Readme File

The Short-Term and Localized Effect of Gun Shows: Evidence from California and Texas by Mark Duggan, Randi Hjalmarsson and Brian A. Jacob

The following table identifies the Stata do file and input data sets used for the creation of each figure and table in the paper.

Table/Figure	**.do File**	**Input Data Set(s)**
Table 1	rr_table1.do, table1_houston.do	rr_new_vars.dta, new_vars_houston2.dta
Table 2	rr_tables2_6dd.do	rr_new_vars.dta
Table 3	rr_tables2_6dd.do	rr_new_vars.dta
Table 4	rr_tables2_6dd.do	rr_new_vars.dta
Table 5	rr_tables2_6dd.do	rr_new_vars.dta
Table 6	rr_tables2_6dd.do	rr_new_vars.dta
Table 7	tables2_5_houst_dd.do	new_vars_houston2.dta
Appendix Table 1	rr_table1.do	rr_new_vars.dta
Appendix Table 2	rr_tables2_6dd.do	rr_new_vars.dta
Appendix Table 3	rr_tables2_6dd.do	rr_new_vars.dta
Appendix Table 4	rr_apptable4_dd.do	rr_new_vars.dta
Appendix Table 5	rr_tables2_6dd.do	rr_new_vars.dta
Appendix Table 6	rr_tables2_6dd.do	rr_new_vars.dta
Appendix Table 7	rr_tables2_6dd.do	rr_new_vars.dta
Appendix Table 8	rr_tables2_6dd.do	rr_new_vars.dta
Appendix Table 9	tables2_5_houst_dd.do	new_vars_houston2.dta
Appendix Table 10	table_app10_houst_dd.do	new_vars_houston2.dta
Appendix Table 11	tables2_5_houst_dd.do	new_vars_houston2.dta
Figure 1		annually aggregated data for figure 1.xls
Figure 2	rr_figures.do	rr_new_vars.dta
Figure 3	houst_figures2.do	new_vars_houston2.dta

As we mentioned in our original September 2008 letter to the Editors at the Review of Economics and Statistics, we cannot share the individual-level mortality data or the aggregate week * zip code mortality data used in our paper. However, we describe below how we constructed these data sets. Additionally, one can apply for this data directly from the states of CA and TX, respectively, using the information at the following sites:

http://www.cdph.ca.gov/data/dataresources/requests/Pages/DeathDataFiles.aspx

http://www.dshs.state.tx.us/chs/vstat/

California and Texas Mortality Analysis

The data set used for the California and Texas mortality consists of 573 weeks of data (January 7, 1994 to December 30, 2004) for 1,861 zip codes in Texas and 1,664 zip codes in California. This data set was created by combining data sets of mortality information with data sets of gun show information for both Texas and California. We also used information from the 2000 Census.

Below is a description of the sources and creation of each piece of data.

Texas Mortality Data

The Texas vital statistics data was sourced from the Center for Health Statistics in the Texas Department of State Health Services. The raw data included 1,609,203 deaths from 1994 to 2004. The data cleaning process resulted in us dropping 30,492 deaths from the raw data and a mortality data set of 1,578,711 deaths.

Specifically, cleaning of the data set resulted in the following data reductions. We dropped 7,902 deaths with zip codes that had fewer than 5 digits. We dropped 912 deaths that had zip codes beginning with a digit other than 7, which all Texas zip codes are supposed to begin with. All remaining death were recorded with 5 digit zip codes. Since some of these zip codes could contain typos, we merged in a list of zip codes that were in either the 1990 or 2000 census (or both). We then eliminated deaths that are not in zip codes listed in either census. Note that while this eliminates zip codes with typos, it also eliminates zip codes with populations too small to be listed in the census. Unfortunately, we cannot distinguish between the two. This is not a big concern, however, as this results in the loss of just 18,119 additional deaths. Finally, our analysis sample is all weeks (defined as Friday to Thursday) in 1994 through 2004. That is, the sample period begins on January 7, 1994 and ends on December 30, 2004. Thus, we drop 3,109 deaths that occur on January 6, 1994 or before and 450 deaths that occur December 31, 2004.

