House Price Index Analysis - Affordability and Expectation

Elisa Carucci - Yuelin Shen - Zhuoyi Jiang - Princeton Huang
Columbia University (Prof. Ali Hirsa; Mr. Josh C. Panknin) & JP Morgan

Abstract

We seek to deepen our understanding of **Housing Price Index (HPI)** dynamics in various urban contexts, focusing on four cities: Dallas, Denver, Miami, and San Diego. Exploring influential microeconomic criteria, industry-level data, building permits, and categorical variables representing regimes (R2-RD technique) allows us to build a comprehensive analytical model.

We leverage previously developed frameworks by using a **1-Layer Gated Recurrent Unit (GRU) with random search** to predict HPI. We identify the importance of integrating lagging and regime-switching mechanisms developed with the HMM model into the GRU, so as to enhance its predictive accuracy for HPI, while also observing performance variability between cities.

Objective

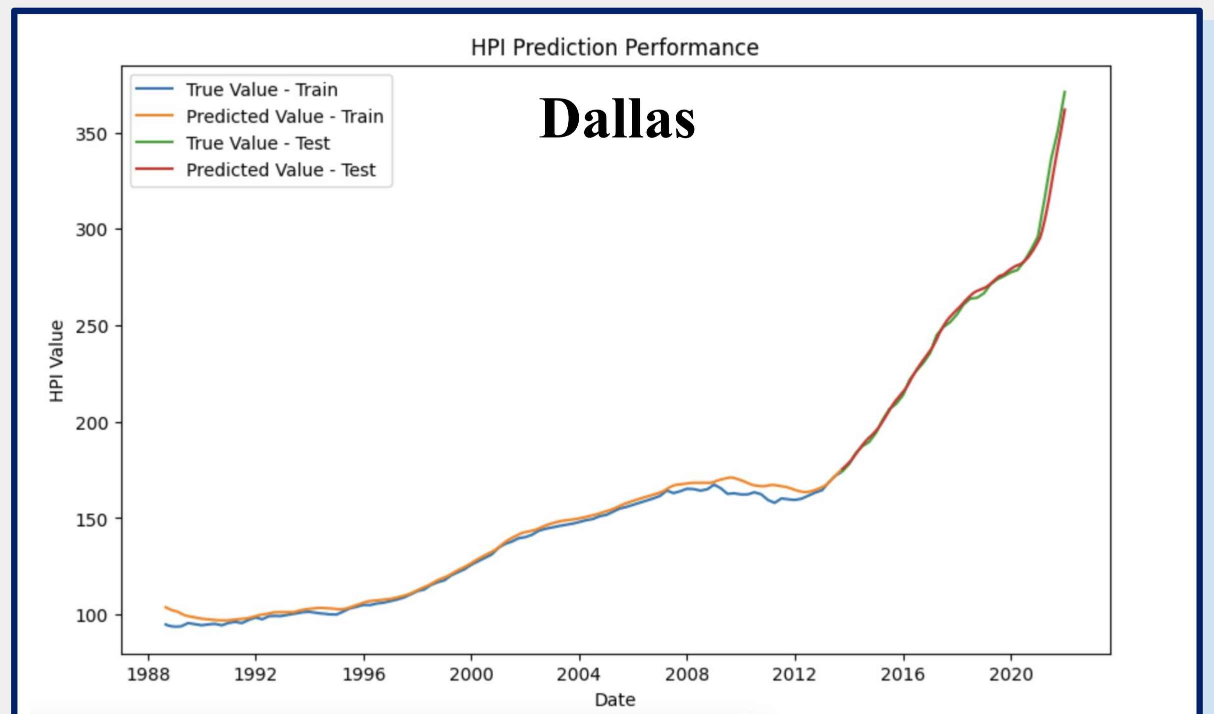
We endeavor to gain insights surrounding the factors which influence housing markets in urban settings:

