

# Detroit Crime and Unemployment Time Series Analysis

## Abstract

Detroit monthly violent and property crime counts (1990–2025) and Wayne County monthly unemployment rates were analyzed to describe long-run crime structure and test whether the unemployment–crime relationship changed after the Covid-19 pandemic. Time series plots showed pronounced downward trends and clear annual seasonality in both crime series, along with greater variability when crime levels were higher. Guided by the sample ACF and PACF, we fit low-order seasonal ARMA models with exogenous covariates to log-transformed crime counts. For violent crime, a SARIMAX(1, 0, 1)  $\times$  (1, 0, 1)<sub>12</sub> model with a linear trend and Covid period indicators provided an adequate fit; residuals were largely uncorrelated and approximately normal, with minor tail deviations near the post-Covid cutoff. Property crime was similarly captured by a low-order seasonal model. In both series, Covid indicators explained meaningful level shifts, while unemployment was not significant after accounting for trend and seasonal dependence. Overall, the results emphasize persistence, seasonality, and structural disruptions in crime dynamics, as well as the importance of time-series methods that properly account for these features.

## 1 Introduction

Crime remains a major concern for Detroit residents, with a substantial share of city resources devoted to public safety (Ortiz 2026). Detroit’s total crime rate in 2024 (6,086.6 per 100,000 residents) was 187.2% higher than the national average, ranking third among the 30 largest U.S. cities (Petrino 2025). The city also faces severe economic disadvantage: median household income is far below the national average, and roughly one third of residents live in poverty (U.S. Census Bureau, n.d.a, n.d.b). Prior research frequently links adverse socioeconomic conditions to higher levels of violent and property crime, though evidence on the relationship between unemployment and crime remains mixed (Stucky, Payton, and Ottensmann 2015). The Covid-19 pandemic introduced a major social and economic shock that altered crime patterns in many U.S. cities. While crime reports declined sharply during the initial lockdown period in early 2020 (Abrams 2021), it remains unclear whether crime subsequently returned to pre-pandemic levels or whether persistent changes emerged.

This analysis aimed to answer three specific questions: (i) whether Detroit crime exhibits long-run trends and cyclical structure, (ii) how patterns shifted during and after the Covid-19 pandemic, and (iii) whether contemporaneous or lagged unemployment rates provide additional explanatory power.

## 2 Data

Data reported by the Detroit Police Department from January 1990 to December 2025 was obtained from the official website of the Federal Bureau of Investigation (FBI) (Federal Bureau of Investigation 2025). Data is split into groups as violent (homicide, rape, robbery, aggravated assault) or property (arson, burglary, larceny-theft, and motor vehicle theft). There was a federal shift in the reporting system on January 1st, 2021 and more details about this change are provided in the supplementary material. However, it is not anticipated to affect the analysis. In addition, Wayne County unemployment data from January 1990 through December 2025 was downloaded from the U.S. Bureau of Labor Statistics' (U.S. Bureau of Labor Statistics 2025). Given that Detroit represents over one third of Wayne County's population and serves as its largest urban and economic center, the county level unemployment rate provides a reasonable estimate of labor market trends affecting the city. The FBI crime and unemployment data for Detroit are largely complete, with a single missing unemployment value in October 2025. This value was imputed using linear interpolation, which estimates the missing observation from adjacent months (IBM Corporation 2026). Additional details on the interpolation procedure are provided in the supplementary material.

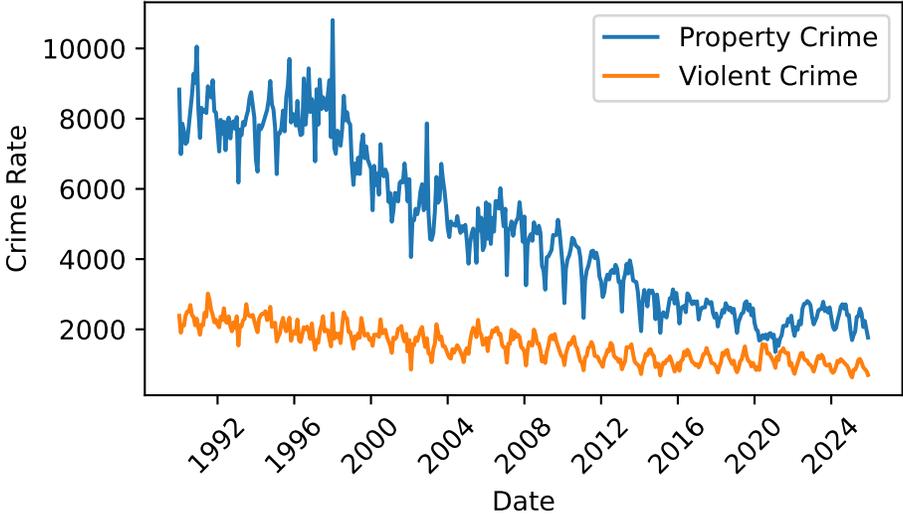


Figure 1: Property and violent crime rates in Detroit over time.

In Figure 1, both property and violent crime in Detroit exhibit a strong long-term downward trend from the early 1990s through 2025, with property crime declining more sharply (from roughly 9,000 to 2,300) than violent crime (from about 2,300 to 900). Property crime remains consistently higher than violent crime and exhibits greater volatility. Both property and violent crime appear to be non-stationary, as they both exhibit seasonal patterns, and they are decreasing over time. There is a small irregular fluctuation within the property crime during 2020, likely due to the lockdown restrictions of the Covid-19 pandemic. The entire crime series exhibits heteroskedastic behavior, with larger fluctuations occurring when crime levels are high. To stabilize the variance, a log transformation was applied in all models.

Unemployment rates during the same time period (Figure 4) exhibit clear seasonal behavior, and there is a dramatic spike in 2020, reflecting the Covid-19 shock. Given the clear disruptions in both crime and unemployment during the Covid-19 period, an indicator was included to distinguish between pre-, during-, and post-pandemic phases to account for this structural break. The pre-Covid period was defined as prior to March 2020, the Covid period as March 2020 through June 2022, and the post-Covid period as July 2022 onward. This indicator allows pandemic-related shifts—driven by lockdowns, remote work, and temporary business closures—to be modeled explicitly as an exogenous influence in subsequent analyses. After plotting the autocorrelation function (ACF) and partial autocorrelation function (PACF) for violent crime and property crime for pre and post-Covid in Figure 2, the pattern observed in the data noticeably changes.

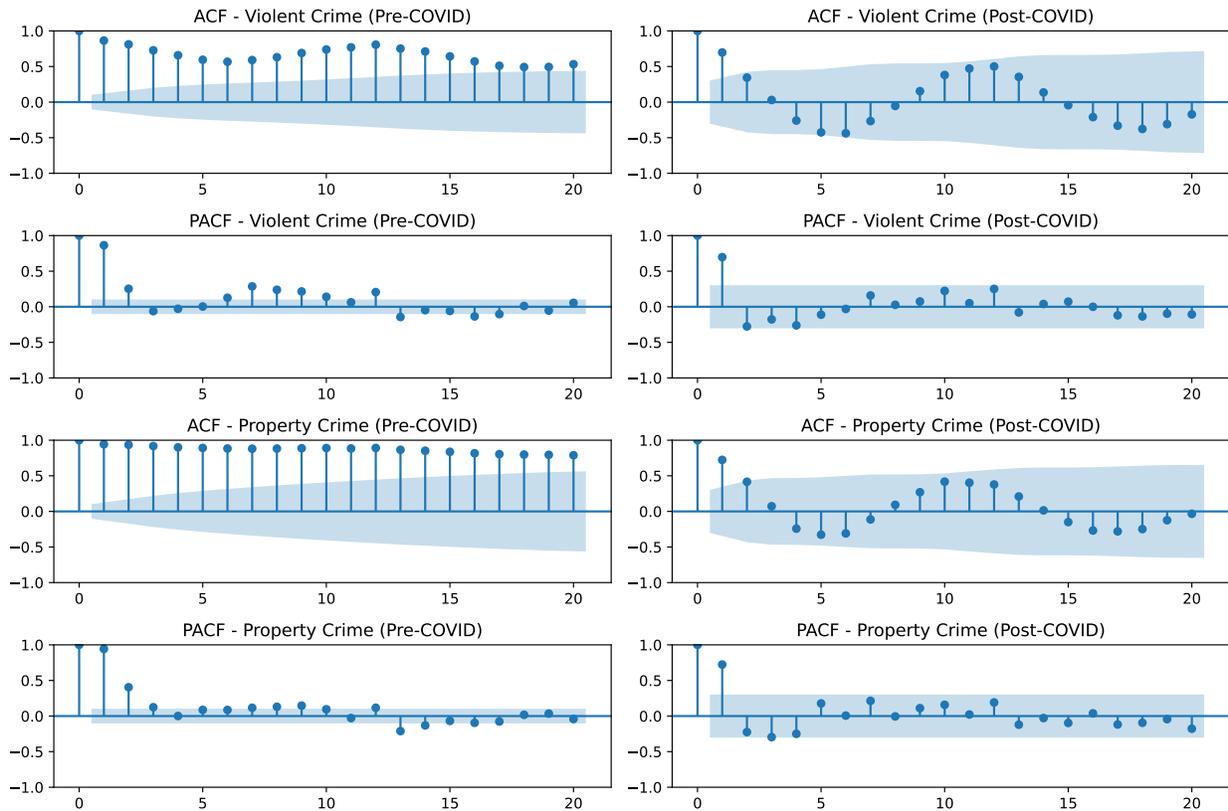


Figure 2: Property and violent crime ACF and PACF Pre-Covid and Post-Covid.

