

## The Impact of Legalized Abortion on Crime: Comment

Christopher L. Foote and Christopher F. Goetz

### Abstract:

This comment makes three observations about Donohue and Levitt's [2001] paper on abortion and crime. First, there is a coding mistake in the concluding regressions, which identify abortion's effect on crime by comparing the experiences of different age cohorts within the same state and year. Second, correcting this error and using a more appropriate per capita specification for the crime variable generates much weaker results. Third, earlier tests in the paper, which exploit cross-state rather than within-state variation, are not robust to allowing differential state trends based on statewide crime rates that pre-date the period when abortion could have had a causal effect on crime.

**JEL Classifications:** I18, J13

---

Christopher L. Foote is a senior economist and policy advisor at the Federal Reserve Bank of Boston. Christopher F. Goetz is a graduate student at the University of Maryland. At the time the original version of this paper was written, Goetz was a senior research assistant at the Federal Reserve Bank of Boston. Their email addresses are [chris.foote@bos.frb.org](mailto:chris.foote@bos.frb.org) and [goetz@umd.edu](mailto:goetz@umd.edu), 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.

This paper could not have been written without the assistance of John Donohue and Steven Levitt, who made their original data and programs available on the Internet and who supplied us with their new data and programs as soon as they became available. We also thank Ted Joyce, Jeffrey Miron, and John Lott for helpful discussions and for sharing their data as well. Comments from three anonymous referees are also appreciated.

This comment is forthcoming (less the figures and the appendix) in the *Quarterly Journal of Economics* ((123):1. February 2008). It is a revision of the 2005 paper "Testing Economic Hypotheses With State-Level Data: A Comment on Donohue and Levitt [2001]."

**This version: January 31, 2008**

## 1. Introduction

This comment revisits a seminal 2001 paper by Donohue and Levitt (henceforth DL) that linked the startling and unexpected decline in crime during the 1990s to the legalization of abortion some 20 years earlier. DL theorize that abortion reduces crime for two reasons. First, holding the number of pregnancies constant, a higher abortion rate today reduces the number of young people in the future. Because younger people commit more crimes than older people, this “cohort-size” effect should reduce crime if the share of young people in the population declines. Second, because a mother can abort a pregnancy more easily when abortion is legal, a child born after legalization is more likely to be wanted than a child born before legalization. If children who are wanted grow up to commit fewer crimes than unwanted children do, then abortion will bring about an additional “selection” effect that further reduces crime.

The strongest evidence in favor of DL’s hypothesis comes from comparing changes in crime rates across U.S. states. The prevalence of abortion differed markedly across states in the years following abortion’s legalization. In the District of Columbia, New York, and California, more than one-third of pregnancies ended in abortion, on average, from 1970-1984. In North Dakota, Idaho, and Utah, however, abortion was used in less than 10 percent of pregnancies over the same period. In the 1990s, high-abortion states experienced bigger declines in crime than low-abortion states, suggesting that abortion reduces crime. Using regressions that are identified by cross-state comparisons of declines in crime, DL attribute about half of the 1990s crime decline to legalized abortion.

Yet statewide crime rates are influenced by other factors besides abortion. Crime in New York is determined by different factors than crime in Utah, so it should not be surprising that crime in the two states diverges over some period. The best way to isolate the true effect of abortion on crime is to use within-state rather than cross-state comparisons. This is done by comparing cohorts of young people who live in the same state in the same year, but whose mothers had different probabilities of aborting an unwanted pregnancy. In other words, the best way to determine if abortion has a causal effect on crime is to compare two people who are in a similar environment today, but who had differing probabilities of being wanted at birth. The most compelling regressions in DL [2001] were the ones that concluded their paper, because these regressions were designed to implement exactly this type of within-state comparison, for cohorts defined on the state-year-age level. In this comment, we offer two reasons why these regressions were implemented incorrectly.

The first flaw in DL’s concluding regressions is that they are missing a key set of regressors because of a computer coding error. The missing regressors would have absorbed

variation in arrests on the state-year level, insuring that the abortion coefficient was identified using within-state comparisons only. Second, unlike the other tests in their paper, the concluding regressions do not model arrests in *per capita* terms. Instead, the dependent variable is the *total* number of arrests attributed to a particular cohort of young persons. Only by using per capita arrest data, however, can we test whether abortion has a selection effect on crime. In Section 2 of this comment, we run the concluding regressions on a per capita basis with the appropriate regressors, and we find that compelling evidence for a selection effect of abortion on crime vanishes. A reader may ask whether the concluding regressions at least show that abortion reduces crime by reducing the number of young people (the cohort-size channel, as opposed to the selection channel).<sup>1</sup> However, we argue below that the concluding regressions do not even provide this partial kind of evidence, owing to the way in which the abortion variable is defined.

At this point, the corrected concluding regressions appear to contradict the other tests in DL's paper, as the concluding regressions no longer suggest that abortion affects crime, while the other tests do. This brings us back to the reason that the concluding regressions are crucial for DL's argument. These regressions are the only formal tests in the paper that cannot be contaminated by time-varying state-level factors that affect both crime and abortion, such as changing aspects of a state's economic circumstances or social and cultural environment. However, it is reasonable to assume that state-specific factors jointly determine both abortion and crime. In Section 3 of this comment, we show that this is indeed the case. First, we show that state-level abortion and crime rates were strongly correlated before 1985, when it was impossible for abortion to have had a causal effect on crime. We then show that accounting for this correlation has damaging consequences for the abortion coefficient in the cross-state regressions that DL use to quantify abortion's effect on crime. In fact, the abortion coefficients in these cross-state regressions are no longer significantly different from zero when a potential proxy for omitted state-year factors is added. Finally, Section 4 concludes with a test that is robust to many of the econometric issues we discuss throughout this comment. This test also provides no evidence that abortion reduces crime.<sup>2</sup>

---

<sup>1</sup> Indeed, this was our first interpretation of these regressions, as discussed in Foote and Goetz [2005]. The same interpretation is lent to the total-arrests regressions by DL [2006]. But Joyce [2006, footnote 12], questions whether the total-arrests regressions are really estimating a cohort-size effect. His arguments were important in developing the line of reasoning we explore in the next section.

<sup>2</sup> We also include an appendix that addresses some of the claims in DL's formal reply to this comment (DL [2008]).

## 2. Correcting DL’s Concluding Regressions

In DL [2001], the concluding regressions presented in Table VII are defined on the state-year-age level:

$$\ln(\text{ARRESTS}_{sta}) = \beta \text{ABORT}_{sb} + \gamma_{sa} + \lambda_{at} + \theta_{st} + \epsilon_{sta}, \quad (1)$$

where  $s$ ,  $t$ ,  $b$ , and  $a$  denote state, year, birth-year, and single year of age, respectively. The *ABORT* variable is the ratio of abortions per 1,000 live births that is relevant for a given cohort of young people aged 15 through 24, as calculated by the Alan Guttmacher Institute (AGI).<sup>3</sup> The three sets of interactions prevent potentially confounding variation from contaminating the estimate of  $\beta$ , and thereby ensure the cleanest possible estimate of abortion’s effect on crime. The state-age fixed effects ( $\gamma_{sa}$ ) allow each state to have a different age profile for arrests. The age-year fixed effects ( $\lambda_{at}$ ) control for nationwide fluctuations in criminal activity for persons of given ages. Finally, the crucial state-year fixed effects ( $\theta_{st}$ ) absorb all state-level variation in both the time-series and cross-sectional dimensions. As a result, including  $\theta_{st}$  insures that  $\beta$  is identified solely by within-state comparisons of arrests by age group. That is, the effect of abortion is estimated by comparing the criminal propensities of two individuals living in the same state in the same year. These individuals differ only in age, and the regression controls for the usual effect of age on criminality. Therefore, these two individuals differ only in their risk of abortion before birth, and therefore in their risk of having been unwanted children.

DL’s coding error was to omit the state-year interactions ( $\theta_{st}$ ) from these regressions. These regressors are especially important because earlier tests in the paper are identified solely by cross-state variation of changes in statewide crime rates. Omission of  $\theta_{st}$  leaves the concluding regressions vulnerable to the same type of state-level omitted variables bias as the paper’s earlier tests.

A second problem is the specification of the *ARRESTS* variable. To test whether abortion has a selection effect, one needs to know whether a person exposed to a high abortion risk *in utero* is less likely to commit a crime. By “less likely,” we mean a lower probability, but the only way to measure a probability is to divide the number of crimes by the number of people who could commit them. In other words, *ARRESTS* must be in per capita terms. DL [2001], however, defines *ARRESTS* as the total number of arrests

---

<sup>3</sup> For example, because 15-year-olds in 1995 were generally conceived in 1979 (= 1995-15-1), the *ABORT* variable for Massachusetts 15-year-olds in 1995 is abortions per 1,000 births in Massachusetts in 1979. In order to line up abortions with future births that are conceived at the same time, AGI measures the number of abortions in a calendar year, divided by the total number of births from July 1 of that year to June 30 of the following year.

for the cohort, “because of the absence of reliable measures of state population by single year of age” (p. 411). In fact, the Census Bureau constructs these population measures for each year beginning in 1980.

While no population estimates are perfect, we believe that estimating the arrests regressions in per capita form is vital, because it is the only way to test for the controversial selection effect of abortion on crime. In fact, it is hard to know what one is estimating when the *ARRESTS* variable is not in per capita form. Recall that  $ABORT_{sb}$  for a cohort that is  $a$  years old is simply the number of abortions over births in its birth year  $b$ . Ignoring the various fixed effects from Equation (1) and noting that  $b = t - a$ , we can write

$$\ln(ARRESTS_{sta}) = \beta \left( \frac{Abortions_{s,t-a}}{Births_{s,t-a}} \right) + \epsilon_{sta}.$$

It is easy to see how  $\beta$  could be negative in this regression, even if abortion has neither a selection nor a cohort-size effect on crime. Yearly fluctuations in births are caused by many factors, with the perceived costs of abortion being only one example. Variation in the other factors determining births will generate movements in the abortion-births ratio that are negatively related to total arrests in a mechanical way. Specifically, abstracting from migration and deaths, an increase in births  $a$  years ago shows up as an increase in the number of people that are  $a$  years old this year. But more people in a cohort is likely to mean more arrests, simply because the cohort is larger. So an increase in births *reduces* the abortion-births ratio, while it *increases* the total number of arrests, via an increase in population. Therefore, arrests and the abortion-births ratio should be negatively related in a total-arrests regression, even if no selection or cohort-size effects exist. A true test of the cohort-size effect would regress the number of births on a discrete indicator of the availability of abortion, then trace out the implications of any decline in birth rates for the nation’s per capita crime rate. It would not regress total arrest counts on abortions *divided by* births.