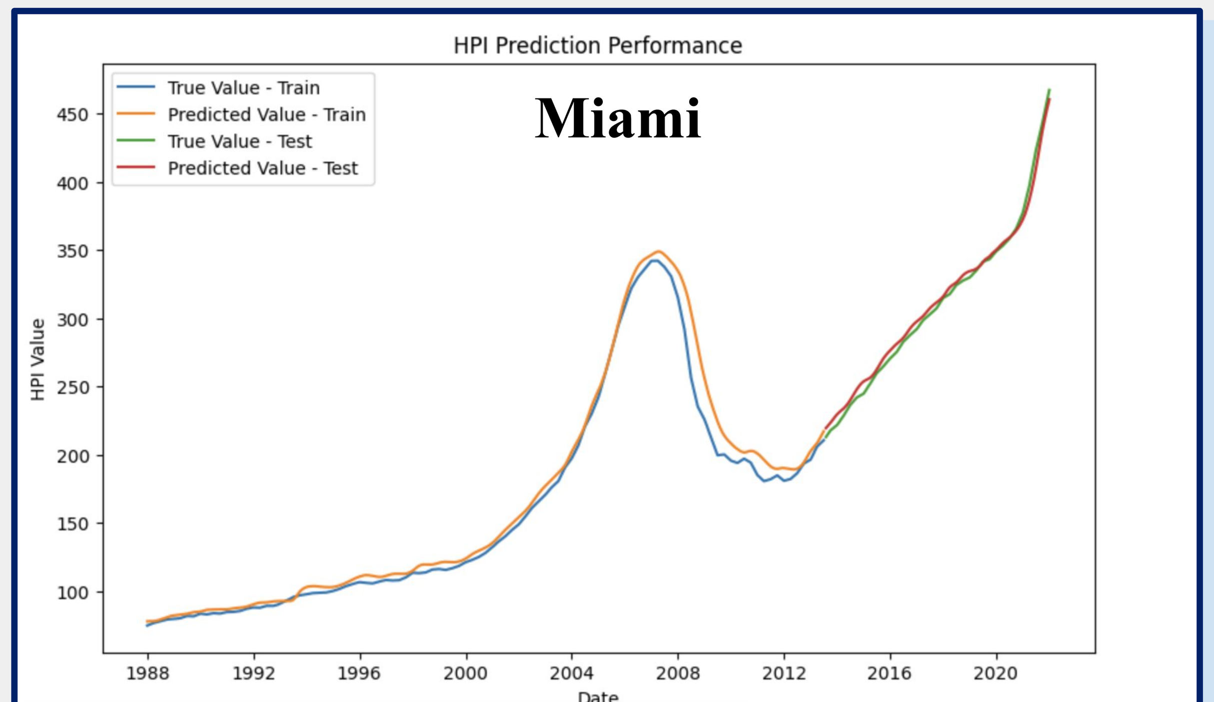
- Develop a **more accurate model** to actively forecast HPI within and across cities.
- Derive **powerful predictive techniques** within the context of real estate modeling.
- Evaluate **related industry trends** in the automotive and luxury goods spheres.

ARIMA Backtest

We employ the AutoRegressive Integrated Moving Average (ARIMA) model to backcast data points for city-level HPI, thereby augmenting our dataset.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1}$$



