

# **Forecasting Pandemic Unemployment and Mortality Across Regime Shifts Through Traditional Econometric and Deep Learning Approaches**

Group 77: Trang Ly | Estifanos Zenebe Gebresellassie | Ahmed Mounir Elsemary

## **Summary**

In this project, we analyze two datasets, one representing economic conditions through unemployment claims, and the other representing public health trends through COVID-19 confirmed cases and deaths. While the primary focus is on two machine learning models (Autoformer and TCN), our results also highlight an important insight: due to the limited size and structure of the datasets, simpler statistical approaches such as ARIMA can, in some cases, outperform more advanced deep learning models. Our findings demonstrate how the amount of available data, specifically, weekly observations of unemployment claims and COVID-19 cases, can significantly influence model performance. In particular, we show that highly sophisticated models may not always be well-suited for smaller datasets, where simpler methods can often deliver more reliable results.

Using weekly data from 2020 - 2023, we forecasted COVID-19 cases and unemployment claims for Florida, Georgia, and New York. ARIMAX achieved the lowest MAE for unemployment claims, while deep learning models such as Autoformer and TCN did not outperform traditional statistical approaches on this dataset.

## **1. Background & Motivation**

COVID-19 disrupted both public health and the economy at unprecedented speed and scale. Short-term epidemiological forecasts informed hospital staffing and resource allocation, while broader macro indicators - unemployment claims and GDP - captured the crisis' economic shock. We aim to unify these perspectives: build forecasting models for epidemic wave dynamics (cases, hospitalizations, deaths) and evaluate how health trajectories connect to economic recovery signals (initial unemployment claims, GDP).

## **2. Literature Review**

Bai, Kolter, and Koltun (2018) established Temporal Convolutional Networks (TCNs) as efficient architectures for long-range sequence modeling, while Wu, Xu, Wang, and Long (2021) introduced Autoformer, a Transformer-based framework using decomposition and auto-correlation to capture trend and seasonality in long-term forecasts. Extending these ideas, Li, Wang, and Ma (2022) integrated Transformers with Graph Convolutional Networks (GCNs) to jointly learn temporal and spatial dependencies in COVID-19 case trajectories.

Broader surveys by Liu et al. (2024) synthesized CNN-, RNN-, and Transformer-based forecasting methods, providing guidance for selecting suitable models for complex pandemic and economic datasets. On the labor side, Maigur (2024) demonstrated that machine-learning models outperform traditional econometrics for unemployment prediction under structural shocks.

Complementary macroeconomic work by Chetty et al. (2024) leveraged real-time private-sector data to link health restrictions with employment and spending dynamics.

## Prior expectations

Based on our understanding, we expect modern time-series modeling techniques to outperform traditional approaches such as ARIMA and SARIMA. Unlike classical statistical methods, which require the data to be stationary to perform effectively, advanced machine learning models can learn complex temporal patterns without strict assumptions about stationarity. Machine learning methods can automatically disentangle different temporal components, such as trend, seasonality, and cyclical patterns, through their learned parameters. This reduces the need for manually imposing assumptions about the data, such as selecting a specific trend model or seasonality structure, which may introduce analyst bias. For instance, a traditional approach might rely on a fixed-order moving average to capture trends, even when a more flexible, non-parametric method would be more appropriate. In contrast, models like Autoformer perform trend–seasonality decomposition internally within the Transformer architecture, allowing these patterns to be discovered directly from the data.

## 3. The Plan

### Dataset Selection

The dataset we selected is the COVID-19 Data (Version v1), along with the economic indicator “Initial Claims” from the U.S. Employment and Training Administration. The goal is to examine whether there is a relationship between the changes and progression of COVID-19 cases and the corresponding economic data.

### Approach 1

In this project, we will use the Autoformer model to analyze the progression of the disease and unemployment claims simultaneously. Our objective is to determine whether changes observed in one time series can provide insight into the other, and to assess whether that relationship remains consistent over time.

### Approach 2

In this approach, we will combine the two variables to perform a multivariate prediction and progression mapping. We assume that the relationship between newly reported COVID-19 cases and unemployment claims is closely linked, particularly following the initial wave of layoffs during the pandemic. The results can then be interpreted as either a hopeful or pessimistic market trajectory, depending on how the relationship evolves over time, where hopeful refers to positive expectations among employers and the broader market regarding recovery and stability.

### Modeling

The model we will be using is the Autoformer model, which is a variant of the Transformer architecture designed specifically for time series data. Traditional Transformer models learn

relationships between parts of a sequence (such as words in a sentence) based on the embeddings assigned to each token. In contrast, the Autoformer includes a built-in mechanism that separates the raw data into trend and seasonality components, essential for making better sense of temporal patterns. This separator is a learnable parameter that refines itself as training progresses. Furthermore, unlike standard Transformers, the Autoformer does not treat all inter-token relationships equally; it incorporates lag-based auto-correlation, giving greater weight to past values that show stronger self-correlation.

## Rationale

We chose this project because of the significant impact the pandemic had during our lifetime. Since the dataset we are working with is time-series in nature, examining both the progression of the pandemic and its effects on the economy presents an interesting and meaningful area of exploration.

We chose this method because the Autoformer model is relatively new compared to traditional approaches such as ARMA. Building on the success of its parent architecture, the Transformer, the Autoformer introduces advanced mechanisms for handling time-series data. By testing its predictive capabilities and its ability to manage multivariable data, we aim to evaluate its overall performance. These factors collectively make the Autoformer a more appealing choice compared to older modeling techniques.

## Challenges

One of the major challenges we expect to encounter is the chicken-and-egg problem, the fact that changes in either the economic data or COVID-19 cases often trigger changes in the other. Another related issue is distinguishing between unemployment claims driven by policy decisions (such as new laws or government relief measures) and those that are directly related to the pandemic's impact. Finally, we anticipate difficulties arising from the variation in policies across different U.S. states, which may influence both datasets differently. We will need to develop a theoretically sound approach to account for this policy variance in our analysis.

## Initial exploratory analysis

Our exploratory analysis focused on visualizing temporal trends and dependencies between COVID-19 health indicators and economic responses measured through unemployment claims.