*Results using original abortion data*

Table I revisits the regressions from Table VII in DL [2001]. The first four columns use the same data DL used, over the same sample period (1985-1996). Panel A presents the results for property crime (the most common type of crime) and Panel B presents results for violent crime. Each of the regressions includes state-age and age-year interactions. Before discussing our main results, we must say a word about the standard errors. We report two sets of standard errors, distinguished by how they are “clustered,” or the extent to which individual residuals are assumed to be independent of one another. DL’s original paper

employs standard errors that are clustered by year-of-birth and state, because the same groups of people are observed at different ages in different years.<sup>4</sup> Since DL published their 2001 paper, however, applied econometricians have begun to worry more about residual-independence assumptions. In this case, as stressed by Joyce [forthcoming], there may be a correlation between the error for, say, 17-year-olds in one year and other age groups (besides 18-year-olds) in the following year, even after entering all the fixed effects. The standard fix for this problem is to cluster the standard errors more widely (Bertrand, Duflo, and Mullainathan [2004]). In the second set of standard errors below, we cluster the standard errors by state.

Consider first the parameter estimates in column 1, which mimics DL’s original specification exactly. We are able to replicate their results for both the abortion coefficients ( $-.025$  for property crime and  $-.028$  for violent crime) and the original standard errors (0.003 and 0.004). The state-clustered standard errors in column 1 are larger than the original ones, suggesting that this specification leaves a great deal of within-state correlation in the residuals. Column 2 corrects DL’s computing error by adding the state-year fixed effects ( $\theta_{st}$ ). Both abortion coefficients drop by more than half. Column 3 adds the population data to the analysis by entering the log of the cohort size as a right-hand-side variable.<sup>5</sup> The log of population enters significantly in both regressions, but the estimated coefficient is less than one, which suggests that arrests and population do not vary proportionately. One possible explanation for this finding is that youths from large cohorts are generally better behaved than youths from small cohorts. A far more likely explanation is that (as DL pointed out) population is measured with error. If so, then well known econometric results predict that the population coefficient will be biased (“attenuated”) towards zero.

Rather than omit the population data, a better choice is to eliminate the attenuation bias by moving the offending variable to the left-hand-side of the regression, transforming the dependent variable from total arrests into arrests per capita. Econometric theory suggests that classical measurement error does not cause bias if it appears on the left-hand-side of the regression. Of course, in our case, the move has the added benefit of permitting us to test if abortion has a selection effect on crime. Column 4 shows that the absolute values of the abortion coefficients fall to essentially zero when this is done.

---

<sup>4</sup> Ignoring migration and deaths, the persons making up the observation for Massachusetts 16-year-olds in 1990 also make up the observation for Massachusetts 17-year-olds in 1991. Residuals from these two observations will not be independent, because they will share any unobserved factors conducive to crime.

<sup>5</sup> We use modified population data constructed by the National Cancer Institute, which is available from 1969 to 2002. Using the unadjusted Census data gave essentially the same results. See SEER [2005] and Ingram et al. [2003] for a discussion of the population data we use.

### *Results using adjusted abortion data: DL [2006]*

After the original version of this comment was released in 2005, DL responded that the corrected regressions do not argue strongly against an abortion-crime link [DL 2006]. Their main concern is measurement error in the abortion data, which arises from three sources. First, the original AGI data measured abortions by the place of occurrence, not the woman's state of residence. Second, because we do not know the due date of the fetus nor the day of the year on which the abortion occurred, we do not know the precise year in which an aborted fetus would have been born. Third, interstate migration means that the abortion exposure relevant for a young person may be a lagged abortion rate in some other state, where he was born. As with the population data, measurement error in the abortion data will bias the abortion coefficients toward zero. DL [2006] re-runs the concluding regressions with abortion data that has been adjusted to address these three issues. That paper's abstract states that "[w]hen one uses a more carefully constructed measure of abortion (e.g., one that takes into account cross-state mobility, or doing a better job of matching dates of birth to abortion exposure), . . . the evidence in support of the abortion-crime hypothesis is as strong or stronger than suggested in our original work."

To evaluate this claim, column 5 of Table I uses the adjusted abortion data from DL [2006]. The sample period is extended by two years, to 1998. As DL also found, the point estimate for abortion's effect on property crime becomes slightly *positive* (.001), though it remains insignificant. The coefficient from the violent-crime regressions moves to -.021, with its statistical significance dependent on the way in which the standard errors are calculated. DL [2006] calculates these errors in the same way as DL [2001], clustering by year-of-birth and state. Using what we would argue is a more appropriate method increases the standard error by about 75 percent, resulting in a t-statistic of only 1.5.<sup>6</sup>

### **3. Reconciling Within-State and Cross-State Results**

The results of equation (1) do not provide evidence for a link between abortion and crime. However, as noted in the introduction, DL [2001] contains other tests besides the

---

<sup>6</sup> DL [2006] also claims that evidence for a selection effect can be resurrected by using a separate measure of abortion, provided by the Centers for Disease Control (CDC). This measure of abortion is an occurrence-based indicator, but DL correct it for interstate migration using the same correction as that used for the residence-based AGI data. This transformed version of the CDC data is then used as an instrument for the AGI measure. Joyce [forthcoming] provides a detailed argument for why this IV is not appropriate. In any case, it has a small effect on the estimates, moving the property-crime coefficient from 0.001 to -0.013 and the violent-crime coefficient from -0.021 to -0.023. Neither IV estimate is statistically significant, no matter how the standard errors are clustered.

concluding regressions. We now illustrate why we should be skeptical of these other tests, which are based on cross-state comparisons, not within-state comparisons.

Consider the regression used in DL [2001] to quantify the effect of abortion on crime. That regression uses observations defined on the state-year level:

$$Crime_{st} = \delta EAR_{st} + \text{other variables} + \phi_s + \phi_t + \varepsilon_{st}. \quad (2)$$

Here,  $Crime_{st}$  is the log per capita crime rate of state  $s$  in year  $t$ , where crime is defined by the number of crimes reported to police, not actual arrests. The  $\phi_s$  and  $\phi_t$  controls are state and year fixed effects, and  $EAR_{st}$  is the state’s “effective abortion rate.” This rate is constructed by weighting abortion rates  $a - 1$  years ago by the fraction of crimes typically committed by persons of age  $a$ .<sup>7</sup> The “other variables” are time-varying, state-level factors such as incarceration rates, per-capita income levels, and gun laws. These controls are potentially important, because many factors affect state-level crime rates besides past abortions. Unlike the concluding regressions, which are designed to *eliminate* cross-state variation via the (inadvertently omitted)  $\theta_{st}$  terms, the regression above is *identified* by cross-state comparisons of changes in crime rates. Of course, we are worried about state-year level factors that are omitted from the equation, not the state-year variables that are included. (The included variables turn out to have little effect on the abortion coefficient.)

In these regressions, the estimate of  $\delta$  in equation (2) will commingle both cohort-size effects and selection effects of abortion. In their regressions, DL [2001] estimated that increasing the abortion ratio by 100 abortions per thousand births reduces per capita crime in a state by about 10 percent. Based on these estimates, DL [2001] surmised that abortion’s selection effect is large, calculating that “those on the margin for being aborted are roughly four times more criminal” than the average 18-24 year old (p. 405).

One potential reason why the cross-state regressions imply evidence for a selection effect — while the concluding within-state regressions do not — is that omitted variables bias remains a problem in the cross-state regressions, despite DL’s attempts to control for it. A good way to determine whether omitted variables bias is possible is to look for a correlation between state-level abortion and crime rates before the mid-1980s, when the

---

<sup>7</sup> Formally, the EAR is

$$EAR_{st} = \sum_a Abortion\ Ratio_{t-a-1} \left( \frac{Arrests_a}{Arrests_{total}} \right),$$

The abortion ratio is abortions over births, as in the state-year-age regressions. The extra “-1” in the subscript for the abortion ratio indicates that the relevant abortion rate of a child born in a given year is the previous year’s abortion rate, because pregnancies last for most of one year.

first cohorts affected by legalized abortion reached adolescence. DL [2001] recognize the potential for concern, stating: “There should be no effect of abortion on crime between 1973-1985. To the extent that high and low abortion states systematically differ in the earlier period, questions about the exogeneity of the abortion rate are raised” (p. 401).

We looked for a pre-1985 relationship between abortion and crime by calculating average abortion and crime rates for each state in the 1970-84 period, then regressing the crime averages on the abortion averages. Figure 1 gives a visual sense of these data. The two panels in this figure plot state-level, 1970-84 averages of per capita property crime (top panel) and violent crime (bottom panel) against average abortion ratios, with pre-legalization abortion ratios set to zero. Both plots indicate that states with high abortion ratios also had high crime rates during this early period. Coefficients from formal regressions of abortion averages on crime averages are positive and highly significant, with large amounts of the variation in crime “explained” by abortion. For example, a population-weighted regression of average property-crime rates on average abortion ratios gives an  $R^2$  of .37 and a p-value for the abortion coefficient of 0.0012. The results for violent crime are even stronger: the  $R^2$  from this regression is .62 and the abortion p-value is zero to four decimal points. When the data are unweighted, the  $R^2$ s for property crime and violent crime are .41 and .67, respectively.<sup>8</sup>

DL also examine the relationship between abortion and crime before 1985, but come to the opposite conclusion: “It is reassuring that the data reveal no clear differences in crime rates across states between 1973 and 1985 as a function of the abortion rate” (p. 401). Our interpretations differ because DL look for a uniform pattern in pre-1985 *changes* in crime rates as a function of the abortion rate, while we focus on the average *levels* of crime and abortion in the early period.<sup>9</sup>

We believe that DL misread the data by focusing on changes rather than levels, because state-level factors that drive crime and abortion may not have constant effects over time. For example, there may be some reason that New York had both a higher crime rate and a higher abortion rate than Utah had before 1985. Perhaps New York’s urban density, its wealth, its demographic structure or some other aspects of its culture offers New Yorkers more chances for interpersonal connections that lead to more crimes and to more unwanted pregnancies. Now consider what would happen if the influence that these factors had on

---

<sup>8</sup> Dropping DC from the unweighted regression generates  $R^2$ s of .36 and .42. The full set of regression statistics appears in Appendix Table II.