For each of the remaining deaths, we identify the date of death and zip code of residence. We then use International Classification of Disease cause-of-death codes to identify the following

sub-sets of deaths: non-gun suicides, gun suicides, non-gun homicides, gun homicides, accidental gun deaths, and gun deaths due to a firearm for which the cause (i.e. accidental or with intent) is undetermined. We then collapse the mortality data so that we find the total number of deaths and the total number of each the above described sub-categories of deaths for each week and zip code. This results in 573 weeks of data for 1,903 zip codes; i.e. all Texas zip codes with at least one death listed in either the 1990 census, the 2000 census, or both. (Note that in Texas, there were 5 census zip codes that had no deaths during this time period.)

The data set used in the analysis further restricts this sample to the set of Texas zip codes listed in the 2000 Census. This is because we use longitude and latitude information from the 2000 Census to calculate the distance between zip code centroids (i.e. to determine whether there are nearby gun shows). We also omitted zip codes with either zero population or zero land area according to the 2000 census. Thus, the final Texas analysis sample has 1,861 zip codes.

California Mortality Data

The California vital statistics data was provided by the Office of Health Information and Research in the California Center for Health statistics (CAHS). The raw data included 2,496,589 deaths from 1994 to 2004. The data cleaning process resulted in us dropping 24,442 deaths from the raw data and a mortality data set of 2,472,147 deaths.

Specifically, cleaning of the data set resulted in the following data reductions. All California deaths had a 5-digit zip code associated with it. Since some of these zip codes could still contain typos, we merged in a list of zip codes that were in either the 1990 or 2000 census (or both). We then eliminated deaths that are not in zip codes listed in either census. Note that while this eliminates zip codes with typos, it also eliminates zip codes with populations too small to be listed in the census. Unfortunately, we cannot distinguish between the two. This is not a big concern, however, as it results in the loss of just 18,894 deaths. Our analysis sample is all weeks (defined as Friday to Thursday) in 1994 through 2004. That is, the sample period begins on January 7, 1994 and ends on December 30, 2004. Thus, we drop 4,934 deaths that occur on January 6, 1994 or before and 614 deaths that occur December 31, 2004.

For each of these deaths, we identify the date of death and zip code of residence. We then use International Classification of Disease cause-of-death codes to identify the following sub-sets of deaths: non-gun suicides, gun suicides, non-gun homicides, gun homicides, accidental gun deaths, and gun deaths due to a firearm for which the cause (i.e. accidental or with intent) is undetermined. We then collapse the mortality data so that we find the total number of deaths and the total number of each the above described sub-categories of deaths for each week and zip code. This results in 573 weeks of data for 1,746 zip codes. This includes all California zip codes with at least one death listed in either the 1990 census, the 2000 census, or both. (Note that in California, all census zip codes had at least one death over this time period.)

The data set used in the analysis further restricts this sample to the set of California zip codes listed in the 2000 Census. This is because we use longitude and latitude information from the 2000 Census to calculate the distance between zip code centroids (i.e. to determine whether there are nearby gun shows). We also omitted zip codes with either zero population or zero land area according to the 2000 census. Thus, the final California analysis sample has 1,664 zip codes.

California and Texas Gun Show Data

Research assistants compiled data on California and Texas gun shows from advertisements in *Gun and Knife Show Calendar*, which is a national magazine that is published quarterly and that lists the dates and locations of gun shows throughout the country. The information on the gun shows is listed in two places in the magazine. First, each show has a full or half-page advertisement in the magazine. Second, there is a list of gun shows at the back of the magazine that are separated by state and arranged in chronological order. For each advertised show, we recorded the start date of the show, the length of the show, and the city, state, and location or venue of the show. The zip code of the show was identified by searching online for the address of the venue. (Note that the magazine often lists a zip code, but this zip code generally did not correspond to the actual location of the show and rather was the zip code of the show organizer.)