The pre-Covid ACFs for both violent and property crime display very slow decay and remain strongly positive across lags, suggesting potential non-stationarity consistent with patterns observed in the initial time series plot. In contrast, the post-Covid ACFs for both violent and property crime decay much more quickly, oscillating between positive and negative values, showing more of a cyclical behavior over the shorter sample. All PACFs show a large spike at lag 1 with remaining lags mostly small, suggesting that once trend and structural shifts are accounted for, the remaining dependence may be best captured through a low order autoregressive component.

### 3 Methods

**SARIMAX with lagged unemployment** In order to identify associations between crime and unemployment before, during, and after the Covid-19 pandemic, unemployment was added as an exogenous variable. The cross correlation function (CCF) between unemployment and property/violent crime were evaluated to assess whether unemployment should be modeled with a lag.

**SARIMAX models for Violent and Property Crime** Due to the seasonal nature of violent and property crime incidents in Detroit in the initial data exploration, using the seasonal extension of ARMA models, SARIMA, was reasonable. The time series for crime in Detroit diverged during the Covid-19 pandemic. Thus, appropriate models for the violent crime and property crime data would account for pre-Covid, post-Covid, and Covid seasons.

Pre is the reference category for dates before March 1, 2020. The indicator variables for the SARIMAX model are

$$D_{\text{during},t} = I(03/01/2020 \leq t < 07/01/2022) \quad D_{\text{post},t} = I(t \geq 07/01/2022)$$

The time plots also suggested that a linear trend with an intercept was appropriate to model these data. Let  $\phi$  and  $\Phi$  denote the non-seasonal and seasonal auto-regressive terms, respectively. Similarly, let  $\theta$  and  $\Theta$  denote the non-seasonal and seasonal moving average terms, respectively. Because the ACF for violent and property crime before and after Covid showed peaks every 12 lags, this was chosen as the period (Lecture 6, Ionides 2026).

$$\phi(B) \Phi(B^{12}) (Y_n - \mu - \delta n - \beta_1 D_{\text{during},n} - \beta_2 D_{\text{post},n} - \beta_3 U_n) = \theta(B) \Theta(B^{12}) \varepsilon_n \quad (1)$$

with polynomials

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi(B^{12}) &= 1 - \Phi_1 B^{12} - \Phi_2 B^{24} - \dots - \Phi_P B^{12P} \\ \theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \\ \Theta(B^{12}) &= 1 + \Theta_1 B^{12} + \Theta_2 B^{24} + \dots + \Theta_Q B^{12Q} \end{aligned}$$

where  $U_n$  is the time series for unemployment,  $\mu$  is the intercept of the equation for crime pre-Covid, and  $\varepsilon_n$  is the white noise process. The mean can vary across the three time periods, but this SARIMAX model did not allow for different trend, seasonal, and non-seasonal terms.

### 4 Results

**Unemployment and Crime** Cross-correlation functions (CCF) (Figure 6) were used to explore potential lead-lag relationships between unemployment and both types of crime. The unemployment series was prewhitened as described in Shumway and Stoffer (2025) pages 34-35 to prevent spurious associations from being identified. The fitting procedure (assisted by generative AI OpenAI (2026a)), plots, and a detailed description of the results are provided in the supplementary materials. In both cases, the majority of lags from -24 to 24 fell within approximately a band of -0.1 to 0.1.

The strongest correlation of 0.15-0.25 was observed at lag 0, so contemporary unemployment was included as an exogenous predictor in each model. No lagged unemployment values were used.

**Model Fitting** Because the sample ACF and time series plots for both violent and property crime show clear seasonality and negative linear trends, the selection of SARIMA models was restricted to  $(p, 0, q) \times (P, 0, Q)_{12}$  with linear trends and intercepts. Two exogenous variables were included: (1) an indicator of time relative to the Covid-19 pandemic (“pre” (reference category), “during”, and “post”) and (2) contemporary unemployment rates. To restrict model checking, AIC values were evaluated for models ranging from SARIMAX(0, 0, 0)  $\times$  (0, 0, 0)<sub>12</sub> to SARIMAX(3, 0, 3)  $\times$  (3, 0, 3)<sub>12</sub>. Within the restricted low-complexity parameter space, the 50 models with the lowest AIC values all had orders satisfying  $p, q, P, Q \leq 2$ . These models and their AIC values are included in the supplementary material (Table 2, Table 5).

**Violent Crime** Models were tested in order of decreasing AIC (and excluding models without seasonal terms). The model selected for the violent crime time series was SARIMAX(1, 0, 1)  $\times$  (1, 0, 1)<sub>12</sub> with an AIC of -57.120. Nested models such as SARIMAX(1, 0, 0)  $\times$  (0, 0, 1)<sub>12</sub> (AIC: -407.197) and SARIMAX(1, 0, 1)  $\times$  (0, 0, 1)<sub>12</sub> (AIC: -139.820) had non-unit roots, but further investigation of the sample ACF of residuals confirmed that these models did not appropriately capture seasonality. AIC comparisons and residual analysis between the three models are included in Supplementary material (Figure 8).

**Property Crime** Similarly, model selection for the property crime time series prioritized checking models with low relative AIC and numerical stability (Table 5, Table 7). However, the AIC table revealed clear indications of likelihood optimization issues in the AIC calculations. For instance, the models with the lowest AIC were MA(2) (AIC: -620.069) and MA(1) (AIC: -585.993), which contain no seasonal terms. The time plot of the property crime time series clearly appeared to be seasonal, and thus, the visual characteristics of the time series led us to choose a different model than those suggested by the AIC table. SARIMAX (1, 0, 1)  $\times$  (1, 0, 0)<sub>12</sub> (AIC: 60.127) was found to be an appropriate model for the property crime time series.

**Residual Analysis** To check the model fit, the residuals of the violent crime model were plotted on a time plot and the sample ACF of the residuals revealed no clear patterns. The sample ACF (Figure 7) of residuals showed that 2 of 40 lags fell outside of the band that measures chance variation (Lecture 2, Ionides 2026). Therefore, based on the data, there is insufficient evidence to suggest that the residuals of the model were not iid. From this analysis, the SARIMAX(1, 0, 1)  $\times$  (1, 0, 1)<sub>12</sub> model captured the trend in the violent crime data and the noise terms can be modeled as independent and identically distributed (Lecture 2, Ionides 2026). The QQ plot suggests that the residuals can be modeled as approximately normally distributed, with deviations in the tails (Figure 3). In particular, there is one large negative residual corresponding to July 1, 2022, which coincides with the cutoff date for the Covid indicator variable. From this analysis, it is possible that the model did not capture the gradual shift in violent crime incidents between the Covid and post-Covid periods (OpenAI 2026b).

The sample ACF of the residuals for the property crime model showed similar results to the violent crime model, suggesting that the SARIMAX(1, 0, 1)  $\times$  (1, 0, 0)<sub>12</sub> model appropriated modeled the property crime data and the residuals could be modeled as iid (Figure 10). The QQ plot in Figure 11

revealed a similar deviation from normality to Figure 3, with the smallest residual occurring on July 1, 2022 and the largest residual occurring on July 1, 2023 (Figure 11).