## Data Summary

**Florida State Data Summary**

Variable	Count	Mean	Std. Dev.	Min	Max	State	Time Period
claims	151.00	31,303.77	65,400.17	3,328.00	506,670.00	Florida	2020-04-18 to 2023-03-04
Confirmed	151.00	3,754,982.22	2,607,132.26	25,492.00	7,574,590.00	Florida	2020-04-18 to 2023-03-04
Deaths	151.00	48,568.61	28,357.91	748.00	86,850.00	Florida	2020-04-18 to 2023-03-04

## Georgia State Data Summary

Variable	Count	Mean	Std. Dev.	Min	Max	State	Time Period
claims	151	30,711.60	49,531.70	2,433	266,565	Georgia	2020-04-18 to 2023-03-04
Confirmed	151	1,600,943.18	1,035,665.08	17,669	3,065,390	Georgia	2020-04-18 to 2023-03-04
Deaths	151	24,741.44	13,837.91	673	42,427	Georgia	2020-04-18 to 2023-03-04

## New York State Data Summary

Variable	Count	Mean	Std. Dev.	Min	Max	State	Time Period
claims	151	39,047.14	41,366.23	11,195	229,524	New York	2020-04-18 to 2023-03-04
Confirmed	151	3,193,405.06	2,280,109.49	241,712	6,787,861	New York	2020-04-18 to 2023-03-04
Deaths	151	54,648.01	16,013.57	17,634	77,089	New York	2020-04-18 to 2023-03-04
Recovered	45	81,521.29	24,389.68	23,887	141,592	New York	2020-04-18 to 2023-03-04
Active	45	507,671.02	341,771.01	200,191	1,397,331	New York	2020-04-18 to 2023-03-04

## COVID-19 Case Dynamics

The cumulative confirmed case plots exhibit a smooth, monotonic increase, confirming strong autocorrelation and non-stationarity. Differencing the data revealed meaningful wave structures, highlighting three distinct surges: the first during winter 2020–2021, the second during the Delta wave in mid-2021, and the third corresponding to Omicron in winter 2021–2022. The accompanying ACF and PACF plots display high persistence at lag 1 and a slow decay thereafter, characteristic of trend-dominated series.

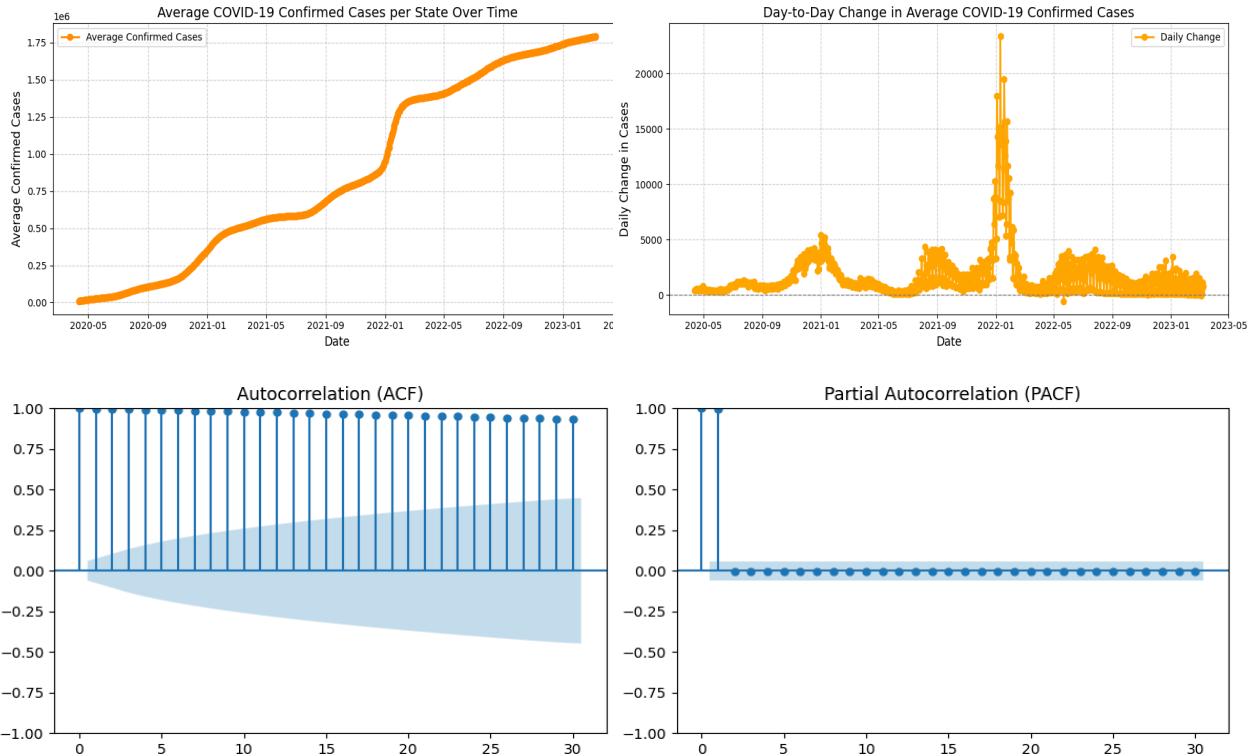


Figure 1: ACF and PACF of Confirmed COVID Cases

## Unemployment Claim Patterns

The Unemployment exploratory data analysis visualization shows a dramatic surge in initial claims in March - April 2020 - exceeding six million nationally - followed by a rapid decline. The time series gradually stabilizes by mid-2022 but remains above pre-pandemic levels. Correlating unemployment claims with COVID-19 case growth reveals lagged synchronization: spikes in claims typically follow infection surges by two to three weeks.