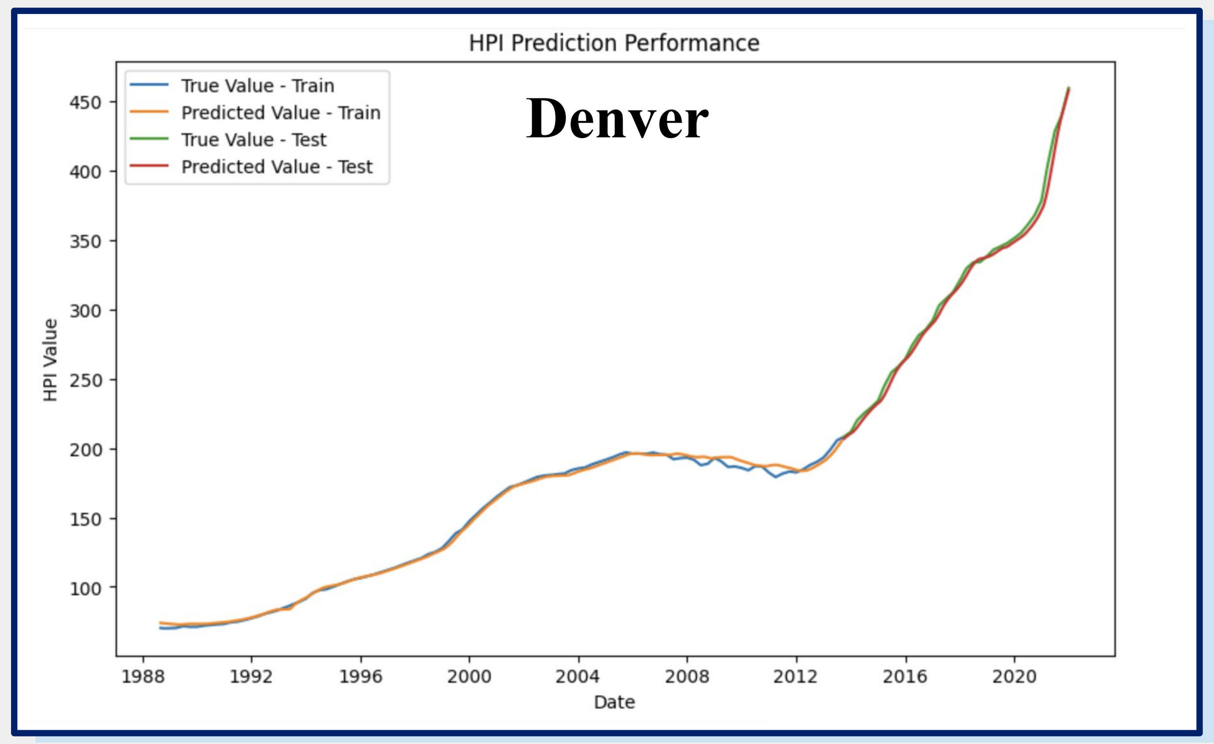


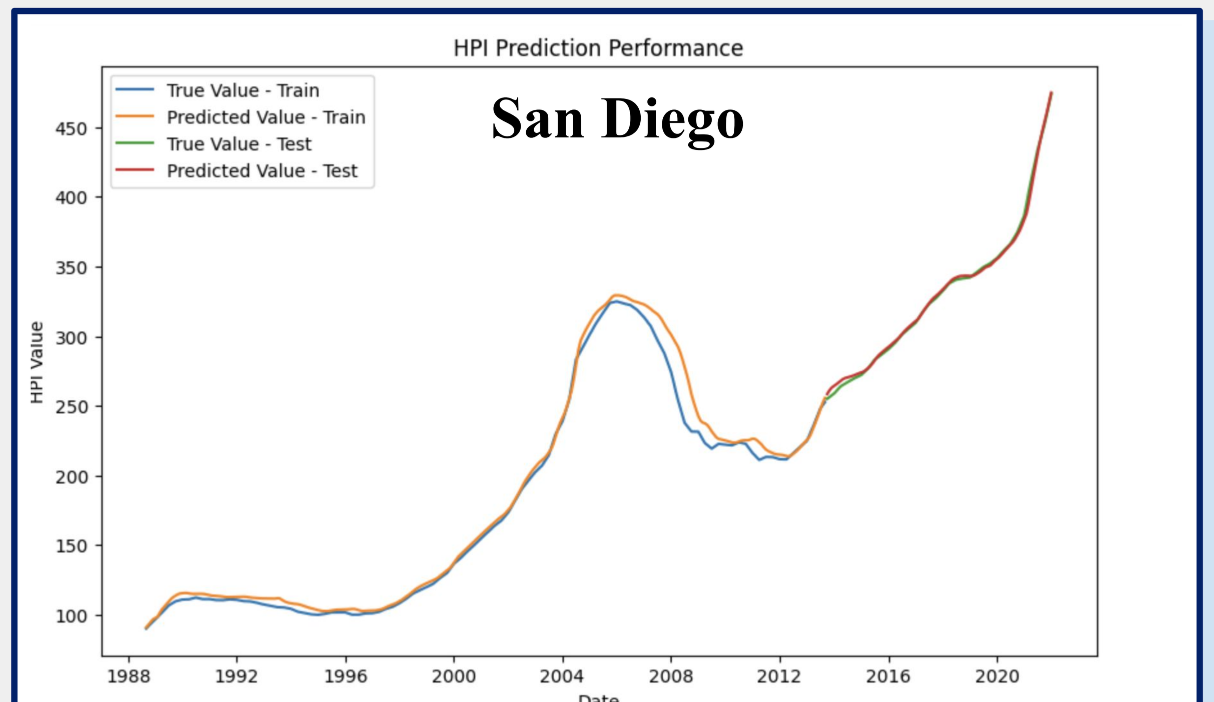
Lagging Detection

The Granger Causality Test is used to determine whether one time series can predict another, enhancing accuracy.

