

Time Series Analysis of Fatal Motor Vehicle Crashes in Michigan, 2008–2023

1 Abstract

This study investigates monthly fatal motor vehicle crashes in Michigan from 2008 through 2023 using modern time series methods. The analysis aims to characterize temporal dependence, seasonal structure, and longer-term variation while developing a forecasting model for short-term crash dynamics. Exploratory visualization and unit root testing indicate strong annual seasonality and nonstationarity, motivating log transformation and seasonal differencing. Autocorrelation diagnostics and spectral analysis consistently identify a dominant 12-month cycle, supporting the use of a seasonal ARIMA framework.

A

$$\text{SARIMA}(1, 1, 1)(1, 1, 1)_{12}$$

model is estimated to capture both short-term dependence and annual periodic behavior. Residual diagnostics, including autocorrelation checks and a Ljung–Box test, suggest that the model adequately represents the stochastic structure of the series. Forecasts preserve seasonal variation while indicating modest stabilization of crash counts in the near term.

Overall, the results highlight persistent seasonal patterns in fatal crash risk alongside moderate long-term variability. The findings demonstrate how classical time series methods, complemented by spectral diagnostics, provide interpretable models for transportation safety data and useful short-term forecasts for policy planning.

2 Introduction

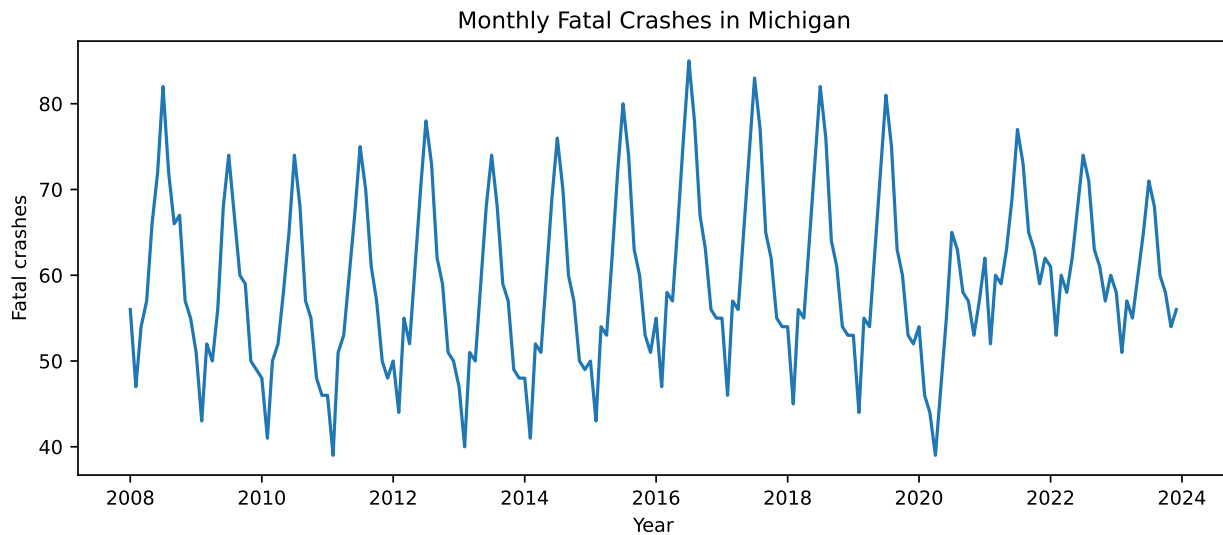
Motor vehicle crashes remain a major public health concern in the United States, with fatalities exhibiting temporal structure driven by seasonal travel behavior, behavioral risk factors, and broader societal changes. The Fatality Analysis Reporting System (FARS) provides comprehensive data for studying these dynamics and their implications for forecasting roadway risk. Understanding temporal variation in crash counts is important for both statistical modeling and transportation safety planning.

This study analyzes monthly fatal motor vehicle crashes in Michigan from 2008 through 2023 using a Box–Jenkins time series framework supplemented with spectral diagnostics. The primary objective is to characterize trend, seasonality, and stochastic dependence while developing an interpretable forecasting model for short-term crash dynamics. Building on prior STATS 531 work examining Michigan fatal crash data, the present analysis places greater emphasis on frequency-domain diagnostics and formal residual assessment to strengthen model validation and interpretation.