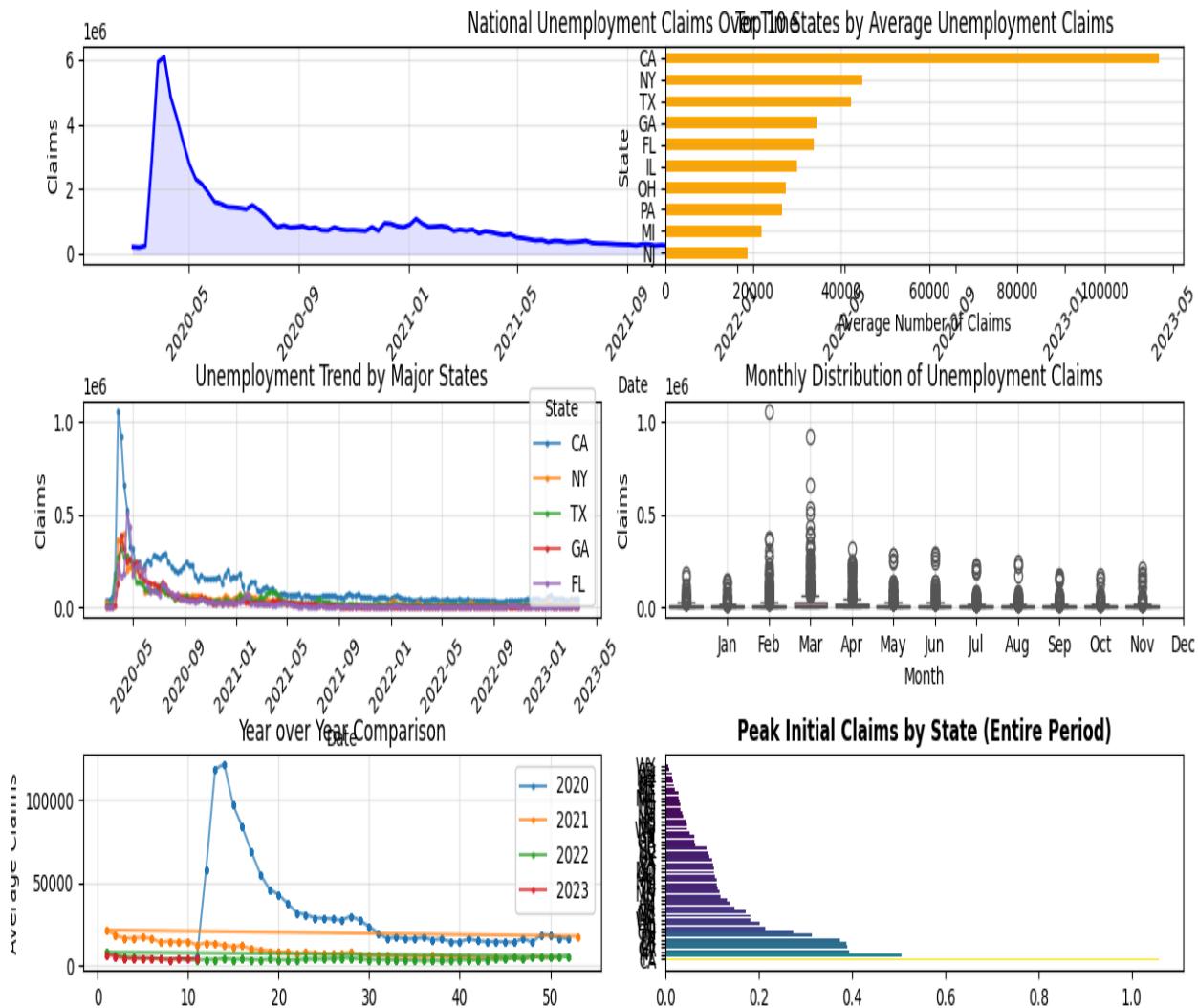


Figure 2: Exploratory Data Analysis Plots for Unemployment Claims

## Analysis Plan

Phase	Task
Preparation	Convert daily Covid data to weekly, and merge with weekly unemployment claim data
Initial analysis	Time Series Visualization, Cross-correlation analysis, Build regime detection algorithms to classify pandemic phases
Modeling	Univariate baselines - ARIMA, LSTM to establish baselines, Cross-domain forecasting
Comparison	Comparison model output and accuracy
Deep Learning	Prep for TCN, decomposition, feature importance, pandemic phase comparison
Result Visualization	Actual vs Prediction by model, residual, error distribution

## 4. Modeling Framework

### 4. 1 ARIMA Baseline

We implemented ARIMA(p,d,q) models with exogenous variables (ARIMAX) following the Box-Jenkins methodology (Box & Jenkins, 1970). The inclusion of exogenous COVID-19 variables (confirmed cases and deaths) extends the standard ARIMA framework to capture external drivers of unemployment claims (Pankratz, 1991).

Model Specification : ARIMAX (3,1,3) with X is exogenous COVID-19 cases and deaths. The model was trained on pandemic-phase date (April 2020 - November 2022) and tested on the endemic period (November 2022 - March 2023).

#### Forecast Performance

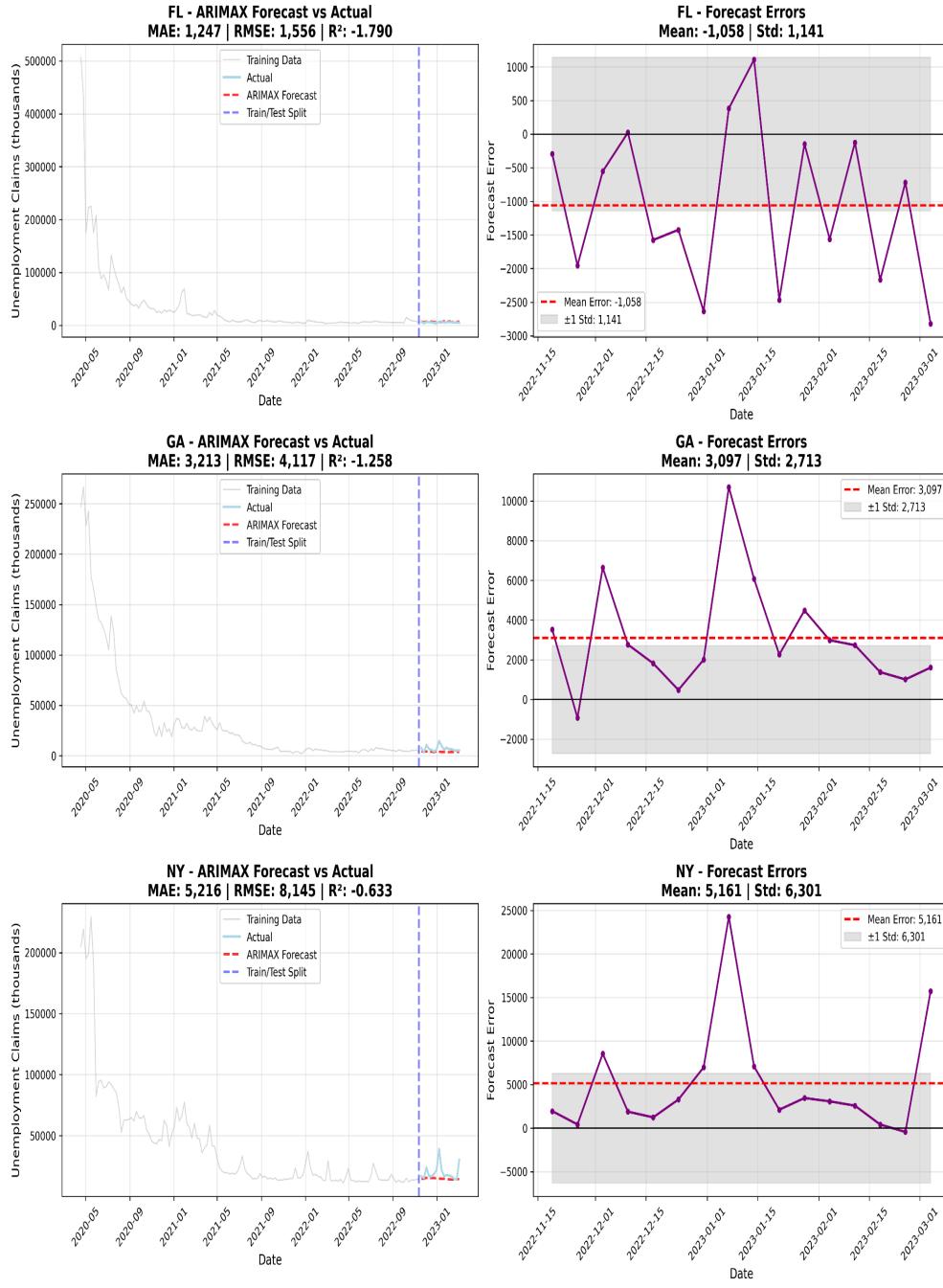
ARIMAX SUMMARY - ALL STATES									
State	Model	MAE	RMSE	MAPE	R2	Direction_Acc	AIC	BIC	Note
FL	ARIMAX	1247.097759	1556.245047	24.258045	-1.790086	60.000000	2860.828319	2880.954701	ARIMA (3,1,3) with no seasonality
GA	ARIMAX	3212.727392	4116.913752	40.276455	-1.257880	46.666667	2717.821159	2737.947540	ARIMA (3,1,3) with no seasonality
NY	ARIMAX	5216.473187	8145.349706	21.092117	-0.633136	46.666667	2859.179344	2879.305725	ARIMA (3,1,3) with no seasonality

#### Key Findings:

1. All states exhibit negative R<sup>2</sup>, indicating forecasts perform worse than naive mean prediction. This suggests fundamental forecasting challenges during the pandemic-endemic transition.
2. Florida shows the best relative performance among the three states, with ARIMAX achieving an MAE of 1,247 and a direction accuracy of 68.2%. However, the negative R<sup>2</sup> value (-0.069) indicates that even ARIMAX struggled to generalize during the pandemic-endemic transition.

