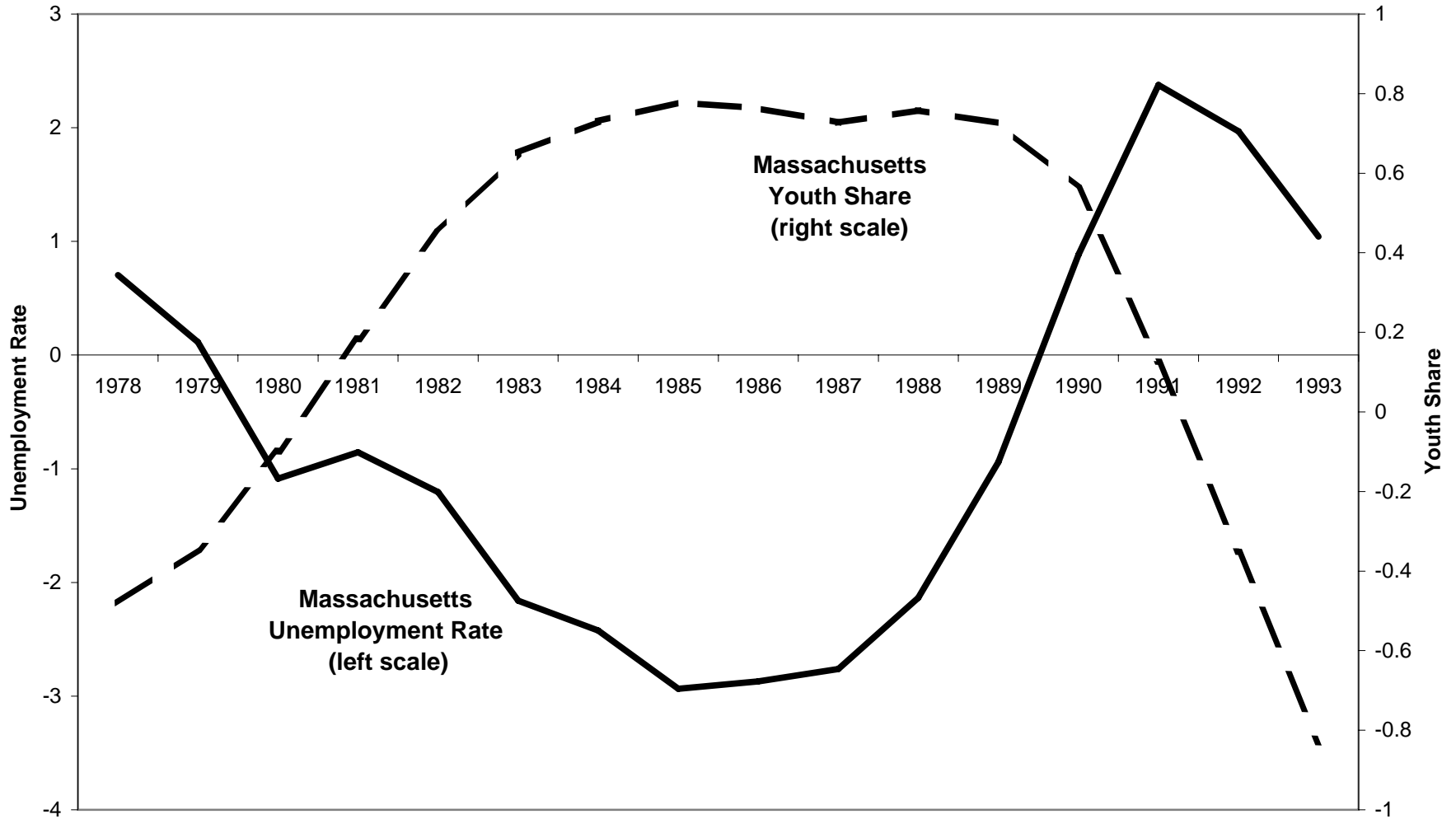
2000, full census count, and Q_{3653} is the original April 1, 2000, postcensal estimate.

