A dynamic factor model to predict homicides with firearm in the United States

Salvador Ramallo[1,3], Máximo Camacho[3], Manuel Ruíz Marín[4], and Maurizio Porfiri[1,2,5*]

[1]Center for Urban Science and Progress, New York University, New York, 11201, USA
[2]Department of Mechanical and Aerospace Engineering, Tandon School of Engineering, New York University, New York, 11201, USA
[3]Department of Quantitative Analysis, University of Murcia, Murcia, 30100, Spain
[4]Department of Quantitative Methods, Law and Modern Languages, Technical University of Cartagena, 30201 Cartagena, Spain
[5]Department of Biomedical Engineering, Tandon School of Engineering, New York University, New York, 11201, USA

ABSTRACT

Purpose

Research on temporal dynamics of crime in the United States is growing. Yet, mathematical tools to reliably predict homicides with firearm are still lacking, due to delays in the release of official data lagging up to almost two years. This study takes a critical step in this direction by establishing a reliable statistical tool to predict homicides with firearm at a monthly resolution, combining official data and easy-to-access explanatory variables.

Method

We propose a dynamic factor model to predict homicides with firearm from 1999 to 2020 using official monthly data released yearly by the Centers for Disease Control and Prevention, provisional quarterly data from the same agencies, media output from newspapers, and crowdsourced information from the Guns Violence Archive.

Results

Statistical findings demonstrate that the dynamic factor model outperforms state-of-the-art techniques (AI and classical autoregressive models). The dynamic factor model offers improved ability to backcast, nowcast, and forecast homicides with firearm, and can anticipate sudden changes in the time-series.

Conclusions

By decomposing the time-series of homicides with firearm on common and idiosyncratic components, the dynamic factor model successfully captures their complex time-evolution. This approach offers a vantage point to policymakers and practitioners, allowing for timely predictions, otherwise unfeasible.

Keywords— AI, autoregressive process, dynamic factor model, gun violence, mathematical modeling, time-series analysis

Introduction

Despite years of drastic reductions in crime levels, certain types of violent crimes have recently increased in the United States (US) (Rosenfeld & Lopez, 2020). Firearm violence is especially on the rise, as suggested

*Corresponding author: Center for Urban Science and Progress (CUSP), New York University Tanson School of Engineeering, 370 Jay Street, Room 1326, Brooklyn, NY 11201. E-mail: mporfiri@nyu.edu

by Albrecht (2022). Lately, we are witnessing a steady increase in non-fatal crimes, as documented by the Police Executive Research Forum (Police Executive Research Forum, 2021) and supported by the Giffords Law Center (Giffords Law Center, 2022). Just as non-fatal shootings are spiking across several cities in the country, fatal shootings continue to increase, as noted by the Pew Research Center (Gramlich, 2022) from data on gun death rates. This dramatic trend has attracted media attention and raised concerns about gun violence among Americans. The 2022 Election Tracking Survey from Ipsos (Ipsos, 2022) identifies gun violence as the second most important issue for Americans.

As an example of this recent trend, Figure 1 shows the 10-year relative increase in homicide offenses rates for different weapons registered on the National Incident-Based Reporting System (NIBRS) (Bureau of Justice Statistics, 2022). While those performed with firearm have dramatically increased, assaults with knives (jackknifes, cutting instruments, etc.), personal weapons (hands, feet, etc.), and other weapons (blunt objects, motor vehicles, explosives, etc.) have remained at the same level, or even decreased, in the last ten years (Rosenfeld & Fox, 2019).

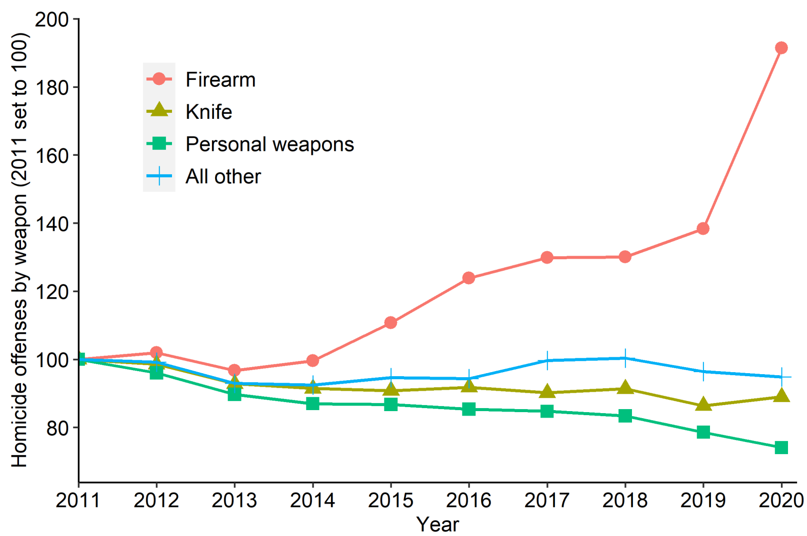


Figure 1: Relative changes in homicide offenses rates by type of weapon in the period from January 2011 to December 2020, from NIBRS. "Firearm" (red, •) refers to any type of gun (shotgun, rifle, handguns), "knife" (dark green, ▲) to cutting instruments, "personal weapons" (light green, ■) to the use of the body (hands, feet, etc.) as weapon, and "all other" (blue, +) to blunt objects, vehicles, and any other object used as a weapon.

The ability of policymakers and practitioners to efficiently make decisions that can reduce crime relies on wide and quick access to real-time data (Neumayer, 2003; Drake et al., 2009; Barreras et al., 2016). Evidence-based legislation is being embraced throughout the world. In the US, the Foundations for Evidence-Based Policymaking Act of 2018 became a law in 2019 (Congress.Gov, 2019), thereby requiring public access to agencies' data[1] and use of statistical evidence for any bill. Likewise, the Organisation for Economic Co-operation and Development points that countries should coordinate their strategies for policy evaluation with those related to evidence and data governance (OECD, 2020). However, several pitfalls persist with respect to availability, quality, and saliency of firearm-related violence data in the US (Strom & Smith, 2017; Roman & Cook, 2021).

According to the July 2014 National Criminal Justice Report from the Bureau of Justice Statistics (Regoeczi et al., 2014), the Wonder database of the National Center for Health Statistics from the Centers for Disease Control and Prevention (CDC) and the Supplementary Homicide Report (SHR) from Federal Bureau of Investigation (FBI) are the only source of detailed information on homicides (Centers for Disease Control and Prevention, 2021). While SHR is published with an annual frequency and suffers from missing or incomplete reports (Fox & Swatt, 2009), the CDC database contains granular data at a monthly resolution. Such

[1] An agency is defined under section 901(b) of title 31.

an improvement in data quality comes at the expense of a significant delay in its release, as acknowledged by Mancik et al. (2021). Specifically, in the CDC database all deaths for each month of the year are released all at once at the end of the following year. In this way, January data is released on December of the following year (23 months later) and December data is released on December of the following year (12 months later). The large amounts of data that need to be collected and processed are partly the reason why this data is published with such a hold-back. As a result, the actual picture of the homicides with firearm in the US can have a delay between 12 to 23 months, depending on the month of the analysis. This time delay cannot be compensated by using other series that are released with shorter delays without losing relevant information, since only 40% of violent crimes are reported to the police (Morgan & Thompson, 2021).

Mathematical tools could be used to overcome some of the practical drawbacks in data access, thereby supporting evidence-based interventions. However, to date, there is a paucity of reliable tools to gauge firearm violence in the US. Several efforts from the scientific community have tackled the problem of modeling temporal dynamics of crime-related rates, but these methods have seldom been applied to homicides and almost never to homicides with firearms – the objective of this study. Back in 2003, Gorr et al. (2003) already proposed a Holt exponential model to forecast crime series, which, however, is questionable when attempting long-term predictions (Chatfield & Yar, 1988; Alonso Brito et al., 2021) and is fragile for its excessive reliance on data extrapolation (Gardner Jr & McKenzie, 1985). Since then, a range of methods based on time-series analysis have been adapted to study crime rates in general (Berk, 2008), including homicides (Phillips, 2016).

Recently, Feng et al. (2018) has suggested that, among artificial intelligence (AI) techniques, tree models may outperform other approaches, such as k-nearest neighbors or naive Bayes for the prediction of crime, an observation which is in line with the previous findings by Nasridinov et al. (2013). In this vein, Berk et al. (2009) used random forests to specifically forecast homicides of paroles within two years after intake. Meskela et al. (2020) and Devi & Kavitha (2021) have proposed the use of a specific type of neural network particularly useful in time-series modeling, the long short-term memory recurrent neural network, to automate crime prediction. While promising, AI techniques are data hungry, whereby their performance is controlled by the richness of their training dataset (Carvalho et al., 2018), so that overfitting is difficult to avoid (Vezhnevets & Barinova, 2007).

In parallel to implementations of AI techniques, Cesario et al. (2016) and Yadav & Sheoran (2018) have used autoregressive integrated moving average (ARIMA) models to predict crime rates. These models are known to systematically revert the forecasts to the mean of the series (Deadman et al., 2001), making it difficult create reliable predictions. Building on the classical ARIMA, researchers have examined the percent change techniques (McDowall, 2002), ARIMA with fan charts (Yim et al., 2020), spatio-temporal autoregressive models (Shoesmith, 2013), and generalized least squares regression to study homicides.

Beyond univariate autoregressive models, classical multivariate autoregressive approaches were also adopted to study crime (Blumstein & Rosenfeld, 2008). The ability to account for multiple drivers in a multivariate sense allows for capturing salient phenomena that would be otherwise missed. For instance, one of the argued reasons for the registered increase in violence with firearm from 2014 to 2018 is the "Ferguson effect" (Hoffman et al., 2021; Cheng & Long, 2022). According to this theory, the police has been more scrutinized following the protests in Ferguson in 2014 in the wake of Michael Brown's death, thereby changing their approach to law enforcement, and, in turn, to crime prevention. Likewise, the number of active law enforcement officers has been suggested to drive the dynamics of crimes (Parker et al., 2017), along with economic fluctuations (Rosenfeld & Fornango, 2007) and firearm possession (Cook & Ludwig, 2006).

In this vein, Pratt & Lowenkamp (2002) related homicides to time-series of coincident economic indicators through a bivariate ARIMA model, whereas Cherian & Dawson (2015) has employed a vector autoregressive (VAR) model to predict several category crimes and Parkin et al. (2020) studied, within the same approach, the relation between deadly force incidents, line of duty deaths, and homicides rates. Although offering a much more complete view of homicide dynamics, most of the series that may be useful for multivariate time-series analysis could have very different sample periods and sampling frequency (so-called "ragged edge" problem identified by Wallis (1986)). Bringing the series to the same sampling period and frequency might cause an excessive reduction in the dataset, thereby hindering the use of any VAR model.