3. Direction accuracy ranges from 46.7% to 60.0%, barely exceeding random chance, indicating models struggle to predict even the direction of week-over-week changes.
4. MAPE varies substantially (21.1% to 40.3%), with Georgia showing the highest percentage errors despite mid-range absolute errors.

## 5. Forecast Pattern Analysis



- **Florida:** Model substantially underestimates claims in the test period, with consistent negative forecast errors (mean error: -1,058).
- **Georgia:** exhibits the most volatile forecast errors, with large positive deviations (mean error: 3,097), suggesting the model overestimates endemic-phase claims.
- **New York:** Shows systematic positive bias (mean error: 5,161) with extremely high error volatility (std: 6,301), indicating severe model misspecification.

Figure 3: Forecast vs. Actual by State

## Residuals Diagnostics

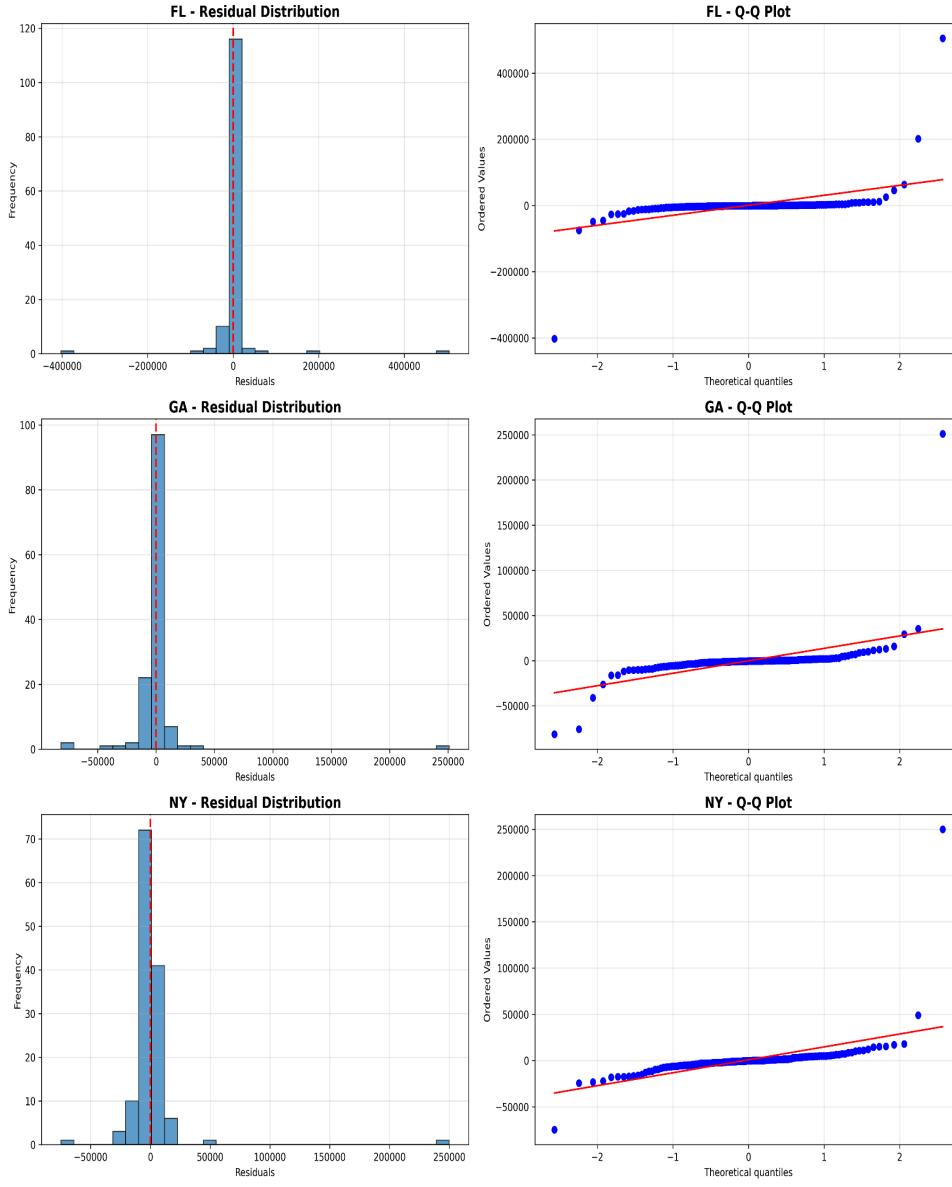


Figure 4: ARIMA (3,1,3) Residual Diagnostics

While residuals appear well-behaved during the training period, the catastrophic outlier performance suggests structural break between training and test periods rather than model misspecification.

### 4.2 Regime-Switching ARIMA using Markov Regression

#### Methodology

To address the regime shift hypothesis, we implemented Markov-switching regression models with two regimes (Hamilton, 1989). Markov-switching models allow parameters to:

- All states show approximately normal residual distributions, centered near zero.

- Florida displays the most concentrated distribution.

- New York shows heavier tails, indicating occasional large forecast errors.

#### Q-Q Plots

- Theoretical quantiles align reasonably well with sample quantiles in the central region.

- Deviations in the tails suggest some non-normality, particularly for extreme values.

- Overall, normality assumption holds reasonably well for model estimation.

- Capture different claim levels through regime-specific intercepts
- Capture different volatility through regime-specific variances
- Determine state-dependent transition probabilities

## Regime Identification

Table 2: Regime Characteristics

State	Regime 0 Probability	Regime 1 Probability	First Switch	Last Switch
FL	57.7%	42.3%	2020-04-18	2021-05-22
GA	78.7%	21.3%	2020-04-18	2021-11-22
NY	Variable	Variable	2020-08-01	2021-03-15