The NCHS and Census Bureau, with help from the National Cancer Institute (NCI), used this method to create bridged-race intercensal data for the 1990s (National Center for Health Statistics, 2004). The NCI, authors of the SEER data, made a further modification by re-estimating the population figures for Hawaii, due to concerns that undercounts were particularly large in this state. These revisions make the SEER data for Hawaii slightly different from the bridged intercensal estimates from the Census Bureau and NCHS.

References

- Donohue, John J. III and Steven D. Levitt (2001). "The Impact of Legalized Abortion on Crime," *Quarterly Journal of Economics*, 116:2, pp. 379–420.
- _____ (2004). "Further Evidence that Legalized Abortion Lowered Crime: A Reply to Joyce," *Journal of Human Resources*, 39:1, pp. 29–49.
- Ingram, D.D.; Weed, J.A.; Parker, J.D.; Hamilton, B.; Schenker, N.; Arias, E.; and Madans J.H. (2003). "United States Census 2000 Population with Bridged Race Categories," *Vital Health Statistics*, 2:135. Hyattsville, Maryland: National Center for Health Statistics.
- Joyce, Ted (2004). "Did Legalized Abortion Lower Crime?" *Journal of Human Resources*, 39:1, pp. 1–28.
- Levitt, Steven D. and Stephen Dubner (2005). *Freakonomics: A Rogue Economist Explains the Hidden Side of Everything*, (New York: HarperCollins).
- National Center for Health Statistics (2004). Bridged-race intercensal estimates of the July 1, 1990–July 1, 1999, United States resident population by county, single-year of age, sex, race, and Hispanic origin, prepared by the U.S. Census Bureau with support from the National Cancer Institute, July 26, 2004.
Available at: <http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm>.
- Shimer, Robert (2001). "The Impact of Young Workers on the Aggregate Labor Market," *Quarterly Journal of Economics*, 116:3, pp. 969–1008.
- Surveillance, Epidemiology, and End Results (SEER) Program Populations for 1969–2002, (2005). National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2005.
Available at www.seer.cancer.gov/popdata

Figure 1: Unemployment Rate and Youth Share in Massachusetts Relative to the Nation: 1978-1993
(Percentage Point Differences from National Means)



Note: The youth share is defined as the number of persons aged 16-24 as a percentage of the population aged 16-64 years.

Table 1: Estimated Effects on (Log of) Violent Crime Arrests

	Original DL Specification	Add State-Year Controls	Add Population	Arrests on Per Capita Basis
<i>No State-Age Controls</i>				
Log of Abortion Exposure	-.0184** (.0030)	.0017 (.0050)	.0263** (.0053)	.0255** (.0048)
Log of Population			1.035** (.090)	
<i>With State-Age Controls</i>				
Log of Abortion Exposure	-.0271** (.0044)	-.0094** (.0034)	-.0032 (.0032)	-.0002 (.0033)
Log of Population			0.670** (.080)	

Table 2: Estimated Effects on (Log of) Property Crime Arrests

	Original DL Specification	Add State-Year Controls	Add Population	Arrests on Per Capita Basis
<i>No State-Age Controls</i>				
Log of Abortion Exposure	-.0403** (.0044)	-.0356** (.0049)	.0005 (.0043)	-.0118** (.0039)
Log of Population			1.515** (.065)	
<i>With State-Age Controls</i>				
Log of Abortion Exposure	-.0283** (.0030)	-.0096** (.0021)	-.0040* (.0020)	-.0004 (.0021)
Log of Population			0.603** (.057)	

Notes to Tables 1 and 2: A single asterisk (*) denotes statistical significance at the 5 percent level (two asterisks denote 1 percent significance). Results correspond to regressions of arrest rates for age-state-year observations on abortion exposure, current population, and various controls. Age-year controls are always included; state-age controls are included in the second rows. State-year controls are included in columns 2–4. The sample includes the 50 states plus the District of Columbia over the period 1985–1998 for 15–24 year-olds. The number of observations is 6,724 in the violent crime regressions and 6,730 in the property crime regressions. Data are not available for some states in some years. As in DL (2001), the abortion exposure relevant to the cohort of age a in state s in year y is the number of abortions per 1,000 live births in state s in year $y - a - 1$. The first column replicates the odd-numbered columns of Table VII (DL 2001), using an updated data set from Donohue’s internet site. As in DL (2001), state population weights are always used, and standard errors account for the error correlation for a given cohort in a particular state over time.