To bestow improved prediction of homicides, we relied on a single-index dynamic factor model (DFM) approach (Sargent et al., 1977; Stock & Watson, 1991), which allows for the integration of information from multiple, easily accessible time-series to predict a variable of interest. Single-index DFM relies on the co-movement of different series, thereby reducing dimensionality compared to a VAR approach and improving the reliability of the identification. The DFM framework allows for the use of data with different frequencies, as shown in Harvey (1990) and Mariano & Murasawa (2003), to create common and idiosyncratic dynamics.

By using Kalman filtering to fill any missing observations (Brockwell & Davis, 2009), DFM forecasts benefit from the different delays of the time-series. The proposed approach solves some of the limitations of the existing state-of-the-art. First, DFM is less prone to overfitting than AI techniques when working with real datasets of crimes that have limited size and potentially missing data (Mitchell & Mitchell, 1997; Ying, 2019; Soybilgen, 2020). Second, the identification of common and idiosyncratic dynamics within DFM avoids the problems of autoreregressive models related to the reversion to the mean (Beshears et al., 2013; Nau, 2014) and the inability to detect abnormal periods (Stock, 1994; Wheeler & Kovandzic, 2018; Yim et al., 2020).

Through the proposed DFM, we are able to predict homicides with firearm in the short term better than other benchmark models, such as tree-based models, neural networks, and classical autoregressive approaches. Specifically, through an out-of-sample exercise, we find that the DFM is the only approach that can perform a better prediction than a benchmark ARIMA model, on average for every month in the year. Furthermore, the model shows an improved ability to timely and accurately capture unexpected changes in the direction of the series, as those experienced in the recent COVID-19 outbreak. Not only does the model offer improved qualitative agreement with real data, but also it begets higher predictive accuracy. The enhanced ability to predict homicides with firearm offers a vantage point to policymakers and practitioners, allowing for timely predictions that would be otherwise unfeasible.

Materials and methods

Data

Homicides with firearm was our main variable of interest. On the Wonder database, we collected data on deaths whose "injury intent" was "homicide" and whose "injury mechanism and all other leading causes" were "firearm" from January 1999 to December 2020. This data is updated by the CDC once a year at a monthly resolution, and it is based on death certificates of US residents from the National Center of Health Statistics. Since there is no official monthly population data, we used the monthly population estimates (Bureau of Economic Analysis, 2022) from the Economic Bureau of Analysis to compute per-capita homicides with firearm.

Our approach is based on the use of a single-index dynamic factor model, which seeks to unveil a general co-movement in the data. To construct a common dynamics in the dynamic factor model, we should consider homicide-related data with a certain degree of co-linearity with homicides with firearm. The model's ability to deal with missing observations allows for taking into account data with different sampling periods. In this vein, the integration of series whose data is released before the official release of data about homicides with firearm is expected to improve predictions. With this in mind, we collected the following additional variables:

- **A provisional estimation of homicides with firearm**. Since January 2017, the CDC publishes a provisional quarterly estimation of the rate of homicides with firearm, released ten months after the end of the quarter (that is, a cumulative monthly series with two missing observations). The rate for the third quarter of 2020 was the highest recorded in our observation window ending in December 2020, with an annualized rate of 15 homicides with firearm per 100,000 inhabitants.
- **Data on deaths in incidents involving guns**. From January 2014, the Guns Violence Archive (Guns Violence Archive, 2022), a non-profit corporation, registers daily gun violence incidents from law enforcement, media, and commercial sources. Based on the incidents, they report the victims of gun violence, including murders, accidents, or suicides. We specified the filter selection "Shot-Dead (murder, accidental, suicide)" as "Incident Characteristic" to create a monthly series of deaths in incidents with guns from January 2014 to December 2020. The highest number of deaths was recorded on July 2020, with 1,964 deaths.
- **Data on homicides from three of the main cities in the US**. We collected daily crime rates in New York City, Chicago, and Philadelphia from the police departments. These data are publicly available from January 2006, weekly updated, and registered within the Unified Crime Report guidelines from the FBI. The Census Bureau Population data for each city was used to estimate the per-capita rates, which were further averaged among each other to create a baseline estimate. In our observation window from January 2006 to Decemeber 2020, the highest average rate was recorded on July 2020, with 2.41 homicides per 100,000 inhabitants.
- **Media output**. According to Phillips & Hensley (1984), when publicity is given in mass media to violence events, an increase in mortality is likely to follow. We collected media output from January

1999 to December 2020, using the Proquest Database, by searching for the number of news articles containing the words "homicide" and "shot" from the New York Times (2,997 news articles) and the Washington Post (6,175 news articles). Daily data was aggregated to create a monthly time-series where the highest media output was recorded on December 2020, with 107 news articles. Similarly, we collected the number of news containing the word "riots" (24,801 news articles) and those containing the word "unemployment" (6,925 news articles). The highest media output on riots was recorded in June 2020, when 744 news articles were published, whereas the peak of news articles about unemployment was in July 2020, with 195 news articles registered.

- **Google Trends**. Building on recent work on crime analysis with Google Trends (Gamma et al., 2016), we collected data from Google Trends with the search term "homicide" from January 2004 to December 2020 at a monthly resolution. The highest output was recorded on January 2013 (index of 100). We repeated the procedure with the term "gun", for which the highest output was observed on December 2012.
- **Background checks**. This classical proxy of firearm sales (Lang, 2013; Wallace, 2015) is made available by the FBI at a monthly resolution via the National Instant Criminal Background Check System (NCIS) (FBI, 2022a). In our observation window from January 1999 to December 2020, the highest number of background checks was recorded on December 2020, when 3,937,066 background checks were performed throughout the nation.
- **Economic uncertainty**. The Economic Policy Uncertainty (EPU) index, available at a monthly resolution from January 1999 from Baker et al. (2016), was chosen as a proxy of economic uncertainty. In our observation window from January 1999 to December 2020, the highest score was registered in May 2020, with an index of 350.
- **Microblogging data**. Building on recent work (Chen et al., 2015) on the use of Twitter data in modeling crime, we collected the number of daily geo-located tweets in the US containing the word "homicide" from January 2010 to December 2020. We collected a total of 146,661 posts and created a monthly time-series, whose peak was registered on May 2015 with 1,862 tweets. Recent work (Chen et al., 2015) has demonstrated the use of Twitter data in modeling crime.

For all the series mentioned except of the media output, a monthly seasonal adjustment was performed to remove seasonal effects. Besides, all the series except of homicides with firearm, monthly and provisional, and deaths in incidents with guns, were also detrended. The resulting time-series are plotted in Figures 2a to 2l. All the raw and processed data are available at Github.

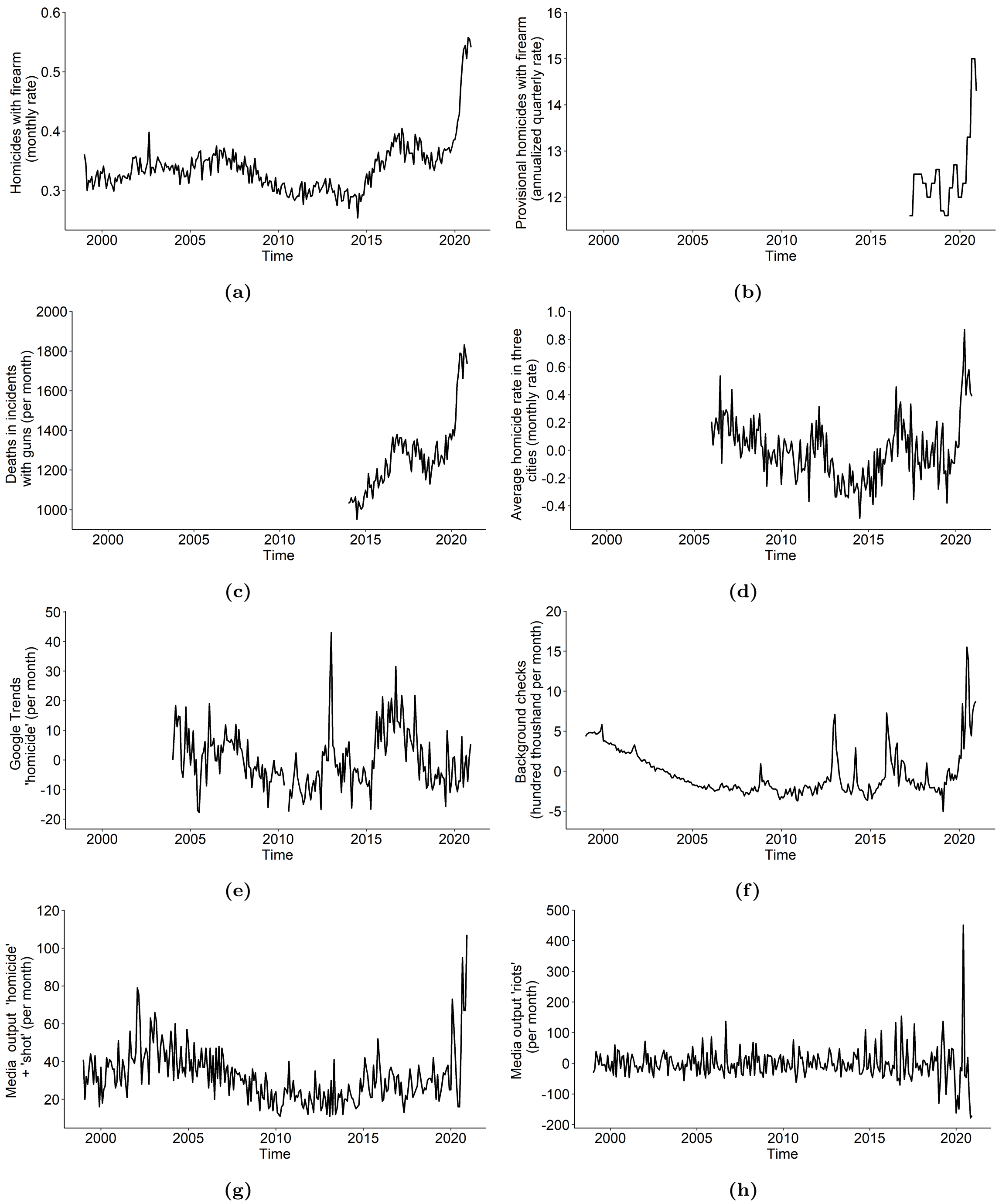
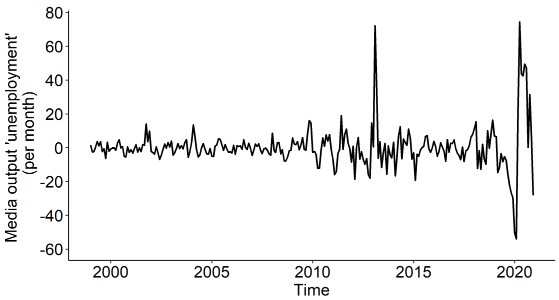
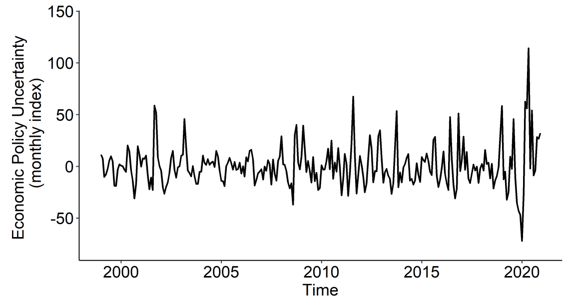