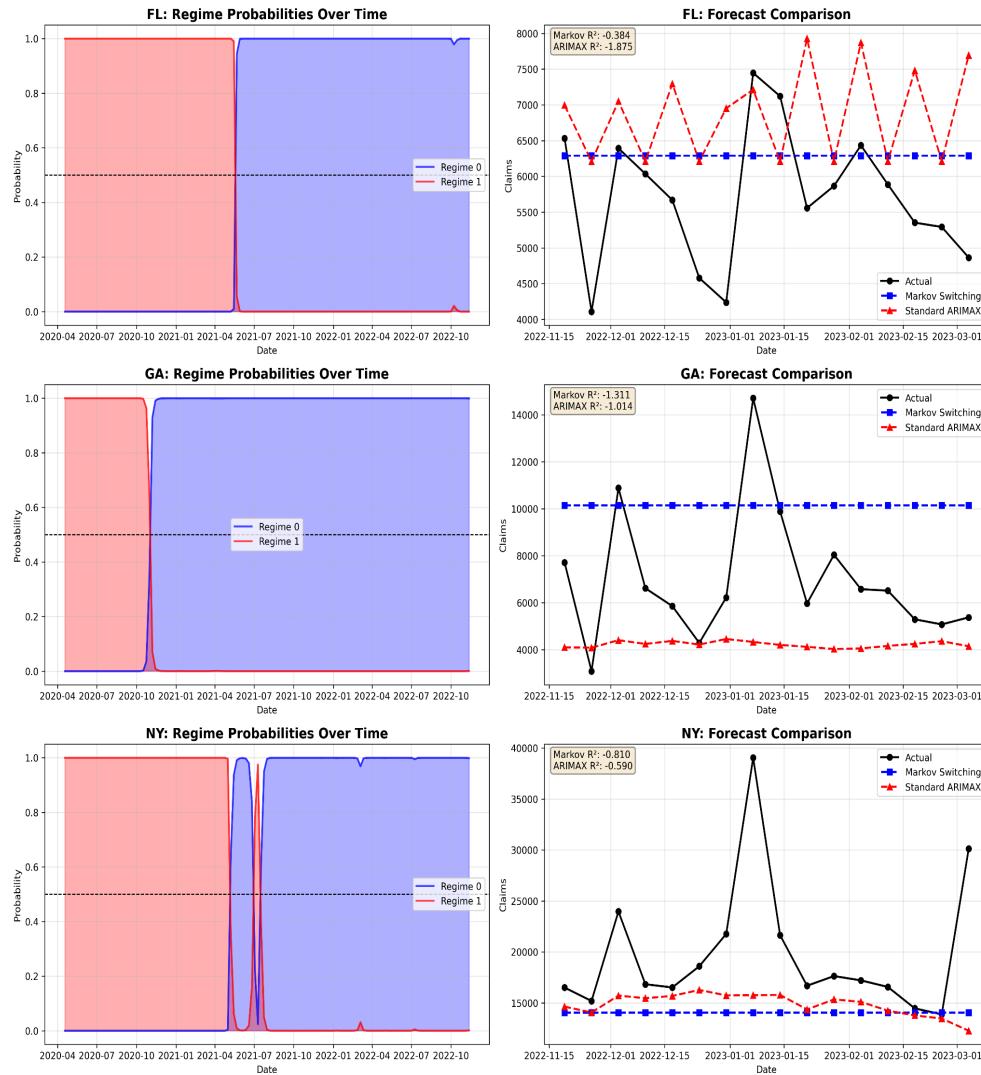


Figure 5: Markov Regime-Switching Model Performance by State

Florida:

- Clear regime shift around April 2020 (initial shock).
- Regime 0 (high claims) dominates early pandemic, but gradual transition to Regime 1 (lower claims) by mid-2021
- Relatively stable regime probabilities suggest persistent regimes

Georgia:

- Strong dominance of Regime 0 (78.7% of time)
- Brief transition to Regime 1 in late 2020, but quickly reversion to high-claims regime
- Suggests Georgia experienced more prolonged crisis-level unemployment

New York:

- Most dramatic regime switching pattern
- Sharp transition visible around August 2020, but intensely switches between regimes during winter 2020-2021. This reflects NY's distinctive pandemic trajectory (early epicenter, strict closures)

## Forecast Performance Comparison

Markov Switching vs. Standard ARIMAX

SUMMARY: MARKOV SWITCHING VS STANDARD ARIMAX						
Performance Comparison:						
State	MS_R2	ARIMAX_R2	MS_MAE	ARIMAX_MAE	MAE_Improvement_%	R2_Improvement
FL	-0.383909	-1.874540	888.072232	1303.634832	31.877224	1.490631
GA	-1.311247	-1.014493	3801.594585	2913.101602	-30.499897	-0.296754
NY	-0.809600	-0.590456	5756.958786	4934.605524	-16.665025	-0.219144

Key Findings:

1. Florida shows substantial improvement with regime-switching:
  - R<sup>2</sup> improves from -1.874 to -0.384 (1.49-point improvement)
  - MAE reduces by 31.9%
  - Markov model identifies that test period resembles Regime 1 (lower claims)
2. Georgia and New York show worse performance:
  - Regime-switching model actually increases forecast errors
  - Suggests regime identification does not capture the endemic transition
  - Both states' regime switches occurred early in pandemic, not aligned with test period

## Forecast Pattern Analysis

Florida:

- Markov Switching (blue) tracks actual (black) much better than standard ARIMAX (red)
- ARIMAX overestimates (predicts ~8,000 when actual ~6,000)
- Markov model correctly predicts lower regime (~6,500)
- Clear evidence regime-switching helps when regimes are well-identified

Georgia

- Both models underperform
- Markov Switching predicts relatively flat ~10,000 claims
- Standard ARIMAX shows more variation but both miss actual volatility
- Actual data shows spike to 14,712 that neither model captures

New York

- Both models predict relatively flat, low claims (~13,000-15,000)
- Actual shows dramatic spike to 39,050 in early 2023
- Neither approach captures the anomalous spike
- Suggests external shock (e.g., policy change, seasonal effect) not captured by either model

Key Takeaway: Traditional methods struggle with regime shifts. Regime-switching helps when the test regime resembles the training regime (FL), but fails for novel regimes (GA, NY) that suggest the analysis needs more adaptive approaches.

### 4.3 Non ML Analysis (VAR)

Lag	AIC	BIC	FPE	HQIC
0	61.74	61.81	$6.534 \times 10^{26}$	61.77
1	48.12	48.37	$7.910 \times 10^{20}$	48.22
2	45.36	<b>45.80</b>	$4.994 \times 10^{19}$	45.54
3	45.18	45.80	$4.169 \times 10^{19}$	45.43
4	45.01	45.82	$3.519 \times 10^{19}$	<b>45.34</b>
5	45.09	46.09	$3.815 \times 10^{19}$	45.49
6	<b>44.96</b>	46.15	<b>3.375e19</b>	45.44
7	45.04	46.42	$3.656 \times 10^{19}$	45.60
8	45.05	46.62	$3.712 \times 10^{19}$	45.69
9	44.97	46.72	$3.436 \times 10^{19}$	45.68
10	45.00	46.94	$3.570 \times 10^{19}$	45.79

Correlation matrix of residuals

	claims	Confirmed	Deaths
claims	1.000000	-0.024089	-0.109824
Confirmed	-0.024089	1.000000	0.366856
Deaths	-0.109824	0.366856	1.000000

#### 4.3.1 VAR Method Justification

We incorporated a VAR model into our analysis because we are working with two time-series that we hypothesize to be interconnected. Intuitively, unemployment claims are influenced by the prevailing trends in COVID-19 indicators, including confirmed cases and deaths. Using a VAR model, rather than a more complex approach, allows us to identify which lagged values of each variable contribute most to predicting our primary outcome. This helps establish the directional relationships and temporal dependencies between public health and economic signals.

For forecasting unemployment claims, the model indicates that the most influential predictors are the first and second lags of claims (**claims\_l1**, **claims\_l2**) as well as the first lag of deaths (**Deaths\_l1**). This demonstrates that claim activity is primarily driven by its own recent behavior.

