

Review of

“Firearm Purchasing and Firearm Violence in the First Months of the Coronavirus Pandemic in the United States” by Julia P. Schleimer, MPH, Christopher D. McCort, MS, Veronica A. Pear, MPH, Aaron Shev, PhD, Elizabeth Tomsich, PhD, Rameesha Asif-Sattar, BS, Shani Buggs, PhD, MPH, Hannah S. Laqueur, PhD, MA, MPA, Garen J. Wintemute, MD, MPH

Reviewed by

Jared D. Fisher, PhD, and David W. Puelz, PhD

for *Rapid Review: COVID-19*, August 4, 2020

This paper examines the relationship between firearm purchases and firearm violence during the COVID-19 pandemic using a two stage model. Using state-level firearm purchasing data from January 2011 to February 2020, the authors forecast what firearm purchasing behavior would have been in March, April, and May of 2020. The forecast residuals (the differences between these forecasts and the actual purchasing levels) are used to estimate the excess firearm purchases as a result of the shock of the pandemic. Finally, they model states' incidences of firearm violence using the excess firearm purchases and some control variables. This model finds a statistically significant relationship between excess firearm purchases and excess firearm violence.

We are statisticians, thus for our Rapid Review, we will be focusing on the modeling decisions, statistical models, and causal inference components of this paper. On Rapid Review's scale, we find this paper to be between *not informative* and *potentially informative*. It is possible that the results may be correct, or that the truth is the opposite of the results. In fact, the authors concede as much in the limitations section of the paper. If the opposite direction of association is true (i.e., if firearm violence is causing unexpected firearm purchases), the entire discussion should be tempered. Decision-makers should not throw out this work but should ask for further investigation since results of *association* are rarely meaningful. This paper can shed light on what to look for to garner appropriate evidence to make a strong conclusion, as close to *causal* as possible, in future work.

Here, we focus on four issues. The first issue regards the selection of their seasonal autoregressive integrated moving average (SARIMA) model used to predict firearm purchases. The complexity of their SARIMA model was chosen by a published¹ stepwise algorithm, and its fit tested with a standard test. However, both of these model checks are in-sample, while the crucial use of the model is out-of-sample: how well does the model forecast data it has not seen? To this end, we would recommend that the authors demonstrate that the model indeed has good forecast performance before the pandemic. Forward-rolling time series cross-validation would be an appropriate way to measure this, at each of the one-/two-/three-month horizons. This is important, because the remainder of the paper and the remainder of the model rely on the excess firearm purchases being appropriately predicted out-of-sample. Their SARIMA model may perform great out-of-sample, but presently there is no evidence presented of this nature.

Second, the study focuses on state-level data, missing out on heterogeneity within each state. The ideal study scenario would be one where the analyst is able to check whether all who commit firearm violence (during March-May 2020) personally had increased firearm availability

due to purchasing activity. The authors note this is difficult to establish (“we have no information on whether the excess firearms acquired were those used in violence”) and have gathered what is likely the best-case data they can get on a state-level. However, consider a likely scenario in which most unexpected firearm purchases are in the suburbs and most firearm violence is located in cities. The authors’ approach would still find an association at an aggregate (state) level, but is this result meaningful if the purchasers are largely independent of the groups that are committing violence? To mitigate this, more granular data could be used. While individual-level data is unlikely, is the data available on the level of county or census tract?

Third, whether the data is state-level, individual-level, or anywhere in between, we have concerns about the covariates the authors have opted to leave out of the final model. In the exploratory analysis subsection, the authors comment on having tried many variables but left them out of the final model due to insignificance. However, some of the covariates left in the final model also lack statistical significance. We would ask that the authors include results with these other variables included in the models, perhaps in a table or set of tables. This would aid the clarity and transparency of the paper’s main point and robustness checks. This practice is common in economics papers when they use regression models to wrestle with associative and causal problems via observational data, as is the case here. Furthermore, it would seem that the pandemic’s effect on mental/emotional health and economic indicators would play large roles in this causal scenario, but these are not included in the final model nor the output.

Our last point regards causal inference. The authors do a good job of ensuring that non-causal words are used (“association”) and even include the blanket disclaimer in the limitations section. It might be worth noting this earlier in the paper, instead of near the end, and use language that is appropriate for the many non-technical readers interested in the paper (i.e. beyond “causal” and “association”). Related to this and our discussion above, with the current set of covariates and data, we are not entirely convinced the association is even meaningful. As a type of robustness check that could be performed, if “excess firearm purchases” was replaced with “excess online retail purchases”, would we still find an association between Amazon.com’s revenue and firearm violence using this method? It would be helpful for the authors to address these critiques through a more thorough data analysis and discussion in the text, beyond just citing other papers with similar conclusions and leaving these concerns for the limitations section at the end of the paper.

In conclusion, we suggest the authors include more descriptions of their statistical methods to convince readers of the soundness of their modeling choices. We need to see if the SARIMA model produces reasonable forecasts. We also need to see what the inclusion/exclusion of different covariates do to the statistical model and, importantly, the main exposure. Of course, if possible, data at a more granular level would help the argument along.

References:

Hyndman RJ, Khandakar Y. Automatic time series forecasting: the **forecast** package for *R*. *J Stat Softw*. 2008;27(3). doi:10.18637/jss.v027.i03