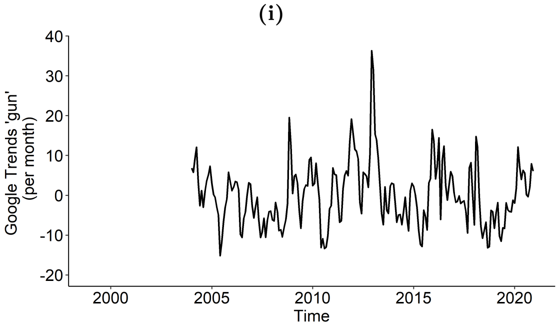
Figure 2: Monthly time-series collected for the period from January 1999 to December 2020. (a) Monthly homicides with firearm rate per 100,000 inhabitants from CDC (seasonally adjusted). (b) Provisional quarterly annualized estimate rate per 100,000 inhabitants for homicides with firearm from CDC. (c) Deaths in incidents involving guns from Guns Violence Archive (seasonally adjusted). (d) Monthly homicides rate for the aggregated cities of New York, Chicago, and Philadelphia (seasonally adjusted and detrended). (e) Monthly Google Trends for the word “Homicide” (seasonally adjusted and detrended). (f) Monthly background checks from NCIS (seasonally adjusted and detrended). (g) Media output for the news containing the words “homicide” with “shot”. (h) Media output for the news containing the word “riots” (seasonally adjusted and detrended).



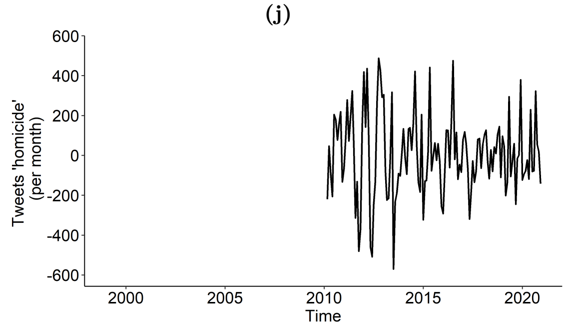
(i)



(j)



(k)



(l)

Figure 2: Monthly time-series collected for the period from January 1999 to December 2020 (continued). (i) Media output for the news containing the word "unemployment" (seasonally adjusted and detrended). (j) Monthly Economic Policy Uncertainty (EPU) index (seasonally adjusted and detrended). (k) Monthly Google trends for the word "gun" (seasonally adjusted and detrended). (l) Monthly geo-located tweets in the US containing the word "homicide" (seasonally adjusted and detrended).

The data differ in their resolution and release date throughout the year (Table 1). While variables such as background checks and media output are available in the same time-span of homicides with firearm (starting January 1999), Google Trends' data is available only from 2004 and homicides data in the three cities from 2006. All the variable, except of homicides with firearm and their provisional estimates, are released with no delay. Firearm homicides have a delay of 12 up to 23 months; for example data from January 2020 to December 2020 of homicides with firearm was released in December 2021. The provisional quarterly estimation is released with a fixed ten months of delay. Other variables available at yearly resolutions and released with similar delays as homicides with firearm (number of police officers (FBI, 2019), the National Crime Victimization Survey (Bureau of Justice Statistics, 2017), or the FBI violent crime index (FBI, 2019)), were not taken into account as they would not improve predictive power. Likewise, the registered murders in the recently released Quarterly Tables from NIBRS (FBI, 2022b) were not used due to its short historic period (three editions released up to date).

Dynamic factor model

Single-index dynamic factor models decompose the dynamics of the chosen observable variables $y_{i,t}$, for $i = 1, ..., n$ and $t = 1, ..., T$, as the sum of two unobservable and orthogonal components: one affecting all the time-series, f_t, and the other one accounting for their idiosyncratic variation, $u_{i,t}$. More specifically, we write

$$y_{i,t} = b_i f_t + u_{i,t}, \tag{1}$$

where b_i is the loading factor of each variable on the common factor. Any variable at a coarser resolution than the variable of interest (like the provisional estimate of homicides with firearm in Table 1) can be written as the sum of unobserved variables at the chosen resolution with proper delays[2]. The dynamics of the common

[2]For example, let $i = 2$ be a time-series with a quarterly resolution, then, we would write $y_{2,t} = b_2 \left(f_t + f_{t-1} + f_{t-2}\right) + u_{2,t}$.

Variable	Frequency	Sample	Delay
Homicides with firearm	monthly	1999M1-2020M12	12-23 months
Provisional estimate of homicides with firearm	quarterly	2017Q1-2020Q4	10 months
Deaths in incidents with guns	monthly	2014M1-2020M12	no delay
Monthly homicides (averaged in three cities)	monthly	2006M1-2020M12	no delay
Google Trends on "homicide"	monthly	2004M1-2020M12	no delay
Background checks	monthly	1999M1-2020M12	no delay
Media output on "homicide"+"shot"	monthly	1999M1-2020M12	no delay
Media output on "riots"	monthly	1999M1-2020M12	no delay
Media output on "unemployment"	monthly	1999M1-2020M12	no delay
Economical Policy Uncertainy index	monthly	1999M1-2020M12	no delay
Google Trends on "gun"	monthly	2004M1-2020M12	no delay
Number of tweets about "homicide"	monthly	2010M1-2020M12	no delay

Table 1: Sampling frequency, time-interval, and delays (in months) for each of the variables forming the dataset used to model and predict homicides with firearm at a monthly resolution.

factor and idiosyncratic components are described as autoregressive processes of orders p and q, that is,

$$\begin{aligned} f_t &= \beta_1 f_{t-1} + \ldots + \beta_p f_{t-p} + \epsilon_t^f \\ u_{i,t} &= c_{i,1} u_{i,t-1} + \ldots + c_{i,q} u_{i,t-q} + \epsilon_{i,t}^u, \end{aligned} \tag{2}$$

where ϵ_t^f captures the errors in the common factor and $\epsilon_{i,t}^u$ the errors in the ith idiosyncratic terms. Errors in the common factor and in the ith idiosyncratic term are considered independent and identically distributed (i.i.d.) in cross-section and time, following $\mathcal{N}(0, \sigma^2_{\epsilon^f})$ and $\mathcal{N}(0, \sigma^2_{\epsilon_i^u})$ distributions, respectively, where σ_{ϵ^f} and $\sigma_{\epsilon_i^u}$ are the standard deviations. Since the common factor is not observed either in mean and variance, σ_{ϵ^f} must be chosen to make the model identifiable (in particular, we set it equal to one).

The model in equations (1) and (2) can be cast in a state-space representation of the form

$$Y_t = H h_t, \tag{3}$$

where h_t is the state vector created from the common and idiosyncratic components in time. The transition equation is written as

$$h_t = F h_{t-1} + \epsilon_t, \tag{4}$$

with ϵ_t being i.i.d. $\mathcal{N}(0,Q)$ and $Q = \mathrm{diag}(\sigma^2_{\epsilon^f}, 0, 0, \sigma^2_{\epsilon_1^u}, 0, 0, \ldots, \sigma^2_{\epsilon_n^u}, 0, 0)$.

For example, considering a model with $p, q = 2$ and all variables at a monthly resolution expect of $y_{2,t}$ chosen to be quarterly, equation (3) reads

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ \cdot \\ \cdot \\ \cdot \\ y_{n,t} \end{pmatrix} = \begin{pmatrix} b_1 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ b_2 & b_2 & b_2 & 0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ \cdot \\ \cdot \\ \cdot \\ b_n & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ u_{1,t} \\ u_{1,t-1} \\ u_{1,t-2} \\ u_{2,t} \\ \cdot \\ \cdot \\ \cdot \\ u_{n,t} \\ u_{n,t-1} \\ u_{n,t-2} \end{pmatrix}, \tag{5}$$

and equation (4) becomes

$$\begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ u_{1,t} \\ u_{1,t-1} \\ u_{1,t-2} \\ u_{2,t} \\ \cdot \\ \cdot \\ \cdot \\ u_{n,t} \\ u_{n,t-1} \\ u_{n,t-2} \end{pmatrix} = \begin{pmatrix} \beta_1 & \beta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{1,1} & c_{1,2} & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_{2,1} & c_{2,2} & 0 & \dots & 0 & 0 & 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & c_{n,1} & c_{n,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ u_{1,t-1} \\ u_{1,t-2} \\ u_{1,t-3} \\ u_{2,t-1} \\ \cdot \\ \cdot \\ \cdot \\ u_{n,t-1} \\ u_{n,t-2} \\ u_{n,t-3} \end{pmatrix} + \begin{pmatrix} \epsilon^f_t \\ \epsilon^f_{t-1} \\ \epsilon^f_{t-2} \\ \epsilon^u_{1,t} \\ \epsilon^u_{1,t-1} \\ \epsilon^u_{1,t-2} \\ \epsilon^u_{2,t} \\ \cdot \\ \cdot \\ \cdot \\ \epsilon^u_{n,t} \\ \epsilon^u_{n,t-1} \\ \epsilon^u_{n,t-2} \end{pmatrix}, \tag{6}$$

where, due to the definition of ϵ_t and Q, the error terms of the period $t-1$ will be zero. The terms in these equations could be estimated by maximum likelihood through a Kalman filter.

In the case of missing observations, such as when treating quarterly data within our monthly resolutions, we use the approach given by Mariano & Murasawa (2003). Specifically, the missing values are replaced by random draws, α_t, from a distribution that does not depend on the parameter space that characterizes the filter, for example $\mathcal{N}(0, \sigma^2_\alpha)$.