Unrestricted: $Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j X_{t-j} + \epsilon_t$

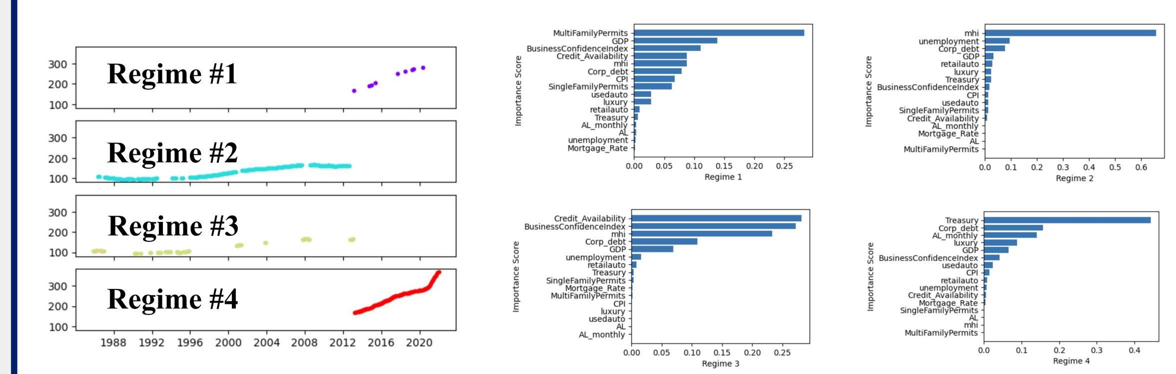
Restricted: $Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \eta_t$