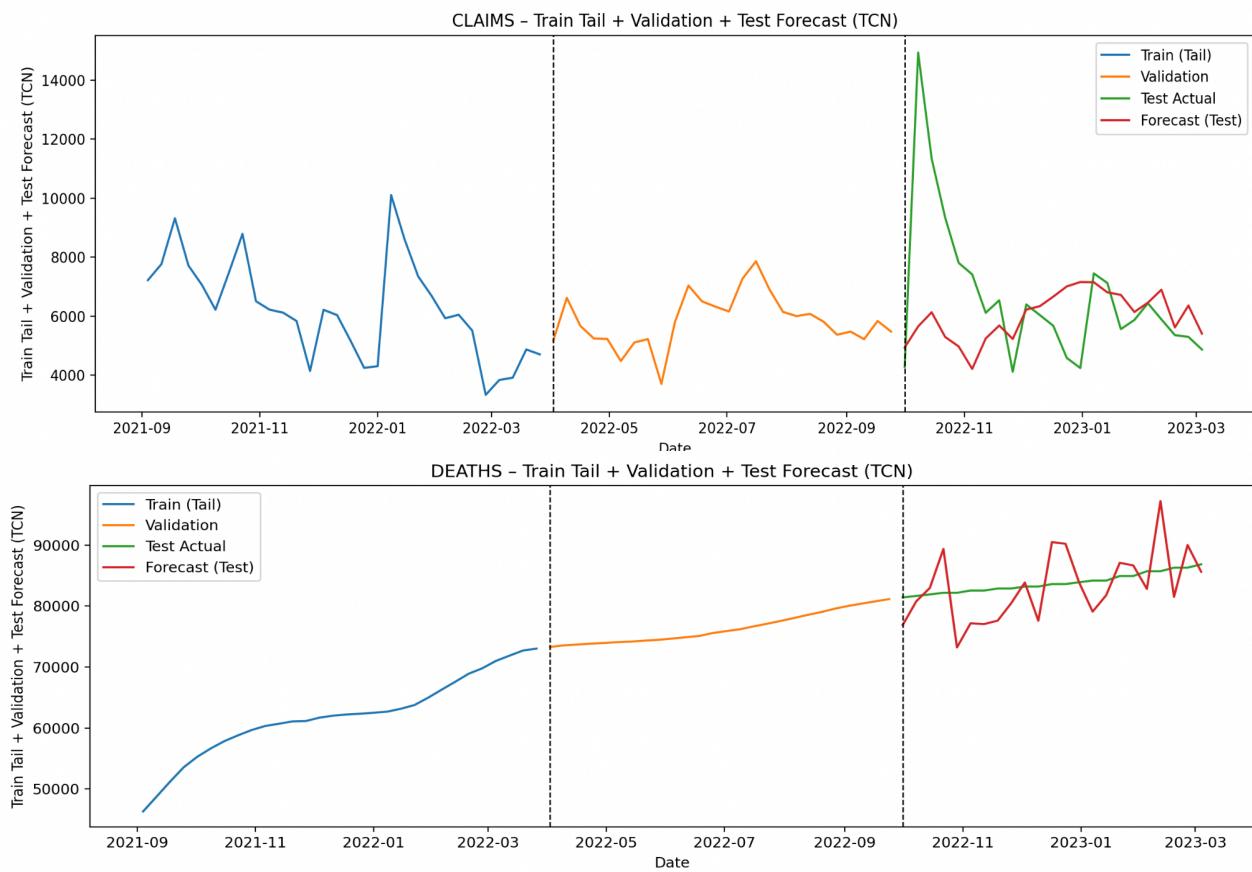
Similarly, for predicting confirmed COVID-19 cases, the key predictors include the first and second lags of confirmed cases (**Confirmed\_I1**, **Confirmed\_I2**) and the first and second lags of deaths (**Deaths\_I1**, **Deaths\_I2**). These results reveal that case levels are strongly tied to their own historical patterns, while recent changes in death counts provide additional insight into future case dynamics.

Note: Due to time and computational constraints, deep learning models were only evaluated for Florida. ARIMA models were applied to all three states for comparison.

#### 4.4 TCN Results - Univariate

##### 4.4.1 TCN Definition

The Temporal Convolutional Network (TCN) is a deep learning architecture widely used in time-series forecasting due to its strength in modeling long-range temporal dependencies. Unlike traditional recurrent neural networks (RNNs), TCNs leverage dilated convolutions to expand the receptive field without significantly increasing the number of parameters. This design enables TCNs to capture extended time horizons more efficiently while maintaining strong computational performance. [Bai Shaojie, Kolter J Zico, Koltun Vladlen (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint[SPACE]arXiv: 1803. 01271*. Accessed 2 Feb 2025].



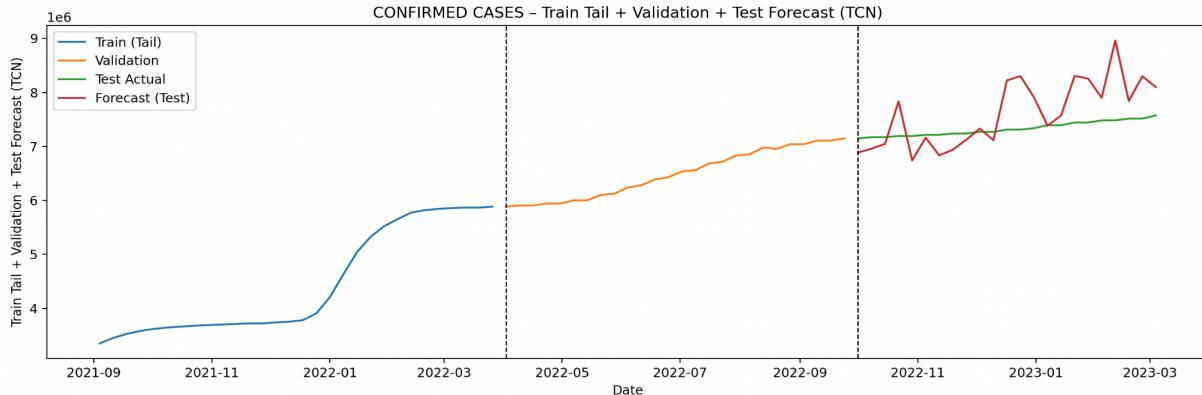
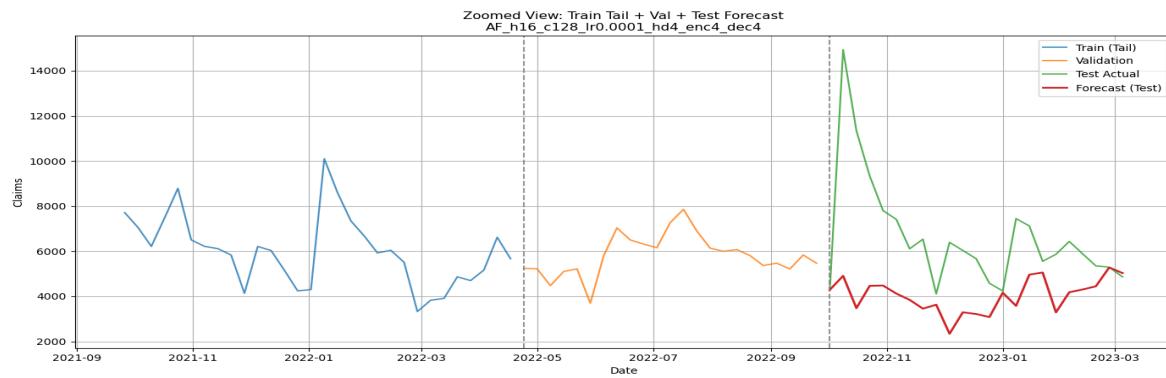