3 Data

The dataset contains 192 monthly observations derived from the Fatality Analysis Reporting System, a census of U.S. fatal crashes widely used in transportation safety research [1]. Monthly aggregation provides sufficient temporal resolution to identify seasonal structure while still capturing longer-term behavioral and policy driven changes. Michigan specific traffic safety summaries further contextualize the observed patterns, highlighting persistent seasonal variation and recent pandemic era disruptions [2].

4 Exploratory Analysis



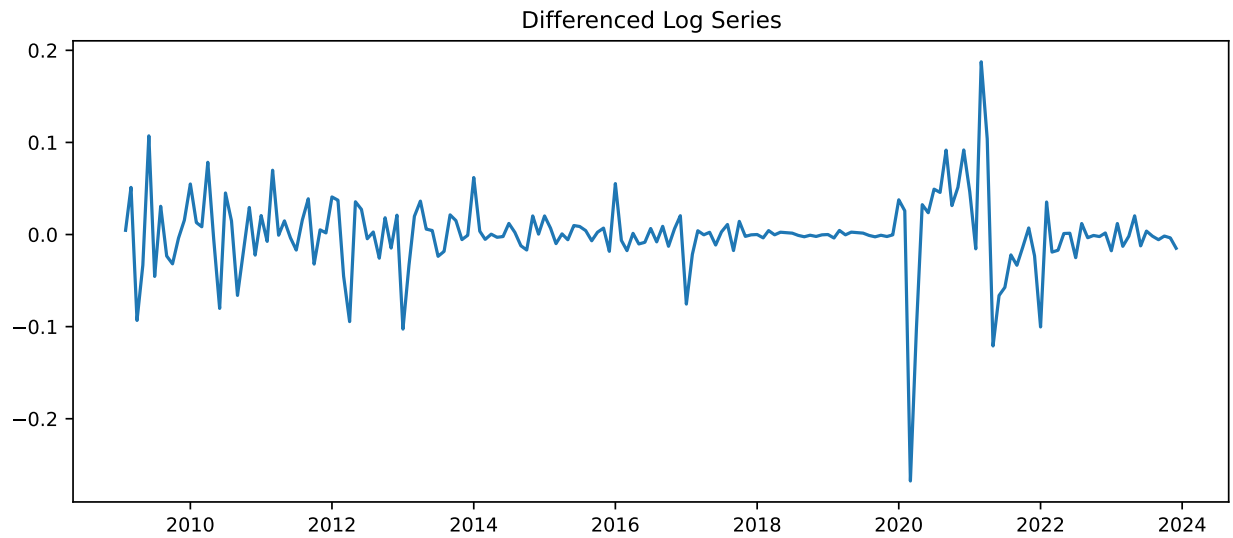
Our exploratory workflow mirrors the staged decomposition and dependence analysis used in [12], beginning with seasonal decomposition before proceeding to autocorrelation and spectral diagnostics.

Visual inspection reveals strong seasonal variation with peaks in summer months and troughs in winter. Such patterns are consistent with increased travel exposure and behavioral risk factors documented in transportation safety literature [2]. A moderate upward shift following 2020 suggests potential pandemic era behavioral changes influencing roadway risk, coinciding with policy interventions such as mobility restrictions and subsequent rebound travel patterns [3]. The presence of trend and seasonality implies that raw data are unlikely to be stationary.

