

Cse150

Final Project

11/12/2025

Team 102 - Watermelon

Dennis Sun

Nick Pham

Noe Gonzalez

Wenrong Zheng

## Final Project Plan

### **Problem Description:**

Research Question: Predicting the housing market regime from FHFA House Price Index.

Description: The problem we are addressing is figuring out a way to predict the housing market regime from the FHFA House Price Index. Predicting housing market regimes are important because different regimes affect housing affordability, household balance sheets, mortgage default risk, and financial stability.

- In an expansion regime, house prices keep rising with relatively high growth rates, possibly accompanied by higher leverage and bubble risk;
- In a correction regime, price growth slows or becomes negative, which may lead to default and wealth losses.
- A neutral regime corresponds to relatively mild growth and risk.

In the real world, it is difficult to determine the current housing market regime. There is no direct label such as “this quarter is a boom” or “this quarter is a correction”. We can only observe the price index itself and determine the underlying regime from that.

This makes the problem very suitable for probabilistic reasoning and learning because of the inherent uncertainty and latent structure of the data. The true market regime is not directly observable and must be inferred from the observed indicator such as the House Price Index.

Therefore, this project directly applies probabilistic modeling, latent-variable learning, and sequence inference methods from the course to a real economic and financial data setting.

**Dataset Source:**

Source: FHFA House Price Index

<https://www.fhfa.gov/data/hpi/datasets>

The FHFA is a weighted repeat sales index covering all 50 U.S. states and more than 400 cities, with data dating back to the mid-1970s, and is available on a monthly and quarterly basis.

Primary CSV: U.S. Summary.xlsx

<https://www.fhfa.gov/data/hpi/datasets?tab=quarterly-data>

**Methodology:**

1. Variables & Pre-processing:

Select scope, national series, or state-level panel; use seasonally adjusted, purchase series for baseline. Compute log returns as emissions and standardize by rolling mean if necessary. Handle structure breaks or outlines, and align the calendar.

Train split along time (e.g., pre-2024), validation for likelihood and stability checks(2024-2025 YTD).

2. Model:

We use a Hidden Markov Model (HMM) to describe the hidden state sequence  $S_1, S_2 \dots S_T$ : the housing market regime (Expansion / Neutral / Correction), and the observation sequence  $O_1, O_2 \dots O_T$ : “Up / Stable / Down” obtained by discretizing changes in the house price index.

3. Evaluation:

Two complementary evaluation dimensions are used: Quantitatives, out-of-sample log-likelihood; BIC/AIC; regime persistence metrics; turning-point dating stability. Qualitative, do decoded regimes align with macro narratives? Like tightening cycles. Discuss limits and failure modes, per the rubric emphasis on analysis or interpretation over raw metrics.