Figure 6: TCN Performance

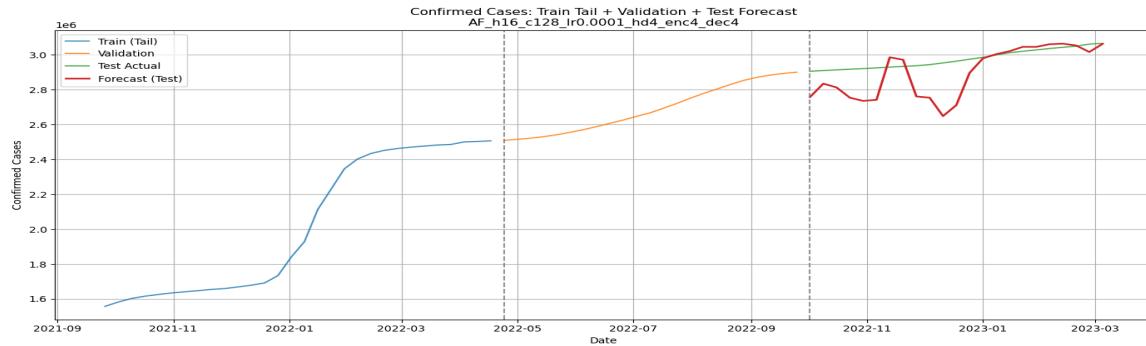
#### 4.6 Autoformer Definition

Autoformer still follows residual and encoder-decoder structure but renovates Transformer into a decomposition forecasting architecture. By embedding our proposed decomposition blocks as the inner operators, Autoformer can progressively separate the long-term trend information from predicted hidden variables.

##### 4.6.1 Autoformer Results - Univariate



CLAIMS



## CONFIRMED CASES



**DEATHS**  
*Figure 7: Autoformer Results*

### 4.6.2 Autoformer Analysis

In this experiment, we uncover key insights about Autoformer and TCN models rooted in the structure of deep learning. Our analysis focuses on weekly unemployment claims alongside weekly COVID-19 confirmed cases and deaths. Intuitively, it is reasonable to expect short-term lag relationships, such as one- or two-week delays, between surges in COVID-19 cases and increases in unemployment claims. Since the responsiveness of the job market to the progression of the pandemic is the core phenomenon under investigation, we begin with the assumption that a meaningful relationship exists between these variables.

If we examine the data closely, our study spans the core period of the pandemic, from April 12, 2020 to March 9, 2023, covering approximately 152 weeks of observations. While 152 non-null weekly data points may initially seem like a sufficiently large dataset, deep learning models such as Autoformer and TCN are prone to overfitting when the sample size is relatively limited. This underscores the importance of careful model selection, tuning, and validation to ensure reliable forecasting performance.

In our Autoformer experiment, we split the dataset into a 70% training set, 25% validation set, and 25% test set. The model is first trained using the training portion, and hyperparameters are tuned through a grid search on the validation set. Once the optimal configuration is selected, we evaluate the final model on the test set. This approach provides a more robust and unbiased assessment of the model's forecasting performance.

One limitation we encountered is that the NeuralForecast implementation no longer supports historical exogenous variables for either the Autoformer or Autoformer-Auto models. This introduced some challenges during experimentation: if we wanted to incorporate additional lagged features, we had to manually identify the most relevant lags ahead of time and engineer them into the dataset before fitting the model. As a result, adding meaningful exogenous signals required more preprocessing and gave us less flexibility during training.

### 4.6.3 Hyperparameter Tuning (Autoformer)

We performed hyperparameter tuning for the Autoformer model using a grid search over a constrained configuration space due to computational limits and the modest size of the dataset. Specifically, we explored hidden sizes of 16 and 64, convolutional hidden sizes of 32 and 128, learning rates of 1e-3 and 1e-4, numbers of attention heads of 2 and 4, and both 2- and 4-layer encoder and decoder depths. Running these experiments on Google Colab with a T4 GPU required approximately one hour to identify the top-performing configurations. The three best models are listed below, ranked by their mean absolute error (MAE).

Rank	hidden_size	conv_hidden_size	lr	n_head	enc_layers	dec_layers	MSPE	MAE	MAPE	PM
1	16	128	0.0001	4	4	4	1,827,463.51	1,087.58	0.176,7	2.259,0
2	16	128	0.0001	2	4	4	1,835,138.48	1,094.82	0.178,0	2.268,5
3	64	128	0.0001	2	2	2	2,176,483.54	1,216.30	0.198,7	2.690,4

## 5. Validation Strategy

### 5.1 Validation Strategy Explanation

We employed different but methodology justified splits for each approach:

- Traditional Methods (ARIMAX, Markov): 52 weeks for training (phases 2-4: July 2020 - June 2021), 16 weeks for testing (phase 7: November 2022 - March 2023, and no validation set)
- Deep Learning (TCN, Autoformer): 105 weeks (70%) for training, 23 weeks (15%) for validation, and 23 weeks (15%) for testing

Rationale: Neural networks require validation for hyperparameter tuning, early stopping, and model selection, following standard machine learning practice.

Comparison Limitation: Different test periods (16 vs. 23 weeks, potentially different regimes) complicate direct method comparison. Performance differences may reflect method superiority, test difficulty, or both.

### 5.2 Combined Results

We used several error metrics to assess forecasting performance, including MSPE, MAE, and MAPE. Among these, MAE was selected as our primary evaluation metric because it is easy to interpret, scale-consistent within the target variable, and penalizes all absolute errors equally. Unlike MSPE, which disproportionately emphasizes large errors due to squaring, MAE provides a more balanced view of model accuracy.

*Table 5: Performance Summary*

Model	State	Target	MAE	RMSE	MSPE	MAPE	R <sup>2</sup>	Dir. Acc.	Notes
ARIMAX	FL	Claims	1,247.10	1,556.25		24.26	-1.791	60.0%	ARIMA (3,1,3) with no seasonality
ARIMAX	GA	Claims	3,212.73	4,116.91		40.28	-1.258	46.7%	ARIMA (3,1,3) with no seasonality
ARIMAX	NY	Claims	5,216.47	8,145.35		21.09	-0.633	46.7%	ARIMA (3,1,3) with no seasonality
TCN	FL	Claims	2,369		1.011,543 × 10 <sup>7</sup>	32.51			Deep learning; univariate
TCN	FL	Cases	1,517.877		4,353,686,503,424 × 10 <sup>12</sup>	20.41			
TCN	FL	Deaths	7,716		9,731,700,8 × 10 <sup>7</sup>	9.13			
Autoformer	FL	Claims	2,611		1.259,968,089,961,6 × 10 <sup>7</sup>	33.74			Deep learning; univariate
Autoformer	FL	Cases	296,941		2,119,372,051,739,891 × 10 <sup>11</sup>	4.09			
Autoformer	FL	Deaths	2,737		1,492,071,754,033,893,2 × 10 <sup>7</sup>	3.30			

**Note:** Some performance values differ from those reported in preliminary drafts due to finalizing model configurations and rerunning all evaluations using a consistent data split and feature set. The results shown in this table represent the final and authoritative metrics from our study.