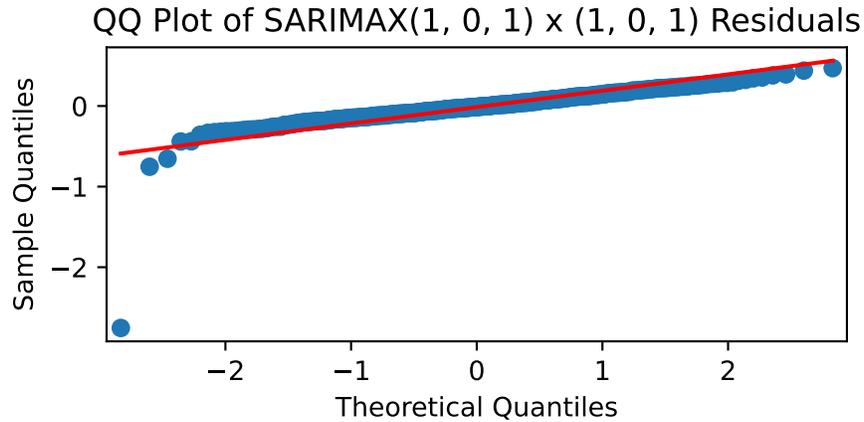


Figure 3: QQ Plot of SARIMAX(1, 0, 1) x (1, 0, 1) Residuals

### Comparison of Coefficients Estimates for Violent vs Property Crime

Parameter	Violent $(1, 0, 1) \times (1, 0, 1)_{12}$	Property $(1, 0, 1) \times (1, 0, 0)_{12}$
$\beta_0$	0.2733	0.3462
$\delta$	-0.0003	-0.0002
$\beta_{\text{during}}$	-0.4206	-0.9669
$\beta_{\text{post}}$	3.0096	3.2109
$\beta_{\text{unemployment}}$	-0.0043	0.0104
$\phi_1$	0.9082	0.9095
$\theta_1$	0.1292	0.0814
$\Phi_1$	0.6541	0.5702
$\Theta_1$	-0.3388	—
$\sigma^2$	0.0486	0.0511

**Drift** The drift for both models was approximately 0.0003. Since the number of incidents per month was log transformed for both violent and property crime, the drift term for violent crime suggested that there was an expected decrease of  $|(e^{-0.0003} - 1) \times 100|\% \approx 0.03\%$  in the expected number of violent incidents per month, holding other variables constant. Similarly, the drift term for property crime suggested that there was an expected decrease of  $|(e^{-0.0002} - 1) \times 100|\% \approx 0.02\%$  in the expected number of property incidents per month, holding other variables constant. Annually, this equated to an expected decrease in violent crime by  $|(e^{12 \times -0.0003} - 1) \times 100|\% \approx 0.36\%$  a year, and an expected decrease in property crime by  $|(e^{12 \times -0.0002} - 1) \times 100|\% \approx 0.24\%$  a year.

**AR terms** Both models contained a relatively large non-seasonal AR coefficient ( $\hat{\phi}_1 \approx 0.91$ ) thus both property and violent crime counts (log-transformed) at month  $n$  were heavily related to the log

of incident counts from month  $n - 1$ , suggesting that both time series were highly autocorrelated. The estimates for the seasonal AR terms were relatively large ( $\hat{\Phi}_1 \geq 0.55$ ) for both models as well. This aligned with the annual seasonality seen on the time plot for both violent and property crime. The number of crime incidents (log-transformed) at month  $n - 12$  may be a strong predictor of the log of incident counts at time  $n$  for both property and violent crime.

**MA terms** The non-seasonal MA coefficients were much smaller  $\hat{\theta}_1 < 0.13$ , which means that the log of incident counts from month  $n$  were not heavily associated with the error terms from month  $n - 1$ . Similarly, the seasonal MA terms for the violent crime model was moderately sized ( $|\hat{\Theta}_1| \approx 0.3388$ ). The log of incident counts from month  $n$  were moderately associated with the error terms from month  $n - 12$ . Although the estimated non-seasonal MA(1) coefficient,  $\hat{\theta}_1$ , in the property crime model was relatively small, diagnostics from the nested model excluding this term revealed auto-correlated residuals. Thus, the final model for the property crime time series kept the non-seasonal MA term (Figure 12).

**Unemployment** The models differ in the estimate for the unemployment coefficient. For violent crime, there was an expected *decrease* of  $|(e^{-0.0043} - 1) \times 100|\% \approx 0.4291\%$  in violent crime incidents per month for every one-percent increase in unemployment rate, holding all other variables constant. However, the SARIMAX model for property crime suggested that for every one-percent increase in unemployment rate, there was an expected *increase* of  $|(e^{0.0104} - 1) \times 100|\% \approx 1.045\%$  in property crime incidents per month, holding all other variables constant.

**Crime during Covid** During the Covid-19 pandemic, the average count for violent crime was about  $|(e^{-0.4206} - 1) \times 100|\% \approx 34.33\%$  lower than pre-Covid whereas the average count for property crime per month was about  $|(e^{-0.9669} - 1) \times 100|\% \approx 61.97\%$  lower than pre-Covid, holding unemployment and trend constant.

**Crime post Covid** The post Covid coefficient is much larger than expected and if interpreted naively as a one-period effect, it would represent a 1924% increase from pre-Covid. Further research indicated this could be due to the large AR root ( $\approx 0.95$ ) causing a long run multiplier effect (Buteikis, n.d.). With highly persistent data, (large AR roots), the strong dependence on past values can cause surprisingly large coefficients for permanent indicator variables. The coefficient should be interpreted as reflecting a sustained change in the baseline of the series, not a short-run effect.

Bootstrap simulations of the coefficient estimates are in the supplementary material (Figure 13, Figure 14) and show wide confidence intervals, including 0, for the Covid indicator variables. While the impact of Covid is evident in the time series of both crime and unemployment, the effect is not well detected by the model. In both models, the non-seasonal AR term is close to 1 with small bootstrap confidence intervals excluding 0. Thus, the crime counts from month  $n$  are highly associated with the counts from month  $n - 1$ . Similarly, after accounting for the Covid indicator and seasonal dependence terms, unemployment does not look to be associated with either property or violent crime.

## 5 Conclusions

Across 36 years of property and violent crime data, strong seasonal and long term patterns emerged. For property crime, there was a strong downward trend from the late 1990s to through the mid

2000s. Violent crime also showed a downward trend, although less sharply. This is consistent with slow influences such as improvements in policy and policing practices that can take months or years to fully implement. Both types of crime showed declines during the Covid-19 pandemic, but property crime was more drastic. While the Covid-19 indicator variable did identify post-Covid consequences, the extent of the changes are unclear and need to be explored further. In addition, despite unemployment showing a modest contemporaneous correlation with both types of crime, it did not appear significant in the model. This aligns with many previous findings of a complex and unstable relationship between crime and unemployment (Greenberg 2001). It does not mean crime and unemployment are unrelated, but it does suggest that once long-term trends and seasonality are accounted for, unemployment does not provide much additional explanatory power for monthly crime.

One of the major difficulties faced in fitting a model to this data was the high persistence, leading to autoregressive roots close to one in nearly all assessed models. This is a known phenomenon in analysis of crime data, with crime sometimes being called a unit root process. Crime is also often compared to random walks, although this is theoretically questionable (Greenberg 2001). In addition, the SARIMAX models assume stationarity, while the near unit root behavior in the models indicate this assumption may not be appropriate for the Detroit crime data. More advanced and flexible methods should be considered to best account for the persistence. This project highlights the importance of using time series methods that consider high persistence and structural change when evaluating crime trends. Ignoring these can lead to misleading conclusions about the effectiveness of policies or the role of economic conditions.

## 5.1 Scholarship

Only two previous midterm projects have focused on crime in other major cities: Seattle and Chicago respectively (“The Monthly Amount of Criminal Incidents in Chicago” 2018; “Time Series Analysis on Monthly Crime Numbers in Seattle from 2008 to 2018” 2020). In particular, the Chicago crime project also assessed the relationship between crime and unemployment. However, they spend minimal time discussing the relationship and do not attempt to use unemployment as a predictor in their model, whereas our project goes into greater detail. Both projects only consider data prior to the Covid-19 pandemic, while understanding the impact of Covid-19 is a primary question in our analysis and we explicitly modeled for Covid with exogenous variables in our SARIMAX model.