ADF p-value: 0.053

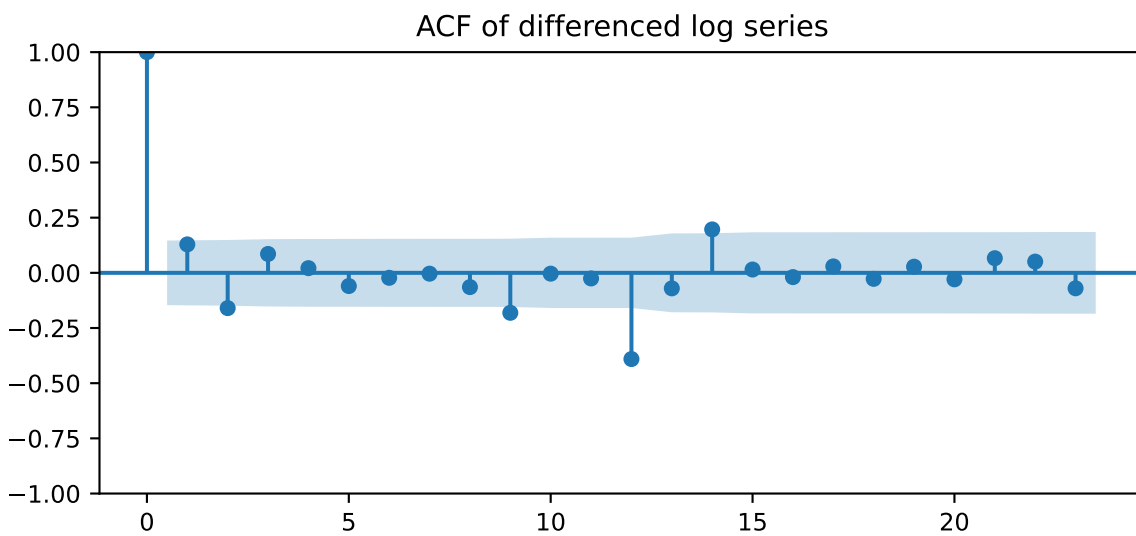
The Augmented Dickey–Fuller test fails to reject non stationarity, consistent with visual evidence of seasonal and trend components. Unit root testing provides a formal statistical basis for differencing decisions within the Box–Jenkins modeling framework [10]. Consequently, transformations are required before model identification.

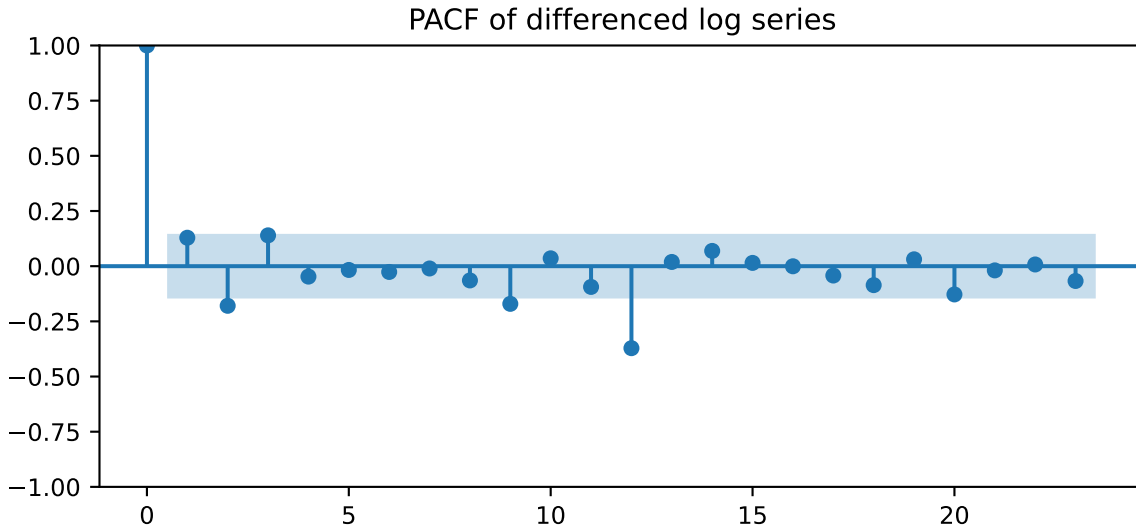
5 Transformations



Log transformation stabilizes variance, while first and seasonal differencing remove long term and periodic structure. Seasonal differencing is a standard technique for addressing deterministic annual cycles in monthly data, enabling stochastic modeling of residual temporal dependence [6]. The resulting series appears mean reverting with reduced amplitude fluctuations, suggesting approximate stationarity.