The estimation algorithm for obtaining the parameters can be summarized as follows. Let $h_{t|\tau}$ be the estimate of h_t with information up to period τ, that is, the expected value of the state vector conditioned on the past. Denoting $P_{t|\tau}$ its covariance matrix, the prediction equations for the Kalman filter are

$$h_{t|t-1} = F h_{t-1|t-1}, \tag{7a}$$

$$P_{t|t-1} = F P_{t-1|t-1} F' + Q, \tag{7b}$$

where a prime indicates transposition.

Then, the error in the prediction is defined as

$$\eta_{t|t-1} = Y_t - H_t h_{t|t-1}, \tag{8}$$

and its covariance matrix is

$$\chi_{t|t-1} = H_t P_{t|t-1} H_t^{'} - R_t, \tag{9}$$

where R_t is the covariance matrix of the added noise that the approach of Mariano & Murasawa (2003) uses to treat missing values in equation (3). Hence, the Gaussian log-likelihood function can be evaluated as

$$l_t = -\frac{1}{2} \ln(2\pi |\chi_{t|t-1}|) - \frac{1}{2} \eta^{'}_{t|t-1} (\chi_{t|t-1})^{-1} \eta_{t|t-1}. \tag{10}$$

The next step in the Kalman filter is updating the estimation with the Kalman gain, typically defined as $K_t = P_{t|t-1} H_t^{'} (\chi_{t|t-1})^{-1}$, such that

$$h_{t|t} = h_{t|t-1} + K_t \eta_{t|t-1}, \tag{11}$$

$$P_{t|t-1} = P_{t|t-1} - K_t H_t P_{t|t-1}. \tag{12}$$

The initial parameters used to start the filter are typically a vector of zeros for (11) and a diagonal matrix for (12); the parameters that ultimately minimize the log-likelihood function in equation (10) are used as model fit parameters. The minimization is carried out through the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (Liu & Nocedal, 1989). Finally, missing values can be added at the end for doing forecast. This operation can be done since if not observed at period τ, the updating equation will be $h_{\tau|\tau} = h_{\tau|\tau-1}$, which will not change the dynamics of the model.

Comparison models and performance metrics

To quantify the ability of the DFM to predict the dynamics of homicides with firearm, its performance is compared to other benchmark models. Specifically, we examined models from two different families: AI and classic autoregressive models. Details about model implementation are presented in Appendix A and related codes are available at Github.

Within AI models, we considered two tree-based family models (random forest, RF, following Berk et al. (2009), and gradient boosting trees, GBOOST, following Kim et al. (2018)) and a long short-term memory recurrent neural network (LSTM), following Muthamizharasan & Ponnusamy (2022). In all cases, we created predictions by accounting for time variations in the availability of data. Specifically, in each time-segment, the models were trained using all the covariates available and predictions were made using models trained up to the latest available datapoint.

With respect to classical autoregressive models, we considered the univariate ARIMA model as the standard benchmark for time-series modeling. To acknowledge variations in levels and trends of the homicides with firearm series, we also implemented a Holt-Winters model (Winters, 1960; Gardner Jr & McKenzie, 1985). For completeness, we considered a VAR modeling as the benchmark for multivariate analysis, following Parkin et al. (2020) – Results are presented in Appendix B.

The main measurement used for comparison is the mean absolute error (MAE) of the model for the entire year at every month of prediction. To compare the predictive powers of the methods, we considered the HLN-statistic based on the difference in the model residuals, established by Diebold & Mariano (1995) and refined by Harvey et al. (1997) to deal with short series. To assess the ability of the forecasts to accurately anticipate changes in the series' directionality we used the PT-statistic (Pesaran & Timmermann, 1992), by which one can monitor the extent to which a model can anticipate the sign of the variation between two different time-steps of a series.

Results

In-sample model estimation

The smallest model that achieved convergence in the optimization of the parameters over the entire sample period consisted of the following four variables: monthly homicides with firearm, provisional quarterly estimates of homicides with firearm, deaths in incidents with guns from Guns Violence Archive, and media output of "homicide" + "shot". To improve the model accuracy, we systematically added one of the remaining variable if: i) its inclusion did not affect convergence, and ii) the new loading factor of the DFM was statistically significant at a confidence level of $\alpha = 0.05$. Such a procedure for model enrichment was conducted on both lagged and non-lagged variable. The final model also included the variables of homicides from three cities and background checks with seven months of lag. The b_i's parameters in equation (2) and their p-values are shown in Table 2.

	HF	PQE HF	GVA	H3	MOH	BCs(-7)
b_i	0.34	0.07	0.20	0.21	0.16	0.10
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$

Table 2: Loading factors for the model fitted over the entire sample period from January 1999 to December 2020. HF refers to monthly homicides with firearm from CDC, PQE HF to provisional quarterly annualized estimate of homicides with firearm from CDC, GVA to deaths in incidents with guns from Guns Violence Archive, H3 to monthly homicides in the three cities, MOH to monthly media output with "homicide"+"shot", and BCs(-7) to monthly background checks lagged seven periods.

The fitted model explained 59.80% of the variance, as estimated through linear regression of monthly homicides with firearm and the common factor. The resulting common factor, along with the reconstructed series of monthly homicides with firearm, are shown in Figure 3 for the sample period.

To ascertain test the robustness and appropriateness of the obtained model, we inspected the features of the residuals of the decomposed series. Specifically, we checked whether the residuals were normally distributed and serially uncorrelated. With respect to nornality, we utilized four different tests settled in the literature: the Kolmogorov-Smirnov test (Massey Jr, 1951) and its corrected version by Lilliefors (1967), the non-parametric

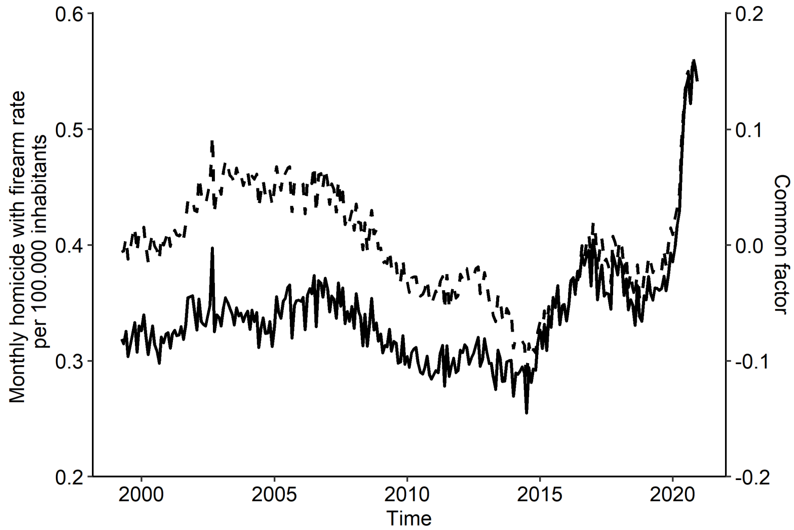


Figure 3: Adjusted model fitted for the whole sample period from January 1999 to December 2020. The dashed line represents the common factor (right vertical axis) and the black line the reconstructed series from the model on monthly homicides with firearm.

entropy-based test by Vasicek (1976), and the Shapiro-Francia test (Shapiro & Francia, 1972; Royston, 1993) for censored data after censoring extreme data. To test for serial correlation, we employed both the Box-Pierce (Box & Pierce, 1970) and the Ljung-Box (Ljung & Box, 1978) tests.

Results in Table 3 offer evidence in favor of normality and non-serial correlation. First, all of the normality tests yield large p-values for the null that the residuals are Gaussian, with the unique exception of background checks. Such a departure from normality should not be treated as a major concern. In fact, according to Durbin & Koopman (2012), if the true distribution of the error is non-Gaussian, then the Kalman filter would still provide the minimum variance linear unbiased estimator of the state variables, especially when only a handful of the residuals depart from normality (Barigozzi & Luciani, 2019). Second, all the residuals are serially uncorrelated, with the unique exception of those of quarterly annualized estimates of homicides with firearm – which should be expected given the time resolution of this variable. Similar to isolated lack of normality, serial correlation of a few variables have limited effect on the estimators and forecast in DFMs Stock & Watson (2002).

Out-of-sample comparison

An out-of-sample pseudo-real-time exercise was performed to evaluate DFM predictions. For each period of the sample history, a database was created with data available in that period. For each database, we estimated model parameters and recorded the corresponding forecasts. We compared forecasts with the true values and computed the MAE for each forecasted year at every month. The DFM was utilized to forecast future homicides with firearms and to provide missing values in the previous (backcasting) and present years (nowcasting). For example, in March 2017, the last available data for homicides with firearm is for December 2015, although information about other covariates can be available until March 2017. Through the DFM, we backcast from January 2016 until February 2017, nowcast March 2017, and forecast from April 2017 till December 2017. In what follows, we organize our comparisons between DFM predictions and real data in terms of MAE values for the previous (backcasting) and present year (backcasting, nowcasting, and forecasting for all months except January and December).

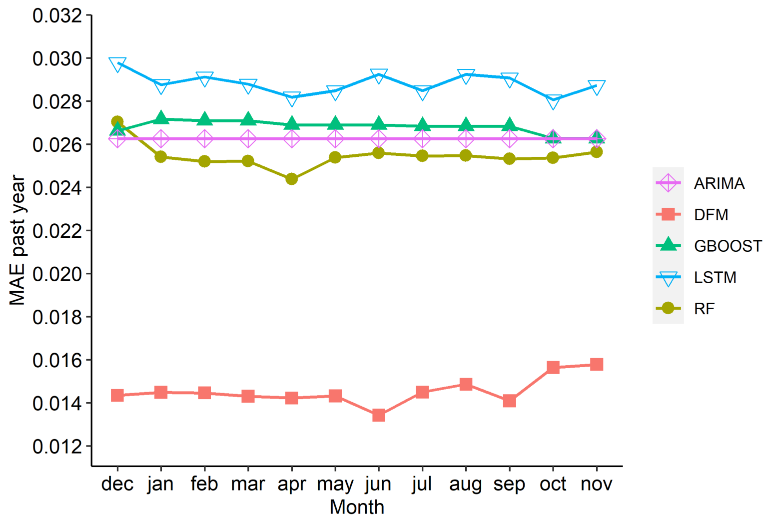
The same analysis was carried out for AI models. Since these models do not systematically impute missing values, we interpolated quarterly data using splines with the *imputeTS* R package. For the case of RF and GBOOST, multiple models were trained according to the availability of data. Should a variable not be available at a given month, it was excluded from the model training. For the case of LSTM, the procedure was based on training with sequences of three periods, using the *Keras* R package. The parameters of the