We also used one other project from the Winter 2025 semester regarding Ireland Covid 19 cases (“Ireland Covid Cases” 2025) and a Winter 2024 project on Federal Unemployment Rates and Interest Rates (“Unemployment and Federal Interest Rate” 2024) as inspiration for some of our analyses. The Ireland Covid project helped us decide how to format our report and we appreciated their attention to current events and trying to use vaccine roll out dates to explain patterns in the data. We explored how meaningful events such as Covid-19 influenced society at a broader level through crime counts. The Unemployment and Interest Rates project led to us assessing the Cross Correlation function, although we took a different approach by using pre-whitening on one of our series whereas their approach used differencing.

## Acknowledgments

This template builds on a template document by [Shao-Ting Chiu](#). AI was used for debugging via Google AI overview. Vscod with the Quarto extension was used for editing.

## Bibliography

- Abrams, David S. 2021. “COVID and Crime: An Early Empirical Look.” *Journal of Public Economics* 194. <https://doi.org/10.1016/j.jpubeco.2020.104344>.
- Buteikis, Andrius. n.d. “05 Regression with Time Series Variables.” [https://web.vu.lt/mif/a.buteikis/wp-content/uploads/2020/04/Lecture\\_05.pdf?](https://web.vu.lt/mif/a.buteikis/wp-content/uploads/2020/04/Lecture_05.pdf?)
- Federal Bureau of Investigation. 2025. “UCR Crime Data Explorer.” 2025. <https://cde.ucr.cjis.gov/LATEST/webapp/>.
- Greenberg, David F. 2001. “Time Series Analysis of Crime Rates.” *Journal of Quantitative Criminology* 17 (4): 291–327. <https://doi.org/10.1023/A:1012507119569>.
- IBM Corporation. 2026. “Time Series Node Missing Value Handling Options — IBM SPSS Modeler Documentation.” 2026. <https://www.ibm.com/docs/en/spss-modeler/saas?topic=options-time-series-node-missing-value>.
- Ionides, Edward. 2026. “Notes for STATS 531, Modeling and Analysis of Time Series Data.” <https://ionides.github.io/531w26/>.
- “Ireland Covid Cases.” 2025. [https://ionides.github.io/531w25/midterm\\_project/project04/blinded.html](https://ionides.github.io/531w25/midterm_project/project04/blinded.html).
- OpenAI. 2026d. “ChatGPT.” 2026. <https://chatgpt.com/c/6998b7f9-68d0-832b-af2b-badf693d6371>.
- . 2026a. “ChatGPT.” 2026. <https://chatgpt.com/c/6998ed18-2e34-8331-b8a9-318b73ba511c>.
- . 2026b. “ChatGPT.” 2026. <https://chatgpt.com/c/6998b7f9-68d0-832b-af2b-badf693d6371>.
- . 2026c. “ChatGPT.” 2026. <https://chatgpt.com/c/6998b7f9-68d0-832b-af2b-badf693d6371>.
- Ortiz, Douglas. 2026. “Crime in the City: Identifying Detroit’s Peer Cities.” January 22, 2026. <https://crcmich.org/crime-in-the-city-identifying-detroit-peer-cities>.
- Petrino, Gene. 2025. “Most Dangerous Cities: 2026 Rankings for the 30 Largest U.S. Cities.” October 10, 2025. <https://www.security.org/resources/most-dangerous-cities/>.
- Shumway, Robert H., and David S. Stoffer. 2025. *Time Series Analysis and Its Applications: With r Examples*. 5th ed. New York: Springer.
- Stucky, Thomas D., Seth B. Payton, and John R. Ottensmann. 2015. “Intra- and Inter-Neighborhood Income Inequality and Crime.” *Journal of Crime and Justice* 39 (3): 345–62. <https://doi.org/10.1080/0735648X.2015.1004551>.
- “The Monthly Amount of Criminal Incidents in Chicago.” 2018. [https://ionides.github.io/531w18/midterm\\_project/project46/mid\\_proj.html](https://ionides.github.io/531w18/midterm_project/project46/mid_proj.html).
- “Time Series Analysis on Monthly Crime Numbers in Seattle from 2008 to 2018.” 2020. [https://ionides.github.io/531w20/midterm\\_project/project46/MT2.html](https://ionides.github.io/531w20/midterm_project/project46/MT2.html).
- “Unemployment and Federal Interest Rate.” 2024. [https://ionides.github.io/531w24/midterm\\_project/project03/blinded.html](https://ionides.github.io/531w24/midterm_project/project03/blinded.html).
- U.S. Bureau of Labor Statistics. 2025. “Local Area Unemployment Statistics: Detroit–Warren–Dearborn, MI.” 2025. <https://data.bls.gov/dataViewer/view/timeseries/LAUCN261630000000003>.

U.S. Census Bureau. n.d.a. “QuickFacts: Detroit City, Michigan (Population Estimates and Other Statistics).” <https://www.census.gov/quickfacts/fact/table/detroitcitymichigan/PST045224>.  
———. n.d.b. “QuickFacts: United States (Population Estimates and Other Statistics).” <https://www.census.gov/quickfacts/fact/table/US/SEX255224>.

# 6 Supplementary material

## 6.1 Supplementary Details of Interpolation of Missing Data

Linear interpolation estimates our missing observation by drawing a line between adjacent known values (September and November 2025 unemployment rates, in this case) and assigning an intermediate value accordingly. Since there was a singular missing value, linear interpolation provides a reasonable approach to preserve the continuity of our series. Linear interpolation works using the formula:

$$y = y_0 + (x - x_0) \frac{(y_1 - y_0)}{(x_1 - x_0)}$$

Where  $(x_0, x_1)$  are the known timepoints, with  $(y_0, y_1)$  as the known values, respectively. In simpler terms, interpolation averages the years since they are adjacently spaced in time.