<sup>9</sup> DL find sharp differences between changes in crime across states with high and low abortion rates, but they dismiss their importance because these changes are not uniform across different types of crime.

crime were to diminish over time. States with high abortion rates in the past would see their crime rates fall the most, because their high abortion rates would proxy for those factors that are losing their significance in driving crime. If so, DL’s regressions would detect the steeper decline in crime in high-abortion states and erroneously assign the credit for this drop to past abortions, when changes in other, unobserved factors are in fact responsible.

The problem becomes even more pernicious when we note that the standard fix for unobserved factors in state-level regressions can make the problem worse. In a crucial robustness check, DL enter interactions between the yearly dummies and dummies for the country’s nine Census divisions. These division-year interactions will allow an unbiased estimate of the abortion coefficient as long as all the potentially confounding factors are determined on a geographic basis, operating *between* and not *within* Census divisions. But the confounding factors that jointly determine abortion and crime turn out to be even more important in driving crime when we compare states within the same Census division. We discovered this by first regressing the early state-level abortion and crime averages on a slate of Census division dummies, then regressing the resulting crime residuals on the resulting abortion residuals. The residuals used in these regressions are graphed in the two panels of Figure 2. In both panels, the positive relationship between early-period abortion and crime appears even stronger than in the unadjusted data from Figure 1. This impression is confirmed by population-weighted regressions that imply abortion “explains” 60 percent of the within-division variation in property crime and an astonishing 72 percent of within-division variation in violent crime in the 1970-84 period.<sup>10</sup>

All in all, omitted variables bias is likely to be a serious problem in DL’s cross-state regressions. The early correlations between abortion and crime suggest that these variables are driven by common factors, because abortion can have no causal effect on crime before 1985. Cross-state estimates of the effect of abortion on crime are invalid if the effect of these driving forces on crime changes over time. If we try to account for these complicating factors by using geographic interactions, we can turn a bad regression into a worse one, because the remaining within-division variation in abortion rates is even more tightly linked to remaining variation in past crime rates. Without some direct control for omitted factors that drive this correlation, entering division-year dummies in these regressions will exacerbate, not solve, the omitted variables problem.

---

<sup>10</sup> As we report in Appendix Table II, the  $R^2$ s from the within-division property- and violent-crime regressions without population weights are .59 and .78, respectively. Dropping DC from the unweighted regressions generates  $R^2$ s of .58 and .54.

### *Revisiting DL's cross-state regressions*

To see if this criticism of DL's cross-state regressions is empirically relevant, Table II presents cross-state regressions with some new data and new specifications.<sup>11</sup> Column 1 uses DL's original specification and original abortion data.<sup>12</sup> Column 2 employs the same specification, but uses the new residence-based abortion data to construct the effective abortion rates. Column 3 updates the sample to end in 2003 rather than 1997. All of these regressions generate significantly negative abortion coefficients.

As an ostensible control for potentially omitted variables, we enter the division-year interactions in the regressions of column 4. As DL found in their original paper, including these controls does not materially affect the abortion coefficients. Recall, however, that the effect of omitted variables may be worse when using within-division variation alone to identify abortion's effect. We cannot eliminate this bias without a model that identifies the omitted variables. Yet we can at least reduce the bias by entering an appropriate proxy. This proxy must be correlated with factors that caused crime in the past, but whose intensity changed after 1985.

Accordingly, in column 5, we enter an interaction between the mean of the state's log per capita crime rate from 1970-84 and a linear trend. A negative coefficient on this variable indicates that states with relatively high crime rates from 1970 to 1984 experience relatively steeper crime declines after 1985. Two points of discussion about this variable are important. First, entering this crime-trend interaction requires an estimate of only one additional coefficient, so it is far more parsimonious than entering 51 unrestricted state-specific trends. (In a robustness check, DL show that there is not enough variation in the data to estimate separate state-specific trends.) Second, because this proxy is not perfect, it will not eliminate omitted variables bias completely. Nevertheless, the evidence provided by the cross-state regressions will be much less convincing if adding the proxy reduces the importance of the abortion coefficient.

---

<sup>11</sup> As in DL [2001], we use a Prais-Winsten method to account for first-order serial correlation in the residuals. Unlike DL's regressions, our regressions cluster the standard errors by state. We found that state-clustered standard errors were larger than simple White-type robust standard errors of DL [2001], probably because the AR(1) corrections do not purge the regressions of all serial correlation in the residuals (Bertrand, Duflo, and Mullainathan [2004]).

<sup>12</sup> The regressions in column 1 include only the effective abortion rates and the state and year fixed effects, so they are comparable to the regressions in columns 1, 3, and 5 of Table IV in DL [2001]. Our estimates are marginally different than those in DL [2001] for four reasons: We used a slightly different methodology for calculating the first-order autocorrelation parameter, we allowed the Prais-Winsten procedure to iterate on this parameter, we used constant population weights within each state, and we used updated estimates of crime and population. The corresponding point estimates in DL [2001, p. 404] are -.095 for property crime, -.137 for violent crime, and -.108 for murder.

Column 5 shows that the abortion coefficients weaken sharply when this proxy is added. The effective abortion coefficient drops by about 77 percent in absolute value in the property crime regression (from -.131 to -.030), by about 52 percent in the violent crime regression, and by about 42 percent in the murder regression. None of the abortion coefficients are statistically significant. By contrast, the coefficients on the new interaction terms are strongly significant for both property and violent crime. The coefficient for murder, while not significant, is about the same size as that for violent crime in general.<sup>13</sup> All told, our results suggest that the estimated abortion effect in cross-state regressions is sensitive to controls for omitted variables. It is therefore crucial to absorb potential omitted variables bias on the state-year level, using controls like the state-year interactions ( $\theta_{st}$ ) that can be included in the concluding state-year-age regressions.<sup>14</sup>

#### 4. Conclusion

DL [2001] suggests alternative ways of studying the abortion-crime relationship, but different methods give different answers. Their concluding state-year-age regressions, when run correctly, provide little evidence for a selection effect of abortion. Their cross-state regressions, by contrast, imply a large selection effect. Each method has its drawbacks: DL contend that measurement error plagues the concluding regressions, whereas we argue that the cross-state results are not robust to controls for omitted variables.

Fortunately, there is a way to test the abortion-crime hypothesis that simultaneously addresses DL’s concerns about measurement error and our worries about omitted state-year factors. By collapsing the state-year-age data into nationwide age-year means, we can then regress each national cohort’s per capita criminal propensity on the appropriate national abortion rate, along with age and year fixed effects. Using the same notation as before, the equation is:

$$\ln(\text{ARRESTS PER CAPITA}_{ta}) = \beta \text{ABORT}_b + \phi_a + \phi_t + \epsilon_{ta}.$$

---

<sup>13</sup> Robustness checks for these regressions appear in the appendix.

<sup>14</sup> In addition to the cross-state regressions, DL [2001] includes two other cross-state tests. One test argues that five states that legalized abortion in 1970 saw crime decline sooner than the rest of the country, which legalized three years later. But these early-legalizers also tend to be high-crime states, so this test involves both the timing of crime declines and the amount by which crime fell in each state. As DL point out, the source of identification is not independent from the cross-state regressions. The other test uses arrests data to calculate a per capita arrest rate for those over 25 and under 25 in each state. Then the difference between the two rates is regressed on state-level EARs. This method would appear to be an improvement over the other cross-state regressions (and the implied effects of abortion are indeed smaller). Yet the data used are arrests, not crimes reported, so it is hard to see how this method is an improvement over linking arrests to population and abortion exposure by single year of age, as is done in the concluding regressions.

This age-year regression has a number of advantages. It gives a direct test of the controversial selection effect of abortion, because it is run with per capita data on well-defined age cohorts. The use of nationwide data also obviates DL's measurement-error concerns, which are caused by the difficulty of measuring arrests, abortion ratios, population, and migration on the state level. And because the identifying variation in this regression is national, it is not biased by the omission of any state-level factors.

As pointed out by numerous previous authors [Sailer 1999; Lott and Whitley 2007; Joyce 2004, 2006], using age-year variation generates no support for the abortion-crime hypothesis. To illustrate this point, these authors often use graphs like Figures 3a and 3b, which show that the criminal activity of different age cohorts does not appear to decline when these cohorts begin to be affected by legalized abortion.<sup>15</sup> Table III formalizes this point with regressions of national age-specific arrest rates on a nationwide version of the abortion ratio. Column 1 shows that for both property and violent crime, abortion exposure appears to raise criminality, not lower it.

DL [2001] points out a drawback to this approach. Just as the cross-state regressions can be contaminated by omitted state-year effects, the regressions in Table III are susceptible to omitted age-year effects. If there is some shock that raises criminality for cohorts with relatively high abortion exposures, then the age-year tests will be misleading. Footnote 21 of DL [2001] points out that the crack wave of the late 1980s and early 1990s may have delivered these shocks.<sup>16</sup> In column 2 of Table III, we exclude data from the zenith of the crack wave (1985-1992). Contrary to what DL's theory would imply, the coefficients from age-year regressions become slightly more positive. Placing these results alongside those from the corrected concluding regressions and our expanded cross-state analysis, we find no compelling evidence that abortion has a selection effect on crime.

---

<sup>15</sup> For example, Figure 3a shows that the property-crime rate of 21-24 year-olds begins to decline around 1989, while that of 15-17 year-olds keeps rising. But if abortion were truly affecting crime rates, we would expect the crime rate of the younger cohort to fall relative to that of the older cohort, because the younger cohort begins to be affected by legalized abortion at about this time.