	Normality				Serial correlation	
	K-S	L	V	S-F	B-P	L-B
ϵ_t^f	0.036 (0.883)	0.036 (0.554)	0.098 (0.062)	1.315 (0.094)	1.327 (0.249)	1.343 (0.247)
ϵ_{1t}^u	0.077 (0.950)	0.077 (0.775)	0.153 (0.911)	-0.527 (0.701)	9.536 (0.002)	10.252 (0.001)
ϵ_{2t}^u	0.054 (0.439)	0.054 (0.066)	0.091 (0.174)	-0.068 (0.527)	1.274 (0.259)	1.288 (0.256)
ϵ_{3t}^u	0.063 (0.884)	0.063 (0.592)	0.130 (0.445)	-0.543 (0.707)	0.124 (0.725)	0.128 (0.720)
ϵ_{4t}^u	0.043 (0.749)	0.043 (0.320)	0.120 (0.078)	1.720 (0.044)	0.106 (0.745)	0.107 (0.744)
ϵ_{5t}^u	0.042 (0.913)	0.042 (0.624)	0.104 (0.504)	-0.592 (0.723)	2.194 (0.139)	2.231 (0.135)
ϵ_{6t}^u	0.154 (<0.001)	0.154 (<0.001)	0.414 (<0.001)	2.116 (0.017)	2.231 (0.345)	0.135 (0.342)

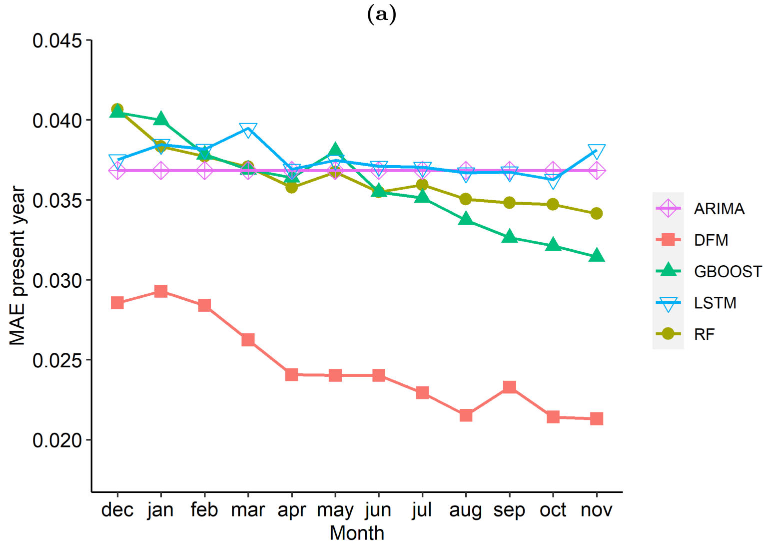
Table 3: Diagnostic tests on the DFM residuals. ϵ_t^f refers to the error of the common factor and ϵ_{it}^u to the errors of the idiosyncratic components: i=1 for provisional quarterly annualized estimate of homicides with firearm from CDC; i=2 for monthly homicides with firearm from CDC; i=3 for deaths in incidents with guns from Guns Violence Archive; i=4 for monthly media output with "homicide"+"shot"; i=5 for monthly homicides in the three cities; and i=6 for monthly background checks lagged seven periods. Numbers in parentheses are p-values. We use the following acronyms: K-S (Kolmogorov-Smirnov test), L (corrected version of the Kolmogorov-Smirnov by Lilliefors test), V (non-parametric entropy-based test by Vasicek), S-F (Shapiro-Francia test), B-P (Box-Pierce test), and L-B (Ljung-Box test).

RF and GBOOST models were identified using cross-validation over the whole sample period; for the latter model we used a Gaussian loss function. The networks' composition of the LSTM was based on training over the whole sample. A dropout layer was included after each LSTM layer, with a final time-distributed layer, and an *adam* optimizer (Kingma & Ba, 2015) was used. In addition to AI models, an ARIMA model was included as a benchmark. The number of parameters was chosen through the Akaike information criterion, yielding an ARIMA(3,0,3) as the model that minimized the score. Since new data for homicides with firearm appears in December of every year, such a univariate model provides the same prediction for any month from January to December.

The MAE associated with predictions of the past and present year at every month is shown in Figure 4. With respect to backcasting the series of homicides with firearm in the previous year, the DFM outperformed every other model, which all showed a comparable performance to an ARIMA. Likewise, the DFM performed better than any other model in the prediction of homicides with firearm in the present years. Importantly, the DFM shows an improved learning ability than any other model, whereby its MAE decays faster than any other model as a function of the month within the year. In other words, the DFM is effective in using information about homicides with firearm and all its covariates in the past to draw accurate predictions in the present year. Additionally, we performed a comparison between DFM and alternative autoregressive models (Holt-Winters and VAR models), whose results, displayed in Appendix B, offer further backing to the predictive power of DFM for backcasting, nowcasting, and forecasting homicides with firearm.



(a)



(b)

Figure 4: Out-of-sample pseudo real-time comparison between the DFM, AI models, and ARIMA for the period from January 2008 to December 2020, in terms of MAE: (a) previous and (b) present years. Each point represents an MAE value for either the entire year preceding the month at which the prediction is made (a) or the entire year when the prediction is made (b). DFM (red, ■), RF (dark green, •), GBOOST (light green, ▲), LSTM (light blue, ▽), and ARIMA(3,0,3) (violet, $\oplus$).

The qualitative observations drawn from the study of Figure 4 were further supported by statistical analyses using the HLN-statistic, as shown in Table 4. We tested whether the accuracy of the DFM was better than any of the other four models (RF, GBOOST, LSTM, and ARIMA). In agreement with our expectations, we registered a consistent improvement in forecasting using DFM, whereby any comparison yielded a significant difference in the HLN-statistic. With regards to backcasting, the DFM offered improved predictive capacity, yet, some of the comparisons failed to reach statistical significance. In particular, comparisons with ARIMA and GBOOST pointed at a marginal improvement bestowed by DFM.

We also tested the ability of the models to capture changes in directionality through the PT-statistic, that is, we studied the extent to which they were able to forecast changes in direction over different prediction periods. In particular, the analysis of Table 5 offers strong support for the systematic ability of DFM to predict the changes in directionality of homicides with firearm. In addition to reliable predictions of changes

	DFM vs RF	DFM vs GBOOST	DFM vs LSTM	DFM vs ARIMA
		Backcast		
		$h = 1$		
HLN	**-3.01**	-1.21	**-3.00**	-1.54
p-value	$< 10^{-2}$	0.23	$< 10^{-2}$	0.12
		$h = 2$		
HLN	**-2.23**	-1.42	**-3.64**	**-1.54**
p-value	0.03	0.15	$< 10^{-2}$	0.01
		$h = 3$		
HLN	-1.77	-1.92	-1.12	-1.82
p-value	0.08	0.05	0.07	0.07
		Forecast		
		$h = 1$		
HLN	**-4.94**	**-5.54**	**-5.32**	**-5.12**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 2$		
HLN	**-4.78**	**-4.87**	**-4.57**	**-4.55**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 3$		
HLN	**-4.78**	**-4.80**	**-4.74**	**-4.70**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 4$		
HLN	**-4.95**	**-5.89**	**-4.51**	**-4.22**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 5$		
HLN	**-5.62**	**-5.59**	**-4.56**	**-4.17**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 6$		
HLN	**-5.01**	**-5.33**	**-4.43**	**-4.04**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$

Table 4: Statistical analysis of DFM performance in backcasting and forecasting against other AI models and ARIMA using the HLN-statistic. The comparisons are carried out for different forecast/backcast horizons (h); bold values indicate a significant statistic at $\alpha = 0.05$.

in directionality for the first backcast period, the DFM also yielded some predictive abilities for longer backcast periods. Importantly, the DFM was the only model that was able to reliably predict such changes for any forecast horizon.

Discussion

In the 2010s, violent crimes exhibited a downward trend in the US (Friedman et al., 2017). However, some types of crimes, mainly those related to firearms, have increased in the period 2015-2020 (Albrecht, 2022), thereby attracting media attention and fueling gun policy debates (Barry et al., 2019). The spark in homicides with firearm during the lockdown in 2020 was the dramatic culmination of such a trend, which, however, remained officially undetected for two years – until official data was released. Such a systematic delay represents a major hurdle for policymakers and practitioners to promptly react to changes in violence. Establishing reliable, statistical tools to predict homicides with firearm is a key, open challenge.

	DFM	RF	GBOOST	LSTM	ARIMA
			Backcast		
			h =1		
PT	**2.38**	**1.83**	1.14	**3.06**	**1.83**
p-value	$< 10^{-2}$	0.03	0.13	$< 10^{-2}$	0.03
			h = 2		
PT	1.06	0.61	0.61	0.10	-0.61
p-value	0.14	0.27	0.27	0.46	0.73
			h = 3		
PT	**2.34**	0.61	**1.83**	-0.61	-1.83
p-value	$< 10^{-2}$	0.27	0.03	0.73	0.97
			Forecast		
			h = 1		
PT	**5.42**	0.72	1.02	-2.32	0.00
p-value	$< 10^{-2}$	0.24	0.15	0.99	0.5
			h = 2		
PT	**4.67**	-1.02	-1.41	-0.61	-1.99
p-value	$< 10^{-2}$	0.85	0.92	0.73	0.98
			h = 3		
PT	**4.34**	0.20	0.51	-0.71	-2.27
p-value	$< 10^{-2}$	0.42	0.31	0.76	0.99
			h = 4		
PT	**3.82**	0.28	0.58	-2.23	-1.64
p-value	$< 10^{-2}$	0.39	0.28	0.99	0.95
			h = 5		
PT	**3.90**	0.94	0.89	-0.75	-1.85
p-value	$< 10^{-2}$	0.17	0.19	0.77	0.97
			h = 6		
PT	**3.85**	0.27	-0.10	-0.68	-1.18
p-value	$< 10^{-2}$	0.39	0.54	0.75	0.88

Table 5: Statistical analysis of direction predictive accuracy of DFM, AI models, and ARIMA using the PT-statistic. The comparisons are carried out for different forecast/backcast horizons (h); bold values indicate a significant statistic at $\alpha = 0.05$.

Standard methods, like univariate autoregressive models, are known to yield unreliable predictions, whose error grows over time (Huang et al., 2020). Multivariate vector autoregressive models could, in principle, stabilize error growth, but the limited length of the time-series challenges the estimation of salient model parameters. The short length of the time-series also hampers the use of AI methods, which may overfit the data and fail to capture underlying patterns. Such an issue is further exacerbated by the wide difference in the range of the available time-series, which further restricts the dataset available for training. To address these issues, we propose a dynamic factor model, as a parsimonious monthly representation of the dataset that is updated in real time as any new information about homicides with firearm and any other explanatory variable (provisional quarterly estimates of homicides with firearm, deaths in incidents with guns from Guns Violence Archive, and media output of "homicide" + "shot") becomes available.