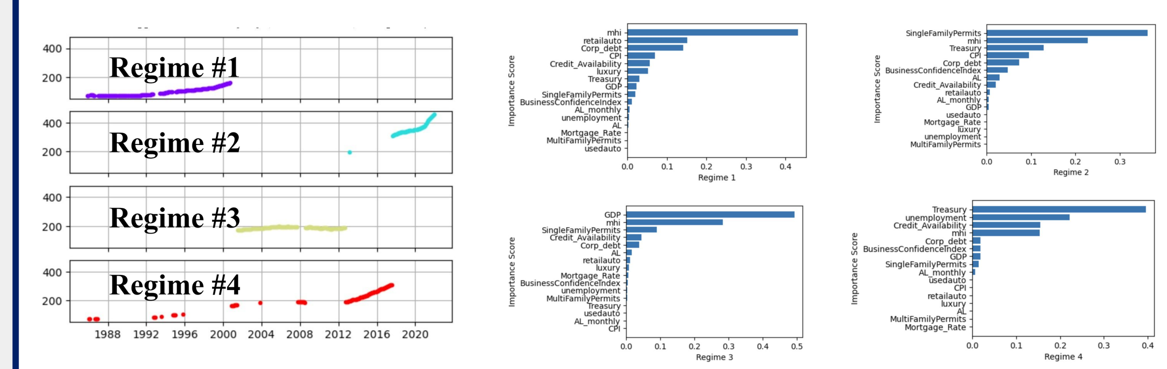


Regimes with HMM

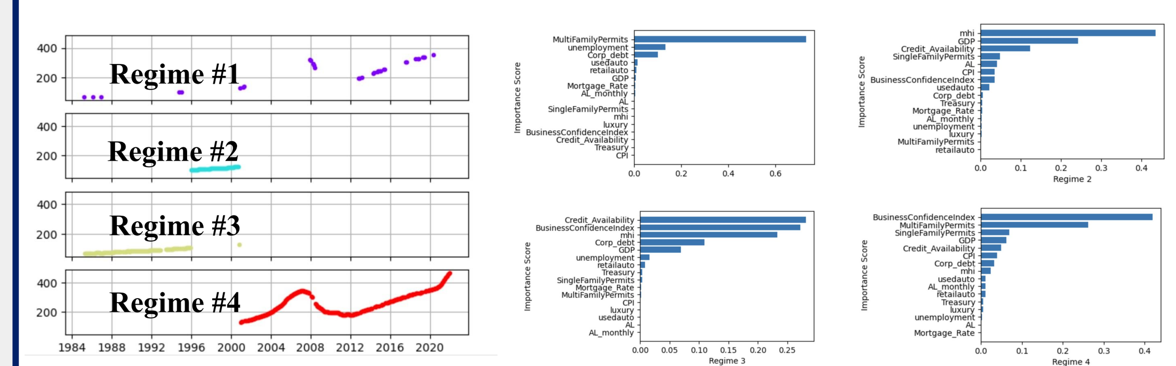
Dallas



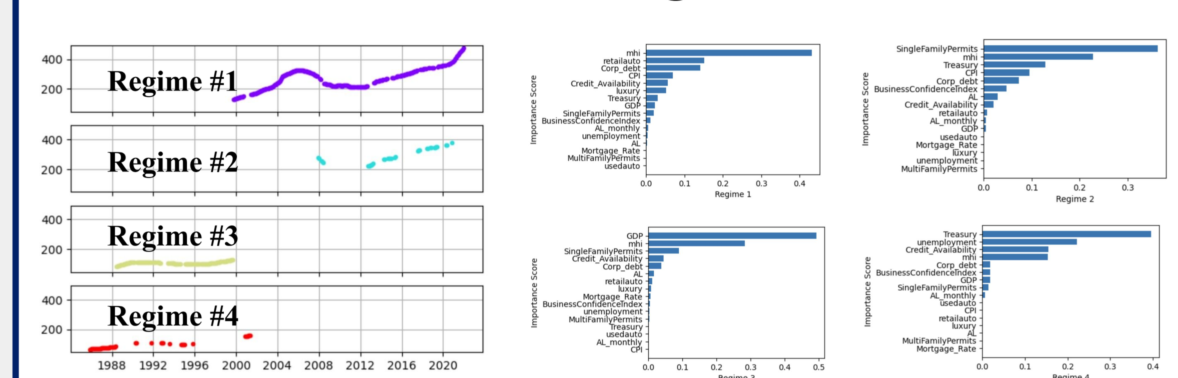
Denver



Miami



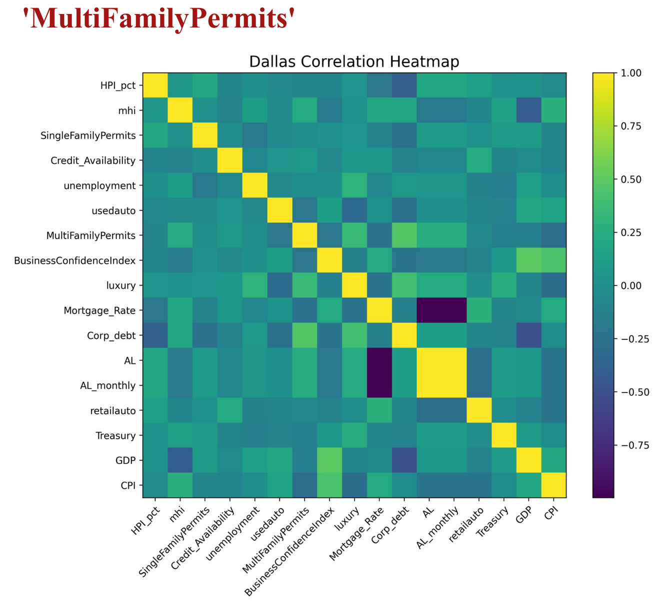
San Diego



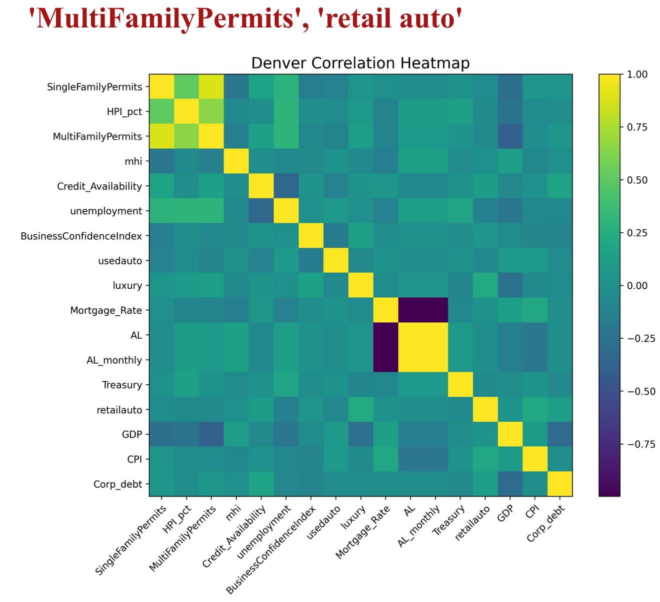
Model Evaluation

Dallas Train RMSE: 3.1891 Test RMSE: 4.0339 Test R^2: 0.9927	**Denver** Train RMSE: 2.4122 Test RMSE: 5.0714 Test R^2: 0.9937
Miami Train RMSE: 10.2503 Test RMSE: 5.3365 Test R^2: 0.9923	**San Diego** Train RMSE: 8.4900 Test RMSE: 2.4027 Test R^2: 0.9979

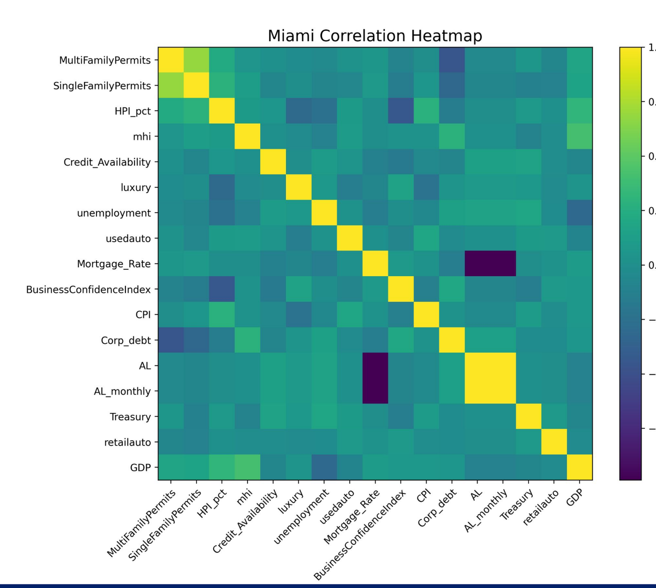
Predictors selected for Dallas:
'BusinessConfidenceIndex','Credit_Availability', 'GDP', 'Corp_debt', 'Treasury', 'mhi', 'HPI', 'MultiFamilyPermits'



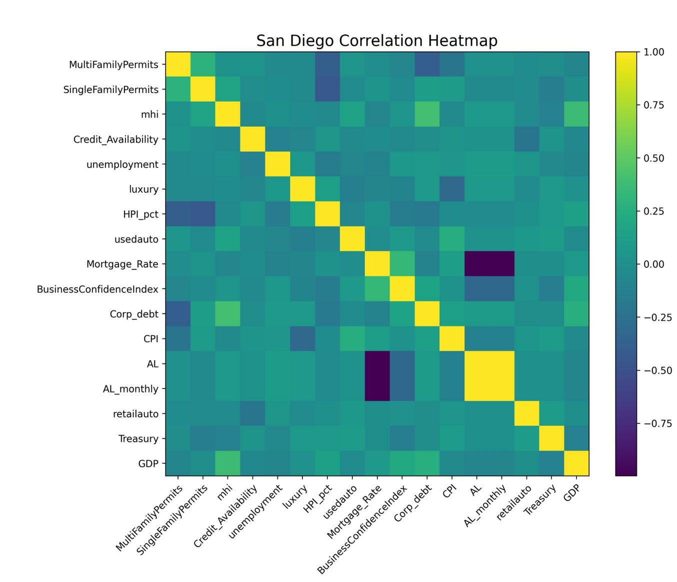
Predictors selected for Denver:
'SingleFamilyPermits', 'Treasury', 'CPI', 'unemployment', 'mhi', 'HPI', 'MultiFamilyPermits', 'retail auto'



Predictors selected for Miami:
'BusinessConfidenceIndex', 'GDP', 'Corp_debt', 'unemployment', 'mhi', 'HPI', 'MultiFamilyPermits'



Predictors selected for San Diego:
'SingleFamilyPermits', 'CPI', 'GDP', 'mhi', 'HPI', 'Credit_Availability'



Future Work

- Increase **data frequency** via advanced techniques.
- Explore **more complex deep learning models** beyond the GRU model that we set forth.
- Test **additional regions** with different features.
- Identify **stronger industry-level** data that can help with predictive modeling.
- Assess different methodologies for employing **regime-switching models.**