<sup>16</sup> For the crack wave to contaminate the age-year tests, however, it is not enough for crack to raise criminality for all 15-24 year-olds in some years, because the year dummies account for shared influences of this type. Crack must raise criminality for various years and ages in ways that coincidentally line up with cohort-specific abortion exposure and mask the large selection effects of abortion from showing up in age-specific arrests.

## 5. Appendix

In this appendix, we respond to various points made in DL [2008], the formal reply to our comment published in the February 2008 issue of the *Quarterly Journal of Economics*.

### *State-year-age regressions of Table I*

One area of contention between DL and us is how to calculate the standard errors in the state-year-age regressions. In their reply, DL write that our preferred method, clustering by state, may exaggerate the size of the standard errors. The implication is that state-clustering may incorrectly reduce the abortion coefficients' t-statistics to insignificance, when in fact the coefficients are significantly different from zero. This dispute hinges on how pervasive the error-correlation patterns in state-year-age data are likely to be. DL recognize that there is likely to be a correlation among residuals corresponding to the same groups of people over time. This correlation is accounted for by clustering the standard errors by *year-of-birth*  $\times$  *state*, as in DL's original paper [2001]. But other correlation patterns are also possible. Joyce [forthcoming] highlights the potential for serial correlation, which will arise if the errors for a given age group in a given state are correlated over time.<sup>17</sup> Serial correlation requires clustering by *state*  $\times$  *age*. Moreover, while we have data for ten separate age groups at the state-year level, we may not obtain truly independent variation from all ten of these groups in every state and year. For example, the factors determining crime for a state's 15- and 16-year-olds might be very similar, so that the variation supplied by these two groups is essentially identical. Clustering by *state*  $\times$  *year* accounts for this possibility.<sup>18</sup> The advantage of clustering by state is that all of these patterns (*year-of-birth*  $\times$  *state*, *state*  $\times$  *age*, and *state*  $\times$  *year*) are accounted for at the same time. In fact, state-clustering accounts for *any* correlation pattern among residuals from the same state.

Appendix Table I shows that correlation patterns are widespread in the state-year-age regressions, so clustering by state is warranted. The two regressions presented in this table use per capita arrests data, all fixed effects, and DL's updated abortion data. They are therefore identical to the last column of our Table I, but are rounded to four decimal points rather than three. The first standard errors in the table (.0035 and .0046) are the Huber-White "robust" errors. These errors do not account for any correlation patterns among

---

<sup>17</sup> Serial correlation would arise if (say) the error for Massachusetts (MA) 17-year-olds in 1991 is correlated with the error for MA 17-year-olds in 1992. These errors are generated by different groups of people, but the factors that determine crime for MA 17-year-olds may move slowly over time.

<sup>18</sup> Correlation that is shared across all ten age groups within a state and year is not a problem for the regression, as long as the state-year dummies are included. Only correlation that is present across some age groups but not others requires state-year clustering, since this type of correlation will remain even after the state-year dummies absorb variation that is common to all age groups in the given state and year.

residuals. However, they do allow individual residuals to be heteroskedastic, so they can serve as useful reference points for what follows. The next errors (.0051 and .0082) are clustered by year-of-birth  $\times$  state. The increase in these errors relative to the Huber-White errors indicates that the correlation pattern captured by this method is likely to be important, as DL correctly foresaw by calculating their original standard errors in this way. The other two rows cluster by state  $\times$  age and state  $\times$  year, respectively. These errors are also larger than the Huber-White errors; in fact, they are generally comparable to the errors that cluster by year-of-birth  $\times$  state. Finally, the state-clustered errors presented in the last row (.0082 and .0139) account for all state-specific patterns of correlation, including the three examples above. As we note in Section 2, the violent crime coefficient is no longer significantly different from zero when state-clustered errors are used.<sup>19</sup>

At this point, it is useful to make two remarks. First, the statistical theory on which the cluster method is based requires a large number of clusters in order to generate appropriate standard errors. When clustering along narrowly-defined criteria (like year-of-birth  $\times$  state), the number of clusters is often large. But when the presence of multiple correlation patterns forces us to cluster along more widely-defined criteria (like state), then the number of clusters is reduced, and we run the risk of having too few clusters.<sup>20</sup> Papers by MacKinnon and White (1985) and Bell and McCaffrey (2002) point out one consequence of having too few clusters — the resulting standard errors are likely to be too small. Using too few clusters would therefore cause us to reject a null hypothesis of “no effect” more often than warranted, given the nominal size of the statistical test. Note that this consequence of too few clusters goes in DL’s favor, because it would cause us to accept their claim of an abortion-crime link more often than we should. A second consequence of using too few clusters, pointed out by Hansen (forthcoming), is that the variance of the estimated standard errors increases. This second consequence means that DL are technically correct to claim that using the state cluster may exaggerate the size of the standard errors. But it is also correct to state that using too few clusters may *underestimate* the size of the standard errors, especially in light of the small sample bias discussed in MacKinnon and White (1985) and Bell and McCaffrey (2002). The most relevant question for our purposes is whether the use of state-clustered errors delivers significance tests of appropriate size.

---

<sup>19</sup> It is possible to construct a multi-way cluster estimator that addresses only the three correlation patterns discussed in this paragraph, leaving other state-level patterns unaccounted for (see Cameron, Gelbach, and Miller [2006] and Thompson [2006]). This method generates standard errors of .0076 for the property-crime regression and .0111 for the violent-crime regression.

<sup>20</sup> When clustering by year-of-birth  $\times$  state, for example, there are 1160 useable clusters in the state-year-age data. When clustering by state, there are 51 clusters, one for each state.

Simulations in Kézdi (2004) and Bertrand, Duflo and Mullainathan (2004) suggest that using 51 clusters is in fact likely to do so.<sup>21</sup>

A second remark is that we can use a Prais-Winsten AR1 correction in an attempt to purge one of the correlation patterns (serial correlation) from the data. If the serial correlation is in fact AR1, then this correction will deliver more efficient estimates.<sup>22</sup> This correction turns out to have minor effects on the coefficients. The AR1-corrected property-crime coefficient is .0028 (with a state-clustered standard error of .0067) while the violent-crime coefficient is  $-.0199$  (.0121).<sup>23</sup>

### *State-year regressions of Table II*

State-level abortion and crime averages were correlated before legalized abortion could have had a causal effect on crime. This suggests that cross-state tests of the abortion-crime hypothesis might suffer from omitted variables bias. We discuss these early-period correlations in Section 3, and Figures 1 and 2 present scatterplots of the data. In Appendix Table II, we present some regression statistics that formally establish these correlations. In columns 1 and 2 of this table, the regression is

$$\overline{Crime}_s^{1970-84} = \beta \overline{Abort}_s^{1970-84} + \epsilon_s,$$

where the dependent variable is the average per capita crime rate for state  $s$  from 1970 to 1984, and the regressor is the average number of abortions per 1,000 births during the same period. These two columns show the strong positive relationship between average

---

<sup>21</sup> The statistical theory justifying the use of the cluster method has traditionally been based on assuming that the number of clusters (for example, cross-sectional units) goes to infinity while the number of observations in the other dimension (for example, time periods) is fixed. Hansen (forthcoming) investigates the case where the number of clusters ( $N$ ) is fixed while the other dimension ( $T$ ) goes to infinity. In this situation, he suggests that the clustered covariance matrix be normalized by  $\frac{N}{N-1}$ , which is quite close to the  $\frac{NT-1}{NT-k} \frac{N}{N-1}$  normalization in the STATA software package, which we use for this paper.

<sup>22</sup> The state-year regressions in DL [2001] and in our Table II also use AR1 corrections. In state-year-age data, the AR1 correction quasi-differences the observations corresponding to the same age group and state in adjacent years.

<sup>23</sup> The estimated AR1 parameter for the property-crime regression is .59 and that for the violent-crime regression is .32. Nickell [1981] points out that estimated AR1 parameters are biased down in panel data when the number of time periods is small. Hansen (2007) provides a bias correction, but we were unsure of how to apply this correction in population-weighted data. Using the tables in Solon (1984) to produce back-of-the-envelope corrections when the time periods number about 15 moves the property-crime AR1 coefficient from .59 to .73 and the violent-crime AR1 coefficient from .32 to .45. Using these corrected parameters generates abortion coefficients (and state-clustered standard errors) of .0028 (.0065) in the property-crime regression and  $-.0181$  (.0110) in the violent-crime regression. Finally, it should be noted all of our Prais-Winsten regressions use population weights that are constant within each state. This has trivial effects on the coefficients.

abortion and crime levels that is depicted visually in Figure 1. In columns 3 and 4, the regression is

$$\widetilde{Crime}_s^{1970-84} = \beta \widetilde{Abort}_s^{1970-84} + \varepsilon_s,$$

where we have now pre-whitened the abortion and crime averages by regressing them against a slate of Census division dummies. The larger  $R^2$ s in these last two columns show that the positive relationship between abortion and crime becomes even stronger when we focus the comparison on states within the same Census division, as was seen by comparing Figure 2 to Figure 1.<sup>24</sup>

These stronger correlations suggest that not all of the potentially confounding effects in state-level abortion-crime regressions operate between different geographic areas of the country. Some effects operate within geographic areas. The augmented regressions in the last column of our Table II control for confounding within-area effects by including interactions between each state's 1970-1984 crime average and a linear trend. Influences that operate between areas are held constant by the division-year interactions. These augmented state-year regressions imply that the true impact of abortion on crime is likely to be much smaller than the impact implied by state-year regressions that have fewer controls for omitted variables bias.

DL's reply to our comment offers some criticisms of these regressions. First, they claim that our results are unduly dependent on the District of Columbia, where the quality of the abortion data is poor. Appendix Table III shows that the augmented regressions are, in fact, quite robust to the omission of various states, including DC. All of the regressions in this appendix table include division-year interactions. Column 2 adds the crime-trend interactions. Hence, the first two columns of this appendix table replicate the last two columns of Table II. They also reiterate its main lesson: The importance of abortion variables in state-year crime regressions is sharply reduced when the crime-trend interactions are included. The next two columns of the appendix table show that the same lesson emerges when DC is omitted from the sample.<sup>25</sup> The last two columns of the appendix table repeat the exercise while omitting DC, NY, and CA, with similar results.