## 6.2 Supplementary Material for Unemployment Times Series and ACF/PACF

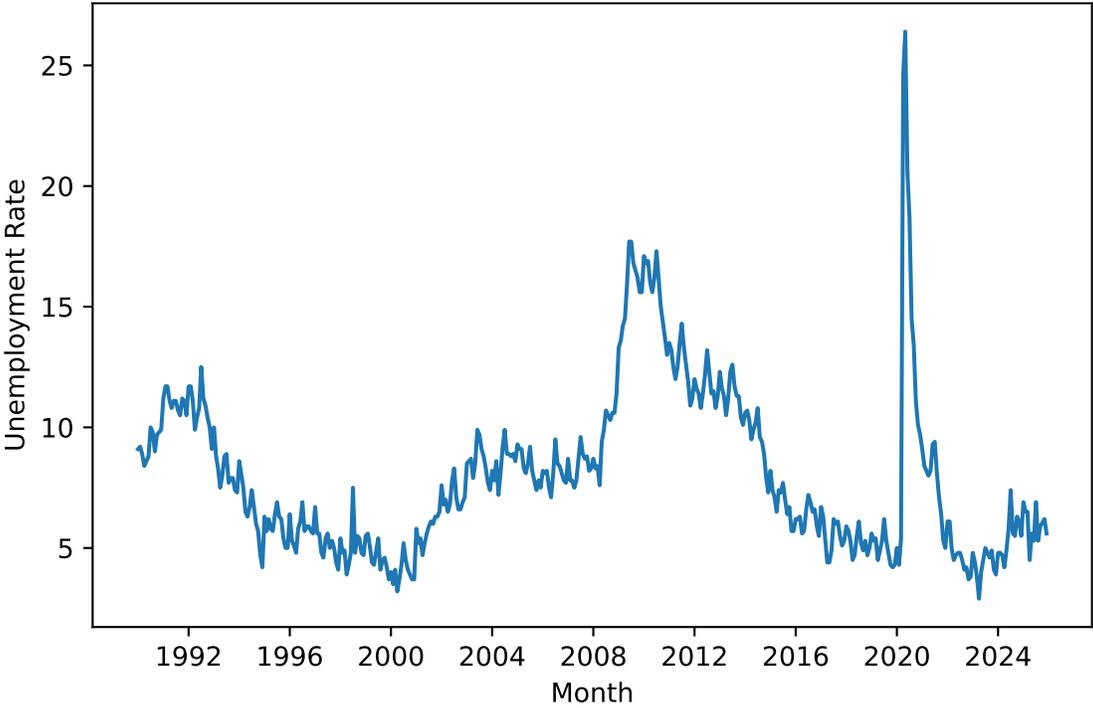


Figure 4: Time Series of Unemployment 1990-2025

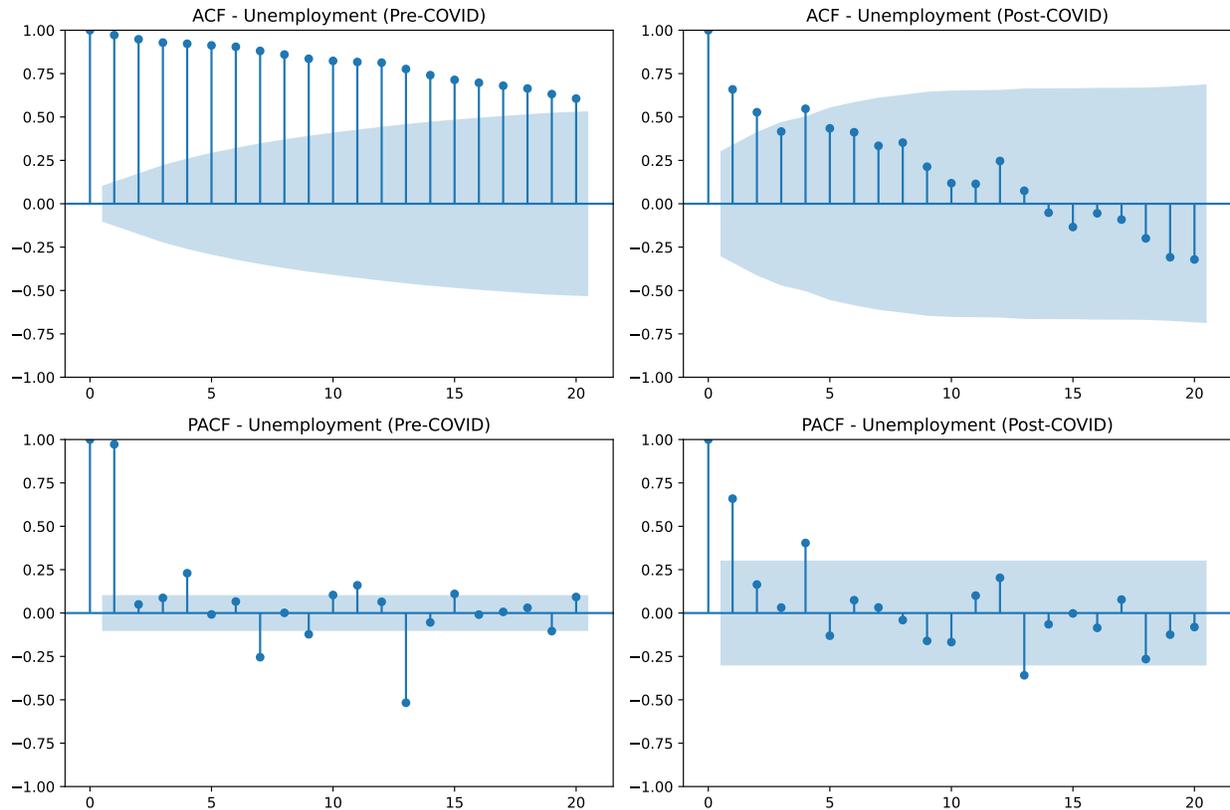


Figure 5: ACF and PACF Plots of Unemployment

From this plot, we can notice that the unemployment rate exhibits clear seasonal behavior over time, with major spikes due to harsh economic downturns. Unemployment rates were elevated during the early 1990s, declined through the early 2000s, and then promptly rose during the Great Recession around 2008-2010, reaching levels around 17%. After falling through the 2010s, unemployment dramatically spiked in 2020, reflecting the Covid-19 shock, before quickly declining again. Outside of these periods, unemployment generally fluctuates around 4-9%, displaying moderate short-term variation. Overall this time series suggests strong trends and seasonality, with temporary disruptions due to economic crises. Based on our plot and our ACFs (below), we did not log transform our data because the variance appears to be relatively stable over time.

### 6.3 Supplementary Material for Cross Correlation of Unemployment and Crime

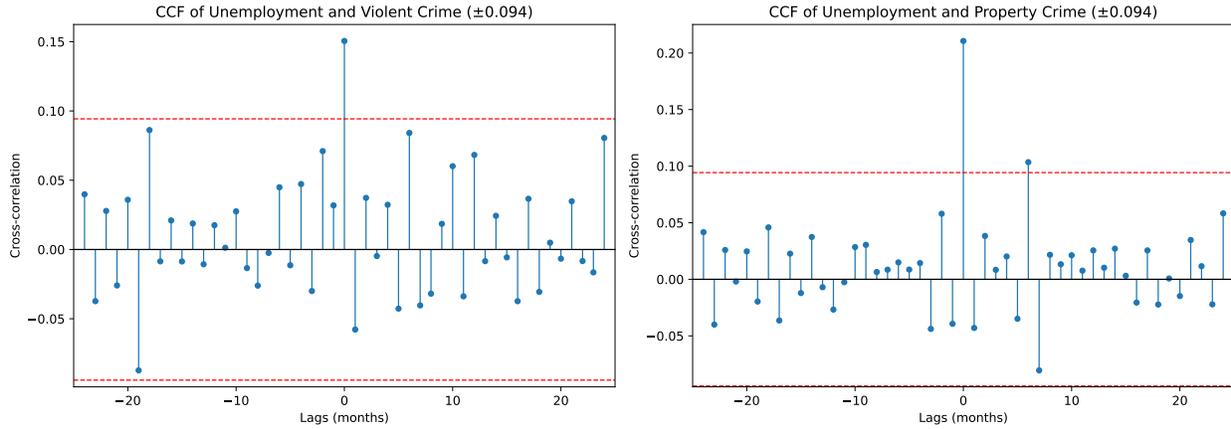


Figure 6: Cross Correlation Function of Unemployment, Property, and Violent Crime

The Cross-Correlation Function (CCF) was created with the help of generative AI (OpenAI 2026a) can be used to visualize the correlation between crime and unemployment at different lags. This helps identify whether changes in unemployment tend to precede or follow changes in crime. To prewhiten unemployment, a  $SARIMA(1, 0, 1) \times (1, 0, 1)_{12}$  model was used to try to explain any existing patterns in the data based on seasonality and prior unemployment. A few of the correlations fall slightly outside the confidence band around 0, meaning this could be a random chance correlation. There is no clear pattern in the correlations across lags. The strongest correlation is approximately 0.20 when lag is 0, indicating a contemporaneous association between higher unemployment and higher violent crime within the same month. This could be due to economic shocks which impact both unemployment and crime in negative ways.

The CCF plot of property crime and unemployment follows a similar pattern. The strongest correlation is approximately 0.25, again at lag is 0, indicating a contemporaneous association is also useful to consider for unemployment and property crimes.

### 6.4 Supplementary Material for Violent Crime

(OpenAI 2026c)

Table 2: AIC for SARIMAX for Monthly Violent Crime

	AR(p)	MA(q)	SAR(P)	SMA(Q)	AIC
0	1	0	2	2	-426.423365
1	1	0	0	1	-407.196520
2	0	1	0	0	-404.888847
3	2	1	2	0	-402.817728
4	1	2	0	2	-383.434984
5	1	2	0	1	-381.828292

Table 2: AIC for SARIMAX for Monthly Violent Crime

	AR(p)	MA(q)	SAR(P)	SMA(Q)	AIC
6	2	0	0	0	-371.138027
7	0	1	1	1	-346.937849
8	0	2	2	1	-317.989259
9	0	0	1	2	-310.594128
10	1	1	2	0	-301.501886
11	0	0	0	0	-293.715478
12	0	2	1	1	-254.752692
13	0	2	2	2	-237.529791
14	0	0	0	2	-228.992487
15	2	0	2	1	-219.110797
16	2	0	1	2	-218.934810
17	0	2	1	0	-209.068433
18	1	1	2	2	-190.267641
19	2	0	2	2	-189.042047
20	0	0	2	2	-185.894552
21	1	0	1	0	-181.537711
22	0	0	1	1	-172.695665
23	1	2	1	0	-171.827063
24	2	1	0	2	-165.876735

Table 3: Full Summary Table of SARIMAX (1, 0, 1) x (1, 0, 1) 12 for Violent Crime

Parameter	Coefficient	Std. Error	<i>p</i> -value
$\beta_0$	0.2733	0.111	0.014
$\delta$	-0.0003	0.000	0.033
$\beta_{\text{during}}$	-0.4206	0.105	0.000
$\beta_{\text{post}}$	3.0096	0.192	0.000
$\beta_{\text{unemp}}$	-0.0043	0.020	0.829
$\phi_1$	0.9082	0.027	0.000
$\theta_1$	0.1292	0.064	0.044
$\Phi_1$	0.6541	0.114	0.000
$\Theta_1$	-0.3388	0.113	0.003
$\sigma^2$	0.0486	0.005	0.000

