Internet Appendix for

Car Seat Laws as Contraception

Jordan Nickerson
(Massachusetts Institute of Technology)

David Solomon
(Boston College)

Contact at jordo@mit.edu and david.solomon@bc.edu respectively. All remaining errors are our own.

A. Sample Construction

Our final dataset represents a panel, constructed from repeated cross-sectional snapshots that are back-filled in time. For each cross-section, we begin with either the ACS or decennial sample provided by IPUMS USA. First, we remove all group quarters when applicable. Next, we remove all non-core individuals, including parents and grand-children of the head-of-household. Then, we re-classify an un-married partner (`related=1114`) to be a spouse if no spouse is present. Next, we take steps to segment each census household into individual 19+ year old women. For households with multiple qualifying women and no children present, we simply assign all secondary adult-age women a new household identifier. For households with multiple adult-age women and children present, we are able to split the household by adult-age woman if only one such woman lists children. After splitting census households into individual adult-age women, in instances where a male is listed as the head-of-household, no spouse is listed, and there is one adult-age woman present, we re-classify the woman as a spouse.

Next, we drop all households with no adult-age women present. This includes all male head-of-households with no adult-age woman present, regardless of the status of children. Next, we drop households with inmates (`relate=13`) or children-in-law (`relate=4`) present. The result is a cross-section of family units, each with a single adult-age woman. We retain all family units where the woman is between ages 19 and 35. When computing household characteristics that potentially vary across individuals within a household (e.g., education) we take the highest value.

We convert the resulting cross-sections to a panel by back-filling. However, this presents two challenges. First, because we restrict the sample to women 35 years of age or younger as of the snapshot date, prior to the earliest ACS snapshot the maximum observable female age in the panel decreases as the panel extends backwards in time. Intuitively, the sample will exclude a woman who is 33 years old in 1997, as she will be 36 years old in the earliest snapshot (observed in 2000). For this reason, we demonstrate the robustness of our analysis in supplemental tests by truncating our final panel to begin in 2000, ensuring that each panel year contains a full cross-section of female ages. Second extending each snapshot backwards in time results in overlapping samples, with more overlap in earlier panel years. For example, while only one snapshot is used to create the 2017 panel year, the 2014 panel year is generated from four snapshots. Panel A of Figure A1 illustrates this, reporting the histogram of observations in the final sample by panel year. The panel demonstrates the steady decline in observations by year, as fewer ACS samples overlap. Similarly, the decreased sample sized from 2001 to 2004 reflects the lower sampling rate in early ACS surveys. Finally, the changing sample sizes prior to 2000 reflect the exclusion of women older than 35 years in the survey year. The remaining panels show how the number of observations across panel years also varies with a woman's age. Panel B of Figure A1 reports the number of observations in which a woman is 22 years old as of the panel year, while Panel C reports the histogram for 27 year olds. For ease of illustration, when producing the figures we truncate the sample to begin in 2000. Intuitively, while the data used to infer fertility decisions for a woman who is 22 years old in 2005 might come from any ACS vintage from 2005-2017, the 2013 ACS is the most recent vintage from which we infer births for a 27-year-old in 2005.

We re-weight observations in our empirical test to account for un-even sampling. Specifically, for a given (panel year, female age) tuple we sum the sampling rates over all cross-sectional snapshots that contribute to that tuple. For example, 25-year-old women in 2014 will have a total sampling rate of 4%, given the 1% sampling rate of ACS surveys in recent years and the four ACS vintages used. In contrast, 25-year-old women in 2013 stem from one of five ACS vintages, and would thus have a total sampling rate of 5%. We assign a sampling weight equal to the inverse of the total sampling rate of a (female age, panel year) tuple.