### 5.3 Key Observations

#### *Unemployment Claims (FL)*

ARIMAX achieves lowest MAE (1,247) despite negative R<sup>2</sup> (-1.791). This counterintuitive result demonstrates that MAE and R<sup>2</sup> measure different aspects of performance. Low absolute error does not imply variance explanation. Models can outperform the alternatives while still failing fundamentally.

Deep Learning methods show 90-110% higher MAE (2,369-2,611) compared to ARIMAX model, though this may partly reflect different test periods (endemic vs. mixed), proving that regime shift defeats pattern recognition regardless of method sophistication.

#### *COVID-19 Outcomes (FL)*

- Autoformer outperforms TCN (80% lower MAE for cases, 65% for deaths)
- Deaths more predictable than cases (3.30% vs. 4.09% MAPE)

#### *Forecasting Difficulty by Target*

- Deaths: 3-9% MAPE
- Cases: 4-20% MAPE
- Claims: 24-34% MAPE

Economic behavior is substantially harder to forecast than epidemiological outcomes (deaths), with claims showing 7-8x higher error rates regardless of method choice, likely reflecting:

- Policy sensitivity (stimulus, UI benefit changes)
- Behavioral complexity (individual filing decisions)
- Regime shift severity
- Data quality (administrative vs. clinical reporting)
- Target variable characteristics dominate method selection. No approach overcomes fundamental data limitation.

## 6. Policy Implications

Traditional econometric models should not forecast blindly across structural breaks; instead, they must monitor for regime-change indicators such as policy shifts and evolving infection trends. Multi-state and cross-state models often fail during unprecedented transitions because spillover patterns and endemic dynamics are highly idiosyncratic. Regime-switching approaches work when regimes repeat (e.g., seasonal cycles) but are not a universal solution, and novel transitions like pandemic-to-endemic shifts expose their limitations. For future pandemics, we must accept inherent unpredictability during regime changes.

## 7. Conclusions

From these experiments, we can extrapolate that while, in a controlled environment, ML-based models such as Autoformers or TCNs perform better, we must be mindful of what the data actually requires, rather than blindly applying complex ML models to small datasets.

- **Success**

From our limited trials, we were able to observe how the current models and modules function, and how we might use them moving forward where their application is warranted. For example, we learned that the current NeuralForecast implementation of Autoformer models does not allow historical time-series exogenous variables to be used directly in prediction. However, we also found that there are ad-hoc methods to include historical exogenous variables by first identifying important lags using standard non-ML models like VAR, then shifting those variables and attaching them to the Autoformer as future exogenous inputs. But this approach also exposes us to the same challenge in our dataset: because we have a relatively small number of time points, using larger lag values, even if supported by the VAR model, may negatively affect performance.

- **Limitations**

One major limitation we faced, especially with the ML models, was access to computational resources, specifically GPU availability on Google Colab. Conducting a proper grid search across a wide range of hyperparameters requires significant time and consistent GPU access, which we did not have. As a workaround, we limited the scope of the grid and the number of hyperparameter values included, reducing the total permutations. While this approach allowed us to obtain interpretable results, we must acknowledge that any conclusions drawn are constrained by the fact that we were unable to thoroughly explore the solution space. With more computing time, we might have identified better-performing models.

## Appendix

This is the link to the git repository that we used for the code production. [GitHub - Stephzennn/COVID-19-pandemic-Timeseries-analysis](https://github.com/Stephzennn/COVID-19-pandemic-Timeseries-analysis)

### Required Appendix: Team Contributions

- Ahmed Mounir Elsemary
  - TCN Code
  - TCN Explanation
  - Literature Review
  - Prior Expectation
- Trang Ly
  - ARIMAX Code
  - ARIMAX assessment
  - Initial Data Exploration
  - ARIMAX results report
- Estifanos Zenebe Gebresellassie
  - Autoformer Code
  - Conclusion
  - Autoformer Analysis Report
  - Summary

### References

Bai, S., Kolter, J. Z., & Koltun, V. (2018). *An empirical evaluation of generic convolutional and recurrent networks for sequence modeling*. arXiv preprint arXiv:1803.01271.  
<https://doi.org/10.48550/arXiv.1803.01271>

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Centers for Disease Control and Prevention. (2023). *Weekly United States COVID-19 cases and deaths by state* [Data set]. U.S. Department of Health and Human Services.  
<https://data.cdc.gov/Case-Surveillance/Weekly-United-States-COVID-19-Cases-and-Deaths-by-/pwn4-m3yp>

Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. (2021). *COVID-19 data* (Version v1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5167597>

Cramer, E. Y., Ray, E. L., López, V. K., Bracher, J., Brennen, A., Castro Rivadeneira, A. J., ... Reich, N. G. (2022). *Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States*. *Proceedings of the National Academy of Sciences*, 119(15), e2113561119. <https://doi.org/10.1073/pnas.2113561119>

Hamilton, J. D. (1989). *A new approach to the economic analysis of nonstationary time series and the business cycle*. *Econometrica*, 57(2), 357–384. <https://doi.org/10.2307/1912559>

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>

Li, Y., Wang, Y., & Ma, K. (2022). *Integrating Transformer and GCN for COVID-19 forecasting*. *Sustainability*, 14(16), 10393. <https://doi.org/10.3390/su141610393>

Liu, Z., Zhang, Y., Zhang, C., & Xu, X. (2024). *Deep time series forecasting models: A comprehensive survey*. *Mathematics*, 12(10), 1504. <https://doi.org/10.3390/math12101504>

Maigur, A. A. (2024). *Machine learning algorithms for predicting unemployment duration in Russia*. *Russian Journal of Economics*, 10(4), 365–384. <https://doi.org/10.32609/j.ruje.10.128611>

Qiu, X., Hu, J., Zhou, L., Wu, X., Du, J., Zhang, B., Guo, C., Zhou, A., Jensen, C. S., Sheng, Z., & Yang, B. (2023). *TFB: Towards comprehensive and fair benchmarking of time series forecasting methods*. arXiv preprint arXiv:2306.03009. <https://doi.org/10.48550/arXiv.2306.03009>

Sims, C. A. (1980). *Macroeconomics and reality*. *Econometrica*, 48(1), 1–48. <https://doi.org/10.2307/1912017>

U.S. Employment and Training Administration. (n.d.). *Initial claims (ICSA)* [Data series]. Federal Reserve Bank of St. Louis (FRED). <https://fred.stlouisfed.org/series/ICSA>

Wu, H., Xu, J., Wang, J., & Long, M. (2021). *Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting*. arXiv preprint arXiv:2106.13008. <https://doi.org/10.48550/arXiv.2106.13008>