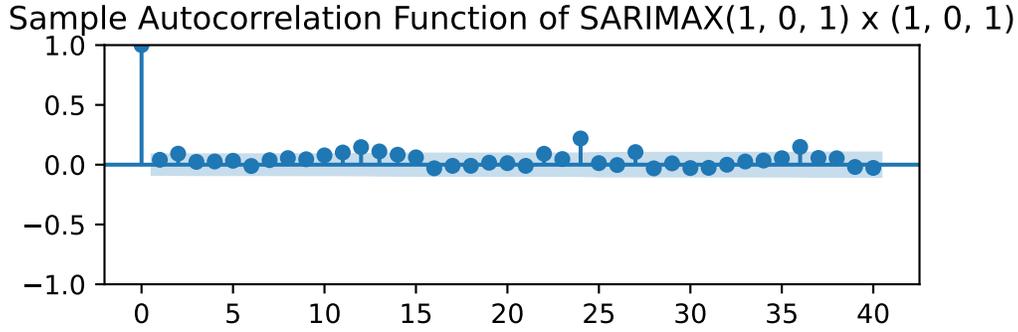


Figure 7: Sample Autocorrelation Function of SARIMAX(1, 0, 1) x (1, 0, 1)

### 6.4.1 Comparison of SARIMAX Models for Violent Crime

SARIMAX(1, 0, 0) × (0, 0, 1)<sub>12</sub> and SARIMAX(1, 0, 1) × (0, 0, 1)<sub>12</sub> were two reasonable models for the time series of violent crime. However, they presented challenges absent from the final model choice SARIMAX(1, 0, 1) × (1, 0, 1).

In terms of numerical instability, the models were similar with roots close to 1.1 with the exception of the non-seasonal MA terms for the SARIMAX(1, 0, 1) × (1, 0, 1) and SARIMAX(1, 0, 1) × (0, 0, 1) models (Table 4).

Table 4: Check for Unit Roots

Model (p,q,P,Q)	AR (non-seasonal)	MA (non-seasonal)	Seasonal AR	Seasonal MA
(1,1,1,1)	1.10111893	7.73816896	1.03600467	1.09439959
(1,1,0,1)	1.08504322	13.72208234	—	1.20004983
(1,0,0,1)	1.06870966	—	—	1.11992932

However, the residual analysis indicated that the nested models did not fully account for the seasonality present in violent crime data whereas residuals of the SARIMAX (1, 0, 1) × (1, 0, 1) model could reasonably be modeled as iid (Figure 8). In other words, the residuals could be described as uncorrelated.

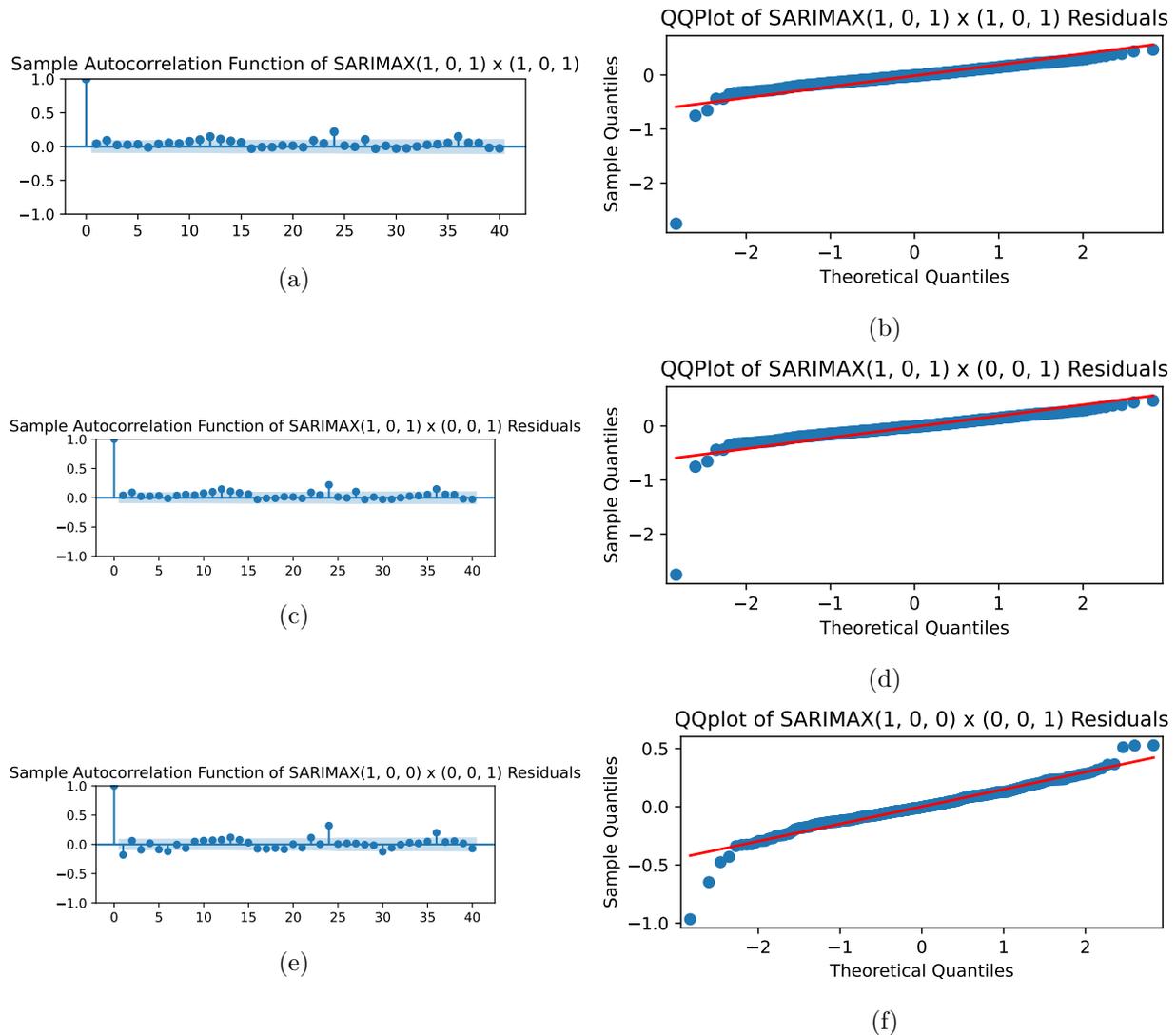


Figure 8: Comparison of sample ACF and QQplots for residuals of SARIMAX(1, 0, 1) x (1, 0, 1)<sup>12</sup> and alternative models

#### 6.4.2 Removing the AR term with a near unit root

Although the non-seasonal AR term has a near unit root, removing this term does not provide a model that sufficiently explains the underlying data.

In particular, the sample ACF suggests that the residuals should not be modeled as uncorrelated or iid (Figure 9). The QQplot also shows severe deviations from normality at the extremes. Thus, SARIMAX(1, 0, 1) x (1, 0, 1) still proved to be a more reasonable model for the violent crime time series.

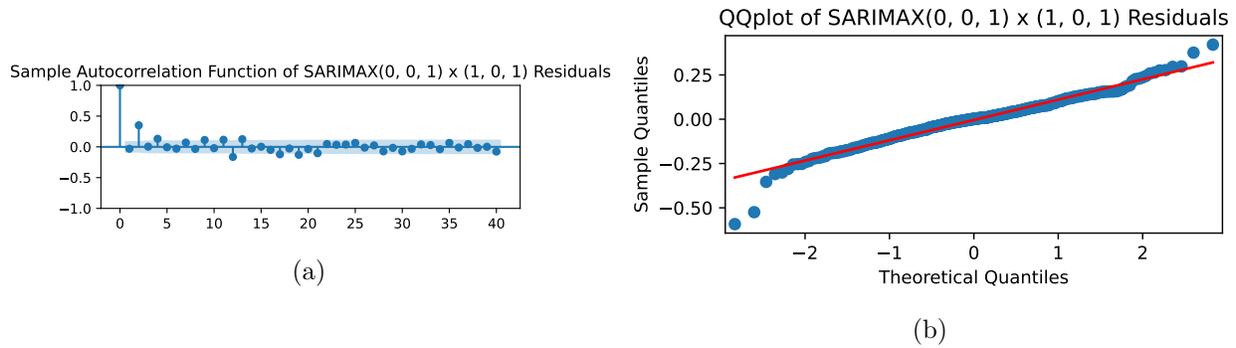


Figure 9: Sample Autocorrelation Function of SARIMAX(0, 0, 1) x (1, 0, 1) Residuals