B. Simulation Details and Validation of Fixed Effect Estimation

Our simulations rely on point estimates for both variables of interest (e.g., a dummy denoting two children present, both bound by safety belt laws) and an assortment of fixed effects. Yet the granularity of fixed effects considered, which include county-year-number children and each unique combination of children ages, introduces a potential obstacle. While we are able to estimate fixed effects using a panel regression for the full set of true outcomes in the data, we are unable to estimate fixed effects for any "off-equilibrium" outcomes. For instance, we are not able to estimate the fixed effects for a woman with 5 children in Cooke County, Texas in 1994 if no-such woman in our sample meets those criteria. This is problematic if such a scenario occurs at some point in our counterfactual simulations. To side-step this issue, we model the estimated fixed effect of any interaction term (e.g., county-year-number of children) as the linear combination of the un-interacted terms from the set of fixed effects that we can estimate using our original sample. More precisely, we first regress each estimated fixed effect observed in our sample on the set of un-interacted covariates that make up the interaction term. We then use the estimated coefficients from this regression to generate fitted values for any permutation of the interaction that is not observed in the data.

Similarly, recall that our panel regressions restrict the sample to women younger than 36 years old. Thus, we do not observe a full cross-section of 18 to 35 year olds in the years prior to the 2000 decennial census. Yet, we are still able to simulate a full cross-section of women in each of these years. More specifically, our simulations begin each woman at age 18 with zero children. Thus, the only information needed to estimate the effect of laws on a 35-year-old in 1983 with a given set of characteristics (i.e., county-race-household income) is the weighted number of women meeting that criteria observed in our data (e.g., women that are 42 in 1990 or 52 in 2000). With that information, we are able to simulate such women beginning at age 18 in 1976 through to age 35 in 1983.[1] Finally, as we cannot directly estimate the fixed effects for a 35 year old in 1983 using our sample, we impute such point estimates using the approach outlined above.

Given the large dimension of high-density fixed effects we employ (17 fixed effects in some specifications), we estimate the model using `REGHDFE`. While the package allows a user to extract estimates for individual fixed effects, caution is given in doing so. Correia (2017) notes that fixed effects may not be identified by the estimating procedure.

Given the use of point estimates in our simulation framework, we briefly examine the robustness of fixed effect estimates in our context. We begin by estimating the full model described in Section 4.2, retaining all covariate and fixed effect point estimates. Using these point estimates, we compute the fitted probability of giving birth for each observation in our final panel. Using these fitted probabilities, we then simulate a new draw of births based on these predicted values. In other words, we generate a new set of child births which exhibits similar variation with respect to each covariate, and for which we know the true probability of birth. Finally, we re-estimate the same model using this new set of simulated births, and collect the point estimates from all covariates and fixed effects. From these point estimates, we construct a second set of fitted probabilities for each observation in the simulated panel. The result is a simulated panel for which we know the true probability of birth for each observation (used to simulate the panel), and have an estimated probability (from the second regression).

[1] The only restriction we do impose here is that we truncate each census snapshot at age 55 to avoid large distortions due to heterogeneous mortality rates.

To validate our use of fixed effect estimates, we report a scatter plot of true probabilities versus probabilities recovered from the second regression in Figure A2. In the figure, we first partition our full panel into 100 bins based on the true birth probability. For each bin, we then compute the mean of the true probability and that recovered by the second regression which uses data from the simulated births. The figure presents an extremely tight fit, with monotonically increasing recovered probabilities for an increase in true birth probabilities. This result suggests that the use of extracted fixed effects does not represent a significant issue in the context of our simulation approach.