Our approach is not free of limitations. One of the key shortcomings of the proposed dynamic model is that, as a single index model, it relies on unique common dynamics. In principle, multiple grouped dynamics might coexist when working with large datasets, so that more than one common factor would be needed. In such a case, homicides with firearm could be associated with more than a single common factor, thereby challenging the use of a dynamic factor model with a single factor for reliable predictions. Likewise, the loading factors of the dynamic factor model could be time-varying; for example, during a certain period, homicides with firearm might be closely related to social unrest and in another period, they might be tied to gun-related dynamics. Such time-varying dynamics might have occured in the last section of the observation, entailing the COVID-19 lockdown, when it is tenable that additional social and economic variables might have played a role. In addition to these methodological drawbacks, we should acknowledge limitations in the process of data curation, whereby data from media can be noisy and crowdsource databases are not officially verified, leading to potentially inaccurate estimations.

In the future, several research directions can be pursued. First, Bayesian estimation following Del Negro & Otrok (2008), or extended/unscented Kalman filters can be leveraged to cope with richer dynamics and nonlinear behaviors. Second, the approach can also be adapted to a spatio temporal setting, as in Lopes et al. (2011), working with time-series of homicides with firearm for distinct US states. Lastly, we envision the inclusion of further data, like the Quarterly Tables from NIBRS (FBI, 2022b) to enrich the existing dataset and enhance the accuracy of our predictions.

Conclusions

The investigation carried out in this effort bears key methodological and practical insights that should be highlighted. From a methodological point of view, we find that decomposing the time-series of homicides with firearm into common and idiosyncratic components through a dynamic factor model yields superior predictions than standard autoregressive and AI models. The superior performance of the dynamic factor model is likely due to asynchronous reporting calendar for official mortality statistics by the CDC, which strain the applicability of the existing machinery. The dynamic factor model thrives on this asynchronicity, where its Kalman filter allows for the incorporation of any information as it becomes available, thereby easing the processes of backcasting, nowcasting, and forecasting.

Despite their increased computational costs, AI techniques failed to contribute significant improvement with respect to a benchmark ARIMA model, in contrast with the dynamic factor model that systematically outperformed ARIMA in backcasting, nowcasting, and forecasting in terms of HLN-statistic (Harvey et al., 1997). The improved ability to detect trends of the dynamic factor model comes with the further additional advantage of reliably anticipating changes in directionality in the time-series of homicides with firearms, a key element to effective policymaking. Through PT-statistic (Pesaran & Timmermann, 1992), we demonstrated an improved ability to detect such changes for both backcasting and nowcasting, in contrast with any other model that can only be reliably used for very narrow backcasting periods.

To our surprise, a new research has been recently conducted by AH firm (AH Datalitycs, 2022), as highlighted by the New York Times (New York Times, 2022), pointing at a reduction of homicides this year. Gun deaths, injuries, and mass shootings are also down this year, compared to the previous year. The New York Times provides an interesting interpretation of the findings, relating to a return to normalcy after a long pandemic, and touches on some “bad news bias” that have prevented such a funding to reach a wider audience. Data by the CDC regarding this very same time period will not be available until December 2023: our new approach to prediction of homicides with firearm could be key in accelerating the validation stage of any new findings like those by AH firm. More in general, the superior performance of the dynamic factor model has important, practical implications for policymakers and practitioners who are tasked with making timely decisions on pressing topics around violence with limited and often outdated data. In this vein, the model may be used to fill data gaps and anticipate outbreaks of violence, thereby offering a concrete aid to evidence-based interventions.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the National Science Foundation under award CMMI-1953135. S.R. acknowledges the support of a Spain-U.S. Fulbright grant co-sponsored by Fundación Séneca. M.C. and M.R. are grateful for the financial support of Ministerio de Ciencia e Innovación of Spain under grants PID2022-136547NB-I00, and PID2019-107800GB-I00 funded by MCIN/AEI/ 10.13039/ 50110001103, respectively. The authors are also thankful to Drs. Barak-Ventura and Boldini for their careful reading of the manuscript.

Data and codes availability

All data and codes are available at Github.

References

AH Datalitycs. (2022). *YTD Murder comparison.* (https://www.ahdatalytics.com/dashboards/ytd-murder-comparison/)

Albrecht, J. F. (2022). Increasing gun and community violence in the United States: Causes and analyses. In *Understanding and preventing community violence* (pp. 119–130). Springer.

Alonso Brito, G. R., Rivero Villaverde, A., Lau Quan, A., & Ruíz Pérez, M. E. (2021). Comparison between SARIMA and Holt–Winters models for forecasting monthly streamflow in the western region of Cuba. *SN Applied Sciences*, *3*(6), 1–12.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Barigozzi, M., & Luciani, M. (2019). Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm. *arXiv preprint arXiv:1910.03821*.

Barreras, F., Daz, C., Riascos, A., & Riber, M. (2016). Comparison of different crime prediction models in Bogotá. *Documentos CEDE*, *34*.

Barry, C. L., Stone, E. M., Crifasi, C. K., Vernick, J. S., Webster, D. W., & McGinty, E. E. (2019). Trends in public opinion on US gun laws: majorities of gun owners and non–gun owners support a range of measures. *Health Affairs*, *38*(10), 1727–1734.

Berk, R. (2008). Forecasting methods in crime and justice. *Annual Review of Law and Social Science*, *4*, 219–238.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009). Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(1), 191–211.

Beshears, J., Choi, J. J., Fuster, A., Laibson, D., & Madrian, B. C. (2013). What goes up must come down? Experimental evidence on intuitive forecasting. *American Economic Review*, *103*(3), 570–74.

Blumstein, A., & Rosenfeld, R. (2008). Factors contributing to US crime trends. In *Understanding crime trends: Workshop report* (Vol. 2, pp. 13–44). National Academies Press.

Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, *65*(332), 1509–1526.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brockwell, P. J., & Davis, R. A. (2009). *Time series: theory and methods.* Springer Science & Business Media.

Bureau of Economic Analysis. (2022). *GDP personal income table. BEA's National Economic Account.* (https://apps.bea.gov/iTable/index_nipa.cfm)

Bureau of Justice Statistics. (2017). *National crime victimization survey.* Inter-university Consortium for Political and Social Research.

Bureau of Justice Statistics. (2022). *National Incident-Based Reporting System, 2020: Extract Files.* Inter-university Consortium for Political and Social Research.

Carvalho, J., Santos, J., Torres, R., Santarém, F., & Fonseca, C. (2018). Tree-based methods: Concepts, uses and limitations under the framework of resource selection models. *Journal of Environmental Informatics*, *32*(2).

Centers for Disease Control and Prevention. (2021). *National Vital Statistics System, Mortality 1999-2020 on CDC WONDER online database. Multiple cause of death files, 1999-2020.* (https://wonder.cdc.gov/ucd-icd10.html)

Cesario, E., Catlett, C., & Talia, D. (2016). Forecasting crimes using autoregressive models. In *2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, 14th International Conference on Pervasive Intelligence and Computing, 2nd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (pp. 795–802). IEEE.

Chatfield, C., & Yar, M. (1988). Holt-Winters forecasting: some practical issues. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *37*(2), 129–140.

Chen, X., Cho, Y., & Jang, S. Y. (2015). Crime prediction using Twitter sentiment and weather. In *2015 Systems and Information Engineering Design Symposium* (pp. 63–68). IEEE.

Cheng, C., & Long, W. (2022). The effect of highly publicized police killings on policing: Evidence from large US cities. *Journal of Public Economics*, *206*, 104557.

Cherian, J., & Dawson, M. (2015). RoboCop: Crime classification and prediction in San Francisco. *Forest*, *15*.

Congress.Gov. (2019). *Foundations for evidence-based policymaking Act of 2018.* (https://www.congress.gov/bill/115th-congress/house-bill/4174)

Cook, P. J., & Ludwig, J. (2006). The social costs of gun ownership. *Journal of Public Economics*, *90*(1-2), 379–391.

Deadman, D., et al. (2001). *Forecasting trends in recorded crime.* University of Leicester.

Del Negro, M., & Otrok, C. (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. *FRB of New York Staff Report*(326).

Devi, J. V., & Kavitha, K. (2021). Automating time series forecasting on crime data using RNN-LSTM. *International Journal of Advanced Computer Science and Applications*, *12*(10).

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Drake, E. K., Aos, S., & Miller, M. G. (2009). Evidence-based public policy options to reduce crime and criminal justice costs: Implications in Washington State. *Victims and Offenders*, *4*(2), 170–196.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.

FBI. (2019). *2019 Crime in the United States.* (https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/violent-crime)

FBI. (2022a). *NICS firearm checks: month/year.* (https://www.fbi.gov/file-repository/nics_firearm_checks_-_month_year.pdf/view)

FBI. (2022b). *Quarterly Uniform Crime Report.* (https://crime-data-explorer.fr.cloud.gov/pages/explorer/crime/quarterly)

Feng, M., Zheng, J., Han, Y., Ren, J., & Liu, Q. (2018). Big data analytics and mining for crime data analysis, visualization and prediction. In *International conference on brain inspired cognitive systems* (pp. 605–614). Springer.

Fox, J. A., & Swatt, M. L. (2009). Multiple imputation of the supplementary homicide reports, 1976–2005. *Journal of Quantitative Criminology*, *25*(1), 51–77.

Friedman, M., Grawert, A. C., & Cullen, J. (2017). *Crime trends, 1990-2016.* Brennan Center for Justice at New York University School of Law New York, NY.

Gamma, A., Schleifer, R., Weinmann, W., Buadze, A., & Liebrenz, M. (2016). Could Google Trends be used to predict methamphetamine-related crime? An analysis of search volume data in Switzerland, Germany, and Austria. *PLoS ONE*, *11*(11).

Gardner Jr, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, *31*(10), 1237–1246.

Giffords Law Center. (2022). *Surging in Gun Violence.* (https://giffords.org/lawcenter/report/surging-gun-violence-where-we-are-how-we-got-here-and-where-we-go/)

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting*, *19*(4), 579–594.

Gramlich, J. (2022). What the data says about gun deaths in the US. *Pew Research Center*.

Guns Violence Archive. (2022). *Guns Violence Archive database.* (https://www.gunviolencearchive.org/query)

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen netzen. *Diploma, Technische Universität München*, *91*(1).