## 6.5 Supplementary Material for Property Crime

Table 5: AIC for SARIMAX for Monthly Property Crime

	AR(p)	MA(q)	SAR(P)	SMA(Q)	AIC
0	0	2	0	0	-620.069393
1	0	1	0	0	-585.993320
2	1	0	2	0	-514.872524
3	1	0	0	0	-512.089723
4	2	0	0	1	-478.365250
5	0	0	0	0	-475.596104
6	0	2	2	0	-430.619417
7	0	2	1	1	-423.472542
8	0	2	1	0	-410.988575
9	0	0	0	1	-368.070251
10	0	2	2	1	-364.626457
11	0	0	1	0	-363.500137
12	1	0	1	0	-323.329053
13	0	0	0	2	-233.581142
14	1	2	2	0	-219.324080
15	0	0	2	0	-208.162637
16	1	2	1	0	-205.178895
17	1	0	2	2	-202.696587
18	1	0	1	1	-167.161953
19	2	2	0	0	-164.792371
20	2	0	1	2	-160.663216
21	1	2	0	0	-160.212842
22	0	0	1	2	-125.840172
23	2	0	1	1	-116.519278
24	2	2	2	1	-110.586662

Table 6: Full Summary Table of SARIMAX (1, 0, 1) x (1, 0, 0) 12 for Property Crime

Parameter	Coefficient	Std. Error	<i>p</i> -value
$\beta_0$	0.3462	0.109	0.002
$\delta$	-0.0002	0.000	0.497
$\beta_{\text{during}}$	-0.9669	0.081	0.000
$\beta_{\text{post}}$	3.2109	0.207	0.000
$\beta_{\text{unemp}}$	0.0104	0.008	0.169
$\phi_1$	0.9095	0.021	0.000
$\theta_1$	0.0814	0.076	0.286
$\Phi_1$	0.5702	0.016	0.000
$\sigma^2$	0.0511	0.006	0.000

Table 7: Unit Root Checks for Property Crime

Component	Magnitude
Seasonal AR	1.1487
Non-seasonal AR	1.0751
Non-seasonal MA	2.6186

Sample Autocorrelation Function of SARIMAX(1, 0, 1) x (1, 0, 0)

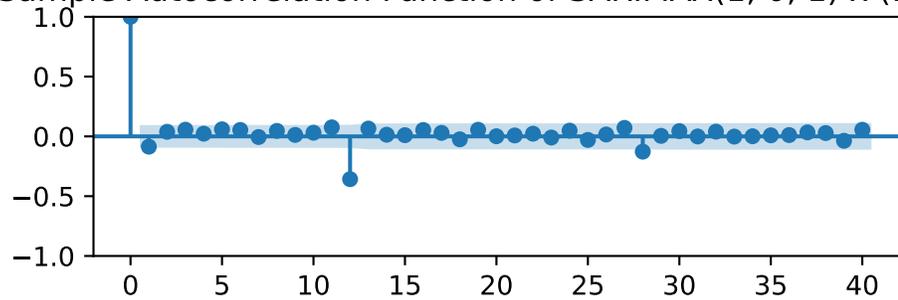


Figure 10

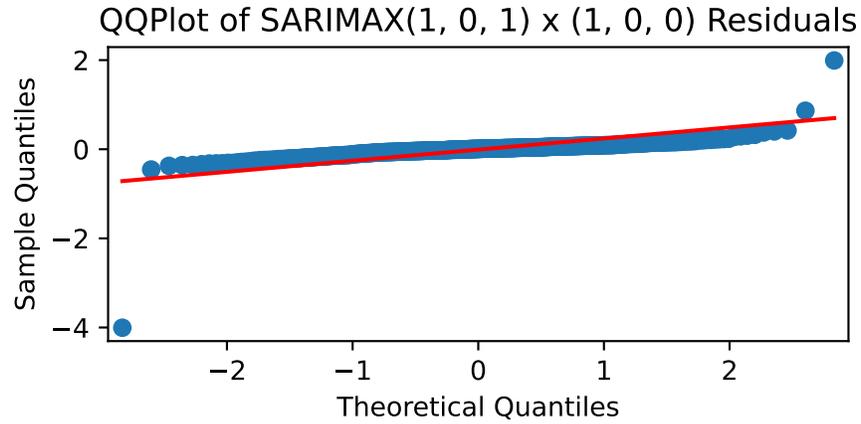


Figure 11

### 6.6 Removing MA term from the property model

The SARIMAX(1, 0, 1)  $\times$  (1, 0, 0)<sub>12</sub> for property crime had a small coefficient for the non-seasonal MA term  $\hat{\theta}_1 = 0.0814$ . For this reason, it was reasonable to consider the smaller SARIMAX(1, 0, 0)  $\times$  (1, 0, 0)<sub>12</sub> model. However, after removing the term, the residuals' sample autocorrelation function showed peaks at seasonal lags (for example, 12, 24, and 36). This suggested that the non-seasonal MA term helped model the seasonal structure in the error terms. Because the residuals of the smaller model could not be modeled as iid, the non-seasonal MA term was kept in the model.

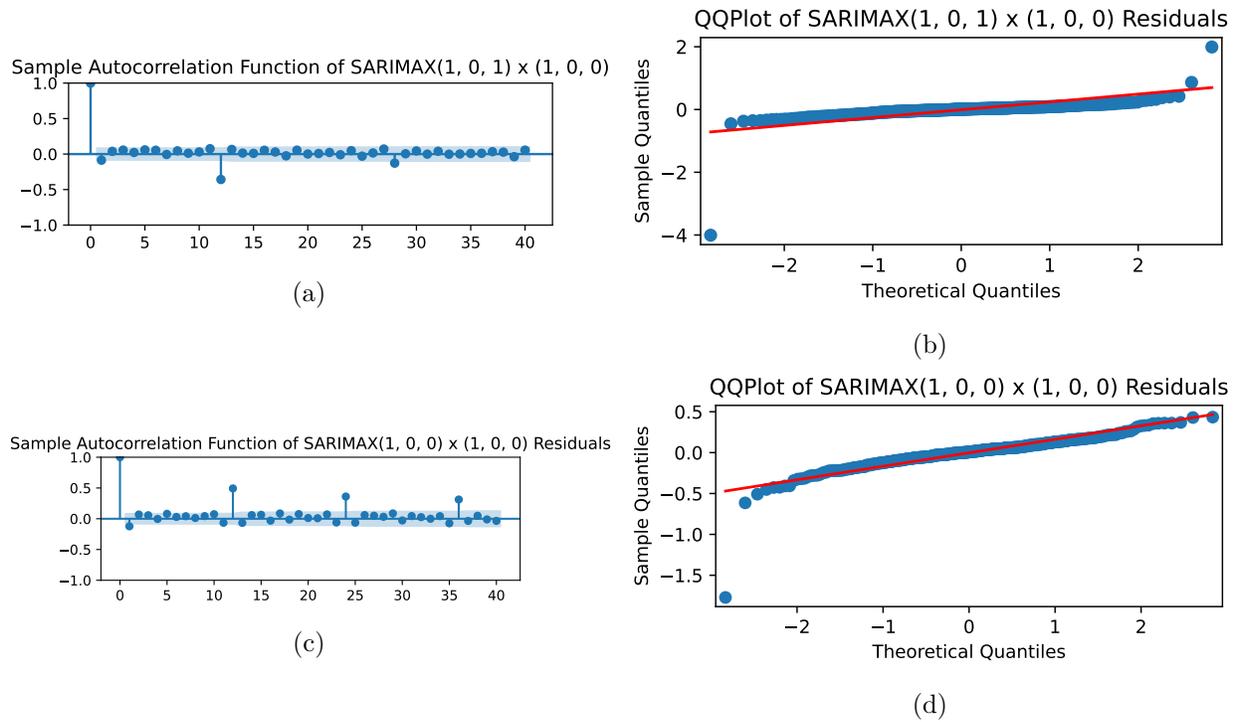


Figure 12: Comparison of sample ACF and QQplots for residuals of SARIMAX(1, 0, 1) x (1, 0, 0)<sub>12</sub> and SARIMAX(1, 0, 0) x (1, 0, 0)<sub>12</sub> for Property Crime

## 6.7 Bootstrap Results For Property and Violent Crime Models

The bootstrap function was created using a code from lecture (Ionides 2026), which were modified to apply to multiple coefficients and a SARIMAX model with assistance from AI (OpenAI 2026d). The best fitting model for property crime (SARIMAX(1,0,1) × (1,0,0)<sub>12</sub>) and violent crime (SARIMAX(1,0,1) × (1,0,1)<sub>12</sub>) were simulated 300 times each and the parameter estimates were plotted. The original fitted model was used to generate data in each simulation and were conditional on the observed exogenous covariates. In addition, a burn period of 100 observations was used to reduce sensitivity to initial conditions.