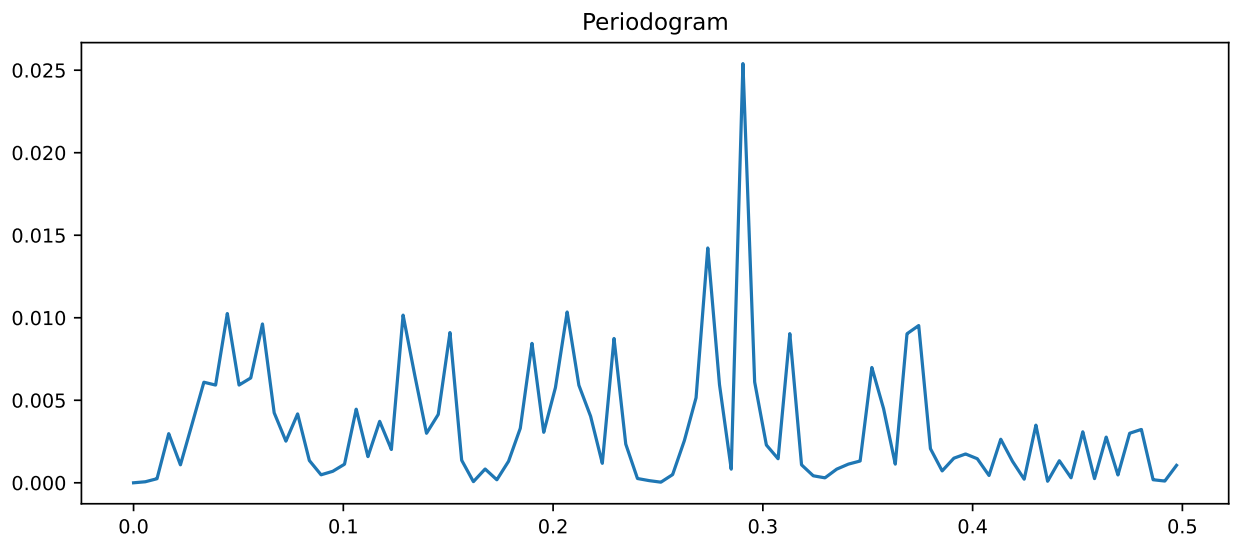
6 ACF and PACF Diagnostics

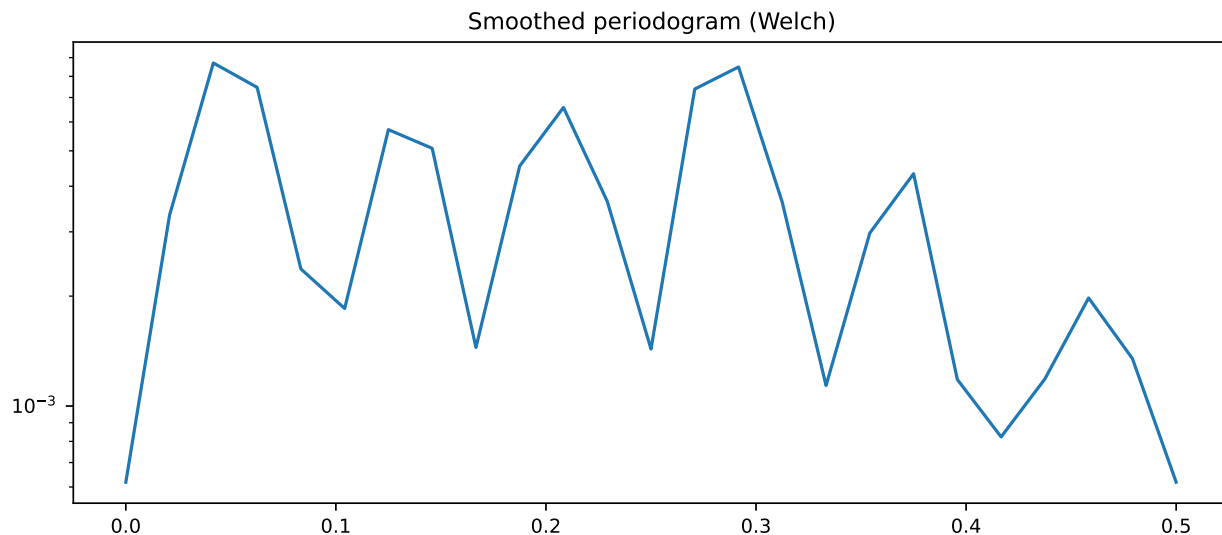




Autocorrelation patterns indicate a mixed autoregressive and moving average structure with seasonal dependence at lag 12. Examination of ACF and PACF functions remains central to ARMA model identification, providing insight into lag structure and guiding parsimonious specification [5,7]. These diagnostics motivate a SARIMA specification consistent with the observed annual cycle.