DL's second criticism of our augmented state-year regressions is that they include division-year interactions. They point out that dropping these interactions restores the

---

<sup>24</sup> For example, the  $R^2$  for the population-weighted property-crime regression using the raw averages is .37 (column 1 of first panel of Appendix Table II), while that for the corresponding within-division regression is .60 (column 3 of first panel).

<sup>25</sup> The replication of this pattern is clearer in our Appendix Table III than it is in DL's reply, because DL do not report the coefficients on the crime-trend interactions. Also, DL do not report the regression when DC is omitted from the sample and the crime-trend interactions are not included.

significance of the abortion coefficients. It is true that if shared geographic factors are not important determinants of state-level crime rates, then the division-year interactions should not be included in the regressions. Additionally, even if geographic factors are important, interacting 8 (=9-1) Census division dummies with each of the yearly dummies generates a lot of new regressors. Including them all could rob the regressions of useful variation and inflate the standard errors.

We chose to use division-year interactions as geographic controls because they were also included in a particular specification in DL's original paper. But more parsimonious geographic controls also undermine support for an abortion-crime link. To show this, we interact the region or division dummies with trends rather than yearly dummies. Interacting the Census dummies with one linear trend term, or two quadratic trend terms, generates far fewer additional regressors than interacting these dummies with each of the yearly dummies. Just as importantly, using trends rather than yearly interactions allows us to test for geographic influences on crime with formal F-tests, even when the standard errors are clustered by state.<sup>26</sup> If the inclusion of the trends is supported by these tests, then some type of geographic controls should be included in the state-year crime regressions, regardless of what happens to the estimated abortion coefficients.

Results appear in Appendix Table IV. Column 1 enters the abortion variable by itself, while column 2 adds the crime-trend interaction. Neither of these columns includes any geographic controls, so the abortion coefficients in column 2 match the coefficients in column 4 of Table III in DL's reply [2008]. Columns 3 and 4 add regional trends. For both property crime and violent crime, the regional trends enter significantly and cause the abortion coefficients to lose statistical significance. In the murder regressions, regional trends reduce the absolute value of the abortion coefficient, but their inclusion is not supported by significance tests. Columns 5 and 6 repeat this exercise using divisional rather than regional trends. While this hardly changes the results yielded by the property-crime and violent-crime regressions, the use of quadratic trends is now supported by a significance test in the murder regression.<sup>27</sup>

In short, the statistical significance of the geographic trends indicates that state-year crime regressions require controls for geographic influences. But the importance of the

---

<sup>26</sup> As is well known, the state-clustered covariance matrix is singular when state and year fixed effects are also included. This defect is usually inconsequential, because the coefficients on the fixed effects are typically of little interest. However, since Census regions and divisions are mutually exclusive groupings of states, it is impossible to perform F-tests on region-year or division-year interactions with a state-clustered covariance matrix.

<sup>27</sup> Note that the quadratic divisional trend specification is closest to the division-year setup we use in Table II.

abortion coefficients is still reduced even when parsimonious controls are used.<sup>28</sup>

### *Age-year regressions of Table III*

In our Figures 3a and 3b and Table III, we aggregate the state-year-age data across states, in order to reduce both the measurement error and the omitted variables bias that arises from the use of state-level data. In footnote 3 of their reply, DL claim that this aggregated analysis is unlikely to yield meaningful insights, because the resulting age-year regressions remain susceptible to confounding age-year shocks, of which the crack wave of the late 1980s and early 1990s is a possible example. The second column of our Table III uses a sample period that omits the main years of the crack wave, but we can also address DL’s concerns about age-year shocks by aggregating only to the regional or divisional level, rather than all the way to the national level. Partially aggregated regressions can include the age-year fixed effects that DL want to include as controls for the crack wave.<sup>29</sup> Additionally, partial aggregation will also reduce measurement error in the abortion variable. While young people may move out of their birth state before reaching adolescence, they are more likely to move to nearby states than to states that are far away. Hence, the abortion variable is more likely to be accurately measured on the regional or divisional level as compared to the state level.

The resulting regressions are presented in Appendix Table V. Column 1 aggregates the data to the divisional level, while column 2 aggregates to the regional level.<sup>30</sup> Three

---

<sup>28</sup> A final criticism of our state-year regressions is DL’s claim that a specification that nests our crime-trend interactions resuscitates an effect of abortion on crime. Specifically, the last columns of their Table III include regressions with state-specific trends and alternative sample periods. As DL found in their original [2001] paper, the use of state-specific trends in state-year regressions causes erratic changes in the estimated coefficients, because the regression must estimate 50 additional coefficients with limited data. (Our crime-trend interaction requires the estimation of only one additional coefficient.) In DL [2008], the estimated abortion-crime effect using state-specific trends ranges from 0.008 (property crime in 1985-2003 sample) to -.741 (murder in 1993-2003 sample). Moreover, as DL found in their original paper, the standard errors rise considerably.

<sup>29</sup> In addition to the age-year dummies, the partially aggregated regressions can also include interactions between dummies for the particular geographic area (region or division) and both age and year fixed effects. Like the regressions aggregated to the national level, the partially aggregated regressions cannot include state-year or state-age fixed effects. Hence, if the true model of arrests is a state-level model, so that *state*-year influences on crime are not well captured by *division*-year dummies, then the partially aggregated regressions will be misspecified and potentially biased. Analysis of the tradeoff between reduced measurement error and potential misspecification in aggregated regressions has a long history in econometrics; a classic reference is Grunfeld and Griliches (1960).

<sup>30</sup> Because there are nine Census divisions, ten age groups, and 14 years in the sample period (1985–1998), there are  $(9 \times 10 \times 14 =)$  1260 observations in the regression of column 1. Column 2 aggregates up to the four Census regions, so there are  $(4 \times 10 \times 14 =)$  560 observations in these regressions. The regressions are clustered by year-of-birth  $\times$  division in column 1 and year-of-birth  $\times$  region in column 2. This clustering pattern may not capture all of the relevant error correlations in these regressions, but none of the coefficients are significant with this limited clustering pattern in any case.

of the four coefficients in this table are positive. None is statistically significant. Like the nationally aggregated regressions in Table III, these partially aggregated regressions provide little support for a negative effect of abortion on crime.

## References

- Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates?", *Quarterly Journal of Economics* 119:1, pp. 249-275.
- Bell, Robert M., and Daniel F. McCaffrey (2002). "Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples," *Survey Methodology*, 28:2, pp. 169-181.
- Cameron, A. Colin, Jonah B. Gelbach and Douglas L. Miller (2006). "Robust Inference with Multi-Way Clustering," NBER Technical Working Paper No. 327 (September).
- 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.
- \_\_\_\_\_ (2006). "Measurement Error, Legalized Abortion, and the Decline in Crime: A Response to Foote and Goetz," NBER Working Paper No. 11987.
- \_\_\_\_\_ (2008). "Measurement Error, Legalized Abortion, and the Decline in Crime: A Response to Foote and Goetz," *Quarterly Journal of Economics*, February.
- Foote, Christopher L. and Christopher F. Goetz (2005). "Testing Hypotheses With State-Level Data: A Comment on Donohue and Levitt," Federal Reserve Bank of Boston Working Paper No. 05-15, November.
- Grunfeld, Yehuda, and Zvi Griliches (1960). "Is Aggregation Necessarily Bad?" *Review of Economics and Statistics*, 42:1, pp. 1-13.
- Hansen, Christian (2007). "Generalized Least Squares Inference in Panel and Multilevel Model with Serial Correlation and Fixed Effects," *Journal of Econometrics*, 140:2, pp. 670-694.
- Hansen, Christian (forthcoming). "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when  $T$  is Large," *Journal of Econometrics*, December.
- 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.
- \_\_\_\_\_ (2006). "Further Tests of Abortion and Crime: A Response to Donohue and Levitt (2001, 2004, 2006)," NBER Working Paper No. 12607.
- \_\_\_\_\_ (forthcoming). "A Simple Test of Abortion and Crime," *Review of Economics and Statistics*.
- Kézdi, Gábor (2004). "Robust Standard Error Estimation in Fixed-Effects Panel Models," *Hungarian Statistical Review*, Special English Volume No. 9, pp. 95-116.
- Lott, John R. Jr., and John E. Whitley, (2007). "Abortion and Crime: Unwanted Children and Out-of-Wedlock Births," *Economic Inquiry*, 45:2, pp. 304-324.
- MacKinnon, James G., and Halbert White (1985). "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29:3, pp. 309-325.
- Nickell, Stephen J. (1981). "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49:6, pp. 1417-26.
- Sailer, Steven (1999). "Does Abortion Prevent Crime?" *Slate Magazine*. Available at <http://www.slate.com/id/33569/entry/33571/>
- Solon, Gary (1984). "Estimating Autocorrelations in Fixed Effects Models," NBER Technical Working Paper No. 32.
- 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. ([www.seer.cancer.gov/popdata](http://www.seer.cancer.gov/popdata))
- Thompson, Samuel B. (2006). "Simple Formulas for Standard Errors that Cluster by Both Firm and Time," Harvard University mimeo.