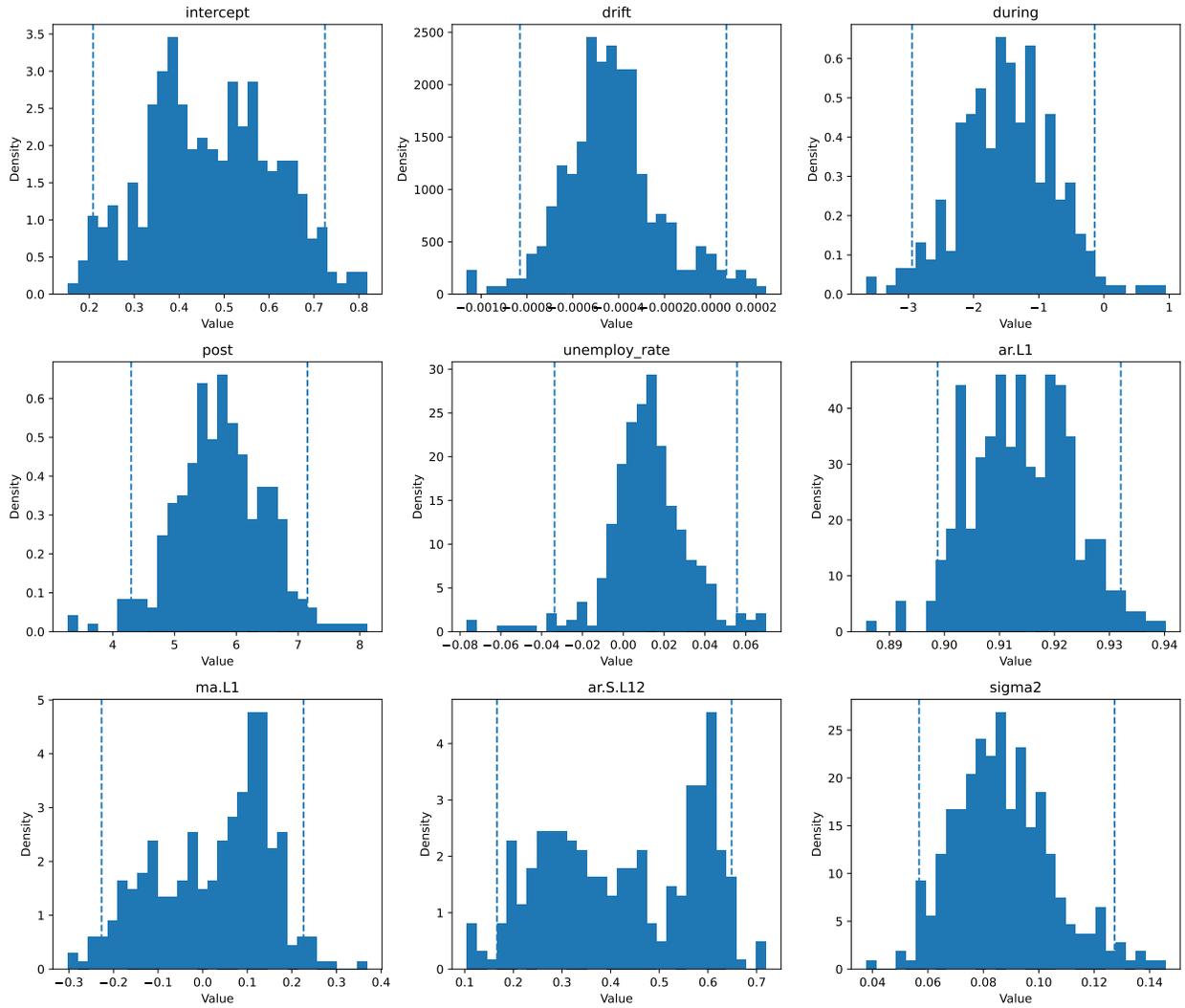


Figure 13: Bootstrap for SARIMAX(1,1,1,0) For Property Crime Model Coefficients.

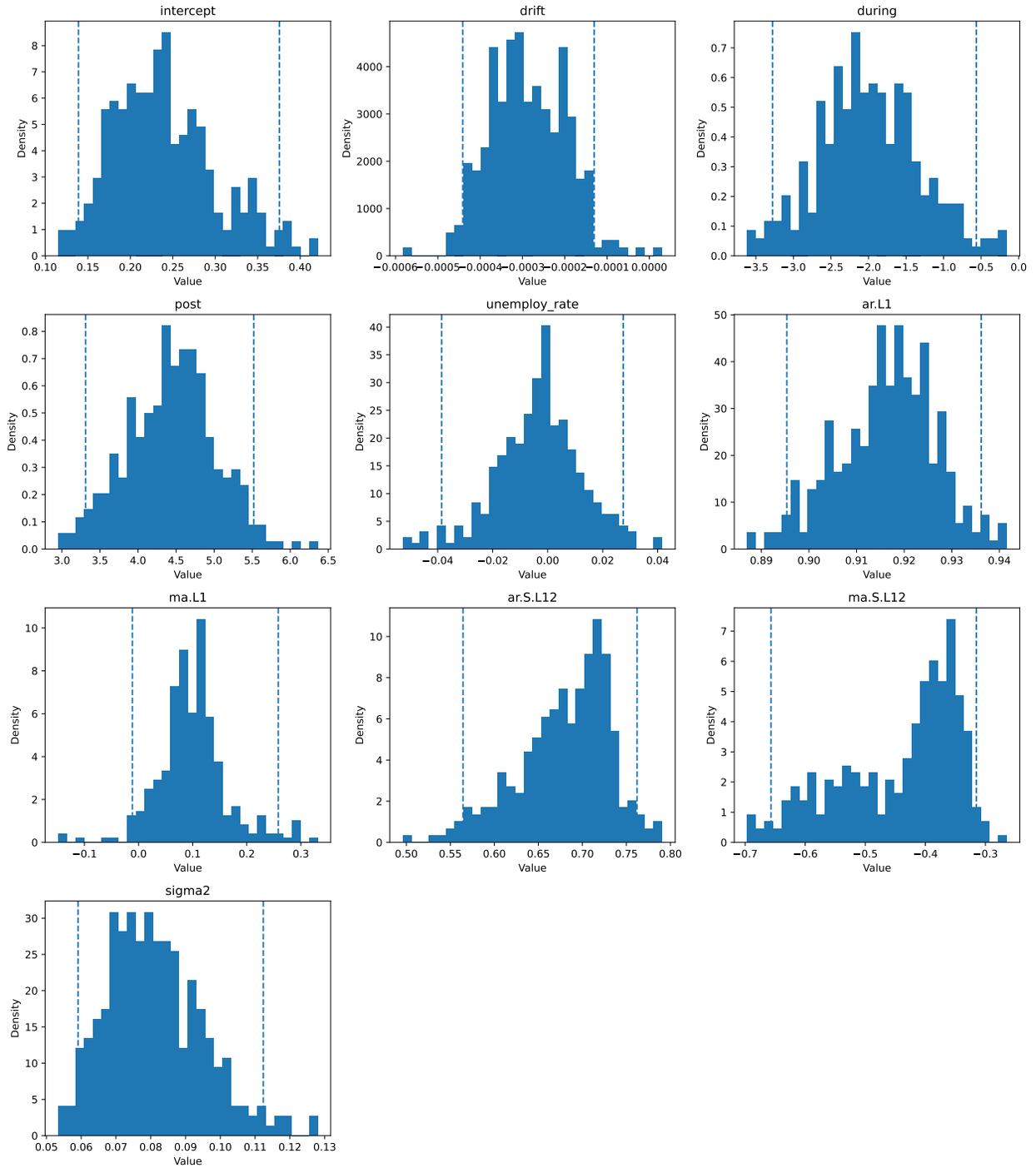


Figure 14: Bootstrap for SARIMAX(1,1,1) For Violent Crime Model Coefficients.