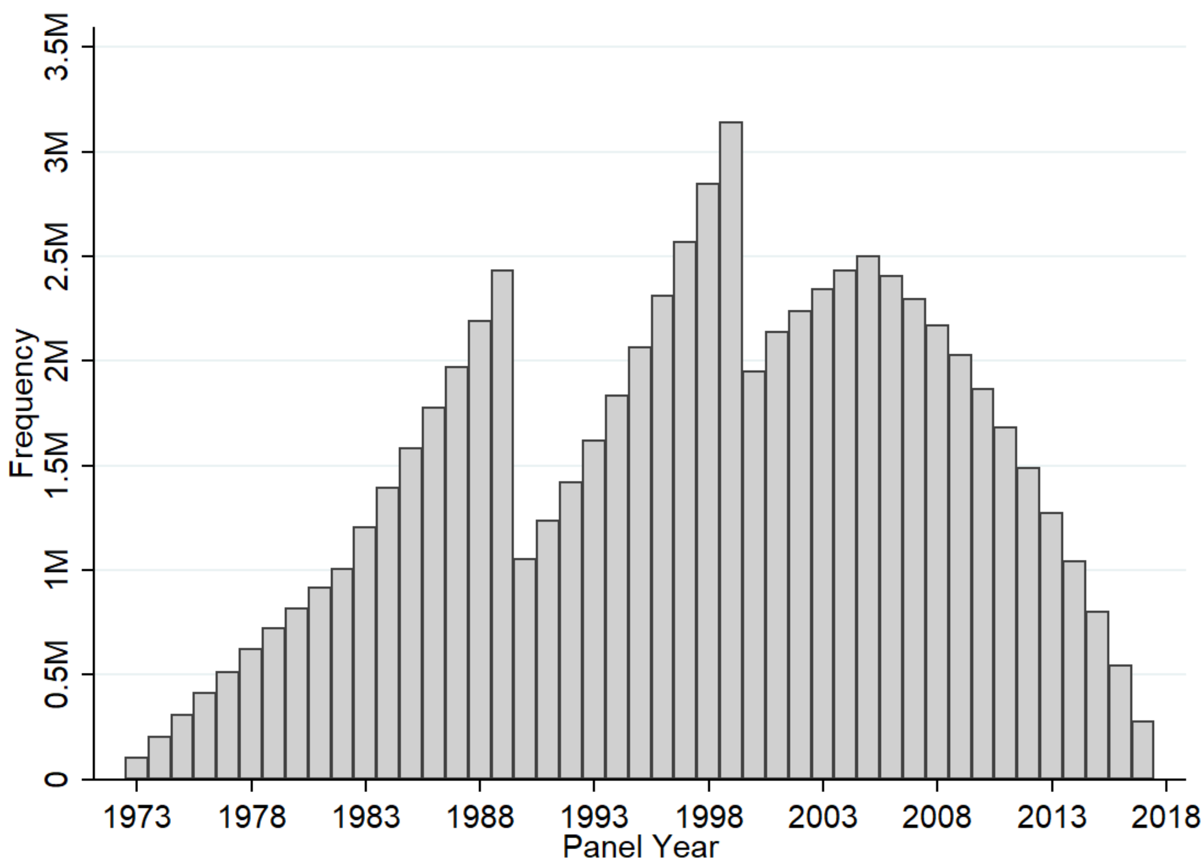
References

Correia, Sergio, 2017. "reghdfe: Stata module for linear and instrumental-variable/GMM regression absorbing multiple levels of fixed effects." *Statistical Software Components s457874*, Boston College.

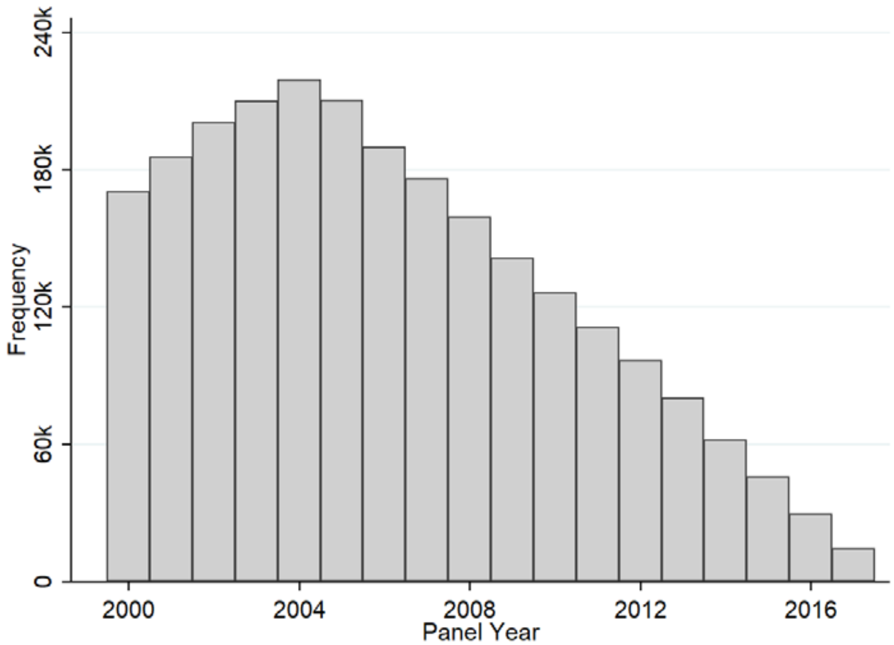
Figure A1 – Distribution of Panel-Year Observations