7 Spectral Analysis





The periodogram displays a dominant frequency corresponding to a 12 month cycle, reinforcing the seasonal interpretation derived from time domain diagnostics. Spectral methods provide complementary insight by representing temporal variation in the frequency domain, allowing identification of periodic structure not immediately apparent from autocorrelation alone [4]. Spectral smoothing highlights a clear annual peak while reducing high frequency noise, providing frequency domain confirmation of the seasonal differencing used in the SARIMA specification.

8 Model Estimation

```
/cloud/python/lib/python3.12/site-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning
  self._init_dates(dates, freq)
/cloud/python/lib/python3.12/site-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning
  self._init_dates(dates, freq)
```

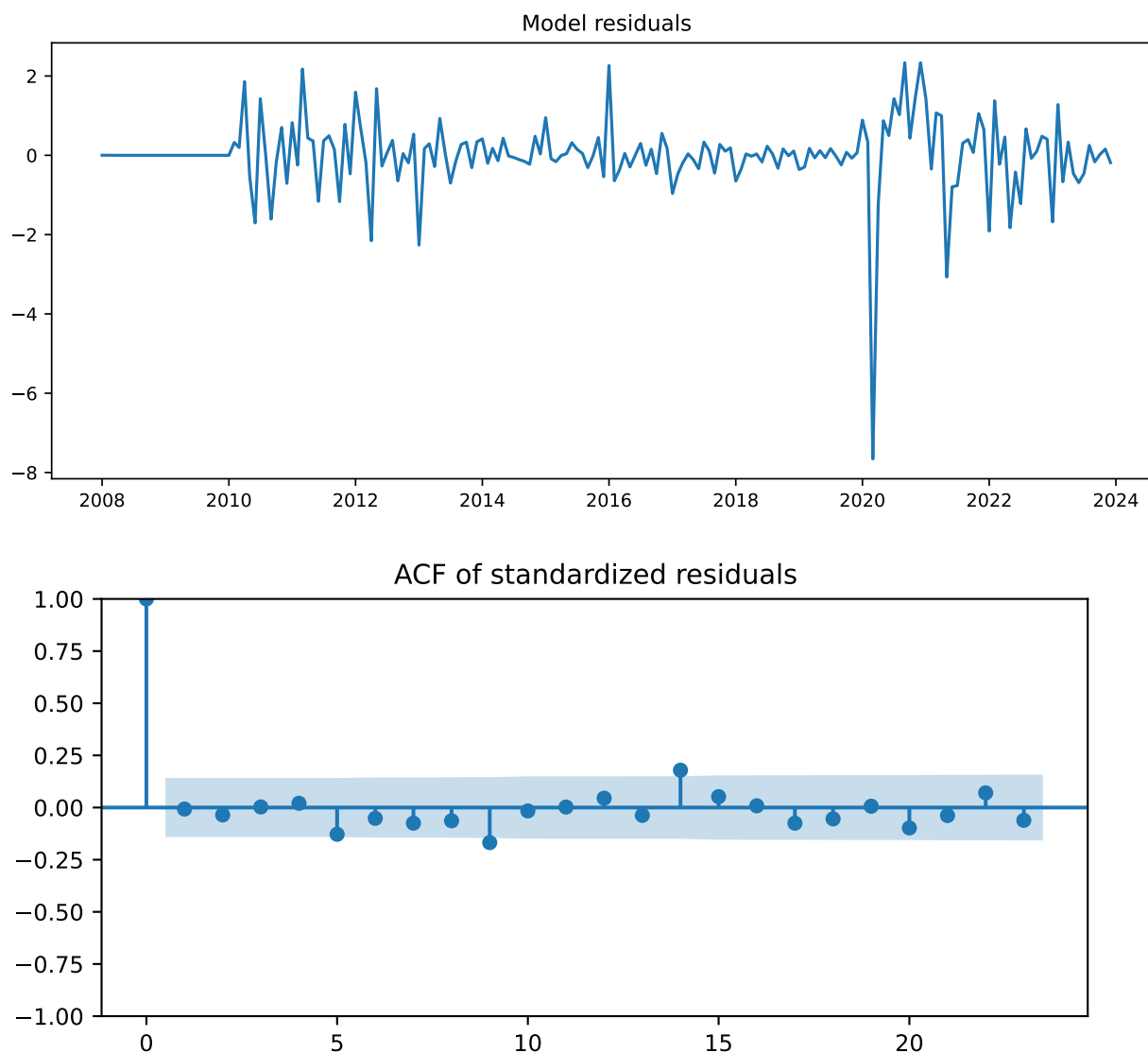
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.2272	0.340	-0.667	0.504	-0.894	0.440
ma.L1	2.2206	1.626	1.366	0.172	-0.966	5.407
ar.S.L12	-0.0404	0.125	-0.323	0.747	-0.286	0.205
ma.S.L12	-0.5436	0.146	-3.715	0.000	-0.830	-0.257
sigma2	0.0002	0.000	0.689	0.491	-0.000	0.001