Hoffman, C. Y., Hinkle, J. C., & Ledford, L. S. (2021). Beyond the "Ferguson Effect" on crime: Examining its influence on law enforcement personnel. *CJC Publications. 36*.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–13.

Huang, Y., Xu, C., Ji, M., Xiang, W., & He, D. (2020). Medical service demand forecasting using a hybrid model based on arima and self-adaptive filtering method. *BMC Medical Informatics and Decision Making*, *20*(1), 1–14.

Ipsos. (2022). *FiveThirtyEight/Ipsos 2022 Election Tracking Survey.* (https://www.ipsos.com/en-us/news-polls/FiveThirtyEight-2022-midterm-election)

Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. (2018). Crime analysis through machine learning. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 415–420).

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations.* ICLR.

Lang, M. (2013). Firearm background checks and suicide. *The Economic Journal*, *123*(573), 1085–1099.

Lilliefors, H. W. (1967). On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, *62*(318), 399–402.

Liu, D. C., & Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, *45*(1), 503–528.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Lopes, H. F., Gamerman, D., & Salazar, E. (2011). Generalized spatial dynamic factor models. *Computational Statistics and Data Analysis*, *55*(3), 1319–1330.

Mancik, A. M., Hawk, S. R., Jarvis, J. P., & Regoeczi, W. C. (2021). Evaluating fluctuations in homicide: Crowdsourcing trends and assessing sentiments of change. *Journal of Crime and Justice*, *44*(5), 553–578.

Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443.

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68–78.

McDowall, D. (2002). Tests of nonlinear dynamics in US homicide time series, and their implications. *Criminology*, *40*(3), 711–736.

Meskela, T. E., Afework, Y. K., Ayele, N. A., Teferi, M. W., & Mengist, T. B. (2020). Designing time series crime prediction model using long short-term memory recurrent neural network. *International Journal of Recent Technology and Engineering*, *9(4)*, 402-405.

Mitchell, T. M., & Mitchell, T. M. (1997). *Machine learning* (Vol. 1) (No. 9). McGraw-hill New York.

Morgan, R. E., & Thompson, A. (2021). *Criminal victimization, 2020* (Vol. 4). Bureau of Justice Statistics.

Muthamizharasan, M., & Ponnusamy, R. (2022). Forecasting Crime Event Rate with a CNN-LSTM Model. In *Innovative data communication technologies and application* (pp. 461–470). Springer.

Nasridinov, A., Ihm, S.-Y., & Park, Y.-H. (2013). A decision tree-based classification model for crime prediction. In *Information technology convergence* (pp. 531–538). Springer.

Nau, R. (2014). The mathematical structure of ARIMA models. *Duke University Online Article*, *1*(1), 1–8.

Neumayer, E. (2003). Good policy can lower violent crime: Evidence from a cross-national panel of homicide rates, 1980–97. *Journal of Peace Research*, *40*(6), 619–640.

New York Times. (2022). *A Continuing Drop in Murders.* (https://www.nytimes.com/2022/12/30/briefing/crime-murders-us-decline.html)

OECD. (2020). *How can governments leverage policy evaluation to improve evidence informed policy making ?* (https://www.oecd.org/gov/policy-evaluation-comparative-study-highlights.pdf)

Parker, K. F., Mancik, A., & Stansfield, R. (2017). American crime drops: Investigating the breaks, dips and drops in temporal homicide. *Social Science Research*, *64*, 154–170.

Parkin, W. S., Bejan, V., & Hickman, M. J. (2020). Police, public and community violence: Exploring the relationships between use of deadly force, law enforcement killed, and homicide rates in the United States. *Criminology, Criminal Justice Law and Society*, *21*, 1.

Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, *10*(4), 461–465.

Phillips, D. P., & Hensley, J. E. (1984). When violence is rewarded or punished: The impact of mass media stories on homicide. *Journal of Communication*, *10*(3), 101–116.

Phillips, M. D. (2016). Time series applications to intelligence analysis: a case study of homicides in Mexico. *Intelligence and National Security*, *31*(5), 729–745.

Police Executive Research Forum. (2021). *Survey on 2020 gun crime and firearm recoveries.* (https://www.policeforum.org/criticalissues26jan21)

Pratt, T. C., & Lowenkamp, C. T. (2002). Conflict theory, economic conditions, and homicide: A time-series analysis. *Homicide Studies*, *6*(1), 61–83.

Regoeczi, W. C., Banks, D., Planty, M., Langton, L., & Warner, M. (2014). The nation's two measures of homicide. *Sociology and Criminology Faculty Publications. 124*.

Roman, J. K., & Cook, P. (2021). Improving data infrastructure to reduce firearms violence. *Chicago, IL: NORC at the University of Chicago*.

Rosenfeld, R., & Fornango, R. (2007). The impact of economic conditions on robbery and property crime: The role of consumer sentiment. *Criminology*, *45*(4), 735–769.

Rosenfeld, R., & Fox, J. A. (2019). Anatomy of the homicide rise. *Homicide Studies*, *23*(3), 202–224.

Rosenfeld, R., & Lopez, E. (2020). Pandemic, social unrest, and crime in US cities. *Council on Criminal Justice*.

Royston, P. (1993). A pocket-calculator algorithm for the shapiro-francia test for non-normality: An application to medicine. *Statistics in Medicine*, *12*(2), 181–184.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Sargent, T. J., Sims, C. A., et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In *New methods in business cycle research* (Vol. 1, pp. 145–168). Minneapolis: Federal Reserve Bank of Minneapolis.

Shapiro, S. S., & Francia, R. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, *67*(337), 215–216.

Shoesmith, G. L. (2013). Space–time autoregressive models and forecasting national, regional and state crime rates. *International Journal of Forecasting*, *29*(1), 191–201.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48.

Soybilgen, B. (2020). Identifying US business cycle regimes using dynamic factors and neural network models. *Journal of Forecasting*, *39*(5), 827–840.

Stock, J. H. (1994). Unit roots, structural breaks and trends. *Handbook of econometrics*, *4*, 2739–2841.

Stock, J. H., & Watson, M. W. (1991). A probability model of the coincident economic indicators. In *Leading economic indicators: New approaches and forecasting records.* Cambridge University Press.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Strom, K. J., & Smith, E. L. (2017). The future of crime data: The case for the National Incident-Based Reporting System (NIBRS) as a primary data source for policy evaluation and crime analysis. *Criminology & Public Policy*, *16*(4), 1027–1048.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(1), 54–59.

Vezhnevets, A., & Barinova, O. (2007). Avoiding boosting overfitting by removing confusing samples. In *European Conference on Machine Learning* (pp. 430–441).

Wallace, L. N. (2015). Responding to violence with guns: Mass shootings and gun acquisition. *The Social Science Journal*, *52*(2), 156–167.

Wallis, K. F. (1986). Forecasting with an econometric model: The 'ragged edge'problem. *Journal of Forecasting*, *5*(1), 1–13.

Wheeler, A. P., & Kovandzic, T. V. (2018). Monitoring volatile homicide trends across US cities. *Homicide Studies*, *22*(2), 119–144.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324–342.

Yadav, R., & Sheoran, S. K. (2018). Crime prediction using auto regression techniques for time series data. In *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)* (pp. 1–5). IEEE.

Yim, H.-N., Riddell, J. R., & Wheeler, A. P. (2020). Is the recent increase in national homicide abnormal? Testing the application of fan charts in monitoring national homicide trends over time. *Journal of Criminal Justice*, *66*.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*(2).

Appendix A: Description of methods used for comparison

AI and ARIMA models

Here, we present a brief description of the AI and ARIMA models that are used in the main manuscript to assess the performance of the DFM. First, we offer some intuition behind tree-based methods (random forest, RF and gradient boosting trees, GBOOST) and then turn our attention to long short-term memory neural networks (LSTM) and the ARIMA model.

RF. Tree-based methods (Breiman, 2001) involve the segmentation/partition of the predictor space into a finite set of regions by minimizing a given loss function. Toward improved robustness, RF and GBOOST produce multiple trees that are combined to yield a single consensus prediction. More specifically, RF builds multiples decision trees on bootstrapped training samples and averages them to obtain the prediction. To avoid the creation of correlated trees that would grow the variance of the final predictor, the algorithm considers a random sample of m predictors among the full available set at each split of any tree. From cross-validation of the whole sample, we set $m = 48$, the number of bootstrapped trees equal to 200, and the minimum node size equal to 5.

GBOOST. GBOOST does not involve bootstrap sampling. Instead, the trees are grown sequentially and each of them is fit to a modified version of the original dataset. Specifically, the algorithm has three parameters: the number of trees B, the shrinkage parameter λ, and the number of splits d in each tree, which controls the complexity of the boosted ensemble. The algorithm implements the following steps:

1. Initialize the algorithm with a null tree ($\hat{f} = 0$) and residuals equal to the observations of the dependent variable ($r_i = y_i$, for all i in the training set).
2. For $b = 1, 2 \ldots, B$, create multiple trees by
   (a) fitting a tree $\hat{f}^b$ with d splits;
   (b) update the tree by adding a shrunken version of the new tree,
   $$\hat{f} \leftarrow \hat{f} + \lambda \hat{f}^b,$$
   (c) update the residuals
   $$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i),$$
   where x_i is the value of the independent variable in the training set.
3. Output the boosted model
$$\hat{f} = \sum_{b=1}^{B} \lambda \hat{f}^b.$$

After cross validation with the whole sample, we choose $B = 20$, $\lambda = 0.5$, and $d = 3$.

LSTM. The long short-term memory networks, LSTM, are a particular case of recurrent neural networks (RNN) that were proposed by Rumelhart et al. (1986) to process variable length sequences of inputs. The LSTM network is created with different layers where a back-propagation algorithm (typically through the stochastic gradient descent procedure) is used to estimate the network parameters during the training process. In contrast with standard RNN, the LSTM networks do not suffer from the so-called "long short-term memory problem", as pointed out in Hochreiter (1991), from which during back-propagation the gradient might vanish or explode and therefore artificially affecting the training process. In our case, the *adam* optimizer was used, which is an extension of the standard stochastic gradient descent procedure, to minimize the mean absolute error (MAE). Moreover, the networks composition were first trained with the whole sample, minimizing the error the configuration with two LSTM layers, with 50 and 20 neurons respectively. A dropout layer was included after each LSTM layer, with a dropout rate of 0.5 to avoid overfitting, and a final time-distributed layer was added at the end. The training was performed through 10 epochs.