This figure shows the distribution of unweighted panel-year observations for women used in the sample from 2000 to 2017. Annual observations are created on a backward-looking basis by using snapshots of female ages plus number and ages of children, as described in Section 2.4 and 2.5 of the paper. Panel A plots the distribution of observations by panel year for all female ages. Panel B plots the distribution of observations for women who are 22 years old at the time of the panel year, while Panel C plots the distribution of observations for women who are 27 years old at the time of the panel year. Panel B and Panel C truncate the sample to begin in 2000 for expositional purposes.

Panel A – All Observations



Panel B – Women Aged 22 in Panel Year



Panel C – Women Aged 27 in Panel Year

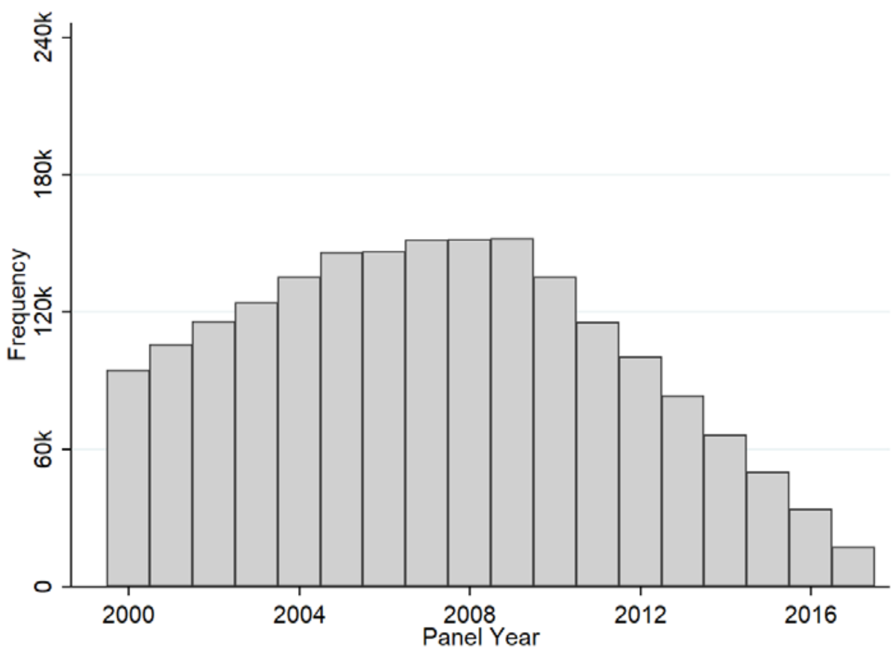


Figure A2
Recovered Estimated Probabilities Versus True Probabilities From Simulated Data

This figure shows the ability of the Stata REGHDFE procedure to recover the true coefficients in simulated versions of our main panel data. Using the procedures described in Section 4.2 of the paper, we produce simulated versions of the underlying data that generate a probability of birth for each woman/year observation. We then run the same procedures on the simulated data, and bin observations into 100 groups. The figure plots the scatter plot of true probabilities in the simulated data, versus the recovered probabilities from using the same REGHDFE procedures off which the original estimates were generated.