**Table I: Arrests Regressions on the State-Year-Age Level**

	(1)	(2)	(3)	(4)	(5)
Arrests as Per Capita?	No	No	No	Yes	Yes
State-Year Fixed Effects Included?	No	Yes	Yes	Yes	Yes
Abortion Data Used	Original	Original	Original	Original	Adjusted
Sample Period	85-96	85-96	85-96	85-96	85-98
<b>Panel A: Log of Property Crime Arrests</b>					
Abortion Ratio/100	-.025	-.010	-.004	-.001	.001
Std Err Clustered by:					
Birth Year $\times$ State	(.003)*	(.002)*	(.002)*	(.002)	(.005)
State	(.005)*	(.003)*	(.003)	(.004)	(.008)
Population Coefficient			.605		
Std Err Clustered by:					
Birth Year $\times$ State			(.062)*		
State			(.135)*		
N	5740	5740	5740	5740	6730
<b>Panel B: Log of Violent Crime Arrests</b>					
Abortion Ratio/100	-.028	-.013	-.007	-.004	-.021
Std Err Clustered by:					
Birth Year $\times$ State	(.004)*	(.004)*	(.003)*	(.004)	(.008)*
State	(.012)*	(.005)*	(.004)	(.005)	(.014)
Population Coefficient			.686		
Std Err Clustered by:					
Birth Year $\times$ State			(.086)*		
State			(.220)*		
N	5737	5737	5737	5737	6724

**Notes:** Each observation in the data set is a cohort of 15- to 24-year-olds defined by state, year and age (for example, Massachusetts 17-year-olds in 1991). Results correspond to OLS regressions of the log of the cohort's arrests (or log per capita arrest rates in columns 4 and 5) on the cohort's *in utero* abortion exposure and various interactions. An asterisk denotes statistical significance at the 5% level. Age-year and state-age interactions are always included; state-year interactions are included in columns 2-5. In columns 1-4, abortions are measured by place of occurrence (as in DL [2001]), not by the state of residence of the mother. Column 5 uses the adjusted abortion data described in DL [2006], which uses residence-based abortion data and makes further adjustments to account for migration and statistical uncertainty about the timing of births and arrests within a calendar year. The sample period for columns 1-4 is 1985-1996 (as in DL [2001]), and the sample period for column 5 is 1985-1998. Not all states report arrest data for all years. The abortion ratio is divided by 100 in all regressions. State-level population weights are always used.

Figure 1: **Average Per Capita Crime Rates and Abortion Ratios: 1970-1984.** The abortion ratio is calculated by the Alan Guttmacher Institute as the number of abortions per 1,000 live births, and is based on the state of residence of the woman, not the state in which the abortion occurred. The crime rate is measured as incidents per 1,000 population and is not logged.

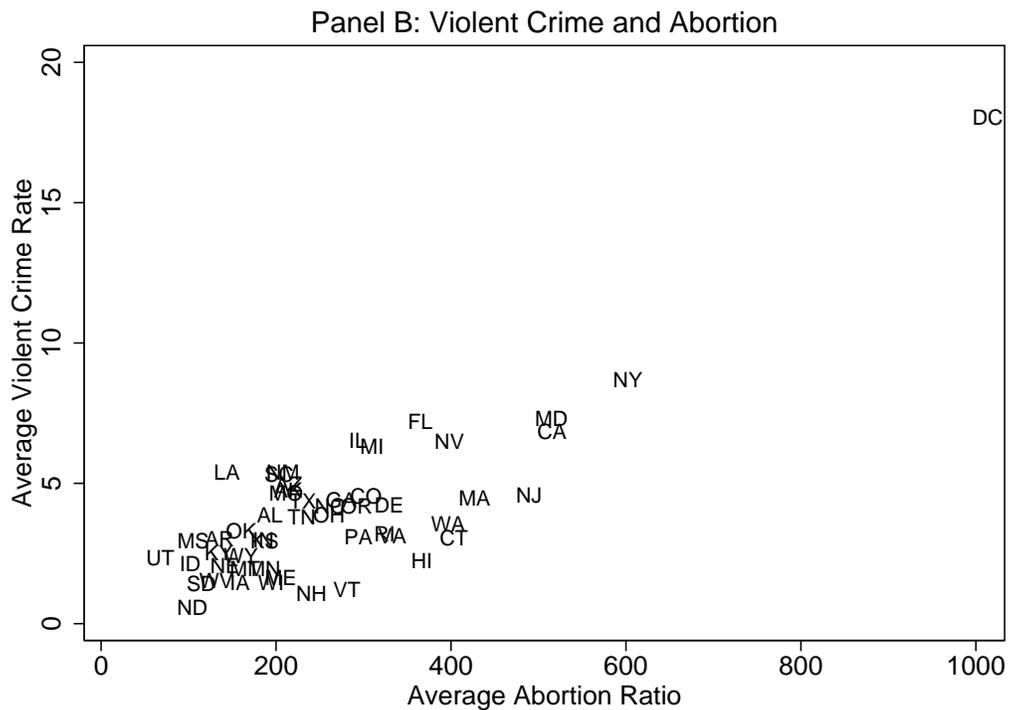
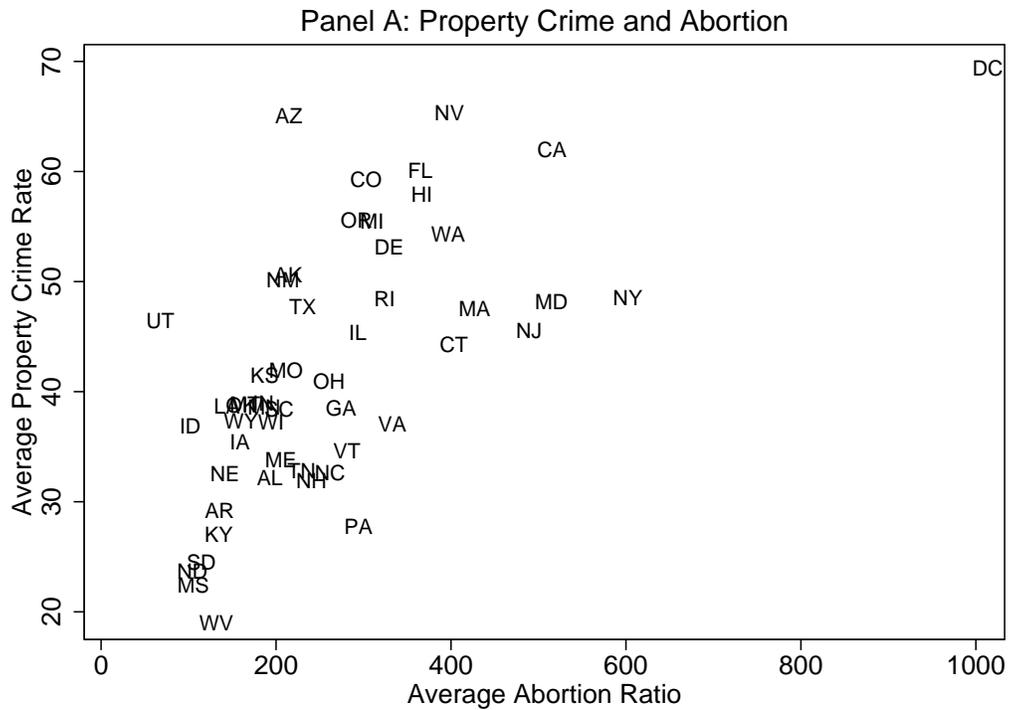
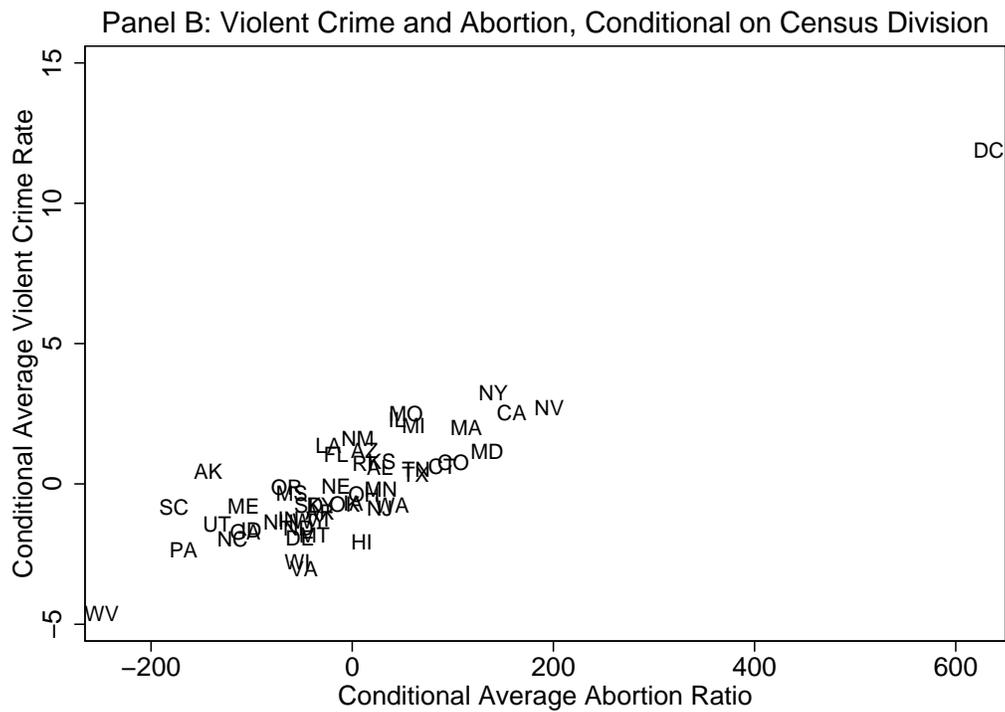
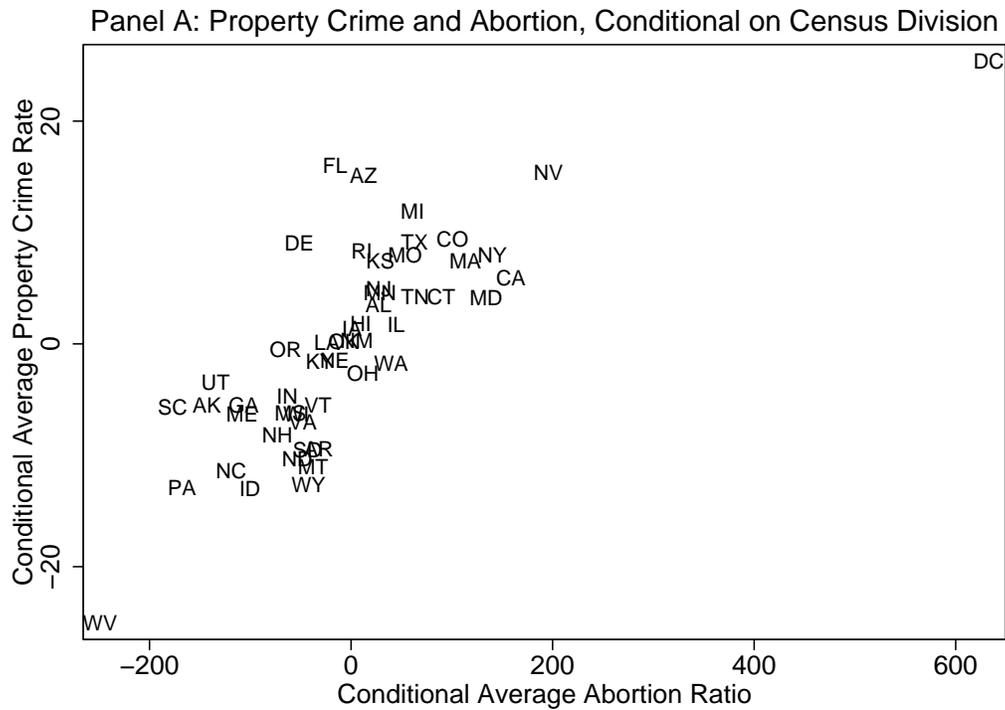


Figure 2: **Average Per Capita Crime Rates and Abortion Ratios, Conditional on Census Division: 1970-1984.** Figures are plots of residuals from regressions of the abortion and crime averages from Figure 1 on Census division dummies. See the notes to Figure 1 for details on the construction of the abortion and crime averages.