This resulted in 2,197 shows in Texas and 1,192 shows in California from 1994 to 2004. We omitted 10 shows in Texas and 13 shows in California that had been specified as knife-only. Zip code could not be obtained for 16 of the Texas gun shows and 5 of the California gun shows because the magazine did not provide information on the location. Since our sample period ranges from January 7, 1994 to December 30, 2004, we drop those gun shows that began before and after these dates, respectively.

We then collapse the gun show data by week (defined as Friday to Thursday) and zip code and rectangularize the resulting data set. This results in 573 weeks of data for 126 Texas zip codes and 101 California zip codes with at least one gun show over the sample period. Specifically, it indicates the number of gun shows in each zip code * week observation for the sample of zip codes with at least one gun show.

Finally, we combine the gun show data with data from the 2000 Census on the longitude and latitude of the centroid of each zip code to identify the number of gun shows in each week in nearby zip codes. Specifically, we calculate the number of gun shows within a 5, 10, and 25 mile radius of each zip code.

2000 Census Data

We use zip code level data from the 2000 Census for two purposes. First, as described above, we record the longitude and latitude of each zip code to use in determining the number of nearby gun shows. Second, we merge in demographic and socioeconomic variables for each zip code. These variables are used to split the sample for heterogeneity analyses and to assess the ways in which locales with gun shows differ from non-gun show zip codes.

Houston Crime Data Analysis

The data set used for the Houston crime analysis is titled "houston_weektract_base.dta". It was created by combining crime data obtained from an Open Records Request to the Houston Police Department, gun show data for Harris County, Texas, and 1990 Census Tract data. The resulting data set includes 573 weeks of data (i.e. each Friday to Thursday week from 1994 to 2004) for 446 Houston census tract. For each week*census tract observation, we record the number of Part I crime incidents the number of gun shows (in each census tract and in nearby tracts). Each underlying component of the final data set is described below.

Houston Crime Data

We obtained Part I crime incident data for each year from 1994 to 2004 through an Open Records Request to the Houston Police Department. The Houston Police Department provided a vast amount of information for each incident, including the date that it occurred, the address at which it occurred, the type of premise, the type of offenses, the number of suspects, and the number of victims. The variables on which we focus in our analysis are the date that each incident was reported, the six digit offense code, and the 1990 census tract of where the incident occurred. We use the six digit offense code to identify the following crime categories: homicides, gun homicides, rapes, gun rapes, robbery, gun robberies, assault, gun assaults, burglary, theft, and auto thefts. We also create the aggregated categories of violent offenses (homicide, rape, robbery, and assault) and property offenses (burglary, thefts, and auto thefts). Of the 1,495,518 Houston crime incidents from 1994 to 2004, there were just 1,161 incidents for which a census tract is not identified.

We collapse the crime data by weeks and census tracts and end up with a data set of 447 census tracts with at least one Part I crime incident from 1994 to 2004. That is, we have 256,131 census tract by week observations for which we have identified the number of crimes in each of the above categories.

Houston Gun Show Data

Gun show information is again obtained from *Gun and Knife Show Calendar*. Specifically, research assistants identified the date and 1990 census tract of each gun show in Harris County, Texas, which contains Houston. For each week and Houston census tract, we then identified the number of shows in that census tract as well as the number of shows within 1 mile, 3 miles, 5 miles, and 10 miles of that census tract. Specifically, there are eight census tracts in Harris County that had at least one gun show from 1994 to 2004. Using the latitude and longitude from the 1990 Census for the centroid of each census tract, we determine how far each census tract is from each of the eight tracts with a show. We then determine the number of shows within a specified distance each week. (Note that we use the 1990 Census longitude and latitude rather

than the 2000 census, since this is the unit of analysis provided by the Houston Police Department.)