Given the clear seasonal behavior observed in exploratory analysis, we extend the baseline ARMA framework to seasonal models, following the strategy outlined in [12]. The

$$\text{SARIMA}(1, 1, 1)(1, 1, 1)_{12}$$

model balances parsimony with the need to capture both short term dependence and annual cyclic structure. Seasonal ARIMA models extend classical ARMA formulations by incorporating differencing and seasonal components to address non stationary periodic processes [6,10]. The seasonal moving average component is strongly significant, while the nonseasonal AR and MA parameters exhibit weaker statistical support. Despite this, the combined specification captures both seasonal persistence and short term dependence, yielding a parsimonious yet adequate model. Model order was selected using iterative inspection of ACF/PACF diagnostics combined with parsimony considerations rather than automated information criterion search, reflecting the classical Box–Jenkins identification philosophy [10].

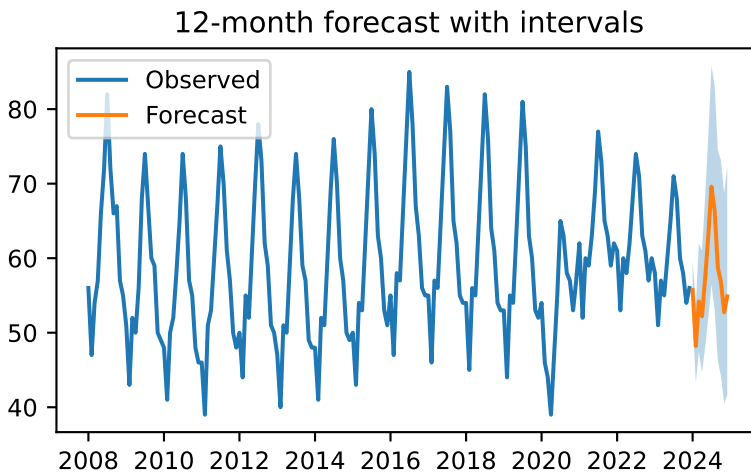
9 Diagnostics



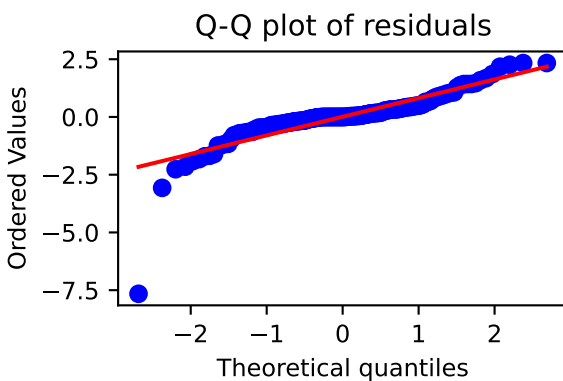
	lb_stat	lb_pvalue
12	12.315457	0.420689

Residual diagnostics indicate approximate white noise behavior. The Ljung–Box test provides a formal assessment of residual autocorrelation, serving as a standard goodness of fit diagnostic for time series models [9]. The residual ACF exhibits no systematic structure, reinforcing adequacy of the SARIMA specification.

10 Forecasting & Q-Q Plot



Forecasts primarily reflect seasonal persistence and mean reversion typical of SARIMA models, suggesting no strong evidence of continued structural change beyond the observed seasonal pattern. Such seasonal persistence is consistent with historical transportation safety patterns in Michigan and nationally [2]. The ability of SARIMA models to propagate both seasonal and stochastic dynamics makes them valuable tools for short term crash forecasting and policy planning [10].