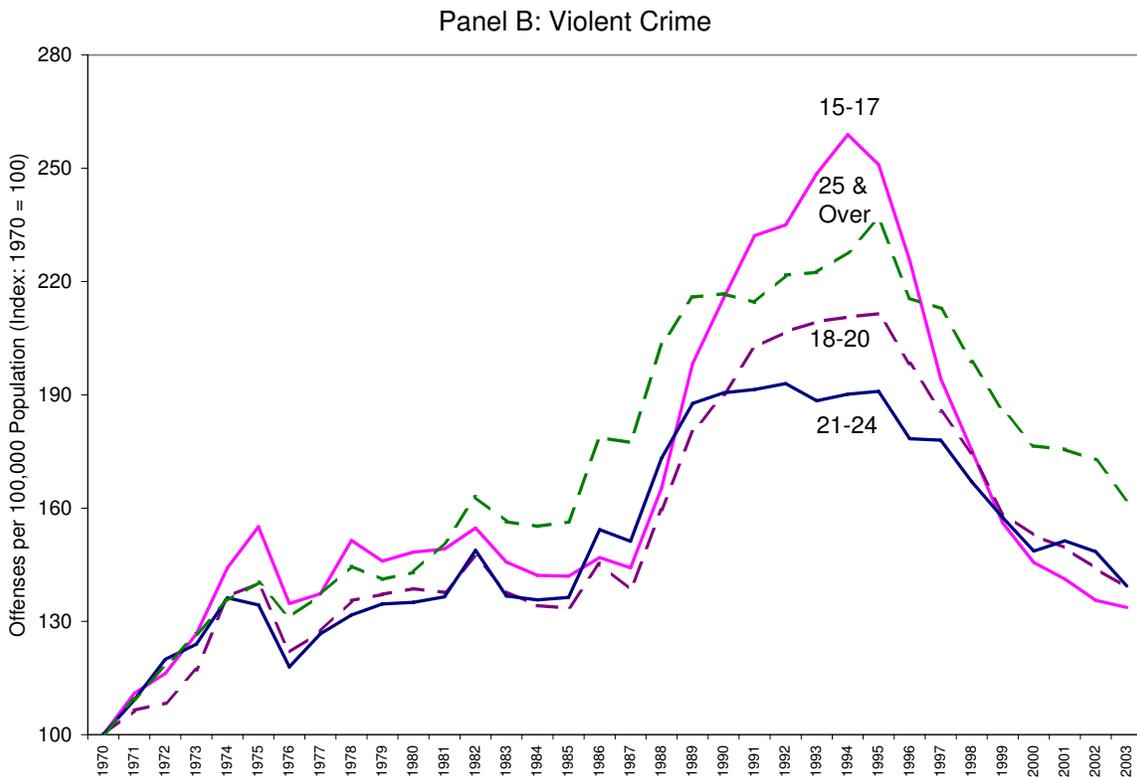
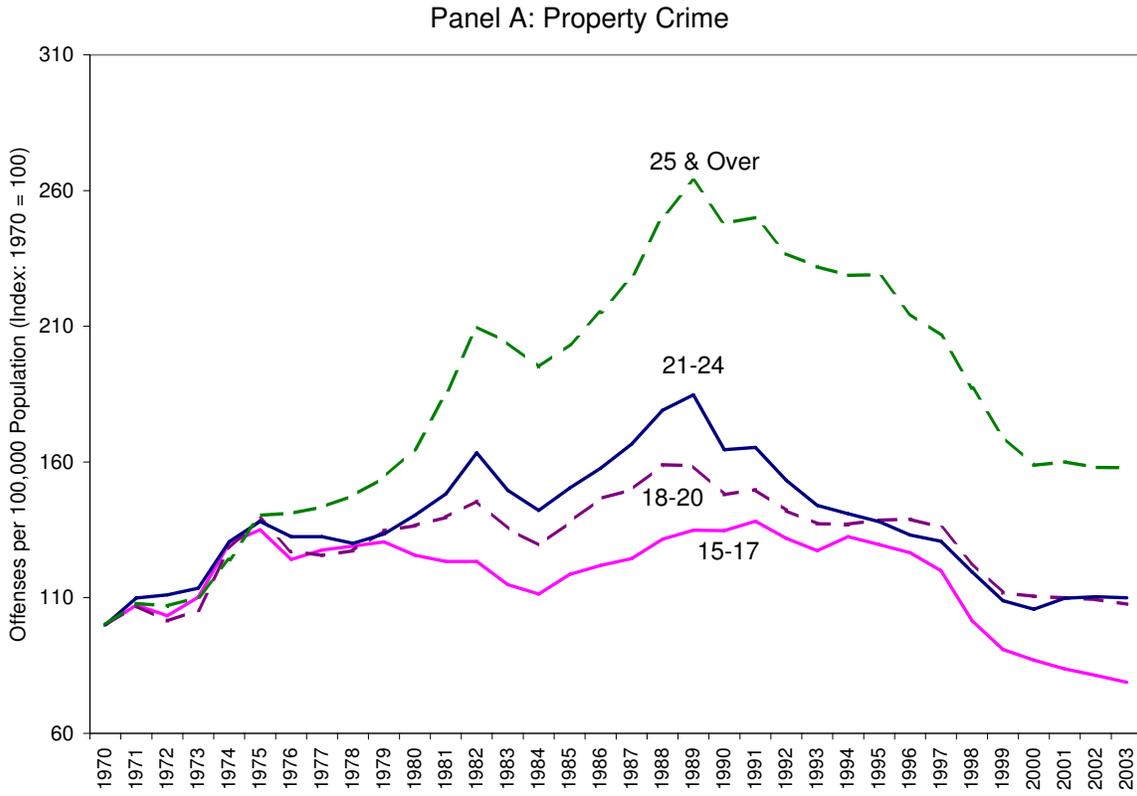


**Table II: Per Capita Crime Regressions on the State-Year Level**

	(1)	(2)	(3)	(4)	(5)
Sample Period	85-97	85-97	85-03	85-03	85-03
N	663	663	969	969	969
Abortion Data: Occurrence or Residence?	Occ	Res	Res	Res	Res
Geographic Controls	None	None	None	Division × Year	Division × Year
<b>Panel A: Log of Per Capita Property Crime Rate</b>					
Effective Abortion Ratio/100	-0.096 (.022)*	-0.114 (.026)*	-0.133 (.025)*	-0.131 (.043)*	-0.030 (.036)
1970-84 Log Per Capita Property Crime Average × Trend					-0.034 (.008)*
<b>Panel B: Log of Per Capita Violent Crime Rate</b>					
Effective Abortion Ratio/100	-0.137 (.032)*	-0.159 (.044)*	-0.165 (.035)*	-0.182 (.068)*	-0.087 (.073)
1970-84 Log Per Capita Violent Crime Average × Trend					-0.014 (.004)*
<b>Panel C: Log of Per Capita Murder Rate</b>					
Effective Abortion Rate/100	-0.115 (.045)*	-0.139 (.067)*	-0.121 (.053)*	-0.139 (.097)	-0.081 (.105)
1970-84 Log Per Capita Murder Average × Trend					-0.013 (.007)

**Notes:** Each observation in the data set corresponds to a group of persons defined by state and year (for example, all Massachusetts residents in 1991). Results correspond to Prais-Winsten regressions of the natural log of the state’s per capita crime rate on the corresponding effective abortion ratio and state and year fixed effects. Crime is defined by crimes reported to police, not actual arrests. An asterisk denotes statistical significance at the 5% level. Interactions between the year fixed effects and Census division dummies are included in columns 4 and 5. Column 5 also adds a trend that varies by state, constructed by multiplying the state’s mean annual log per capita crime rate from 1970-1984 with a linear time trend. State and year fixed effects are always included, and constant state population weights are always used. Standard errors are clustered by state, to account for the serial correlation of residuals within each state that remains after the Prais-Winsten quasi-differencing procedure.

Figure 3: **Per Capita Arrests Rates By Age Group: 1970-2003.** Source: Bureau of Justice Statistics (<http://www.ojp.usdoj.gov/bjs/data/arrests.wk1>).



**Table III: Per Capita Arrests Regressions on the Age-Year Level**

	(1)	(2)
Sample Period	1985-2003	1993-2003
N	190	110
<b>Panel A: Log of Per Capita Property Crime Rate</b>		
Abortion Ratio/100	.026 (.015)	.030 (.020)
<b>Panel B: Log of Per Capita Violent Crime Rate</b>		
Abortion Ratio/100	.057 (.014)*	.062 (.017)*

**Notes:** Each observation in the data corresponds to a cohort of persons aged 15 to 24 years in one calendar year (for example, all U.S. 17-year-olds in 1991). Results correspond to Prais-Winsten regressions of the natural log of the cohort's per capita arrest rate on its average abortion exposure and year and age fixed effects. An asterisk denotes statistical significance at the 5% level. The per capita arrest rates are calculated by dividing national age-specific arrest totals from various issues of *Crime in the United States* by population for the age-year cell. The abortion ratio for each birth cohort is constructed by averaging the appropriate residence-based abortion ratio by state, then making a further adjustment for statistical uncertainty about the date of arrests and births within a calendar year. Standard errors are clustered by year of birth. Regressions are not weighted by population.