ARIMA. The univariate autoregressive model assumes that the dynamics of a time series, y_t, is driven by its own past. In particular, assuming the number of lag-observations to be equal to p and mean value to be equal to μ, the time-series can be written as

$$y_t = \mu + a(L)(y_{t-1} - \mu) + \varepsilon_t, \tag{13}$$

where L is the lag-operator, $a(L) = (a_1 + a_2 L + ... + a_p L^{p-1})$, and the errors ε_t are i.i.d. following $\mathcal{N}(0, \sigma^2)$. By choosing the order of the moving average to be equal to q and the degree of differencing to be equal to d, we obtain ARIMA(p, d, q), defined as

$$(1 - a_1 - a_2 L - \dots - a_p L^{p-1})((1-L)^d y_t - \mu) = (1 + \theta_1 + \theta_2 L + \dots + \theta_q L^q)\varepsilon_t, \qquad (14)$$

For the time-series of homicides with firearm, we used an ARIMA(3,0,3) as a parsimonious, yet descriptive model; the estimated coefficients and their respective p-values are presented in Table 6.

	a_1	a_2	a_3	θ_1	θ_2	θ_3
Value	-0.09	0.12	0.95	0.68	0.64	-0.35
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$

Table 6: Estimated coefficients for the ARIMA(3,0,3) for homicides with firearm. The intercept was not statistically significant and the resulting AIC was −7510.

Alternative autoregressive models

In addition to the ARIMA examined in the main manuscript, we considered two alternative autoregressive models: the Holt-Winters' method, initially proposed by Holt (1957) and extended to account for seasonality by Winters (1960), and the vector autoregressive (VAR) model, the natural extension of univariate models to multivariate settings Sims (1980).

Holt-Winters. The Holt-Winters' method uses exponential smoothing to encode the values of a time-series, y_t, as a combination of three components: the level (l_t), trend (b_t), and seasonal factor (s_t). The three components are determined using smoothing methods as follows:

$$\begin{aligned} l_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}, \end{aligned} \qquad (15)$$

where α, β, and γ are the corresponding smoothing parameters, and m denotes the frequency of the seasonality. The parameters of the model can be estimated by minimizing the residual sum of squares (in our case, $\alpha = 0.3, \beta = 0.1$, and $\gamma = 0.1$), and l_t, b_t, and s_t can be obtained by simply initializing at the first time-step. For forecasting h time-steps in the future, we implement the following:

$$y_{t+h} = l_t + hb_t + s_{t+h-m(k+1)}, \qquad (16)$$

where k is the integer part of $(h-1)/m$.

VAR. Let Y_t be a vector of n stationary time- series $y_{1,t}, \dots, y_{n,t}$ with expected value ν. The reduced form of a VAR model with p lags, VAR(p), is

$$Y_t = \nu + C(L)(Y_{t-1} - \nu) + u_t, \qquad (17)$$

where $C(L) = (C_1 + C_2 L + \dots + C_p L^{p-1})$ is the matrix of lag polynomials and the errors are serially uncorrelated with zero mean and covariance matrix Ω. In our database, the in-sample estimation is performed with only four variables: homicides with firearm, media output on "homicide" + "shot", homicides from three cities, and background checks, while deaths in incidents with guns from the Guns Violence Archive and provisional quarterly annualized estimate rate for homicides with firearm from the CDC are omitted because they start very late The model can be estimated by ordinary least squares in each equation and the lag-length is typically chosen by using model selection criteria (in our case, we used AIC and obtained $p = 3$). In-sample parameter estimates are shown in Table 7.

Appendix B: Comparison with alternative autoregressive models

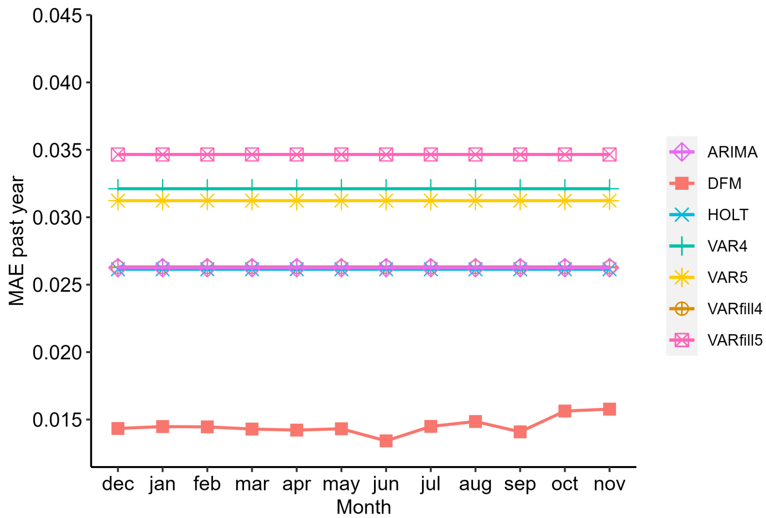
Here, we compare the DFM with Holt-Winters' method and VAR.Similar to ARIMA, the Holt-Winters' method yields the same forecast for every estimation period from January to December as new data for homicides with firearm appear in December every year. With the VAR model, we implemented two different

	HF_1	MOH_1	$H3_1$	BCs_1	HF_2	MOH_2	$H3_2$	BCs_2	HF_3	MO_3	$H3_3$	BCs_3	ν
HF	$4{\cdot}10^{-1}$	$2{\cdot}10^{-4}$	$2{\cdot}10^{2}$	$2{\cdot}10^{-8}$	$3{\cdot}10^{-1}$	$2{\cdot}10^{-5}$	$-7{\cdot}10^{2}$	$-1{\cdot}10^{-9}$	$1{\cdot}10^{-1}$	$1{\cdot}10^{-4}$	$1{\cdot}10^{2}$	$1{\cdot}10^{-2}$	$6{\cdot}10^{-10}$
MOH	$2{\cdot}10^{1}$	$3{\cdot}10^{-1}$	$1{\cdot}10^{5}$	$-8{\cdot}10^{-6}$	$1{\cdot}10^{-2}$	$9{\cdot}10^{-2}$	$-3{\cdot}10^{5}$	$7{\cdot}10^{-6}$	$4{\cdot}10^{-1}$	$3{\cdot}10^{-3}$	$3{\cdot}10^{5}$	$-3{\cdot}10^{-7}$	$-3{\cdot}10^{1}$
H3	$2{\cdot}10^{-6}$	$9{\cdot}10^{-9}$	$2{\cdot}10^{-1}$	$2{\cdot}10^{-12}$	$1{\cdot}10^{-5}$	$-2{\cdot}10^{-9}$	$2{\cdot}10^{-1}$	$-2{\cdot}10^{-12}$	$-1{\cdot}10^{-5}$	$2{\cdot}10^{-8}$	$3{\cdot}10^{-1}$	$7{\cdot}10^{-13}$	$-2{\cdot}10^{-6}$
BCs	$1{\cdot}10^{6}$	$2{\cdot}10^{3}$	$6{\cdot}10^{9}$	$8{\cdot}10^{-1}$	$-1{\cdot}10^{6}$	$4{\cdot}10^{2}$	$-6{\cdot}10^{9}$	$-2{\cdot}10^{-1}$	$1{\cdot}10^{6}$	$2{\cdot}10^{3}$	$-1{\cdot}10^{10}$	$2{\cdot}10^{-1}$	$-4{\cdot}10^{-6}$

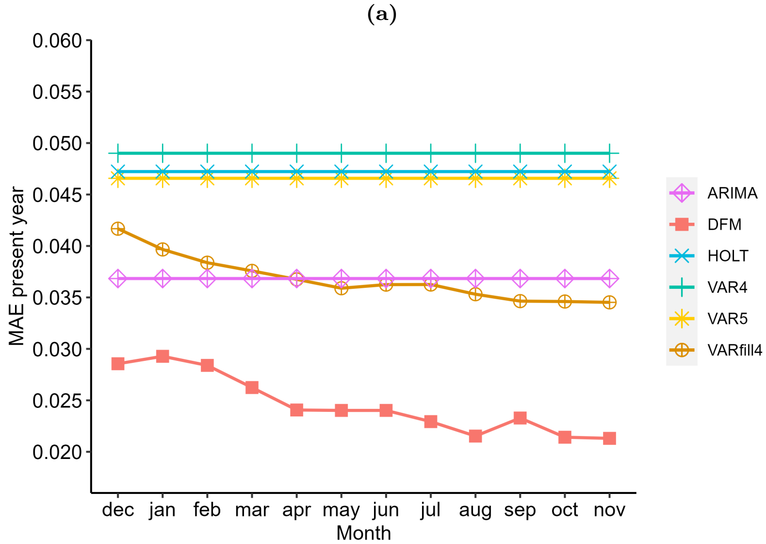
Table 7: Estimated coefficients for the VAR(3) model. HF refers to monthly homicides with firearm from CDC, H3 to monthly homicides in the three cities, MOH to monthly media output with "homicide"+"shot", and BCs to monthly background checks. Subscript refers to the lag of the variable in the model.

approaches for prediction. In the first, we performed an iterative forecast from the last time period where all observations for all the four variables were available and we refer to these results as VAR4. For predictions after 2017, there is the possibility to include deaths in incidents with guns from the Guns Violence Archive; the VAR model accounting for this extra variable is referred to as VAR5. In the second approach, which we refer to as VAR "fill", we performed a one-step-ahead forecast from the period where all observations were available. For the next period, if some variables became available, we disregarded the forecast of those variables and used observations in their place. In this case, the forecasts computed with the four original variables are called VARfill4, while those that include also deaths in incidents with guns from the Guns Violence Archive are referred to as VARfill5.

The results of the out-of-sample analysis are displayed in Figure 5. As in the case of AI models, the DFM was the only model with better performance than the ARIMA benchmark model. Holt-Winters and VARfill4 performed similarly to the ARIMA model in the previous years, while just VARfill4 did it in the present years. Finally, VAR4, VAR5, and VARfill5 predictions perform poorly.



(a)



(b)

Figure 5: Out-of-sample pseudo real-time comparison between DFM and other autoregressive models in the period from January 2008 to December 2020, in terms of MAE: (a) previous and (b) present years. Each point represents an MAE value for either the entire year preceding the month at which the prediction is made (a) or the entire year when the prediction is made (b). DFM (red, ■), VAR with four variables and "fill" schema (orange, ⊕), VAR with four variables (turquoise, +), VAR with five variables and "fill" schema (pink, ⊠), VAR with five variables (yellow, $\ast$), Holt-Winters' method (light blue, ×), and ARIMA(3,0,3) (violet, ⟐). In (b), VAR predictions with five variables with "fill" scheme are not shown for their higher range. VAR and VAR models with "fill" scheme do not contain predictions for years 2013 and 2008, respectively, due to their outlier performance.