The Q–Q plot suggests approximate central normality with noticeable tail deviations, a common feature when modeling log transformed count data. These departures do not appear severe enough

to undermine inference or forecasting performance. While Gaussian assumptions underpin many ARIMA estimation procedures, minor deviations are generally tolerated without substantial impact on forecast performance [10].

11 Discussion

The analysis demonstrates that Michigan fatal crash counts exhibit stable seasonal structure alongside moderate long term variation. Seasonal peaks align with increased travel exposure and behavioral risk factors during warmer months, consistent with transportation safety findings reported in statewide summaries [2]. Importantly, pandemic era deviations highlight how exogenous shocks can temporarily alter roadway risk dynamics without fundamentally changing seasonal structure, reflecting behavioral responses to mobility restrictions and subsequent travel rebounds [3].

From a policy perspective, results emphasize the continued importance of targeted summer safety interventions. Forecast stability suggests that existing mitigation strategies may be maintaining baseline risk levels, though sustained monitoring remains essential.

11.1 Limitations

The analysis relies on aggregated monthly counts, which may obscure heterogeneity in crash circumstances. Additionally, exogenous factors such as traffic exposure, enforcement intensity, and weather conditions were not incorporated. Consequently, while the SARIMA model captures temporal dependence, causal interpretation remains limited.

12 Conclusion

This study applied a comprehensive time series framework to monthly fatal crash data in Michigan from 2008 to 2023. After addressing non stationarity through transformation and differencing, a SARIMA model effectively captured seasonal and stochastic dynamics, consistent with established Box–Jenkins methodology [10]. Our findings broadly align with the modeling insights of [12], reinforcing the importance of seasonal structure in fatal crash time series while highlighting differences in model stability and parameterization. Diagnostic checks supported model adequacy, and forecasts indicated persistent seasonal behavior with modest long term stabilization.

Future work could incorporate exogenous predictors such as traffic volume, weather conditions, or policy interventions using SARIMAX models. Such extensions would allow deeper causal interpretation and potentially improve forecast accuracy, providing valuable insights for transportation safety planning.

13 Acknowledgments

Data are from the NHTSA Fatality Analysis Reporting System, accessed via the NHTSA CrashAPI (state code 26). Building on a prior STATS 531 project examining Michigan fatal crash data, the present analysis places greater emphasis on spectral diagnostics and formal residual assessment, providing stronger validation of seasonal structure and model adequacy. Analysis was conducted in Python using `pandas` [13], `numpy`, `matplotlib`, `scipy`, and `statsmodels` [11], compiled with

Quarto. AI assistance (Claude, Anthropic) was used for code scaffolding, debugging, and editing support, in accordance with the STATS 531 course policy on AI acknowledgment.

14 References

- [1] National Highway Traffic Safety Administration. *Fatality Analysis Reporting System (FARS) Analytical User's Manual, 1975–2022*. DOT HS 813 417. Washington, DC: NHTSA, 2023.
- [2] National Highway Traffic Safety Administration. *Michigan Traffic Safety Facts 2022*. DOT HS 813 494. Washington, DC: NHTSA, 2024.
- [3] State of Michigan. *Executive Order 2020-21: Temporary requirement to suspend activities that are not necessary to sustain or protect life*. Lansing, MI, April 2020.
- [4] Ionides, E.L. *Chapter 7: Introduction to Time Series Analysis in the Frequency Domain*. STATS 531 Course Notes, University of Michigan, Winter 2026.
- [5] Ionides, E.L. *Chapter 5: Parameter Estimation and Model Identification for ARMA Models*. STATS 531 Course Notes, Winter 2026.
- [6] Ionides, E.L. *Chapter 6: Extending the ARMA Model: Seasonality, Integration and Trend*. STATS 531 Course Notes, Winter 2026.
- [7] Ionides, E.L. *Chapter 4: Linear Time Series Models and the Algebra of ARMA Models*. STATS 531 Course Notes, Winter 2026.
- [8] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- [9] Ljung, G.M. and Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65(2), 297–303.
- [10] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., and Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley.
- [11] Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proc. 9th Python in Science Conf.*, 57–61.
- [12] STATS 531 Winter 2025 Project 15. *Time Series Analysis on Fatal Car Accidents in Michigan*. University of Michigan. https://ionides.github.io/531w25/midterm_project/project15/blinded.html
- [13] McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proc. 9th Python in Science Conf.*, 51–56.