**Appendix Table I:**  
**State-Year-Age Per Capita Arrests Regressions and Alternative Standard Errors**

	(1)	(2)
Type of Crime	Property Crime	Violent Crime
<b>Estimated Abortion Coefficient:</b>	.0012	-.0209
<b>Standard Error Clustered by:</b>		
Nothing (Huber-White robust errors)	(.0035)	(.0046)
Year-of-birth $\times$ state	(.0051)	(.0082)
State $\times$ age	(.0053)	(.0067)
State $\times$ year	(.0054)	(.0072)
State	(.0082)	(.0139)

**Notes:** The regressions in the table replicate the state-year-age regressions in last column of Table I, though the coefficients and standard errors are rounded to four decimal points rather than three. See the notes to Table I for details of the specification. The first standard errors reported are the Huber-White errors, which allow individual residuals to be heteroskedastic but do not account for any correlation patterns among residuals. The next four rows of errors account for various patterns using the cluster method. The first of these rows, which clusters by year-of-birth  $\times$  state, accounts for the fact that the same groups of people are observed in multiple observations in the data. The next row (state  $\times$  age) accounts for serial correlation among the errors corresponding to a given age group in a given state. The next row (state  $\times$  year) accounts for the use of data from 10 separate age groups for each state and year in the data, even though not all of these age groups may generate independent variation. The last row clusters by state, which accounts for all three potential correlation patterns, as well as any other correlation pattern among residuals from the same state.

**Appendix Table II:**  
**Pre-1985 Correlations Between State-Level Abortion and Crime Averages**

	Raw Averages		Averages Conditional on Census Division	
	(1)	(2)	(3)	(4)
	Property Crime	Violent Crime	Property Crime	Violent Crime
Weighted by Population (N=51)				
Abortion Coefficient	.045	.011	.067	.017
Standard Error	(.013)	(.001)	(.009)	(.001)
P-Value	[.0012]	[.0000]	[.0000]	[.0000]
R-Squared	.37	.62	.60	.72
Unweighted, excluding DC (N=50)				
Abortion Coefficient	.056	.010	.077	.014
Standard Error	(.011)	(.002)	(.011)	(.002)
P-Value	[.0000]	[.0000]	[.0000]	[.0000]
R-Squared	.36	.42	.58	.54
Unweighted, all states (N=51)				
Abortion Coefficient	.047	.014	.056	.017
Standard Error	(.008)	(.002)	(.010)	(.002)
P-Value	[.0000]	[.0000]	[.0000]	[.0000]
R-Squared	.41	.67	.59	.78

**Notes:** Results correspond to univariate regressions of  $\overline{Crime_s^{1970-84}}$  on  $\overline{Abort_s^{1970-84}}$ , where the dependent variable is the average per capita crime rate for state  $s$  from 1970 to 1984 inclusive, and the regressor is the average abortions-per-births ratio over the same period. Abortions in pre-legalization years are set to zero. Huber-White robust covariance matrices are used to generate standard errors and P-values. For the conditional results in columns 3 and 4, both the crime and the abortion averages are pre-whitened by regressing them on a slate of Census division dummies before running the univariate regressions reported. By the properties of OLS, the abortion coefficients in these columns are numerically identical to those from one-step regressions of crime averages on abortion averages and the Census division dummies. In columns 3 and 4, the standard errors and P-values are generated by the one-step regressions, so that the implied degrees of freedom for these statistics will be correct. However, the  $R^2$ 's from columns 3 and 4 are generated by the univariate regressions performed in the two-step procedure, so that they are not inflated by the presence of the Census division dummies. See Figures 1 and 2 for scatterplots of the data used in these regressions.

**Appendix Table III:  
State-Year Crime Regressions Omitting Particular States**

	Omitted States	None (Cols 4-5 of Table II)	DC	DC	DC, NY & CA	DC, NY & CA
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Log of Per Capita Property Crime Rate</b>						
Effective Abortion Ratio/100	-.131 (.043)*	-.030 (.036)	-.165 (.036)*	-.063 (.036)	-.142 (.041)*	-.030 (.041)
1970-84 Log Per Capita Property Crime Average $\times$ Trend		-.034 (.008)*		-.030 (.008)*		-.030 (.008)*
<b>Panel B: Log of Per Capita Violent Crime Rate</b>						
Effective Abortion Ratio/100	-.182 (.068)*	-.087 (.073)	-.221 (.065)*	-.130 (.079)	-.187 (.056)*	-.110 (.069)
1970-84 Log Per Capita Violent Crime Average $\times$ Trend		-.014 (.004)*		-.012 (.004)*		-.011 (.004)*
<b>Panel C: Log of Per Capita Murder Rate</b>						
Effective Abortion Ratio/100	-.139 (.097)	-.081 (.105)	-.181 (.103)	-.123 (.119)	-.099 (.095)	-.057 (.098)
1970-84 Log Per Capita Murder Average $\times$ Trend		-.013 (.007)		-.012 (.008)		-.012 (.006)

**Notes:** As in Table II, each observation in the data corresponds to a group of persons defined by state and year. Results correspond to Prais-Winsten regressions of the natural log of the state's per capita crime rate on the corresponding effective abortion ratio (divided by 100), state fixed effects, and interactions between year fixed effects and Census division dummies. The sample period is 1985-2003. Crime is defined as crimes reported to police, not actual arrests. An asterisk denotes statistical significance at the 5% level. Columns 2, 4, and 6 add a trend that varies by state, constructed by multiplying the state's mean annual log per capita crime rate from 1970-1984 with a linear time trend. Constant state population weights are used for all regressions. Standard errors are clustered by state, to account for the serial correlation of residuals within each state that remains after the Prais-Winsten quasi-differencing procedure.

**Appendix Table IV:**  
**State-Year Crime Regressions Using Region-Specific and Division-Specific Trends**

	(1)	(2)	(3)	(4)	(5)	(6)
Type of Trend Included	None	None	Linear Region	Quadratic Region	Linear Division	Quadratic Division
Number of Regressors Added	0	0	3	6	8	16
<b>Panel A: Log of Per Capita Property Crime Rate</b>						
Effective Abortion Ratio/100	-0.133 (.025)*	-0.084 (.032)*	-0.013 (.030)	-0.011 (.028)	-0.010 (.033)	-0.022 (.030)
1970-84 Log Per Capita Property Crime Average $\times$ Trend		-0.023 (.005)*	-0.039 (.007)*	-0.039 (.007)*	-0.038 (.008)*	-0.036 (.008)*
P-values for Exclusion of All Trend Terms			[.0017]	[.0001]	[.0000]	[.0000]
<b>Panel B: Log of Per Capita Violent Crime Rate</b>						
Effective Abortion Ratio/100	-0.165 (.035)*	-0.113 (.035)*	-0.070 (.041)	-0.070 (.041)	-0.088 (.063)	-0.097 (.065)
1970-84 Log Per Capita Violent Crime Average $\times$ Trend		-0.011 (.003)*	-0.015 (.003)*	-0.015 (.003)*	-0.014 (.003)*	-0.013 (.003)*
P-values for Exclusion of All Trend Terms			[.0219]	[.0157]	[.0235]	[.0000]
<b>Panel C: Log of Per Capita Murder Rate</b>						
Effective Abortion Ratio/100	-0.121 (.053)*	-0.102 (.054)	-0.088 (.079)	-0.088 (.079)	-0.097 (.097)	-0.100 (.098)
1970-84 Log Per Capita Murder Average $\times$ Trend		-0.011 (.006)	-0.009 (.007)	-0.009 (.007)	-0.011 (.006)	-0.011 (.007)
P-values for Exclusion of All Trend Terms			[.5804]	[.8836]	[.3466]	[.0005]

**Notes:** As in Table II, each observation in the data set corresponds to a group of persons defined by state and year. The sample period is 1985-2003, and all regressions have 969 observations. Crime is defined by crimes reported to police, not actual arrests. Estimates are generated from Prais-Winsten regressions of the natural log of a state's per capita crime rate on the corresponding effective abortion ratio (divided by 100) and state and year fixed effects. Standard errors are clustered by state, to account for serial correlation that remains after the Prais-Winsten quasi-differencing procedure. Columns 1 and 2 employ no additional geographic controls. Columns 3 and 4 use interactions between linear and quadratic trends that are specific to the four Census regions of the country. Columns 5 and 6 replicate the previous two columns using divisional rather than regional trends. Columns 2-6 also enter a crime-trend variable, constructed by multiplying the state's average annual log per capita crime rate from 1970-1984 with a linear time trend. Constant state population weights are used in all regressions. Asterisks denote statistical significance at the 5% level.

**Appendix Table V:  
Partially Aggregated Per Capita Arrests Regressions**

	(1)	(2)
Unit of Aggregation	Census Division	Census Region
Number of Areas	9	4
N	1260	560
<b>Panel A: Log of Per Capita Property Crime Rate</b>		
Abortion Ratio/100	.014 (.007)	.005 (.010)
<b>Panel B: Log of Per Capita Violent Crime Rate</b>		
Abortion Ratio/100	-.001 (.014)	.021 (.020)
Fixed-Effects	Age-Year	Age-Year
Included	Age-Division	Age-Region
	Division-Year	Region-Year

**Notes:** Each observation in the data set corresponds to a cohort of young persons aged 15 to 24 years, defined by either a U.S. Census region or division, a calendar year, and a single year of age (for example, 17-year-olds living in the New England Census division in 1991). Results correspond to Prais-Winsten regressions of the log of the cohort's per capita rate on the cohort's *in utero* abortion exposure and various interactions. Regressions use the adjusted abortion data described in DL [2006], which is based on residence-based abortion data and which makes further adjustments to account for migration and statistical uncertainty about the timing of births and arrests within a calendar year. The sample period is 1985-1998. The abortion ratio is divided by 100 in all regressions. Regressions are not weighted by population. Standard errors are clustered by year-of-birth  $\times$  division in column 1 and year-of-birth  $\times$  region in column 2.